MASS SPECTROMETRY-BASED CHARACTERIZATION OF PROTEOMIC AND GLYCOSYLATION ALTERATIONS IN LUNG CANCER TISSUES AND EXTRACELLULAR VESICLES

PhD thesis

Mirjam Balbisi

Semmelweis University Doctoral College

Pharmaceutical Sciences and Health Technologies Division

Supervisor: Lilla Turiák, PharmD, Ph.D

Official reviewers: László Márk, D.Sc
Olivér Ozohanics, Ph.D

Head of the Complex Examination Committee:
Zoltán Benyó, MD, D.Sc

Members of the Complex Examination Committee:
Krisztina Ludányi, Ph.D
Gitta Schlosser, Ph.D

Budapest
2026

1. Introduction

Proteomics aims to characterize the proteome, the complete set of proteins expressed by a genome. A major source of proteome complexity arises from post-translational modifications (PTMs), which expand protein functionality beyond the genetic code. Among analytical techniques, nano-ultra-high-performance liquid chromatography coupled to tandem mass spectrometry (nano-UHPLC-MS/MS) has become the central tool for proteomic investigations.

Protein glycosylation is one of the most structurally diverse PTMs in mammalian systems, affecting more than half of all proteins. It plays essential roles in protein folding, stability, receptor activation, and cell-cell communication and is increasingly recognized as a key regulator of disease processes, including cancer.

N-glycosylation involves the covalent attachment of oligosaccharides to asparagine residues within a consensus sequence. All *N*-glycans share a common core structure, which can be extended to form oligomannose, hybrid, or complex *N*-glycans. Two complementary analytical strategies are commonly applied: *N*-glycomics,

which profiles released *N*-glycans at the global level, and *N*-glycoproteomics, which enables site-specific characterization of intact *N*-glycopeptides.

Another major class of glycosylated biomolecules is proteoglycans, which consist of a protein core modified with glycosaminoglycan (GAG) chains. GAGs are linear polysaccharides composed of repeating disaccharide units. Chondroitin sulfate/dermatan sulfate (CS/DS) contains glucuronic acid/iduronic acid and *N*-acetylgalactosamine residues and can be *O*-sulfated at two positions, while heparan sulfate (HS) consists of glucuronic acid/iduronic acid and *N*-acetylglucosamine, and can be *O*-sulfated at two positions and *N*-sulfated at one position. In general, enzymatic digestion is performed, followed by HPLC-MS analysis of the GAG disaccharides. The analysis of GAGs is analytically challenging due to their negative charge, structural heterogeneity, and labile sulfate groups.

Lung cancer is the leading cause of cancer-related mortality worldwide and is characterized by extensive heterogeneity. Non-small cell lung cancer, the most common form, includes multiple subtypes such as adenocarcinoma, squamous cell carcinoma, and large cell

carcinoma. In many adenocarcinoma cases, an oncogenic driver mutation is present, e.g. in ALK gene, which also determines the treatment options.

Importantly, lung tumors display marked spatial heterogeneity, where different morphological regions (e.g., tubular, papillary, and solid tumor) may coexist within the same tissue. Conventional bulk tissue analysis fails to capture this complexity. To address this limitation, spatially resolved workflows have been developed, including on-surface digestion approaches that enable localized proteomic and GAG analysis directly from tissue sections.

Extracellular vesicles (EVs) have emerged as important mediators of intercellular communication. Small EVs (sEVs) carry proteins, lipids, and nucleic acids that reflect the molecular state of their cells of origin and play key roles in cancer progression. While EV-based biomarker research is primarily focused on proteins and nucleic acids, the role of protein glycosylation in EV biology remains insufficiently explored, largely due to analytical challenges and limited sample amounts.

2. Objectives

The aim of this work was to investigate protein expression and glycosylation alterations in lung cancer using integrated HPLC-MS-based analytical approaches.

The dissertation is based on two projects. The first study focused on the comprehensive analysis of proteomic and GAG (CS/DS and HS) profiles in different regions of lung tissue sections carrying ALK rearrangement. The aim was to identify and quantify proteins and GAG disaccharides, followed by statistical and bioinformatic analyses to characterize molecular differences between tumor and non-tumor regions, as well as between tumor regions with distinct morphologies and microenvironmental features.

The second project investigated the proteomic, *N*-glycoproteomic, and CS/DS GAG profiles of sEVs derived from A549 lung adenocarcinoma and BEAS-2B non-tumorigenic epithelial cells. The aim was to develop and apply integrated analytical workflows, perform statistical and bioinformatic evaluation, and compare tumor- and non-tumor-derived sEVs to identify molecular differences, and assess the relevance of sEV glycosylation for future biomarker studies.

3. Methods

3.1. Tissue study

For the tissue study, 22 regions from 7 formalin-fixed, paraffin-embedded lung tissue sections were analyzed. The regions were characterized based on morphology, mucin and stromal content. Following deparaffinization and antigen retrieval, regions of interest were subjected to parallel on-surface digestion workflows for proteomic and GAG analyses.

In proteomics, disulfide bonds were reduced and alkylated, followed by sequential Lys-C/trypsin and trypsin digestion directly on the tissue surface, with a total digestion time of 200 min. Peptides were extracted using 10% acetic acid, purified by C_{18} solid-phase extraction, and analyzed on a Bruker Maxis II Q-TOF coupled to a Dionex Ultimate 3000 nano-UHPLC system using data-dependent acquisition. Protein identification was performed using Byonic against the Swiss-Prot human database, followed by label-free quantification in MaxQuant. Statistical analysis and data visualization were carried out in R, while functional interpretation of altered proteins was supported by STRING network analysis.

For GAG analysis, CS/DS and HS chains were digested directly on tissue using chondroitinase-ABC and a combination of heparin lyases, respectively. CS/DS digestion was performed over 20 h, while HS digestion lasted 48 h. The resulting disaccharides were extracted with 1% ammonia solution and purified using a two-step workflow combining cotton and porous graphitic carbon solid-phase extraction. Measurements were carried out on a Waters Select Series cyclic ion mobility mass spectrometer coupled to a Waters Acquity I-class UHPLC system using hydrophilic interaction chromatography and weak anion exchange-based separation. Quantification and peak integration were performed using MassLynx, followed by visualization in R.

3.2. Extracellular vesicle study

In the sEV study, sEVs were isolated from A549 lung adenocarcinoma and BEAS-2B epithelial cell cultures grown in F12 and BEGM media, respectively. After culturing under serum-free conditions, supernatants were subjected to centrifugation and filtration to remove cell debris, apoptotic bodies, and microvesicles, followed by size exclusion chromatography for sEV isolation. Small

EVs were characterized by microfluidic resistive pulse sensing (MRPS) and transmission electron microscopy (TEM). The solvent was exchanged to ammonium bicarbonate solution, and sEVs were lysed by repeated freeze-thaw cycles. For the next steps, 6 replicates from both sEV types and 3 control samples from both culture media were processed.

Proteomic analysis involved reduction, alkylation, and overnight enzymatic digestion using Lys-C/trypsin and trypsin. *N*-glycopeptides were enriched by acetone precipitation, where the pellet was used for *N*-glycoproteomic analysis and the supernatant for proteomics after C_{18} purification. Peptides were analyzed by a Bruker timsTOF HT instrument coupled to a Dionex Ultimate 3000 nano-UHPLC using data-independent acquisition. Protein identification and quantification were performed using DIA-NN, followed by statistical analysis in R and gene set enrichment analysis for biological processes.

N-glycopeptides were analyzed on an Orbitrap Exploris 240 MS coupled to a Waters nanoAcquity system. In parallel, released *N*-glycans were analyzed by capillary

electrophoresis-mass spectrometry using a Sciex CESI 8000 Plus system coupled to a Bruker Q-TOF instrument to generate an *N*-glycan database. *N*-glycopeptide identification and quantification were performed using GlycReSoft, followed by statistical evaluation in R.

CS/DS GAG analysis of sEVs was carried out using chondroitinase-ABC digestion, followed by the same purification, UHPLC-MS measurement, and data processing workflow as applied for tissue samples. Statistical analysis and visualization were performed in R.

.

4. Results

4.1. Tissue study

Proteomic and GAG profiles were characterized in ALK rearranged lung tissue regions, including tumor areas of different morphologies and adjacent normal lung (ANL). A total of 2092 proteins were quantitatively analyzed. The number of differentially expressed proteins ranged from 250 to 541 in comparisons between ANL and tumor types, while fewer changes (40-199 proteins) were observed between tumor morphologies, with the largest difference detected between papillary and solid regions. Two proteoglycan core proteins, agrin and perlecan, showed reduced abundance in all tumor regions compared to ANL. Functional analysis indicated that differences between tumor morphologies were mainly associated with protein synthesis-related processes, including RNA splicing and translation. In contrast, comparisons between tumor and ANL revealed broader biological alterations, involving extracellular matrix organization, mitochondrial proteins, protein folding, and protein synthesis.

Multivariate analyses demonstrated a clear separation between ANL and tumor samples, while tumor subtypes

showed partial overlap. Solid tumor regions displayed relatively homogeneous proteomic profiles, and the mucin content, unlike the stromal content, contributed to the formation of subgroups.

Tumor regions showed a marked increase in total GAG abundance, with CS/DS levels rising up to 4.4-fold and HS levels up to 8.5-fold compared to ANL. In parallel, the average rate of CS/DS and HS sulfation increased by 4.1-4.7 and 2.1-2.7 in all tumor types. In the case of HS, a considerable decrease was observed in *N*-sulfation in tumor regions. Among tumor subtypes, the most pronounced differences were observed in total HS amounts.

Global GAG profiling clearly distinguished ANL from tumor regions, although tumor subtypes largely overlapped. Just as with the proteomic data, mucin content strongly influenced clustering, while stromal content had minimal impact.

4.2. Extracellular vesicle study

Small EVs derived from A549 lung adenocarcinoma and BEAS-2B non-tumorigenic epithelial cells were analyzed using integrated proteomic, *N*-glycoproteomic, and GAG

approaches. EVs were characterized using TEM and MRPS, confirming the presence of predominantly spherical particles below 200 nm. BEAS-2B sEVs displayed a broader size distribution and lower particle concentration compared to A549 sEVs.

Proteomic analysis resulted in the quantification of 2334 proteins, of which 945 were retained for statistical evaluation after filtering. The presence of 77-91 proteins from the Exocarta reference list supported the high quality of the isolates. A total of 408 proteins showed differential abundance between the two sEV types, including 313 upregulated proteins and 95 downregulated ones. Among proteoglycan core proteins, distinct expression patterns were observed: versican and testican-1 were elevated, while CSPG4, aggrecan, and syndecan-4 were reduced in A549 sEVs. Functional enrichment analysis indicated increased representation of pathways related to cell cycle regulation, nucleic acid metabolism, and protein synthesis, while immune-related processes were downregulated.

N-glycomic profiling identified substantially more *N*-glycan compositions in sEVs than in control media, confirming their vesicular origin. 1659 *N*-glycopeptides

were quantified, of which 227 *N*-glycoforms were retained after filtering, and 152 differed significantly. Proteins such as versican, laminin subunits, and galectin-3-binding protein showed extensive site-specific *N*-glycosylation changes. Normalization to protein abundance indicated that these alterations were not solely driven by protein-level differences but also reflected modifications in *N*-glycan abundance. When summarized at the level of general *N*-glycan features, a shift toward increased complex-type and decreased oligomannose-type *N*-glycans was observed in A549 sEVs.

CS/DS analysis revealed a 3.4-fold higher total disaccharide content in A549 sEVs, accompanied by a significantly increased 4S sulfation and reduced 6S sulfation. These changes suggest remodeling of GAG structure in A549 sEVs and may indicate the presence of cancer-associated glycan motifs, such as oncofetal CS.

Principal component analysis and hierarchical clustering demonstrated a clear and complete separation between the two sEV populations across all three molecular profiles, indicating that not only proteomic but also glycosylation features reflect tumor-associated differences.

5. Conclusions

This thesis provides a comprehensive characterization of lung cancer-associated molecular alterations by integrating proteomic and glycosylation-focused analytical approaches across tissue and sEV systems.

In ALK rearranged lung tissue, both protein expression and GAG composition were markedly altered in tumor regions compared to ANL. While proteomic differences between tumor morphologies were primarily associated with protein synthesis-related processes, GAG analysis revealed substantial increases in total CS/DS and HS amounts. Importantly, these alterations were largely independent of proteoglycan core protein levels, indicating that differences arise mainly from modifications in glycan biosynthesis rather than protein abundance. The strong association between mucin content and GAG profiles further highlights the influence of the local microenvironment on glycosylation.

In the sEV study, A549 and BEAS-2B sEVs showed distinct molecular signatures across proteomic, *N*-glycoproteomic, and GAG data. Significant alteration of the protein cargo was accompanied by widespread,

site-specific changes in *N*-glycosylation, affecting key extracellular matrix proteins such as versican, laminins, and galectin-3-binding protein. In parallel, CS/DS analysis revealed increased total GAG content and a shift toward enhanced 4S sulfation.

Across both models, glycosylation changes consistently complemented proteomic alterations and provided additional discriminatory power. Overall, this work highlights the importance of integrating proteomic and glycosylation approaches to achieve a more comprehensive understanding of cancer biology and supports the potential of glycosylation-based features, particularly in EVs, for future biomarker development.

6. Bibliography of the candidate's publications

6.1. Publications related to the dissertation

1. M. Balbisi, T. Langó, V.N. Horváth, D. Pál, G. Schlosser, G. Kecskeméti, Z. Szabó, K. Ilyés, N. Nagy, O. Tóth, J. Zheng, G.S.M. Lageveen-Kammeijer, T. Visnovitz, Z. Varga, B.G. Vértessy, L. Turiák: Distinct proteomic and glycosylation signatures differentiate A549 tumor and BEAS-2B non-tumor cell line derived small extracellular vesicles. *Molecular & Cellular Proteomics* 2026, 25(3): 101524.

IF: 5.5; D1

2. M. Balbisi, S. Sugár, G. Schlosser, B. Szeitz, J. Fillinger, J. Moldvay, L. Drahos, A.M. Szász, G. Tóth, L. Turiák: Inter- and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. *Scientific Reports* 2023, 13: 6268.

IF: 3.8; D1

6.2. Review publication related to the dissertation

3. M. Balbisi, S. Sugár, L. Turiák: Protein glycosylation in lung cancer from a mass spectrometry perspective. *Mass Spectrometry Reviews* 2026, 45(3): 455-475.

IF: 6.6; D1

6.3. Publications unrelated to the dissertation

4. D. Lenzinger, L. Lankovics, I. Dudás, T. Bárkai, Zs. Szász, M. Balbisi, A. Y. Wu, K. V. Vukman, K. Fletcher, A. Csomos, Z. Mucsi, E. Bugyik, Cs. Cserép, Á. Dénes, Sz. Bősze, L. Turiák, C. P. Lai, E. I. Buzás, T. Visnovitz: Stress-induced switch in small extracellular vesicle secretion: from constitutive "torn bag mechanism" to exocytosis. *Journal of Extracellular Vesicles* 2026, in press.

   IF: 21.7; D1

5. D. Virág, K. M. Karvaly, A. Molnár, M. Balbisi, J. Németh, L. Turiák: Mass spectrometry insights into post-translational modifications in extracellular vesicles. *Mass Spectrometry Reviews* 2026, in press.

   IF: 6.6; D1

6. F. Bugyi, M. Balbisi, S. Sugár, L. Váncza, E. Regős, I. Kovalszky, I. Laczó, T. Harkó, G. Kecskeméti, Z. Szabó, J. Moldvay, L. Drahos, L. Turiák: Unveiling unique protein and phosphorylation signatures in lung adenocarcinomas with and without ALK, EGFR, and

KRAS genetic alterations. *Molecular Oncology* 2025, 19: 3243-3265.

IF: 4.5; D1

7. B. György, R. Szatmári, T. Ditrói, F. Torma, K. Pálóczi, M. Balbisi, T. Visnovitz, E. Koltai, P. Nagy, E.I. Buzás, S. Horvath, Zs. Radak: The protein cargo of extracellular vesicles correlates with the epigenetic aging clock of exercise sensitive DNAmFitAge. *Biogerontology* 2025, 26: 35.

   IF: 4.1; Q1

8. B. György, K. Pálóczi, M. Balbisi, L. Turiák, L. Drahos, T. Visnovitz, E. Koltai, Zs. Radák: Effect of the 35 nm and 70 nm size exclusion chromatography (SEC) column and plasma storage time on separated extracellular vesicles. *Current Issues in Molecular Biology* 2024, 46(5): 4337-4357.

   IF: 3.0; Q2

9. B. Szeitz, T. Glasz, Z. Herold, G. Tóth, M. Balbisi, J. Fillinger, Sz. Horváth, R. Mohácsi, H.J. Kwon, J. Moldvay, L. Turiák, A.M. Szász: Spatially resolved proteomic and transcriptomic profiling of anaplastic lymphoma kinase-rearranged pulmonary

adenocarcinomas reveals key players in inter- and intratumoral heterogeneity. *International Journal of Molecular Sciences* 2023, 24(14): 11369.

IF: 4.9; D1

10. G. Tóth, S. Sugár, M. Balbisi, B.A. Molnár, F. Bugyi, K.D. Fügedi, L. Drahos, L. Turiák: Optimized sample preparation and microscale separation methods for high-sensitivity analysis of hydrophilic peptides. *Molecules* 2022, 27(19): 6645.

    IF: 4.6; Q1

11. M. Balbisi, R.A. Horváth, M. Szőri, P. Jedlovszky: Computer simulation investigation of the adsorption of acetamide on low density amorphous ice. An astrochemical perspective. *The Journal of Chemical Physics* 2022, 156: 184703.

    IF: 4.4; Q1

12. M. Balbisi, R.A. Horváth, M. Szőri, P. Jedlovszky: Adsorption of acetamide on crystalline and amorphous ice under atmospheric conditions. A grand canonical Monte Carlo simulation study. *Journal of Molecular Liquids* 2022, 354: 118870.

    IF: 6.0; Q1

13. S. Góbi, M. Balbisi, Gy. Tarczay: Local and remote conformational switching in 2-fluoro-4-hydroxy benzoic acid. *Photochem* 2022, 2(1): 102-121.
    IF: -