Quantitative proteomic analysis: from data-independent acquisition to targeted measurements

Gábor Kecskeméti

Summary of Ph.D. Thesis

Supervisors:

Tamás Janáky Ph.D., D.Sc.

Zoltán Szabó Ph.D.

Doctoral School of Theoretical Medicine

Department of Medical Chemistry

University of Szeged

Szeged, Hungary

2022

1. Introduction

With the recent developments in separation and mass spectrometry methods, liquid chromatography coupled with mass spectrometry (LC–MS) has become a fundamental analytical tool in proteomics. Proteomics refers to the characterization of the proteome, including the expression, function, structure, modifications, and interactions of proteins. Proteomics is key to early diagnosis, prognosis and monitoring disease progression and it also play an important role in drug development. To choose the right MS/MS data acquisition technique, compromises are needed between time, reproducibility, and the depth of the examined protein composition of the sample. The most widely used data collection strategy for proteomic profiling is data-dependent acquisition (DDA) when the mass spectrometer produces MS/MS spectra from the top N most intense ions in a given cycle, which meet the preset criteria. DDA is suitable for examination of the most complete proteome as possible, however quantification is only possible at precursor level (MS1). Targeted methods can be used to analyze fewer peptides, including MS1-based (SIM) and fragment(MS2)-based (SRM, MRM, and PRM) quantification. In general, the MS1 signal is often not specific enough and may contain interferences. Using the fragment ions of peptides, a significantly better signal-to-noise ratio can be achieved, which enables much more selective and specific data acquisition. Data independent acquisition (DIA) is a kind of compromise between DDA and PRM. The goal with this method is to analyze the sample at the deepest available level with better reproducibility than DDA, however, MS2-based quantitation becomes feasible. One of the most challenging limitation factors of DIA, is the wide precursor isolation window, but thanks to modern instruments and techniques, this is becoming less of a problem.

Every step of a standard bottom-up proteomic workflow is important, but sample preparation is of particular importance, as it has one of the greatest effects on the final result. Since protein isolation, extraction and the applied digestion protocol have a significant influence on the peptide composition of a given sample, their investigation and optimization are key points in the development of quantitative methods.

For proteome quantification, DIA is a powerful tool, because of the high throughput and good reproducibility. However, the gold standards for the absolute quantitation of proteins are the SRM and PRM methods. For these strategies, the targets must be defined in advance. Information from preliminary DIA experiments can play a significant role in their selection. The quantitative signal selection has two levels in bottom-up proteomics: it is necessary to select peptides that can specifically reflect the quantity of a protein in the sample, and their most sensitive and robust precursor and fragment ions should be used for quantitation.

2. Aims

In my thesis I present two different projects, both based on the DIA data acquisition. In the first project our main goal was to verify the applicability of a new tear sampling protocol using phenol red treated cotton thread (PRT). The aims for this project were to: 1) examine the effect of the used spectral library on the results of the DIA measurements, 2) evaluate different protein extraction protocols to achieve efficient sample preparation, 3) compare the new procedure with the two most frequently used methods based on proteome content and reproducibility of collected samples.

In the second project our main goal was to develop a targeted LC–MS (PRM) method for targeted measurements of different membrane transporter proteins. Based mostly on experimental processes using DIA acquisition, we aimed to: 4) compare two membrane enrichment and digestion protocols, which are frequently used in LC–MS proteomics studies, 5) develop a method to select peptides of a given protein that can be used to create the most sensitive quantitative method, 6) determine the absolute amounts of overexpressed transporter proteins in vesicles and cell lines used in different membrane transport assays.

3. Materials and Methods

3.1. Samples

3.1.1. Tear samples

Tear fluid samples were collected from both right and left eyes using three different sampling techniques using glass capillaries (CAP), phenol red thread (PRT), and Schirmer's strips. During sample preparation, the Schirmer's strips were divided into lower (SL), which were in direct contact with the eyeball and upper (SU) parts, which were the following 10 mm parts of the strips.

3.1.2. Cell and membrane samples

Total cell (TC) lysates of HEK293 cells, membrane protein fractions of Organic Anion Transporting Polypeptide 1B3 (OATP1B3) overexpressing HEK293 cells (PE, enriched using ProteoExtract® Native Membrane Protein Extraction kit) and Breast Cancer Resistance Protein (BCRP) overexpressing human membrane vesicle preparations (BCRP-HEK293, BCRP-M) and human BCRP-overexpressing insect vesicle preparations (BCRP-Sf9, BCRP-Sf9-HAM) were used for the analyzes.

3.2. Extraction of tear proteins and the evaluation of protein extraction methods from PRT

A pooled tear sample was collected from three healthy persons using glass capillaries. Nine pieces of 5 cm long PRT threads were inserted into the pooled tear sample for 15 sec. After that the treads were divided into three groups. To examine the protein extraction efficiency from PRTs 5% sodium dodecyl sulfate (SDS) in ammonium bicarbonate (AmBic), AmBic, and 1% acetic acid in water were tested. All the SL, SU, and experimental PRT samples were extracted individually using AmBic containing 5% SDS. The capillary tear samples were diluted with the same extraction solution before determining the protein concentration.

3.3. Digestion protocols

Prior to digestion, the protein contents of all samples were determined using BCA Protein Assay. For all tear samples and corresponding cell and membrane samples 10 μg protein were processed based on a modified on-pellet digestion (SEPOD). The rest of cell and membrane samples 10 μg (25 μg for targeted BCRP measurements) protein were processed based on a modified Filter Aided Sample Preparation (FASP) method.

3.4. NanoLC-MS measurements

All of the measurements were carried out on a Waters ACQUITY UPLC M-Class LC system coupled with a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer. Chromatographic separation of peptides was accomplished by gradient elution using water and acetonitrile, both containing 0.1% formic acid were used as mobile phases.

3.5. DIA acquisition and processing

The data required to generate sample-specific spectral libraries were collected from LC–MS analysis of sample-specific pool samples. The mass spectrometer was configured to acquire six gas-phase fractionated (GPF–DIA) acquisitions with 4 Th precursor isolation windows applying an overlapping window pattern. The experimental sample-specific libraries were created by searching the GPF measurements of each sample groups separately against the predicted spectral library. The experimental combined spectral library was built by searching all the project specific groups of GPF measurements in one against the same predicted spectral library. For the quantitative analysis of individual samples, a single LC–MS run with DIA acquisitions (DIA-Q) from 395 to 1020 Th were used with 27×22 Th overlapping precursor isolation windows.

3.6. PRM acquisition and processing

The mass spectrometer was configured to acquire fragment ion spectra from the +2 charged peptide precursors of interest with 2 Th isolation windows.

For the comparison of the sensitivity of methods using different peptides relative limit of detection (LOD) and relative limit of quantitation (LOQ) values were calculated based on calibration curves. These were generated by the dilution of overexpressing HEK293 digests with control HEK293 digest at 5 different ratios (5-fold to 100-fold dilution).

All quantitative comparisons were conducted using Welch T-test with a significance limit of p-value<0.05.

4. Project I.: Deep proteome profiling of tears using DIA acquiring

4.1. Theoretical background

Human tear fluid has attracted increasing interest in the last decades as a potential source of biomarkers of pathophysiological states, due to its accessibility, non-invasive nature of its sampling, moderate complexity, and responsiveness to ocular and systemic diseases. The most frequently used tear sampling methods for proteomics analysis in both clinical and research settings involve the direct collection of tear fluid into a glass microcapillary tube or via an absorbent material such as Schirmer's paper strips. Both sampling methods have advantages and disadvantages. They are often time demanding, uncomfortable for the patient, require medical professional assistance, and sometimes do not provide a sufficient amount of samples for analysis. We need a simple, fast, non-invasive, and reliable tear collection procedure that provides unbiased tear samples even from low-volume sampling. The method of tear secretion measurement using PRT might meet the above requirements.

4.2. Results

A qualitative and quantitative proteomics study was performed on samples collected from the left and right eyes of all donors: altogether 20 PRT, 20 SU, 19 SL, and 18 CAP tear samples were analyzed. Both the volume and the total protein content of tears collected with the CAP sampling procedure showed large variance. The total protein content after the extraction of SL strips, 10 mm long SU strips, and whole PRT samples were more reproducible. These sampling methods could collect sufficient samples for proteomic analysis in all cases, while in two cases with CAP insufficent samples could be collected.

Three different extraction solutions (5% SDS in AmBic, AmBic, and 1% acetic acid in water) were tested to evaluate the efficiency and reproducibility of protein extraction from PRT

samples. There were no significant differences observed in the total amount of extracted proteins; however, the lowest reproducibility was found using acetic acid. The number of detected and quantified proteins, reproducibility (CV) of protein intensities and correlation to the original capillary sample were similar using the SDS and AmBic protocols, but the acetic acid protocol performed worst regarding all these measures. Because of the presence of a tenside may help the protein extraction and for comparability with the protocol applied for Schirmer's strips, we used the SDS approach in our analysis.

Data of DIA acquisitions collected from the analysis of DIA-Q samples were searched against both the sample-specific and combined spectral libraries of different sample types. The number of quantified peptides were 117.5%, 84.7%, and 62% higher for SU, CAP, and PRT samples when the search was carried out on the combined library instead of the sample-specific ones, but it had a lower effect for SL samples. The peptide increment was quantifiable in a significant proportion of DIA-Q samples. A very similar result was obtained at the level of proteins.

For the comparison of the proteomes detectable in our CAP, SL, SU, and PRT samples, the combined spectral library created from a GPF–DIA analysis of sample-specific pooled samples was chosen, as this demonstrates the deepest available protein coverage. The lowest number of proteins could be identified in the pooled CAP sample (422 proteins). In the two indirect pooled tear samples (PRT and SU), a similar number of proteins could be detected (1439 and 1225) with a high overlap, while in the SL sample 2493 proteins were identified. Altogether, 341 protein groups could be detected in all the four sample types. To study the quantitative similarity of different sample types, Pearson's correlation coefficient values were calculated which were the strongest between PRT and SU samples ($r = 0.90$), but the pooled CAP sample also correlates satisfactorily with these samples ($r = 0.76–0.78$).

The statistical classification of proteins based on sample-specific detectability was carried out on the DIA-Q samples. A k-means cluster analysis — based on sample-type specific detection frequency (in %) — was performed to classify the 1144 quantifiable proteins into four clusters (denoted as Cluster A to Cluster D). In Cluster A, there were 195 proteins that could be measured with high frequency in all sample types. Cluster B includes 242 proteins that were repeatedly measurable in samples from indirect procedures (SL, SU, and PRT), but they were quantifiable in only a few CAP samples. The other two clusters consist of proteins that could be measured with high frequency in the SL samples, but in the SU and PRT samples only with medium (Cluster C, 312 proteins) or low (Cluster D, 395 proteins) frequency. Proteins of Cluster A and B may be considered as common tear proteins. The possible contaminating eye-

surface proteins (Clusters C and D) were found in the largest quantities in the SL samples. Proteins of Cluster B were detected only in a few CAP samples, while their relative abundance was highest in SL samples again. Members of Cluster A contribute the most to CAP sample protein intensity (95%), and slightly less than 50% in SL samples. The average abundance of these proteins is higher in PRT (83%) than in SU (76%) samples.

In order to give biological classification of the proteins in tear samples, each protein was annotated with available Gene Onology (GO) and UniProt terms. Proteins associated with specific intracellular localization GO terms, are enriched in Cluster D, which are specific to SL samples, e.g., 21% of Cluster D proteins are from mitochondrion, contrary to cluster A, which includes only 1% of such proteins. The general Cytoplasm GO subcellular localization is enriched in all the Clusters B, C, and D (40–71%), while in the Cluster A, which was the only cluster effectively sampled using capillaries, there were only 19% of such proteins. Sixty-three percent of proteins in Cluster A are secreted, while only 9% of Cluster D proteins are annotated with this Uniprot Keyword. In addition to those general ontological annotations, some more eye-specific information was also added to clarify the origin of proteins in different clusters. The EyeOME database collects a list of proteins identified in different parts of the human eye. There is a large overlap in these assignments, 1213 proteins are common to tears and the eye surface (Cornea, Sclera) in that database. Most of the members of Cluster A are common (124), or specific to tears (31), thus we suggest classifying these as common tear fluid proteins. It must be noted that all Igs can be considered as such also, as those are found in Cluster A, but excluded from the EyeOME. Most of the proteins of Cluster B (228 of 239) are common to tears and eye surface in EyeOME, rarely detected in CAP samples, therefore we can consider them as proteins of the lower layer of tear fluid and proteins easily and reproducibly collected from the eye surface using the indirect sampling methods. Altogether, these 437 proteins in Cluster A and B (392 in EyeOME) proteins we would classify as regular tear sample proteins, independent of origin.

To identify sampling induced effects, the protein composition of tear samples of the same person was compared using Pearson correlation of intensities. The highly contaminated SL samples were excluded from this examination. The strongest intrapersonal correlation was found in the CAP samples, while PRT samples, and SU samples presented weaker correlations. Distribution of protein intensity ratios measured in the two eyes was the narrowest in the case of CAP samples, and widest is in Schirmer's samples. The median CV of the 195 proteins common to all sample types (Cluster A) was lower in PRT samples (64%), than in SU (77%) or CAP (70%) samples.

4.3. Discussion

Several tear sampling methods are available, but as each sampling method has advantages and disadvantages, it is not easy to choose the appropriate one. The most often used tear sampling methods for proteomics analysis in both clinical and research settings involve direct collection of tear fluid into a glass microcapillary tube or via an absorbent material such as Schirmer's paper strips. Phenol red thread, like the previous two tests, is a widespread clinical test for measuring tear volume but has not yet been investigated as a proteomic tear sampling technique.

According to the literature, the capillary samples contain a higher percentage of proteins originating from extracellular region, while an increased number of cell- and organelle-specific (intracellular) proteins can be found in Schirmer's strip samples, due to the contact of the paper with the highly vascularized conjunctiva and the possibility of injuring its surface and microvasculature.

It was confirmed that proteins can be efficiently extracted from PRT, using either the 5% SDS or AmBic protocol, 84–86% of proteins were recovered, and protein composition was not biased, the extract correctly demonstrated the original composition of tear with excellent reproducibility.

Our findings are consistent with the literature, as more proteins were identified in the samples collected with the Schirmer's strips compared to the capillary samples. The number of identified proteins in the novel PRT samples was similarly high, as in the SU samples, and protein intensities showed a strong correlation with other sample types. This proves the applicability of the PRT method to efficiently collect samples for proteomics LC–MS analysis with a composition comparable to samples from other methods.

In our experiments, the identification of such a large number of proteins was made possible by using the GPF-DIA LC–MS method. The application of a combined spectral library of different sample types, increased the number of proteins that could be quantified in the tear fluid in a significant portion of the experimental samples. It may therefore be effective to use multiple type of samples to generate the library, regardless of the applied sample collection protocol for the quantitative experiment.

Based on detection frequency in all sample types, we could identify four clusters of proteins, and by comparing of these clusters to the EyeOME dataset we identified 437 proteins (Cluster A + B) which can be considered as common tear fluid proteins, but only 155 of those (in addition to immunoglobulins) can be effectively sampled by our capillary protocol. PRT has little higher efficiency in sampling of all those proteins than the Schirmer's strip. The summed

relative intensity of the possible contaminant proteins originating from the eye surface (Cluster D) is the lowest in the PRT samples (less than 1%). This may be a consequence of the smaller diameter and the smaller and smoother surface of PRT fibers compared with Schirmer's paper strips.

In order to validate the application of PRT in tear biomarker analysis, we collected tear biomarkers from recent reviews of literature data. We identified 87 proteins in our dataset that were previously assigned as putative biomarkers, 90% of which (78 proteins) are among the proteins which were commonly detected by the indirect sampling methods. Considering these results, PRT is a suitable sampling method for the studying biomarkers of both eye-specific and systemic diseases.

We have found a stronger correlation and smaller differences between samples from the two eyes of the same person using the PRT method compared to the Schirmer's test (SU samples). The interpersonal protein intensity variances within all the healthy subjects were the lowest in the PRT samples, considerably lower than in the SU samples, thus making it more suitable for comparative analysis.

5. Project II.: DIA acquisition as a preliminary experiment for targeted measurements and the application of the established targeted methods

5.1. Theoretical background

Transmembrane drug transporters have a significant effect on the pharmacokinetics and pharmacodynamics of several drugs. ATP-binding cassette (ABC) and solute carrier (SLC) transporters are the most responsible for the influx and efflux delivery of drugs across the cell membrane. Quantification of these transporters expand and significantly increase the utility of in vitro drug metabolism assays, as the measured transporter activities are influenced by their abundance. LC–MS/MS based targeted quantitative proteomics is an appropriate technique to quantify multiple proteins of interest in a single analysis.

Besides some *in silico* selection criteria, in targeted LC–MS measurements the response factor of a peptide must be suitable for sensitive detection. The peptide must be produced in sufficiently large amounts by enzymatic digestion for reproducible detection and its sequence must be specific to the target. A combined approach based on *in silico* and experimental (LC–MS DIA) techniques may be the most appropriate to select peptides with a high degree of confidence.

Since the studied transporters are found and function in the plasma membrane, and are often expressed at a very low level, for their proper detection, in addition to using the most

sensitively measurable peptide, some membrane enrichment method may be also required. However, it has been shown that it is difficult to separate the plasma membrane from other membranes with currently available membrane purification techniques. A further difficulty in the detection of membrane proteins is their frequently large size and most notably their hydrophobic characteristics. For their proper handling and keeping them in solution, the use of a relatively large amount of detergent is unavoidable, which may complicate further sample processing and analysis.

5.2. Results

The aim was to compare the protein composition of the membrane that can be isolated from cells using PE kit with the membrane preparation used for vesicular transport studies. Both types of membrane-enriched samples were compared with the TC lysates, in terms of number and quantity of the proteins and peptides after SEPOD digestion. In the enriched samples higher number of membrane proteins and peptides with higher summed intensities were measured than in the TC lysates. The relative intensity of membrane proteins and membrane peptides did not differ between the two membrane samples. The number of ABC + SLC proteins was approximately 1.5–1.6-fold higher in the enriched samples (PE–vesicle samples) than in the TC samples. The ratio of the total intensity of these membrane proteins increased approximately 2.6–3.8-fold compared to that measured in TC samples. To characterize the sample composition obtained with the isolation techniques and comparing to TC lysates, specific marker proteins were selectively evaluated from the DIA measurements. The cytosolic protein content of both the PE and vesicle samples were significantly less than of the TC samples, along with this, plasma membrane protein content increased significantly applying both enrichment protocols. As a result of both membrane isolation, the nucleus proteins were present in a lower amount and the mitochondrial membrane proteins in a higher amount while the Golgi membrane proteins were significantly enriched only in the vesicle samples, than in TC.

Because of the often-large size and extreme hydrophobic nature of membrane proteins the use of a significant amount of detergents cannot be avoided. A sample preparation/digestion protocol capable of removing these substances is needed, such as FASP and SEPOD. However, the peptide composition of the sample is highly dependent on the digestion method used, so it is important to investigate its effects. More proteins and peptides could be quantified using the SEPOD protocol than with FASP in both PE and vesicle samples, however the difference was higher in case of PE samples. With both methods, the average ratio of the number of membrane

proteins and peptides compared to all was approximately 40%, while this number was around 2% for ABC + SLC transporters. The differences between the digestion protocols were more prominent at the level of peptides than proteins. The reproducibility of the measured intensities was better using SEPOD for both membrane and ABC + SLC proteins and peptides.

The selection of the most suitable peptides for targeted protein quantification was based both on sequence based in silico analysis and quantitative experimental studies. Peptide sequences were analyzed for the presence of probably problematic amino acids, known post-translational modifications and possible missed enzymatic cleavage sites within and around the particular peptide sequence. Location of the peptides relative to transmembrane regions was also considered.

To establish the most sensitive method, it is important to know the detection (LOD) and quantification limit (LOQ) values that can be achieved applying the different peptides. As these values are highly dependent on the sample matrix, we developed a method in which a dilution series was prepared by mixing the digestion of an overexpressing sample with the digestion of a non-overexpressing control sample of the same cell line in several replicates. The relative LOD and LOQ values were calculated from the linear regression data of the obtained calibration curves. Considering all aspects, the NQTANLTNQGK peptide proved to be the most applicable for the absolute quantification of OATP1B3. Peptides were also selected for BCRP protein quantification using the method presented above. Since the aim was to quantify BCRP expressed by other species in addition to human BCRP, another peptide (LLSDLLPMR) was used in addition to the SSLLDVLAAR peptide, as it originates from an evolutionarily conserved region. The methionine sulfoxide of the LLSDLLPMR peptide was monitored from DIA measurements. The relative methionine oxidation was 5.65 ± 0.49% during the whole experiment assuming the same response factor for oxidized and non-oxidized forms.

The expression stability of OATP1B3 overexpression in HEK293-OATP1B3-LV cells was investigated during tissue flask culture and multiple passaging using the selected NQTANLTNQGK peptide. For this purpose, membrane fractions were prepared from the cells using PE kit to be examined from three biological replicates (A, B, and C) at the time of culturing (0 passage), and after 8 and 16 passage (approximately 1 and 2 months). Since the specific absolute quantities are not published, the initial quantity of one of the samples was taken as 100%. The quantity measured in the other samples was related to that amount. There was no significant difference in the expression of investigated transporter between the replicates and time points. The relative standard deviation (RSD) of OATP1B3 expression between the samples were low, 7.7% and 11.7% at the initial point of the experiment and after 8 passages,

but a higher value (21.9%) was observed after 16 passes. The LOQ of the established targeted method was 0.014 pmol / mg membrane protein and the RSD of the heavy isotope labelled NQTANLTNQGK standard during the measurements was 8.5%.

The absolute amount of BCRP transporter protein was determined in four different vesicle types (HEK293, MCF-7/MX, Sf9 and SF9-HAM), all overexpressing the same human BCRP protein. The average of two proteospecific peptides (SSLLDVLAAR and LLSDLLPMR) was used for quantitation as the amounts of the two peptides were similar in these samples. No significant difference was detected in the BCRP expression of the different types of vesicle samples. The LOQ of the established targeted method was 0.030 pmol / mg membrane protein (using SSLLDVLAARK) and the RSD of the heavy isotope labelled SSLLDVLAARK and LLSDLLPMR peptides were 8.0 % and 10.8 %, respectively.

5.3. Discussion

One of the key parts in the sample preparation for transporter quantification is the isolation of pure cell membrane fraction. Therefore, PE membrane isolates were compared with vesicular preparations to investigate the differences in the obtained proteome, especially in terms of the amount of ABC and SLC transporters. The membrane protein content of the samples was effectively increased using both membrane enrichment protocols, along with this, the measured number, and the relative intensities of the ABC and SLC transporters also increased notably. Both methods were able to decrease the proportion of cytosolic proteins in the samples compared to the whole cell samples. Although there were differences in the enrichment of different subcellular membranes (ER, Golgi, mitochondrion) in the two protocols, the intensities of membrane proteins showed strong correlation.

In bottom-up proteomics, another important influencing factor for the results is the applied digestion protocol. The difference in the number of detected proteins and peptides between FASP and SEPOD was smaller for the vesicle samples than for the PE samples, probably because the vesicle samples did not contain detergents. Although an extra detergent removal step was used in the FASP protocol, for PE samples containing large amounts of detergent, the use of SEPOD was preferable as a higher amount of protein and peptides could be quantified than with the FASP protocol. There was no difference in the repeatability of the relative intensities between PE and vesicle samples using the same digestion protocol, while CVs using SEPOD were lower for both sample types.

Choosing the sufficient peptides for quantifying proteins have a significant effect on the final result. The response factor of peptides within proteins could differ on average by over

three orders of magnitude. To create the most robust, reliable and sensitive targeted methods, the most suitable peptide should be selected based on several criteria. The peptides applied for the absolute quantification of OATP1B3 and BCRP proteins were selected by a combined criteria system. The relative LOQs of the most intense peptides were determined using calibration curves. If the digests of the overexpressed samples were diluted with control digests of the same cell type, the matrix effect was constant during the relative LOQ determination. With this method precise information could obtained about what dilution of the overexpressed amount can be detected in the given matrix with the tested peptides. Taking into account a combination of sequence restrictions, online available information and experimental DIA data (MS2 intensity, relative LOQ value), the most suitable peptides were selected for PRM measurements.

Using the targeted methods established, we were able to confirm the stability of expression of OATP1B3 protein in HEK293-OATP1B3-LV cells over 16 passages for approximately 2 months. There was no significant difference in the amount of expressed BCRP in the different BCRP overexpressing vesicles prepared from different cell lines, so that all these membranes can be used to perform activity assays.

6. Summary of the new findings

1) Application of a combined spectrum library of tear samples collected from multiple sampling procedure instead of sample-specific spectrum libraries was found to be effective to increase the proteome coverage in DIA analyzes.
2) We have confirmed that the novel PRT method is highly efficient for sample collection of tear fluid proteins. Proteins collected by this method can be efficiently and reproducibly extracted from the threads. Unlike microcapillaries, it can also be used for proteomic analyses with small amounts of tears.
3) The protein composition of samples collected using this method strongly correlate with samples collected using other common methods. The intra- and inter-personal variance with the PRT method was lower than with the other sampling procedures, since the PRT method is fast, non-irritating and can be used to collect small sample volumes with low level of eye surface contamination.
4) Cluster analysis was used to classify proteins according to the frequency of occurrence in the different tear sample types. On this basis, we identified common tear fluid proteins as well as eye surface proteins.

5) The proteomes of membrane isolates prepared from cells with PE kit were compared with membrane preparations used for vesicular transport assays. There were differences between the membranes based on marker proteins, however, relative quantities of membrane proteins and peptides showed a strong correlation.
6) The SEPOD digestion protocol was found to be more effective in digestion of samples with tenside content even without separate steps towards their removal. We were able to quantify more membrane proteins and ABC and SLC transporters with better reproducibility using this method from membrane enriched samples, than with FASP.
7) We have established a method for relative LOQ determination, which can help in the selection of the suitable peptide for absolute quantification. For this purpose, digests from protein-overexpressing cells were diluted with control digests of the same cell type.
8) Our targeted methods were suitable for the determination of absolute amounts of target proteins with satisfactory reproducibility for measurement of biological variability both in membrane vesicles and cell lines used in membrane transport assays. The stability of overexpressed OATP1B3 in OATP1B3-HEK293-LV cells over 16 passages was confirmed, and the targeted BCRP quantification contributed to the interpretation of activity assays of different BCRP overexpressing vesicle samples.

List of publications related to this thesis:

I. **Kecskeméti, G.**; Tóth-Molnár, E.; Janáky, T.; Szabó, Z. An Extensive Study of Phenol Red Thread as a Novel Non-Invasive Tear Sampling Technique for Proteomics Studies: Comparison with Two Commonly Used Methods. *Int. J. Mol. Sci.* **2022**, 23, 8647, doi:10.3390/ijms23158647. **IF: 6.208 (2021)**

II. Sáfár, Z.; **Kecskeméti, G.**; Molnár, J.; Kurunczi, A.; Szabó, Z.; Janáky, T.; Kis, E.; Krajcsi, P. Inhibition of ABCG2/BCRP-Mediated Transport–Correlation Analysis of Various Expression Systems and Probe Substrates. *Eur. J. Pharm. Sci.* **2021**, 156, 105593, doi:10.1016/j.ejps.2020.105593. **IF: 5.112**

List of publications not related to this thesis:

1. Jójárt, R.; Pécsy, S.; Keglevich, G.; Szécsi, M.; Rigó, R.; Ozvegy-Laczka, C.; **Kecskeméti, G.**; Mernyák, E. Pd-Catalyzed Microwave-Assisted Synthesis of Phosphonated 13α-Estrones as Potential OATP2B1, 17β-HSD1 and/or STS Inhibitors. *Beilstein J. Org. Chem.* **2018**, 14, 2838–2845, doi:10.3762/bjoc.14.262. **IF: 2.592**

2. Bacsa, I.; Konc, C.; Orosz, A.B.; **Kecskeméti, G.**; Rigó, R.; Zvegy-Laczka, C.; Mernyák, E. Synthesis of Novel C-2- or C-15-Labeled BODIPY—Estrone Conjugates. *Molecules* **2018**, 23, 821, doi:10.3390/molecules23040821. **IF: 3.060**
3. Tömösi, F.; **Kecskeméti, G.**; Cseh, E.K.; Szabó, E.; Rajda, C.; Kormány, R.; Szabó, Z.; Vécsei, L.; Janáky, T. A Validated UHPLC–MS Method for Tryptophan Metabolites: Application in the Diagnosis of Multiple Sclerosis. *J. Pharm. Biomed. Anal.* 2020, 185, 113246, doi:10.1016/j.jpba.2020.113246. **IF: 3.935**

4. Gieszinger, P.; Stefania Csaba, N.; Garcia-Fuentes, M.; Prasanna, M.; Gáspár, R.; Sztojkov-Ivanov, A.; Ducza, E.; Márki, Á.; Janáky, T.; **Kecskeméti, G.**; et al. Preparation and Characterization of Lamotrigine Containing Nanocapsules for Nasal Administration. *Eur. J. Pharm. Biopharm.* **2020**, 153, 177–186, doi:10.1016/j.ejpb.2020.06.003. **IF: 5.571**

5. Ambrus, R.; Gieszinger, P.; Gáspár, R.; Sztojkov-Ivanov, A.; Ducza, E.; Márki, Á.; Janáky, T.; Tömösi, F.; **Kecskeméti, G.**; Szabó-Révész, P.; et al. Investigation of the Absorption of Nanosized Lamotrigine Containing Nasal Powder via the Nasal Cavity. *Molecules* **2020**, 25, 1065, doi:10.3390/molecules25051065. **IF: 4.412**

6. Bartos, C.; Ambrus, R.; Kovács, A.; Gáspár, R.; Sztojkov-Ivanov, A.; Márki, Á.; Janáky, T.; T m si, F.; **Kecskeméti, G.**; Szabó-Révész, P. Investigation of Absorption Routes of Meloxicam and Its Salt Form from Intranasal Delivery Systems. *Molecules* **2018**, 23, 784, doi:10.3390/molecules23040784. **IF: 3.060**

7. Katona, G.; Balogh, G.T.; Dargó, G.; Gáspár, R.; Márki, Á.; Ducza, E.; Sztojkov-Ivanov, A.; Tömösi, F.; **Kecskeméti, G.**; Janáky, T.; et al. Development of Meloxicam-Human Serum Albumin Nanoparticles for Nose-to-Brain Delivery via Application of a Quality

by Design Approach. *Pharmaceutics* **2020**, 12, 97, doi:10.3390/pharmaceutics12020097. **IF: 6.321**

8. Tuka, B.; Nyári, A.; Cseh, E.K.; Körtési, T.; Veréb, D.; Tömösi, F.; **Kecskeméti, G.**; Janáky, T.; Tajti, J.; Vécsei, L. Clinical Relevance of Depressed Kynurenine Pathway in Episodic Migraine Patients: Potential Prognostic Markers in the Peripheral Plasma during the Interictal Period. *J. Headache Pain* **2021**, 22, doi:10.1186/s10194-021-01239-1. **IF: 8.592**

9. Jójárt, R.; Laczkó-Rigó, R.; Klement, M.; Kőhl, G.; **Kecskeméti, G.**; Özvegy-Laczka, C.; Mernyák, E. Design, Synthesis and Biological Evaluation of Novel Estrone Phosphonates as High Affinity Organic Anion-Transporting Polypeptide 2B1 (OATP2B1) Inhibitors. *Bioorg. Chem.* **2021**, 112, 104914, doi:10.1016/j.bioorg.2021.104914. **IF: 5.307**

10. Annus, Á.; Tömösi, F.; Rárosi, F.; Fehér, E.; Janáky, T.; **Kecskeméti, G.**; Toldi, J.; Klivényi, P.; Sztriha, L.; Vécsei, L. Kynurenic Acid and Kynurenine Aminotransferase Are Potential Biomarkers of Early Neurological Improvement after Thrombolytic Therapy: A Pilot Study. *Adv. Clin. Exp. Med.* **2021**, 30, 1225–1232, doi:10.17219/acem/141646. **IF: 1.736**

Number of publications: 12 (1 first author)

Cumulative IF: 55.905

Number of independent citations (MTMT2): 53

https://m2.mtmt.hu/gui2/?type=authors&mode=browse&sel=10063503&view=pubTable

Hirsch index: 6