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

Gábor Kecskeméti

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

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

Table of contents

1. Introduction 1

1.1. Common proteomics workflows 1

1.2. Peptide identification based on tandem mass spectra 3

1.3. Data acquisition strategies using LC–MS 4

1.4. Development of DIA data acquisition technique for proteomics 7

1.5. DIA data evaluation 9

1.6. Applications of DIA in proteomics 10

2. Aims 11

3. Materials and Methods 11

3.1. Materials/reagents 11

3.2. Samples 12

3.2.1. Tear samples 12

3.2.2. Cell and membrane samples 13

3.3. Extraction of tear proteins and the evaluation of protein extraction methods from PRT 14

3.4. Digestion protocols 14

3.4.1. Surfactant Cocktail-Aided Extraction/Precipitation/On-Pellet Digestion (SEPOD) method 14

3.4.2. Filter Aided Sample Preparation (FASP) method 15

3.5. NanoLC–MS measurements 16

3.6. DIA acquisition and processing 16

3.7. PRM acquisition and processing 17

4. Project I.: Deep proteome profiling of tears using DIA acquiring 18

4.1. Theoretical background 18

4.2. Results 20

4.2.1. Comparison of the effectiveness of sampling methods 20

4.2.2. Evaluation of the proteomics protocol 20

4.2.3. Impact of MS data acquisition and evaluation method 21

4.2.4. Comparison of the tear proteomes of different sample types 23

4.2.5. Clustering of proteins in different types of samples 25

4.2.6. Classification of tear proteins 27

4.2.7. Intra- and interpersonal variances of different tear samples 29

4.3. Discussion 30

5. Project II.: DIA acquisition as a preliminary experiment for targeted measurements and the application of the established targeted methods. 35

5.1. Theoretical background 35

5.2. Results 38

5.2.1. Comparison of different membrane enrichment protocols 38

5.2.2. Comparison of different digestion protocols 41

5.2.3. Selection of suitable peptides for targeted measurements 42

5.2.4. Absolute quantification of OATP1B3 and BCRP in different membrane fractions 44

5.3. Discussion 46

6. Summary 49

7. Summary of the new findings 49

8. Acknowledgements 51

9. References 52

10. Supplementary material 63

List of abbreviations

ABC	ATP-binding cassette
AGC	automatic gain control
AmBic	ammonium bicarbonate
BCRP	Breast Cancer Resistance Protein
CAP	capillary
CV	coefficient of variation
DDA	data dependent acquisition
DDT	dithiothreitol
DIA	data independent acquisition
EyeOME	The Human Eye Proteome Project
FA	formic acid
FDR	false discovery rate
HCD	higher-energy collisional dissociation
IAA	iodoacetamide
LC–MS	liquid chromatography–mass spectrometry
LOD	limit of detection
LOQ	limit of quantitation
M	mol/L
m/z	mass/charge
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS1	precursor level mass spectrometry
MS2	fragment level mass spectrometry
nanoLC–MS	nano-scale liquid chromatography–mass spectrometry
OATP1B3	Organic Anion Transporting Polypeptide 1B3
PE	ProteoExtract™
PRT	phenol red thread
RSD	relative standard deviation
SDS	sodium dodecyl sulfate
SL	Schirmer’s strip lower part
SLC	solute carrier
SU	Schirmer’s strip upper part
TC	total cell

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 biological and medical research. LC–MS enables qualitative and quantitative analysis of a wide range of biomolecules in a high-throughput manner, which has led to significant advances in many research areas, such as biomarker discovery and systems biology [1–3]. The mass spectrometer produces ions from the sample substances, separates them according to their mass-to-charge-ratio (m/z) and records the relative abundance of each type of ions. The results are presented as a mass spectrum, a plot of intensity versus the mass-to-charge ratio. The selected ions (precursors) can be broken down into smaller pieces, called fragment ions, using various tandem mass spectrometry (MS/MS) methods. Fragments of a molecule have a unique pattern in the mass spectrum; therefore, these data facilitate the identification of a wide variety of biomolecules (proteins, carbohydrates, lipids etc.) in MS/MS experiments [4].

The proteome can be defined as the collection of proteins present in cells, tissues, biofluids and reflects the functional state of the biological system [5]. The proteome is not constant, it changes over time, from cell to cell, or even in response to an external stimulus. In eukaryotic cells, proteomics is complex because of the post-translational modifications that occur in different ways at different sites [6]. 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 also play an important role in drug development [7].

1.1. Common proteomics workflows

The most frequently used proteomic strategies for protein identification are based on various combinations of separation techniques, mass spectrometry (MS) and bioinformatics tools. Recently, three MS-based proteomics approaches are widely used for protein identifications. During the top-down strategy [8], intact proteins are measured in their original state, without digestion, with high-resolution MS. Applying the middle-down approach [9], the proteins are digested to a limited extent, the resulting peptides can be analyzed relatively easily by MS, however, their sequence is still quite long and may contain several post-translational modifications that allow the identification of protein variants. The focus of this thesis is limited to the most widely used, “peptide-centric” bottom-up strategy [10], when complex mixtures of proteins are first digested to peptides via chemical or

enzymatic cleavage, then the resulting peptide products are separated by LC and analyzed using MS/MS.

The main steps of a standard bottom-up proteomics experiment are the following: i) isolation of proteins from a sample; ii) fractionation to remove contaminants and proteins that are not of interest (not always needed); iii) digestion of proteins into peptides; iv) post-digestion separations to reduce the number of peptides to be detected at the same time (e.g. with LC); v) analysis by MS and vi) data processing and evaluation with bioinformatics tools [11]. Each step of the whole process is important to achieve a successful result, but sample preparation is a key step as the quality of the sample has a significant impact on the final results.

Detergents, such as SDS, are routinely used to denature and solubilize proteins, especially membrane proteins. However, even at very low concentrations, these chemicals can interfere with subsequent protease digestion and MS analysis and are difficult to remove from solution [12]. For controlled digestion of proteins into peptides, extracted and solubilized proteins are digested with a sequence-specific protease. Trypsin is gold standard because it specifically cleaves after lysines or arginines, leaving a positively charged amino acid at the newly formed C-terminal, which is an advantage for ionization and fragmentation [5]. Different digestion protocols have a significant influence on the peptide composition of the formed digests, therefore each step of the sample preparation should be examined and optimized.

The detection of MS signals that are interference-free, selective and have a good signal-to-noise ratio — which is essential for quantitative measurements — is facilitated by liquid chromatography, which is able to resolve complex samples, thus reducing the number of components to be analyzed simultaneously in the MS. In the field of bottom-up proteomics, peptide separations are generally carried out using a solvent gradient flowing through a packed non-polar stationary phase, most often hydrophobic carbon chain bonded silica (C8–C30) with particle sizes varying between 1.7 to 3.5 μm. The analytes eluting from the column are transferred to an appropriate ESI ion source, which ensures the soft ionization of liquid phase analytes to gaseous phase ions, which are able to enter the mass spectrometer [13]. Mass spectrometers can be equipped with several types of mass analyzers, of which the orbitrap mass analyzer is a reasonable choice for proteomic studies: it is suitable for MS/MS fragmentation, has high resolution and mass accuracy, and can be combined with other analyzers (quadrupoles and iontraps) to perform different data acquisition modes and increase selectivity and sensitivity [14].

The peptides are measured with peptide-specific precursor and/or fragment ions, and bioinformatics tools help to compile those detections into proteins. The intensity of the detected m/z signal at any given retention time depends on the amount of peptides in the sample, which can be integrated over time to quantify the peptides [15]. In LC–MS analysis, matrix components may co-elute with the analyte and interfere with the ionization of the analyte in the mass spectrometer. This matrix effect can cause ionization enhancement or suppression, which can negatively affect the results of LC–MS analysis [16]. Therefore, 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.

1.2. Peptide identification based on tandem mass spectra

The peptide identifications could be based on matching the MS/MS spectra against *in silico* fragmentation of possible combinations of amino acids (*de novo* sequencing) or against *in silico* digested peptides from the protein sequence database (database search). Sequence database search engines generate theoretical spectra of all possibly existing peptides based on a given sequence database even if they were never detected with MS. Usually only m/z value of canonical fragment ions are predicted and used for identification by these search engines [17]. A recently evolving approach to identify peptide MS/MS spectra is spectral library searching. For identifying the observed MS/MS spectra, spectral library search engines use spectral libraries of identified, generally experimental MS/MS spectra. Since the experimental spectral libraries contain peaks and measured intensity values even of non-canonical fragment ions, a higher identification rate can be achieved, especially for low-resolution data sets and mixture (chimera) MS/MS spectra. In addition, these spectrum libraries only contain information about peptides that can be detected by MS [17]. For this reason, spectrum library search engines are able to identify more peptides in significantly less time than database search engines [18], but it is limited to the spectra in the library. Optimally it contains the spectra of all peptides that can occur in the sample (no false negatives) and only a few that have a low chance of occurring (reduces false positives).

Each of the aforementioned approaches require sophisticated bioinformatics tools, the choice of which depends on MS instrumentation (eg. resolution) and MS/MS acquisition method. More and more new algorithms are appearing, or existing ones are constantly being improved in order to be more efficient and reliable, as well as to be able to take advantage of the newer, more advanced and faster instruments and data collection techniques [19].

1.3. Data acquisition strategies using LC–MS

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 [15].

The most commonly 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 (Figure 1-A), which meet the preset criteria [20]. In order to generate MS/MS spectra from fragment ions that originate from a single peptide, only a narrow m/z range, 1–4 Th around the mass of the peptide is monitored with this approach at a time. As it is possible to limit the measurements to only the m/z range where the signals are derived from peptides, deep proteome coverage can be achieved using DDA [21,22] in a single LC–MS run [23]. One of the major disadvantages of this acquisition method is the stochastic nature of precursor selection for fragmentation, since it is based on actual measured intensities, minor differences between the samples and the time of data acquisition may result in different peptides being selected for MS/MS, thereby negatively affecting the reproducibility of the measurements. With the DDA acquisition method the peptides are dynamically excluded from being fragmented multiple times during its elution. This dynamic exclusion increases the number of unique peptide detections, but prevents the identification of closely eluting peptidoforms. In the absence of repeated MS/MS data, quantification based on the integration of the area under the curve can only be implemented at the MS1 level [15].

While DDA aims to examine the most complete proteome as possible, targeted methods deal with measuring a limited number of peptides. In selected ion monitoring [24] (SIM) acquisition mode, the mass analyzer is set to measure one peptide ion, which can be constantly monitored over the time. With this method, data is only collected at the precursor level (MS1), not at the fragment (MS2) level. Applying selected reaction monitoring [25] (SRM or MRM) or parallel reaction monitoring [26] (PRM) MS/MS data from pre-defined m/z ranges are consistently generated (Figure 1-B), regardless of whether the given precursor is present or not. The fast cycle time of modern MS instruments allows collecting several MS/MS spectra during the elution of a peptide, which can be used for quantitative determination of peptides/proteins not from MS1 but MS2 signals.

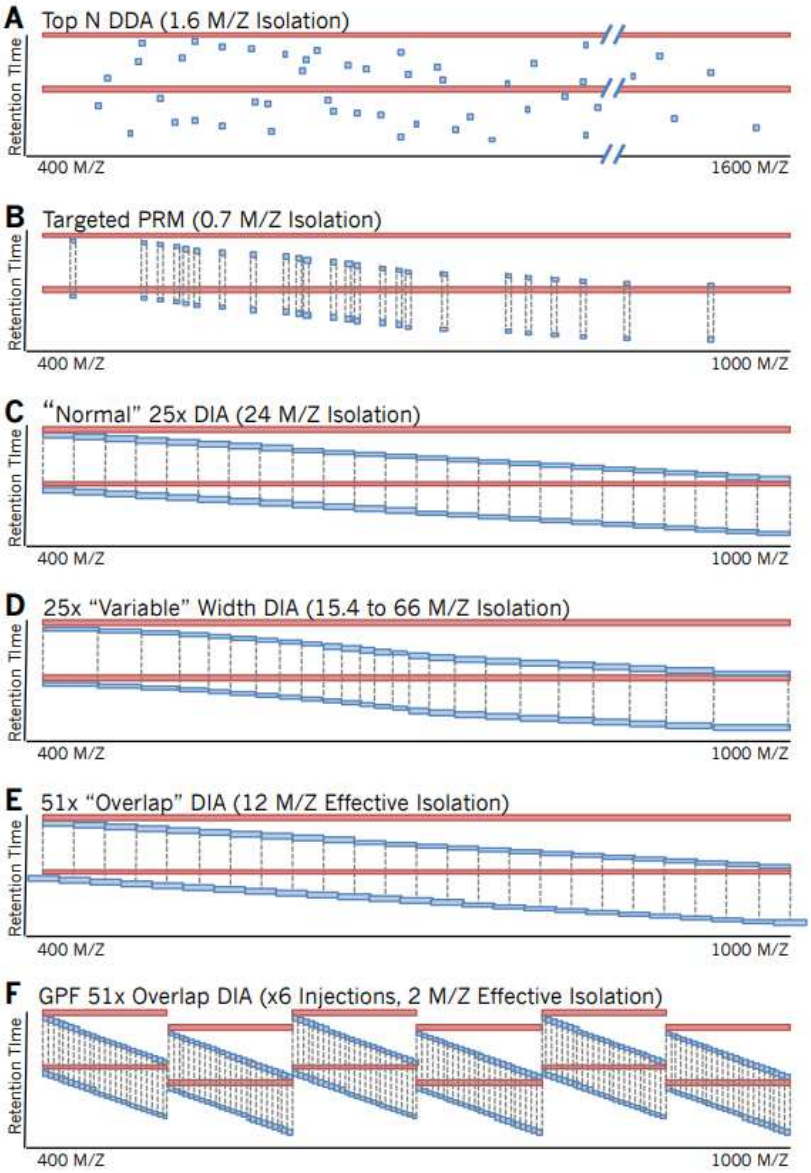


Figure 1. Summary of MS/MS-based proteomic data acquisition methods. Source: Brian C. Searle, 2018 [15]

(A) DDA method: first, an MS1 scan (red line) is collected from the whole preset m/z range (e.g. 400 to 1600 Th), and then MS/MS spectra are generated with narrow precursor isolation of the N (e.g. 12) most intense ions (blue squares).

(B) Targeted method: produces repeated MS/MS spectra with a narrow isolation window, regardless of the presence of the precursor.

(C) Normal DIA: collects MS/MS spectra from the whole mass range in fix-width wide-windows at consistent intervals.

(D) Variable Width DIA: the width of the MS/MS precursor isolation window depends on the number of ions in the investigated m/z range and is adjusted accordingly.

(E) Overlapped DIA: collects the data in the same way as Normal DIA, the only difference is that it slides half the window size in each cycle, thus reducing the number of isolated precursors with deconvolution between cycles.

(F) Gas phase fractionated DIA (GPF): it requires several (e.g. 6) runs from one sample, each has a different mass range of interest, but the precursor isolation window can be as narrow as in DDA and PRM.

In general, the MS1 signal of a peptide is much more intense than the MS2 signal, but its usability is more limited. Since different peptides are built from the same building blocks, different precursors can have the same m/z value. For this reason, the MS1 signal is often not specific enough and may contain interferences (undistinguishable MS signals from other molecules) (Figure 2-B1).

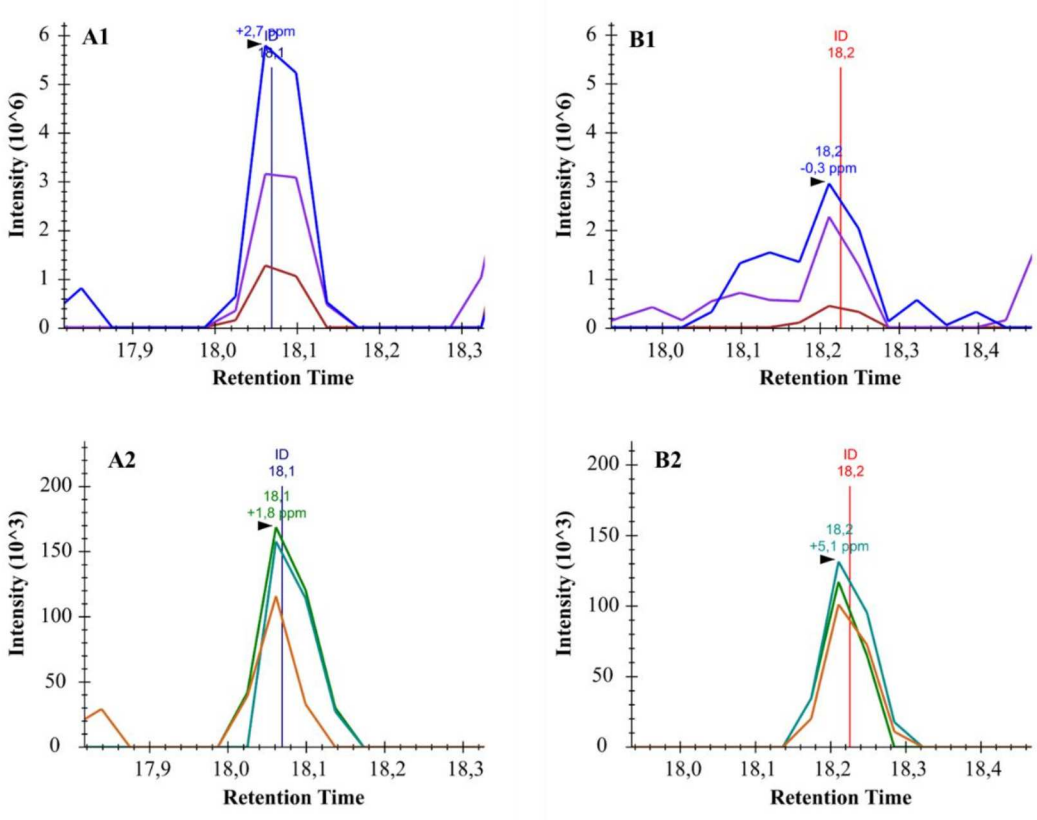


Figure 2. The fragment ions of a peptide are often free from interference even when the precursor ion interferes with signals from other ions. MS1 (A1, B1) and MS2 (A2, B2) data from the same peptide in two different measurements.

Interferences are less probable in MS2, resulting in a significantly better signal-to-noise ratio. This enables acquisition of much more selective and specific data. For both SRM and PRM, it is also necessary to pre-select the MS signals to be measured. In addition, the number of peptides that can be measured simultaneously is limited, so they are usually

scheduled to different retention time windows, or several runs can be multiplexed, as a result of which even hundreds or thousands of peptides can be measured in one experiment [27].

Data independent acquisition [28] (DIA) is a kind of compromise between DDA and PRM (Figure 1-C). Like DDA, the goal with this method is to analyze the sample at the deepest available level, identify and quantify as many peptides as possible. However, like PRM, serial MS/MS spectra are collected from all precursor ions in a given m/z range, so MS2-based quantitation becomes feasible.

In the original DIA method, Venable *et al.* used 10 Th wide precursor isolation window with a cycle time of 35 sec. Due to the long cycle time, only a few MS/MS spectrum was acquired from each detectable peptide, therefore quantitation could only be solved at the MS1 level. The much shorter cycle time of modern devices makes it possible to produce sufficient MS/MS spectra from a precursor ion to enable quantification at MS2 level. Despite the availability of such tools, DIA has some limiting factors. First, by keeping the optimal precursor isolation window width and optimal cycle time, a limited m/z range can be covered, thus some peptides cannot be detected in their optimal charge state. In addition, while with the DDA and PRM methods it is possible to optimize the collision energy (CE) for each given precursor to achieve the appropriate fragmentation, in the DIA method only one collision energy value can be set for co-eluting peptides in a precursor isolation window. It is common that the simultaneously eluting peptide precursor ions have different masses and charges, consequently not all precursors can be fragmented with optimal collision energy. Third, since the goal is to collect data on as many peptides as possible in one measurement, a wide mass range must be measured with wide precursor isolation windows, so the collected MS/MS spectra are highly chimeric, derived from multiple co-eluting peptides. For the accurate and reliable peptide identification and quantification, these signals must be deconvoluted. Despite these difficulties, DIA also has significant advantages: it has better reproducibility than DDA, which is a distinct advantage in experiments with a large number of samples, and, unlike PRM, it is possible to re-evaluate the MS/MS data if later examination of previously not investigated precursors become important.

1.4. Development of DIA data acquisition technique for proteomics

One of the most challenging limitation factors of DIA, is the wide precursor isolation window. Several strategies have emerged to overcome this difficulty. One such approach was developed by Zhang *et al.* [29], in which instead of constant-width isolation windows, variable-width windows were used (Figure 1-D). Taking advantage of the fact that the

distribution of the number of peptides in the investigated mass range is not uniform, the width of the isolation windows was optimized for the expected number of peptide precursors. It would be the best to measure all possible peptides, but from a practical point of view it is worth to optimize the cycle time and the mass range of the measurement. Thus, it is possible to collect the most MS/MS spectra from a precursor and create the most reliable quantitative measurements. Although the signal of some peptides are more intense below 400 Th or above 900 Th, Pino *et al.* [30] found that 93% of all peptides in the Pan-Human library [31] can produce a measurable signal in the mentioned range (Figure 3.).

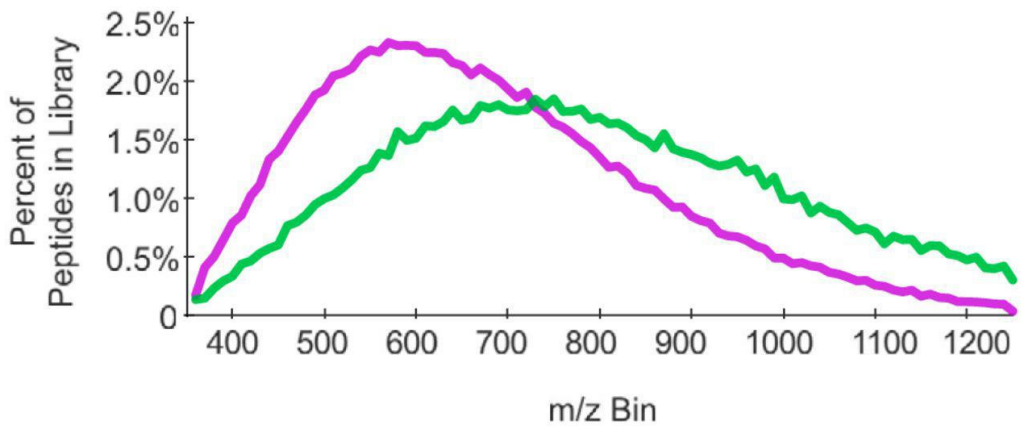


Figure 3. The distribution of peptide precursor m/z in the Pan-Human library (purple) and in the HeLa phosphopeptide library (green). Source: Pino *et al.*, 2020 [30]

Thus, if measurements are focused only on this range, the width of the isolation window can be narrower, which can reduce the number of precursor interferences. In this way, simpler deconvolution and more reliable results can be achieved. However, only 77% of the peptides in a human phosphopeptide library [32] can be measured in the same range, so it can be concluded that it is always worth optimizing the measurement range for peptides that are important in the given experiment. For general proteomic studies, this m/z range is from 400 to 1000 Th. A multiplexing approach was devised by Egertson *et al.* [33], where narrow precursor isolation windows were randomly mixed, then the data were demultiplexed after measurement. This approach is only feasible with certain mass spectrometers, but a variation of it can be performed with any DIA-capable instrument. In this, overlapping precursor isolation windows are used (Figure 1-E), which are repeated in every second cycle, then the data can be deconvoluted to virtually decrease the width of the isolation window. Finally, the PAcIFIC [34] approach, which uses gas phase fractionation (Figure 1-F), whereby the isolation windows are narrow, but multiple injections can cover the entire mass

range of interest. Of course, this method is not optimal for experiments with a large number of samples, but it is suitable for creating a spectral library, which enhance the evaluation of the wide-window DIA [35].

1.5. DIA data evaluation

Two approaches are used to interpret DIA data, identify, and quantify peptides. One of these is spectrum-centric analysis, the essence of which is to try to demultiplex several spectra derived from different precursors in each MS/MS spectrum [36,37] and assign those fragment ions to a "pseudo" spectrum whose signal intensity co-vary by the retention time, since these fragments probably arise from the same precursor. The great advantage of the method is that, like DDA, these pseudo spectra can be used by a wide variety of MS/MS software for further data evaluation. The drawback is that the assignment of fragments to peptides is not always straightforward, often MS1 signals are not even collected, and the peaking of MS2 signals is cumbersome. In addition, the identification of less abundant analytes from mixed spectra is difficult because their signal is suppressed by the more abundant analytes. In contrast the peptide-centric [38] approach searches for specific fragment ions of peptides in each spectrum of the precursor isolation window.

Peptide-centric software tools for interpreting DIA data can be divided into two classes: library-based or library-free approaches (Table 1) [39]. A spectral library is usually composed of a set of different coordinates including the precursor and fragment ion m/z values, relative intensity of fragment ions and standardized retention time (RT) for each peptide precursor and its fragments [40] and used to extract the ion chromatograms of the targeted peptides from the DIA spectra.

Libraries can be built from acquired data with several tools, such as MaxQuant [53], SpectraST [54], Skyline [44], Pulsar [55], but ready-to-use libraries are also available online, e.g. SWATHatlas [56], SRMAtlas [57], PeptideAtlas [58], and NIST [59]. With the recent advances in deep learning, predicted fragment ion spectra has reached a comparable level of accuracy as spectra generated by empirical approaches. Suitable tools for prediction, for example DeepMass [60], Prosit [61], and pDeep [62]. Library-based approaches are sensitive but limited the data interrogation to only analytes present in the library. However, using too large library significantly reduces the speed of the database search and increases the probability of false results. Using the right size spectrum library is critical to successfully evaluate DIA data. Spectral library is usually generated by multiple fractionated analysis of the same or same type of sample of interest [39]. An opportunity is mixing the experimental

samples into a pooled sample, which may have the advantage of obtaining a larger library from fewer measurements as a result of samples with slightly different compositions. However, a disadvantage could be if the concentration of a peptide which occurs only in a few samples may be diluted below its detection limit with sample mixing. If a peptide is missing from the library, it can only be identified with an untargeted, library-free approach, however these methods are less sensitive [39].

Table 1. Summary of the often-used library-based and library-free peptide-centric software tools.

Library-based DIA analysis tools	**Applicable instruments**
OpenSWATH [41]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap, Bruker timsTOF Pro
Spectronaut [42]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap, Bruker timsTOF Pro
EncyclopeDIA [43]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap
Skyline [44]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap
PeakView [45]	SCIEX Triple TOF 5600/6600/6600+
Specter [46]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap
SWATHProphet [47]	SCIEX Triple TOF 5600/6600/6600+
PIQED [48]	Thermo Orbitrap
DIA-NN [49]	SCIEX Triple TOF 5600/6600/6600+, Thermo Orbitrap
Library-free DIA analysis tools	**Applicable instruments**
DirectDIA (Spectronaut)	SCIEX Triple TOF 5600/6600, Thermo Orbitrap, Bruker timsTOF Pro
Peaks [50]	Thermo Orbitrap
DIA-Umpire [37]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap
Group-DIA [51]	SCIEX Triple TOF 5600/6600, Thermo Orbitrap
PECAN [52]	Thermo Orbitrap

A recent study showed that DIA-NN [49] is one of the best performing software packages even if a predicted or an experimental library from GPF–DIA measurements was used [63]. DIA-NN is a software suite, which easy to use and apply a novel signal correction and quantification strategy and deep neural network to evaluate DIA data. Protein inference and peptide uniqueness can be performed within the software, therefore the results can be obtained at the level of genes, protein groups and precursors.

1.6. Applications of DIA in proteomics

It is widely known that compared to mRNA expression, protein expression correlates much better with the activity of proteins [64], therefore quantitative determination of proteins could be applied for functional study of proteins. LC–MS techniques are perfectly suited for this assignment, due to their high throughput and the ability to analyze complex protein mixtures. Deep protein profiling can provide a huge amount of information relevant

to many research areas where accurate, consistent, and large-scale protein quantitation data is needed. The list is not complete, but perhaps the most important, for example drug screening, biomarker discovery, and personalized medicine. For proteome quantification, DIA is a powerful tool, because of the high throughput and good reproducibility [65]. However, the gold standard for the quantitative measurement of proteins is the SRM and PRM methods [66,67]. For these strategies, the targets must be defined in advance. Information from DIA preliminary experiments can play a significant role in their selection.

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). Therefore, we aimed 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. Materials/reagents

Reagents, such as ammonium bicarbonate (AmBic), dithiothreitol (DTT), iodoacetamide (IAA), sodium dodecyl sulfate (SDS), sodium deoxycholate (SDC), and tris(hydroxymethyl)aminomethane (Tris) were purchased from Sigma-Aldrich (Darmstadt, Germany), acetone from Merck (Darmstadt, Germany), trypsin, and formic acid (FA) from

Thermo Scientific (Rockford, IL, US), octylphenoxypolyethoxyethanol (IGEPAL CA-630) from Alfa Aesar (Ward Hill, MA, US). Water, acetonitrile, ethyl acetate, acetic acid, and sodium chloride (NaCl) were delivered by VWR (Debrecen, Hungary).

Calibrated glass microcapillary tubes (20 µL) were manufactured by Drummond Scientific Company (Broomall, PA, United States), Schirmer's strips (I-Dew Tearstrips) by Entod Research Cell UK Ltd. (London, UK) and PRT (Zone-Quick test) by Showa Yakuhin Kako Co. Ltd. (Tokyo, Japan).

Stable isotope (13C and 15N on N-terminal K or R) labeled proteospecific peptide fragments of OATP1B3 (NQTANLTNQGK (305–315) and BCRP (SSLLDVLAAR (87–96) and LLSDLLPMR (474–482)) were ordered from JPT Peptide Technologies (Berlin, Germany). These SpikeTides™ TQL peptides contains QTag, which is a small chemical tag attached to the peptides, allowing robust and reproducible quantitation. The labelled peptides must be added to the samples before digestion, because the QTag is released by cleaving with trypsin and does not interfere with further analyses.

3.2. Samples

3.2.1. Tear samples

Five healthy female and five healthy male volunteers participated in the experiment with age between 20 and 33 years. All subjects included in this experiment were enrolled at the Department of Ophthalmology, University of Szeged. Approval for the human experiment was granted by the local Ethical Committee of the University of Szeged (108/2019-SZTE), and the experiment protocol adhered to the tenets of the most recent revision of the Declaration of Helsinki for experiments involving human subjects. All subjects enrolled in this experiment provided voluntary written informed consent.

For each subject, tear fluid samples were collected from both right and left eyes using three different sampling techniques during the same morning visit. More specifically, tear samples were collected (in the order of sampling) using glass capillaries (CAP), phenol red thread (PRT), and Schirmer's strips. For CAP sampling glass microcapillary tubes were introduced into the ventral cul-de-sac of the conjunctiva of the opened eyes for an average of 2 minutes without contact to the cornea, conjunctiva, and lower eyelid. PRTs and Schirmer's strips were placed over the lid margin at the junction of the lateral and middle thirds of the lower eyelids. The sampling lasted 15 seconds for PRTs and 5 minutes for Schirmer's strips while subjects closed their eyes without an anesthetic. In order to get tear samples from the resting eyes, there was a 30-min break between the different sample

collection methods. Samples were immediately placed into sterile plastic microtubes and frozen at −80 °C until the analytical procedure. During sample preparation, the Schirmer's strips were divided into lower (SL) and upper (SU) parts by cutting strips at the zero line and at the sign of 10 mm, respectively. The SL represent the rounded portions of the paper that were in direct contact with the eye surface, while the SU samples are the subsequent 10 mm sections. The rest of the strips were not processed.

3.2.2. Cell and membrane samples

Whole cell lysates were produced by an ultrasonic homogenization of 5 million HEK293 cells in 500 µL buffer containing 2% SDS, 1% SDC, and 2% IGEPAL CA-630.

Three biological replicates (A, B, C) of HEK293-OATP1B3-LV cell type were cultured on separated tissue culture flasks and passaged (recultured on a new flask) two times a week using a 0.25% trypsin–EDTA solution and incubated at 37°C in a humidified atmosphere containing 5% CO_2. For the expression stability experiment, all 3 biological replicates were sampled (3.75 million cells each) at the initial time, after 8 and 16 passages. The membrane protein fractions of these OATP1B3-overexpressing (gene: SLCO1B3) HEK293 cells were enriched using ProteoExtract® Native Membrane Protein Extraction kit (Merck) according to the manufacturer's protocol. The membrane-enriched buffer II fractions were used for the further analyzes. For digestion protocol comparison, samples were produced from BCRP overexpressing HEK293 cells with the same kit.

All BCRP-overexpressing human membrane vesicle preparations (BCRP-HEK293, BCRP-M) and human BCRP-overexpressing insect vesicles (BCRP-Sf9, BCRP-Sf9-HAM) contained the human wild-type (482R) version of the ABCG2 transporter (Accession number NM_004827). Membrane vesicles were prepared from cells according to the method described by Sarkadi *et al*., [68]. Briefly, the cells were lysed using a glass-Teflon tissue homogenizer in a mannitol containing buffer and after debris removing the membrane fraction was obtained via ultracentrifugation (60 min at 100,000 x g). HAM (high activity membrane) vesicles were cholesterol-loaded during the process of preparing membranes from Sf9 cells [69].

All cell and vesicle samples were obtained from SOLVO Biotechnology (Szeged, Hungary, http://www.solvo.com). The samples were stored at −80°C until the day of use. The protein contents of the samples were determined using the BCA Protein Assay (Thermo Scientific) following the manufacturer's protocol. Aliquots of 10 µg or 25 µg protein were used for each sample preparation.

Vesicle (n=5), ProteoExtract (PE, n=4) and TC (n=4) samples of HEK293 cells were used to test the membrane preparation methods.

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

All the SL, SU, and PRT samples were extracted individually by adding 100 µL 5% SDS in AmBic (pH 8.0, 100 mM) containing 0.25% protease inhibitor cocktail (Sigma-Aldrich, Rockford, IL, United States). The samples were vortexed and incubated in a thermal shaker at room temperature for 1 hour at 350 RPM. After a centrifugation at 12 000 x g, 10 min, 4 °C, the supernatants of the samples were transferred into new tubes. The capillary tear samples were diluted using AmBic (pH 8.0, 100 mM) containing 5% SDS and 0.25% protease inhibitor cocktail before determining the protein concentration.

To examine the protein extraction efficiency from PRTs, different extraction solutions were tested. A pooled tear sample was collected in a low protein binding tube 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. The protein contents of the threads were extracted using 100 µL of 1) 5% SDS in AmBic (pH 8.0, 100 mM), 2) AmBic (pH 8.0, 100 mM), 3) 1% acetic acid in water. All extraction liquids contained 0.25% protease inhibitor cocktail and were tested on 3 replicates. The samples were incubated in a thermal shaker at room temperature for 1 hour at 350 RPM. After a centrifugation at 12 000 x g, 10 min, 4 °C, the supernatants of the samples were transferred into new tubes.

3.4. Digestion protocols

Prior to digestion, the protein contents of all samples were determined using BCA Protein Assay (Thermo Scientific, Rockford, IL, United States) according to the manufacturer's protocol. All digestion methods were tested in 3 technical replicates.

3.4.1. Surfactant Cocktail-Aided Extraction/Precipitation/On-Pellet Digestion (SEPOD) method

For all tear samples and corresponding cell and membrane samples 10 µg protein were processed based on a modified on pellet digestion (SEPOD) [70]. Briefly, the samples were reduced with 10 mM DTT at 60 °C for 30 min and alkylated with 20 mM IAA in dark at room temperature for 30 min. The protein content was precipitated by adding a 7-fold volume of ice-cold acetone and incubated at −20 °C overnight. After centrifugation with 15 000 x g, 10 min, 4 °C the supernatant was discarded. The protein pellet was washed twice

with 500 µL acetone/water (85:15, v/v) mixture. After centrifugation with 14,000 x g, 10 min, 4 °C, the protein pellet was dissolved in 15 µL RapiGest SF Surfactant (Waters, Milford, MA, United States) and was incubated at 100 °C for 5 min. After being cooled to room temperature, 65 µL AmBic (pH 8.0, 100 mM) and 0.25 µg trypsin in 5 µL AmBic (pH 8.0, 100 mM) were added to the mixtures. The samples were incubated at 37 °C for 30 min and another 0.25 µg trypsin in 5 µL AmBic (pH 8.0, 100 mM) was added, and the mixture was digested at 37 °C for 5.5 hours. Digestion was stopped by the addition of 1 µL concentrated formic acid. After centrifugation with 12,000 x g, 10 min, 4 °C the supernatant was injected to the LC–MS system.

Samples prepared for targeted measurements were spiked with stable isotope labeled (SIL) peptides (1 pmol/peptide) prior to digestion.

3.4.2. Filter Aided Sample Preparation (FASP) method

For the corresponding 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 [71]. Briefly, the samples were extracted five times with 1 mL of ethyl acetate to remove the detergents. Proteins were reduced by the addition of DTT in 4% m/V of SDS to a final concentration of 50 mM followed by incubation at 95 °C for 5 min and cooling at room temperature for 10 min. The Microcon Ultracel 30 kDa (Millipore, Burlington, MA, USA) units were prepared by spinning two times 100 µL water through the filter at 12,000 × g 10 min. The samples were transferred to the filters and spun at 12,000 × g for 10 min. Buffer exchange was performed in two successive washes with 8 M urea in 100 mM Tris–HCl pH 8.5 with a 10 min centrifuge at 12,000 x g. Proteins were then alkylated in 50 µL 50 mM of IAA for 20 min in darkness at room temperature. Samples were spun at 12,000 × g for 10 min and washed two times 100 µL urea. Urea was removed by two washes with 100 µL Tris–HCl with centrifugation at 12,000 × g for 10 min. Protein digestion was performed by adding 100 ng of trypsin (250 ng for targeted BCRP measurements), in 40 µL 50 mM AmBic and incubating overnight at 37 °C. Peptides were recovered using an initial spin of 12,000 × g for 10 min followed by two centrifugations with 50 µL of 0.5 M NaCl and 100 µL AmBic. Digestion was stopped by the addition of 3 µL concentrated FA. Recovered peptides were evaporated to dryness with gentle stream of nitrogen. Samples were resuspended in 91 µL (60 µL for targeted BCRP measurements) of the starting eluent.

3.5. NanoLC–MS measurements

NanoLC–MS/MS analysis was carried out on a Waters ACQUITY UPLC M-Class LC system (Waters, Milford, MA, United States) coupled with a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). Symmetry® C18 (100 Å, 5 µm, 180 µm × 20 mm) trap column was used for trapping and desalting the samples. Chromatographic separation of peptides was accomplished on an ACQUITY UPLC® M-Class Peptide BEH C18 analytical column (130 Å, 1.7 µm, 75 µm × 250 mm) at 45 °C by gradient elution (linear gradient from 3% solvent B to 40% solvent B in 70 min, followed by a 30 min washing and equilibrating gradient for DIA measurements, while linear gradient from 3% solvent B to 40% solvent B in 18 min, followed by a 14 min washing and equilibrating gradient for PRM measurements). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid were used as mobile phases at a flow rate of 350 nL/min. The sample temperature was maintained at 5 °C. The injected volumes were 5 µL for cell and membrane samples, while 2 µL for tear samples. The mass spectrometer was operated using the equipped Nanospray Flex Ion Source.

Data acquisition was performed using Xcalibur™ 4.1 (Thermo Fisher Scientific, Waltham, MA, United States).

3.6. DIA acquisition and processing

Sample-specific pool samples were produced by mixing experimental samples prepared in the same way for both tear samples and cell/membrane samples. The data required to generate sample-specific spectral libraries were collected from LC–MS analysis of the aforementioned pool samples. The mass spectrometer was configured to acquire six gas-phase fractionated (GPF–DIA) acquisitions with 4 Th precursor isolation windows at 17,500 resolution to hit an AGC target 1×10^6 with a maximum inject time of 120 ms. In the GPF–DIA measurements, an overlapping window pattern was adjusted from the previously optimized mass ranges (395–505, 495–605, 595–705, 695–805, 795–905, and 895–1005 Th).

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 at 17,500 resolution to hit an AGC target 2×10^5 and the maximum inject time was set to auto.

DIA-NN 1.8 software [49] was applied for database search. First, a predicted spectral library was generated from the Uniprot Human Reference FASTA proteome (20,575 genes,

one protein per gene) with the following settings: cysteine carbamidomethylation was set as a constant and methionine oxidation as a variable modification, the N-terminal methionines were excised, the maximum number of missed cleavages were set to two with a maximum of two variable modifications per peptide, the precursor m/z range was 380 to 1020 Th. 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. All the individual samples were searched against the combined spectral libraries to detect all peptides and proteins that can be quantified. DIA-NN was configured with the default settings with the following modifications: the maximum number of missed cleavages and variable modifications were set to 2, methionine oxidation was added, and precursor m/z range was adjusted to 380-1020 Th. Search results were filtered for 1% FDR rate at the level of unique genes. Further data processing, statistical analysis and creation of figures were carried out in Perseus (version 1.6.15.0) [72], Instantclue (version 0.1.1) [73], InteractiVenn [74] and Microsoft® Excel® 2016 software.

For tear samples, only proteins identified with at least 2 unique peptides were included in all quantitative statistical analysis. Because of the large biological variance among samples, for better comparability, not the raw intensities of individual proteins but their relative proportion in the samples were used. Therefore, intensities of the proteins were divided by the sum of intensities of all proteins as normalization and log2 transformation was performed on all datasets.

The analyzed proteins were annotated from Gene Ontology [75] and UniProt (https://www.uniprot.org). Proteins that had the “membrane” Uniprot Subcellular Location term were defined as membrane proteins.

Relative methionine oxidation (%) was calculated for LLSDLLPMR peptide using the following equation: relative oxidation (%) = intensity of the methionine-oxidized peptide / (intensity of the monoisotopic mass of methionine-oxidized peptide + intensity of the methionine-non oxidized peptide) × 100 [76].

3.7. 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 at 17,500 resolution to hit an AGC target 2×10^5 with a maximum inject time of 60 ms. To reach the maximum

sensitivity, precursor ions were fragmented using optimized collision energies. Skyline 20.1 [44] was used for targeted data evaluation.

To generate calibration samples, overexpressing HEK293 digests were diluted with control HEK293 digest at 5 different ratios (5-fold to 100-fold dilution). The peak areas of the peptides and dilution factors (1 / n dilution fold) were used to construct the 7-point calibration curves for OATP1B3 peptides. The limit of detection (LOD) and limit of quantitation (LOQ) of each peptide were calculated based on the standard error of the intercept. LOD and LOQ were calculated using the formulas $3.3 \times \alpha / S$ and $10 \times \alpha / S$, respectively, where α is the standard error of the y-intercept and S is the slope of the calibration curve.

The absolute amount of BCRP and OATP1B3 proteins was calculated as follows: amount of target protein (pmol / mg protein) = endogenous / heavy peptide peak area ratio × heavy peptide amount (pmol) × (1 / digested amount of proteins (mg)).

Comparisons between the biological replicates and between the different time points were conducted using Welch T-test and one-way analysis of variance (one-way ANOVA). A p-value of <0.05 was considered statistically significant.

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

4.1. Theoretical background

The paradigm change from reactive to predictive, preventive, and personalized medicine requires novel and reliable methods that can provide a more precise diagnosis and patient stratification, detect early disease, elucidate the risk of disease, and predict disease outcome, response to therapy, and permit monitoring of therapeutic management. To this end, the discovery and validation of specific biomarkers/biomarker panels would be a promising approach. The source of biomarkers is crucial for the specificity, sensitivity, accuracy, and reliability of diagnostic tests and treatment targets [77]. 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 [78]. Tear fluid contains proteins, such as enzymes, mucins, hormones, immunoglobulins, growth factors, neuropeptides, and cytokines along with lipids, salts, and carbohydrates [79]. This comprehensive biomolecule repertoire in human tears serves as a good source for biomarker discovery for diseases. The biggest advantage is that tears are proximal to the disease location (such as ocular surface disease, lacrimal gland disease, etc.) contrary to, e.g., cancer

biomarkers in blood, where the related biomarker molecules could be distant from the source, and are highly diluted [80,81]. The tear fluid proteomic profile has been found to provide basic biological information for many ocular diseases, such as dry eye syndrome, blepharitis, keratoconus, thyroid eye disease, vernal keratoconjunctivitis, diabetic retinopathy, and primary open angle glaucoma [77,82]. Even though tear components are mainly derived from secretory glands, such as the lacrimal glands, the change of tear film composition is not only regulated by its secretion units. Protein molecules can enter the tear fluid through conjunctival blood vessels and, due to the overlap between the tear and plasma proteome, there may be opportunities to observe systemic responses in the tears. Several such examples thus far include breast cancer, type 2 diabetes, Alzheimer's disease, multiple sclerosis, and rheumatoid arthritis [77,82]. Some systemic diseases may affect the eye so we can use 'tears' as a 'window' to assess systemic as well as ocular 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, threads, ophthalmic sponges, and polyester rods [83]. During capillary sampling, one end of the glass capillary is placed in the meniscus of the tear fluid and due to the capillary action, the tear flows from the conjunctival sac into the interior of the glass capillary. The Schirmer's strip was originally a standard clinical tool used in many places to measure tear fluid volume, but it was found to be suitable for collecting tear samples too. 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 (e.g., aqueous tear-deficient patients or experimental animals). Kurihashi *et al.* established a method of tear secretion measurement using fine cotton threads [84] which might meet the above requirements, and already proved to be an applicable sample collection method for analysis of small molecules [85].

The advancement in nano-scale liquid chromatography coupled MS (nanoLC–MS) that provides improved chromatographic separation of peptides, higher sensitivity, and extended dynamic ranges to identify > 1500 proteins, has opened up the possibility of tear biomarker research [82]. There are numerous studies and reviews in the literature on the investigation of protein profiles from differently collected tear samples using liquid chromatography coupled to mass spectrometry (LC–MS) [77,78,80,81,83]. It was shown

that samples collected by Schirmer's strip and capillary method have large quantitative and qualitative differences in their protein composition [82], therefore any new sample collection method should be evaluated in that respect too.

4.2. Results

4.2.1. Comparison of the effectiveness of sampling methods

Our goal was to introduce a reliable new sampling procedure, which provides us with a convenient way of collecting enough tear samples for quantitative proteomic studies. In this comparative experiment tears of the same 10 healthy donors were collected by three sampling methods and analyzed by the same proteomics procedure. General protocols for the use of glass capillary tubes (CAP) and Schirmer's strips [86] were followed, and PRT as a new tear sampling device was included in this experiment. After sample collection, Schirmer's strips were divided into a lower section (SL) which was in direct contact with the surface of the eyeball and eyelid, and the following 10 mm long upper section (SU) and processed separately.

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. The total volume or protein content of all PRT, SU, and SL samples was satisfactory for our protocols; however, one SL sample was damaged during processing. In the case of two donors, the volume of tears collected with glass capillary tubes was insufficient (less than 1 µL) to determine the total protein content and perform proteomic sample preparation.

Both the volume and the total protein content of tears collected with the CAP sampling procedure showed large variance (7.9 ± 7.0 µL and 64.1 ± 44.9 µg, respectively). Tears are collected slowly and erratically, with interruptions due to blinking and eye movement, what—apart of individual variance—may explain the large fluctuation [80]. The total protein content after the extraction of SL strips, 10 mm long SU strips, and whole PRT samples proved to be 42.0 ± 11.4 µg, 57.8 ± 15.5 µg, and 28.9 ± 11.2 µg, respectively.

4.2.2. Evaluation of the proteomics protocol

The nanoLC–MS reproducibility was determined by triplicate injection of a randomly selected PRT sample. The number of identifications and the reproducibility of the relative intensities were compared at both protein and precursor levels; 3003 ± 21 precursors were quantified with the median CV of 0.17 in these measurements, corresponding to 548 ± 4

proteins with the median CV of 0.15. Based on these results, our procedure was suitable for the study of sample preparations and sampling procedures.

In order to evaluate the efficiency and reproducibility of protein extraction from PRT samples, three different extraction solutions (5% SDS, 100 mM AmBic, and 1% acetic acid in water) were tested. Nine PRTs were soaked in a pooled capillary tear sample to ensure the same initial protein concentration and composition. Note that this experiment was limited by the number of proteins detectable from capillary samples, but comparison to the original pooled capillary sample provides satisfactory data on recovery and reproducibility for major tear proteins. There were no significant differences observed in the total amount of extracted proteins as determined by BCA assay (Table 2); 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.

Table 2. Comparison of tear protein extraction methods from PRT.

	Total amount of proteins (µg)	**Number of quantified proteins**	**Percentage of the total number of proteins**	**Median CV of intensity**	**Pearson correlation to capillary**
Capillary	–	236	100.0	0.17	–
SDS	41.5 ± 3.3	198	83.9	0.19	0.941
AmBic	36.4 ± 1.16	203	86.0	0.15	0.946
Acetic acid	40.9 ± 6.78	152	64.4	0.23	0.851

Using either the SDS or AmBic protocol, 84–86% of proteins can be recovered, and Pearson correlation with the original sample demonstrated, that the protein composition was not biased, the extract correctly demonstrates the original composition. Although AmBic extraction seems to be suitable for protein extraction from PRT samples, the presence of a tenside (e.g., SDS) may help the extraction of the intracellular fraction that was excluded by our evaluation approach. Because of that, and for comparability with the protocol applied for Schirmer's strips, we used the SDS approach in our analysis.

4.2.3. Impact of MS data acquisition and evaluation method

Mass-spectrometry-based proteomics enables us to identify and quantify hundreds of thousands of proteins from different samples. However, the quality of the results is highly dependent on the experimental and computational workflows. A large number of proteomics software tools and algorithms have been published for data-independent acquisition (DIA)

proteomics data processing (detailed in the Introduction section). We applied DIA-NN software for spectral library creation, protein identification, and quantification [87]. Our DIA-NN workflow first was set up to use sample-specific spectral libraries generated by refining predicted libraries using six gas-phase-fractionated acquisitions (GPF–DIA) with 4 Th precursor isolation windows from all four types of pooled samples. These sample-specific spectral libraries demonstrate the deepest proteome coverage of a given sample type. A combined spectral library was built by searching all four groups of GPF measurements in one against the same predicted spectral library. This combined library contains a total of 2583 proteins (1742 with at least two unique peptides), any of which were present in a quantifiable amount in at least one pooled tear sample.

Data of DIA acquisitions collected from the analysis of individual samples (DIA-Q) were searched against both the sample-specific and combined libraries. The number of peptides that could be quantified in SU samples increased to the greatest extent (by 117.5%), as the 3312 peptides detected with sample specific library increased to 7205 using the combined library. A smaller but significant increase was observed for CAP (84.7%) and PRT (62.0%), where the measured number of peptides increased from 834 to 1540 and from 3322 to 5380. In the case of SL samples, only a 0.2% increase was observed from 8337 to 8354 quantified peptides. Figure 4 shows the distribution of the detected peptide among the different sample types. The use of a combined library facilitated better comparability by reducing the number of proteins and peptides that could only be quantified in one sample type.

It was also investigated that in how many percent of the samples were the peptides measurable that could only be quantified using the combined library but not with the sample-specific ones to evaluate their quantitative value. These peptides could be measured on average in 53.6 ± 31.4% of CAP samples, 20.2 ± 17.9% of SU samples, while 24.5 ± 20.8% of PRT samples. As for the SL samples, the peptides which were identified with the combined library only were measured in 40.2 ± 28.1 % of the samples, but the number of these peptides was negligible compared to the number of peptides detected with both types of libraries. Regarding the SL samples, the peptide increase was measurable in 40.2 ± 28.1 % of the samples, however, the number of these peptides was negligible compared to the number of peptides detectable with both types of libraries. A very similar result was obtained when the examination was carried out at the level of proteins, in which only proteins quantified with at least 2 proteotypic peptides were included.

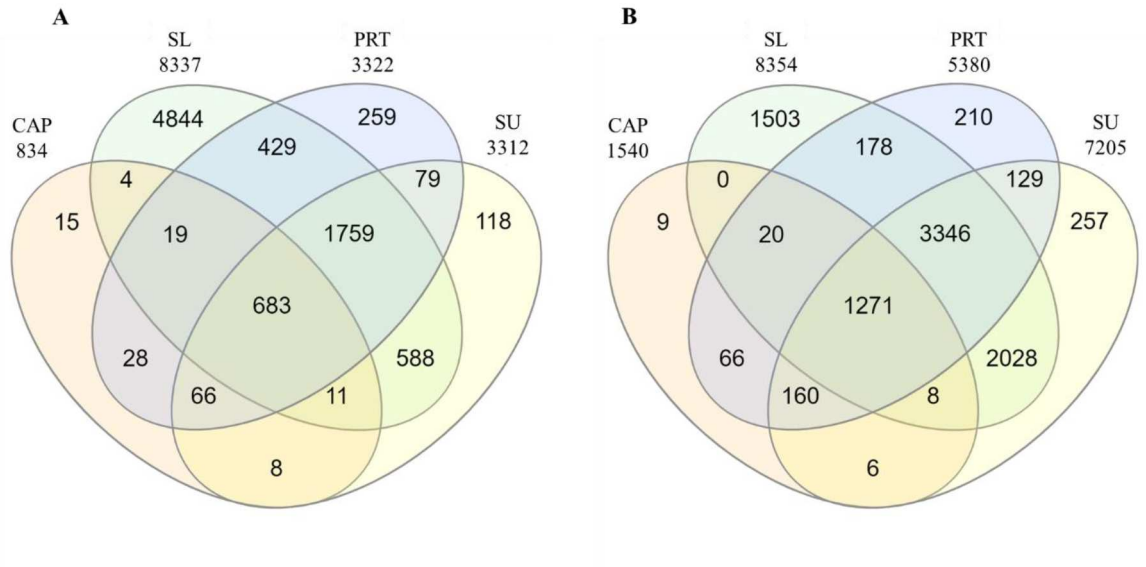


Figure 4. The number of quantified peptides in different types of tear samples by searching against the sample-specific library (**A**) or the combined library (**B**).

Although the combined library originating from the GPF–DIA measurements from four pooled samples contain 2583 proteins (1742 with at least two proteotypic peptides), in the individual samples 1144 could be quantified with at least two peptides in the DIA-Q measurements.

4.2.4. Comparison of the tear proteomes of different sample types

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. All quantifiable proteins were included in this comparison, regardless of the number of peptides detected. The dynamic range and contribution of proteins in different types of samples to the combined library are shown in Figure 5. The library covers proteins with a range of more than six orders of magnitude in summed signal intensity. Number of proteins shown on the Venn diagram refers to number of protein groups, which may consist of more than one protein, and cannot be differentiated based on identified peptides. The lowest number of proteins could be identified in the pooled CAP sample (422 protein), while in the two indirect pooled tear samples (PRT and SU), a similar number of proteins could be detected (1439 and 1225) with a much higher overlap (1092 shared protein identifications). The SL samples were only included in the investigation to detect possible contaminants originating from the strong direct contact of the Schirmer's strip with the ocular surface. This assumption is supported by the presence of nearly 1000 proteins, which were identified only in this type of

pooled sample. These proteins are present in low abundance, according to their summed intensity (green dots in Figure 5). It must be noted that the parts of the PRTs in contact with the eye were not removed, which may explain the approximately 300 proteins shared by only PRT and SL samples (red dots in Figure 5). The majority of the most intense (106–1010) proteins could be identified in all sample types, but there are many common proteins with lower intensity. Altogether, 341 protein groups could be detected in all the four sample types (ochre dots in Figure 5).

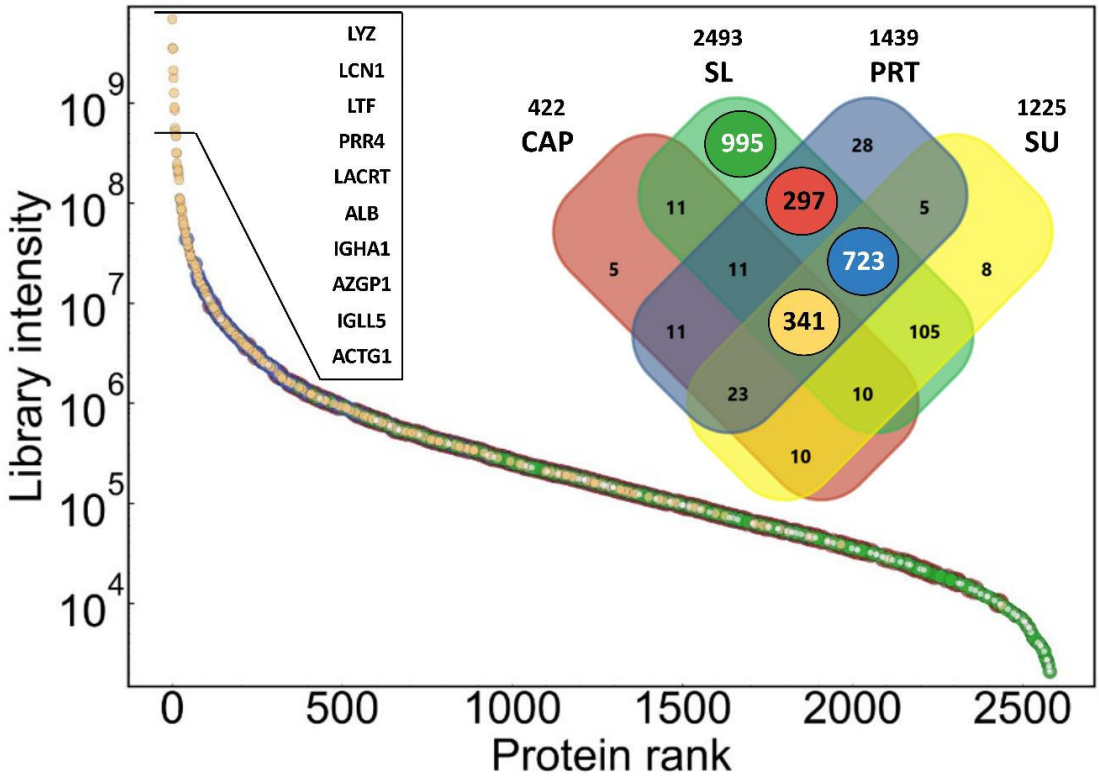


Figure 5. Summed intensity of proteins in combined spectral library as a function of protein intensity rank, demonstrating the dynamic range of identified proteins. Data points are colored according to highlighted subsets of the library as shown in the Venn diagram insert, to demonstrate contribution of different sample types to the library. Gene names of the top proteins are shown in the order of intensity rank.

To study the quantitative similarity of different sample types, Pearson correlation coefficient values were calculated for the proteins quantified in sample type-specific pooled samples (Figure 6). This assay shows a strong correlation between PRT and SU samples (r = 0.90), but the pooled CAP sample also correlates satisfactorily with these samples (r = 0.76–0.78).

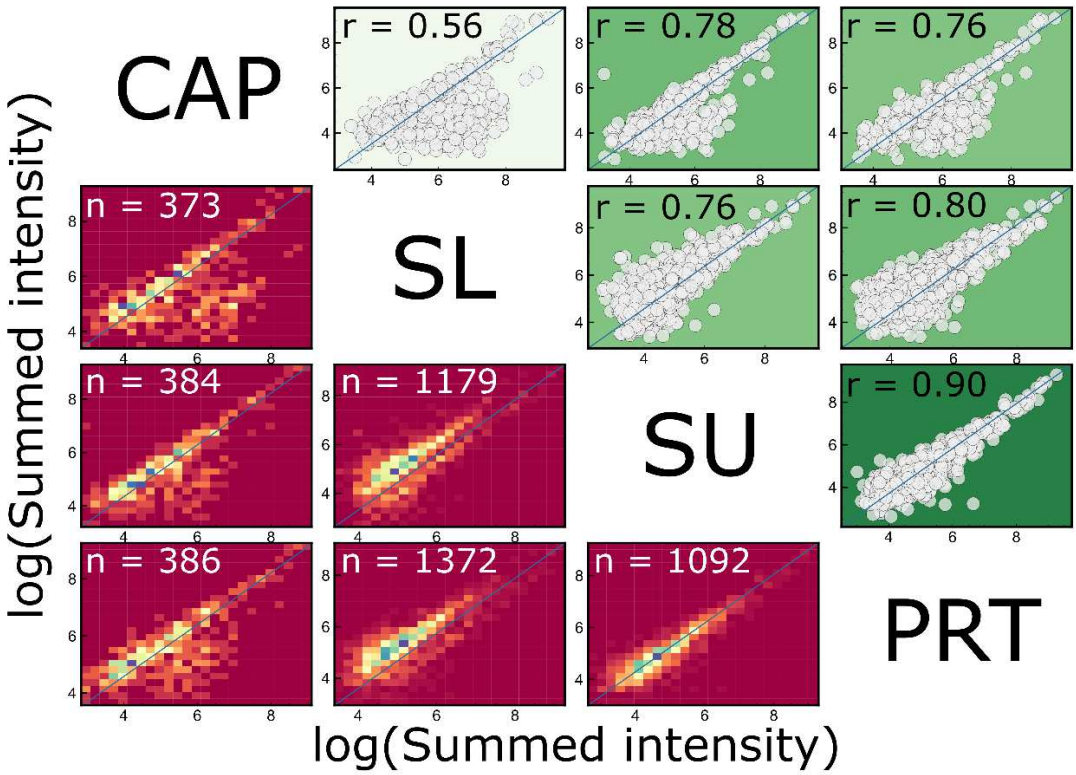


Figure 6. Correlation of intensities of proteins in different pooled samples. The depth of the background colors of scatter points on top-right is proportional to the value of actual Pearson correlation coefficients, which are shown on the 2D histograms at bottom-left. The number of the common sets of proteins (n) are given.

4.2.5. Clustering of proteins in different types of samples

The analysis of the spectral library created from the analysis of four different types of pooled samples in the previous section already highlighted the diversity of proteomes determined in samples collected by different methods, but the statistical classification of proteins based on sample-specific detectability is only possible on analysis of individual samples. It must be noted that the applied DIA approach, with match-between-run identifications enabled, can minimize the technical reasons for missing values, therefore this analysis reflects heterogeneity in sample composition.

A k-means cluster analysis — based on sample-type specific detection frequency (in %) — was performed to classify the 1144 quantifiable proteins into four clusters (the number of clusters was predicted using the elbow method [88]) (Figure 7).

In Cluster A (pink), there were 195 proteins that could be measured with high frequency in all sample types (86–92% of the samples within any sample type). Cluster B (dark green) includes 242 proteins that were repeatedly measurable (84–98%) in samples

from indirect procedures (SL, SU, and PRT), but they were quantifiable in only a few CAP samples (8%). 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. Using this approach, proteins within the latest cluster may be marked as possible contaminants of tear samples. They have been found mainly in the SL samples, but rarely in any other sample group. On the other hand, proteins of Cluster A and B may be considered as common tear proteins, with the remark that CAP sampling provides reproducible detection for members of only Cluster A. It must be noted that PRT sampling produces the highest detection rate for proteins of those two clusters (see distributions on boxplots in Figure 7).

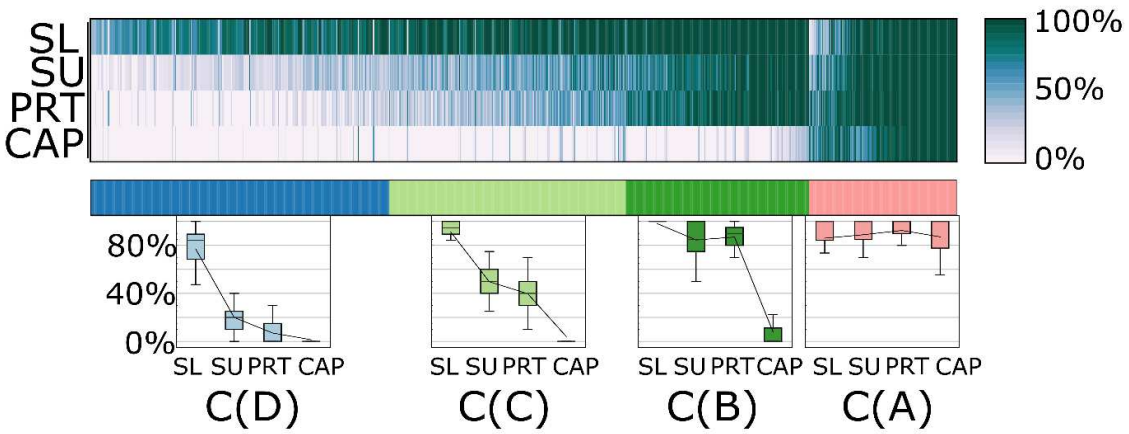


Figure 7. k-means (k = 4) cluster analysis of the quantified proteins based on the percentage of observations in samples of the different sampling procedures. Detection frequency of each protein is given as a heatmap on top. The boxplots on the bottom show the distributions of detectability of proteins in different sample types within each cluster.

We have compared the summed relative intensities of the identified protein clusters in each type of sample (Figure 8). As expected, the highest level of the possible contaminating eye-surface proteins (395 and 317 proteins in Clusters D and C, respectively) were found in the SL samples, where on average 6.0% of the total intensity of all proteins is given by the proteins of Cluster D. CAP, SU, and PRT samples contain an average of 1.2, 1.5, and 0.7% proteins of that cluster. Proteins of Cluster C can be found at higher abundance in each sample type. Their values are still less than 5%, except for the SL samples (13% of total intensity). Proteins of Cluster B were detected only in a few CAP samples, and their intensity was low (on average 2%), while their relative abundance was highest in SL samples again (32% average). 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. For all clusters, the variance of PRT samples is lower than those in SU, but slightly higher than in CAP samples, which, however, have a less complex protein composition.

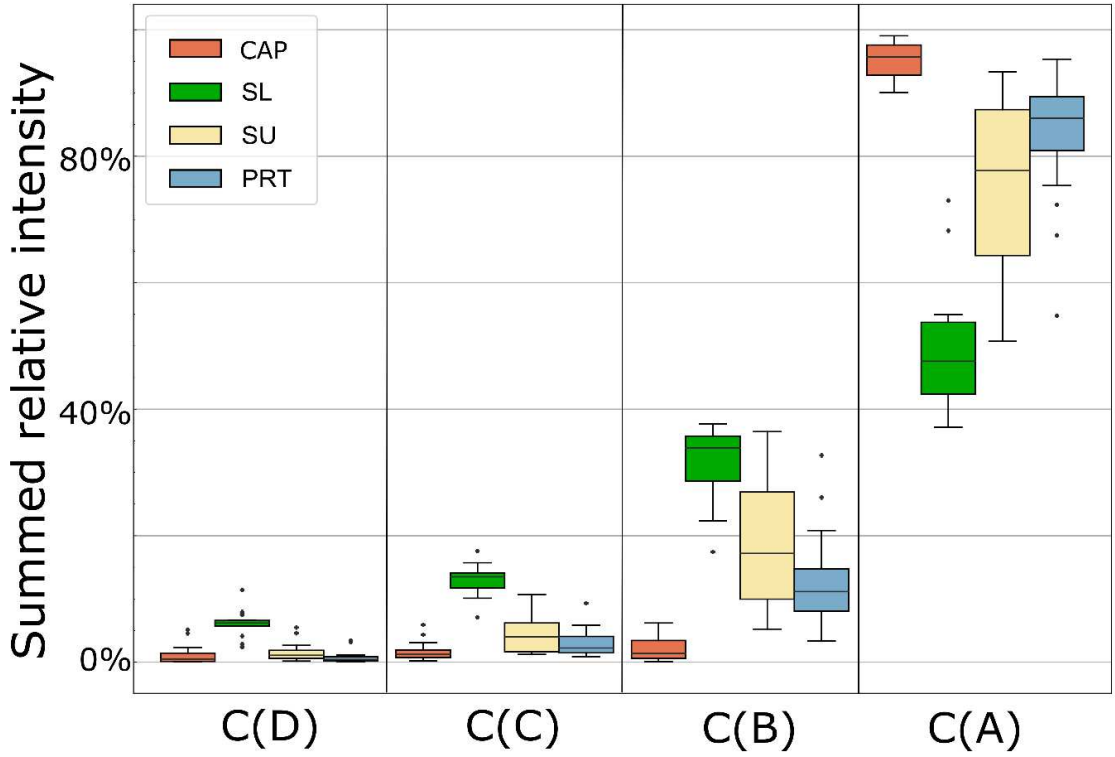


Figure 8. Boxplot of summed relative intensity of the four protein clusters in each sample type (nCAP = 18, nSL = 19, nSU = 20, nPRT = 20).

4.2.6. Classification of tear proteins

Based on the detection frequency of proteins in samples collected using different sampling methods, we identified four clusters of proteins in the previous section. In order to give biological classification of the proteins in tear samples, each protein was annotated with available GO and UniProt terms. Fisher's enrichment analysis was performed on these terms to identify common properties of proteins within each cluster. Here we discuss a few examples of the significantly enriched categories (Benjamini–Hochberg corrected FDR < 2%), which are different in different clusters and relevant to tear and tear sampling.

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. Eighty-five percent of the different Ig chains (36 of 40) detected in our samples are in Cluster A, which makes 17% of the proteins in this cluster. Most of the identified keratins (10 out of 12) frequently occur in all sample types, thus they are found in Cluster A.

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 [89] database collects a list of proteins identified in different parts of the human eye. For the classification of proteins in tear samples based on their possible origin, two groups were created: proteins which can be found in the 'Tears' section (including 1506 proteins) and an eye surface group from the 'Cornea' or 'Sclera' sections of the EyeOME database (1469 and 1895 proteins respectively). There is a large overlap in these assignments, 1213 proteins are common to tears and the eye surface in that database. Immunoglobulins are excluded from EyeOME, that is the main reason that not all, but 1093 proteins of the database were among the 1144 quantified proteins from our experiments, 846 of which are common to tears and the eye surface. Based on the overlap of clusters identified in Section 4.2.5 and the EyeOME assignments (shown in Figure 9), we can make further refinements of protein classification.

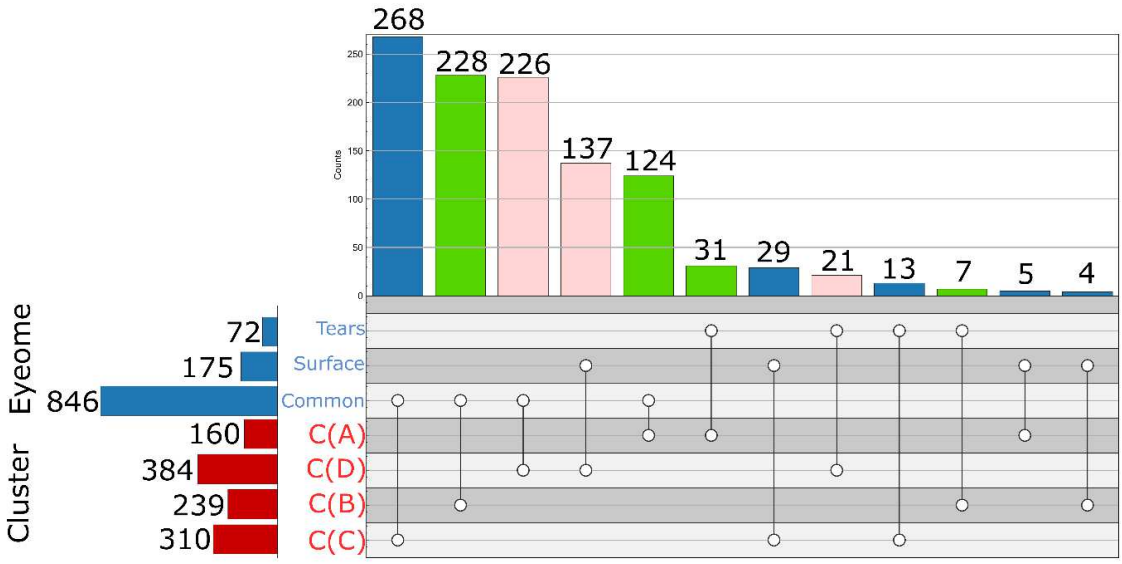


Figure 9. Count plot representation of the intersections of protein clusters and sections of the EyeOME [89] database. Clusters are those identified in Section 4.2.5. Proteins in the 'Cornea' or 'Sclera' section of EyeOME were classified as Surface.

The majority 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 (green in Figure 9).

4.2.7. Intra- and interpersonal variances of different tear samples

The differences in the composition of the tear samples collected from different persons, or from the two eyes of the same subject, may be a combination of several simultaneous reasons, including the effect of the sampling procedure on eye surface and tear secretion. In order to identify sampling induced effects, the protein composition of tear samples of the same person was compared using different methods, excluding the highly contaminated SL samples. Similar distributions of absolute protein abundance differences were observed in all sampling methods, however the Pearson correlation of intensities showed marked differences in the three tear fluid samples. The strongest intrapersonal correlation was found in the CAP samples (average of coefficients was 0.91, median 0.92), while PRT samples (average of 0.77, median 0.81), and SU samples (average of 0.73, median 0.76) presented weaker correlations. (Figure 10-A). The ratio of the protein MS intensities measured in the two eyes relative to the average of the eyes was also calculated to represent the differences between the eyes. The log2 transformed distribution is the narrowest around 0 in the case of CAP samples, and widest is in Schirmer's samples (Figure 10-B). Eleven percent of data points has an absolute value higher than 1 (at least twofold difference relative to eye average) in the case of CAP samples, in Schirmer's samples 21% and in PRT samples 14% has that high difference.

We calculated the overall variance (CV) of protein intensities in the whole sets of sample types. The CV distribution of the common 195 proteins of all sample types (Cluster A) are shown in Figure 10-C. According to this, these proteins have a lower median CV in PRT samples (64%) than in either SU (77%) or CAP samples (70%).

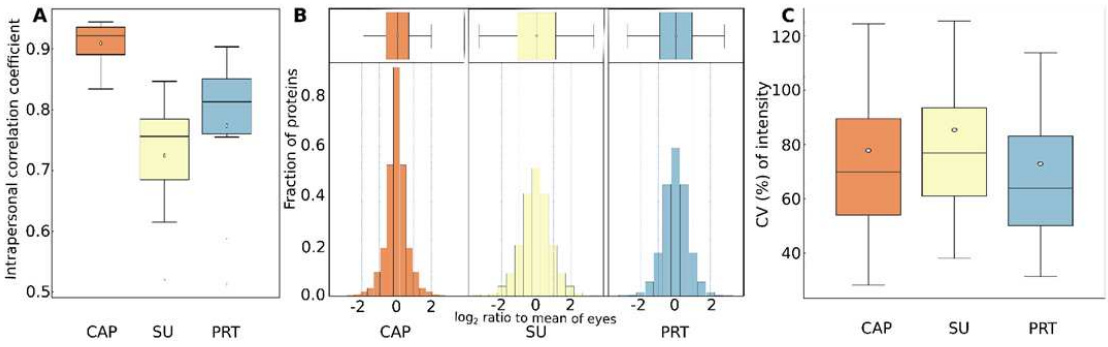


Figure 10. (**A**). Intrapersonal (left-to-right eye) Pearson correlation coefficients of samples of different sampling methods. (**B**). Distribution of protein log2 transformed intensity ratios measured in the two eyes relative to the average of the two eyes. (**C**). Distribution of coefficients of variation (CV %) for intensities of proteins in Cluster A in all samples of different sampling methods.

4.3. Discussion

Protein concentrations in normal tear fluid range from 6 to 11 mg/mL [81], but the protein concentration and composition of tear fluid samples are greatly influenced by the following factors [83,90]: (i) tear collection device (glass capillary tubes, Schirmer's strips, threads, ophthalmic sponges, and polyester rods; (ii) types of collected tears (reflex, or non-stimulated tears), which might be affected by the tear sampling procedure, irritating stimuli like environmental fluctuations, physiological status or rubbing the skin with alcohol, anesthesia; (iii) location of tear sampling (the inferior temporal tear meniscus near the external canthus of the eyes, or the inferior conjunctival sac); (iv) whether sampling occurred from open or closed eye; (v) last, but not least, the sample processing (recovery from the sampling device) and the further analytical procedure and data analysis. As can be seen from the above, tear sampling method is definitely a major challenge and has the greatest significant influence on the precision and reproducibility of the analytical results.

Several tear sampling methods are available, and every sample collection method used must be assessed since it has a significant influence on the precision and reproducibility of the analytical results. Each sampling method has advantages and disadvantages; therefore, 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, threads, ophthalmic sponges, and polyester rods. For determination of the protein profile of the tear sample the capillary tube and the Schirmer's strip are used most frequently [83,90].

As both methods are routinely used for measuring tear volume in ophthalmology, and in many tears proteomics studies, much information has been gathered on their application. Phenol red thread, like the previous two tests, is a widespread clinical test for measuring tear volume [91,92]. A comparison of the capillary, Schirmer's strip and the new PRT tear collection methods is provided in Supplementary Table 1.

Capillary samples contain a higher percentage of proteins originating from extracellular region, protein containing complexes, and membrane [80,86,93]. In biological processes, immune response, complement pathway, and tissue development proteins dominate more frequently in capillary samples. In contrast, an increased number of cell- and organelle-specific (intracellular) proteins contaminating tear samples have been reported in samples collected with Shirmer's strips [86], due to the contact of the paper with the highly vascularized conjunctiva and the possibility of injure its surface and microvasculature.

Collection of tears with Schirmer's strip is an indirect method carried out using absorbing supports. Proteomics analysis of these kinds of samples requires either centrifugation of tear fluid or extraction of proteins (before digestion) or peptides (after in-strip digestion) from the paper strips. A disadvantage of this absorbent-based sampling method is that different extraction procedures may result in varying protein profiles. It was recently reported [94] that the elution of proteins from Schirmer's strips varies significantly between different brands of the filter paper because of their distinct absorptive properties. Apart from the variety of clinical procedures used to perform Schirmer's test, it seems likely that one of the causes of the variability of Schirmer's test between studies is related to the use of different Schirmer's test strips. Unfortunately, there is no standardization of commercial strips, even though the need for standardization was recognized over fifty years ago. The maximum volume of absorption on 10 mm long piece of Schirmer's strip was found to be around 9 μL [95], but it may also depend on strip material. Phenol red thread is produced only by one company (Showa Yakuhin Kako Co., Ltd., Tokyo, Japan), so using the same tear collection device in different clinical and laboratory studies would reduce potential variables in tear analysis. Phenol red thread collects around 3 μL of tear fluid [96], and this volume is low compared to an average basal tear volume. The inability of PRT to absorb the entire volume of basal tears means that the collected initial sample is almost pure basal tears; however, it can be used for the collection of reflex tears if applied after stimulus [96], which can be a big advantage over other tear collection methods.

In our experiments, both absorbent-based approaches (SU and PRT samples) allowed the collection of low-volume samples, and the amounts of proteins were sufficient to perform

several proteomic analyses even from those two donors whose CAP tear samples were not sufficient. The safe application of capillary approach without touching the eye surface, requires larger tear volume with a thick tear film. This makes it less suitable for the collection of tear samples from aqueous tear-deficient patients [93,97], but there is no such limitation with the indirect methods, where the soft sampling material is immersed in the tear film.

Sample processing is one of the most crucial processes in proteomics research; this pre-analytical step can bias both protein composition and quantification. Absorbent-based tear collection requires a pre-analytical step to elute/extract proteins from a paper strip, sponge, thread, etc. Several extraction conditions were proposed, by varying extraction solvent, volume, agitation time, and temperature to recover proteins from Schirmer's strips [80,98] and small bioactive molecules from PRT [96].

Tear proteins absorb on PRT by varying strengths of intermolecular interactions with thread cellulose. The small surface and volume of PRT compared to a Schirmer's strip results in less interactions between proteins and thread. This can make it easier to elute/extract proteins from PRT. Our results supporting this theory because using either the SDS or AmBic protocol, 84–86% of proteins can be recovered, and protein composition was not biased, the extract correctly demonstrated the original composition with excellent reproducibility.

The observable proteome with the two most frequently used tear fluid sampling methods have been compared several times; most of these studies conclude that although both methods can be used, the capillary and Schirmer's strip tear collection methods still result in different protein compositions [83,86,93,99]. It was assumed that this is because Schirmer's strip results in an increased tear production due to possible irritation and contains proteins not only from tear fluid but also from tissues via direct eye contact [100].

Our findings are consistent with these results, as more proteins were identified in the samples collected with the Schirmer's strips compared to the capillary samples. The SL samples contain an even larger number of proteins than SU samples being in direct contact with the ocular surface. The number of identified proteins in the novel PRT samples was similarly high (1439), 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.

Ma *et al*. recently summarized [82] proteomics datasets from tear films using either capillary or Schirmer's strips by different research groups. They collected 1892 proteins

from 11 publications and found 435 proteins common to capillary and Schirmer's strip samples. Based on gene names, we matched those proteins to our dataset, to evaluate the overlap of our data with those recent results. Of all those proteins, 1656 were identified in our experiments, most of them in SL samples (1623), while 1238, 1089, and 382 proteins were found in the PRT, SU, and CAP samples, respectively. Regarding the 435 proteins designated as common by Ma and coworkers [82], we could identify 432 in total, 426, 406, 392, and 192 in SL, PRT, SU, and CAP samples, respectively. These results highlight the comparable proteomics usability of PRT not just with SU samples processed and analyzed in our laboratory, but with numerous other methods.

Akkurt Arslan *et al*. [101] processed their Schirmer's samples in a similar manner and found that 1153 proteins were identifiable in their SL samples, while 1107 proteins were identified in their SU samples. The significantly larger number of proteins identified in our SL samples may be explained by the MS method (DIA in our case vs. DDA) [82] and by the larger number of samples (20 SL samples from 10 individuals in our experiment vs. four SL samples from two individuals). Contamination occurs randomly, thus increasing the number of samples increases the chance of identification. It must be noted that only pooled samples were analyzed in that study and no quantitative comparisons were made, so no information is available on identification repeatability and quantitative variability.

In our experiments, the identification of such a large number of proteins was made possible by using the GPF–DIA LC–MS method. By the application of a combined spectral library, it was possible to quantify a higher number of useful proteins present in the tear fluid under normal conditions in a single nanoLC–MS run. Using a combined library, proteins and peptides that did not met the identification criteria with sample-specific libraries could be quantified in a significant proportion of samples. Searching against the combined library had a lower effect for SL samples than the PRT, SU, and CAP, as these samples contained the highest number of quantifiable peptides and proteins, therefore contributed to the greatest extent to the size of the library. This confirms the importance of using an optimally sized library. It may therefore be effective to use multiple samples, including Schirmer's strips, to generate the library, regardless of the applied sample collection protocol for the experimental samples. A similar result was observed by Nättinen *et al*. [86], that almost twice as much protein could be quantified in tear samples collected by capillary using a combined spectral library instead of capillary type-specific one. They also did not find increased number of proteins in Schirmer's type samples searched in a combined library. Green-Church *et al*. [93] have previously demonstrated that although the proteins detected in capillary samples

were mainly extracellular, tear samples collected by Schirmer's strip contained a large number of additional cellular proteins. Using a combined spectrum library, those proteins that could not be detected with the sample-specific libraries became easier to identify. Thus, it can be used as a quality control to identify tear samples that contain higher levels of contaminating proteins.

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 however 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 is very interesting because the entire thread was processed; the part in direct contact with the surface of the eyeball and the eyelids was not removed, unlike in the case of SU samples. This may be a consequence of the smaller diameter and the smaller and smoother surface of PRT fibers compared with Schirmer's paper strips.

According to the Gene Ontology analysis of protein clusters, we can conclude that intracellular proteins originating from the eye surface and lower layers of the tear film are increasing the size of the proteome sampled by the indirect methods compared to capillary. These proteins are most effectively collected on the surface of the lower part of the Schirmer's strip which is in direct contact with the surface of the eyeball and eyelids. Therefore, we conclude that those proteins may be designated as contaminants, if study of tear fluid is the goal, but may provide diagnostic information on eye surface, if specific sample collection methods can be applied for their reproducible analysis.

We can compare our clustering with the dataset of Ma *et al*. [82] described above, in which they found 435 proteins common in the literature data to Schirmer's strip and capillary samples. We could quantify 387 proteins of those, 68% of which can be found in Cluster A and B, while only 11% are in Cluster D. This shows high similarity between the classifications, despite the different approaches. They concluded that those common proteins are present at high abundance regardless of sample collection, which we also demonstrated (Figure 4). Our observation, that the rarely identified, low intensity proteins can be found with high frequency and at higher intensity in the SL samples, confirms the ocular surface origin of those proteins.

In order to validate the application of PRT in tear biomarker analysis, we collected tear biomarkers from recent reviews of literature data [82,102–106]. 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 also evaluated the effect of sampling on variance of results. 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 (median of 64% for the common the proteins), considerably lower than in the SU samples (median of 77%).

These lower intra- and interpersonal variances may be attributed to the lower induction of reflex tear formation during sampling compared to Schirmer's strip [91,107]. At the same time, the lower volume quickly collected by PRT ensures the collection of reproducible pure basal tear [96], 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, such as the drug metabolizing enzymes (DME) have a significant effect on the pharmacokinetics and pharmacodynamics of several drugs. Two families of transporters are the most responsible for the influx and efflux delivery of drugs across the cell membrane. Transporters from the ATP-binding cassette (ABC) superfamily mediate efflux of drugs and metabolites from the cells, while solute carrier (SLC) transporters are rather uptake or bidirectional transporters [108]. One important member of the ABC protein family is the Breast Cancer Resistance Protein (BCRP), which was first isolated from multidrug-resistant human breast cancer cells [109]. It is a common ABC transporter with broad substrate specificity, expressed in various tissues and plays an important role in the export of various endogenous and exogenous substrates including drugs [110]. An important representative of SLC transporters is the Organic Anion Transporting Polypeptide 1B3 (OATP1B3), which is expressed at a high level in liver and play important

role in the uptake of numerous therapeutic reagents, although it is also expressed in tumor tissues [111].

Some of the most effective and widely used tools for studying individual transporters often rely on the overexpression of specific transporters in host cell lines, as these methods provide plenty of information on drug interactions, including substrate determination, kinetics, and inhibitory effects. Inside-out membrane vesicles produced from cells expressing a transporter of interest are a powerful tool for studying drug transporters, their substrates and inhibitors and suited for high-throughput applications. These vesicles are generally utilized to study ABC transporters. The examinations are based on the ATP-dependent translocation of low permeability substances into these vesicles. Uptake transporters are usually examined using cell lines that have been transfected with cDNA encoding a particular transporter. The uptake of compounds into transfected cell lines is compared with that of control cells, allowing the contribution of the transfected transporter to be measured [112].

A prior study has shown that quantification of these DMEs and transporters by targeted absolute LC–MS/MS proteomics expand and significantly increase the utility of in vitro drug metabolism assays [113], as the measured transporter activities are influenced by the their abundance. In general, only mRNA assays are performed in these studies, while it is known, that the mRNA levels of DME and transporter proteins do not directly correlate with the concentration and functional activity of the proteins [114] expressions need to be determined at the protein level to observe real relations. Since both vesicles and cells have batch-to-batch variation, monitoring of transporter expression stability is also a quality control tool for such experiments.

In a quantitative measurement the change in the measured signal is proportional to the change in the amount of the analyte [115]. To confirm the correlation, a calibration curve is used, which is generated by a systematic dilution of the analyte. This curve can be used to determine the linearity of the method, as well as the LOD and LOQ values. According to the U.S. Food and Drug Administration (FDA) the LOD is the smallest measured concentration of the analyte from which the presence of the analyte in the tested sample can be deduced with acceptable certainty, while the LOQ is the smallest measured concentration of the analyte above which the analyte content can be determined with specified accuracy and precision [116]. Several methods are accepted and used to determine these values, as well as variously defined criteria. Generally, acceptable precision is when the relative standard deviation (RSD) is below 20% and the accuracy is good when the back-calculated error of

the concentration is smaller than 20%. The peptide signal must be accurately measurable and above the LOQ. Because of the significant role of matrix effects in LC–MS measurements, the calibration curves must be established in suitable sample matrix [117]. Signals below the LOQ may still be informative for the determination of relative differences, although the differences in the signals are not proportional to the quantitative differences of the analytes [118].

The proteolytic digestion of proteins allows an indirect measurement of their quantity via peptides. The usability of a peptide for targeted LC–MS measurements has three basic requirements: its response factor is suitable for sensitive detection, is formed in sufficiently large quantities by enzymatic digestion for reproducible detection and its sequence should be specific to the target. In addition, some *in silico* selection criteria can be used to select out peptides with inappropriate properties, supporting the selection of the most suitable peptide for LC–MS measurements. The length of the peptide should be between 6–16 amino acids, in order to have the suitable m/z value for MS detection. In addition, shorter peptides are often non-specific and longer ones may not fragment properly. For a unique MS signal, the candidates should not contain posttranslational modifications and single nucleotide polymorphism because of the multiple possible biological variants. Peptides should not derive from transmembrane region due to the difficult reproducibility of its release from the membrane and the aggregation tendency of extremely hydrophobic peptides. Arginine and lysine should not be included in the peptide sequence (missed cleavage sites), or in the close vicinity of the peptide's N- or C-terminal (possible miss cleaved forms). Presence of methionine (possible oxidation) and cysteine (possibly not completely alkylated) [119] may also lead to presence of different peptide forms. It is also preferable if the peptide is detectable only in one charge states. The mentioned properties are crucial to present a single form of a peptide in the samples instead of the quantitative distribution between several forms. A combined method based on *in silico* and experimental techniques can also be used to select the most reliable peptides. While the sequences of the proteins of interest are digested *in silico* and then filtered based on the discussed criteria, in parallel, samples containing the protein sequence are experimentally digested, and the usability of the peptides are confirmed by LC–MS [120].

Since the studied transporters are found and function in the plasma membrane, and are often expressed at a very low level [119], 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 [121,122]. An additional difficulty in the detection of membrane proteins is their frequently large size and most notably their hydrophobic characteristics [123]. 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

5.2.1. Comparison of different membrane enrichment protocols

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 were compared with the whole cell lysate samples, in terms of number and quantity of the proteins and peptides after SEPOD digestion.

In order to compare the membrane isolation efficiency of the protocols, the measured number and quantity (summed relative intensity) of membrane proteins and peptides, as well as ABC and SLC transporters and peptides were investigated. In the enriched samples (PE–vesicle) 5.5–7.1 % more membrane proteins and 8.3–10.9 % more membrane peptides were measured than in the TC lysates. The average number of membrane proteins were 2120 in PE and 2402 in vesicle samples compared to 1980, in TC samples, respectively. These numbers for peptides were 17689 in PE, 22620 in vesicle and 1490 in TC samples. The summed intensity of membrane proteins was increased by 15.3–18.0 % and 13.7–18.5 % at the peptide level. The relative intensity of membrane proteins and membrane peptides did not differ between the two membrane samples (Figure 11).

In both membrane isolation techniques, a greater enrichment of ABC and SLC transporters was observed than for all membrane proteins. The ratio of the number of ABC and SLC proteins was approximately 1.5–1.6-fold higher in the enriched samples (PE–vesicle samples) than in the TC samples, while this number for the ABC and SLC peptides was 1.8–2-fold. The ratio of the total intensity of these proteins increased approximately 2.6–3.8-fold and their peptides increased 2.9–4.8-fold compared to that measured in TC samples (Table 3). The average number of ABC and SLC transporters were 122 in PE and 145 in vesicle samples compared to 85, in TC samples, respectively. These numbers for peptides were 771 in PE, 1011 in vesicle and 456 in TC samples.

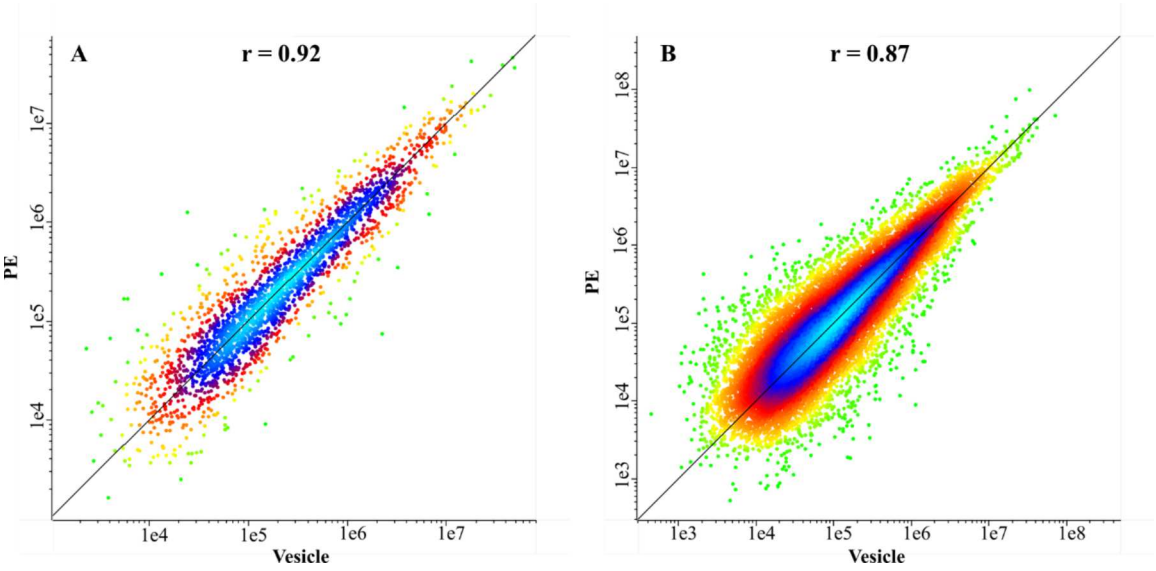


Figure 11. Comparison of the intensities of the measured membrane proteins (**A**) and peptides (**B**) in the two membrane samples after SEPOD protocol. The mean intensities measured in samples on a log10 scale are plotted. Pearson correlation coefficients are shown at the top of the figure.

Table 3. The fraction (% of the total) of the summed intensity and number of membrane and ABC + SLC transporter proteins in the different sample types.

	Membrane proteins		**Membrane peptides**	
	Intensity (%)	**Number (%)**	**Intensity (%)**	**Number (%)**
TC	24.67 ± 1.21	32.56 ± 0.76	21.83 ± 0.83	28.16 ± 0.95
PE	40.01 ± 6.29	38.00 ± 1.72	35.50 ± 6.80	36.46 ± 2.99
Vesicle	42.62 ± 1.53	39.64 ± 0.18	40.32 ± 1.76	39.09 ± 0.45
	ABC + SLC proteins		**ABC + SLC peptides**	
	Intensity (%)	**Number (%)**	**Intensity (%)**	**Number (%)**
TC	0.83 ± 0.14	1.46 ± 0.07	0.75 ± 0.12	0.86 ± 0.04
PE	2.16 ± 0.41	2.19 ± 0.21	2.17 ± 0.48	1.59 ± 0.19
Vesicle	3.15 ± 0.23	2.39 ± 0.03	3.63 ± 0.27	1.75 ± 0.05

The mean intensities, their relative standard deviation and the number of the measured proteins and peptides (all, membrane, and ABC + SLC) can be found in Supplementary Table 2.

Since our target proteins are located in the plasma membrane, in our case the aim of the enrichment procedures is to increase the proportion of plasma membrane proteins and remove other components from the samples. Although both membrane isolation techniques have been shown to be effective, it is difficult to separate the different membranes present in cells. As well as plasma membrane, other membranes may also be enriched by the

isolation techniques. Therefore, our aim was to investigate whether there is a difference in the membrane composition of the samples produced by the two techniques. For this investigation, cytosol (GAPDH), cell membrane (ATP1A1), endoplasmic reticulum (CALR), Golgi (GOSR1), nucleus (H4C2), and mitochondria (TIMM23) specific marker proteins were selectively evaluated from the DIA measurements. The protein intensities measured from the samples of each enrichment procedure were compared to the TC samples. With the calculated enrichment factors, it was possible to characterize the obtained sample composition (Figure 12).

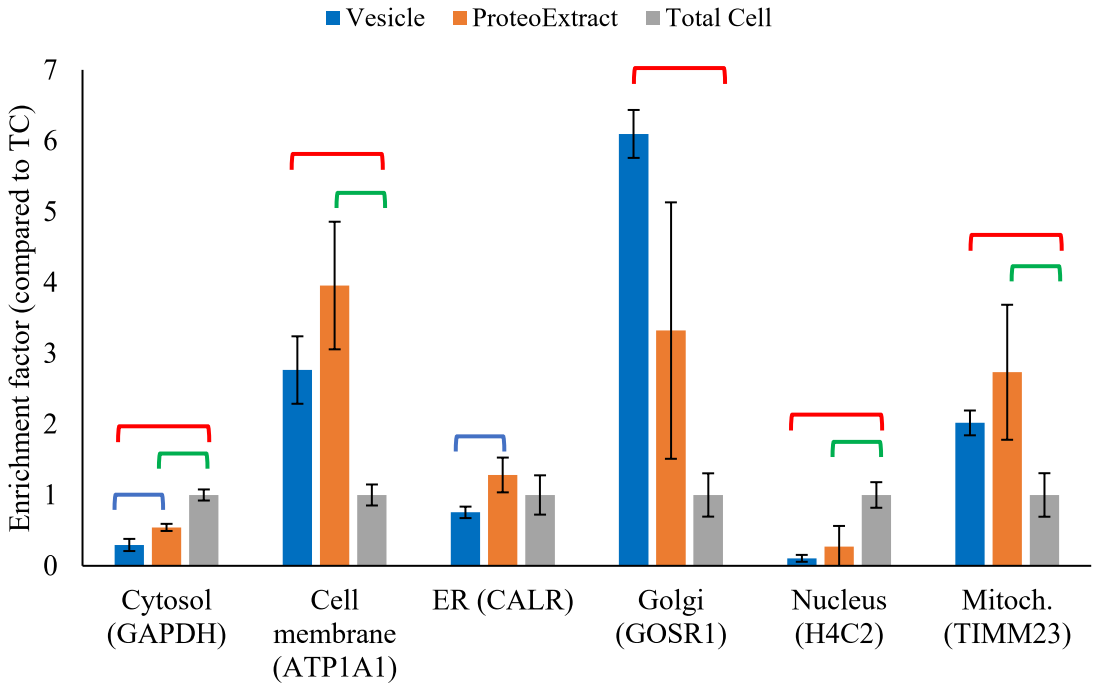


Figure 12. The enrichment factors of the markers of different cell compartments calculated for different membrane enrichment protocols (Vesicle displayed with blue, PE with orange) compared to the TC lysate (with grey). Brackets represent significant differences with Welch's T-test $p<0.05$. Red bracket means a significant difference between TC and vesicles, green between TC and PE, while blue between PE and vesicle samples.

The cytosolic protein content of both the PE and vesicle samples were significantly less than in the TC samples, along with this, cell membrane protein content increased significantly during both enrichment protocols. As a result of both enrichments, nucleus proteins were present in a lower amount and the mitochondrial membrane proteins in a higher amount than in TC lysates. The portion of Golgi membrane proteins was significantly higher in the vesicle samples, while the ER membrane proteins did not differ from TC with

any of the protocols. However, the amount of ER proteins was greater in PE samples than in vesicle samples. The exact values are summarized in Supplementary Table 3.

5.2.2. Comparison of different digestion protocols

Due to the often-large size and extreme hydrophobic nature of membrane proteins, it is necessary to use a large amount of detergents and other components to promote cell disruption and keep proteins in solution during sample preparation. However, it complicates LC–MS detection as well as contaminates the instrument and reduces the lifetime of the LC column if it is not removed from the samples. Therefore, it is not possible to prepare these samples by a simple, direct in-solution digestion method, without any molecular contaminant removal. A sample preparation/digestion is required that can be effective for producing a tenside-free sample with or without a separate tenside removal step, such as FASP or SEPOD (Supplementary Figure 1). However, the peptide composition of the sample is highly dependent on the digestion method used, so it is important to investigate its effects.

Slightly more proteins and peptides could be quantified using the SEPOD protocol than with FASP in vesicle samples, but the difference between the two methods was greater in PE samples. The average fraction of the number of membrane proteins and peptides compared to all was approximately 40% by all methods, while this number was 2% for ABC + SLC transporters (exact numbers in Table 4). The results also show that the differences are more prominent between the protocols at the level of peptides than proteins. The number of quantified proteins and peptides applied the examined digestion protocols.

Table 4. The number of quantified proteins and peptides applied the examined digestion protocols in PE and vesicle samples.

		All		**Membrane**		**ABC + SLC**	
		Proteins	**Peptides**	**Proteins**	**Peptides**	**Proteins**	**Peptides**
PE	FASP	4930 ± 203	37489 ± 2690	1982 ± 84	14754 ± 1037	113 ± 9	672 ± 43
	SEPOD	5307 ± 486	54582 ± 670	2339 ± 6	21181 ± 283	144 ± 0	945 ± 17
Vesicle	FASP	5206 ± 105	42564 ± 1744	2161 ±41	17651 ± 702	136 ± 1	900 ± 29
	SEPOD	5488 ± 395	56293 ± 1090	2402 ± 13	22709 ± 334	149 ± 3	1095 ± 22

Taking advantage of the DIA data acquisition method, not only the number of detectable proteins and peptides was compared, but also the relative intensities of their MS2 signals. Not large differences were detectable between the methods at the protein level, but more apparent were seen at the level of peptides (Supplementary Figure 2). In general, smaller differences were detectable for more abundant proteins and peptides, so both

digestion protocols can be used for reproducible quantification of overexpressed proteins. The reproducibility of the measured intensities was also investigated (Supplementary Table 4).

5.2.3. Selection of suitable peptides for targeted measurements

Using DIA measurement, MS2 quantitative information was collected on all detectable peptides of OATP1B3 protein from samples digested by SEPOD method. Only those peptides whose relative intensity reached 10% of the most intense peptide were included in the further examinations. Predicted and empirical data were collected about the detected peptides o target proteins from PeptideAtlas [58]. The location of the peptides within the proteins and the known post-translational modifications were examined with Protter [124] (Supplementary Figure 3). Peptides that appeared to be problematic due to their sequence or location were flagged before further examination.

The LC–MS signal of a peptide largely depends on the sample matrix. The co-eluting components, which are thus also analyzed in the MS at the same time, have a significant effect on each other. Various interferences can be observed, so the LC–MS signal of a peptide always depends on the given matrix and the applied method. It is not possible to give exact LOD and LOQ values from an experiment performed in a matrix different from the sample. Therefore, 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. Thus, the matrix effect was similar for a peptide at the different dilutions throughout the study. Data on these peptides were collected using the PRM measurement method, and the relative LOD and LOQ values were calculated from the linear regression data of the obtained calibration curves (Figure 13). These values showed that what dilution of the overexpressed protein amount can be detected and quantified with the given peptide (Table 5). In addition, the linearity and MS2 relative intensities were examined, as well as the greatest dilutions where the CVs were less than 0.2. Considering all aspects, the NQTANLTNQGK peptide proved to be the most applicable for the absolute quantification of OATP1B3.

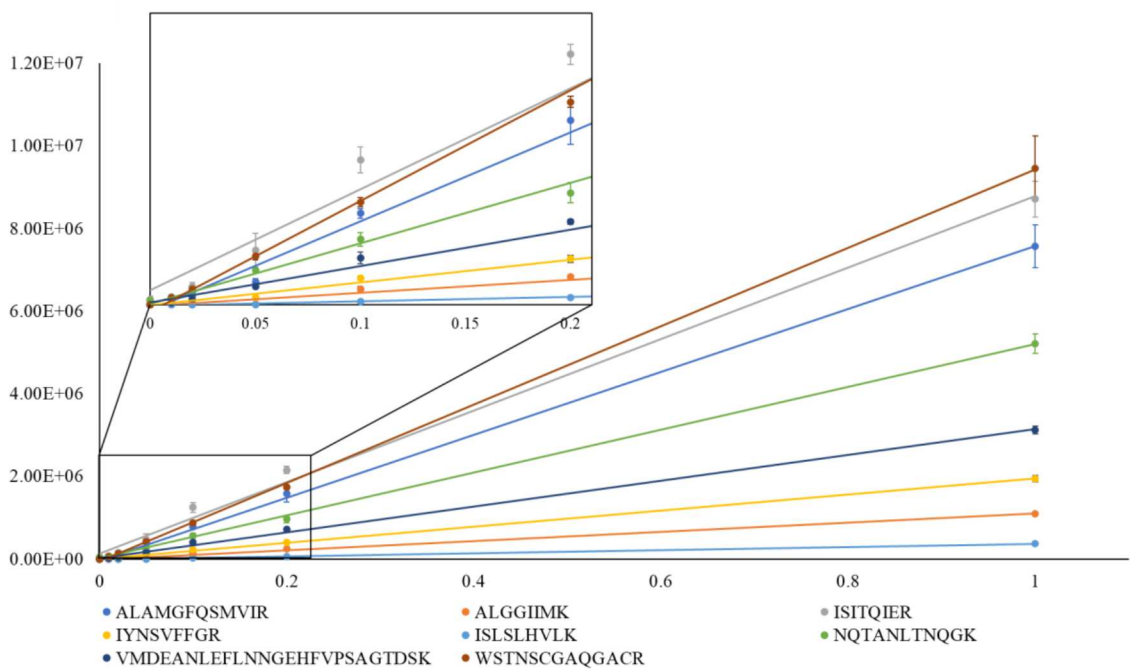


Figure 13. Calibration curves of the top 8 most intense peptides of OATP1B3 in the overexpressed sample diluted with control HEK293 digest (all r^2>0.995). The relative intensities are plotted against the dilution factor (the overexpressed ratio of the whole sample).

Table 5. Summary of the data of the most intensively detectable OATP1B3 peptides. Amino acids generally considered problematic in the literature are highlighted in bold. The peptide selected for absolute quantification is highlighted in green. The MS2 intensity rank is based on the undiluted sample.

Peptid	r^2	LOD*	LOQ*	Intensity Rank (MS2)	Maximum dilutability (CV<0.2)	PSS**	ESS**	N Obs**
WSTNS**C**GAQGA**C**R	0.9997	121.5	40.1	1	50x	–	0.51	193
NQTANLTNQGK	0.9994	77.4	25.6	4	50x	0.09	0.61	427
ALA**M**GFQS**M**VIR	0.9991	59.5	19.6	3	50x	0.51	0.5	184
IYNSVFFGR	0.9987	49.5	16.3	6	20x	0.27	0.41	70
V**M**DEANLEFLNNG--EHFVPSAGTDSK	0.9983	42.5	14	5	100x	0.9	0.77	1058
ISLSLHVLK	0.9979	38.8	12.8	8	5x	–	–	–
ISITQIER	0.9964	29.2	9.6	2	10x	0.07	–	–
ALGGII**M**K	0.9959	27.5	9.1	7	10x	0.11	–	–

* In the absence of stable isotope-labeled peptides, absolute amounts cannot be determined, however, it is possible to calculate how many times the overexpressed amount can be diluted.

** Predicted (PSS) and empirical suitability scores (ESS) and the number of total observations (N Obs) derived from PeptideAtlas [58].

The length of the peptide is sufficient for the selective MS detection, does not contain any known post-translational modifications, and is not derived from the transmembrane region but from the intracellular region. The missed cleavage sites were investigated in the peptide and in the immediate surrounding sequence. It does not contain methionine or cysteine. It was also investigated whether multiple charge states can be detected, since the result could be that several types of precursor ions are formed from one sequence. However, multiple charge states and missed cleavage site containing versions for this peptide could not be detected. It showed good linearity, the relative intensity of this peptide changed proportionally with the dilution, and its reproducibility was sufficient even in the 50-fold dilution. Based on the MS2 intensity, this peptide is ranked only 4th, but based on the relative LOQ value, it is the 2nd, which proves that it is not enough to rely only on the MS2 intensity.

Peptides were also selected for BCRP protein quantification using the method presented above (data not shown). An additional consideration for these samples was that — as the overexpressed protein derived from different species than the background proteome in some of the samples — a peptide from a less conserved protein region should be used as a quality control. Although there is no described BCRP protein in the Uniprot database for *Spodoptera frugiperda*, the commonly used SSLLDVLAAR peptide for human BCRP quantitation is shared with many other species. In a different project for simultaneous determination of mouse and rat BRCP with human (data not shown), we had to use a different BCRP peptide. The other peptide was also used in these measurements, which was the LLSDLLPMR. 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.

5.2.4. Absolute quantification of OATP1B3 and BCRP in different membrane fractions

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 sample A was taken as 100%. The quantity measured in the other samples was related to that amount (Figure 14).

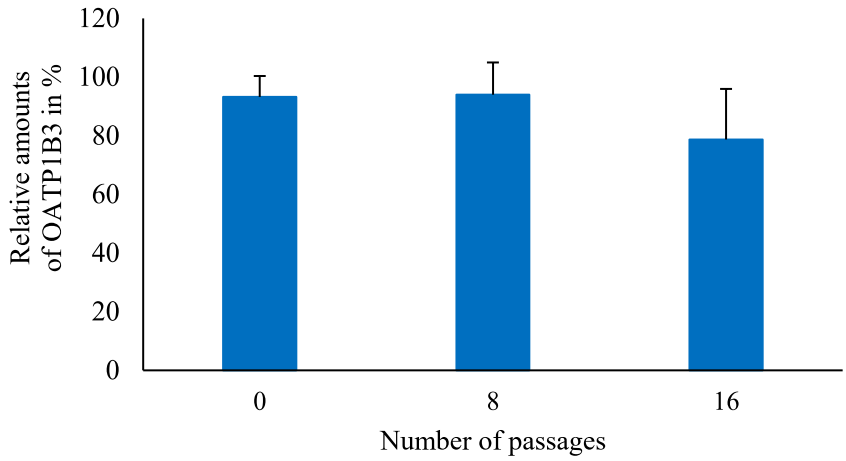


Figure 14. Relative amounts of OATP1B3 at the initial time point and after 8 and 16 passages (approximately 1 and 2 months) related to the initial amount of sample A.

There was no significant difference in the membrane protein concentrations (0.35 ± 0.05 mg/mL), neither in the expression of investigated transporter between the replicates and time points. The results show that the RSD of OATP1B3 expression between the samples were relatively 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 relative standard deviation (RSD) of the isotopically 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. The measured amounts were 165.76 ± 41.98 pmol / mg membrane protein in HEK293, 157.75 ± 26.62 pmol / mg membrane protein in MCF-7/MX samples, 128.77 ± 41.59 pmol / mg membrane protein in Sf9 and 122.51 ± 9.66 pmol / mg membrane protein in Sf9-HAM samples (values in Supplementary Table 5). No significant difference was detected in the BCRP expression of the different types of vesicle samples (Figure 15). 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.

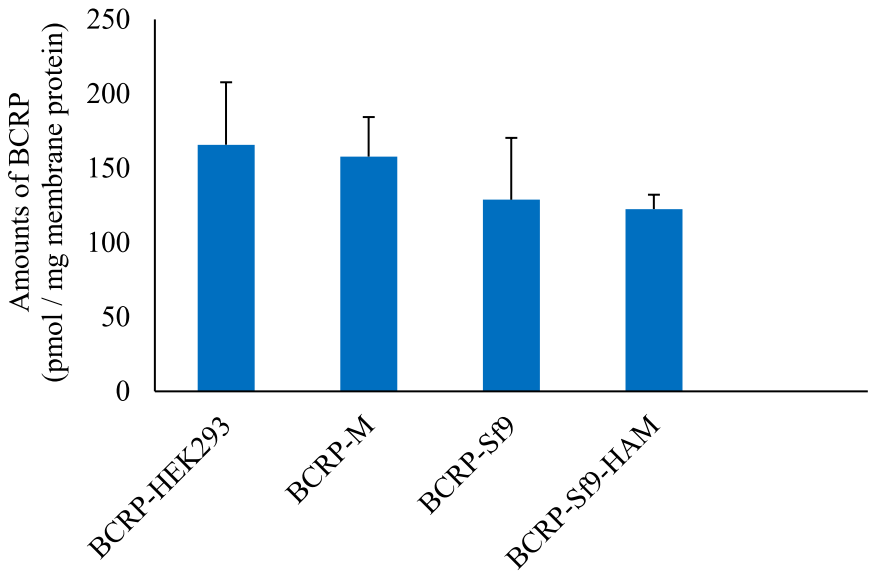


Figure 15. Absolute BCRP protein content of the four types of investigated vesicle samples obtained from BCRP-overexpressing HEK293 (n = 6), MCF-7/MX (n = 4), Sf9 cell lines (n = 3 with cholesterol load (HAM) and n = 3 without cholesterol loading). Data shown as mean + standard deviation from.

5.3. Discussion

Since drug transporter proteins are in the plasma membrane, 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 to what extent the PE sample is representative of the same plasma membrane proteome as the vesicle sample, 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 transporter also increased significantly. Kumar *et al.* [121] found that using membrane extraction kits, not only plasma membrane proteins were enriched, but these samples were often contaminated with proteins from other membranes and cytoplasm. Similar results were experienced by Jankovskaja *et al.* [122] using the ProteoExtract® Native Membrane Protein Extraction kit. The amount of a plasma membrane marker protein (Na^+/K^+-ATPase) was two times higher in samples prepared by an ultracentrifugation method, than in the samples prepared with this kit. In addition, the reproducibility measured in technical replicates was also weaker using the kit. Our results were similar, as the mean CV of measurable membrane proteins was higher, also that of ABC and SLC transporters in PE samples, compared to the vesicle samples. In addition,

different marker (based on [121]) proteins were used to determine the composition of the enriched membrane samples. Although there were differences in the composition of the samples based on specific markers, membrane proteins and their peptides showed a strong correlation in the two membrane preparations. Both methods were able to significantly increase the proportion of plasma membrane proteins and decrease the proportion of cytosolic proteins in the samples compared to the whole cell samples. In accordance with the results found in the literature, the samples revealed with the kit also contained other proteins. However, it was also true for the other examined membrane enrichment protocol. Proteins from other membranes (ER, Golgi, mitochondrion) were also enriched using both methods.

In bottom-up proteomics, another important factor what influences the results is the applied digestion protocol. Due to the poor solubility of membrane proteins it is often necessary to use solvents containing large amounts of detergents to dissolve membranes, disrupt protein–lipid interactions, and achieve protein solubilization [125]. However, these can negatively affect the sensitivity and reproducibility of LC–MS measurements [126], so it may be necessary to remove them during sample preparation. Shen *et al.* [70] found that, using SEPOD, significantly more proteins could be identified in human and mouse tissue samples, with much higher peptide yield, than with the FASP protocol. Klont *et al.* [127] also compared these sample preparations, but concluded that no sample preparation is clearly superior to the other. In these experiments DDA acquisition was applied, so the quantitative comparison was completed at the MS1 level. Our results are derived from DIA measurements, therefore, MS2 relative intensity data was available, which enabled a more selective and reliable quantitative comparison. 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. Therefore, the vesicle samples were prepared using the FASP protocol. 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.

In bottom-up proteomics, choosing peptides for quantifying proteins have a significant effect on the final result. In the response factor of peptides within proteins could differ on average by over three orders of magnitude between the lowest and highest responding

peptides [15]. To create the most robust, reliable, and sensitive targeted methods, the most suitable peptide should be used. As the peptide composition of a sample is dependent on the applied digestion protocol, peptide need to be selected using the same sample preparation, which is used for the experimental samples. Peptides for the absolute quantification of proteins must meet several criteria [119]. Recently, software packages have appeared (e.g. PREGO [128]) that are able to predict the applicability of peptides for targeted measurements, however, peptide selection based on experimental studies still plays an important role. Gröer *et al.* [120] applied a combination of *in silico* predictions and experimental data for determining the appropriate peptides for MRM measurements of transporter proteins. Briefly, after *in silico* digestion, peptides were selected based on their sequences, which were then verified by LC–MS measurements. Similarly to this approach the peptides were examined with a combined strategy in our experiment. The suitable peptides for the absolute quantification of OATP1B3 and BCRP proteins were selected by the combined criteria system. In addition to the criteria previously defined in the literature, the occurrence of missed cleavage sites not only in the peptide sequence but also in its immediate vicinity was investigated, as this would increase the number of possible precursors. The possible occurrence of multiple charge states was also monitored, which could introduce additional variance. It was taken into account that the background proteome and the overexpressed protein are from different species, so it is advisable to choose a peptide that can distinguish between them. Therefore, two peptides were chosen for absolute quantification of BCRP protein. A common problem is that it is not possible to choose a perfect peptide in all aspects and a compromise has to be made. Also in this method, one of the peptides contained methionine, but the oxidized product was monitored during the analysis. It was not found to be problematic in the short-term analysis, as the amount of the methionine sulfoxide version was constant and low, but its long-term applicability requires further investigation.

However, our experimental examination was carried out with DIA measurement, for which it was not necessary to pre-filter the target peptides, e.g. based on their sequence. The selection of the peptides was applied only based on MS2 quantitative information. The relative LOQs of the most intense peptides were determined using calibration curves. This is a challenge for endogenous peptides because a target peptide-free matrix is not available. Several approaches are used, of which one of the most common is using reverse calibration curves. In this method stable isotope-labeled synthetic equivalent of the analyte can be used to predict the responses of a peptide in targeted proteomic experiments [129]. However, the

isotope-labeled peptide has a different m/z than the endogenous, so different ions can cause interference and affect their detection, but it can also be tested whether the possible labeled analogue will have interference in the given matrix. In our experiments, the goal was to determine the amount of overexpressed proteins for which a control cell line was available. 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. A matrix in which the endogenous amount of the peptide of interest is below the LOD is suitable for dilution. 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. Considering the in silico, MS2 intensity, sequence and, relative LOQ data, the most suitable peptides were selected for PRM measurements.

Due to the appropriate reproducibility, the developed target method is also suitable for the detection of small quantitative differences. Applying this method, the expression stability of overexpressed OATP1B3 in HEK293-OATP1B3-LV cells was confirmed, as they are stable cell lines. There was no significant difference detected over 16 passages, for approximately 2 months. This suggests that they can be used to study uptake transporters two months after culturing. There was no significant difference in the amount of expressed BCRP in the different BCRP overexpressing membranes, so that all these membranes can be used to perform comparable activity assays.

6. Summary

In our experiments, DIA data acquisition technique was used to analyze the proteomes of different samples. As a result, we have successfully demonstrated the proteomic applicability of a new tear sampling protocol compared to two commonly used methods. In addition, by using DIA, it was possible to optimize the sample preparation protocol and select the most suitable peptides for targeted quantitative proteomic measurements.

7. Summary of the new findings

1) We created a combined library of tear samples collected from multiple sampling procedure, which was used instead of sample-specific libraries, significantly improving the results.
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, for 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.

8. Acknowledgements

I would like to express my most sincere gratitude to my mentors Prof. Dr. Tamás Janáky and Dr. Zoltán Szabó at the Department of Medical Chemistry, University of Szeged for their guidance and support. Without their outstanding supervision, this Ph.D. thesis would not have been possible. I owe a special thanks to Dr. Edit Tóth-Molnár at the Department of Ophtalmology, University of Szeged, Dr. Péter Krajcsi, Dr. Emese Kis, Dr. Anita Kurunczi, Dr. Judit Molnár and Zsolt Sáfár at SOLVO Biotechnology Zrt. for their collaboration and help. I am grateful to Prof. Dr. Tamás Martinek and Prof. Dr. Gábor Tóth, the current and former heads of the Department of Medical Chemistry, University of Szeged, who gave me the opportunity to work at their department. I also would like to thank Dr. Zoltán Kele for his scientific discussions and valuable insights over the years, and all my colleagues and friends for their help, encouragement, and the great times we had. This work would not have been possible to accomplish without the assistance and work of Dr. Rita Ábrahámné Szendrei. I owe warm thanks to my beloved family for their unconditional love, support, and patience during my doctoral studies; especially to my wife Mariann Ács-Kurucz and my son Kornél Kecskeméti, who taught me the importance of the save button.

9. References

1. Diamandis, E.P. Mass Spectrometry as a Diagnostic and a Cancer Biomarker Discovery Tool: Opportunities and Potential Limitations. *Mol. Cell. Proteomics* **2004**, *3*, 367–378, doi:10.1074/mcp.R400007-MCP200.

2. Gstaiger, M.; Aebersold, R. Applying Mass Spectrometry-Based Proteomics to Genetics, Genomics and Network Biology. *Nat. Rev. Genet.* **2009**, *10*, 617–627, doi:10.1038/nrg2633.

3. Ahrens, C.H.; Brunner, E.; Qeli, E.; Basler, K.; Aebersold, R. Generating and Navigating Proteome Maps Using Mass Spectrometry. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 789–801, doi:10.1038/nrm2973.

4. Xue, J.; Guijas, C.; Benton, H.P.; Warth, B.; Siuzdak, G. METLIN MS2 Molecular Standards Database: A Broad Chemical and Biological Resource. *Nat. Methods* **2020**, *17*, 953–954, doi:10.1038/s41592-020-0942-5.

5. Sinha, A.; Mann, M. A Beginner's Guide to Mass Spectrometry—Based Proteomics. *Biochem. (Lond).* **2020**, *42*, 64–69, doi:10.1042/bio20200057.

6. Aslam, B.; Basit, M.; Nisar, M.A.; Khurshid, M.; Rasool, M.H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **2017**, *55*, 182–196, doi:10.1093/chromsci/bmw167.

7. Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science (80-.).* **2006**, *312*, 212–217, doi:10.1126/science.1124619.

8. Zhang, H.; Ge, Y. Comprehensive Analysis of Protein Modifications by Top-down Mass Spectrometry. *Circ. Cardiovasc. Genet.* **2011**, *4*, 711, doi:10.1161/CIRCGENETICS.110.957829.

9. Sidoli, S.; Lu, C.; Coradin, M.; Wang, X.; Karch, K.R.; Ruminowicz, C.; Garcia, B.A. Metabolic Labeling in Middle-down Proteomics Allows for Investigation of the Dynamics of the Histone Code. *Epigenetics and Chromatin* **2017**, *10*, 34, doi:10.1186/s13072-017-0139-z.

10. Zhang, Y.; Fonslow, B.R.; Shan, B.; Baek, M.C.; Yates, J.R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* 2013, *113*, 2343–2394.

11. Dupree, E.J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B.A.; Darie, C.C. A Critical Review of Bottom-up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020**, *8*, 1–26, doi:10.3390/proteomes8030014.

12. Wang, W.Q.; Jensen, O.N.; Møller, I.M.; Hebelstrup, K.H.; Rogowska-Wrzesinska, A. Evaluation of Sample Preparation Methods for Mass Spectrometry-Based Proteomic Analysis of Barley Leaves. *Plant Methods* **2018**, *14*, doi:10.1186/s13007-018-0341-4.

13. Jami-Alahmadi, Y.; Pandey, V.; Mayank, A.K.; Wohlschlegel, J.A. A Robust Method for Packing High Resolution C18 Rp-Nano-Hplc Columns. *J. Vis. Exp.* **2021**, *2021*, e62380, doi:10.3791/62380.

14. Scigelova, M.; Makarov, A. Orbitrap Mass Analyzer - Overview and Applications in Proteomics. In Proceedings of the Proteomics; 2006; Vol. 1, pp. 16–21.

15. Searle, B.C. Development of Data Independent Acquisition Methods to Systematically Analyze the Human Proteome, University of Washington, 2018.

16. Fang, N.; Yu, S.; Ronis, M.J.J.; Badger, T.M. Matrix Effects Break the LC Behavior Rule for Analytes in LC-MS/MS Analysis of Biological Samples. *Exp. Biol. Med.* **2015**, *240*, 488–497, doi:10.1177/1535370214554545.

17. Griss, J. Spectral Library Searching in Proteomics. *Proteomics* **2016**, *16*, 729–740, doi:10.1002/pmic.201500296.

18. Zhang, X.; Li, Y.; Shao, W.; Lam, H. Understanding the Improved Sensitivity of Spectral Library Searching over Sequence Database Searching in Proteomics Data Analysis. *Proteomics* **2011**, *11*, 1075–1085, doi:10.1002/pmic.201000492.

19. Szabo, Z.; Janaky, T. Challenges and Developments in Protein Identification Using Mass Spectrometry. *TrAC - Trends Anal. Chem.* 2015, *69*.

20. Stahl, D.C.; Swiderek, K.M.; Davis, M.T.; Lee, T.D. Data-Controlled Automation of Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptide Mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 532–540, doi:10.1016/1044-0305(96)00057-8.

21. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature* **2014**, *509*, 582–587, doi:10.1038/nature13319.

22. Kim, M.S.; Pinto, S.M.; Getnet, D.; Nirujogi, R.S.; Manda, S.S.; Chaerkady, R.; Madugundu, A.K.; Kelkar, D.S.; Isserlin, R.; Jain, S.; et al. A Draft Map of the Human Proteome. *Nature* **2014**, *509*, 575–581, doi:10.1038/nature13302.

23. Richards, A.L.; Hebert, A.S.; Ulbrich, A.; Bailey, D.J.; Coughlin, E.E.; Westphall, M.S.; Coon, J.J. One-Hour Proteome Analysis in Yeast. *Nat. Protoc.* **2015**, *10*, 701–714, doi:10.1038/nprot.2015.040.

24. Borràs, E.; Sabidó, E. What Is Targeted Proteomics? A Concise Revision of Targeted Acquisition and Targeted Data Analysis in Mass Spectrometry. *Proteomics* **2017**, *17*, 1700180, doi:10.1002/pmic.201700180.

25. Kondrat, R.W.; McClusky, G.A.; Cooks, R.G. Multiple Reaction Monitoring in Mass Spectrometry/Mass Spectrometry for Direct Analysis of Complex Mixtures. *Anal. Chem.* **1978**, *50*, 2017–2021, doi:10.1021/ac50036a020.

26. Peterson, A.C.; Russell, J.D.; Bailey, D.J.; Westphall, M.S.; Coon, J.J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **2012**, *11*, 1475–1488, doi:10.1074/mcp.O112.020131.

27. Lawless, C.; Holman, S.W.; Brownridge, P.; Lanthaler, K.; Harman, V.M.; Watkins, R.; Hammond, D.E.; Miller, R.L.; Sims, P.F.G.; Grant, C.M.; et al. Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring. *Mol. Cell. Proteomics* **2016**, *15*, 1309–1322, doi:10.1074/mcp.M115.054288.

28. Venable, J.D.; Dong, M.Q.; Wohlschlegel, J.; Dillin, A.; Yates, J.R. Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat. Methods* **2004**, *1*, 39–45, doi:10.1038/nmeth705.

29. Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* **2015**, *14*, 4359–4371, doi:10.1021/acs.jproteome.5b00543.

30. Pino, L.K.; Just, S.C.; MacCoss, M.J.; Searle, B.C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**, *19*, 1088–1103, doi:10.1074/mcp.P119.001913.

31. Rosenberger, G.; Koh, C.C.; Guo, T.; Röst, H.L.; Kouvonen, P.; Collins, B.C.; Heusel, M.; Liu, Y.; Caron, E.; Vichalkovski, A.; et al. A Repository of Assays to Quantify 10,000 Human Proteins by SWATH-MS. *Sci. Data* **2014**, *1*, doi:10.1038/sdata.2014.31.

32. Lawrence, R.T.; Searle, B.C.; Llovet, A.; Villén, J. Plug-and-Play Analysis of the Human Phosphoproteome by Targeted High-Resolution Mass Spectrometry. *Nat. Methods* **2016**, doi:10.1038/nmeth.3811.

33. Egertson, J.D.; Kuehn, A.; Merrihew, G.E.; Bateman, N.W.; MacLean, B.X.; Ting, Y.S.; Canterbury, J.D.; Marsh, D.M.; Kellmann, M.; Zabrouskov, V.; et al. Multiplexed MS/MS for Improved Data-Independent Acquisition. *Nat. Methods* **2013**, *10*, 744–746, doi:10.1038/nmeth.2528.

34. Panchaud, A.; Scherl, A.; Shaffer, S.A.; Von Haller, P.D.; Kulasekara, H.D.; Miller, S.I.; Goodlett, D.R. Precursor Acquisition Independent from Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81*, doi:10.1021/ac900888s.

35. Searle, B.C.; Pino, L.K.; Egertson, J.D.; Ting, Y.S.; Lawrence, R.T.; Villén, J.; Maccoss, M.J. Comprehensive Peptide Quantification for Data Independent Acquisition Mass Spectrometry Using Chromatogram Libraries. *bioRxiv* **2018**, *51*.

36. Li, G.Z.; Vissers, J.P.C.; Silva, J.C.; Golick, D.; Gorenstein, M. V.; Geromanos, S.J. Database Searching and Accounting of Multiplexed Precursor and Product Ion Spectra from the Data Independent Analysis of Simple and Complex Peptide Mixtures. *Proteomics* **2009**, *9*, doi:10.1002/pmic.200800564.

37. Tsou, C.C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.C.; Nesvizhskii, A.I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat. Methods* **2015**, *12*, doi:10.1038/nmeth.3255.

38. Ting, Y.S.; Egertson, J.D.; Payne, S.H.; Kim, S.; MacLean, B.; Käll, L.; Aebersoldo, R.; Smith, R.D.; Noble, W.S.; MacCoss, M.J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics* **2015**, *14*, 2301–2307, doi:10.1074/mcp.O114.047035.

39. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20*, 1900276, doi:10.1002/pmic.201900276.

40. Schubert, O.T.; Gillet, L.C.; Collins, B.C.; Navarro, P.; Rosenberger, G.; Wolski, W.E.; Lam, H.; Amodei, D.; Mallick, P.; Maclean, B.; et al. Building High-Quality Assay Libraries for Targeted Analysis of SWATH MS Data. *Nat. Protoc.* **2015**, *10*, 426–441, doi:10.1038/nprot.2015.015.

41. Röst, H.L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinoviä, S.M.; Schubert, O.T.; Wolski, W.; Collins, B.C.; Malmström, J.; Malmström, L.; et al. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32*, 219–223, doi:10.1038/nbt.2841.

42. Bruderer, R.; Bernhardt, O.M.; Gandhi, T.; Miladinović, S.M.; Cheng, L.Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; et al. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **2015**, *14*, 1400–1410, doi:10.1074/mcp.M114.044305.

43. Searle, B.C.; Pino, L.K.; Egertson, J.D.; Ting, Y.S.; Lawrence, R.T.; MacLean, B.X.; Villén, J.; MacCoss, M.J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2018**, *9*, doi:10.1038/s41467-018-07454-w.

44. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26*, 966–968, doi:10.1093/bioinformatics/btq054.

45. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11*, 1–17, doi:10.1074/mcp.O111.016717.

46. Peckner, R.; Myers, S.A.; Jacome, A.S.V.; Egertson, J.D.; Abelin, J.G.; MacCoss, M.J.; Carr, S.A.; Jaffe, J.D. Specter: Linear Deconvolution for Targeted Analysis of Data-Independent Acquisition Mass Spectrometry Proteomics. *Nat. Methods* **2018**, *15*, 371–378, doi:10.1038/nmeth.4643.

47. Keller, A.; Bader, S.L.; Shteynberg, D.; Hood, L.; Moritz, R.L. Automated Validation of Results and Removal of Fragment Ion Interferences in Targeted Analysis of Data-Independent Acquisition Mass Spectrometry (MS) Using SWATHProphet. *Mol. Cell. Proteomics* **2015**, *14*, 1411–1418, doi:10.1074/mcp.O114.044917.

48. Meyer, J.G.; Mukkamalla, S.; Steen, H.; Nesvizhskii, A.I.; Gibson, B.W.; Schilling, B. PIQED: Automated Identification and Quantification of Protein Modifications from DIA-MS Data. *Nat. Methods* **2017**, *14*, 646–647, doi:10.1038/nmeth.4334.

49. Demichev, V.; Messner, C.B.; Vernardis, S.I.; Lilley, K.S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17*, 41–44, doi:10.1038/s41592-019-0638-x.

50. Tran, N.H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep Learning Enables de Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nat. Methods* **2019**, *16*, 63–66, doi:10.1038/s41592-018-0260-3.

51. Li, Y.; Zhong, C.Q.; Xu, X.; Cai, S.; Wu, X.; Zhang, Y.; Chen, J.; Shi, J.; Lin, S.; Han, J. Group-DIA: Analyzing Multiple Data-Independent Acquisition Mass Spectrometry Data Files. *Nat. Methods* **2015**, *12*, 1105–1106,

doi:10.1038/nmeth.3593.

52. Ting, Y.S.; Egertson, J.D.; Bollinger, J.G.; Searle, B.C.; Payne, S.H.; Noble, W.S.; MacCoss, M.J. PECAN: Library-Free Peptide Detection for Data-Independent Acquisition Tandem Mass Spectrometry Data. *Nat. Methods* **2017**, *14*, 903–908, doi:10.1038/nmeth.4390.

53. Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372, doi:10.1038/nbt.1511.

54. Lam, H.; Deutsch, E.W.; Eddes, J.S.; Eng, J.K.; King, N.; Stein, S.E.; Aebersold, R. Development and Validation of a Spectral Library Searching Method for Peptide Identification from MS/MS. *Proteomics* **2007**, *7*, 655–667, doi:10.1002/pmic.200600625.

55. Pulsar (Biognosys) Available online: https://biognosys.com.

56. SWATHatlas Available online: http://www.swathatlas.org (accessed on 10 August 2022).

57. Kusebauch, U.; Campbell, D.S.; Deutsch, E.W.; Chu, C.S.; Spicer, D.A.; Brusniak, M.Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; et al. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016**, *166*, 766–778, doi:10.1016/j.cell.2016.06.041.

58. Desiere, F.; Deutsch, E.W.; King, N.L.; Nesvizhskii, A.I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S.N.; Aebersold, R. The PeptideAtlas Project. *Nucleic Acids Res.* **2006**, *34*, D655–D658, doi:10.1093/nar/gkj040.

59. NIST Available online: https://chemdata.nist.gov/ (accessed on 10 August 2022).

60. Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K.K.; Deming, L.; Berndl, M.; Brant, A.; Cimermancic, P.; Cox, J. High-Quality MS/MS Spectrum Prediction for Data-Dependent and Data-Independent Acquisition Data Analysis. *Nat. Methods* **2019**, *16*, 519–525, doi:10.1038/s41592-019-0427-6.

61. Gessulat, S.; Schmidt, T.; Zolg, D.P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nat. Methods* **2019**, *16*, 509–518, doi:10.1038/s41592-019-0426-7.

62. Zhou, X.X.; Zeng, W.F.; Chi, H.; Luo, C.; Liu, C.; Zhan, J.; He, S.M.; Zhang, Z. PDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Anal. Chem.* **2017**, *89*, 12690–12697, doi:10.1021/acs.analchem.7b02566.

63. Fröhlich, K.; Brombacher, E.; Fahrner, M.; Vogele, D.; Kook, L.; Pinter, N.; Bronsert, P.; Timme-Bronsert, S.; Schmidt, A.; Bärenfaller, K.; et al. Benchmarking of Analysis Strategies for Data-Independent Acquisition Proteomics Using a Large-Scale Dataset Comprising Inter-Patient Heterogeneity. *Nat. Commun.* **2022**, *13*, 2622, doi:10.1038/s41467-022-30094-0.

64. Jain, K.K. Role of Proteomics in the Development of Personalized Medicine. In Proceedings of the Advances in Protein Chemistry and Structural Biology; Elsevier, 2016; Vol. 102, pp. 41–52.

65. Li, J.; Smith, L.S.; Zhu, H.J. Data-Independent Acquisition (DIA): An Emerging Proteomics Technology for Analysis of Drug-Metabolizing Enzymes and Transporters. *Drug Discov. Today Technol.* **2021**, *39*, 49–56, doi:10.1016/j.ddtec.2021.06.006.

66. Liebler, D.C.; Zimmerman, L.J. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry* **2013**, *52*, 3797–3806, doi:10.1021/bi400110b.

67. Marx, V. Targeted Proteomics. *Nat. Methods* **2013**, *10*, 19–22, doi:10.1038/nmeth.2285.

68. Sarkadi, B.; Price, E.M.; Boucher, R.C.; Germann, U.A.; Scarborough, G.A. Expression of the Human Multidrug Resistance CDNA in Insect Cells Generates a High Activity Drug-Stimulated Membrane ATPase. *J. Biol. Chem.* **1992**, *267*, 4854–4858, doi:10.1016/s0021-9258(18)42909-2.

69. Pál, Á.; Méhn, D.; Molnár, É.; Gedey, S.; Mészáros, P.; Nagy, T.; Glavinas, H.; Janáky, T.; Von Richter, O.; Báthori, G.; et al. Cholesterol Potentiates ABCG2 Activity in a Heterologous Expression System: Improved in Vitro Model to Study Function of Human ABCG2. *J. Pharmacol. Exp. Ther.* **2007**, *321*, 1085–1094, doi:10.1124/jpet.106.119289.

70. Shen, S.; An, B.; Wang, X.; Hilchey, S.P.; Li, J.; Cao, J.; Tian, Y.; Hu, C.; Jin, L.; Ng, A.; et al. Surfactant Cocktail-Aided Extraction/Precipitation/On-Pellet Digestion Strategy Enables Efficient and Reproducible Sample Preparation for Large-Scale Quantitative Proteomics. *Anal. Chem.* **2018**, *90*, 10350–10359, doi:10.1021/acs.analchem.8b02172.

71. Potriquet, J.; Laohaviroj, M.; Bethony, J.M.; Mulvenna, J. A Modified FASP Protocol for High-Throughput Preparation of Protein Samples for Mass Spectrometry. *PLoS One* **2017**, *12*, doi:10.1371/journal.pone.0175967.

72. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13*, 731–740, doi:10.1038/nmeth.3901.

73. Nolte, H.; MacVicar, T.D.; Tellkamp, F.; Krüger, M. Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Sci. Rep.* **2018**, *8*, 12648, doi:10.1038/s41598-018-31154-6.

74. Heberle, H.; Meirelles, V.G.; da Silva, F.R.; Telles, G.P.; Minghim, R. InteractiVenn: A Web-Based Tool for the Analysis of Sets through Venn Diagrams. *BMC Bioinformatics* **2015**, *16*, doi:10.1186/s12859-015-0611-3.

75. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25*, 25–29, doi:10.1038/75556.

76. Liu, H.; Ponniah, G.; Neill, A.; Patel, R.; Andrien, B. Accurate Determination of Protein Methionine Oxidation by Stable Isotope Labeling and LC-MS Analysis. *Anal. Chem.* **2013**, *85*, doi:10.1021/ac403072w.

77. Zhan, X.; Li, J.; Guo, Y.; Golubnitschaja, O. Mass Spectrometry Analysis of Human Tear Fluid Biomarkers Specific for Ocular and Systemic Diseases in the Context of 3P Medicine. *EPMA J.* **2021**, *12*, 449–475, doi:10.1007/s13167-021-00265-y.

78. Ponzini, E.; Santambrogio, C.; De Palma, A.; Mauri, P.; Tavazzi, S.; Grandori, R. Mass Spectrometry-based Tear Proteomics for Noninvasive Biomarker Discovery. *Mass Spectrom. Rev.* **2021**, 1–19, doi:10.1002/mas.21691.

79. Chen, L.; Zhou, L.; Chan, E.C.Y.; Neo, J.; Beuerman, R.W. Characterization of the Human Tear Metabolome by LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 4876–4882, doi:10.1021/pr2004874.

80. Zhou, L.; Beuerman, R.W. Tear Analysis in Ocular Surface Diseases. *Prog. Retin. Eye Res.* **2012**, *31*, 527–550, doi:10.1016/j.preteyeres.2012.06.002.

81. Zhou, L.; Beuerman, R.W. The Power of Tears: How Tear Proteomics Research Could Revolutionize the Clinic. *Expert Rev. Proteomics* **2017**, *14*, 189–191, doi:10.1080/14789450.2017.1285703.

82. Ma, J.Y.W.; Sze, Y.H.; Bian, J.F.; Lam, T.C. Critical Role of Mass Spectrometry Proteomics in Tear Biomarker Discovery for Multifactorial Ocular Diseases (Review). *Int. J. Mol. Med.* **2021**, *47*, 83, doi:10.3892/IJMM.2021.4916.

83. Rentka, A.; Koroskenyi, K.; Harsfalvi, J.; Szekanecz, Z.; Szucs, G.; Szodoray, P.; Kemeny-Beke, A. Evaluation of Commonly Used Tear Sampling Methods and Their Relevance in Subsequent Biochemical Analysis. *Ann. Clin. Biochem.* **2017**, *54*, 521–529, doi:10.1177/0004563217695843.

84. Kurihashi, K.; Yanagihara, N.; Honda, Y. A Modified Schirmer Test: The Fine-Thread Method for Measuring Lacrimation. *J. Pediatr. Ophthalmol.* **1977**, *14*, 390–397, doi:10.3928/0191-3913-19771101-15.

85. Barmada, A.; Shippy, S.A. Quantifying Sample Collection and Processing Impacts on Fiber-Based Tear Fluid Chemical Analysis. *Transl. Vis. Sci. Technol.* **2020**, *9*, 1–10, doi:10.1167/tvst.9.10.23.

86. Nättinen, J.; Aapola, U.; Jylhä, A.; Vaajanen, A.; Uusitalo, H. Comparison of Capillary and Schirmer Strip Tear Fluid Sampling Methods Using Swath-Ms Proteomics Approach. *Transl. Vis. Sci. Technol.* **2020**, *9*, 16, doi:10.1167/tvst.9.3.16.

87. Vashisht, S.; Singh, S. Evaluation of Phenol Red Thread Test versus Schirmer Test in Dry Eyes: A Comparative Study. *Int. J. Appl. Basic Med. Res.* **2011**, *1*, 40–42, doi:10.4103/2229-516x.81979.

88. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. In Proceedings of the IOP Conference Series: Materials Science and Engineering; Institute of Physics Publishing: Surabaya, Indonesia, November 2018; Vol. 336.

89. Ahmad, M.T.; Zhang, P.; Dufresne, C.; Ferrucci, L.; Semba, R.D. The Human Eye Proteome Project: Updates on an Emerging Proteome. *Proteomics* **2018**, *18*, 1700394, doi:10.1002/pmic.201700394.

90. Pieczyński, J.; Szulc, U.; Harazna, J.; Szulc, A.; Kiewisz, J. Tear Fluid Collection Methods: Review of Current Techniques. *Eur. J. Ophthalmol.* **2021**, *31*, 2245–2251, doi:10.1177/1120672121998922.

91. Saleh, T.A.; McDermott, B.; Bates, A.K.; Ewings, P. Phenol Red Thread Test vs

Schirmer's Test: A Comparative Study. *Eye* **2006**, *20*, 913–915, doi:10.1038/sj.eye.6702052.

92. Masmali, A.; Alqahtani, T.A.; Alharbi, A.; El-Hiti, G.A. Comparative Study of Repeatability of Phenol Red Thread Test Versus Schirmer Test in Normal Adults in Saudi Arabia. *Eye Contact Lens Sci. Clin. Pract.* **2014**, *40*, 127–131, doi:10.1097/ICL.0000000000000025.

93. Green-Church, K.B.; Nichols, K.K.; Kleinholz, N.M.; Zhang, L.; Nichols, J.J. Investigation of the Human Tear Film Proteome Using Multiple Proteomic Approaches. *Mol. Vis.* **2008**, *14*, 456–470.

94. García-Porta, N.; Mann, A.; Sáez-Martínez, V.; Franklin, V.; Wolffsohn, J.S.; Tighe, B. The Potential Influence of Schirmer Strip Variables on Dry Eye Disease Characterisation, and on Tear Collection and Analysis. *Contact Lens Anterior Eye* **2018**, *41*, 47–53, doi:10.1016/j.clae.2017.09.012.

95. Bertram, M.; Allbaugh, R.A.; Mochel, J.P.; Peraza, J.; Page, L.; Sebbag, L. Influence of Schirmer Strip Wetness on Volume Absorbed, Volume Recovered, and Total Protein Content in Canine Tears. *Vet. Ophthalmol.* **2021**, *24*, 425–428, doi:10.1111/vop.12876.

96. Avilov, V.; Zeng, Q.; Shippy, S.A. Threads for Tear Film Collection and Support in Quantitative Amino Acid Analysis. *Anal. Bioanal. Chem.* **2016**, *408*, 5309–5317, doi:10.1007/s00216-016-9624-7.

97. Stuchell, R.N.; Feldman, J.J.; Farris, R.L.; Mandel, I.D. The Effect of Collection Technique on Tear Composition. *Investig. Ophthalmol. Vis. Sci.* **1984**, *25*, 374–377.

98. Aass, C.; Norheim, I.; Eriksen, E.F.; Thorsby, P.M.; Pepaj, M. Single Unit Filter-Aided Method for Fast Proteomic Analysis of Tear Fluid. *Anal. Biochem.* **2015**, *480*, 1–5, doi:10.1016/j.ab.2015.04.002.

99. Posa, A.; Bräuer, L.; Schicht, M.; Garreis, F.; Beileke, S.; Paulsen, F. Schirmer Strip vs. Capillary Tube Method: Non-Invasive Methods of Obtaining Proteins from Tear Fluid. *Ann. Anat.* **2013**, *195*, 137–142, doi:10.1016/j.aanat.2012.10.001.

100. Choy, C.K.M.; Cho, P.; Chung, W.Y.; Benzie, I.F.F. Water-Soluble Antioxidants in Human Tears: Effect of the Collection Method. *Investig. Ophthalmol. Vis. Sci.* **2001**, *42*, 3130–3134.

101. Akkurt Arslan, M.; Kolman, I.; Pionneau, C.; Chardonnet, S.; Magny, R.; Baudouin, C.; Brignole-Baudouin, F.; Kessal, K. Proteomic Analysis of Tears and Conjunctival Cells Collected with Schirmer Strips Using Timstof pro: Preanalytical Considerations. *Metabolites* **2022**, *12*, 2, doi:10.3390/metabo12010002.

102. Kalló, G.; Emri, M.; Varga, Z.; Ujhelyi, B.; Tozsér, J.; Csutak, A.; Csosz, É. Changes in the Chemical Barrier Composition of Tears in Alzheimer's Disease Reveal Potential Tear Diagnostic Biomarkers. *PLoS One* **2016**, *11*, e0158000, doi:10.1371/journal.pone.0158000.

103. Jung, J.H.; Ji, Y.W.; Hwang, H.S.; Oh, J.W.; Kim, H.C.; Lee, H.K.; Kim, K.P. Proteomic Analysis of Human Lacrimal and Tear Fluid in Dry Eye Disease. *Sci. Rep.* **2017**, *7*, 13363, doi:10.1038/s41598-017-13817-y.

104. Tamhane, M.; Cabrera-Ghayouri, S.; Abelian, G.; Viswanath, V. Review of

Biomarkers in Ocular Matrices: Challenges and Opportunities. *Pharm. Res.* **2019**, *36*, 40, doi:10.1007/s11095-019-2569-8.

105. Zernii, E.Y.; Golovastova, M.O.; Baksheeva, V.E.; Kabanova, E.I.; Ishutina, I.E.; Gancharova, O.S.; Gusev, A.E.; Savchenko, M.S.; Loboda, A.P.; Sotnikova, L.F.; et al. Alterations in Tear Biochemistry Associated with Postanesthetic Chronic Dry Eye Syndrome. *Biochem.* **2016**, *81*, 1549–1557, doi:10.1134/S0006297916120166.

106. Hagan, S.; Martin, E.; Enríquez-de-Salamanca, A. Tear Fluid Biomarkers in Ocular and Systemic Disease: Potential Use for Predictive, Preventive and Personalised Medicine. *EPMA J.* **2016**, *7*, 15, doi:10.1186/s13167-016-0065-3.

107. Tomlinson, A.; Blades, K.J.; Pearce, E.I. What Does the Phenol Red Thread Test Actually Measure? *Optom. Vis. Sci.* **2001**, *78*, 142–146, doi:10.1097/00006324-200103000-00005.

108. Jetter, A.; Kullak-Ublick, G.A. Drugs and Hepatic Transporters: A Review. *Pharmacol. Res.* 2020, *154*.

109. Austin Doyle, L.; Yang, W.; Abruzzo, L. V.; Krogmann, T.; Gao, Y.; Rishi, A.K.; Ross, D.D. A Multidrug Resistance Transporter from Human MCF-7 Breast Cancer Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 15665–15670, doi:10.1073/pnas.95.26.15665.

110. Toyoda, Y.; Takada, T.; Suzuki, H. Inhibitors of Human ABCG2: From Technical Background to Recent Updates with Clinical Implications. *Front. Pharmacol.* **2019**, *10*, 208, doi:10.3389/fphar.2019.00208.

111. Shitara, Y. Clinical Importance of OATP1B1 and OATP1B3 in Drugdrug Interactions. *Drug Metab. Pharmacokinet.* **2011**, *26*, 220–227, doi:10.2133/dmpk.DMPK-10-RV-094.

112. *The Transporter Book*; Péter, T., Joseph, K.Z., Zsuzsanna, G., Wilde, R. de, Noémi, P., Eds.; 4th Editio.; SOLVO Biotechnology: Budapest, 2021;

113. Schaefer, O.; Ohtsuki, S.; Kawakami, H.; Inoue, T.; Liehner, S.; Saito, A.; Sakamoto, A.; Ishiguro, N.; Matsumaru, T.; Terasaki, T.; et al. Absolute Quantification and Differential Expression of Drug Transporters, Cytochrome P450 Enzymes, and UDP-Glucuronosyltransferases in Cultured Primary Human Hepatocytes. *Drug Metab. Dispos.* **2012**, *40*, 93–103, doi:10.1124/dmd.111.042275.

114. Ohtsuki, S.; Schaefer, O.; Kawakami, H.; Inoue, T.; Liehner, S.; Saito, A.; Ishiguro, N.; Kishimoto, W.; Ludwig-Schwellinger, E.; Ebner, T.; et al. Simultaneous Absolute Protein Quantification of Transporters, Cytochromes P450, and UDP-Glucuronosyltransferases as a Novel Approach for the Characterization of Individual Human Liver: Comparison with MRNA Levels and Activities. *Drug Metab. Dispos.* **2012**, *40*, 83–92, doi:10.1124/dmd.111.042259.

115. *The IUPAC Compendium of Chemical Terminology*; IUPAC, 2019;

116. Food and Drug Administration (FDA) Guidance for Industry Studies to Evaluate the Metabolism and Residue Kinetics of Veterinary Drugs In Studies to Establish Product Withdrawal Periods - VICH GL49(R). **2010**.

117. Pino, L.K.; Searle, B.C.; Yang, H.Y.; Hoofnagle, A.N.; Noble, W.S.; MacCoss, M.J. Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in

Quantitative Proteomics. *J. Proteome Res.* **2020**, *19*, 1147–1153, doi:10.1021/acs.jproteome.9b00666.

118. Savitski, M.M.; Mathieson, T.; Zinn, N.; Sweetman, G.; Doce, C.; Becher, I.; Pachl, F.; Kuster, B.; Bantscheff, M. Measuring and Managing Ratio Compression for Accurate ITRAQ/TMT Quantification. *J. Proteome Res.* **2013**, *12*, 3586–3598, doi:10.1021/pr400098r.

119. Kamiie, J.; Ohtsuki, S.; Iwase, R.; Ohmine, K.; Katsukura, Y.; Yanai, K.; Sekine, Y.; Uchida, Y.; Ito, S.; Terasaki, T. Quantitative Atlas of Membrane Transporter Proteins: Development and Application of a Highly Sensitive Simultaneous LC/MS/MS Method Combined with Novel in-Silico Peptide Selection Criteria. *Pharm. Res.* **2008**, *25*, 1469–1483, doi:10.1007/s11095-008-9532-4.

120. Gröer, C.; Brück, S.; Lai, Y.; Paulick, A.; Busemann, A.; Heidecke, C.D.; Siegmund, W.; Oswald, S. LC-MS/MS-Based Quantification of Clinically Relevant Intestinal Uptake and Efflux Transporter Proteins. *J. Pharm. Biomed. Anal.* **2013**, *85*, 253–261, doi:10.1016/j.jpba.2013.07.031.

121. Kumar, V.; Prasad, B.; Patilea, G.; Gupta, A.; Salphati, L.; Evers, R.; Hop, C.E.C.A.; Unadkat, J.D. Quantitative Transporter Proteomics by Liquid Chromatography with Tandem Mass Spectrometry: Addressing Methodologic Issues of Plasma Membrane Isolation and Expression-Activity Relationship. *Drug Metab. Dispos.* **2015**, *43*, 284–288, doi:10.1124/dmd.114.061614.

122. Jankovskaja, S.; Kamiie, J.; Rezeli, M.; Gustavsson, L.; Sugihara, Y.; Miliotis, T.; Ruzgas, T.; Marko-Varga, G. Optimization of Sample Preparation for Transporter Protein Quantification in Tissues by LC–MS/MS. *J. Pharm. Biomed. Anal.* **2019**, *164*, 9–15, doi:10.1016/j.jpba.2018.10.013.

123. Santoni, V.; Molloy, M.; Rabilloud, T. Membrane Proteins and Proteomics: Un Amour Impossible? *Electrophoresis* **2000**, *21*, doi:10.1002/(SICI)1522-2683(20000401)21:6<1054::AID-ELPS1054>3.0.CO;2-8.

124. Omasits, U.; Ahrens, C.H.; Müller, S.; Wollscheid, B. Protter: Interactive Protein Feature Visualization and Integration with Experimental Proteomic Data. *Bioinformatics* **2014**, *30*, 884–886, doi:10.1093/bioinformatics/btt607.

125. Cañas, B.; Piñeiro, C.; Calvo, E.; López-Ferrer, D.; Gallardo, J.M. Trends in Sample Preparation for Classical and Second Generation Proteomics. *J. Chromatogr. A* **2007**, *1153*, 235–258, doi:10.1016/j.chroma.2007.01.045.

126. Duan, X.; Young, R.; Straubinger, R.M.; Page, B.; Cao, J.; Wang, H.; Yu, H.; Canty, J.M.; Qu, J. A Straightforward and Highly Efficient Precipitation/on-Pellet Digestion Procedure Coupled with a Long Gradient Nano-LC Separation and Orbitrap Mass Spectrometry for Label-Free Expression Profiling of the Swine Heart Mitochondrial Proteome. *J. Proteome Res.* **2009**, *8*, 2838–2850, doi:10.1021/pr900001t.

127. Klont, F.; Bras, L.; Wolters, J.C.; Ongay, S.; Bischoff, R.; Halmos, G.B.; Horvatovich, P. Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. *Anal. Chem.* **2018**, *90*, 5405–5413, doi:10.1021/acs.analchem.8b00600.

128. Searle, B.C.; Egertson, J.D.; Bollinger, J.G.; Stergachis, A.B.; MacCoss, M.J. Using

Data Independent Acquisition (DIA) to Model High-Responding Peptides for Targeted Proteomics Experiments. *Mol. Cell. Proteomics* **2015**, *14*, 2331–2340, doi:10.1074/mcp.M115.051300.

129. Abbatiello, S.E.; Schilling, B.; Mani, D.R.; Zimmerman, L.J.; Hall, S.C.; MacLean, B.; Albertolle, M.; Allen, S.; Burgess, M.; Cusack, M.P.; et al. Large-Scale Interlaboratory Study to Develop, Analytically Validate and Apply Highly Multiplexed, Quantitative Peptide Assays to Measure Cancer-Relevant Proteins in Plasma. *Mol. Cell. Proteomics* **2015**, *14*, 2357–2374, doi:10.1074/mcp.M114.047050.

130. You, J.; Fitzgerald, A.; Cozzi, P.J.; Zhao, Z.; Graham, P.; Russell, P.J.; Walsh, B.J.; Willcox, M.; Zhong, L.; Wasinger, V.; et al. Post-Translation Modification of Proteins in Tears. *Electrophoresis* **2010**, *31*, 1853–1861, doi:10.1002/elps.200900755.

131. Van Setten, G.B.; Stephens, R.; Tervo, T.; Salonen, E.M.; Tarkkanen, A.; Vaheri, A. Effects of the Schirmer Test on the Fibrinolytic System in the Tear Fluid. *Exp. Eye Res.* **1990**, *50*, 135–141, doi:10.1016/0014-4835(90)90223-H.

132. Cho, P.; Yap, M. Schirmer Test. I. A Review. *Optom. Vis. Sci.* **1993**, *70*, doi:10.1097/00006324-199302000-00011.

133. Doughty, M.J.; Whyte, J.; Li, W. The Phenol Red Thread Test for Lacrimal Volume - Does It Matter If the Eyes Are Open or Closed? *Ophthalmic Physiol. Opt.* **2007**, *27*, 482–489, doi:10.1111/j.1475-1313.2007.00500.x.

10. Supplementary material

Supplementary Table 1. Characteristics of tear collection methods for proteomics analysis.

Parameter	Glass capillary (CAP)	Schirmer's strip (SU)	PRT
Average protein amount recovered*	60 µg	58 µg	29 µg
Number of proteins identidfied*	422	1225	1439
Average intracellular „contaminant" protein MS intensity*	1.2%	1.5%	0.7%
Median RSD of major proteins*	70%	77%	64%
Tear volume collected	3–10 µL [130]	~9 µL/cm [95]	~3–4 µL [96]
Tear collection time	Up to 5 min [90]	5 min [91]	15–20 sec [91]
Contact with eye	No/minimal [93,97]	Strong contact with cornea, conjunctiva and lower eye-lid [83,97,131]	Mild contact with cornea, conjunctiva and lower eye-lid
Site of tear sampling	Variable [83,90]	Variable [83,90]	Standard[85,90,91]
Diagnostic repeatibility (volume)		Poor** [91,132]	Good [92]
Open/closed eye	Open	Open/Closed	Open/Closed [133]
Invasiveness	Low [100]	High [100]	Low/Medium
Discomfort (subjective)	No [99]	Yes [91]	No [91,92]
Sensitivity in detecting dry eye disease	Low	Low [91]	High [92]
Induces reflex tearing with higher flow rate	No [83]	Yes [91,100]	Minimally [91]
Sample processing	–	Cutting and protein extraction or centrifugation [78,98]	Protein extraction*
Sampling device material/source	Variable	Variable [94]	Uniform [90]
Risk of injury	Yes	No [99]	No
Cellular/plasma contamination	Low [86,93]	High [80,86,93,97,131]	Low*
Tear collection requires specialist	Yes [99]	No	No
Continuous intervention during sample collection	Yes [99]	No [99]	No

* Data from current work, ** Can be improved with anesthesia (Type II test) [130]

Supplementary Table 2. Summary of the number and summed intensities of quantified proteins and peptides in the different cell disruption protocols. nTC=4; n PE =4; nVesicle=5.

	All proteins			All peptides		
	Number	Sum intensity		Number	Sum intensity	
	Mean ± SD	Mean	RSD	Mean ± SD	Mean	RSD
TC	5803 ± 140	4.87E+09	1.89	52912 ± 2334	3.07E+10	1.49
PE	5567 ± 339	5.00E+09	1.01	48205 ± 5188	2.84E+10	3.32
Vesicle	6061 ± 41	4.58E+09	1.54	57870 ± 1171	2.44E+10	3.17

	Membrane proteins			Membrane peptides		
	Number	Sum intensity		Number	Sum intensity	
	Mean ± SD	Mean	RSD	Mean ± SD	Mean	RSD
TC	1890 ± 88	1.20E+09	3.03	14920 ±1157	6.69E+09	4.22
PE	2120 ±210	2.00E+09	16.64	17689 ± 3011	1.00E+10	16.01
Vesicle	2402 ± 16	1.95E+09	2.27	22620 ± 509	9.83E+09	2.03

	ABC + SLC proteins			ABC + SLC peptides		
	Number	Sum intensity		Number	Sum intensity	
	Mean ± SD	Mean	RSD	Mean ± SD	Mean	RSD
TC	85 ± 6	4.04E+07	14.91	456 ± 41	2.29E+08	16.63
PE	122 ± 18	1.08E+08	19.76	771 ± 148	6.12E+08	19.64
Vesicle	145 ± 2	1.44E+08	6.27	1011 ± 38	8.83E+08	5.06

Supplementary Table 3. Calculated enrichment factors of the marker proteins in the different type of samples.

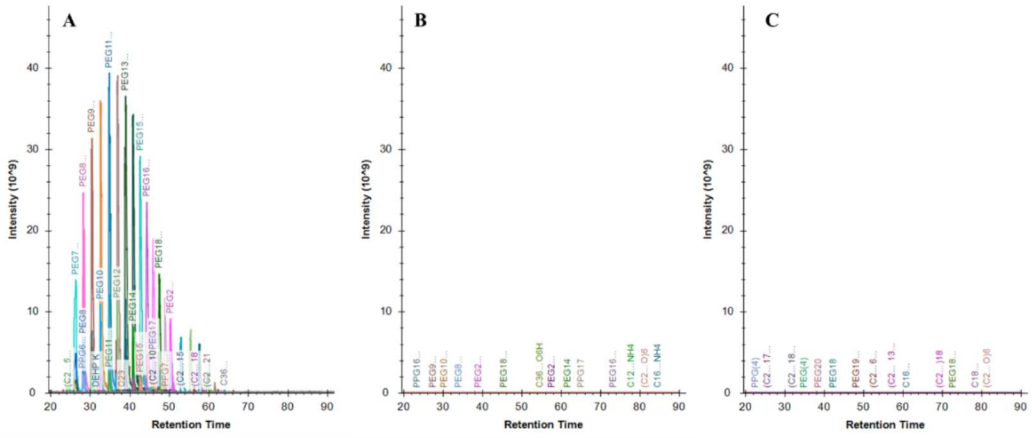
	Vesicle	ProteoExtract	Total Cell
	Mean ± SD	Mean ± SD	Mean ± SD
Cytosol (GAPDH)	0.3 ± 0.1	0.5 ± 0.1	1.0 ± 0.1
Cell membrane (ATP1A1)	2.8 ± 0.5	4.0 ± 0.9	1.0 ± 0.1
ER (CALR)	0.8 ± 0.1	1.3 ± 0.2	1.0 ± 0.3
Golgi (GOSR1)	6.1 ± 0.3	3.3 ± 1.8	1.0 ± 0.3
Nucleus (H4C2)	0.1 ± 0.05	0.3 ± 0.3	1.0 ± 0.2
Mitoch. (TIMM23)	2.0 ± 0.2	2.7 ± 1.0	1.0 ± 0.3

Supplementary Table 4. The reproducibility of the measured intensities (CV) of the peptides and proteins applying different protocols.

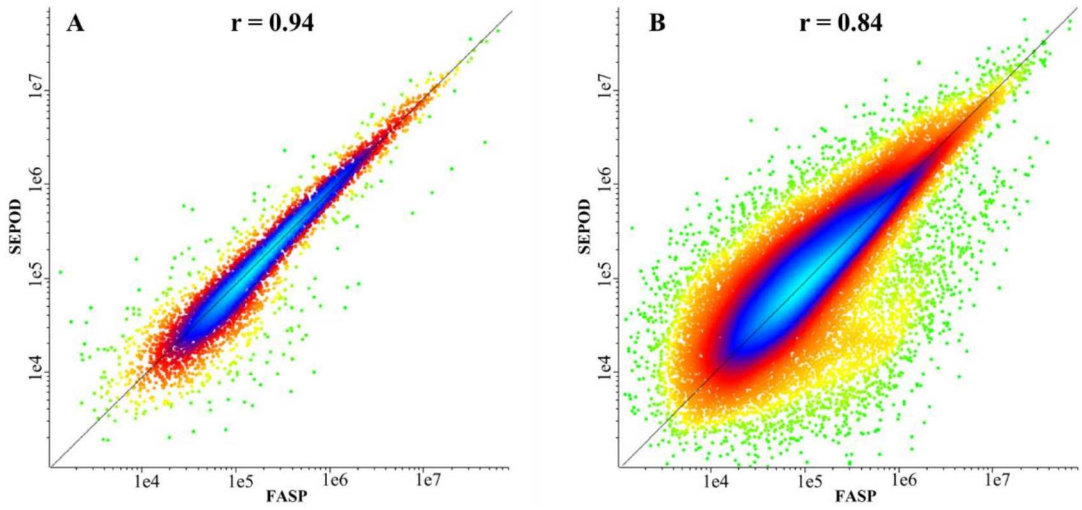
		All		Membrane		ABC + SLC	
		Proteins	Peptides	Proteins	Peptides	Proteins	Peptides
PE	FASP	0.16	0.27	0.16	0.27	0.19	0.25
	SEPOD	0.11	0.20	0.11	0.20	0.14	0.19
Vesicle	FASP	0.17	0.30	0.17	0.30	0.20	0.28
	SEPOD	0.10	0.19	0.09	0.18	0.11	0.18

Supplementary Table 5. BCRP content of the BCRP-containing vesicle samples (pmol / mg membrane protein).

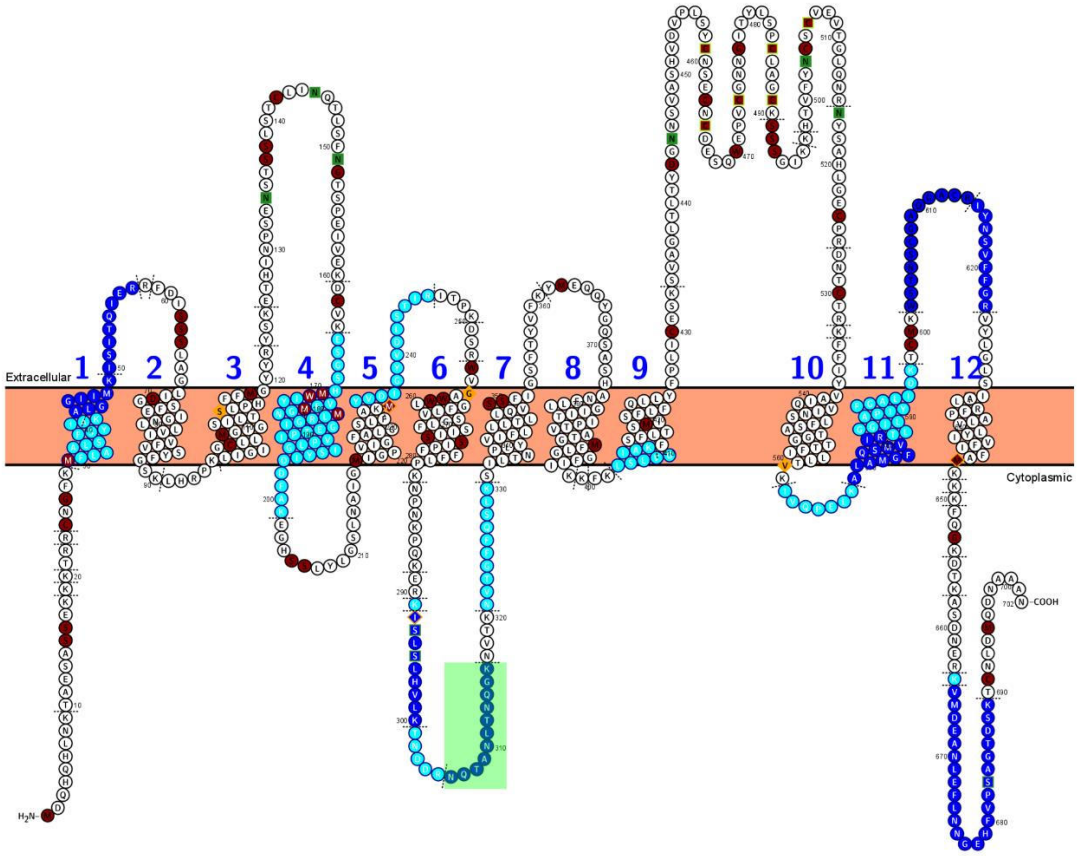
BCRP-HEK293	BCRP-M	BCRP-Sf9	BCRP-Sf9-HAM
95.47	127.884	80.36	109.592
214.976	149.148	181.914	132.832
170.616	153.232	124.038	125.114
179.61	200.738		
205.636			
128.262			



Supplementary Figure 1. Chromatograms of molecular contaminants [128] after direct in-solution (A) and FASP (B) and SEPOD (C) digestion of the same PE membrane sample.



Supplementary Figure 2. Comparison of the intensities of the measured proteins (**A**) and peptides (**B**) with the tested digestion protocols. The mean intensities measured in SEPOD samples on a log10 scale are plotted on the y-axis, against the mean intensities measured in FASP on the x-axis. Pearson correlation coefficients are shown at the top of the figure.



Supplementary Figure 3. Location of OATP1B3 protein in the membrane. The sequences marked in blue could be identified with the DIA measurement, the most intensively detectable sections marked in dark blue were included in the further examinations. The peptide selected for the absolute quantitative measurements marked with green.

APPENDIX I.

International Journal of
Molecular Sciences

MDPI

Article

An Extensive Study of Phenol Red Thread as a Novel Non-Invasive Tear Sampling Technique for Proteomics Studies: Comparison with Two Commonly Used Methods

Gábor Kecskeméti [1], Edit Tóth-Molnár [2], Tamás Janáky [1] and Zoltán Szabó [1,*]

[1] Department of Medical Chemistry, Albert Szent-Györgyi Medical School, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary; kecskemeti.gabor@med.u-szeged.hu (G.K.); janaky.tamas@med.u-szeged.hu (T.J.)
[2] Department of Ophtalmology, Albert Szent-Györgyi Health Centre, University of Szeged, Korányi Fasor 10-11, H-6720 Szeged, Hungary; toth-molnar.edit@med.u-szeged.hu
* Correspondence: szabo.zoltan@med.u-szeged.hu

Abstract: Tear samples are considered in recent publications as easily, noninvasively collectible information sources for precision medicine. Their complex composition may aid the identification of biomarkers and the monitoring of the effectiveness of treatments for the eye and systemic diseases. Sample collection and processing are key steps in any analytical method, especially if subtle personal differences need to be detected. In this work, we evaluate the usability of a novel sample collection technique for human tear samples using phenol red threads (cotton thread treated with the pH indicator phenol red), which are efficiently used to measure tear volume in clinical diagnosis. The low invasiveness and low discomfort to the patients have already been demonstrated, but their applicability for proteomic sample collection has not yet been compared to other methods. We have shown, using various statistical approaches, the qualitative and quantitative differences in proteomic samples collected with this novel and two traditional methods using either glass capillaries or Schirmer's paper strips. In all parameters studied, the phenol red threads proved to be equally or even more suitable than traditional methods. Based on detectability using different sampling methods, we have classified proteins in tear samples.

Keywords: LC-MS; tear; proteomics; data independent analysis; mass spectrometry

Citation: 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. https://doi.org/10.3390/ijms23158647

Academic Editors: De-Kuang Hwang and Shih-Jen Chen

Received: 27 June 2022
Accepted: 30 July 2022
Published: 3 August 2022

1. Introduction

The paradigm change from reactive to predictive, preventive, and personalized medicine requires novel and reliable methods that can provide a more precise diagnosis and patient stratification, detect early disease, elucidate the risk of disease, and predict disease outcome, response to therapy, and permit monitoring of therapeutic management. To this end, the discovery and validation of specific biomarkers/biomarker panels would be a promising approach. The source of biomarkers is crucial for the specificity, sensitivity, accuracy, and reliability of diagnostic tests and treatment targets [1]. 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 [2]. Tear fluid contains proteins, such as enzymes, mucins, hormones, immunoglobulins, growth factors, neuropeptides, and cytokines along with lipids, salts, and carbohydrates [3]. This comprehensive biomolecule repertoire in human tears serves as a good source for biomarker discovery for diseases. The biggest advantage is that tears are proximal to the disease location (such as ocular surface disease, lacrimal gland disease, etc.) contrary to, e.g., cancer biomarkers in blood, where the related biomarker molecules could be distant from the source, and are highly diluted [4,5]. The tear fluid proteomic profile has been found to provide basic biological information for many ocular diseases, such as dry eye syndrome, blepharitis,

keratoconus, thyroid eye disease, vernal keratoconjunctivitis, diabetic retinopathy, and primary open angle glaucoma [1,6]. Even though tear components are mainly derived from secretory glands, such as the lacrimal glands, the change of tear film composition is not only regulated by its secretion units. Protein molecules can enter the tear fluid through conjunctival blood vessels and, due to the overlap between the tear and plasma proteome, there may be opportunities to observe systemic responses in the tears. Several such examples thus far include breast cancer, type 2 diabetes, Alzheimer's disease, multiple sclerosis, and rheumatoid arthritis [1,6]. Some systemic diseases may affect the eye so we can use 'tears' as a 'window' to assess systemic as well as ocular 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, threads, ophthalmic sponges, and polyester rods [7]. During capillary sampling, one end of the glass capillary is placed in the meniscus of the tear fluid and due to the capillary action, the tear flows from the conjunctival sac into the interior of the glass capillary. The Schirmer's strip was originally a standard clinical tool used in many places to measure tear fluid volume, but it was found to be suitable for collecting tear samples too. 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 (e.g., aqueous tear-deficient patients or experimental animals). Kurihashi et al. established a method of tear secretion measurement using fine cotton threads [8] which might meet the above requirements, and already proved to be an applicable sample collection method for analysis of small molecules [9].

The advancement in nano-scale liquid chromatography coupled MS (nanoLC-MS) that provides improved chromatographic separation of peptides, higher sensitivity, and extended dynamic ranges to identify > 1500 proteins, has opened up the possibility of tear biomarker research [6]. There are numerous studies and reviews in the literature on the investigation of protein profiles from differently collected tear samples using liquid chromatography coupled to mass spectrometry (LC-MS) [1,2,4,5,7]. It was shown that samples collected by Schirmer's strip and capillary method have large quantitative and qualitative differences in their protein composition [6], therefore any new sample collection method should be evaluated in that respect too.

The goal of this study was to compare a new tear sampling procedure using phenol red thread (cotton thread treated with the pH indicator phenol red) (PRT) with the two most frequently used methods for sample collection in order to determine the protein profile of tears by liquid chromatography-mass spectrometry.

To the best of our knowledge, this is the first comprehensive study on the comparison of the commonly used Schirmer's strip, microcapillary, and the novel PRT method in human tear proteomics analysis.

2. Results

2.1. The Choice of Sampling Method

Our goal was to introduce a reliable new sampling procedure, which provides us with a convenient way of collecting enough tear samples for quantitative proteomic studies. In this comparative study tears of the same 10 healthy donors were collected by three sampling methods and analyzed by the same proteomics procedure. General protocols for the use of glass capillary tubes (CAP) and Schirmer's strips [10] were followed, and PRT as a new tear sampling device was included in this study. After sample collection, Schirmer's strips were divided into a lower section (SL) which was in direct contact with the surface of the eyeball and eyelid, and the following 10 mm long upper section (SU) and processed separately.

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. The total volume or protein content of all PRT, SU, and SL samples was satisfactory for our protocols; however, one SL sample was damaged during processing. In the case of two donors, the volume of tears collected with glass capillary tubes was insufficient (less than 1 µL) to determine the total protein content and perform proteomic sample preparation.

Both the volume and the total protein content of tears collected with the CAP sampling procedure showed large variance (7.9 ± 7.0 µL and 64.1 ± 44.9 µg, respectively). Tears are collected slowly and erratically, with interruptions due to blinking and eye movement, what—apart of individual variance—may explain the large fluctuation [4]. In contrast, the total protein content after the extraction of SL strips, 10 mm long SU strips, and whole PRT samples proved to be 42.0 ± 11.4 µg, 57.8 ± 15.5 µg, and 28.9 ± 11.2 µg, respectively.

2.2. Impact of MS Data Acquisition and Evaluation Method

Mass-spectrometry-based proteomics enables us to identify and quantify hundreds of thousands of proteins from different samples. However, the quality of the results is highly dependent on the experimental and computational workflows. A large number of proteomics software tools and algorithms have been published for data-independent acquisition (DIA) proteomics data processing. We applied DIA-NN software for spectral library creation, protein identification, and quantification, which is an integrated software suite, that exploits deep neural networks and new quantification and signal correction strategies [11]. Our DIA-NN workflow first was set up to use sample-specific spectral libraries generated by refining predicted libraries using six gas-phase-fractionated acquisitions (GPF-DIA) with 4 m/z precursor isolation windows from all four types of pooled samples. These sample-specific spectral libraries demonstrate the deepest proteome coverage of a given sample type. A combined spectral library was built by searching all four groups of GPF measurements in one against the same predicted spectral library. This combined library contains a total of 2583 proteins, any of which were present in a quantifiable amount in at least one pooled tear sample.

Data of DIA acquisitions collected from the analysis of individual samples (DIA-Q) were searched against both the sample-specific and combined libraries. The number of quantified peptides was 66.1 ± 42.9% higher when the search was carried out on the combined library. Searching against the combined library had a lower effect for SL samples than the PRT, SU, and CAP, as these samples contained the highest number of quantifiable peptides and contributed to the greatest extent to the size of the library.

Although the combined library originating from the GPF-DIA measurements from four pooled samples contain 2583 proteins, in the individual samples 1144 could be quantified with at least two peptides in the DIA-Q measurements.

2.3. Comparison of the Tear Proteomes of Different Sample Types

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. All quantifiable proteins were included in this comparison, regardless of the number of peptides detected.

The dynamic range and contribution of proteins in different types of samples to the combined library are shown in Figure 1. The library covers proteins with a range of more than six orders of magnitude in summed signal intensity. Number of proteins shown on the Venn diagram refers to number of protein groups, which may consist of more than one protein, and cannot be differentiated based on identified peptides. The lowest number of proteins could be identified in the pooled CAP sample (422 protein), while in the two indirect pooled tear samples (PRT and SU), a similar number of proteins could be detected (1439 and 1225) with a much higher overlap (1092 shared protein identifications).

The SL samples were only included in the investigation to detect possible contaminants originating from the strong direct contact of the Schirmer's strip with the ocular surface. This assumption is supported by the presence of nearly 1000 proteins, which were identified only in this type of pooled sample. These proteins are present in low abundance, according to their summed intensity (green dots in Figure 1). It must be noted that the parts of the PRTs in contact with the eye were not removed, which may explain the approximately 300 proteins shared by only PRT and SL samples (red dots in Figure 1). The majority of the most intense (10^6–10^{10}) proteins could be identified in all sample types, but there are many common proteins with lower intensity. Altogether, 341 protein groups could be detected in all the four sample types (ochre dots in Figure 1). The list and annotations of all proteins in the combined spectral library can be found in Table S1 in the Supplementary Materials.

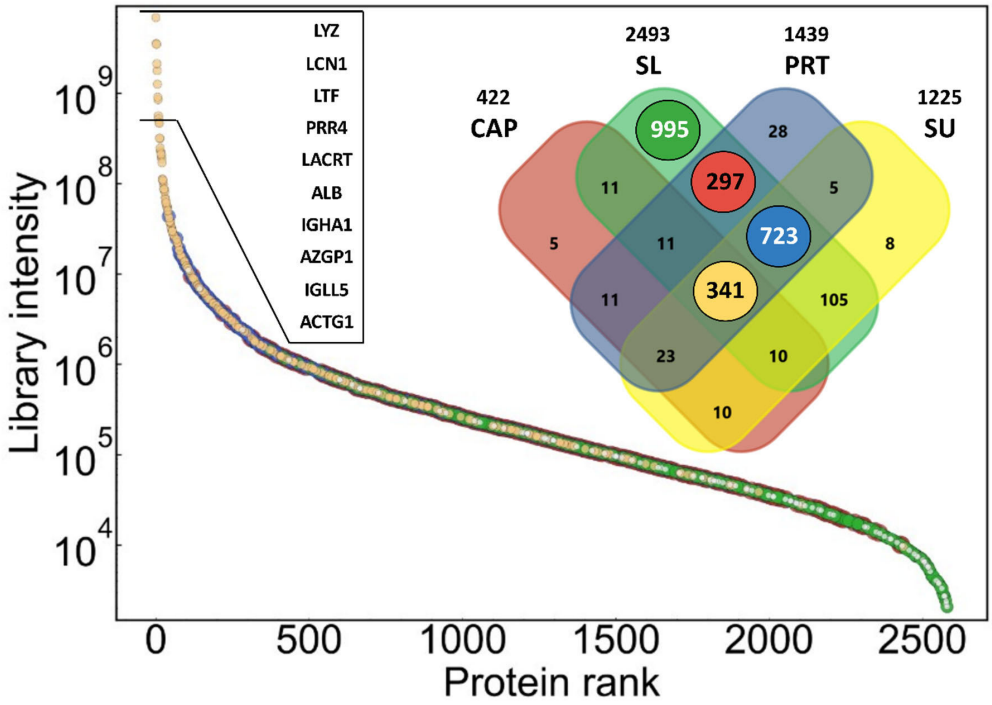


Figure 1. Summed intensity of proteins in combined spectral library as a function of protein intensity rank, demonstrating the dynamic range of identified proteins. Data points are colored according to highlighted subsets of the library as shown in the Venn diagram insert, to demonstrate contribution of different sample types to the library. Gene names of the top proteins are shown in the order of intensity rank.

To study the quantitative similarity of different sample types, Pearson correlation coefficient values were calculated for the proteins quantified in sample type-specific pooled samples (Figure 2). This assay shows a strong correlation between PRT and SU samples (r = 0.90), but the pooled CAP sample also correlates satisfactorily with these samples (r = 0.76–0.78).

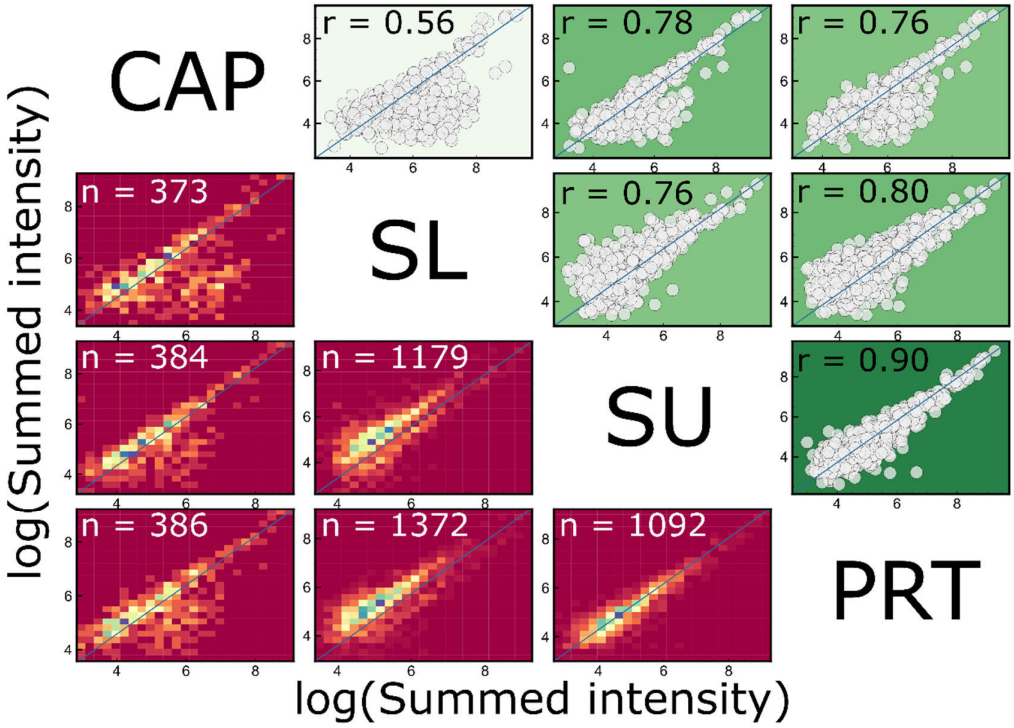


Figure 2. Correlation of intensities of proteins in different pooled samples. The depth of the background colors of scatter points on top-right is proportional to the value of actual Pearson correlation coefficients, which are shown on the 2D histograms at bottom-left. The number of the common sets of proteins (*n*) are given.

2.4. Clustering of Proteins in Different Types of Samples

The analysis of the spectral library created from the analysis of four different types of pooled samples in the previous section already highlighted the diversity of proteomes determined in samples collected by different methods, but the statistical classification of proteins based on sample-specific detectability is only possible on analysis of individual samples. It must be noted that the applied DIA approach, with match-between-run identifications enabled, can minimize the technical reasons for missing values, therefore this analysis reflects heterogeneity in sample composition.

A k-means cluster analysis based on sample-type specific detection frequency (in %)—was performed to classify the 1144 quantifiable proteins into four clusters (the number of clusters was predicted using the elbow method [12]) (Figure 3).

In Cluster A (pink), there were 195 proteins that could be measured with high frequency in all sample types (86–92% of the samples within any sample type). Cluster B (dark green) includes 242 proteins that were repeatedly measurable (84–98%) in samples from indirect procedures (SL, SU, and PRT), but they were quantifiable in only a few CAP samples (8%). 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. Using this approach, proteins within the latest cluster may be marked as possible contaminants of tear samples. They have been found mainly in the SL samples, but rarely in any other sample group. On the other hand, proteins of Cluster A and B may be considered as common tear proteins, with the remark that CAP sampling provides reproducible detection for members of only Cluster A. It must be noted that PRT sampling produces the highest detection rate for proteins of those two clusters (see distributions on boxplots in Figure 3).

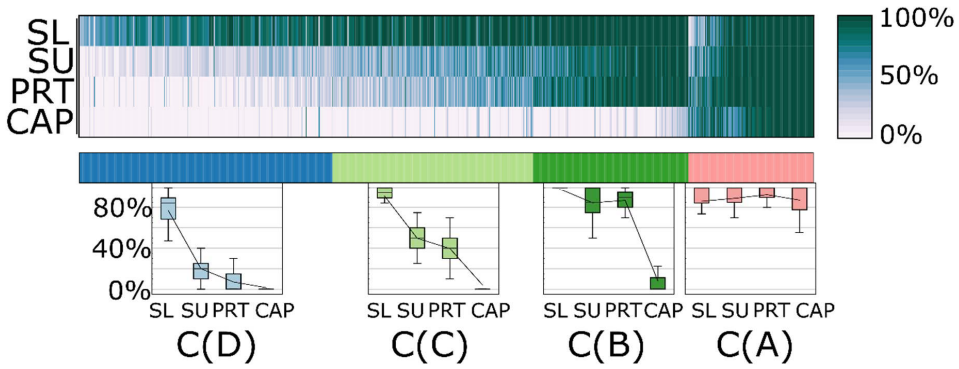


Figure 3. k-means (k = 4) cluster analysis of the quantified proteins based on the percentage of observations in samples of the different sampling procedures. Detection frequency of each protein is given as a heatmap on top. The boxplots on the bottom show the distributions of detectability of proteins in different sample types within each cluster.

We have compared the summed relative intensities of the identified protein clusters in each type of sample (Figure 4). As expected, the highest level of the possible contaminating eye-surface proteins (395 and 317 proteins in Clusters D and C, respectively) were found in the SL samples, where on average 6.0% of the total intensity of all proteins is given by the proteins of Cluster D. CAP, SU, and PRT samples contain an average of 1.2, 1.5, and 0.7% proteins of that cluster. Proteins of Cluster C can be found at higher abundance in each sample type. Their values are still less than 5%, except for the SL samples (13% of total intensity). Proteins of Cluster B were detected only in a few CAP samples, and their intensity was low (on average 2%), while their relative abundance was highest in SL samples again (32% average). 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. For all clusters, the variance of PRT samples is lower than those in SU, but slightly higher than in CAP samples, which, however, have a less complex protein composition.

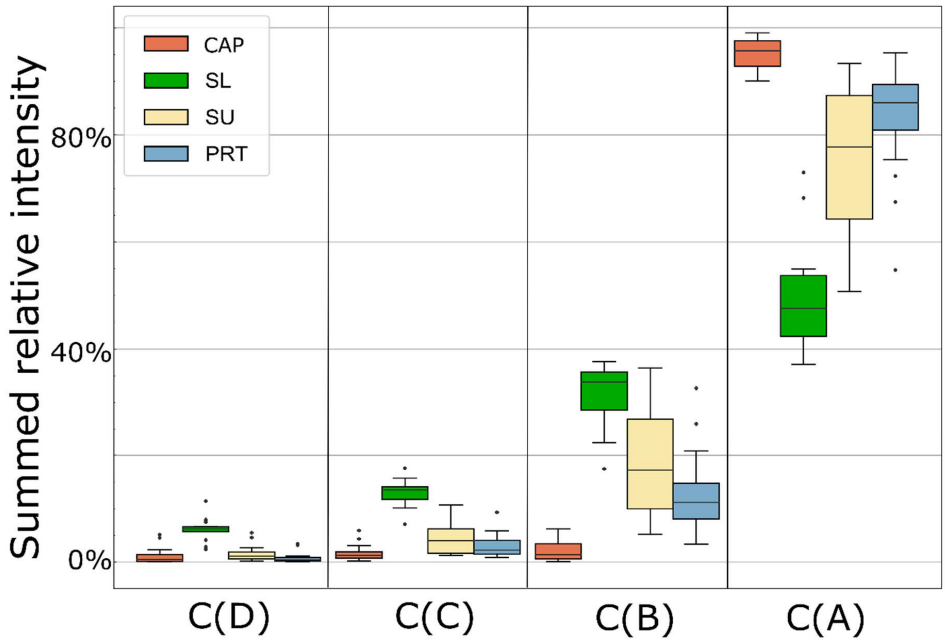


Figure 4. Boxplot of summed relative intensity of the four protein clusters in each sample type (n_{CAP} = 18, n_{SL} = 19, n_{SU} = 20, n_{PRT} = 20).

The list, quantitative data, annotations, and cluster assignments of all proteins quantified in the individual samples can be found in Table S2 in the Supplementary Materials.

2.5. Classification of Tear Proteins

Based on the detection frequency of proteins in samples collected using different sampling methods, we identified four clusters of proteins in the previous section. In order to give biological classification of the proteins in tear samples, each protein was annotated with available GO and UniProt terms. Fisher's enrichment analysis was performed on these terms to identify common properties of proteins within each cluster. Here we discuss a few examples of the significantly enriched categories (Benjamini–Hochberg corrected FDR < 2%), which are different in different clusters and relevant to tear and tear sampling.

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. Eighty-five percent of the different Ig chains (36 of 40) detected in our samples are in Cluster A, which makes 17% of the proteins in this cluster. Most of the identified keratins (10 out of 12) frequently occur in all sample types, thus they are found in Cluster A.

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 [13] database collects a list of proteins identified in different parts of the human eye. For the classification of proteins in tear samples based on their possible origin, two groups were created: proteins which can be found in the 'Tears' section (including 1506 proteins) and an eye surface group from the 'Cornea' or 'Sclera' sections of the EyeOME database (1469 and 1895 proteins respectively). There is a large overlap in these assignments, 1213 proteins are common to tears and the eye surface in that database. Immunoglobulins are excluded from EyeOME, that is the main reason that not all, but 1093 proteins of the database were among the 1144 quantified proteins from our experiments, 846 of which are common to tears and the eye surface. Based on the overlap of clusters identified in Section 2.4 and the EyeOME assignments (shown in Figure 5), we can make further refinements of protein classification. The majority 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 (green in Figure 5).

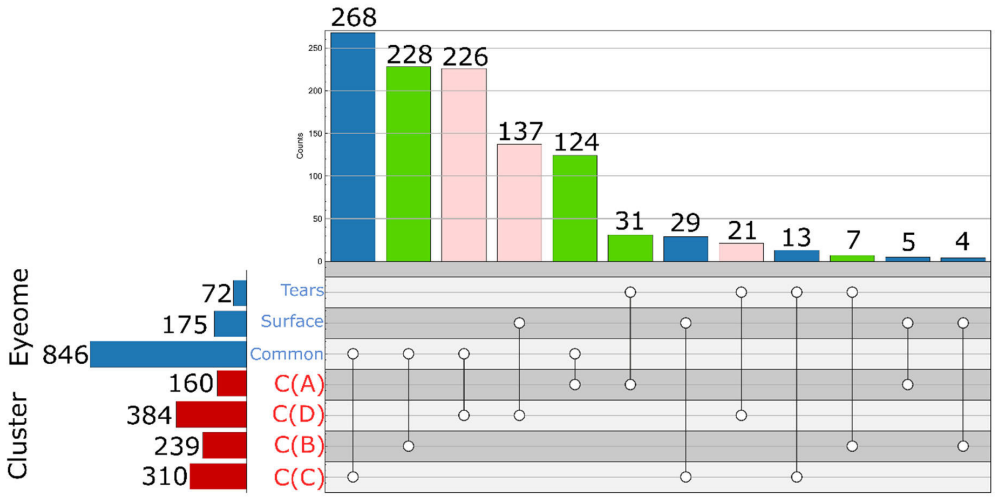


Figure 5. Count plot representation of the intersections of protein clusters and sections of the EyeOME [13] database. Clusters are those identified in Section 2.4. Proteins in the 'Cornea' or 'Sclera' section of EyeOME were classified as Surface.

2.6. Intra- and Interpersonal Variances of Different Tear Samples

The differences in the composition of the tear samples collected from different persons, or from the two eyes of the same subject, may be a combination of several simultaneous reasons, including the effect of the sampling procedure on eye surface and tear secretion. In order to identify sampling induced effects, the protein composition of tear samples of the same person was compared using different methods, excluding the highly contaminated SL samples. Similar distributions of absolute protein abundance differences were observed in all sampling methods, however the Pearson correlation of intensities showed marked differences in the three tear fluid samples. The strongest intrapersonal correlation was found in the CAP samples (average of coefficients was 0.91, median 0.92), while PRT samples (average of 0.77, median 0.81), and SU samples (average of 0.73, median 0.76) presented weaker correlations. (Figure 6A). The ratio of the protein MS intensities measured in the two eyes relative to the average of the eyes was also calculated to represent the differences between the eyes. The log2 transformed distribution is the narrowest around 0 in the case of CAP samples, and widest is in Schirmer's samples (Figure 6B). Eleven percent of data points has an absolute value higher than 1 (at least twofold difference relative to eye average) in the case of CAP samples, in Schirmer's samples 21% and in PRT samples 14% has that high difference.

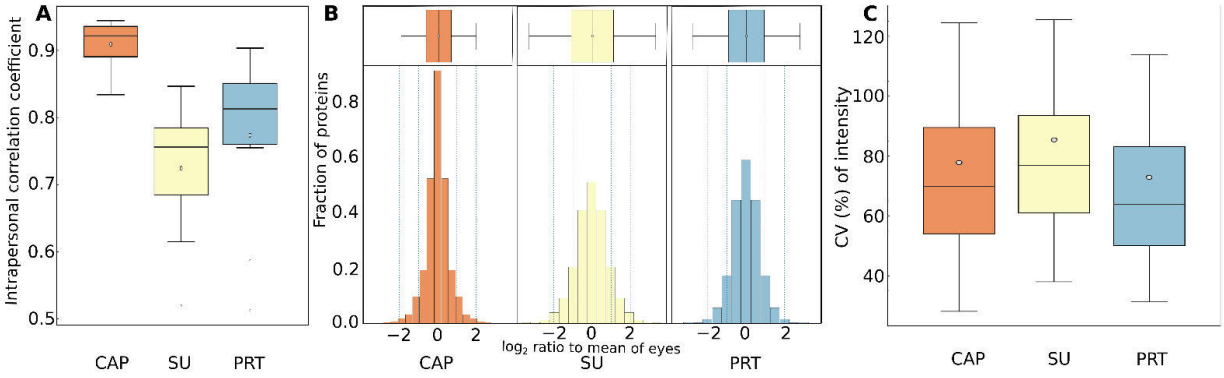


Figure 6. (**A**). Intrapersonal (left-to-right eye) Pearson correlation coefficients of samples of different sampling methods. (**B**). Distribution of protein log2 transformed intensity ratios measured in the two eyes relative to the average of the two eyes. (**C**). Distribution of coefficients of variation (CV%) for intensities of proteins in Cluster A in all samples of different sampling methods.

We calculated the overall variance (CV) of protein intensities in the whole sets of sample types. The CV distribution of the common 195 proteins of all sample types (Cluster A) are shown in Figure 6C. According to this, these proteins have a lower median CV in PRT samples (64%) than in either SU (77%) or CAP samples (70%).

2.7. Evaluation of the Proteomics Rotocol

The nanoLC-MS reproducibility was determined by triplicate injection of a randomly selected PRT sample. The number of identifications and the reproducibility of the relative intensities were compared at both protein and precursor levels; 3003 ± 21 precursors were quantified with the median CV of 0.17 in these measurements, corresponding to 548 ± 4 proteins with the median CV of 0.15. Based on these results, our procedure was suitable for the study of sample preparations and sampling procedures.

In order to evaluate the efficiency and reproducibility of protein extraction from PRT samples, three different extraction solutions (5% SDS, 100 mM ABC, and 1% acetic acid in water) were tested. Nine PRTs were soaked in a pooled capillary tear sample to ensure the same initial protein concentration and composition. Note that this experiment was limited by the number of proteins detectable from capillary samples, but comparison to the original pooled capillary sample provides satisfactory data on recovery and reproducibility for major tear proteins. There were no significant differences observed in the total amount of extracted proteins as determined by BCA assay (Table 1); 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 ABC protocols, but the acetic acid protocol performed worst regarding all these measures.

Table 1. Comparison of tear protein extraction methods from PRT.

	Total Amount of Proteins (µg)	Number of Quantified Proteins	Percentage of the Total Number of Proteins	Median CV of Intensity	Pearson Correlation to Capillary (r)
Capillary	-	236	100.0	0.17	-
SDS	41.5 ± 3.3	198	83.9	0.19	0.941
ABC	36.4 ± 1.16	203	86.0	0.15	0.946
Acetic acid	40.9 ± 6.78	152	64.4	0.23	0.851

Using either the SDS or ABC protocol, 84–86% of proteins can be recovered, and Pearson correlation with the original sample demonstrated, that the protein composition was not biased, the extract correctly demonstrates the original composition. Although ABC extraction seems to be suitable for protein extraction from PRT samples, the presence of a tenside (e.g., SDS) may help the extraction of the intracellular fraction that was excluded by our evaluation approach. Because of that, and for comparability with the protocol applied for Schirmer's strips, we used the SDS approach in our analysis.

3. Discussion

The lacrimal apparatus supplies tear fluid to the surface of the eyeball and the eyelid, minimizing friction and cleaning the eye. Tear fluid and its components are actively secreted by various secretory units, including lacrimal glands, meibomian glands, accessory lacrimal glands, sebaceous glands of Zeis and Moll, and corneal and conjunctival cells. Additionally, the highly vascularized conjunctival tissue contributes with a large number of blood-derived compounds to the composition of tear fluid [14].

The main and accessory lacrimal glands secrete tear fluid into the lateral aspect of the superior conjunctival fornix, and the upper fornix. Then the fluid spread across the entire surface of the eye covering the anterior eyeball with a thin film, called preocular/precorneal tear film. The tear fluid flows from lateral to medial toward the tear drainage system at the inner canthus [15]. Tear production normally is about 0.5–2.2 µL/min, considering the total volume of 6–7 µL it causes 16% turnover per minute [4,16]. In addition to production,

evaporation, absorption, and drainage are responsible for dynamic balance of the preocular tear film.

The preocular tear film is a thin fluid layer (2–6 μm) covering the ocular surface; it is the interface of the ocular surface with the environment. The precorneal tear film is now regarded as a complex blended two-layer structure comprising of a mucoaqueous gel layer lying beneath, but at least partly integrated with an overlying lipid layer [17].

Tears have been classified into three main types: basal, reflex, and psycho-emotional [18]. Basal tears, also known as non-stimulated tears, continuously coat the eye to keep it moist and protected. Reflex tear (or stimulated) is a higher lacrimal flow produced in response to external physical or chemical stimuli. Psycho-emotional tears are produced in response to joy, sadness, fear, and other emotional states. Each type of tears has many variants, and their differences are not always clear, as they sometimes overlap or are imprecise or controversial. It is more accurate to think of tear output as a continuum, whereby the rate of production is proportional to the degree of sensory or emotive stimulation [14]. The aqueous part of all types of tears are produced by the lacrimal glands and accessory lacrimal glands, but differ in their volume and composition [19].

As can be seen from the above, tear fluid is a dynamic mixture of substances with different origins and may have different composition depending on external physical or emotional factors. In addition, to the biological variance in tear composition, samples collected from the tear film may vary depending on the sampling position on the eye surface, depth of sampling of different fluid layers, changes induced by the sampling process (physical or emotional stimuli) and recovery from the sampling device.

Protein concentrations in normal tear fluid range from 6 to 11 mg/mL [5], but the protein concentration and composition of tear fluid samples are greatly influenced by the following factors [7,20]: (i) tear collection device (glass capillary tubes, Schirmer's strips, threads, ophthalmic sponges, and polyester rods; (ii) types of collected tears (reflex, or non-stimulated tears), which might be affected by the tear sampling procedure, irritating stimuli like environmental fluctuations, physiological status or rubbing the skin with alcohol, anesthesia; (iii) location of tear sampling (the inferior temporal tear meniscus near the external canthus of the eyes, or the inferior conjunctival sac); (iv) whether sampling occurred from open or closed eye; (v) last, but not least, the sample processing (recovery from the sampling device) and the further analytical procedure and data analysis. As can be seen from the above, tear sampling method is definitely a major challenge and has the greatest significant influence on the precision and reproducibility of the analytical results.

Several tear sampling methods are available, and every sample collection method used must be assessed since it has a significant influence on the precision and reproducibility of the analytical results. Each sampling method has advantages and disadvantages; therefore, 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, threads, ophthalmic sponges, and polyester rods. For determination of the protein profile of the tear sample the capillary tube and the Schirmer's strip are used most frequently [7,20].

As both methods are routinely used for measuring tear volume in ophthalmology, and in many tears proteomics studies, much information has been gathered on their application. A comparison of the capillary, Schirmer's strip and the new PRT tear collection methods is provided in Table 2.

Collecting tear fluid by capillary tubes is generally found to be more convenient for patients, but because of fear of injury, some find it less desirable than the Schirmer's strip. It can be time consuming, sometimes it can take up to 10 min to collect enough tears, because tears flow slowly and erratically, with interruptions due to blinking and eye movement [21]. Therefore, it less suitable for collecting samples from aqueous tear-deficient patients or animal models. To collect tears, capillary tube is usually placed close to the inferior temporal tear meniscus near the external canthus without touching the cornea,

conjunctiva, and lower eyelid. If the sampling is performed by a specialist who has practice and experience in this collection method, then it does not induce reflex tearing, nor does it involve a potential risk of injury. However, the investigator has to hold the capillary tube for the duration of the sampling procedure, which entails constant and prolonged work on the open eye. Tear collection with capillary frequently requires previous stimulation or instillation of different volumes of saline into the cul-de-sac and collecting after sufficient mixing [7]. Capillary samples contains a higher percentage of proteins originating from extracellular region, protein containing complexes, and membrane [4,10,22]. In biological processes, immune response, complement pathway, and tissue development proteins dominate more frequently in capillary samples.

Schirmer's test is well-established in clinical ophthalmic practice to measure tear secretion [23,24]. Tear collection is performed using a special filter paper strip (5 mm wide and 35 mm long) with the bent end placed between the palpebral conjunctiva of the lower eyelid and the bulbar conjunctiva of the eye. The eye is then closed for 5 min while the tear fluid absorbs into the filter paper. The test can be performed with or without the use of anesthetics. Although Schirmer's strips have been considered a convenient and easy to perform method of tear collection, their use can cause strong irritation, leading to reflex tearing that results in larger but more diluted samples. The strip is in contact with the highly vascularized conjunctiva and can injure its surface and microvasculature. This damage and induction of the secretion of reflex tears by mechanical irritation likely distort the protein-profile of tear samples. Due to the trauma, these kinds of samples contain proteins not only from tears but also from surrounding tissues and blood [4,25]. An increased number of cell and organelle-specific (intracellular) proteins has been reported that contaminate tear samples [10].

Table 2. Characteristics of tear collection methods for proteomics analysis.

Parameter	Glass Capillary (CAP)	Schirmer's Strip (SU)	PRT
Average protein amount recovered *	60 µg	58 µg	29 µg
Number of proteins identified *	422	1225	1439
Average intracellular "contaminant" protein MS intensity *	1.2%	1.5%	0.7%
Median CV of major proteins *	70%	77%	64%
Tear volume collected	3–10 µL [26]	~9 µL/cm [27]	~3–4 µL [21]
Tear collection time	Up to 5 min [20]	5 min [28]	15–20 s [28]
Contact with eye	No/minimal [22,29]	Strong contact with cornea, conjunctiva and lower eye-lid [7,29,30]	Mild contact with cornea, conjunctiva and lower eye-lid
Site of tear sampling	Variable [7,20]	Variable [7,20]	Standard [9,20,28]
Diagnostic repeatability (volume)		Poor ** [28,31]	Good [32]
Open/closed eye	Open	Open/Closed	Open/Closed [33]
Invasiveness	Low [34]	High [34]	Low/Medium
Discomfort (subjective)	No [23]	Yes [28]	No [28,32]
Sensitivity in detecting dry eye disease	Low	Low [28]	High [32]
Induces reflex tearing with higher flow rate	No [7]	Yes [28,34]	Minimally [28]
Sample processing	-	Cutting and protein extraction or centrifugation [2,35]	Protein extraction *
Sampling device material/source	Variable	Variable [36]	Uniform [20]
Risk of injury	Yes	No [23]	No
Cellular/plasma contamination	Low [10,22]	High [4,10,22,29,30]	Low *
Tear collection requires specialist	Yes [23]	No	No
Continuous intervention during sample collection	Yes [23]	No [23]	No

* Data from current work, ** Can be improved with anesthesia (Type II test) [26].

Collection of tears with Schirmer's strip is an indirect method carried out using absorbing supports. Proteomics analysis of these kinds of samples requires either centrifugation of tear fluid or extraction of proteins (before digestion) or peptides (after in-strip digestion) from the paper strips. A disadvantage of this absorbent-based sampling method is that different extraction procedures may result in varying protein profiles. It was recently reported [36] that the elution of proteins from Schirmer's strips varies significantly between different brands of the filter paper because of their distinct absorptive properties. Apart from the variety of clinical procedures used to perform Schirmer's test, it seems likely that one of the causes of the variability of Schirmer's test between studies is related to the use of different Schirmer's test strips. Unfortunately, there is no standardization of commercial strips, even though the need for standardization was recognized over fifty years ago. The maximum volume of absorption on 10 mm long piece of Schirmer's strip was found to be around 9 µL [27], but it may also depend on strip material.

Phenol red thread, like the Schirmer's test, is a widespread clinical test for measuring tear volume [28,32]. In the detection of dry eye, PRT is equally sensitive, but it has many advantages over Schirmer's test. Sampling can be performed in a significantly shorter time (15 s vs. 5 min), which is much more convenient for patients. It can also be performed on children [11] and it has a significantly smaller contact area causing minimal irritation and minimal induction of reflex tearing [37]. It has the advantage that it is not only suitable for human tear sampling but has been shown to better evaluate tear secretion in small animals with lower tear volume such as birds [38] and rodents [39]. Phenol red thread collects around 3 µL of tear fluid [21], and this volume is low compared to an average basal tear volume. The inability of PRT to absorb the entire volume of basal tears means that the collected initial sample is almost pure basal tears; however, it can be used for the collection of reflex tears if applied after stimulus [21], which can be a big advantage over other tear collection methods.

In our experiments, both absorbent-based approaches (SU and PRT samples) allowed the collection of low-volume samples, and the amounts of proteins were sufficient to perform several proteomic analyses even from those two donors whose CAP tear samples were not sufficient. The safe application of capillary approach without touching the eye surface, requires larger tear volume with a thick tear film. This makes it less suitable for the collection of tear samples from aqueous tear-deficient patients [22,29], but there is no such limitation with the indirect methods, where the soft sampling material is immersed in the tear film.

The observable proteome with the two most frequently used tear fluid sampling methods have been compared several times; most of these studies conclude that although both methods can be used, the capillary and Schirmer's strip tear collection methods still result in different protein compositions [7,10,22,23]. It was assumed that this is because Schirmer's strip results in an increased tear production due to possible irritation and contains proteins not only from tear fluid but also from tissues via direct eye contact [34].

Our findings are consistent with these results, as more proteins were identified in the samples collected with the Schirmer's strips compared to the capillary samples. The SL samples contain an even larger number of proteins than SU samples being in direct contact with the ocular surface. The number of identified proteins in the novel PRT samples was similarly high (1439), 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.

Ma et al. recently summarized [6] proteomics datasets from tear films using either capillary or Schirmer's strips by different research groups. They collected 1892 proteins from 11 publications and found 435 proteins common to capillary and Schirmer's strip samples. Based on gene names, we matched those proteins to our dataset, to evaluate the overlap of our data with those recent results. Of all those proteins, 1656 were identified in our experiments, most of them in SL samples (1623), while 1238, 1089, and 382 proteins

were found in the PRT, SU, and CAP samples, respectively. Regarding the 435 proteins designated as common by Ma and coworkers [6], we could identify 432 in total, 426, 406, 392, and 192 in SL, PRT, SU, and CAP samples, respectively. These results highlight the comparable proteomics usability of PRT not just with SU samples processed and analyzed in our laboratory, but with numerous other methods.

Akkurt Arslan et al. [40] processed their Schirmer's samples in a similar manner and found that 1153 proteins were identifiable in their SL samples, while 1107 proteins were identified in their SU samples. The significantly larger number of proteins identified in our SL samples may be explained by the MS method (DIA in our case vs. DDA) [6] and by the larger number of samples (20 SL samples from 10 individuals in our study vs. four SL samples from two individuals). Contamination occurs randomly, thus increasing the number of samples increases the chance of identification. It must be noted that only pooled samples were analyzed in that study and no quantitative comparisons were made, so no information is available on identification repeatability and quantitative variability.

The identification of such a large number of proteins in our experiments was made possible by the application of the GPF-DIA LC-MS method. By the application of a combined spectral library, it was possible to quantify a higher number of useful proteins present in the tear fluid under normal conditions in a single nanoLC-MS run. A similar result was observed by Nättinen et al. [10], that almost twice as much protein could be quantified in tear samples collected by capillary using a combined spectral library instead of capillary type-specific one. They also did not find increased number of proteins in Schirmer's type samples searched in a combined library. Green-Church et al. [22] have previously demonstrated that although the proteins detected in capillary samples were mainly extracellular, tear samples collected by Schirmer's strip contained a large number of additional cellular proteins. Using a combined spectrum library, those proteins that could not be detected with the sample-specific libraries became easier to identify. Thus, it can be used as a quality control to identify tear samples that contain higher levels of contaminating proteins.

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 however 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 is very interesting because the entire thread was processed; the part in direct contact with the surface of the eyeball and the eyelids was not removed, unlike in the case of SU samples. This may be a consequence of the smaller diameter and the smaller and smoother surface of PRT fibers compared with Schirmer's paper strips.

According to the Gene Ontology analysis of protein clusters, we can conclude that intracellular proteins originating from the eye surface and lower layers of the tear film are increasing the size of the proteome sampled by the indirect methods compared to capillary. These proteins are most effectively collected on the surface of the lower part of the Schirmer's strip which is in direct contact with the surface of the eyeball and eyelids. Therefore, we conclude that those proteins may be designated as contaminants, if study of tear fluid is the goal, but may provide diagnostic information on eye surface, if specific sample collection methods can be applied for their reproducible analysis.

We can compare our clustering with the dataset of Ma et al. [6] described above, in which they found 435 proteins common in the literature data to Schirmer's strip and capillary samples. We could quantify 387 proteins of those, 68% of which can be found in Cluster A and B, while only 11% are in Cluster D. This shows high similarity between the classifications, despite the different approaches. They concluded that those common proteins are present at high abundance regardless of sample collection, which we also demonstrated (Figure 4). Our observation, that the rarely identified, low intensity proteins

can be found with high frequency and at higher intensity in the SL samples, confirms the ocular surface origin of those proteins.

In order to validate the application of PRT in tear biomarker analysis, we collected tear biomarkers from recent reviews of literature data [6,41–45]. 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.

Sample processing is one of the most crucial processes in proteomics research; this pre-analytical step can bias both protein composition and quantification. Absorbent-based tear collection requires a pre-analytical step to elute/extract proteins from a paper strip, sponge, thread, etc. Several extraction conditions were proposed, by varying extraction solvent, volume, agitation time, and temperature to recover proteins from Schirmer's strips [4,35] and small bioactive molecules from PRT [21].

Tear proteins absorb on PRT by varying strengths of intermolecular interactions with thread cellulose. The small surface and volume of PRT compared to a Schirmer's strip results in less interactions between proteins and thread. This can make it easier to elute/extract proteins from PRT. Our results supporting this theory because using either the SDS or ABC protocol, 84–86% of proteins can be recovered, and protein composition was not biased, the extract correctly demonstrated the original composition with excellent reproducibility. Finally, phenol red thread is produced only by one company (Showa Yakuhin Kako Co., Ltd., Tokyo, Japan), so using the same tear collection device in different clinical and laboratory studies would reduce potential variables in tear analysis.

We have also evaluated the effect of sampling on variance of results. 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 (median of 64% for the common the proteins), considerably lower than in the SU samples (median of 77%).

These lower intra- and interpersonal variances may be attributed to the lower induction of reflex tear formation during sampling compared to Schirmer's strip [28,37]. At the same time, the lower volume quickly collected by PRT ensures the collection of reproducible pure basal tear [21], thus making it more suitable for comparative analysis.

4. Materials and Methods

4.1. Materials/Reagents

Calibrated glass microcapillary tubes (20 mL) were manufactured by Drummond Scientific Company (Broomall, PA, USA), Schirmer's strips (I-Dew Tearstrips) by Entod Research Cell UK Ltd. (London, UK) and PRT (Zone-Quick test) by Showa Yakuhin Kako Co., Ltd. (Tokyo, Japan).

Reagents, such as ammonium bicarbonate (ABC), dithiothreitol (DTT), iodoacetamide (IAA), and sodium dodecyl sulfate (SDS) were purchased from Sigma-Aldrich (Darmstadt, Germany), acetone from Merck (Darmstadt, Germany), trypsin, and formic acid (FA) from Thermo Scientific (Rockford, IL, USA). Water, acetonitrile, and acetic acid were delivered by VWR (Debrecen, Hungary).

4.2. Samples

Five healthy female and five healthy male volunteers participated in the experiment, aged between 20 and 33 years. All subjects included in this study were enrolled at the Department of Ophthalmology, University of Szeged. Approval for the human study was granted by the local Ethical Committee of the University of Szeged (108/2019-SZTE), and the study protocol adhered to the tenets of the most recent revision of the Declaration of Helsinki for experiments involving human subjects. All subjects enrolled in this study provided voluntary written informed consent.

For each subject, tear fluid samples were collected from both right and left eyes using three different sampling techniques during the same morning visit. More specifically, tear samples were collected (in the order of sampling) using glass capillaries (CAP), PRT, and Schirmer's strips. For CAP sampling glass microcapillary tubes were introduced into the ventral cul-de-sac of the conjunctiva of the opened eyes for an average of 2 min without contact to the cornea, conjunctiva, and lower eyelid. PRTs and Schirmer's strips were placed over the lid margin at the junction of the lateral and middle thirds of the lower eyelids. The sampling lasted 15 s for PRTs and 5 min for Schirmer's strips while subjects closed their eyes without an anesthetic. In order to obtain tear samples from the resting eyes, there was a 30-min break between the different sample collection methods. Samples were immediately placed into sterile plastic microtubes and frozen at −80 °C until the analytical procedure. During sample preparation, the Schirmer's strips were divided into lower (SL) and upper (SU) parts by cutting strips at the zero line and at the sign of 10 mm, respectively. The SL represent the rounded portions of the paper that were in direct contact with the eye surface, while the SU samples are the subsequent 10 mm sections. The rest of the strips were not processed.

4.3. Protein Extraction

All the SL, SU, and PRT samples were extracted individually by adding 100 µL 5% SDS in ABC (pH 8.0, 100 mM) containing 0.25% protease inhibitor cocktail (Sigma-Aldrich, Rockford, IL, USA). The samples were vortexed and incubated in a thermal shaker at room temperature for 1 h at 350 RPM. After a centrifugation at 12,000× g, 10 min, 4 °C, the supernatants of the samples were transferred into new tubes.

The capillary tear samples were diluted using ABC (pH 8.0, 100 mM) containing 5% SDS and 0.25% protease inhibitor cocktail before determining the protein concentration.

4.4. Evaluation of Protein Extraction Methods from PRT

To examine the protein extraction efficiency from PRTs, different extraction solvents were tested. A pooled tear sample was collected in a low protein binding tube 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. The protein contents of the threads were extracted using 100 µL of (1) 5% SDS in ABC (pH 8.0, 100 mM), (2) ABC (pH 8.0, 100 mM), (3) 1% acetic acid in water. All extraction liquids contained 0.25% protease inhibitor cocktail and were tested on 3 replicates. The samples were incubated in a thermal shaker at room temperature for 1 h at 350 RPM. After a centrifugation at 12,000× g, 10 min, 4 °C, the supernatants of the samples were transferred into new tubes.

4.5. Sample Preparation

The protein contents of all samples were determined using BCA Protein Assay (Thermo Scientific, Rockford, IL, USA) according to the manufacturer's protocol. Samples containing 10 µg protein were processed using an on-pellet digestion protocol [46] with slight modifications. Briefly, the samples were reduced with 10 mM DTT at 60 °C for 30 min and alkylated with 20 mM IAA in dark at room temperature for 30 min. The protein content was precipitated by adding a 7-fold volume of ice-cold acetone and incubated at −20 °C overnight. After centrifugation with 15,000× g, 10 min, 4 °C the supernatant was discarded. The protein pellet was washed twice with 500 µL acetone/water (85:15, v/v) mixture. After centrifugation with 14,000× g, 10 min, 4 °C, the protein pellet was dissolved in 15 µL RapiGest SF Surfactant (Waters, Milford, MA, USA) and was incubated at 100 °C for 5 min. After being cooled to room temperature, 65 µL ABC (pH 8.0, 100 mM) and 0.25 µg trypsin in 5 µL ABC (pH 8.0, 100 mM) were added to the mixtures. The samples were incubated at 37 °C for 30 min and another 0.25 µg trypsin in 5 µL ABC (pH 8.0, 100 mM) was added and the mixture was digested at 37 °C for 5.5 h. Digestion was stopped by the addition of 1 µL

concentrated formic acid. The samples were centrifuged with 12,000× g, 10 min, 4 °C, and 2 µL of the supernatant was injected to the nanoLC-MS system.

4.6. NanoLC-MS Measurements

NanoLC-MS/MS analysis was carried out on a Waters ACQUITY UPLC M-Class LC system (Waters, Milford, MA, USA) coupled with a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Symmetry® C18 (100 Å, 5 µm, 180 µm × 20 mm) trap column was used for trapping and desalting the samples. Chromatographic separation of peptides was accomplished on an ACQUITY UPLC® M-Class Peptide BEH C18 analytical column (130 Å, 1.7 µm, 75 µm × 250 mm) at 45 °C by gradient elution (linear gradient from 3% solvent B to 40% solvent B in 70 min, followed by a 30 min washing and equilibrating gradient). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid were used as mobile phases at a flow rate of 350 nL/min. The sample temperature was maintained at 5 °C. The mass spectrometer was operated using the equipped Nanospray Flex Ion Source.

Data acquisition was performed using XcaliburTM 4.1 (Thermo Fisher Scientific, Waltham, MA, USA).

4.7. DIA Acquisition and Data Processing

Data for creating sample-specific (CAP, SL, SU, PRT) spectrum libraries were collected from LC-MS analysis of four sample-group-specific pooled samples. The mass spectrometer was configured to acquire six gas-phase fractionated (DIA-GPF) acquisitions with 4 m/z precursor isolation windows at 17,500 resolution to hit an AGC target 1×10^6 with a maximum inject time of 120 ms. In the GPF-DIA measurements, an overlapping window pattern was adjusted from the previously optimized mass ranges (395–505, 495–605, 595–705, 695–805, 795–905, and 895–1005 m/z).

For the quantitative analysis of individual samples, a single LC-MS run with DIA acquisitions (DIA-Q) from 395 to 1020 m/z was used with 27 × 22 m/z overlapping precursor isolation windows at 17,500 resolution to hit an AGC target 2×10^5 and the maximum inject time was set to auto.

DIA-NN 1.8 software [47] was applied for database search. First, a predicted spectral library was generated from the Uniprot Human Reference FASTA proteome (20,575 genes, one protein per gene). 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 four groups of GPF measurements in one against the same predicted spectral library. All the individual samples were searched against both the experimental sample-specific and combined spectral libraries to detect all peptides and proteins that can be quantified. DIA-NN was configured with the default settings with the following modifications: the maximum number of missed cleavages and variable modifications were set to 2, methionine oxidation was added, and precursor m/z range was adjusted to 380–1020. Search results were filtered for 1% FDR at the level of unique genes. Further data processing, statistical analysis and creation of figures were carried out in Perseus (version 1.6.15.0) [48] and Instantclue (version 0.1.1) [49] software. In all quantitative statistical analysis, only proteins identified with at least two unique peptides were included. Because of the large biological variance among samples, for better comparability, not the raw intensities of individual proteins but their relative proportion in the samples were used. Therefore, intensities of the proteins were divided by the sum of intensities of all proteins as normalization and log2 transformation was performed on all datasets.

The analyzed proteins were annotated from Gene Ontology [50] and UniProt (https://www.uniprot.org) (accessed on 12 April 2022)

5. Conclusions

A crucial starting point for biomarker discovery in the tear proteome is the in-depth investigation of tear proteins in healthy subjects. Thanks to recent methodological advances, MS-based tear proteomics can guarantee the high accuracy and sensitivity necessary for biomarker discovery and single-tear analysis. To determine how the tear fluid sampling method affects the protein profiles obtained, tear samples from healthy persons, collected by using glass capillary, Schirmer's strip and phenol red threads, were analyzed and statistically evaluated. Compared to the two widely used tear collection methods, the PRT was proven to be a fast, reproducible, and reliable sampling procedure, providing tears from patients whose capillary samples were not enough for proteomics analysis. The tear sampling method substantially impacted the proteomic profiles of individual samples determined by nanoLC-MS/MS. The number of proteins considered to be "tear proteins" was higher in the Schirmer's strip and PRT; however, samples collected by PRT contain even lower amounts of "contaminant" proteins than the eye surface. According to our experience, PRT can be used successfully for proteomic analysis not only of human but also of mouse tears (data not shown).

Unique properties of PRT allow fast, non-invasive collection of small amounts of tears and gives almost pure basal tear samples with unbiased protein composition. These characteristics make PRT sampling methods ideal for proteomics analysis of tear fluid and the discovery of protein biomarkers of ocular and systemic diseases, therefore in such studies it could replace current methods of sample collection.

The mass spectrometric data acquisition and evaluation method also influence the proteomics results. Creating and using a combined instead of sample-specific spectrum libraries resulted in a higher number of quantified proteins in each sample. Application of this method not only increased the number of useful proteins but also helped to identify and quantify different clusters of proteins, based on detectability using different sampling techniques. This may help future studies to focus on the most informative sub-proteomes of the human tears and eye surface.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23158647/s1.

Author Contributions: Conceptualization, G.K., T.J., Z.S. and E.T.-M.; methodology, G.K., T.J., Z.S. and E.T.-M.; validation, G.K., T.J., Z.S. and E.T.-M.; formal analysis, G.K. and Z.S.; investigation, G.K., T.J., Z.S. and E.T.-M.; data curation, G.K., T.J. and Z.S.; writing—original draft preparation, G.K., T.J. and Z.S.; writing—review and editing, G.K., T.J., Z.S. and E.T.-M.; visualization, G.K. and Z.S.; supervision, T.J. and Z.S.; project administration, T.J. and Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU and the Hungarian Government, grant number EFOP-3.6.1-16-2016-00008;and by the Albert Szent-Györgyi Medical School, University of Szeged, grant number SZTE ÁOK-KKA No 2018/Tóth-MolnárE.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethical Committee of the University of Szeged (108/2019-SZTE, 5 December 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The combined spectral library and quantitative results from DIA-NN analysis are available through Zenodo at: https://doi.org/10.5281/zenodo.6757957 (accessed on 24 June 2022).

Acknowledgments: This study was supported by Scholarship for Young Talents of the Nation from the Hungarian Prime Minister's Office.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ABC	ammonium bicarbonate
AGC	Automatic gain control
CAP	capillary
CV	coefficient of variation
DDA	data dependent acquisition
DIA	data independent acquisition
DTT	dithiothreitol
EyeOME	The Human Eye Proteome Project
FA	formic acid
FDR	false discovery rate
HCD	Higher-energy collisional dissociation
IAA	iodoacetamide
LC/MS	Liquid chromatography–mass spectrometry
m/z	mass/charge
M	mol/L
MS/MS	tandem mass spectrometry
PRT	phenol red thread
SDS	sodium dodecyl sulfate
SL	Schirmer's strip lower part
SU	Schirmer's strip upper part

References

1. Zhan, X.; Li, J.; Guo, Y.; Golubnitschaja, O. Mass Spectrometry Analysis of Human Tear Fluid Biomarkers Specific for Ocular and Systemic Diseases in the Context of 3P Medicine. *EPMA J.* **2021**, *12*, 449–475. [CrossRef] [PubMed]
2. Ponzini, E.; Santambrogio, C.; De Palma, A.; Mauri, P.; Tavazzi, S.; Grandori, R. Mass Spectrometry-based Tear Proteomics for Noninvasive Biomarker Discovery. *Mass Spectrom. Rev.* **2021**, 1–19. [CrossRef] [PubMed]
3. Chen, L.; Zhou, L.; Chan, E.C.Y.; Neo, J.; Beuerman, R.W. Characterization of the Human Tear Metabolome by LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 4876–4882. [CrossRef] [PubMed]
4. Zhou, L.; Beuerman, R.W. Tear Analysis in Ocular Surface Diseases. *Prog. Retin. Eye Res.* **2012**, *31*, 527–550. [CrossRef] [PubMed]
5. Zhou, L.; Beuerman, R.W. The Power of Tears: How Tear Proteomics Research Could Revolutionize the Clinic. *Expert Rev. Proteom.* **2017**, *14*, 189–191. [CrossRef] [PubMed]
6. Ma, J.Y.W.; Sze, Y.H.; Bian, J.F.; Lam, T.C. Critical Role of Mass Spectrometry Proteomics in Tear Biomarker Discovery for Multifactorial Ocular Diseases (Review). *Int. J. Mol. Med.* **2021**, *47*, 83. [CrossRef] [PubMed]
7. Rentka, A.; Koroskenyi, K.; Harsfalvi, J.; Szekanecz, Z.; Szucs, G.; Szodoray, P.; Kemeny-Beke, A. Evaluation of Commonly Used Tear Sampling Methods and Their Relevance in Subsequent Biochemical Analysis. *Ann. Clin. Biochem.* **2017**, *54*, 521–529. [CrossRef]
8. Kurihashi, K.; Yanagihara, N.; Honda, Y. A Modified Schirmer Test: The Fine-Thread Method for Measuring Lacrimation. *J. Pediatr. Ophthalmol.* **1977**, *14*, 390–397. [CrossRef]
9. Barmada, A.; Shippy, S.A. Quantifying Sample Collection and Processing Impacts on Fiber-Based Tear Fluid Chemical Analysis. *Transl. Vis. Sci. Technol.* **2020**, *9*, 23. [CrossRef]
10. Nättinen, J.; Aapola, U.; Jylhä, A.; Vaajanen, A.; Uusitalo, H. Comparison of Capillary and Schirmer Strip Tear Fluid Sampling Methods Using Swath-Ms Proteomics Approach. *Transl. Vis. Sci. Technol.* **2020**, *9*, 16. [CrossRef]
11. Vashisht, S.; Singh, S. Evaluation of Phenol Red Thread Test versus Schirmer Test in Dry Eyes: A Comparative Study. *Int. J. Appl. Basic Med. Res.* **2011**, *1*, 40–42. [CrossRef]
12. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *336*, 012017. [CrossRef]
13. Ahmad, M.T.; Zhang, P.; Dufresne, C.; Ferrucci, L.; Semba, R.D. The Human Eye Proteome Project: Updates on an Emerging Proteome. *Proteomics* **2018**, *18*, 1700394. [CrossRef]
14. Lawrenson, J.G. Anterior Eye. In *Contact Lens Practice*; Nathan Efron, Ed.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 10–27.e2.
15. Ocran, E. Lacrimal Gland. Available online: https://www.kenhub.com/en/library/anatomy/lacrimal-gland (accessed on 25 June 2022).
16. Mishima, S.; Gasset, A.; Klyce, S.D.; Baum, J.L. Determination of Tear Volume and Tear Flow. *Investig. Ophthalmol.* **1966**, *5*, 264–276.

17. Kopacz, D.; Niezgoda, Ł.; Fudalej, E.; Nowak, A.; Maciejewicz, P. Tear Film—Physiology and Disturbances in Various Diseases and Disorders. In *Ocular Surface Diseases—Some Current Date on Tear Film Problem and Keratoconic Diagnosis*; Dorota Kopacz, Ed.; IntechOpen: London, UK, 2020.
18. Murube, J. Basal, Reflex, and Psycho-Emotional Tears. *Ocul. Surf.* **2009**, *7*, 60–66. [CrossRef]
19. Willcox, M.D.P.; Argüeso, P.; Georgiev, G.A.; Holopainen, J.M.; Laurie, G.W.; Millar, T.J.; Papas, E.B.; Rolland, J.P.; Schmidt, T.A.; Stahl, U.; et al. TFOS DEWS II Tear Film Report. *Ocul. Surf.* **2017**, *15*, 366–403. [CrossRef]
20. Pieczyński, J.; Szulc, U.; Harazna, J.; Szulc, A.; Kiewisz, J. Tear Fluid Collection Methods: Review of Current Techniques. *Eur. J. Ophthalmol.* **2021**, *31*, 2245–2251. [CrossRef]
21. Avilov, V.; Zeng, Q.; Shippy, S.A. Threads for Tear Film Collection and Support in Quantitative Amino Acid Analysis. *Anal. Bioanal. Chem.* **2016**, *408*, 5309–5317. [CrossRef]
22. Green-Church, K.B.; Nichols, K.K.; Kleinholz, N.M.; Zhang, L.; Nichols, J.J. Investigation of the Human Tear Film Proteome Using Multiple Proteomic Approaches. *Mol. Vis.* **2008**, *14*, 456–470.
23. Posa, A.; Bräuer, L.; Schicht, M.; Garreis, F.; Beileke, S.; Paulsen, F. Schirmer Strip vs. Capillary Tube Method: Non-Invasive Methods of Obtaining Proteins from Tear Fluid. *Ann. Anat.* **2013**, *195*, 137–142. [CrossRef]
24. Vandermeid, K.R.; Su, S.P.; Ward, K.W.; Zhang, J.Z. Correlation of Tear Inflammatory Cytokines and Matrix Metalloproteinases with Four Dry Eye Diagnostic Tests. *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 1512–1518. [CrossRef]
25. Van Haeringen, N.J.; Glasius, E. The Origin of Some Enzymes in Tear Fluid, Determined by Comparative Investigation with Two Collection Methods. *Exp. Eye Res.* **1976**, *22*, 267–272. [CrossRef]
26. You, J.; Fitzgerald, A.; Cozzi, P.J.; Zhao, Z.; Graham, P.; Russell, P.J.; Walsh, B.J.; Willcox, M.; Zhong, L.; Wasinger, V.; et al. Post-Translation Modification of Proteins in Tears. *Electrophoresis* **2010**, *31*, 1853–1861. [CrossRef]
27. Bertram, M.; Allbaugh, R.A.; Mochel, J.P.; Peraza, J.; Page, L.; Sebbag, L. Influence of Schirmer Strip Wetness on Volume Absorbed, Volume Recovered, and Total Protein Content in Canine Tears. *Vet. Ophthalmol.* **2021**, *24*, 425–428. [CrossRef]
28. Saleh, T.A.; McDermott, B.; Bates, A.K.; Ewings, P. Phenol Red Thread Test vs Schirmer's Test: A Comparative Study. *Eye* **2006**, *20*, 913–915. [CrossRef]
29. Stuchell, R.N.; Feldman, J.J.; Farris, R.L.; Mandel, I.D. The Effect of Collection Technique on Tear Composition. *Investig. Ophthalmol. Vis. Sci.* **1984**, *25*, 374–377.
30. Van Setten, G.B.; Stephens, R.; Tervo, T.; Salonen, E.M.; Tarkkanen, A.; Vaheri, A. Effects of the Schirmer Test on the Fibrinolytic System in the Tear Fluid. *Exp. Eye Res.* **1990**, *50*, 135–141. [CrossRef]
31. Cho, P.; Yap, M. Schirmer Test. I. A Review. *Optom. Vis. Sci.* **1993**, *70*, 152–156. [CrossRef]
32. Masmali, A.; Alqahtani, T.A.; Alharbi, A.; El-Hiti, G.A. Comparative Study of Repeatability of Phenol Red Thread Test Versus Schirmer Test in Normal Adults in Saudi Arabia. *Eye Contact Lens Sci. Clin. Pract.* **2014**, *40*, 127–131. [CrossRef]
33. Doughty, M.J.; Whyte, J.; Li, W. The Phenol Red Thread Test for Lacrimal Volume—Does It Matter If the Eyes Are Open or Closed? *Ophthalmic Physiol. Opt.* **2007**, *27*, 482–489. [CrossRef]
34. Choy, C.K.M.; Cho, P.; Chung, W.Y.; Benzie, I.F.F. Water-Soluble Antioxidants in Human Tears: Effect of the Collection Method. *Investig. Ophthalmol. Vis. Sci.* **2001**, *42*, 3130–3134.
35. Aass, C.; Norheim, I.; Eriksen, E.F.; Thorsby, P.M.; Pepaj, M. Single Unit Filter-Aided Method for Fast Proteomic Analysis of Tear Fluid. *Anal. Biochem.* **2015**, *480*, 1–5. [CrossRef] [PubMed]
36. García-Porta, N.; Mann, A.; Sáez-Martínez, V.; Franklin, V.; Wolffsohn, J.S.; Tighe, B. The Potential Influence of Schirmer Strip Variables on Dry Eye Disease Characterisation, and on Tear Collection and Analysis. *Contact Lens Anterior Eye* **2018**, *41*, 47–53. [CrossRef] [PubMed]
37. Tomlinson, A.; Blades, K.J.; Pearce, E.I. What Does the Phenol Red Thread Test Actually Measure? *Optom. Vis. Sci.* **2001**, *78*, 142–146. [CrossRef] [PubMed]
38. Smith, S.P.; Barbon, A.R.; Forbes, N.A. Evaluation of the Phenol Red Thread Tear Test in Falconiformes. *J. Avian Med. Surg.* **2015**, *29*, 25–29. [CrossRef]
39. Rajaei, S.M.; Sadjadi, R.; Sabzevari, A.; Ghaffari, M.S. Results of Phenol Red Thread Test in Clinically Normal Syrian Hamsters (Mesocricetus Auratus). *Vet. Ophthalmol.* **2013**, *16*, 436–439. [CrossRef]
40. Akkurt Arslan, M.; Kolman, I.; Pionneau, C.; Chardonnet, S.; Magny, R.; Baudouin, C.; Brignole-Baudouin, F.; Kessal, K. Proteomic Analysis of Tears and Conjunctival Cells Collected with Schirmer Strips Using Timstof pro: Preanalytical Considerations. *Metabolites* **2022**, *12*, 2. [CrossRef]
41. Kalló, G.; Emri, M.; Varga, Z.; Ujhelyi, B.; Tozsér, J.; Csutak, A.; Csosz, É. Changes in the Chemical Barrier Composition of Tears in Alzheimer's Disease Reveal Potential Tear Diagnostic Biomarkers. *PLoS ONE* **2016**, *11*, e0158000. [CrossRef]
42. Jung, J.H.; Ji, Y.W.; Hwang, H.S.; Oh, J.W.; Kim, H.C.; Lee, H.K.; Kim, K.P. Proteomic Analysis of Human Lacrimal and Tear Fluid in Dry Eye Disease. *Sci. Rep.* **2017**, *7*, 13363. [CrossRef]
43. Tamhane, M.; Cabrera-Ghayouri, S.; Abelian, G.; Viswanath, V. Review of Biomarkers in Ocular Matrices: Challenges and Opportunities. *Pharm. Res.* **2019**, *36*, 40. [CrossRef]
44. Zernii, E.Y.; Golovastova, M.O.; Baksheeva, V.E.; Kabanova, E.I.; Ishutina, I.E.; Gancharova, O.S.; Gusev, A.E.; Savchenko, M.S.; Loboda, A.P.; Sotnikova, L.F.; et al. Alterations in Tear Biochemistry Associated with Postanesthetic Chronic Dry Eye Syndrome. *Biochemistry* **2016**, *81*, 1549–1557. [CrossRef]

45. Hagan, S.; Martin, E.; Enríquez-de-Salamanca, A. Tear Fluid Biomarkers in Ocular and Systemic Disease: Potential Use for Predictive, Preventive and Personalised Medicine. *EPMA J.* **2016**, *7*, 15. [CrossRef]
46. Shen, S.; An, B.; Wang, X.; Hilchey, S.P.; Li, J.; Cao, J.; Tian, Y.; Hu, C.; Jin, L.; Ng, A.; et al. Surfactant Cocktail-Aided Extraction/Precipitation/On-Pellet Digestion Strategy Enables Efficient and Reproducible Sample Preparation for Large-Scale Quantitative Proteomics. *Anal. Chem.* **2018**, *90*, 10350–10359. [CrossRef]
47. Demichev, V.; Messner, C.B.; Vernardis, S.I.; Lilley, K.S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17*, 41–44. [CrossRef]
48. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13*, 731–740. [CrossRef]
49. Nolte, H.; MacVicar, T.D.; Tellkamp, F.; Krüger, M. Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Sci. Rep.* **2018**, *8*, 12648. [CrossRef]
50. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]

APPENDIX II.

European Journal of Pharmaceutical Sciences 156 (2021) 105593

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

Inhibition of ABCG2/BCRP-mediated transport–correlation analysis of various expression systems and probe substrates

Zsolt Sáfár[a], Gábor Kecskeméti[b], Judit Molnár[a], Anita Kurunczi[a], Zoltán Szabó[b], Tamás Janáky[b], Emese Kis[a], Péter Krajcsi[a,c,d,e,*]

[a] Solvo Biotechnology, a Charles River Company, 52 Közép fasor, Szeged H-6726, Hungary
[b] Department of Medical Chemistry, Faculty of Medicine, University of Szeged, Dóm tér 8, Szeged H-6720, Hungary
[c] Solvo Biotechnology, a Charles River Company, 4-20 Irinyi J str, Budapest H-1117, Hungary
[d] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter str 50/a, Budapest H-1083, Hungary
[e] Semmelweis University, Faculty of Health Sciences, Vas str 17, Budapest H-1088, Hungary

ARTICLE INFO

Keywords:
BCRP
ABCG2
HEK293-BCRP
vesicular transport
substrate-dependent inhibition

ABSTRACT

BCRP / ABCG2 is a key determinant of pharmacokinetics of substrate drugs. Several BCRP substrates and inhibitors are of low passive permeability, and the vesicular transport assay works well in this permeability space.

Membranes were prepared from BCRP-HEK293, MCF-7/MX, and baculovirus-infected Sf9 cells with (BCRP-Sf9-HAM), and without (BCRP-Sf9) cholesterol loading. K_m values for three substrates - estrone-3-sulfate, sulfasalazine, topotecan - correlated well between the four expression systems. In contrast, a 10-20-fold range in V_{max} values was observed, with BCRP-HEK293 membranes possessing the largest dynamic range. IC_{50} values of the different test systems were similar to each other, with 94.4% of pairwise comparisons being within 3-fold. Substrate dependent inhibition showed somewhat greater variation, as 81.4% of IC_{50} values in the BCRP-HEK293 membranes were within 3-fold in pairwise comparisons.

Overall, BCRP-HEK293 membranes demonstrated the highest activity. The IC_{50} values showed good concordance but substrate dependent inhibition was observed for some drugs.

1. Introduction

BCRP is one of the most important efflux transporters expressed in all major barriers (International Transporter Consortium et al., 2010). Testing of small molecule drug candidates for interaction with BCRP is required by all major regulatory agencies prior to new drug approval and marketing, with BCRP inhibition data a requirement (EMA, 2012; FDA, 2020; PMDA, 2018), and BCRP substrate assessment also either required (FDA, 2020; PMDA, 2018) or suggested (EMA, 2012). BCRP is a broad transporter with many low passive permeable substrates (Poirier et al., 2014; Safar et al., 2019). The membrane-based vesicular transport (VT) assay is generally considered the best test system for low passive permeable substrates in both substrate and inhibition assays (Poirier et al., 2014). However, transport activity of BCRP is significantly potentiated by increased membrane cholesterol content, hence BCRP activity is greater in human cells that contain higher levels of membrane cholesterol levels than in insect cells, which have low levels of membrane cholesterol (Pal et al., 2007). Accordingly, depletion of endogenous cholesterol from human cell membranes has been shown to decrease BCRP activity, and conversely enrichment of insect cell membrane with exogenous cholesterol can increase the activity of BCRP (Pal et al., 2007). Furthermore, BCRP localization is enriched within the plasma membrane in lipid rafts (Szilagyi et al., 2017), and the transporter also interacts with caveolin-1 (Storch et al., 2007). BCRP activity is therefore expected to vary depending on the expression system used to manufacture membranes that are utilized in VT assays. BCRP was also one of the first transporter proteins that was shown to display substrate-specific inhibition (Giri et al., 2009; Pedersen et al., 2017).

Drug-drug interaction testing using membrane vesicle-based assays are one of the options specified for *in vitro* BCRP test systems in regulatory guidances for new drug approvals (FDA, 2017; PMDA, 2018). However, despite evidence in the literature showing the impact of expression system on BCRP function, commercially available vesicles

* Corresponding author.
E-mail addresses: safar@solvo.com (Z. Sáfár), kecskemeti.gabor@med.u-szeged.hu (G. Kecskeméti), molnar.judit@solvo.com (J. Molnár), kurunczi@solvo.com (A. Kurunczi), szabo.zoltan@med.u-szeged.hu (Z. Szabó), janaky.tamas@med.u-szeged.hu (T. Janáky), kis@solvo.com, krajcsi@solvo.com (E. Kis), krajcsi@solvo.com (P. Krajcsi).

https://doi.org/10.1016/j.ejps.2020.105593
Received 3 July 2020; Received in revised form 23 September 2020; Accepted 9 October 2020
Available online 13 October 2020
0928-0987/

Table 1
Enzyme kinetic parameter estimates (mean ± standard deviation for at least three independent experiments) determined for estrone-3-sulfate, sulfasalazine, and topotecan substrates by four membranes.

Substrates	Membranes	V_{max} (pmol/mg protein/min)			K_m (μM)			Cl_{int} (μL/mg protein/min)		
estrone-3-sulfate	BCRP-HEK293	6957.00	±	900.49	9.61	±	0.94	723.93	±	117.49
	BCRP-M	2960.67	±	693.22	12.59	±	2.48	235.16	±	72.01
	BCRP-Sf9-HAM	2890.00	±	694.04	12.64	±	3.65	228.58	±	85.88
	BCRP-Sf9	291.00	±	120.42	23.10	±	6.64	12.60	±	6.35
sulfasalazine	BCRP-HEK293	2291.33	±	268.37	0.36	±	0.02	6364.81	±	825.08
	BCRP-M	644.30	±	120.29	0.26	±	0.05	2510.26	±	676.55
	BCRP-Sf9-HAM	276.30	±	53.37	0.25	±	0.04	1105.20	±	287.59
	BCRP-Sf9	191.87	±	45.03	0.29	±	0.07	654.09	±	212.66
topotecan	BCRP-HEK293	833.53	±	249.51	39.49	±	4.26	21.11	±	6.72
	BCRP-M	157.67	±	30.41	22.62	±	2.33	6.97	±	1.52
	BCRP-Sf9-HAM	123.66	±	29.48	19.02	±	8.80	6.50	±	3.38
	BCRP-Sf9	37.48	±	5.27	64.64	±	11.75	0.58	±	0.13

Table 2
IC_{50} values of estrone-3-sulfate substrate determined in all membrane preparations using fifteen BCRP inhibitors. The IC_{50} values differed by more than 3-fold are highlighted by bold font style. Data were presented as mean ± standard deviation for three independent experiments.

Inhibitor	$IC_{50,\ BCRP\text{-}HEK293}$ (μM)			$IC_{50,\ BCRP\text{-}M}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9}$ (μM)		
KO143	0.0209	±	0.0022	0.0322	±	0.0059	0.0214	±	0.0026	0.0197	±	0.0027
elacridar	0.0214	±	0.0011	0.0328	±	0.0026	0.0243	±	0.0061	0.0192	±	0.0060
eltrombopag	0.0397	±	0.0078	0.0272	±	0.0085	0.0314	±	0.0143	0.0289	±	0.0117
KO134	0.0943	±	0.0142	0.0844	±	0.0047	0.0659	±	0.0204	0.0907	±	0.0347
imatinib	0.2388	±	0.1158	0.4169	±	0.1977	0.3291	±	0.0672	0.3478	±	0.2143
MK571	1.25	±	0.31	1.55	±	0.31	1.38	±	0.41	0.9804	±	0.1170
leflunomide	0.5410	±	0.0932	0.6685	±	0.0959	1.09	±	0.46	0.9513	±	0.2780
ivermectin	0.3109	±	0.0353	0.2113	±	0.0541	0.3308	±	0.0258	0.3570	±	0.1505
prazosin	1.94	±	0.29	2.89	±	0.62	3.50	±	0.97	2.76	±	0.45
rapamycin	1.38	±	0.27	2.84	±	0.51	3.24	±	0.96	1.79	±	0.56
pantoprazole	4.78	±	0.32	13.16	±	4.78	9.31	±	1.43	8.66	±	2.82
sulfasalazine	0.5608	±	0.1757	0.6327	±	0.0810	0.4620	±	0.1384	1.05	±	0.40
ritonavir	7.19	±	1.15	10.10	±	0.87	9.05	±	2.07	6.76	±	1.90
omeprazole	28.30	±	2.55	83.80	±	32.22	34.85	±	3.33	40.82	±	7.46
furosemide	**12.11**	±	**2.71**	22.33	±	4.32	**42.29**	±	**13.90**	26.99	±	1.25

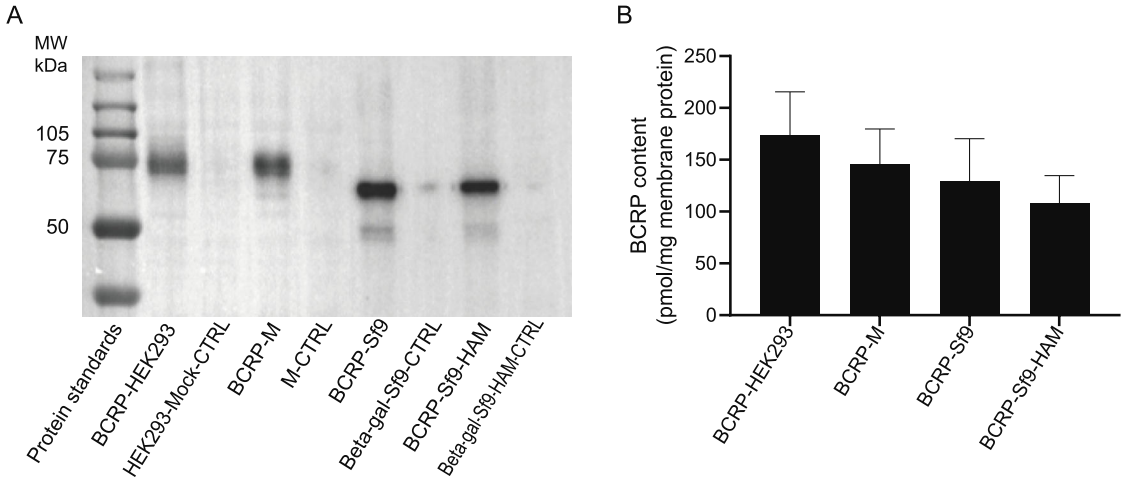


Fig. 1. BCRP protein expression analysis of the membrane preparations purified from the BCRP-overexpressing HEK293, MCF-7/MX and Sf9 cell lines. (A) Expression of BCRP (72 kDa) was assessed using the primary antibody BXP-21 in Western blot technique. The transporter protein was not evincible from the BCRP not-overexpressing negative control membrane samples (HEK293-Mock-CTRL, M-CTRL, Beta-gal-Sf9-CTRL, Beta-gal-Sf9-HAM-CTRL). (B) Absolute BCRP protein content of four membrane preparations were determined by PRM-based quantitative proteomics. Data shown as mean + standard deviation from at least three independent experiments.

generated using different cellular expression systems are widely used in drug research. Furthermore, no consensus substrate has been defined for *in vitro* testing (Safar et al., 2019). In the current study we therefore characterized transport kinetics of three different low passive permeable substrates, estrone-3-sulfate, sulfasalazine and topotecan using BCRP-containing membranes generated from four different expression systems. In addition, we tested and correlated IC_{50} data for fifteen well-known inhibitors of BCRP. Our data show that while expression levels of BCRP were similar by both Western blot- and LC-MS/MS-based proteomics across the four expression systems, the activity of BCRP in

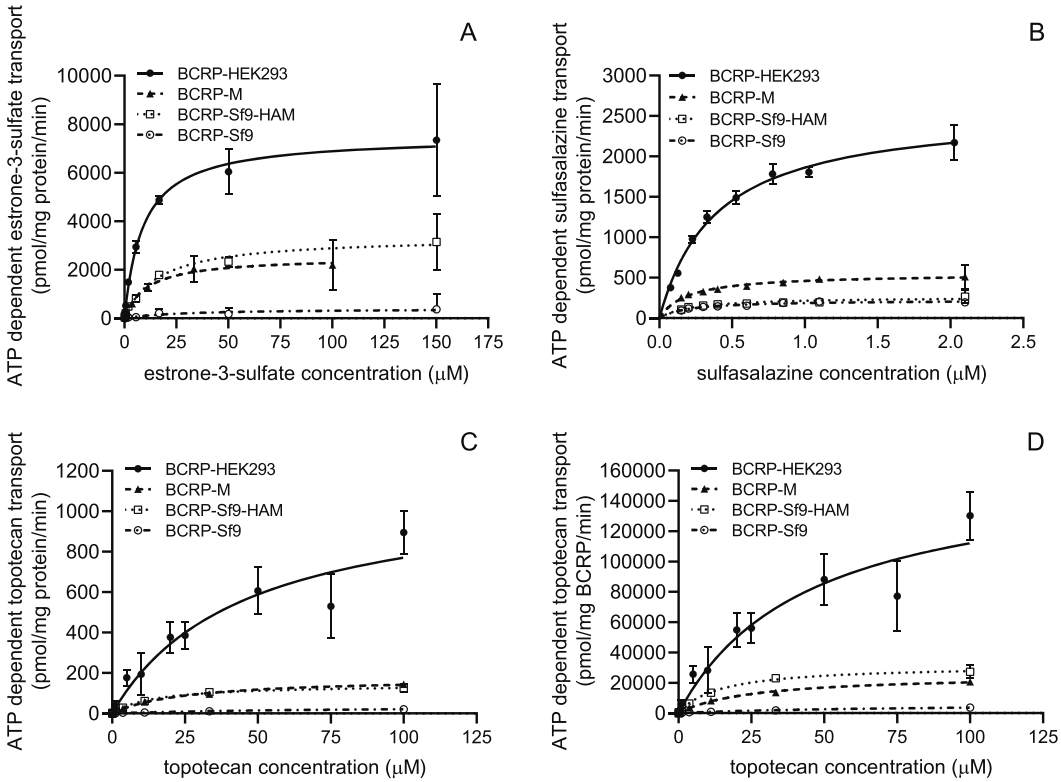


Fig. 2. The enzyme kinetic plots of estrone-3-sulfate (A), sulfasalazine (B), and topotecan (C) substrate concentration versus ATP dependent substrate transport (A, B, C) or BCRP expression correlated ATP dependent topotecan transport (D) measured on BCRP-HEK293 (filled circles and solid line), BCRP-M (filled triangles and dashed line), BCRP-Sf9-HAM (empty squares and dotted line) and BCRP-Sf9 (empty circles and dash-dotted line) membranes. Each data point represents the mean ± standard deviation of three independent experiments (n = 3). The fitted curves derived by non-linear regression using the Michaelis-Menten model.

HEK293-based membranes was superior to BCRP activities measured in the other expression systems. Although IC_{50} values for a few inhibitors showed slight substrate dependence, overall a good concordance was observed among all test systems.

2. Materials and Methods

2.1. Materials

Stable isotope (^{13}C and ^{15}N) labeled proteospecific peptide fragments of BCRP (SSLLDVLAAR (86-95), and LLSDLLPMR (473-481)) were ordered from JPT Peptide Technologies (Berlin, Germany). [^{3}H]-Estrone-3-sulfate (specific activity (S.A.), 51.3 Ci/mmol) and [^{3}H]-Sulfasalazine (S.A., 10 Ci/mmol) were synthesized at laboratory of Dr. Csaba Tömböly (Biological Research Center, Szeged, Hungary). Unlabeled topotecan hydrochloride hydrate (topotecan), estrone-3-sulfate and sulfasalazine were purchased from Sigma-Aldrich (St. Louis, MO, USA). In the inhibition experiments the following compounds were investigated: furosemide, ivermectin, KO143, leflunomide, MK571, omeprazole, pantoprazole sodium hydrate (pantoprazole), prazosin hydrochloride (prazosin) and ritonavir were purchased from Sigma-Aldrich; elacridar, eltrombopag, imatinib mesylate (imatinib) and rapamycin were from Carbosynth Ltd. (Compton, UK). KO134 was a kind gift of Prof. Gábor Tóth (Department of Medical Chemistry, Faculty of Medicine, University of Szeged, Szeged, Hungary). Recombinant baculoviruses encoding wild-type human ABCG2 were also a kind gifts from Prof. Balázs Sarkadi (National Medical Center, Institute of Haematology and Immunology, Budapest, Hungary). Randomly methylated-β-cyclodextrin (RAMEB) and cholesterol complex of RAMEB (cholesterol@RAMEB; cholesterol content, 4.74%) were purchased from Cyclolab (Cyclodextrin Research and Development Laboratory, Budapest, Hungary). Cholesterol contents of membrane preparations were determined using Amplex™ Red Cholesterol Assay KIT provided by ThermoFisher Scientific (Waltham, MA, USA). The selection agent puromycin was purchased from InvivoGen (Toulouse, France), the mitoxantrone from Sigma-Aldrich. Cell culture reagents were provided by ThermoFisher Scientific and Lonza (Basel, Switzerland). All other chemicals were purchased from Sigma-Aldrich unless stated otherwise in the text.

2.2. Cell culture

Sf9 insect cells were cultured and infected with ABCG2 or β-galactosidase encoding recombinant baculoviruses as described earlier (Ozvegy et al., 2002; Sarkadi et al., 1992). The human breast cancer MCF-7/WT and the mitoxantrone-resistant variant MCF-7/MX cells were cultured in mixture medium composed of equal parts of Dulbecco's Modified Eagle's Medium (DMEM) with 1 g/L glucose and DMEM F12 containing 10% v/v fetal bovine serum albumin, 2 mM L-glutamine and 100 μg/mL penicillin-streptomycin (100 unit/mL penicillin, 100 μg/mL streptomycin, Lonza). The MCF-7/MX cells were selected for mitoxantrone as described previously (Nakagawa et al., 1992). HEK293-BCRP and HEK293-Mock cell lines were generated from HEK293 cells (ThermoFisher Scientific) by lentiviral technology (Toth et al., 2018) using ABCG2 and resistance gene, respectively. HEK293 cells were cultured in DMEM (high glucose (25 mM)) with GlutaMAX$^{(TM)}$ containing 3 μg/mL puromycin, 10% v/v fetal bovine serum albumin, and 100 μg/mL penicillin-streptomycin. Human cells were grown on tissue culture flasks (VWR International Ltd., Radnor, PA, USA), passaged two times a week using a 0.25% trypsin-EDTA solution and incubated at 37°C in a humidified atmosphere containing 5% CO_2. The HEK293 cells were long-time cultured and propagated in Ex Cell™ 293 serum free medium (Sigma-Aldrich) in spinner flasks for later membrane preparations according to the suppliers' instructions.

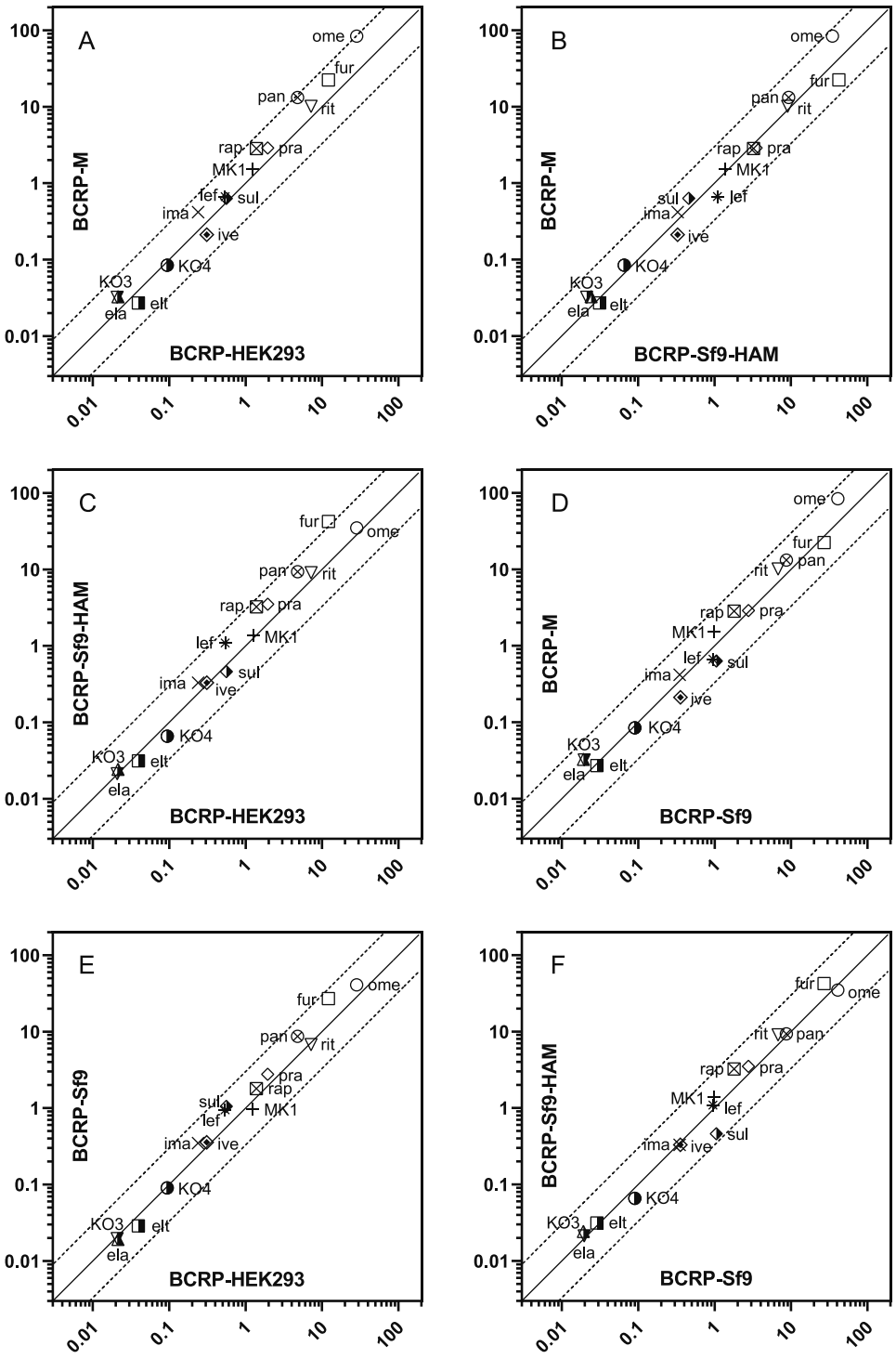


Fig. 3. Pairwise comparisons of IC_{50} (μM) values for the inhibitions of BCRP-mediated transport of estrone-3-sulfate substrate between (A) BCRP-HEK293 and BCRP-M, (B) BCRP-Sf9-HAM and BCRP-M, (C) BCRP-HEK293 and BCRP-Sf9-HAM, (D) BCRP-Sf9 and BCRP-M, (E) BCRP-HEK293 and BCRP-Sf9, (F) BCRP-Sf9 and BCRP-Sf9-HAM membranes. Each IC_{50} value is the mean ± standard deviation of values determined in three independent experiments (n = 3). The line of unity (the equal IC_{50} values for the inhibition of both compared membranes) is illustrated by the solid line, and the ratios of IC_{50} values that differ from one another by three-fold represented by dashed lines. The mean IC_{50} values of the fifteen inhibitor compounds depicted as follows: □ fur - furosemide, ○ ome - omeprazole, ▽ rit - ritonavir, ⊗ pan - pantoprazole, ⊠ rap - rapamycin, ◇ pra - prazosin, ◈ ive - ivermectin, ✳ lef - leflunomide, + MK1 - MK571, × ima - imatinib, ◑ KO4 - KO134, ◨ elt - eltrombopag, ▲ ela - elacridar, ▼ KO3 - KO143, □◆ sul - sulfasalazine.

2.3. Membrane preparation

All BCRP preparations contained the human wild-type (482R) version of the ABCG2 transporter (Accession number NM_004827). All BCRP-overexpressing human membrane vesicle preparations (BCRP-HEK293, BCRP-M), control human membrane preparations (HEK293-Mock-CTRL, M-CTRL), human BCRP transporter containing insect membranes (BCRP-Sf9, BCRP-Sf9-HAM), and control insect membrane vesicles (Beta-gal-Sf9-CTRL, Beta-gal-Sf9-HAM-CTRL) were obtained from SOLVO Biotechnology (Szeged, Hungary, http://www.solvo.com). Purified membrane vesicles from cells were prepared according to the method described previously (Sarkadi et al., 1992). HAM (high activity membrane) vesicles were cholesterol-loaded during the process of preparing membranes from baculovirus-infected Sf9 cells (Pal et al., 2007). The vesicles were stored at -80°C with 5 mg/mL total protein concentration, as determined by the bicinchoninic acid (BCA) method (Smith et al., 1985) (Sigma-Aldrich).

2.4. Western blot analysis of BCRP in the vesicles

Protein expression of BCRP was confirmed by Western blotting as described in detail elsewhere (Pal et al., 2007). Briefly, 5 μg total

Table 3
IC_{50} values of sulfasalazine substrate determined in all membrane preparations using fifteen BCRP inhibitors. The IC_{50} values differed by more than 3-fold are highlighted by bold font style. Data were presented as mean ± standard deviation for three independent experiments.

Inhibitor	$IC_{50,\ BCRP\text{-}HEK293}$ (μM)			$IC_{50,\ BCRP\text{-}M}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9}$ (μM)		
KO143	0.0080	±	0.0013	0.0113	±	0.0005	0.0109	±	0.0003	0.0104	±	0.0010
elacridar	0.0116	±	0.0026	0.0099	±	0.0014	0.0082	±	0.0013	0.0101	±	0.0012
eltrombopag	0.0275	±	0.0075	**0.0111**	±	**0.0029**	**0.0367**	±	**0.0078**	0.0328	±	0.0100
KO134	0.0377	±	0.0040	0.0450	±	0.0072	0.0521	±	0.0246	0.0343	±	0.0122
imatinib	0.2895	±	0.0704	0.2584	±	0.0377	0.2449	±	0.0380	0.2538	±	0.0712
MK571	0.5058	±	0.0807	0.4651	±	0.0564	0.7577	±	0.0776	0.5587	±	0.0400
leflunomide	1.01	±	0.45	**0.5382**	±	**0.1374**	**0.6649**	±	**0.1065**	**2.02**	±	**0.18**
ivermectin	2.33	±	1.35	1.10	±	0.13	1.81	±	0.38	1.33	±	0.08
prazosin	3.29	±	0.67	4.11	±	0.81	4.69	±	0.96	3.21	±	0.51
rapamycin	**9.13**	±	**3.01**	8.10	±	2.75	9.06	±	1.72	**1.69**	±	**0.21**
pantoprazole	10.40	±	1.11	12.19	±	2.28	10.99	±	1.43	10.21	±	2.02
estron-3-sulfate	12.11	±	0.77	17.85	±	6.63	12.48	±	3.11	23.78	±	4.71
ritonavir	18.77	±	3.40	16.17	±	2.86	12.93	±	3.07	12.81	±	1.33
omeprazole	24.46	±	6.30	44.90	±	9.15	33.93	±	8.22	46.90	±	5.24
furosemide	**31.97**	±	**2.48**	34.59	±	10.29	**98.99**	±	**25.54**	65.31	±	15.87

protein was separated in SDS-PAGE technique and blotted to a polyvinylidene fluoride (PVDF) membrane. BXP-21 antibody (Abcam, Cambridge, UK) against a conserved epitope on BCRP was used as the primary antibody in a 1:2000 dilution, followed by a horseradish peroxidase-conjugated anti-mouse IgG-HRP secondary antibody (Sigma-Aldrich), used at the same dilution. Immunoreactive bands were visualized by enhanced chemiluminescence (ECL) using a Chemidoc™ (BioRad, Hercules, CA, USA) imaging system.

2.5. Quantification of BCRP by targeted proteomics

25 μg protein were processed using a modified Filter Aided Sample Preparation (FASP) method (Potriquet et al., 2017). Briefly, proteins were reduced by the addition of DTT in 4 % m/V of SDS to a final concentration of 50 mM followed by incubation at 95°C for 5 min. The samples were transferred to Microcon Ultracel 30 kDa (Millipore, Burlington, MA, USA) centrifugal filter units and buffer exchange was performed in two successive wash with 8 M urea in 100 mM Tris-HCl (pH 8.5) with 10 min spin at 12,000 x g. Proteins were then alkylated in 50 μL 50 mM of IAA for 20 min in darkness at room temperature. Samples were washed two times with 100 μL urea and twice with 100 μL Tris-HCl. Protein digestion was then performed by adding 250 ng of trypsin in 40 μL 50 mM ammonium bicarbonate buffer (ABC) and incubating overnight at 37°C. Stable isotope-labeled (SIL) peptides (1 pmol in 10 μL ABC) were added to the sample prior to digestion. Digestion was stopped by the addition of 3 μL of formic acid (FA) and peptides were collected using an initial spin of 12,000 × g for 10 min followed by two centrifugations with 50 μL of 500 mM NaCl and 100 μL ABC. After evaporation, samples were resuspended in 60 μL of 3% ACN/97 % aqueous FA (0.1 % v/v).

LC-MS/MS analysis was carried out on a nanoACQUITY UPLC system (Waters, Milford, MA, USA) coupled with a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer (ThermoFisher Scientific). Chromatographic separation of 2 μg peptides was accomplished on an ACQUITY UPLC® M-Class Peptide BEH C18 column (130 Å, 1.7 μm, 75 μm × 250 mm) at 45°C by gradient elution. Water (solvent A) and acetonitrile (solvent B), both containing FA (0.1 % v/v) were used as mobile phases at a flow rate of 350 nL/min. The gradient was as follows: 3 % solvent B for 2 min, followed by a linear gradient to 40 % solvent B over 16 min, held at 40 % solvent B for 1 min, then held 2 min at 90 % solvent B, then returned to 3 % solvent B. In each measurement 5 μL sample was injected and trapped with Symmetry® C18 (100 Å, 5 μm, 180 μm × 20 mm) trap column.

After proteomics samples were prepared from each of the membranes, BCRP was quantified by determination of the amount of its two proteotypic peptides (SSLLDVLAAR (86-95) and LLSDLLPMR (473-481)) using LC-MS/MS. The mass spectrometer was operated in positive-ion mode using the equipped Nanospray Flex Ion Source. Parallel reaction monitoring (PRM) method was used to monitor the m/z transitions (y3, y4, y5, y6, y7, y8) for the + 2 charged peptides precursors of interest (522.8 m/z and 529.3 m/z for native peptides and 527.8 m/z and 534.3 m/z for stable isotope labeled ones). To reach the maximum sensitivity, precursor ions were fragmented using manually optimized collision energies. The automatic gain control (AGC) setting was defined as 2×10^5 charges, the maximum injection time was set to 60 ms and resolution at 17,500.

Data acquisition was performed using Xcalibur™ 4.1 (ThermoFisher Scientific), and Skyline 20.1 (MacLean et al., 2010) was used for data evaluation. The average of the two peptides to stable isotope labeled internal standard ratio was used for protein quantification.

2.6. Vesicular transport assay

The VT assay was conducted as described earlier (Jani et al., 2009; Pal et al., 2007) and optimized for each of the membrane vesicles using 3 probe substrates.

Concentration dependent vesicular transport assay for determination of kinetic constants: ATP-dependent estrone-3-sulfate and sulfasalazine transport into BCRP-containing and control vesicles was determined in the presence of 4 mM Mg-ATP or AMP in the incubation buffer (10 mM Tris–HCl, 250 mM sucrose, 10 mM $MgCl_2$, pH 7.4) on 96-well plates at 32°C for a 1 minute incubation time. The reaction mixtures containing 0.05 μM [^{3}H]-Estrone-3-sulfate or 0.1 μM [^{3}H]-Sulfasalazine were adjusted with an 8-step, 3-fold (or irregular) serial dilution with non-radiolabeled compounds to reach the desired final substrate concentrations in each well. The total membrane protein contents were between 6.25 and 25 μg/well, optimized for each assay. The final reaction volume was 75 μL in each well. Transport was stopped by the addition of cold wash buffer (10 mM Tris–HCl, 250 mM sucrose, 100 mM NaCl, pH 7.4) then rapidly filtered through PVDF/glass fiber filters mounted to a 96-well plate (MSFBN6B, Millipore Corporation, Billerica, MA, USA). Filters were washed with 200 μL ice-cold wash buffer five times, dried, and the retained radioactivity was measured by liquid scintillation counting (Wallac MicroBeta TriLux, Perkin-Elmer, Waltham, MA, USA). ATP-dependent topotecan transport measurements were carried out in incubation buffer containing 40 mM MOPS-Tris (pH 7.4), 70 mM KCl and 7.5 mM $MgCl_2$ at 37°C with a 10 minute (or 1 minute for BCRP-HEK293) incubation time. The final concentrations of topotecan were reached by adding non-radiolabeled topotecan solutions from an 8-step, 3-fold (or irregular) serial dilution to the reaction mixtures in the wells. The total membrane protein content was

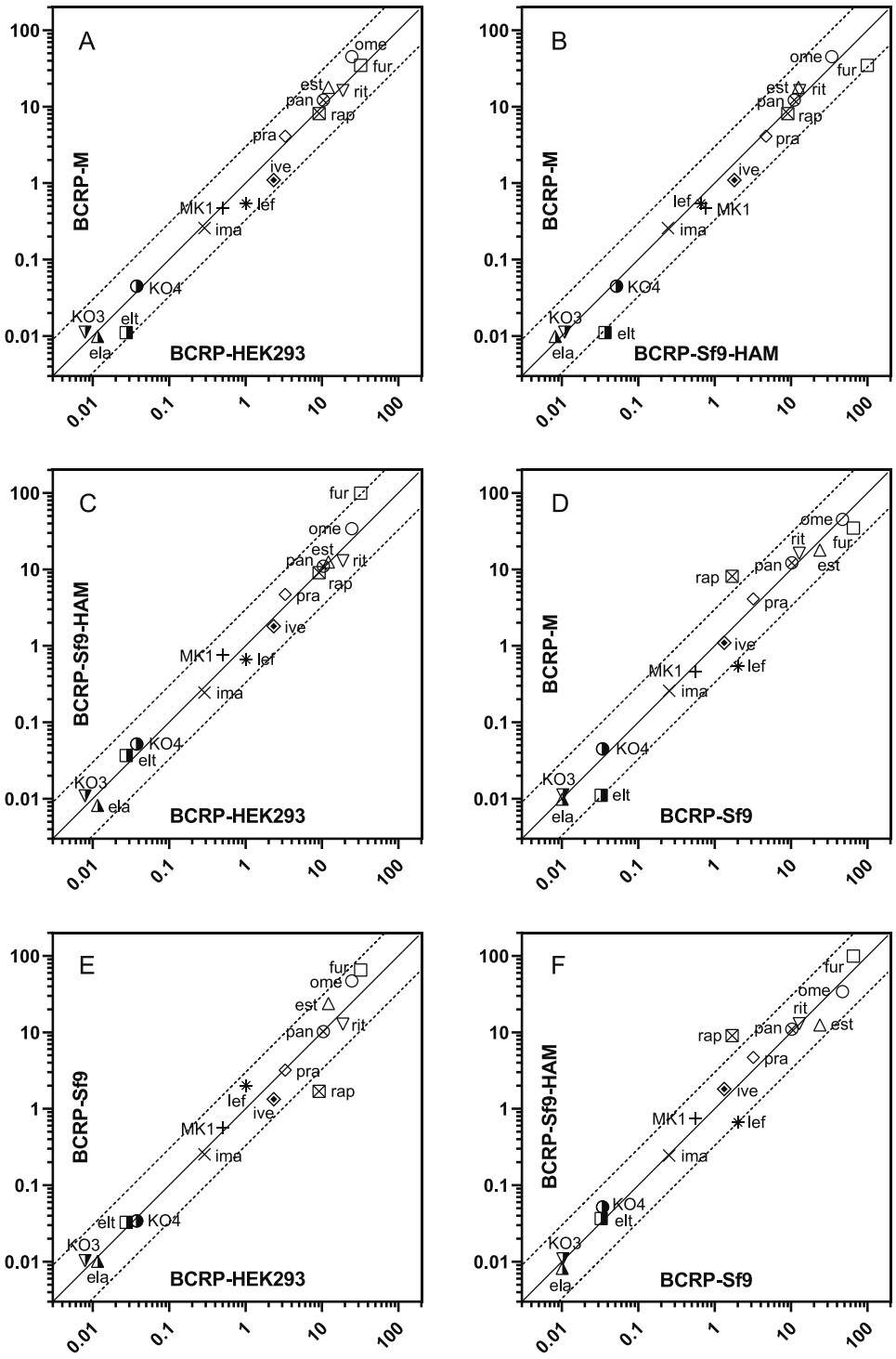


Fig. 4. Pairwise comparisons of IC_{50} (μM) values for the inhibitions of BCRP-mediated transport of sulfasalazine substrate between (A) BCRP-HEK293 and BCRP-M, (B) BCRP-Sf9-HAM and BCRP-M, (C) BCRP-HEK293 and BCRP-Sf9-HAM, (D) BCRP-Sf9 and BCRP-M, (E) BCRP-HEK293 and BCRP-Sf9, (F) BCRP-Sf9 and BCRP-Sf9-HAM membranes. Each IC_{50} value is the mean ± standard deviation of values determined in three independent experiments (n = 3). The line of unity (the equal IC_{50} values for the inhibition of both compared membranes) is illustrated by the solid line, and the ratios of IC_{50} values that differ from one another by three-fold represented by dashed lines. The mean IC_{50} values of the fifteen inhibitor compounds depicted as follows: □ fur - furosemide, ○ ome - omeprazole, ▽ rit - ritonavir, △ est - estrone-3-sulfate, ⊗ pan - pantoprazole, ⊠ rap - rapamycin, ◇ pra - prazosin, ◈ ive - ivermectin, ✳ lef - leflunomide, + MK1 - MK571, × ima - imatinib, ◑ KO4 - KO134, ◨ elt - eltrombopag, ◭ ela - elacridar, ▼ KO3 - KO143.

between 25 and 50 μg/well, optimized to each assay. As described above the incubation reactions were quenched and washed through a 96-well filter plate with ice cold wash buffer (40 mM MOPS-Tris, 70 mM KCl, pH 7.4). Topotecan was eluted from filters into a 96-well Costar plate (Corning), with 100 μL HCl (10 mM) solution, and fluorescence of the eluate was measured using excitation/emission wavelengths 355/590 nm with a FLUOstar Omega microplate reader (BMG Labtech, Ortenberg, Germany). ATP-dependent transport was calculated by subtracting the values obtained in the presence of AMP from those in the presence of ATP.

Inhibition experiments for determination of IC_{50} values: Inhibition experiments were conducted in a similar manner to the concentration dependent VT assays are described above, with some modifications. Vesicles were incubated in the presence of various inhibitors, with 1 μM estrone-3-sulfate, 0.1 μM sulfasalazine or 10 μM topotecan used as probe substrates. Inhibitor stock solutions were prepared, and 8-step, 3-fold serial dilutions were made for each inhibitor in dimethyl sulfoxide (DMSO), which were then diluted 100-fold upon addition to each of the wells in the assay.

Table 4
IC_{50} values of topotecan substrate determined in all membrane preparations using fifteen BCRP inhibitors. The IC_{50} values differed by more than 3-fold are highlighted by bold font style. Data were presented as mean ± standard deviation for three independent experiments.

Inhibitor	$IC_{50,\ BCRP\text{-}HEK293}$ (μM)			$IC_{50,\ BCRP\text{-}M}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ (μM)			$IC_{50,\ BCRP\text{-}Sf9}$ (μM)		
KO143	0.0155	±	0.0066	0.0146	±	0.0045	**0.0053**	±	**0.0041**	**0.0186**	±	**0.0029**
elacridar	**0.0258**	±	**0.0071**	**0.0151**	±	**0.0066**	**0.0046**	±	**0.0013**	**0.0057**	±	**0.0012**
eltrombopag	**0.0578**	±	**0.0078**	**0.1954**	±	**0.0429**	0.0735	±	0.0126	0.1342	±	0.1068
KO134	0.0394	±	0.0065	0.0336	±	0.0079	0.0403	±	0.0135	0.0857	±	0.0227
imatinib	0.0786	±	0.0252	0.1369	±	0.0043	0.1011	±	0.0189	0.1796	±	0.0906
MK571	0.4288	±	0.0580	0.4903	±	0.1097	0.3525	±	0.1084	0.8604	±	0.1982
leflunomide	**0.9453**	±	**0.3056**	**3.11**	±	**0.75**	**0.7453**	±	**0.0244**	1.07	±	0.23
ivermectin	0.1525	±	0.0356	0.1953	±	0.0734	0.1073	±	0.0580	0.2039	±	0.1277
prazosin	0.9304	±	0.3063	0.7684	±	0.1843	1.16	±	0.23	1.80	±	1.08
rapamycin	0.9745	±	0.4039	0.7788	±	0.2482	**0.3460**	±	**0.1858**	**1.25**	±	**0.44**
pantoprazole	3.81	±	0.54	3.25	±	0.22	4.74	±	1.49	6.03	±	1.34
estrone-3-sulfate	10.49	±	2.67	13.55	±	4.44	21.40	±	12.99	25.58	±	5.20
ritonavir	2.46	±	1.30	1.45	±	0.18	3.71	±	1.61	3.79	±	1.59
omeprazole	15.18	±	3.01	45.40	±	5.85	19.07	±	8.60	35.08	±	6.33
furosemide	**47.72**	±	**18.19**	65.64	±	35.67	62.14	±	10.51	**183.07**	±	**69.06**

2.7. Data analysis

A data points for concentration-dependent VT assays were derived from a minimum of three independent experiments. Michaelis–Menten constant (K_m) and maximal velocity (V_{max}) values were calculated from triplicate determinations by non-linear regression analysis. The Michaelis–Menten equation was fitted onto the baseline-corrected accumulation data using GraphPad Prism v7 (GraphPad Inc., San Diego, CA, USA) to obtain best-fit parameters. Intrinsic clearance (Cl_{int}) values were calculated by dividing the V_{max} by the K_m values for each experiment. Inhibition VT assays were conducted with triplicate data points in three independent experiments. IC_{50} values were calculated with nonlinear regression for a sigmoidal dose-response using the four-parameter logistic equation, and best fit was approached with the least square method using GraphPad Prism. Data are expressed as arithmetic means and standard deviations.

3. Results

To study kinetic parameters of transport and IC_{50} values for inhibition by fifteen known inhibitors, membranes overexpressing BCRP were generated from HEK293 cells stably transduced with lentivirus carrying ABCG2 (BCRP-HEK293), mitoxantrone-selected MCF-7/MX cells (Nakagawa et al., 1992) (BCRP-M), Sf9 cells infected with a baculovirus harboring ABCG2 (BCRP-Sf9), and Sf9 cells infected with a baculovirus harboring ABCG2 and loaded with exogenous cholesterol (BCRP-Sf9-HAM). BCRP protein expression in each of the membranes was quantified by Western blotting (Fig 1A) and quantitative proteomics (Fig 1B). BCRP in all four membranes showed a similar level of protein expression by Western blot (Fig 1), with LC-MS/MS proteomics data on BCRP protein content ranging from 107.68 ± 27.01 pmol/mg membrane protein in the BCRP-Sf9-HAM membranes to 173.26 ± 42.16 pmol/mg membrane protein in BCRP-HEK293 membranes (Fig 1B).

Kinetic parameters were determined for estrone-3-sulfate, sulfasalazine and topotecan as substrates (Fig 2A-C). K_m values were determined to be within a 3-fold range for all three substrates across the four membranes (Table 1), but for topotecan in BCRP-Sf9-HAM and BCRP-Sf9 membranes the values were 19.02 μM and 64.64 μM, respectively. Sulfasalazine showed K_m values in the submicromolar range (0.25 – 0.36 μM). K_m values for estrone-3-sulfate ranged from 9.61 to 23.10 μM and for topotecan, the lowest affinity substrate, from 19.02 μM to 64.64 μM. Calculated V_{max} values were, in contrast, much more variable with a greater than 20-fold difference between different membrane preparations using estrone-3-sulfate and topotecan, and more than 10-fold difference between different membrane preparations using sulfasalazine, with the BCRP-HEK293 membrane being the most efficient, and the BCRP-Sf9 being the least efficient membrane for BCRP-mediated transport (Table 1). BCRP-M and BCRP-Sf9-HAM displayed intermediate values that were within 3-fold of each other for all substrates (Table 1). Intrinsic clearance values showed the greatest difference for estrone-3-sulfate (cca 60-fold) and the lowest difference for sulfasalazine (cca 10-fold), with BCRP-HEK293 membranes being the most efficient and BCRP-Sf9 being the least efficient membranes for all substrates (Table 1). When V_{max} values for topotecan were correlated for BCRP expression, the calculated values were 162227 pmol/mg BCRP/min, 25379 pmol/mg BCRP/min, 31621 pmol/mg BCRP/min and 6099 pmol/mg BCRP/min for BCRP-HEK293, BCRP-M, BCRP-Sf9-HAM and BCRP-Sf9 membranes, respectively (Fig 2D). To test if membrane cholesterol content was responsible for the differences observed in V_{max} values cholesterol measurements were carried out (Supplementary Fig 1). As expected, the lowest amount of membrane cholesterol was observed in Sf9 membranes. Both HAM membranes as well as membranes prepared from human cells (HEK293 and MCF-7/MX) contained higher amount of cholesterol than Sf9 membranes.

IC_{50} values were generated for all three substrates in all expression systems using fifteen known BCRP inhibitors (Tables 2-4; Fig 3-5). The vast majority of IC_{50} values were within 3-fold in pairwise comparison between different expression systems for the same probe substrate. IC_{50} values with estrone-3-sulfate as a probe are shown in Table 2 and pairwise comparison of data is shown in Fig 3. A greater than 3-fold difference in these data were observed only for furosemide with $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ / $IC_{50,\ BCRP\text{-}HEK293}$ = 3.49 (Table 2; Fig 3C). IC_{50} values with sulfasalazine as a probe are shown in Table 3 and pairwise comparison of data is shown in Fig 4. For sulfasalazine, four compounds showed IC_{50} values that were outside the 3-fold range in pairwise comparisons (Table 3). IC_{50} values for leflunomide differed more than 3-fold in two pairwise comparisons ($IC_{50,\ BCRP\text{-}Sf9}$ / $IC_{50,\ BCRP\text{-}M}$ = 3.75 (Table 3; Fig 4D) and $IC_{50,\ BCRP\text{-}Sf9}$ / $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ = 3.04 (Table 3; Fig 4F)). One pairwise comparison showed a difference of greater than 3-fold for eltrombopag ($IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ / $IC_{50,\ BCRP\text{-}M}$ = 3.31 (Table 4; Fig 4B)), rapamycin ($IC_{50,\ BCRP\text{-}HEK293}$ / $IC_{50,\ BCRP\text{-}Sf9}$ = 5.4 (Table 3; Fig 4E)) and furosemide ($IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ / $IC_{50,\ BCRP\text{-}HEK293}$ = 3.10 (Table 3; Fig 4C)). IC_{50} data for topotecan showed differences greater than 3-fold in pairwise comparisons for 6 compounds (Table 4). Elacridar and leflunomide displayed differences greater than 3-fold in two pairwise comparisons. For elacridar $IC_{50,\ BCRP\text{-}HEK293}$ / $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ = 5.61 (Table 4; Fig 5C), $IC_{50,\ BCRP\text{-}M}$ / $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ = 3.28 (Table 4; Fig 5B) and $IC_{50,\ BCRP\text{-}HEK293}$ / $IC_{50,\ BCRP\text{-}Sf9}$ = 4.53 (Table 4; Fig 5E) were measured and for leflunomide the two pairwise comparisons with IC_{50} values differing by more than 3-fold were $IC_{50,\ BCRP\text{-}M}$ / $IC_{50,\ BCRP\text{-}HEK293}$ = 3.29 (Table 4; Fig 5A) and $IC_{50,\ BCRP\text{-}M}$ / $IC_{50,\ BCRP\text{-}$

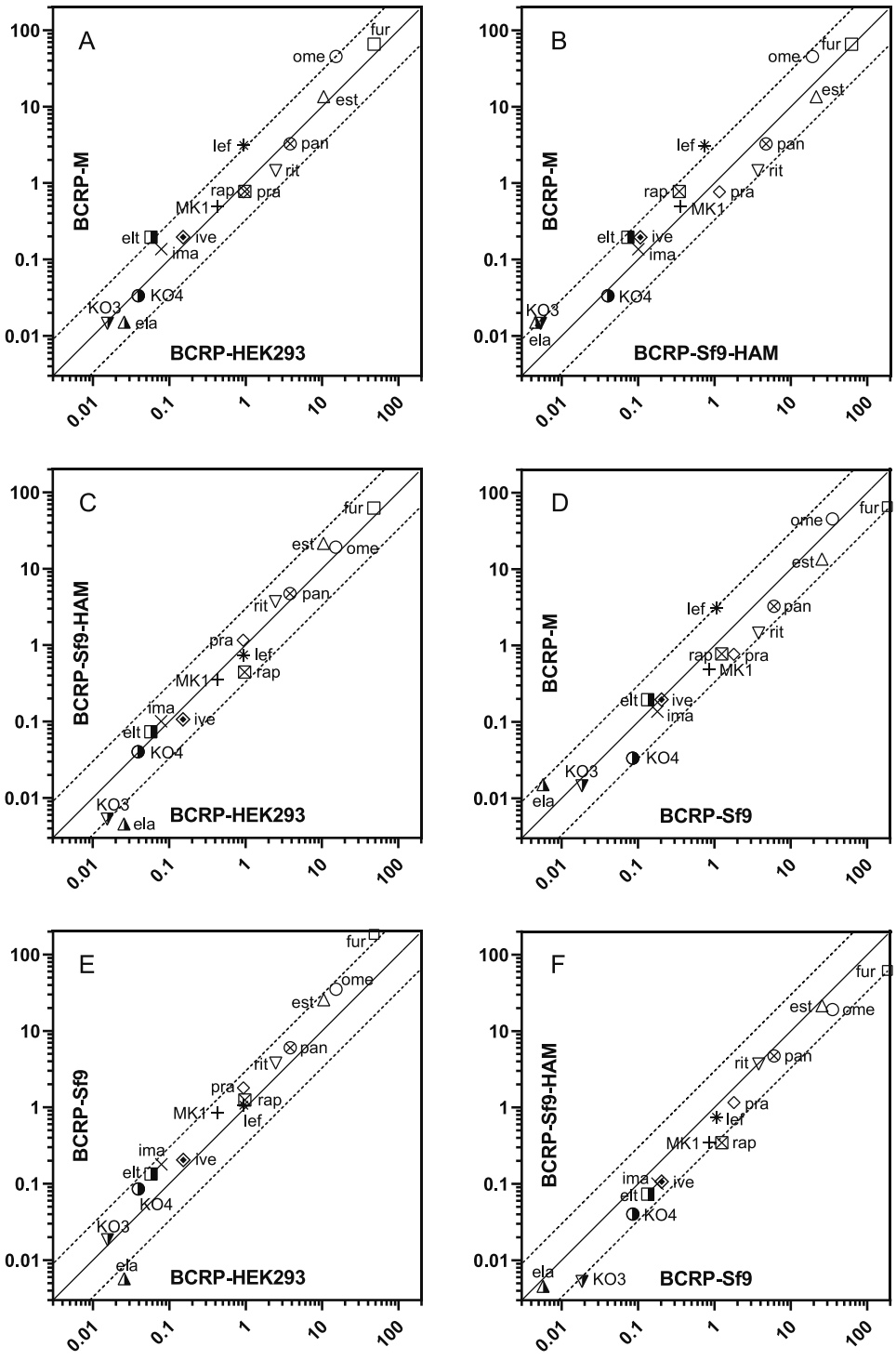


Fig. 5. Pairwise comparisons of IC_{50} (μM) values for the inhibitions of BCRP-mediated transport of topotecan substrate between (A) BCRP-HEK293 and BCRP-M, (B) BCRP-Sf9-HAM and BCRP-M, (C) BCRP-HEK293 and BCRP-Sf9-HAM, (D) BCRP-Sf9 and BCRP-M, (E) BCRP-HEK293 and BCRP-Sf9, (F) BCRP-Sf9 and BCRP-Sf9-HAM membranes. Each IC_{50} value is the mean ± standard deviation of values determined in three independent experiments (n = 3). The line of unity (the equal IC_{50} values for the inhibition of both compared membranes) is illustrated by the solid line, and the ratios of IC_{50} values that differ from one another by three-fold represented by dashed lines. The mean IC_{50} values of the fifteen inhibitor compounds depicted as follows: □ fur - furosemide, ○ ome - omeprazole, ▽ rit - ritonavir, △ est - estrone-3-sulfate, ⊗ pan - pantoprazole, ⊠ rap - rapamycin, ◇ pra - prazosin, ◈ ive - ivermectin, ✳ lef - leflunomide, + MK1 - MK571, × ima - imatinib, ◑ KO4 - KO134, ◨ elt - eltrombopag, ◭ ela - elacridar, ▼ KO3 - KO143.

$_{Sf9\text{-}HAM}$ = 4.17 (Table 4; Fig 5B). KO143, eltrombopag, rapamycin and furosemide showed IC_{50} values that differed by more than 3-fold in one pairwise comparison each. KO143 showed an $IC_{50,\ BCRP\text{-}Sf9}$ / $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ = 3.51 (Table 4; Fig 5F), eltrombopag showed an $IC_{50,\ BCRP\text{-}M}$ / $IC_{50,\ BCRP\text{-}HEK293}$ = 3.38 (Table 4; Fig 5A), rapamycin an $IC_{50,\ BCRP\text{-}Sf9}$ / $IC_{50,\ BCRP\text{-}Sf9\text{-}HAM}$ = 3.61 (Table 4; Fig 5F) and furosemide $IC_{50,\ BCRP\text{-}Sf9}$ / $IC_{50,\ BCRP\text{-}HEK293}$ = 3.84 (Table 4; Fig 5E).

In sum, of the 270 pairwise comparisons only 15 (5.6 %) fell outside the 3-fold limit and out of these pairwise comparisons only 4 (1.5 %) differed by more than 4-fold (Table 2-4).

When IC_{50} data determined in the most efficient BCRP-HEK293 membranes was correlated between the three substrates, the best correlation was found between IC_{50} data measured with estrone-3-sulfate and topotecan (Table 5; Fig 6B), as only furosemide IC_{50} values differed by more than 3-fold (12.1 μM and 47.7 μM respectively). IC_{50} values measured using estrone-3-sulfate and sulfasalazine as probe substrates correlated also well (Table 5; Fig 6A) as they differed more than 3-fold only for ivermectin (0.311 μM and 2.33 μM respectively) and for rapamycin (1.38 μM and 9.13 μM respectively). The greatest difference was observed in the sulfasalazine and topotecan as probe substrates comparison (Table 5; Fig 6C) as IC_{50} data for five inhibitors such as imatinib (0.290 μM and 0.079 μM respectively), ivermectin (2.33 μM and 0.153 μM respectively), prazosin (3.29 μM and 0.930 μM respectively), ritonavir (18.8 μM and 2.46 μM respectively) and

Table 5
IC_{50} values of estrone-3-sulfate, sulfasalazine and topotecan substrate determined in BCRP-HEK293 membrane preparation using sixteen BCRP inhibitors all together. The IC_{50} values differed by more than 3-fold are highlighted by bold font style. Data were presented as mean ± standard deviation for three independent experiments.
N/D* - IC_{50} values were not determined in case of identity of substrate and inhibitor.
N/D** - IC_{50} value was not determined due to the possible interference of sulfasalazine in detection of fluorescence of topotecan.

Inhibitor	$IC_{50,\ BCRP\text{-}HEK293}$ (μM) estrone-3-sulfate			sulfasalazine			topotecan		
KO143	0.0209	±	0.0022	0.0080	±	0.0013	0.0155	±	0.0066
elacridar	0.0214	±	0.0011	0.0116	±	0.0026	0.0258	±	0.0071
eltrombopag	0.0397	±	0.0078	0.0275	±	0.0075	0.0578	±	0.0078
KO134	0.0943	±	0.0142	0.0377	±	0.0040	0.0394	±	0.0065
imatinib	0.2388	±	0.1158	**0.2895**	±	**0.0704**	**0.0786**	±	**0.0252**
MK571	1.25	±	0.31	0.5058	±	0.0807	0.4288	±	0.0580
leflunomide	0.5410	±	0.0932	1.01	±	0.45	0.9453	±	0.3056
ivermectin	**0.3109**	±	**0.0353**	**2.33**	±	**1.35**	**0.1525**	±	**0.0356**
prazosin	1.94	±	0.29	**3.29**	±	**0.67**	**0.9304**	±	**0.3063**
rapamycin	**1.38**	±	**0.27**	**9.13**	±	**3.01**	**0.9745**	±	**0.4039**
pantoprazole	4.78	±	0.32	10.40	±	1.11	3.81	±	0.54
estrone-3-sulfate		N/D*		12.11	±	0.77	10.49	±	2.67
ritonavir	7.19	±	1.15	**18.77**	±	**3.40**	**2.46**	±	**1.30**
omeprazole	28.30	±	2.55	24.46	±	6.30	15.18	±	3.01
furosemide	**12.11**	±	**2.71**	31.97	±	2.48	**47.72**	±	**18.19**
sulfasalazine	0.5608	±	0.1757		N/D*			N/D**	

rapamycin (9.13 μM and 0.975 μM respectively) differed more than 3-fold. A greater than 10-fold difference (2.33/0.153 = 15.2) was only observed for ivermectin inhibition when comparing sulfasalazine and topotecan as probe substrates.

In sum, in 43 pairwise comparisons IC_{50} data for 14 pairs differed by more than 3-fold (Table 5; Fig 6).

4. Discussion

The use of VT assay for substrate and inhibitor testing of transport proteins has been highlighted by multiple studies (Bentz et al., 2013; Glavinas et al., 2008; Heredi-Szabo et al., 2013; Lumen et al., 2010; Poirier et al., 2014; Safar et al., 2019), and this method of testing potential drug interactions with BCRP has also been encouraged by regulatory agencies (EMA, 2012; FDA, 2020; PMDA, 2018). Testing for substrates and inhibitors of BCRP using VT assays is of particular importance as BCRP is known to have many low passive permeable substrates and inhibitors (Poirier et al., 2014), which can present technical challenges for other *in vitro* transport assay types. A 15 nm/s transcellular permeability measured in MDCKII-BCRP cells in the presence of inhibitor has been suggested as the permeability threshold (Poirier et al., 2014) for determining BCRP substrates in a bidirectional transcellular assay format, and furthermore a dependence of IC_{50} values on the test system used (cellular vs vesicular) has also been shown (Poirier et al., 2014; Szeremy et al., 2011). In general, higher IC_{50} values were observed in cellular assays than in VT assays for many low permeable inhibitors (Poirier et al., 2014; Szeremy et al., 2011). Therefore, utilization of VT assays for low permeable substrates and inhibitors is justified.

Membrane preparations generated from a variety of different cellular backgrounds are commercially available for use in transporter studies. Membranes generated from Sf9 cells infected with baculovirus harboring ABCG2 cDNA were extensively used for early studies (Glavinas et al., 2007). However it was soon discovered that for efficient BCRP transport, Sf9 cells are not the optimal host cells as the composition of insect membranes is quite different from that of mammals, and in particular that insect cell membranes have low levels of membrane cholesterol that is deleterious for the transport activity of BCRP (Pal et al., 2007). Therefore, Sf9 membranes enriched with cholesterol (high activity membranes (HAM); BCRP-Sf9-HAM) or membranes prepared from selected mammalian MCF-7/MX cells (BCRP-M) have been used instead (Pal et al., 2007). In a study on another cholesterol-sensitive ABC transporter, P-glycoprotein (P-gp), it was found that membrane cholesterol is likely not the only determinant for efficient transporter activity, as membrane vesicles generated from P-gp-overexpressing K562-MDR cells displayed a greater transport of N-methyl-quinidine (NMQ) than insect cell-derived P-gp-Sf9-HAM vesicles despite much lower expression of the transporter in the mammalian vesicles (Heredi-Szabo et al., 2013). We therefore included membranes generated from an additional cell line, HEK293 cells stably expressing BCRP (BCRP-HEK293) in our study. HEK293 cells are commonly used in transporter studies and grow in suspension culture. Our data show that BCRP-HEK293 membranes contain similar amounts of cholesterol as BCRP-M membranes do (Supplementary Fig 1). However, V_{max} values of BCRP-HEK293 membranes for all substrates were 2-5-fold higher than V_{max} values of BCRP-M membranes. V_{max} value for topotecan corrected for BCRP expression was about 6-fold higher in BCRP-HEK293 membrane than the corresponding value in BCRP-M membrane (162227 pmol/mg BCRP/min and 25379 pmol/mg BCRP/min respectively). Therefore, other factors contribute to activity of BCRP. Membrane lipid content is different for many membrane components between HEK293 and Sf9 membranes (Dawaliby et al., 2016). However, it is not known how the membrane lipid classes other than cholesterol affect activity of BCRP.

It has also been shown that the potency of BCRP inhibition may depend on the probe substrate used (Poirier et al., 2014). Therefore, for the purpose of the study we selected three different substrates: estrone-3-sulfate as a physiological substrate, and two drugs sulfasalazine and topotecan that represent different indications (inflammatory and cancer, respectively) and different chemical spaces. Estrone-3-sulfate and sulfasalazine are currently used as *in vitro* BCRP probe substrates (Safar et al., 2019) and sulfasalazine is also a clinical probe substrate. Topotecan is a chemotherapeutic drug and as such its use in clinical drug-drug interaction (DDI) studies is limited. Nevertheless, topotecan was the victim and elacridar was the perpetrator in the first documented BCRP-mediated DDI study (Kruijtzer et al., 2002). It is also fluorescent and thus commonly used in *in vitro* studies (Safar et al., 2019). As it is a chemotherapeutic agent it represents a therapeutic class where multidrug transporters are thought to play an important role in drug resistance (Klukovits and Krajcsi, 2015). In addition, it is structurally very different from estrone-3-sulfate and sulfasalazine. Our data show that K_m values were within 3-fold for all substrates in all membrane systems with the notable exception of topotecan, for which the K_m determined using BCRP-Sf9 membranes was 3.4-fold higher than

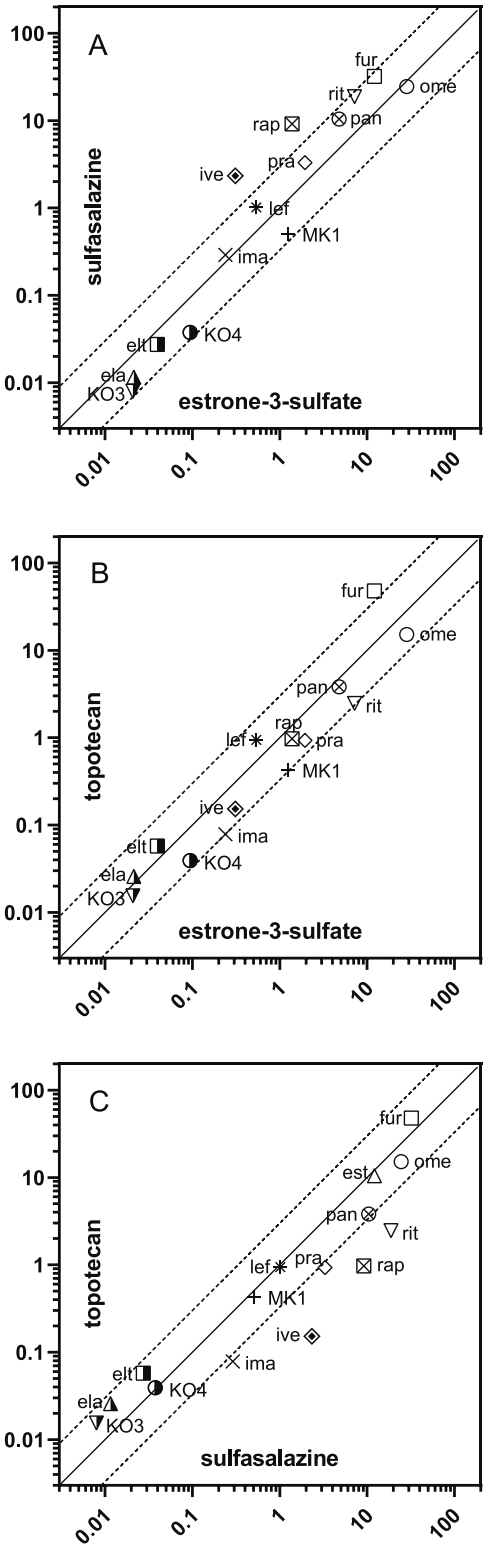


Fig. 6. Pairwise comparisons of IC_{50} (μM) values for the inhibitions of BCRP-mediated transport of estrone-3-sulfate and sulfasalazine (A), estrone-3-sulfate and topotecan (B), sulfasalazine and topotecan (C) substrates respectively measured in BCRP-HEK293 membrane preparation. Each IC_{50} value is the mean ± standard deviation of values determined in three independent experiments (n = 3). The line of unity (the equal IC_{50} values for the inhibition of both membranes compared) is illustrated by the solid line, and the ratios of IC_{50} values that differ from one another by three-fold represented by dashed lines. The mean IC_{50} values of inhibitor compounds are depicted as follows: □ fur - furosemide, ○ ome - omeprazole, ▽ rit - ritonavir, △ est - estrone-3-sulfate, ⊗ pan - pantoprazole, ⊠ rap - rapamycin, ◇ pra - prazosin, ◈ ive - ivermectin, ✳ lef - leflunomide, + MK1 - MK571, × ima - imatinib, ◑ KO4 - KO134, ◨ elt - eltrombopag, ◭ ela - elacridar, ▼ KO3 - KO143.

when measured using BCRP-Sf9-HAM membranes.

The K_m data for estrone-3-sulfate generated in this study, which ranged between 9.61 μM and 23.1 μM, were in good agreement with the K_m of 16.4 μM reported by in the literature by other researchers using a similar VT assay format (Xia et al., 2005). For sulfasalazine, the K_m data from this study ranged between 0.25 μM and 0.36 μM, which is slightly below the K_m of 0.7 μM described in an earlier VT study (Jani et al., 2009). Furthermore, the K_m values for topotecan, which we found to range between 19.0 μM and 64.6 μM, were significantly below the K_m of 213.3 μM (Li et al., 2008) previously determined using MDCKII-BCRP cells in a bidirectional monolayer assay setup. The lack of correlation between K_m values generated for some low passive permeable substrates in VT and monolayer systems may indicate the limitation of cellular systems in studying kinetics of transport of low passive permeable substrates (Poirier et al., 2014). Accurate determination of K_m values in monolayer experiments for low passive permeable substrates in particular may require a model that is based on unbound intracellular concentration (Tachibana et al., 2010).

In contrast to the K_m values, when V_{max} was calculated for each of the different membrane preparations, there was a marked difference for all 3 substrates with values for the highest activity membranes (BCRP-HEK293) over 20-fold higher than those determined for the lowest activity membranes (BCRP-Sf9). Therefore, while affinities of substrates for BCRP did not differ significantly between expression systems, the increased dynamic range of the assay revealed that BCRP-HEK293 was clearly superior to the other expression systems studied.

IC_{50} values showed good concordance too with 94.4 % (255 out of 270) of values within the 3-fold range in pairwise comparisons and only 4 of 270 (1.5 %) showing more than 4-fold difference. Especially good correlation was observed between IC_{50} values determined in the two expression systems with human cellular origin (BCRP-HEK293 and BCRP-M), as eltrombopag was the only inhibitor showing greater than 3-fold difference (3.38-fold) when tested using topotecan as probe substrate. All inhibitors tested showed a measurable IC_{50} that correlated well between test systems for the same probe substrate, indicating that the VT assay is applicable across a broad chemistry and permeability space to determine IC_{50} values.

When IC_{50} values determined in BCRP-HEK293 using different probe substrates were correlated, a somewhat lower concordance was observed as only 81.4 % of pairwise comparisons showed less than 3-fold difference, and ivermectin and rapamycin showed around 9- to 15-fold differences (Table 5; Fig 6). BCRP has a large substrate binding region with multiple binding sites predicted from docking studies on homology models (Ferreira et al., 2017). Mutagenesis studies have also identified key amino acids such as R482 (Honjo et al., 2001) and F439 (Gose et al., 2020) that affect BCRP-mediated transport of many drugs. From the data available, binding sites for the substrates used in this study cannot be clearly delineated. However, the compounds showing the greatest differences such as ivermectin (Mw: 875), rapamycin (Mw: 914) and ritonavir (Mw: 720) are all large compounds, and therefore steric constrains may be behind their substrate-dependent inhibitory effect.

In summary, the kinetics data generated for 3 substrates and 15 inhibitors showed good correlation for K_m and IC_{50} values confirming applicability of the VT method for substrate assays using low passive permeable substrates, and inhibition assays in a broad permeability space. V_{max} data suggest that the dynamic range of the HEK293 based expression system (BCRP-HEK293) is superior to other expression systems.

Conflict of Interest

Z. S, J. M., A. K., E. K., P. K. are employees of Solvo Biotechnology, a Charles River Company that specializes in development of transporter technology applications. Solvo Biotechnology owns a patent on cholesterol-enriched insect cell membranes (WO2007132279A2), termed in this article as High Activity Membranes (HAM) and E. K. and P. K. are co-inventors on this patent. Outside of the submitted work P.K. is a consultant, co-founder and co-owner of Habilitas LLC (Budapest, Hungary), a company that provides ADME consultation services. No other authors declared any conflict of interest.

Acknowledgments

Help of Fanni Fülöp BSc in preparation of the manuscript and help of Joseph K Zolnerciks PhD with reviewing and language editing of the manuscript is acknowledged.

Funding

This work was supported by grant: GINOP-2.2.1-15-2016-00009.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2020.105593.

References

Bentz, J., O'Connor, M.P., Bednarczyk, D., Coleman, J., Lee, C., Palm, J., Pak, Y.A., Perloff, E.S., Reyner, E., Balimane, P., Brannstrom, M., Chu, X., Funk, C., Guo, A., Hanna, I., Heredi-Szabo, K., Hillgren, K., Li, L., Hollnack-Pusch, E., Jamei, M., Lin, X., Mason, A.K., Neuhoff, S., Patel, A., Podila, L., Plise, E., Rajaraman, G., Salphati, L., Sands, E., Taub, M.E., Taur, J.S., Weitz, D., Wortelboer, H.M., Xia, C.Q., Xiao, G., Yabut, J., Yamagata, T., Zhang, L., Ellens, H., 2013. Variability in P-glycoprotein inhibitory potency (IC(5)(0)) using various in vitro experimental systems: implications for universal digoxin drug-drug interaction risk assessment decision criteria. Drug Metabol. Disposition: Biol. Fate Chem. 41, 1347–1366.

Dawaliby, R., Trubbia, C., Delporte, C., Noyon, C., Ruysschaert, J.M., Van Antwerpen, P., Govaerts, C., 2016. Phosphatidylethanolamine is a key regulator of membrane fluidity in eukaryotic cells. J. Biol. Chem. 291, 3658–3667.

EMA, 2012. Guideline on the investigation of drug interactions. https://www.ema.europa.eu/documents/scientific-guideline/guideline-investigation-drug-interactions_en.pdf.

FDA, 2017. In Vitro Metabolism and Transporter Mediated Drug-Drug Interaction Studies Guidance for Industry. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM581965.pdf.

FDA, 2020. In vitro drug interaction studies — cytochrome P450 enzyme- and transporter-mediated drug interactions guidance for industry. https://www.fda.gov/media/134582/download.

Ferreira, R.J., Bonito, C.A., Cordeiro, M., Ferreira, M.U., Dos Santos, D., 2017. Structure-function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. Scientif. Reports 7, 15534.

Giri, N., Agarwal, S., Shaik, N., Pan, G., Chen, Y., Elmquist, W.F., 2009. Substrate-dependent breast cancer resistance protein (Bcrp1/Abcg2)-mediated interactions: consideration of multiple binding sites in in vitro assay design. Drug Metabol. Disposition: Biol. Fate Chem. 37, 560–570.

Glavinas, H., Kis, E., Pal, A., Kovacs, R., Jani, M., Vagi, E., Molnar, E., Bansaghi, S., Kele, Z., Janaky, T., Bathori, G., von Richter, O., Koomen, G.J., Krajcsi, P., 2007. ABCG2 (breast cancer resistance protein/mitoxantrone resistance-associated protein) ATPase assay: a useful tool to detect drug-transporter interactions. Drug Metabol. Disposition: Biol. Fate Chem. 35, 1533–1542.

Glavinas, H., Mehn, D., Jani, M., Oosterhuis, B., Heredi-Szabo, K., Krajcsi, P., 2008. Utilization of membrane vesicle preparations to study drug-ABC transporter interactions. Expert Opin. Drug Metab. Toxicol. 4, 721–732.

Gose, T., Shafi, T., Fukuda, Y., Das, S., Wang, Y., Allcock, A., Gavan McHarg, A., Lynch, J., Chen, T., Tamai, I., Shelat, A., Ford, R.C., Schuetz, J.D., 2020. ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket. FASEB J. 34, 4890–4903.

Heredi-Szabo, K., Palm, J.E., Andersson, T.B., Pal, A., Mehn, D., Fekete, Z., Beery, E., Jakab, K.T., Jani, M., Krajcsi, P., 2013. A P-gp vesicular transport inhibition assay - optimization and validation for drug-drug interaction testing. European journal of pharmaceutical sciences. Off. J. Eur. Feder. Pharmaceutical Sci. 49, 773–781.

Honjo, Y., Hrycyna, C.A., Yan, Q.W., Medina-Perez, W.Y., Robey, R.W., van de Laar, A., Litman, T., Dean, M., Bates, S.E., 2001. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. Cancer Res. 61, 6635–6639.

International Transporter Consortium,Giacomini, K.M., Huang, S.M., Tweedie, D.J., Benet, L.Z, Brouwer, K.L., Chu, X., Dahlin, A., Evers, R., Fischer, V., Hillgren, K.M., Hoffmaster, K.A., Ishikawa, T., Keppler, D., Kim, R.B., Lee, C.A., Niemi, M., Polli, J.W., Sugiyama, Y., Swaan, P.W., Ware, J.A., Wright, S.H., Yee, S.W., Zamek-Gliszczynski, M.J., Zhang, L., 2010. Membrane transporters in drug development. nature reviews. Drug Discov 9, 215–236.

Jani, M., Szabo, P., Kis, E., Molnar, E., Glavinas, H., Krajcsi, P., 2009. Kinetic characterization of sulfasalazine transport by human ATP-binding cassette G2. Biol. Pharm. Bull. 32, 497–499.

Klukovits, A., Krajcsi, P., 2015. Mechanisms and therapeutic potential of inhibiting drug efflux transporters. Expert Opin. Drug Metab. Toxicol. 11, 907–920.

Kruijtzer, C.M., Beijnen, J.H., Rosing, H., ten Bokkel Huinink, W.W., Schot, M., Jewell, R.C., Paul, E.M., Schellens, J.H., 2002. Increased oral bioavailability of topotecan in combination with the breast cancer resistance protein and P-glycoprotein inhibitor GF120918. J. Clin. Oncol. 20, 2943–2950.

Li, H., Jin, H.E., Kim, W., Han, Y.H., Kim, D.D., Chung, S.J., Shim, C.K., 2008. Involvement of P-glycoprotein, multidrug resistance protein 2 and breast cancer resistance protein in the transport of belotecan and topotecan in Caco-2 and MDCKII cells. Pharm. Res. 25, 2601–2612.

Lumen, A.A., Acharya, P., Polli, J.W., Ayrton, A., Ellens, H., Bentz, J., 2010. If the KI is defined by the free energy of binding to P-glycoprotein, which kinetic parameters define the IC50 for the Madin-Darby canine kidney II cell line overexpressing human multidrug resistance 1 confluent cell monolayer? Drug Metabol. Disposition: Biol. Fate Chem. 38, 260–269.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968.

Nakagawa, M., Schneider, E., Dixon, K.H., Horton, J., Kelley, K., Morrow, C., Cowan, K.H., 1992. Reduced intracellular drug accumulation in the absence of P-glycoprotein (mdr1) overexpression in mitoxantrone-resistant human MCF-7 breast cancer cells. Cancer Res. 52, 6175–6181.

Ozvegy, C., Varadi, A., Sarkadi, B., 2002. Characterization of drug transport, ATP hydrolysis, and nucleotide trapping by the human ABCG2 multidrug transporter. Modulation of substrate specificity by a point mutation. J. Biol. Chem. 277, 47980–47990.

Pal, A., Mehn, D., Molnar, E., Gedey, S., Meszaros, P., Nagy, T., Glavinas, H., Janaky, T., von Richter, O., Bathori, G., Szente, L., Krajcsi, P., 2007. Cholesterol potentiates ABCG2 activity in a heterologous expression system: improved in vitro model to study function of human ABCG2. J. Pharmacol. Exp. Ther. 321, 1085–1094.

Pedersen, J.M., Khan, E.K., Bergstrom, C.A.S., Palm, J., Hoogstraate, J., Artursson, P., 2017. Substrate and method dependent inhibition of three ABC-transporters (MDR1, BCRP, and MRP2). Eur. J. Pharm. Sci. 103, 70–76.

PMDA, 2018. Drug Interaction Guideline for Drug Development and Labelling Recommendations. https://www.mhlw.go.jp/hourei/doc/tsuchi/T180724I0150.pdf.

Poirier, A., Portmann, R., Cascais, A.C., Bader, U., Walter, I., Ullah, M., Funk, C., 2014. The need for human breast cancer resistance protein substrate and inhibition evaluation in drug discovery and development: why, when, and how? Drug Metabol. Disposition: Biol. Fate Chem. 42, 1466–1477.

Potriquet, J., Laohaviroj, M., Bethony, J.M., Mulvenna, J., 2017. A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry. PLoS One 12, e0175967.

Safar, Z., Kis, E., Erdo, F., Zolnerciks, J.K., Krajcsi, P., 2019. ABCG2/BCRP: variants, transporter interaction profile of substrates and inhibitors. Expert Opin. Drug Metab. Toxicol. 15, 313–328.

Sarkadi, B., Price, E.M., Boucher, R.C., Germann, U.A., Scarborough, G.A., 1992. Expression of the human multidrug resistance cDNA in insect cells generates a high activity drug-stimulated membrane ATPase. J. Biol. Chem. 267, 4854–4858.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement of protein using bicinchoninic acid. Anal. Biochem. 150, 76–85.

Storch, C.H., Ehehalt, R., Haefeli, W.E., Weiss, J., 2007. Localization of the human breast cancer resistance protein (BCRP/ABCG2) in lipid rafts/caveolae and modulation of its activity by cholesterol in vitro. J. Pharmacol. Exp. Ther. 323, 257–264.

Szeremy, P., Pal, A., Mehn, D., Toth, B., Fulop, F., Krajcsi, P., Heredi-Szabo, K., 2011. Comparison of 3 assay systems using a common probe substrate, calcein AM, for studying P-gp using a selected set of compounds. J. Biomol. Screen 16, 112–119.

Szilagyi, J.T., Vetrano, A.M., Laskin, J.D., Aleksunes, L.M., 2017. Localization of the placental BCRP/ABCG2 transporter to lipid rafts: Role for cholesterol in mediating efflux activity. Placenta 55, 29–36.

Tachibana, T., Kitamura, S., Kato, M., Mitsui, T., Shirasaka, Y., Yamashita, S., Sugiyama, Y., 2010. Model analysis of the concentration-dependent permeability of P-gp substrates. Pharm. Res. 27, 442–446.

Toth, B., Jani, M., Beery, E., Heslop, T., Bayliss, M., Kitteringham, N.R., Park, B.K., Weaver, R.J., Krajcsi, P., 2018. Human OATP1B1 (SLCO1B1) transports sulfated bile acids and bile salts with particular efficiency. Toxicol. Vitro 52, 189–194.

Xia, C.Q., Liu, N., Yang, D., Miwa, G., Gan, L.S., 2005. Expression, localization, and functional characteristics of breast cancer resistance protein in Caco-2 cells. Drug Metabol. Disposition: Biol. Fate Chem. 33, 637–643.