SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Cross-linking mass spectrometry on P-glycoprotein

by Gabriella Gellen

Supervisor: Zsolt Bacso, Ph.D

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2023

Cross-linking mass spectrometry on P-glycoprotein

By Gabriella Gellén, PharmD

Supervisor: Zsolt Bacsó, PhD

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

Head of the **Defense Committee**:	István Balogh, PhD, DSc
Reviewers:	László Drahos, PhD János András Mótyán, PhD
Members of the Defense Committee:	Tamás Janáky, PhD, DSc András Penyige, PhD

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen at 1PM 26th of January, 2024.

1. Introduction

According to the World Health Organization (WHO), around 17 % of all deaths in 2020 was caused by some form of cancer. The majority of therapies are chemotherapeutic agents or combinations of them, often combined with surgical procedures and radiotherapy, which are often used successfully. Traditional cytostatic drugs alone can be effective, especially in hematological malignancies (leukemias, lymphomas). However, cancer cells are not only able to select and become resistant through the body's natural defense mechanisms but can also fight cytostatic treatments. In a process known as multidrug resistance (MDR), the mutant cells become insensitive to many different types of chemotherapeutic agents.
Transmembrane ATP-binding cassette (ABC) transporters are major contributors to resistance, of which the P-glycoprotein (Pgp) encoded by the *mdr1* gene, also known as ABCB1, MDR1 protein, as well as ABCC1 or MRP1 (multidrug resistance-associated protein 1) and ABCG2 or BCRP (breast cancer resistance protein) are the most studied. Pgp clearly plays a role in the development of multifactorial multidrug resistance by contributing to the clearance of hydrophobic and amphipathic chemotherapeutic agents. Pgp may also be a prognostic factor for certain tumor types, as its expression levels often correlate with disease stage. There has been intensive investigations on Pgp function, structure and the development of potential Pgp targeting agents, but no specific inhibitor has yet been identified that could be used therapeutically in MDR.

Studying the function, molecular structure and protein interactions of the membrane embedded Pgp is therefore essential for a better understanding and successful treatment of multidrug resistance. However, the experimental study of membrane proteins, including Pgp, is challenging, because their structure and function are highly influenced by the lipids in the membrane. Their isolation and purification are difficult, and may cause alterations in their structure, therefore their function may be lost due to changes in the lipid composition. It is no coincidence that the Protein Data Bank (PDB) contains information only about 2 % on transmembrane proteins, even though 25 % of protein coding genes encode some transmembrane protein. A better understanding of plasma membrane proteins could help to develop targeted therapies, identify tumor markers, and block signaling pathways and mechanisms involved in resistance mediated by membrane proteins. Nevertheless, it is important to note that 60 % of drugs used for non-cancer treatment target some membrane-associated protein. Thus, it is crucial to develop and set up experimental procedures that allow for a more feasible study of these proteins, optimally while maintaining the physiological membrane environment.

1.1. The structure and catalytic cycle of Pgp

The 170 kDa human Pgp, contains 12 transmembrane helices (TMH), which form 2 transmembrane domains (TMD), 6 extracellular loops (ECL). The TMDs bind both directly and indirectly to the ATP-binding domains (nucleotide-binding domains, NBDs). There is direct binding via the alpha helixes between TMH6 and NBD1 or TMH12 and NBD2. Indirect, non-covalent binding occurs along 4 intracellular coupling helices (ICH1-4) extending from the membrane towards the cytoplasm. The ICHs play a signaling role sending information about changes in the membrane to the NBDs and thus may influence the proper binding and hydrolysis of ATP. The ATP-binding pockets (cassettes) (nucleotide-binding sites, NBS) are formed by intracellular loops and helices of the two opposite NBDs. The ATP-binding pocket is formed between the Walker regions, Q-loop and A-loop, on one NBD and the D-loop and C-loop (signature motif) on the other NBD, by intercalation of a water molecule and Mg^{2+}. The sequences of these specific ATP binding regions are evolutionarily conserved. The two homologous halves of the protein are linked by a disordered linker region, which is flexible enough to readily establish contact with protein partners and to form a link between the two

NBDs during the catalytic cycle of Pgp. Due to its flexibility, the linker region is often missing from protein structure models and its different conformations have been controversial so far.

In the inner cavity of the TMDs, up to approximately the height of the inner plate of the membrane bilayer, a substrate-binding pocket is formed with the participation of each of the TMHs, in which a structurally diverse set of compounds can bind at different sites. Mainly hydrophobic interactions are involved in the binding, but in some cases hydrogen bonds has also been described. Several studies have been published on the mechanism of the transport of substrates and the dynamics of structural changes of Pgp, but a unified molecular model has not yet been established. The reasons for this are the difficulties in crystallizing large transmembrane proteins and the changes in their spatial structure during their isolation. What we know is that substrate binding leads to increased ATP hydrolysis and substrate translocation occurs upon hydrolysis; that the ICHs and the conserved nucleotide-binding motifs are required for dimerization of NBDs; that dimerization occurs upon ATP binding; and that the nucleotide-binding pocket is open after ATP hydrolysis. The one-way transport mechanism involves structural signaling between NBDs, ICHs and the membrane. In contrast to previous results, which investigated the function of Pgp in its native environment, recent studies under physiological conditions show that a single functional NBS is sufficient for ATPase activity and substrate transport.

Based on 3D structural models, Pgp adopts two main conformations, the substrate-binding inward-facing (IF) conformation and the substrate-releasing outward-facing (OF) conformation. A further intermediate dominant structure, the pre-hydrolytic (occluded) structure, suggests that Pgp is closed in both outward- and inward-facing directions. In addition to the main conformations, several smaller and rather slowly evolving structural shifts have been observed, mainly influenced by the membrane and its lipid composition.

1.2. The effects of the membrane environment on Pgp's structure and function

The proper function and structure of integral membrane proteins is clearly determined by the composition and physical state of the membrane. Membrane proteins, and thus also in the case of Pgp, the lipid ring surrounding the hydrophobic transmembrane region (annular lipids) directly influence the proper function of the protein. Removal of lipids by detergents inhibits the ATPase activity of Pgp. In contrast, the presence of cholesterol, which stabilizes the lipid bilayer, is not essential for Pgp activity, but modifies its basic and substrate-stimulated ATPase activity, drug binding and transport capacity, and flippase functions.

Membrane microdomains (lipid rafts and a subtype of rafts, the caveolae) which are typically rich in sphingolipids and cholesterol, play important roles in, for example, signal transduction, transport of membrane proteins and substrates. Pgp is present in both raft and non-raft membrane microdomains, and its distribution may depend on the cell type. Pgp is less abundant in caveolae and is likely to interact only indirectly with the 'hairpin-like' caveolin that stabilizes the caveolae. Cytoskeletal proteins play an important role in the assembly and stabilization of lipid rafts and may also act as a link between membrane proteins, influencing their function. The functions and substrate specificity of Pgp may vary in different membrane microdomains. Furthermore, our laboratory has previously found that the substrate-binding, IF Pgp conformer is more abundant in lipid rafts.

Cholesterol can affect the structure and function of Pgp not only indirectly, by altering the physicochemical properties of the membrane, but also directly, by binding to Pgp. In the case of the ABCG2 protein, cholesterol binds to cholesterol-binding domains (Cholesterol Recognition Amino Acid Consensus domains: CRAC, CRAC-like, CARC, and CARC-like) with specific sequence patterns. Such sequences can also be found in the transmembrane helices and extracellular regions of Pgp, to most of which cholesterol binding has been proved by cryo-EM experiments.

In MDR cells, the proportion of raft-associated Pgp is about 22-40 % of the total Pgp amount, depending on the cell type. Besides, the number of raft- and caveolae-associated proteins and lipids is increased in MDR in general. This may be because chemotherapeutic agents used in tumor therapy can in many cases alter the lipid composition of the membrane. Pgp may stabilize membrane microdomains through its lipid flippase function, although its cholesterol transporter function looks unlikely. Our group found that, when plasma membrane cholesterol levels are decreased, the selective transport of raft-conformer Pgp to the plasma membrane is increased in MDR cells compared to control cells. The results confirm that Pgp together with its attaching cholesterols, through cellular trafficking processes (exocytosis of Golgi-bound vesicles, endosomes and lysosomes), help to promote membrane repair mechanisms, thus forming a more resistant phenotype in MDR.

1.3. Pgp-recognizing antibodies

The structure of Pgp and the orientation of ECLs are often investigated with extracellular monoclonal antibodies (mAbs), which allow the monitoring of structural changes of Pgp *in vivo* under physiological conditions. Generally, antibodies can label all cell surface Pgps, regardless of its major conformational changes. However, the conformation sensitive UIC2 IgG_{2a} mAb favors the ATP-deficient, substrate-binding IF conformation. The discontinuous epitope of UIC2 is composed of regions at ECL 1, 3 and 4, which are closely aligned in the IF conformation. In the absence of a substrate, UIC2 labels about 10-40 % of cell surface Pgp (Pool 1), but upon substrate binding, all remaining Pgps are also labelled (Pool 1 + Pool 2). This so-called UIC2 shift phenomenon is caused by the increasing number of available binding sites, which can increase up to 10-fold.

The binding affinity of the 15D3 IgG_1 mAb, which labels both Pool 1 and Pool 2 cell surface Pgp, is reduced by cholesterol extraction with β-cyclodextrins. This is because the structural cholesterols closely associated with Pgp are required to maintain the orientation of ECLs that the 15D3 antibody recognizes. The epitope of the 15D3 antibody and the cholesterol-sensitive ECLs have not been previously determined. It is known that 15D3 and UIC2 are competitive with each other, so their epitopes are presumed to overlap partially. Nevertheless, UIC2 binding is not affected by cholesterol extraction and our previous experiments suggest that the binding site of 15D3 is also a conformational epitope, i.e. not a continuous peptide sequence. Thus, 15D3 can bind to Pgp throughout the whole catalytic cycle, whereas UIC2 prefers the IF structure when ECL 1 and 4 are close together. Structural data suggest that ECL 1 and ECL 6 remain close to each other throughout the catalytic cycle, also, ECL 1 partially overlaps with the epitope of the UIC2 antibody, raising the possibility that 15D3 binds to these regions.

1.4. Structural proteomics methods

Protein crystallography is the most common method used to study the 3D structure of proteins, however, with these conventional methods it is difficult or impossible to investigate membrane proteins, intrinsically disordered proteins (IDPs) or proteins containing intrinsically disordered regions (IDRs). This is because membrane proteins are amphipathic and require a physiological membrane environment to maintain their original structure. Their purification and analysis in this way is challenging. IDP and IDR proteins are not crystallizable and therefore X-ray diffraction analysis is not possible. Because of their flexibility, the structure of these proteins can be characterized as conformational ensemble, and their analysis by cryo-electron microscopy (cryo-EM) and NMR spectroscopy is difficult, also, disordered regions are often missing from the structural data. Theoretical calculations and predictions play an increasingly important role in describing the structure and movement of proteins. The neural network-based machine learning algorithm AlphaFold (DeepMind) allows reliable prediction of the structure of protein types (such as membrane proteins) that were not present in the database used for training. However, the analysis of flexible regions with these algorithms is

still limited and more experimentally defined data is needed to generate appropriate 3D structures. Such information can be obtained, for example, using different methods of structural proteomics.

Over the last 30 years, the continuous development of mass spectrometry (MS) has opened more and more possibilities to study the higher order structure of proteins by this method. Mass spectrometry analysis does not require freezing, or crystallization of the protein sample and the size of the protein does not usually limit the analysis. Using MS, major conformational ensembles can be separated from each other, so that not only the ensemble average can be investigated, but also transitions between conformational states and protein movement can be followed, and the binding sites and binding dynamics of different ligands, lipids or protein partners can be described by mass spectrometry. Mass spectrometry based structural methods are excellent complementors to traditional structural biochemistry methods, providing insights into protein properties that cannot be detected by conventional methods.

1.4.1. Cross-linking mass spectrometry

In recent years, the XL-MS technique has become a widely established complementary method to conventional structural biology techniques, especially for difficult-to-analyze proteins such as membrane proteins or IDP and IDR proteins.

The cross-linking reaction typically takes place under physiological conditions *in vitro* in isolated protein solutions, gels, or immunoprecipitation beads, but can also be performed *in vivo* in living cells and tissues. Cross-linking molecules form a covalent bond between two amino acid side chains in close proximity to each other. Then, after this fixation step, the formed bond remains stable even during stringent sample preparation steps and ionization in the mass spectrometer. Chemical cross-linking agents usually contain two reactive groups separated by a spacer of defined length. The length of the spacer provides information on the distance constraint between the side chains of specific amino acids within a protein or between different protein-protein partners. Amino acids cross-linked on the polypeptide chain of a protein are then typically determined by bottom-up proteomic approaches, i.e. derivatization (denaturation, reduction, alkylation) and enzymatic digestion followed by LC/ESI-MS/MS.

The targets of cross-linking agents can be, for example, sulfhydryl (cysteine), carboxyl (aspartic acid, glutamic acid) or amino (lysine) groups. The most commonly used cross-linkers are amine-reactive N-hydroxysuccinimide (NHS) esters, which primarily target the ε-nitrogen of lysine side chains. To a lesser extent, they can also react with serine (12.5 % of the total reactive groups), tyrosine (4.3 %) and threonine (3 %). In our experience, however, NHS reagents reacted with these residues only to a negligible extent. Additionally, according to the recommendations of the XL-MS community it is more advantageous to compromise considering only lysines as the reactive site of NHS esters in order to avoid false positive identifications on a complex dataset.

The efficiency of the cross-linking reaction is about 1-5 % and the amount of lysine side chains spaced at a suitable distance from each other is limited. Thus, a significant proportion of the sample contains unmodified peptides and other products of the cross-linking reaction. A large number of the modified products are the mono-links which are formed when only one end of the cross-linker reacts with a lysine residue, the other end is hydrolyzed in the aqueous medium or reacted with a quenching buffer (e.g. Tris, glycine). Mono-links can provide useful information about the hydrophobic regions of the protein and lysines which are the most readily reactive. Intra-peptide cross-links (loop-link) within a peptide are less common and are typically formed in α-helix or loops. Cross-links are also distinguished whether they are formed within a protein (intraprotein) or between two different proteins (interprotein).

There are several commercially available NHS ester cross-linkers of different solubilities and lengths, which can provide complementary information about the protein

structure. One of the most common water-soluble cross-linkers is BS2G (bis(sulfosuccinimidyl)glutarate, 7.7 Å), a non-cleavable cross-linker because it fragments only at higher voltages when fragmented in the gas phase (CID). For such cross-linking, the MS/MS spectrum contains fragments of both cross-linked peptides plus the linker, thus increasing the detection space quadratically (n^2 problem), significantly increasing the computational burden of the evaluation and the probability of misidentifications. In order to avoid a quadratic increase in the search space (n), cleavable crosslinks have been developed which give characteristic peaks in the spectra, when fragmented at lower voltages than those required to cleave the peptide backbone. For example, the most used DSSO (disuccinimidyl sulfoxide) when fragmented in MS2 gives 2 intense signature ion pairs with a mass shift of 32 Da. In a suitable mass spectrometer it is possible to select one peptide from the pair based on the signature ions, and then further fragment them in MS3 to obtain fragments of only one peptide within a spectrum (n search space). For non-cleavable crosslinks, deuterium labeling is used for unambiguous crosslink identification. Thus, a 1:1 ratio of labeled to unlabeled cross-linkers gives a characteristic mass shift in the MS1 and MS2 mass spectra.

In the case of transmembrane proteins, difficulties arise in labeling hydrophobic amino acids in the intramembrane region and the residues sterically embedded by the membrane. XL-MS technique has been successfully applied mostly on those membrane proteins, that have large extramembrane domains. In vivo cross-linking reactions generate a highly complex data set, making the identification of the not so abundant cross-linked species challenging. Novel structural proteomics approaches targeting membrane proteins are under extensive development, however, no standard approach has yet been established for specific cell surface cross-linking.

2. Aims of the study

It is known that the structure and function of P-glycoprotein is strongly influenced by the membrane microenvironment, with lipids and cholesterols directly bound to the transmembrane protein. Our group has previously shown that the monoclonal antibody 15D3 binds to the extracellular region of P-glycoprotein in a cholesterol-dependent manner, but its exact epitope is still unknown. In order to investigate the structure of P-glycoprotein and determine the epitope of 15D3 while preserving the physiological membrane microenvironment, different cross-linking mass spectrometry (XL-MS) based experimental approaches were used.

1. Our goal was to optimize robust sample preparation methods that can be used to study membrane proteins in a reproducible manner while preserving the natural membrane environment.

2. Optimize settings of the mass spectrometer we use and to set up reliable and efficient computational data analysis to identify the most modified peptides.

3. We wanted to use our optimized methods to map the dynamic structure of P-glycoprotein and regions that are difficult to study using conventional structural biology methods.

4. We aimed to establish a theoretical tertiary structure of 15D3 antibody based on its variable region's amino acid sequence.

5. Using theoretical protein docking and cross-linking mass spectrometry results, we aimed to determine the epitope of the 15D3 monoclonal antibody on the extracellular regions of P-glycoprotein.

6. In addition, we aimed to identify the direct and indirect protein partners of P-glycoprotein in 3T3 cells and to compare the protein lists obtained by enrichment with 15D3 and UIC2 antibodies.

3. Materials and methods

3.1. Cell Culture

Human *Mdr1* gene-transduced 3T3 mouse fibroblast cells (NIH 3T3 MDR1, a gift from M. Gottesman, stably expressing a high level of cell surface human P-glycoprotein in mouse cells) were cultured at 37 °C in a humidified gas incubator containing 5 % CO_2. Cells were subdivided every second day in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 25 µg/mL gentamicin, and 670 nM doxorubicin. Two days before the experiments, the culture medium was replaced by a doxorubicin-free culture medium. Cells were regularly tested for mycoplasma.

3.2. Antibody Purification

UIC2 (IgG2a; ATCC No. HB-11027) and 15D3 (IgG1; ATCC No. HB-11342) anti-ABCB1 mAbs were prepared from supernatants of hybridoma cells. Supernatants were enriched and purified using Protein G affinity chromatography. Hybridoma cell lines were obtained from American Type Tissue Culture Collections (Manassas, VA, USA). Briefly, 15D3 hybridoma cells were grown on a macrophage feeder layer from BALB/c mouse peritoneal fluid. Briefly, 15D3 hybridoma cells (1 million cells/mouse) were intra-peritoneally injected into male BALB/c or SCID Swiss nude mice aged 7–11 weeks old using 0.2 mL IFA/mouse (incomplete Freund's adjuvant). Mice were kept in the Animal Core Facility at the University of Debrecen, housed separately in cages, had ad libitum access to water and chow, and were kept in a 12 h light/dark cycle with a controlled temperature of 22 ± 1 °C. The isolated 15D3 mAb protein concentration was measured at 280 nm using NanoDrop One^C microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Budapest, Hungary). If not frozen and stored at −20 °C, Na-azide (from 2% in 100× dilution) was added to the purified antibody and stored at 4 °C.

3.3. Living Cell Cross-Linking Reactions

3.3.1. BS2Gd_0/d_4 (Bis(Sulfosuccinimidyl) Glutarate-d0/d4)

NIH 3T3 MDR1 cells were grown in T75 flasks in complete media. Briefly, ~90 % confluency cells were trypsinized (0.05 % trypsin and 0.02 % EDTA in PBS, pH 7.4) for 2 min, and trypsinization was stopped with complete media. Cells were washed twice with serum-free media supplemented with 0.1 % molecular-biology-grade BSA to prepare for PNGase F treatment. Furthermore, ~8×10^7 cells/mL in a bacterial Petri dish were treated with 100 U/mL PNGase F (New England Biolabs, Ipswich, UK) diluted with serum-free media supplemented with 0.1 % molecular-biology-grade BSA. Cells were incubated for 4 h at 37 °C in a humidified gas incubator containing 5 % CO_2. Cells were collected in 15 mL centrifuge tubes and washed twice with glucose–PBS (phosphate-buffered saline extended with 8 mM glucose). For extracellular antibody cross-linking samples, 10^7 cells/mL were incubated with 100 µg/mL 15D3 or UIC2 for 30 min at 37 °C with mild agitation. After washing twice with glucose-PBS, at RT (room temperature), 3×10^7 cells per sample were resuspended in PBS, pH 7.0, and BS2Gd_0/d_4 (Thermo Fisher Scientific) was added in 1:1 ratio for a final concentration of 5mM dissolved in PBS, pH 7.0, in a final volume of 1 mL. The cross-linking reaction was performed for 60 min at 37 °C, and it was quenched for 5 min at RT with 1 M Tris, pH 7.5, reaching a 20 mM final concentration in the reaction mixture.

3.3.2. DSSO (Disuccinimidyl Sulfoxide)

Living cells were treated with PNGase F as described above and 3×10^7 cells per sample were resuspended in PBS, pH 7.0.; 50 mM DSSO (Thermo Fisher Scientific) stock solution was prepared in DMSO and further diluted with the samples to reach 0.5 mM, 0.5 mM, 1 mM, 5 mM, and 10 mM final concentrations in a final volume of 1 mL. The cross-linking reaction was performed for 60 min at 37 °C, and then samples were treated as described above.

3.4. On-Bead Cross-Linking

Reductive methylation of Protein G was performed with sodium cyanoborohydride ($NaBH_3CN$) on Dynabeads prior to immunoprecipitation and cross-linking. Briefly, 500 µL of beads were washed three times with PBS, pH 6, then 0.1 M paraformaldehyde (with no methanol), and finally, 0.1 M $NaBH_3CN$ was added. The mixture was incubated for 30 min, at RT with mild agitation. Beads were washed with PBS, pH 7.4, three times, then placed in a new tube. Membrane preparation and immunoprecipitation with 15D3 or UIC2 antibody were performed as described below. Subsequent to washing with RIPA buffer, beads were washed with PBS, pH 7.0, twice, and the second wash was placed in a new tube. Furthermore, 1 mM DSSO was added to the samples, and had been previously dissolved in DMSO with a 50 mM stock concentration. The cross-linking reaction was carried out for 60 min, at RT with mild agitation, and it was quenched for 5 min with 1 M ammonium hydrogen carbonate (AMBIC), pH 7.5, reaching a 20 mM final concentration in the reaction mixture. Then, beads were washed twice with 25 mM AMBIC (for safety and environmental considerations, these experiments were carried out under a fume hood, and all supernatants and remaining solvents containing $NaBH_3CN$ were neutralized with sodium hypochlorite.)

3.5. Membrane Preparation and Immunoprecipitation Enrichment

Membrane preparation was performed in accordance with the manual of Thermo Fisher Scientific, for Mem-PER™ Plus Membrane Protein Extraction Kit (#89842). For immunoprecipitation, 100 µL of Dynabeads Protein G (Thermo Fisher Scientific) was washed four times in PBS supplemented with 5 mg/mL BSA (PBS/BSA). Cell lysates with 500 µg of protein of a 1 mg/mL concentration were added to the beads and rotated at 4 °C overnight. The following day, the beads were washed nine times in 1 mL of RIPA buffer at 4 °C. Then, the beads were washed twice in 1 mL of freshly made ice-cold 100 mM AMBIC, and the second wash was placed in a new tube. Beads were frozen at this point and stored at −80 °C until trypsin digestion and LC-MS analysis.

3.6. Flow Cytometry

Briefly, 500,000 cells per sample, previously treated with different concentrations of PNGase F (0, 50, 100, and 200 U/mL) as described below, were placed into sorter tubes in 500 µL of glucose-PBS. Cells were labeled with primary antibodies, 15D3 or UIC2, at a concentration of 10 µg/mL for 30 min at 37 °C. After 3 washes in PBS, cells were labeled with 1 µg/mL Alexa-488 conjugated anti-mouse goat secondary antibody (Thermo Fisher Scientific) for 1 h on ice. Cells were washed 2 times in PBS and resuspended in 100 µL of ice-cold 4 % PFA. Cells were analyzed using a NovoCyte flow cytometer (ACEA Biosciences, San Diego, CA, USA). In each sample, 20,000 cells were collected. Cells were gated for single cells according to the FSC-H/SSC-H and SSC-H/SSC-A plots, and then homogenous populations according to the FSC-H/FL1-H dot plot were selected. The FL1-H signal of the Alexa-488 dye was plotted on overlayed histograms. Data analysis and graphs were made using FCS Express version 6 (De Novo Software, Glendale, CA, USA).

3.7. Immunoblot Analysis

Cell lysates (7 µg protein/sample) were diluted in SDS sample buffer (0.31 M Tris-HCl, pH 6.8, 50 % glycerol, 10 % SDS, 100 mM DTT, 0.01 % bromophenol blue, and 1 M β-mercaptoethanol) and incubated at 65 °C for 10 min with shaking. Proteins were separated via electrophoresis on 7 % SDS-polyacrylamide gel in Laemmli buffer and electro-transferred in Towbin buffer onto a nitrocellulose membrane (Bio-Rad). The membrane was blocked with 5 % non-fat dried milk in PBS for 60 min, at RT. Pgp was labeled with D3H1Q (Cell Signaling Technology) human anti-Pgp rabbit monoclonal primary antibody (1:1000 diluted in 5 % non-fat dried milk in PBS) for 60 min, at RT with continuous agitation. Non-bound antibodies were

washed with PBS containing 0.1 % Tween-20 3 times for 10 min, at RT with continuous shaking. They were then labeled with anti-rabbit goat IgG secondary antibody conjugated with horseradish peroxidase (diluted in a ratio of 1:5000 in 5 % non-fat dried milk in PBS) for 60 min, at RT with continuous shaking. The unbound antibodies were washed 3 times, then the immunoblots were developed with SuperSignal West Femto ECL reagent (Thermo Fisher Scientific) and images were recorded using the Chemidoc imaging system (Bio-Rad Hungary Ltd., Budapest, Hungary).

3.8. Trypsin Digestion and LC-MS Analysis

Beads were resuspended in 25 mM AMBIC, and protein disulfide bonds were reduced with 1 mM tris(2-carboxyethyl)phosphine (TCEP) for 10 min at room temperature. Resulting sulfhydryls were derivatized with 2.4 mM S-methyl methanethiosulfonate (MMTS) for 10 min at room temperature. Protein digestion with side-chain-protected porcine trypsin (Promega Corporation, Madison, USA) proceeded overnight at 37 °C; then, it was terminated by acidifying the samples to obtain a final 0.2 % trifluoroacetic acid (TFA). The digests were analyzed using a LC-MS/MS on an Orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific) mass spectrometer on-line coupled to an ACQUITY UPLC M-Class system (Waters Corporation, Milford, UK) or an Evosep One HPLC system (Evosep). In the ACQUITY UPLC M-Class setup, samples were loaded onto a Symmetry C18 Trap column (100 Å; 5 µm; 180 µm × 20 mm; 2D; V/M) with a flow rate of 5 µL/min for 5 min with 1 % solvent B, then separated on a Peptide BEH 130 C18 column (130 Å, 1.7 µm, 100 µm × 100 mm) or a Peptide CSH C18 column (130 Å, 1.7 µm, 75 µm × 250 mm) at a flow rate of 400 nL/min with a gradient of solvent B from 5 to 10 % in 2 min, from 10 to 35 % in 50 min, from 35 to 50 % in 15 min, and then up to 90 % in 8 min keeping the column temperature at 45 °C. In the Evosep One setup (Evosep Biosystems, Odense, Denmark), samples were loaded onto Evotip Pure tips and separated using the Extended 15 SPD (samples per day) method on EV-1106 analytical column (C18; 1.9 µm; 150 µm × 150 mm). In both setups, solvent A was 0.1 % formic acid in water, and solvent B was 0.1 % formic acid in acetonitrile. MS data acquisition was performed in a data-dependent fashion on multiply charged precursors. Non-cross-linked and BS2G-cross-linked samples were analyzed via MS2-HCD fragmentation only. In DSSO cross-linked samples, targeted MS3-HCD spectra were collected whenever the DSSO fragmentation-specific mass difference was detected in MS2-CID, and here, the data acquisition was complemented with MS2-EThcD and MS2-HCD fragmentation as well. MS1 and MS2-CID data were measured in the orbitrap with a resolution of 60k and 30k, respectively, and MS2-HCD data were measured either in the orbitrap with a resolution of 15k or in the ion trap with a rapid scan rate, whereas MS3-HCD and MS2-EThcD data were measured in the ion trap with a rapid scan rate.

3.9. Data Processing and Interpretation

Peak lists from raw data were generated using Proteome Discoverer (v1.4 or v2.4) and then submitted for a database search via ProteinProspector (v5.24.0, v6.2.1, or v6.3.1). As a first stage, data were searched against mouse entries in the Uniprot 2017.11.01 database (83,889 entries) or the SwissProt 2021.06.18 database (17089 entries) to identify proteins present in the sample. Sequences of human Pgp, Protein G, the known stretches of UIC2 and 15D3 antibodies, and common contaminants (e.g., human keratins or bovine serum proteins) were also considered. Precursor mass tolerance was set to 5 ppm and fragment mass tolerance was set to 20 ppm (orbitrap data) or 0.6 Da (ion trap data). Only fully tryptic peptides were taken into consideration, with a maximum of two missed cleavages. Methylthio modification of Cys residues was used as a fixed modification and Met oxidation, acetylation of protein N-termini, pyroglutamic acid formation of peptide N-terminal Gln residues, and hydrolyzed DSSO or $BS2Gd_0/d_4$ derivatives on lysine residues were used as variable modifications. Protein hits were accepted with a protein/peptide FDR of < 1 % and at least two unique peptide identifications

per protein. For the semiquantitative analysis of identified proteins, SPCs were normalized to all PSMs in the given analysis. In the second stage, proteins confidently identified with at least 2 or 5 unique peptides (cross-linked in living cells or on-bead, respectively) were considered to identify cross-links between lysine residues, either with DSSO or BS2G, using Protein Prospector or the XlinkX node in Proteome Discoverer (v3.0.1.27). In the DSSO cross-linked samples, the alkene, unsaturated thiol, and sulfenic acid derivatives (resulting upon fragmentation) were considered for MS3-HCD data as variable modifications on uncleaved lysine residues. Eventually, MS3 data were paired with their respective MS2 data, and cross-links were also manually evaluated for Pgp.

3.10. Molecular Docking Analysis

The 15D3 antibody structure was modeled using ABodyBuilder2 (RMSD for CDR-H3 of 2.81 Å) the web application using the 15D3 variable heavy- and light-chain sequence from United States Patent US5849877. RMS prediction errors below the 5 Å threshold were computed using the deep learning algorithm of ImmuneBuilder.

Protein–protein docking was performed using the ClusPro 2.0 web server in the Antibody mode and with automatic non-CDR masking. The PDB structure of 15D3 previously created with ABodyBuilder2 and the human MDR1 PDB structure 6qex was uploaded together with a masking PDB file created in PyMOL for using only extracellular loops of human MDR1 for docking. The best fit, cluster 0, with 177 members, was used for further interpretations via PyMOL.

4. Results

4.1. Development of sample preparation methods for XL-MS analysis of transmembrane proteins

Two different sample preparation approaches were used for our mass spectrometry studies. (i) Live 3T3 MDR cells were treated with cross-linking agents and then, after membrane preparation, immunoprecipitation enrichment with monoclonal antibodies against Pgp was performed on Protein G magnetic beads. (ii) 3T3 MDR cells were first subjected to membrane preparation and then Pgp was enriched by immunoprecipitation. Subsequently, cross-linking was performed on Protein G purified Pgp and its tightly coupled protein partners.

Our sample preparation methods were originally based on the Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) protocol steps. According to this protocol, a strong RIPA buffer wash is applied after cross-linking of live cells and immunoprecipitation enrichment of proteins, allowing us to remove several non-specific binding contaminants prior to mass spectrometry measurement. In our protocol, we complemented the RIME method, which originally focuses on the isolation of nuclear proteins, with membrane preparation. After trying several different membrane preparation methods, the Thermo Fisher Scientific Mem-PER™ Plus membrane protein extraction kit proved to be the most suitable for the preparation of large amounts of membrane proteins in a reproducible manner. This method allows the isolation of proteins with one or more transmembrane domains and their associated peripheral and cytosolic proteins. The solubilization buffer containing zwitterionic detergent retains some of the lipid molecules around Pgp, so that the protein can retain ATPase activity.

4.1.1. Optimisation of the cross-linking reaction conditions and verification of the efficiency of immunoprecipitation

In order to identify as many cross-links as possible and to minimize artefact formation, several cross-linking concentrations and cross-linking times were tested and verified by Western blot analysis. As the concentration of cross-linking agents and reaction time were increased, the Pgp signal gradually appeared in the higher (250 kDa) molecular weight ranges. Simultaneously, the signal of 15D3 mAb gradually disappears, presumably because it is cross-linked into the complex. We found that above 5 mM concentration of cross-linkers, the number of identified modifications decreases due to excessive cross-linking of proteins. Thus, for our living cell experiments, we have chosen the concentration of 5 mM for the future. For the on-bead DSSO experiments, a concentration of 1 mM was used, since in this case Pgp and its protein partners are more directly reached by the cross-linker molecules, thus excessive cross-linking can occur at lower concentrations.

The efficiency of immunoprecipitation enrichments was followed by Western blotting. It was possible to label less Pgp when enriching with 15D3 compared to the enrichment with UIC2. This is probably due to the cholesterol sensitivity of 15D3, as the level of physiologically occurring cholesterol in the cell membrane may be slightly reduced after membrane preparation.

4.1.2. Effect of PNGase-F treatment on cell surface binding of 15D3 and UIC2 monoclonal antibodies

For live cell cross-linking, we included PNGase-F (Peptide-N-Glycosidase F) treatment prior to incubation with the antibody, as we found in our flow cytometry measurements that it improves the binding affinity of 15D3 and UIC2 monoclonal antibodies. The first extracellular loop on Pgp (ECL1) has three N-glycosylation sites (asparagines 91-, 94-, and 99), whose N-glycan pattern can vary widely depending on cell line and pathotype. PNGase-F cleaves between the innermost GlcNAc (N-acetylglucosamine) and asparagine on N-glycosylated

proteins. The N-glycans of Pgp presumably sterically inhibit antibody docking, so removal of N-glycans slightly increased cell surface binding of 15D3 and UIC2 monoclonal antibodies. Previous experiments suggest that removal of glycans does not affect the proper function of Pgp.

As a consequence of glycosylation during the separation of Pgp by SDS-PAGE, the migration of Pgp in the electrophoretic field is altered and often gives a diffuse band in Western blot analysis. Deglycosilation altered Pgp's migration in the electrophoretic field as detected via Western blot analysis. The small shift in the electrophoretic mobility of the Pgp appeared at molecular weights of approximately 140 kDa, which is consistent with previously published data. Western blot analysis, in addition to flow cytometry measurements, demonstrated that PNGase-F treatment did indeed cleave N-glycans on Pgp.

These results suggest that PNGase-F treatment may also be beneficial for cell surface labeling of other extracellularly glycosylated membrane proteins.

4.1.3. Reductive methylation of protein G beads

We have applied our on-bead cross-linking sample preparation approach in order to cross-link Pgp more directly and efficiently. However, in our initial attempts, we found that the cross-links formed on Protein G were much more abundant than the cross-links identified on the antibody or the enriched Pgp. Thus, due to the high intensity of the identifications of Protein G, the detection of cross-links on Pgp and antibodies by mass spectrometry became essentially impossible. With the methylation of Protein G beads, the immunoglobulin G-binding protein (Protein G), previously identified with 1704 spectral counts (SPCs), was reduced to 13 SPCs. The number of spectra associated with DSSO cross-linked peptide identifications (cross-link spectra match, CSM) was reduced from 61 to 0. Moreover, the number of identified spectra of antibodies binding to Protein G did not decrease after methylation, so the modification presumably did not significantly affect the binding affinity of the antibody to Protein G. In this way, the mass spectrometric analysis of the proteins we investigated was simplified and made more efficient.

4.2. Identified mono-links

Mono-links are formed in much higher concentrations than cross-links and can be used to map hydrophobic and hydrophilic segments of the protein, similar to hydrogen-deuterium exchange (HDX), thus helping to map the protein's spatial structure.

A total of 28 unique mono-links were identified on the surface of Protein G beads cross-linked with DSSO, while 19 unique mono-links were detected after living cell DSSO treatment and 9 unique mono-links were detected after $BS2G_{d0/d4}$ treatment. On 15D3 and UIC2 antibodies, 10-10 mono-links were identified using the on-bead sample preparation method. The lysines K271-, K380-, K515-, K808-, K915- and K1150- were detected by all three methods, suggesting that these lysine side chains are the most readily available in polar solvent. We also identified mono-links on the Walker A regions on the two NBDs on lysines K433 and K1076 using the on-bead DSSO cross-linking method, and the mono-link on K1076 was also found using the living cell DSSO method. We also identified a mono-link and cross-links on lysine K536 located directly adjacent to the C loop.

4.3. Identified cross-links

The two cross-linkers we use have different hydrophobicity and spacer arm lengths. While DSSO has a length of 7.7 Å and is a more hydrophobic and membrane permeable compound, BS2G has a length of 10.3 Å and is easily water soluble. These, and the fragmentation differences mentioned in the introduction, may contribute to the different results obtained with different sample preparation methods.

The WESA online program can be used to predict the water accessible amino acids in a given protein. Thus, it was found that out of a total of 85 lysines in Pgp, 70 are readily water accessible on the protein surface and 15 are embedded in the tertiary structure of the protein. Among the lysines exposed on the surface, 34 (49 %) could be chemically modified by cross-linkers, and of the more embedded lysines, cross-linking or mono-linking was detected on 5 (33 %) lysines. Thus, 39 (46 %) of the total lysines reacted with a cross-linker in all sample preparation procedures. The majority of the occluded lysines were located close to the membrane both intra- and extracellularly, and the transmembrane region is inaccessible to water. The labelling of the less water accessible lysines near the membrane in the inner cavity, such as lysine K826, suggests that cross-linking molecules may enter the interior of the protein in a way similar to Pgp substrates, forming covalent modifications there.

The majority of crosslinks were identified in the on-bead crosslinked samples with 10 unique crosslinks, while 4 unique crosslinks were detected in the living cell DSSO crosslinked samples and none in the $BS2G_{d0/d4}$-treated samples. Extracellularly, 1 cross-link was identified on the constant heavy chain of the UIC2 antibody. All of the crosslinks found in living cell samples were also present in the on-bead samples, except one, between lysines K279 and K786. One of the most frequently identified crosslinks was between lysines K1220 and K1150 near the D-loop, the latter very often carrying a mono-link modification as well. Differences in sample preparation procedures may cause variation in the structure of Pgp, and thus there may be differences in cross-linking patterns.

No cross-links were identified unambiguously between Pgp and the antibodies, and none of them formed interprotein cross-links with protein interaction partners. Furthermore, we did not identify interprotein cross-linking between different Pgp molecules, which is confirmed by our Western blot studies, where no dimer formation was identified. Thus, in the present work, all of the discussed cross-links refer to intraprotein cross-links.

The amount of cross-linking identified in our experiments is of the same order of magnitude as similar cross-linking investigations with NHS esters performed by other research groups. Membrane proteins being difficult to handle, the number of identified cross-links in the present study is considered efficient enough and some functional conclusions could be drawn.

The distance between cross-linked lysines was accepted between 5 Å and 30 Å based on recommendations from previous publications. In this way, we could fit all crosslinks to the PDB structure of Pgp 6qex. Since lysines are highly flexible, and disordered regions (IDRs) can often be found on proteins in general, a more permissive interval for the length of the bond created by the cross-linkers may be necessary and relevant.

IDR loops in proteins are often difficult to retain during sample preparation by conventional structural biology methods. Cross-linking mass spectrometry can provide more structural information about these regions, as cross-linking occurs in a neutral pH buffer solution close to physiological conditions. In our experiments we found several mono-links and cross-links on disordered regions of Pgp. Mono-links were identified at the N-terminal portion of the elbow helix on lysine K31 and on lysines K645 and K685 of the linker region. These regions are all missing from the 6qex cryo-EM structure in the Pgp PDB database. Cross-linking was identified between lysines K685 of the linker region and K536 next to the C-loop.

In our experiments, we found cross-links between lysines K826 and K786 in the inner cavity and between K826 and K808 on the ICH3 loop, the latter located directly above NBD2 and in close structural and functional association with it. In addition, lysine K808 was also found to be cross-linked with lysine K536 adjacent to the C-loop on NBD1, which is also in close proximity to the Walker A region on NBD1, on which we identified a mono-link modification too. The ICH4 helix is located directly above NBD1, and not far from it in the *N*-terminal direction, lysine K895 was cross-linked to K536. Also, a mono-link on K915 directly adjacent to ICH4 in the *C*-terminal direction was identified. Thus, the C-loop on NBD1 is in close

proximity to ICH4 and the opposite ICH3, which are directly linked to extracellular loops ECL5 and ECL6. These cross-linked regions are in close proximity to the predicted binding site of the 15D3 antibody, and in close proximity to cholesterols binding to Pgp in transmembrane regions.

4.4. Docking of 15D3 monoclonal antibody to the extracellular loops of Pgp

The observed cross-links related to the cholesterol-affected NBD connections and the lack of detecting unambiguous cross-links between the ECLs of Pgp and the applied mAbs forced us to model these protein interactions. Since no cross-links were identified between Pgp and the antibodies, this information could not be used for molecular docking.

The structure of the 15D3 mAb variable region was estimated from its amino acid sequence (Patent US 5849877) using ABodyBuilder2 software. This estimated structure was aligned to the structure of human Pgp 6qex PDB on the ClusPro 2.0 web server. The alignment revealed that the complementarity-determining regions (CDRs) of the 15D3 antibody on substrate-bound, inward-facing Pgp are located primarily in proximity to the first and sixth extracellular loops (ECL1 and ECL6). In comparison, the UIC2 antibody docks to the first, third and fourth extracellular loops (ECL1, ECL3, ECL4), thus approaching Pgp from the opposite direction.

That is, the cholesterol-dependent 15D3 approaches the extracellular part of Pgp from ECL5 and ECL6, also reaching ECL1. Cross-links and mono-links were identified between intracellular regions connected to these loops, which were described in previous chapters.

4.5. Identification of protein partners of P-glycoprotein by cross-linking and affinity purification MS

Several known protein partners of Pgp were identified by mass spectrometry in our immunoprecipitation experiments.

Although cross-links between Pgp and other proteins were not identified, since intense washes of the magnetic beads with RIPA buffer were applied after enrichment, it is assumed that not many contaminants and weak interaction partners remained on the beads. This is evidenced by the fact that bovine serum albumin (BSA), with which protein G magnetic beads were blocked prior to immunoprecipitation, was detected with relatively low SPC after washing.

Our experiments were performed on NIH 3T3 mouse fibroblast cells transfected with the human *mdr1* gene, suggesting that the known partners we identified are linked along conserved sequences that are identical in the human and mouse proteomes.

We identified an order of magnitude more proteins in the on-bead cross-linked samples compared to living cell samples, as the interference of protein G with the bead samples was reduced by reductive methylation.

For our bead samples, 368 proteins were identified by mass spectrometry with enrichment with the 15D3 antibody and 222 proteins with UIC2 with at least 10 SPC. Of these, 208 proteins were identified in both types of samples. The 15D3 antibody, because it binds to Pgp independently of the catalytic cycle, can probably enrich more protein partners even though UIC2 could enrich about 7 times more Pgp based on the SPCs identified. This may be because of the reduced binding affinity of the 15D3 antibody to the cholesterol-depleted Pgp after membrane preparation. In contrast, UIC2 was unaffected by the change in cholesterol levels after membrane preparation, but only recognizes the IF, substrate binding conformation, therefore only proteins that bind to this conformation could be purified in this way.

Based on the protein lists, network analysis was performed using the STRING database and Cytoscape software. Networks are assembled based on protein-protein interactions verified using experimental and predictive data, or estimated using data and text mining techniques, visualizing the presence or absence of a link. The network nodes are the proteins present in our list, and the edges between nodes are the knowledge of protein-protein interactions.

Interactome networks of human orthologs of mouse proteins identified in the samples enriched with 15D3 and UIC2 cross-linked on-bead with DSSO were analyzed using the STRING database and Cytoscape software. The analysis showed that about 2% of the hits were not connected to the network. For enrichment with the 15D3 antibody, 8 proteins were distinct, and 5 proteins were distinct for UIC2.

In addition to network visualization, Cytoscape is also suitable for functional and cell compartment distribution analysis, for which different applications are available. ClueGO program was used to group given protein lists according to their gene ontology (GO) classification (The Gene Ontology Consortium 2013) and intracellular component. The software lists the GO cellular compartment names corresponding to the statistically significantly abundant proteins that are most enriched in the groups. Optionally it displays the percentage of genes (proteins) assigned to the groups relative to the total number of proteins queried and grouped. A protein can belong to more than one GO cell compartment grouping, so since our database was relatively large, fusion of the groups was set in order to reduce redundancy for the evaluation.

5. Discussion

The structure and function of the integral membrane protein Pgp is strongly influenced by the lipids and cholesterols in the membrane environment. Cholesterol binds directly to Pgp not only in the vicinity of lipid rafts and caveolae, but also influences its intracellular and extracellular spatial structure.

The main aim of our experiments was to determine the epitope of the 15D3 monoclonal antibody that binds to the extracellular regions of Pgp in a cholesterol-dependent manner. This could shed light on which extracellular regions are affected by cholesterol. XL-MS technique was chosen for our studies. This method allowed us to study Pgp while preserving its physiological membrane microenvironment. After optimization of our methods, we followed two main experimental approaches: (i) Living cells were treated with DSSO or $BS2G_{d0/d4}$ cross-linkers, followed by membrane preparation using a zwitterionic detergent and immunoprecipitation of the cross-linked complexes using 15D3 or UIC2 antibodies. (ii) Membrane preparation was first performed on the cells and then Pgp was purified by immunoprecipitation using 15D3 or UIC2 antibodies. Then, after stringent washing of the beads, the cross-linking reaction was performed on the beads. The effect of physiologically occurring cholesterol is present during the living cell cross-linking during fixation. However, during on-bead cross-linking, this naturally occurring cholesterol may be reduced. In the membrane preparation procedure we used, some of the cholesterol binding to Pgp is lost, but since immunoprecipitation worked with the 15D3 antibody, even if at a reduced level compared to UIC2, there is presumably still some cholesterol remained in the samples connected to Pgp.

Based on the modifications identified in our experiments, the two methods overlap by about 40 %, confirming that both methods may be suitable for investigating the structure of Pgp. We identified the same number of unique mono-links in the on-bead samples as in the living cell samples (DSSO and BS2Gd0/d4 treatments combined), but 60% more cross-links in the on-bead prepared samples. In the case of living cell samples, some of the cross-linking molecules are hydrolyzed until they enter the cell, or that they cannot access the lysines near the membrane surface. In contrast, in the case of on-bead samples, the cross-linker reaches Pgp more directly and can form a covalent bond between lysine side chains in close proximity to each other.

The lipid composition of the samples prepared by our methods (live cell and on-bead) is significantly different. In the case of living cell cross-linking, the cells are fixated by cross-linking after removal of the cell culture medium, so that the lipid composition of the membrane proteins remains essentially unchanged during the cross-linking reaction. During the on-bead sample preparation the hydrophobic membrane proteins were first isolated by applying a mild zwitterionic detergent and then, later, were cross-linked. In this case, the lipid profile of the proteins involved may change slightly; e.g., they can lose cholesterol, but at the same time, the immediate advantage is that the amount of the non-membrane proteins in the preparation is significantly reduced. Thus, by applying the two approaches, we can compare the effect of membrane protein structural changes caused by the reduced cholesterol amount or the slightly altered lipid environment with the membrane protein structure that can be expected in an entirely physiological lipid environment under normal cholesterol conditions.

Mono-link modifications, by providing information on lysines available to water, can help to study structural changes in Pgp while maintaining physiological conditions. Intensive research is ongoing to understand the structure and function of Walker A regions, where mono-links have been found in both living cell and on-bead samples. Most of the mono-linked lysines we identified have been previously described to be deuterated during HDX, supporting that mono-links do indeed provide information on water accessibility. Moreover, we have also identified modified lysine residues (K411, K550, K1093, K1099 and K1212), which were previously found to be cholesterol-sensitive in HDX. Furthermore, the mono-linked lysine

K1002, which is linked to ECL6, is located close to a CRAC-like domain where binding cholesterol was indeed determined by cryo-EM. Although HDX and cryo-EM techniques have provided insight into the structure of Pgp, these methods do not provide information on the connection between spatial arrangements of different residues, also, flexible regions are usually completely absent, so no information on their dynamics is available. In addition, these experiments are often performed on recombinant mouse Pgp instead of human Pgp, because it is more difficult to isolate human Pgp in a way that preserves its original folding. Even if these recombinant proteins are incorporated into lipid nanodiscs, they only provide an approximate model of the physiological lipid and cholesterol environment. The methods we use are much more robust than those mentioned above, so they are easy to follow and can be performed relatively quickly, making our results more reproducible. In this way, by maintaining a near-physiological medium, we can identify modified residues that act as beacons to indicate points of structural change in response to modulation of the lipid environment.

The cross-links we detected are located on the intracellular side of Pgp, many of them on regions forming the ATP-binding pockets and on the linker peptide, which are difficult to study. The cross-linking between lysines K536 and K685 may complement studies on the linker's orientation, since predictions of this region have been controversial so far. The linker region is known to contain several conserved phosphorylation sites, to bind with high affinity to tubulin, and is likely to have a role in the catalytic cycle. Since the lysine K536 is located directly adjacent to the C-loop on NBS2, this cross-link may support the idea that the linker is associated with NBDs and regulates ATP hydrolysis and substrate specificity.

Molecular dynamics calculations and IM-MS experiments suggest that lysine K826 interacts electrostatically with the phosphate group of phospholipids. In our experiments, lysine K826 was cross-linked with lysine K786 and lysine K808 on ICH3. By HDX measurements on nanodisc-embedded Pgp, it was found that the ICH3 resting on NBD2 is protected from deuterium exchange in the presence of cholesterol. Lysine K808 was also found to be cross-linked to K536 next to the C-loop on NBD1, which is also close to the Walker A region, and cross-linked to K895 adjacent to ICH4. The Walker A region and ICH4 are known to interact in the presence of lipids. The ICH4 helix, which rests on NBD1, is directly connected to ECL5 and ECL6 via TMH10 and TMH11, and Clouser et al. showed that ICH4 is more easily deuterated in the presence of cholesterol, compared to ICH3, which is more difficult to access. The identified cross-links on these regions of Pgp are in parallel with previous findings demonstrating that cholesterol and membrane composition has a complex effect of the intracellular structure of Pgp and its proper function. Surrounded by cholesterols, ECL5 and ECL6 bind directly to ICH3 and ICH4 helices through transmembrane regions, which shift asymmetrically in the presence of cholesterol. This raises the possibility that the spatial orientation of ECL5 and ECL6 is also altered by cholesterol.

Interestingly, K1220 was cross-linked to lysine K1150 near the cholesterol-sensitive D-loop, also, K1150 was the most readily available for the cross-linkers we used. All the cross-linked lysines identified in our experiments have been previously described as deuteratable in HDX experiments, suggesting that these lysines are indeed accessible to cross-linkers dissolved in aqueous buffer.

In the living cell experiments, the native lipid environment remains intact during the cross-linking reaction, but the cross-linking reagents hydrolyze sooner than they reach a lysine on the protein of interest. Thus, less information can be extracted in this case. The on-bead cross-linking method retains less of the native environment, although the zwitterionic detergent that was used retains some the lipid molecules around the Pgp, so it does not lose its ATPase activity. The similarity of the mono-link and cross-link identifications obtained by the two methods

suggests that the on-bead cross-linking method is equally suitable for membrane protein structure analysis.

Our experiments provide insight into how results obtained by crystallography and mass spectrometry methods (XL-MS, HDX-MS, IM-MS) complement each other even for complex data generated by processing membrane proteins.

Molecular docking predicts that the 15D3 antibody binds near ECL1 and ECL6. 15D3 connects to Pgp independently of structural changes in the catalytic cycle, so 15D3 is likely to bind to extracellular loops that remain close to each other throughout the cycle. Indeed, the ECL1 and ECL6 loops are close to each other and are not affected by the major structural changes in Pgp. Furthermore, Pgp and UIC2 are known to bind to partially overlapping epitopes, so if one of the binding sites of 15D3 is on ECL1, this would explain the competition between the two antibodies.

Only two of the mono-links identified in living cell experiments fall within the linker-ICH3-ICH4 regions. These mono-links were also found in the on-bead cross-linked samples and in this case three additional mono-links (K645, K685, K626) and five cross-links were identified in these regions. These differences suggest that rearrangements may have occurred in the structure of Pgp prepared by the two methods. The more frequent identifications of mono-links and cross-links on the linker-ICH3-ICH4 regions in the cholesterol-depleted on-bead samples may indicate that these regions rearrange in a cholesterol-sensitive manner. As described in previous HDX-MS experiments, cholesterol-induced compaction and decompaction occurs along the ICH3 and ICH4 regions. Distribution of mono-links in other areas of Pgp was not substantially rearranged in living cell or on-bead preparations, so it may be an internal control of our observation, as a more intense random behavior of mono-links compared to that of cross-links is to be expected.

This suggests that a vertical rearrangement occurs in the absence of cholesterol, which exactly affects the linker, ICH3 and ICH4 cytoplasmic sequences, attached directly to the ECL4, ECL5 and ECL6 through rigid transmembrane helices. This rearrangement can pull the ECL5 or ECL6 loops vertically into the plane of the membrane, while in the presence of cholesterol, the same loops can protrude more from the membrane. The ECL4 extracellular loop is longer than the ECL5 and ECL6 one, and thus its relative motion may be less significant. Cholesterol may exert its effect through cholesterol-binding motifs, concentrated in excess in the nearby ECL4, ECL5 and ECL6, through a mechanism similar to that of surface tension. In cholesterol abundance, cholesterol can crawl up the “wall” of the cholesterol-binding motifs containing protein, which can deform the horizontal plane of the transporter protein in the level of the membrane. This mechanism is a possible explanation for the cholesterol-dependent behavior of the 15D3 antibody.

A link between the ECL6 loop and the ICH-NBD interfaces has been suggested by molecular dynamics calculations, as it was found that a mutation (F978A) in the modulator binding site (M-site) near ECL6 reduces ICH3-NBD2 contacts but increases ICH4-NBD1 interactions.

Several known protein partners of Pgp were identified using affinity purification MS. The high abundance of cytoskeletal proteins identified supports the idea that Pgp is strongly cytoskeleton associated. Actin, tubulins, and filamin A have previously been described to possess a direct or indirect connection with Pgp. AHNAK, which was identified in all samples with a high number of SPCs, had been shown to have altered expression levels in drug-resistant cells where Pgp was overexpressed.

Moreover, heat shock proteins Hsp90 and Hsp70, which were found in all our approaches, are also known to be associated with MDR and Pgp. These chaperones are associated with

membrane lipid rafts and cholesterol, and crosstalk between Hsp90 and cholesterol might foster the activity of Pgp.

Among the 160 uniquely identified proteins of 15D3 enrichment, many belong to the ubiquitin system or caveolae.

Different ubiquitin E3 ligases were primarily identified in samples enriched with 15D3 mAb. The most abundant among them was NEDD4 which promotes the ubiquitination and degradation of Pgp. Whether or not there is a reason why E3 ligases are more definite in the case of pull-down by the cholesterol-sensitive 15D3 antibody has not been investigated; however, it has been demonstrated that increased cholesterol inhibits the degradation of ABC transporters ABCA1 and ABCG1 via the ubiquitin–proteasome pathway. Additionally, RING finger protein 2, another E3 ubiquitin ligase, was identified with both 15D3 and UIC2 mAbs, which was reported to be in direct connection with Pgp's linker peptide.

In addition, caveolae-associated protein 1 (Cavin-1) identification was also more notable in samples enriched with 15D3 than in those enriched with UIC2. Cavin-1 stabilizes caveolae in a cholesterol-dependent manner by connecting to Caveolin-1 through its scaffolding region in the lateral side of the plasma membrane, enriched with cholesterol and acidic lipids. Caveolae reside in the cholesterol-rich lipid rafts of the plasma membrane, and the increased number of these specific membrane regions promote more aggressive tumor cell growth and MDR. Cavin-1 has an increased expression level in MDR along with Caveolin-1 and Pgp, which was claimed to be necessary for fortifying lipid rafts and MDR. A decent number of Pgps are localized in membrane lipid rafts and in certain cancer cell types; a physical interaction between Pgp and Caveolin-1 and regulation of Pgp by Caveolin-1 was proven. Hinrichs et al. claimed that in other cell types, Pgp and Caveolin-1 had different solubility behaviors in Triton X-100, and Pgp does not dwell in caveolae. Based on our experiments, it can be hypothesized that Pgp has a stronger connection with Cavin-1 and cytoskeletal proteins, and through these connections, Pgp could be regulated by Caveolin-1. Since the binding affinity of the 15D3 antibody is dependent on the presence of cholesterol in the plasma membrane, the enrichment of Pgp by this antibody can pull down more cholesterol-sensitive interactions.

6. Summary

The structure and protein interaction partners of Pgp were investigated by XL-MS, applying cross-linkers on living cells or on immunoaffinity magnetic beads. To optimize our methods, we supplemented previously published techniques and protocols with cell surface PNGase F treatment, membrane preparation, and methylation of immunoprecipitation beads. Our novel experimental procedures allowed us to draw the following conclusions:

- The mono-links we identified indicate the solvent accessibility of Pgp and the lysines that are more likely to bind to protein partners.

- The cross-linking between the C-loop on NBD1 and the linker supports the regulatory effect of the linker on ATP hydrolysis.

- The identified cross-links suggest that the C-loop is in close proximity to ICH4 and ICH3 at the opposite side of Pgp, which are known to have cholesterol-sensitive orientations. ICH3 and ICH4 are directly linked to ECL5 and ECL6, which bind cholesterols nearby.

- Our studies suggest that the linker-ICH3-ICH4 regions shift in a cholesterol-depleted environment, consistent with previous results describing asymmetric compaction and decompaction along the ICH3 and ICH4 regions when cholesterol levels were increased.

- Based on our results, we hypothesize a vertical rearrangement in response to cholesterol depletion, which may pull ECL5 and ECL6 in towards the membrane, whereas in the presence of cholesterol these loops may protrude from the membrane.

- The cholesterol-sensitive 15D3 mAb binds to ECL1, ECL5 and ECL6 based on our molecular docking results, which is consistent with the cholesterol-dependent conformational changes in these extracellular loops hypothesized from our cross-linking results.

- The compartmental distribution analysis of the protein partners we identified suggests a strong cytoskeleton association of Pgp.

- We identified higher amounts of E3 ubiquitin ligases and cavin-1 in 15D3-enriched samples.

- Our experiments suggest that Pgp interacts with the protein cavin-1, which is specific for lipid rafts.

7. Publications

UNIVERSITY of DEBRECEN

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

Registry number: DEENK/366/2023.PL
Subject: PhD Publication List

Candidate: Gabriella Gellén
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology
MTMT ID: 10079791

List of publications related to the dissertation

1. Gutay-Tóth, Z., **Gellén, G.**, Doan-Xuan, Q. M., Eliason, J. F., Vincze, J., Szente, L., Fenyvesi, F., Goda, K., Vecsernyés, M., Szabó, G., Bacsó, Z.: Cholesterol-Depletion-Induced Membrane Repair Carries a Raft Conformer of P-Glycoprotein to the Cell Surface, Indicating Enhanced Cholesterol Trafficking in MDR Cells, Which Makes Them Resistant to Cholesterol Modifications.
   Int. J. Mol. Sci. 24 (15), 1-23, 2023.
   DOI: http://dx.doi.org/10.3390/ijms241512335
   IF: 5.6 (2022)
2. **Gellén, G.**, Klement, É., Biwott, K., Schlosser, G., Kalló, G., Csősz, É., Medzihradszky-Fölkl, K., Bacsó, Z.: Cross-Linking Mass Spectrometry on P-Glycoprotein.
   Int. J. Mol. Sci. 24 (13), 1-24, 2023.
   DOI: http://dx.doi.org/10.3390/ijms241310627
   IF: 5.6 (2022)

List of other publications

3. Nagy, K., **Gellén, G.**, Papp, D., Schlosser, G., Révész, Á.: Optimum collision energies for proteomics: The impact of ion mobility separation.
   J. Mass Spectrom. Epub, 1-12, 2023.
   DOI: http://dx.doi.org/10.1002/jms.4957
   IF: 2.3 (2022)
4. Mammadova, R., Maggio, S., Fiume, I., Bokka, R., Moubarak, M., **Gellén, G.**, Schlosser, G., Adamo, G., Bongiovanni, A., Trepiccione, F., Guescini, M., Pócsfalvi, G.: Protein Biocargo and Anti-Inflammatory Effect of Tomato Fruit-Derived Nanovesicles Separated by Density Gradient Ultracentrifugation and Loaded with Curcumin.
   Pharmaceutics. 15 (2), 1-21, 2023.
   DOI: http://dx.doi.org/10.3390/pharmaceutics15020333
   IF: 5.4 (2022)

5. Sárvári, É., **Gellén, G.**, Sági-Kazár, M., Schlosser, G., Solymosi, K., Solti, Á.: Qualitative and quantitative evaluation of thylakoid complexes separated by Blue Native PAGE.
Plant Methods. 18 (23), 1-19, 2022.
DOI: http://dx.doi.org/https://doi.org/10.1186/s13007-022-00858-2
IF: 5.1

6. Mammadova, R., Fiume, I., Bokka, R., Kralj-Iglič, V., Božič, D., Kisovec, M., Podobnik, M., Zavec, A. B., Hočevar, M., **Gellén, G.**, Schlosser, G., Pócsfalvi, G.: Identification of Tomato Infecting Viruses That Co-Isolate with Nanovesicles Using a Combined Proteomics and Electron-Microscopic Approach.
Nanomaterials. 11 (8), 1-19, 2021.
DOI: http://dx.doi.org/10.3390/nano11081922
IF: 5.719

7. Bulyáki, É., Kun, J., Molnár, T., Papp, A., Micsonai, A., Vadászi, H., Márialigeti, B., Kovács, A. I., **Gellén, G.**, Yamaguchi, K., Lin, Y., So, M., Józsi, M., Schlosser, G., Lee, Y. H., Liliom, K., Goto, Y., Kardos, J.: Pathogenic D76N Variant of beta(2)-Microglobulin: Synergy of Diverse Effects in Both the Native and Amyloid States.
Biology (Basel). 10 (11), 1-22, 2021.
DOI: http://dx.doi.org/10.3390/biology10111197
IF: 5.168

Total IF of journals (all publications): 34,887
Total IF of journals (publications related to the dissertation): 11,2

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

03 August, 2023