TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

UNVEILING EARLY PROTEOMIC EVENTS IN B CELL RECEPTOR ACTIVATION

Luqman Awoniyi

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1972 | MEDICA – ODONTOLOGICA | TURKU 2026

University of Turku

Faculty of Medicine
Institute of Biomedicine
Pathology and Immunology
Turku Doctoral Programme of Molecular Medicine (TuDMM)

Supervised by

Docent, Pieta Mattila, PhD
Lymphocyte Cytoskeleton Group Institute of Biomedicine
University of Turku
Turku, Finland

Reviewed by

Professor, Maria Vartiainen
Institute of Biotechnology.
University of Helsinki
Helsinki, Finland

Professor, Yolanda R. Carrasco
Centro Nacional de Biotecnologia
CNB)-CSIC
Madrid, spain

Opponent

Professor, Tuula Nyman
Department of Immunology
Institute of Clinical Medicine
University of Oslo
Oslo, Norway

ISBN 978-952-02-0693-2 (PRINT)
ISBN 978-952-02-0694-9 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Grano Oy, Vantaa, Finland, 2026

Dedicated to my beloved parents, my wife, and our children, whose unwavering support and encouragement made this journey possible. I am also deeply grateful to my supervisor for her guidance and support.

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Pathology
AWONIYI LUQMAN: Unveiling Early Proteomic Events in B cells Receptor Activation
Doctoral Dissertation, 131 pp.
Turku Doctoral Programme of Molecular Medicine (TuDMM)
June 2026

ABSTRACT

B cells are an indispensable component of the adaptive immune system, essential for effective protection against pathogens. B cells rely on B cell receptor (BCR) to interact with antigens and to trigger various biological processes and signaling pathways. Defects in BCR signaling can lead to diseases such as immunodeficiency, cancer, and autoimmune disorders.

While the roles of BCR signaling in antigen processing, internalization, and presentation, antibody production and B cell survival are well documented, the mechanisms by which BCR activation controls these multiple biological processes remain poorly understood.

In this study, we employed three distinct approaches to gain unprecedented insight into the identity, dynamics and environment of BCR signaling proteins, as well as the multifaceted biological processes or pathways that are concomitantly activated in response to BCR stimulation.

Firstly, we utilized proximity biotinylation by ascorbate peroxidase 2 (APEX2) to obtain high-coverage mass spectrometric data on the identity and dynamics of proteins in the vicinity of the BCR upon activation. This method leverages the efficient promiscuous biotinylation activity of the APEX2 enzyme, which enables kinetic assessment of the local proteome with high spatiotemporal resolution, within approximately 20 nm and a 1-min time window.

Secondly, we developed a new in silico pipeline, AutoCoEV, for analyzing large protein datasets and to predict inter-protein co-evolution. This pipeline was piloted to explore potential functional interactions or co-dependencies among selected lipid raft resident proteins identified in the APEX2 study.

Lastly, we developed a methodology to achieve a comprehensive overview of the proteome at the immunological synapse (IS) in primary mouse B cells. This technique uses magnetic beads to emulate antigen-presenting cells, which initiated the formation of the IS, allowing efficient isolation of bead-associated cell adhesions.

The results presented in this thesis provide a valuable reference for the early proteomic events following BCR activation, enhancing our understanding of the complex signaling mechanisms that govern B cell function.

KEYWORDS: B cell, B cell receptor, APEX2, mass spectrometry, lipid rafts, SUMOylation, protein coevolution, immunological synapse.

TURUN YLIOPISTO
Lääketieteellinen tiedekunta - Biolääketieteen laitos
Patologia
LUQMAN AWONIYI: B-solureseptorin aikaisen signaloinnin proteomiikkataso karakterisointi
Väitöskirja, 131 s.
Turun Molekyylilääketieteen tohtoriohjelma (TuDMM)
Kesäkuu 2026

TIIVISTELMÄ

B-solut ovat olennainen osa opittua immuunijärjestelmää muodostaen vasta-aineita. B-solut tunnistavat spesifisiä antigeenejä B-solureseptorinsa (BCR) välityksellä. Tämä tunnistus käynnistää monimuotoisen signaalinvälityksen ja solunsisäisen vasteen soluissa, johon kuuluu muun muassa välitön antigeenin internalisaatio ja antigeenin peptidi-esittelyn käynnistäminen lisäsignaalien vastaanottamiseksi. Näiden prosessien seurauksena B-solut erilaistuvat vasta-aineita tuottaviksi plasma soluiksi tai muistisoluiksi. BCR-signalointi on tarkoin säädeltyä ja sen häiriintyminen voi johtaa useisiin eri tautitiloihin, kuten immuunipuutoksiin, syöpään tai autoimmuunisairauksiin. Vaikka BCR-signaloinnin tärkeä rooli B-solujen vasteessa on kiistaton, ne molekyylitason mekanismit, jotka säätelevät BCR-signaloinnin soluvastetta, ovat edelleen puutteellisesti ymmärrettyjä.

Tässä työssä käytimme kolmea erilaista lähestymistapaa tutkiaksemme antigeenin välittämää BCR-signalointia. Pystytimme kaksi proteomiikka-pohjaista menetelmää BCR-signalointiin vastaavien proteiinien identifioimiseksi, sekä tietokonepohjaisen, automatisoidun, proteiinin ko-evoluutioon pohjautuvan interaktio-analyysin. Hyödynsimme solukalvolle kohdennetun askorbaattiperoksidaasi 2 entsyymi (APEX2) aikaansaamaa epäspesifistä välittömän lähistön proteiinien biotinylaatiota (proximity biotinylation) elävissä soluissa BCR signaloinnin alkuvaiheissa, yhden minuutin aikajänteellä. Massaspektrometrinen biotinyloitujen proteiinien analyysi valaisi antigeenin tunnistamisen aiheuttamia proteiinitason muutoksia solukalvolla. Toisessa osatyössä kehitimme uuden tietokoneavusteisen analyysimenetelmän, AutoCoEv, joka ennustaa proteiinien välisiä interaktioita, niiden koevoluutioon perustuen, suurista proteiiniaineistoista. Lopuksi kehitimme menetelmän, jolla pystyimme eristämään B-solujen immunologisia synapseja proteomiikka-analyysiä varten. Tämä hiiren primäärisiin B-soluihin sovellettu menetelmä mahdollisti B-solujen immunologisen synapsin laajan proteiinisisällön kartoittamisen.

Tässä väitöskirjassa esitetyt tulokset tarjoavat uusia menetelmiä sekä arvokkaita aineistoja BCR-signaloinnin ja sitä seuraavien solubiologisten tapahtumien parempaan ymmärtämiseen. Tulokset syventävät ymmärrystämme B-soluaktivaation monimutkaisesta säätelystä ja sitä kautta B-solujen toiminnasta osana immuunijärjestelmäämme.

AVAINSANAT: B-solu, B-solureseptori, APEX2, massaspektrometria, lipidilautat, SUMOylaatio, proteiinien koevoluutio, immunologinen synapsi.

Table of Contents

Abbreviations **8**

List of Original Publications **10**

1 Introduction **11**

2 Review of the Literature **13**
2.1 B cell and its receptor 13
2.1.1 Functions and structure of the BCR complex 14
2.1.2 BCR signaling and regulation 16
2.1.3 BCR-mediated antigen endocytosis 17
2.1.4 B cell immunological synapse 18
2.2 The plasma membrane and its role in BCR signaling 18
2.2.1 Plasma membrane compartmentalization 18
2.2.2 Lipid raft domains 19
2.2.2.1 Role of lipid raft as BCR signaling domain 20
2.2.2.2 Fatty acylation 22
2.2.2.2.1 Protein myristoylation and palmitoylation 22
2.3 Proximity labeling techniques 24
2.3.1 Enzyme-initiated proximity labeling: concept and mechanism 25
2.3.2 APEX2 technique 26
2.3.3 Mechanism of action of APEX2 27
2.4 Quantitative mass spectrometry analysis 29
2.4.1 Label-free and labelled protein quantification 30
2.4.2 Analyzing and Interpreting Proteomics Data 31
2.4.2.1 Missing values and imputation 32
2.4.2.2 Data normalization 33
2.4.2.3 Batch correction 33
2.4.2.4 Differential enrichment analysis 34
2.5 Protein coevolution 35

3 THE AIMS OF THIS THESIS **37**

4 MATERIALS AND METHODS **38**

5 RESULTS AND DISCUSSION **47**

5.1 Identification of lipid raft associated regulators of BCR signaling (Study I). 47
5.1.1 Results: Proximity labeling technique 47
5.1.1.1 Assesment of the effects of H_2O_2 on the BCR sinalling cascade 48
5.1.1.2 Exploratory data analysis 50
5.1.1.3 BCR proteome identification and dynamics 51
5.1.1.3.1 BCR components were among top differentially enriched proteins. 51
5.1.1.3.2 BCR proteome identification 51
5.1.1.3.3 Characterization of Lipid Raft-Associated Proteins in B cells 52
5.1.1.4 Monitoring of pathway dynamics upon BCR activation 53
5.1.1.4.1 Dynamics of BCR signaling proteins in the lipid rafts following antigen stimulation 53
5.1.1.4.2 Dynamics of proteins involved in endocytosis 54
5.1.1.4.3 The role of SUMO1 in BCR signaling and regulation. 56
5.1.2 Discussion 58
5.2 A High-Throughput In Silico Pipeline for Predicting Inter-Protein Coevolution (study II) 61
5.2.1 AutoCoEv 61
5.2.2 Discussion 64
5.3 Proteomic analysis of the Immunological Synapse (Study III) 67
5.3.1 Results: Immunological Synapse Isolation 67
5.3.2 Discussion 69

6 Conclusion 71

References 73

Original Publications 85

Abbreviations

aa	amino acid
Ab	antibody
Ag	antigen
APC	antigen presenting cells
AP	affinity purification
APX	ascorbate peroxidase
APEX2	ascorbate peroxidase 2
BCR	B cell receptor
BioID	proximity-dependent biontin identification
BIS2	blocks in sequences
CAPS	coevolution analysis using protein sequences
DCA	direct coupling analysis
EGFR	epidermal growth factor receptor
F(ab')2	2 antigen binding fragment
GO	gene onthology
HEL	hen egg lysozyme
HRP	horseradish peroxidase
Ig	immunoglobulin
Igα	immunoglobulin alpha
Igβ	immunoglobulin beta
IgH	immunoglobulin heavy chain
Ighm	immunoglobulin heavy constant Mu
IgK-V	immunoglobulin kappa chain
IgL	immunoglobulin light chain
IgM	immunoglobulin M
ITAM	immunoreceptor tyrosine-based kinase
LC-MS	liquid chromatography-mass spectrometry
LIMMA	linear models for microarray data
IS	immunological synapse
MOA	mechanism of action

MSA	multiple sequence alignment
MS/MS	tandem mass spectrometry
MS	mass spectrometry
NUC	nuclear pore complex
NUP	nucleoporin
GOLGA 3	golgin subfamily A member 3
PCR	polymerase chain reaction
PI3K	phosphoinositide 3-kinase
PPI	protein protein interaction
PTM	post translational modification
PTK	protein tyrosine kinase
SL	surrogate light chain
SUMO	small ubiquitin-related modifier

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Awoniyi, L. O., Cunha, D. M., Sarapulov, A. V., Hernández-Pérez, S., Runsala, M., Tejeda-González, B., Šuštar, V., Balci, M. Ö., Petrov, P., & Mattila, P. K. (2023). B cell receptor-induced protein dynamics and the emerging role of SUMOylation revealed by proximity proteomics. Journal of Cell Science, 136(15), jcs261119. https://doi.org/10.1242/jcs.261119

II Petrov, P. B., Awoniyi, L. O., Šuštar, V., Balci, M. Ö., & Mattila, P. K. (2022). AutoCoEv—A High-Throughput In Silico Pipeline for Predicting Inter-Protein Coevolution. International Journal of Molecular Sciences, 23(6), Article 6. https://doi.org/10.3390/ijms23063351

III Cunha, D. M., Hernández-Pérez, S., Awoniyi, L. O., Carisey, A. F., Jacquemet, G., & Mattila, P. K. (2023). Proteomic profiling of isolated immune synapses from primary mouse B cells (p. 2023.02.23.529674). Preprint on bioRxiv. https://doi.org/10.1101/2023.02.23.529674

1 Introduction

Imagine catching the flu from a coworker during the cold season. Despite your best efforts to wash hands and staying warm, sometimes pathogens, the likes of virus, bacteria etc. bypass our basic defenses and invade the body. In this situation, the immune system kicks into action surveying your body to quickly respond not only to external threats but also internal ones, such as rogue cancerous cells. This constant surveillance and response system is critical to keeping us healthy.

The immune system is the core biological system for protecting individuals against pathogens. The immune system relies on two key components: the innate immune system and the adaptive immune system. The innate immune system provides the first level/immediate defense against pathogens. It is quick in response but short-lived. In contrast, the adaptive immune system provides a delayed but more robust and long-lasting defense against more complex, and persistent pathogens that need to be neutralized. The process of adaptive immune response is facilitated by specialized adaptive immune cells such as the T helper cells and B lymphocytes, and their products. Adaptive immune response involves regulated communication between T helper cells and B lymphocytes where T helper cells activate B cells. This activation facilitates B cells to produce pathogen-specific antibodies and differentiate into plasma B cells and memory B cells.

Antibodies are a soluble form of B cell receptor (BCR) that are secreted by plasma B cells. The biological process that leads to high affinity antibody production comprises critical stages that are strictly regulated. Impairment in these regulatory processes may in some instances lead to life threatening disease. One important step in the B cell differentiation towards antibody producing cells is the rearrangement of immunoglobulin gene locus. This ensures production of diverse antibodies that have ability to neutralize varieties of pathogens.

A critical component of B cells, central to their role in the adaptive immune response against pathogens, is the BCR. Understanding the interactions of a protein, such as the BCR with its binding partners, whether direct or indirect, stable or transient, as well as the complexes it forms and the dynamic expression levels of its interacting partners in response to extracellular or intracellular changes, is crucial. These insights are fundamental for unraveling the biological functions of the BCR and for identifying the mechanisms by which it regulates immune responses. Studying these aspects of the BCR is vital because it provides a deeper understanding of how adaptive immunity is orchestrated, ultimately guiding the development of therapeutic interventions for immune-related diseases.

In this thesis, we aim to provide new insights into early events that follow BCR antigen engagement using three different approaches that focus on identifying BCR interacting partners and the dynamics of its neighborhood. The approaches adopted in the thesis offer advantages over traditional affinity purification (AP) techniques in identifying novel interactomes. These approaches help in unraveling novel molecular signatures and dynamics of BCR signaling molecules and its associated biological processes. In this study we approached these questions using multifaceted methods; I) by employing novel proximity labeling technique using Ascorbate peroxidase 2 (APEX2), which has not been previously used in studying BCR signaling responses and dynamics, II) by developing AutoCoev pipeline to predict functional interactions in the proteomics dataset, used here to the lipid raft resident proteins orginating from the APEX2 approach above, and lastly III) using chemical cross-linking approach to identify proteins forming the cell contact upon surface-bound antigen stimulation. These multifaceted methods help unravel novel molecular signatures and dynamics of BCR signaling processes, advancing our understanding of BCR signaling and adaptive immune response.

2 Review of the Literature

2.1 B cell and its receptor

B (*b*ursal or *b*one marrow-derived) lymphocytes, also known as B cells, are immune cells characterized by surface expression of clonally diverse antigen-specific surface immunoglobulin receptors, called BCRs (LeBien & Tedder, 2008). The main functions of B cells include antibody production, antigen presentation, cytokine secretion and immune regulation (Wen et al., 2019). B cells are notorious for possessing the capability to discriminate between pathogens and self-antigens, they produce antibodies to neutralize infections and develop into memory B cells that provide long-lasting immunity against antigens they had previously encountered (Kwak et al., 2019a). If B cells, for some reason, are rendered unable to carry out these functions, the individual is exposed to a great health risk.

BCR is a multifunctional receptor used by B cells to sense and probe their environment for foreign substances. BCR activation triggers numerous biological processes in B cells, such as antigen internalization and processing, cytoskeleton reorganization, signal transduction etc. Additionally, BCR signalling is essential for B cell proliferation, survival, development, and differentiation into memory or plasma cells (Tanaka & Baba, 2020). Essentially, recognition of antigens via BCR needs to be specific and not promiscuous, to avoid self-reactive B cells which play a pathological role in diseases such as systemic lupus erythematosus (SLE), rheumatoid arthritis, and diabetes.

The development of B cells from progenitor cells happens in the primary lymphoid tissues (fetal liver and bone marrow). They progress through stages including progenitor B cells, pre-B cells, immature B cells, transitional B cells, and mature B cells. Upon activation, naïve B cells undergo differentiation into either plasma cells (antibody-producing cells) or memory B cells, a process that primarily takes place in the secondary lymphoid tissues such as the lymph nodes and the spleen. During both development and differentiation stages, one the prominent facilitator is the

functional combinatorial rearrangement of Immunoglobulin loci (Brack et al., 1978; LeBien & Tedder, 2008). The combinatorial rearrangement of Ig loci through the rearrangement of the V, D, and J gene segments of the heavy chain locus and the V and J gene segments in the light chain locus is essential for an unhindered progression of B cells development (Brack et al., 1978; LeBien & Tedder, 2008). This re-combinatory rearrangement is the source of a diverse repertoire of functional VDJH and VJL rearrangements that encodes the B cell receptor (BCR). Notably, during B cell development, both exogenous and intrinsic signals are needed for proper B cell proliferation and differentiation (Herzog & Jumaa, 2012). At the early stage of development, pro-B cells express precursor B-cell receptor (pre-BCR), a type of BCR that is composed of two μ-heavy chains that are linked with surrogate light chain and Igα and Igβ (Rickert, 2013; Winkler & Mårtensson, 2018). Pre-B cells differentiate from pro-B cells, during which the Ig heavy chain genes undergo recombinant rearrangement. Pre-B cells lack complete BCR (Mårtensson et al., 2010; Rickert, 2013; Wen et al., 2019). Pre-BCRs serve as a major checkpoint in B cell development (Mårtensson et al., 2010). After transitioning of a pre-B cell into an immature B cell, light chain gene rearrangement leads to the formation of a complete BCR (IgM) molecule that is responsive to activation by antigens. Subsequently, mature B cells develop from immature B cells via transitional stages. Mature B cells express both IgM and IgD on their membrane surface (Lutz et al., 1998; Wen et al., 2019). Pathogen stimulation of mature B cells subsequently leads to the cell proliferation and differentiation into both short-lived antibody producing plasma cells and long-lived plasma cells that persist in the bone marrow to maintain long-term humoral immunity. Notably, after stimulation and activation by antigens, some B cells are differentiated into memory B cells, by exiting the cell cycle, and remain in a resting state capable of rapid activation upon re-exposure to the antigen. These B cells lose their IgD expression but have a longer life span. They remain primed for rapid activation and differentiation into plasma cells whenever they encounter the same antigen again (Wen et al., 2019).

2.1.1 Functions and structure of the BCR complex

BCR is a transmembrane protein that regulates multiple processes in the life cycle of B cells, from immature to mature B cells (Seda & Mraz, 2015). Binding of antigens to BCR initiates a cascade of intracellular signaling that

subsequently leads to antigen internalization, followed by antigen processing and presentation to T cells (Treanor, 2012).

BCR consists of two immunoglobulins (Ig) heavy chain (IgH) and two light (IgL) chains (Fig. 1). These Ig chains possess unique complementarity-determining regions (CDRs), generated via VDJ recombination that enable BCRs to recognize a diverse range of antigens (Tanaka & Baba, 2020). The BCR complex encompasses membrane bound immunoglobulin (mIg), Igα (CD79A) and Igβ (CD79B). The BCR alone lacks the ability to transduce downstream signals, they only gain this capability when they form complex with CD79A and CD79B. CD79A and CD79B possess intracellular immunoreceptor tyrosine-based activation motifs (ITAMs) called immunoreceptor tyrosine-based kinase (Fig. 1).

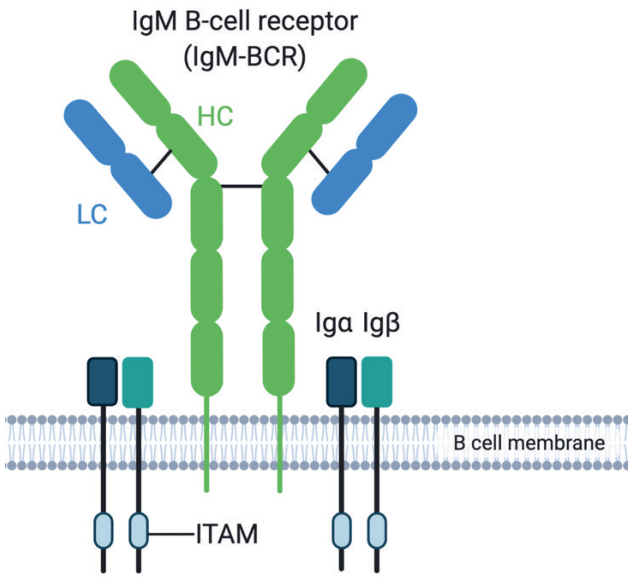

Figure 1: **Structure of BCR**: The B cell receptor is an assembly of two heavy chains (HCs) and two light chains (LCs) associated with the signaling units, Igα and Igβ. The HCs and LCs are linked together by disulphide bonds. At the cytoplasmic side the Igα and Igβ possess immunoreceptor tyrosine-based activation motifs (ITAM), which are sites of phosphorylation upon BCR activation.

2.1.2 BCR signaling and regulation

BCR signaling is a critical process that ensures the precise activation of B cells. This process is an essential prerequisite for the adaptive immune response. Antigen-initiated activation of B cells requires impeccable specificity to avoid antibody recognition of host cells, which can be pathogenic and lead to autoimmune diseases.

BCR is linked to various downstream networks of kinases. These kinases help to augment BCR activation (Seda & Mraz, 2015). Upon antigen engagement, the BCR changes conformation and forms clusters on the B cell plasma membrane. This subsequently leads to the phosphorylation of ITAM domains that then serve as a docking site for protein tyrosine kinases (PTKs) and SH-2 domain containing proteins such as SYK, BLK, BTK (Kurosaki, 2011; Tanaka & Baba, 2020). This is then followed by activation of neighboring signaling molecules, leading to the formation of the signalosome (central signaling complex). Pathways that are activated following BCR stimulations include phospholipase C gamma 2 (PLCγ2)-calcium–nuclear factor of activated T cells (NFAT), the inhibitor of κB kinase (IKK) nuclear factor κB (NFκB) pathway, the Ras–extracellular signal-regulated kinase (ERK) pathway and phosphoinositide 3-kinase (PI3K) Pathway (Fig. 2).

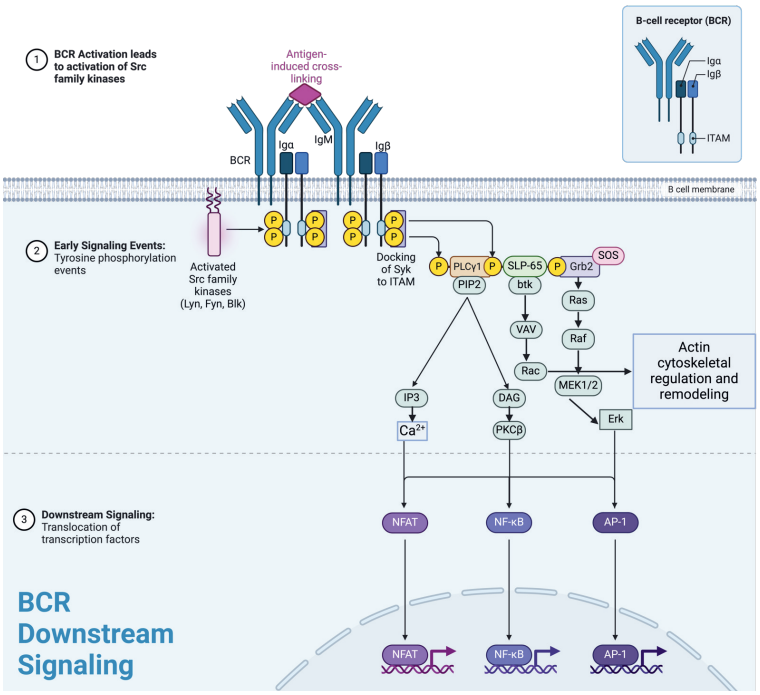

Figure 2: Schematic representation of the BCR signaling pathway. Upon antigen binding, the B cell receptor (BCR) undergoes phosphorylation at the immunoreceptor tyrosine-based activation motifs (ITAMs) within the Igα/Igβ complex by Src family kinases such as Lyn. This triggers the recruitment and activation of Syk, leading to the assembly of the signalosome, which includes adaptor proteins like BLNK (SLP-65) and effector molecules such as PLCγ1. Activated PLCγ1 hydrolyzes PIP_2 into IP_3 and DAG, promoting Ca^{2+} mobilization and PKCβ activation, respectively. These events further activate downstream signaling cascades, including the NF-κB pathway. Image adapted Abbas et al., 2014.

2.1.3 BCR-mediated antigen endocytosis

Stimulation of BCR typically triggers antigen internalization (Hou et al., 2006). Antigen internalization process is aided by local actin polymerization and Myosin IIa, which provides the contractile force needed for membrane invagination (Hoogeboom & Tolar, 2016; Natkanski et al., 2013).

Clathrin-mediated endocytosis (CME) is the main mechanism by which the BCR internalizes antigen. Clathrin and AP-2 are recruited to the site of BCR-antigen engagement, and depletion of either Clathrin or AP-2 diminishes the ability of B cells to internalize both soluble and surface-linked antigens.

After internalization, antigens are transferred to the specialized late endosomal compartments that gradually mature into lysosome-like structures, which also accumulate major histocompatibility complex class II

(MHCII) molecules. This maturation process involves a series of dynamic changes that prepare these compartments for optimal antigen processing. As the compartments mature, they acquire properties similar to lysosomes but are distinct in their role and functionality, being specifically tailored for antigen degradation.

Following antigen degradation, the resulting peptides are then loaded onto MHCII molecules for presentation to T helper cells, facilitating the adaptive immune response (Busman-Sahay et al., 2013; Hämälistö et al., 2024; Hoogeboom & Tolar, 2016; Natkanski et al., 2013; Stoddart et al., 2005).

2.1.4 B cell immunological synapse

Immunological synapse is a transient interface between an immune cell and other cells, consolidated by immune receptors and adhesion molecules. It can form, for example, between B cells and T helper cells, or between cytotoxic T lymphocytes (CTLs) or natural killer (NK) cells and any target cell, such as virus-infected cells or tumor cells (Dustin, 2014, 2016; Kuokkanen et al., 2015a). B cells can be stimulated by both soluble and surface-bound antigens. It is nowadays considered that surface-bound antigens are the predominant form of B cell stimulation in vivo.

During B cell activation, the IS serves as a site where B cell recognizes, extracts and internalizes antigens from antigen presenting cells (APCs) (Kuokkanen et al., 2015). Recognition of antigens, followed by IS formation, antigen extraction and subsequent initiation of BCR downstream signaling cascade and differentiation into antibody-producing cells is initiated and consolidated by BCR and co-immune receptors (Dustin, 2014, 2016; Kuokkanen et al., 2015).

2.2 The plasma membrane and its role in BCR signaling

2.2.1 Plasma membrane compartmentalization

The plasma membrane is an organelle that separates the intracellular and the extracellular space. Its functions include serving as a barrier that controls the

entry and exit of macromolecules. The plasma membrane also participates in the transduction of biochemical signals (Lu & Fairn, 2018).

The membrane is composed of lipids and proteins. The fluid mosaic model of biological membrane was proposed in 1972 by Singer and Nicolson (Singer & Nicolson, 1972; Sonnino & Prinetti, 2013). The model described biological membrane as matrix that is composed of mainly fluid bilayer of phospholipids with integral membrane proteins and glycoproteins. However, recent studies have reported higher level of organization at the plasma membrane and the biological importance of specialized membrane domains, as well as roles played by various molecular components such as actin cytoskeleton and glycocalyx, in limiting the diffusion and movement of membrane components (Möckl, 2020; Sedwick & Altman, 2002).

One of the major characteristics of eukaryotic cells is their compartmentalization and possession of organelles that perform specific functions. The basis for lipid layer formation and separation of intracellular and extracellular spaces is the tendency of the hydrophobic moieties of the lipids to self-associate and exclude aqueous molecules, while the hydrophilic moieties preferentially associate with aqueous molecules (Simons & Sampaio, 2011; Sonnino & Prinetti, 2013). In vertebrates, the cell membrane is mainly composed of glycerophospholipids (GPLs), sphingolipids (SLs) and cholesterol (Sonnino & Prinetti, 2013).

Most cellular membrane-associated processes require transient compartmentalization of proteins, sugars and lipids. The transient nature of membrane compartments coupled with the continuous active dynamic processes such as membrane transport and metabolism, as well as polymerization and motor driven constriction of cell cortex, makes it challenging to study how the interaction between membrane components and underlying cell cortex give rise to the structure and function of cellular membranes (Honigmann & Pralle, 2016).

2.2.2 Lipid raft domains

Lipid rafts are specialized membrane domains that are detergent-resistant and enriched in sphingolipids and cholesterol (Sedwick & Altman, 2002). The idea of raft and non-raft domains on plasma membrane first began with Van Meer et al. 1987 study on the sorting of glycoshingolipid in the trans-Golgi and their transport to the apical membrane in polarized epithelia cells (MDCK cells) (Lu & Fairn, 2018; Simons & Ikonen, 1997; van Meer et al., 1987). The concept was further solidified after discovering that glycosphingolipid-rich membrane regions (a specific subset of sphingolipids that contain sugar molecules attached to the lipid head group) were resistant

to the solubilization by nonionic detergents such as Triton X-100 at 4ºC (Butters & Hughes, 1974; Lu & Fairn, 2018). The lipid raft model has been further strengthened over the years through the observation that model membranes made up of saturated phospholipids, (glyco)sphingolipids and cholesterol are also resistant to detergent solubilization. In contrast, membranes enriched in unsaturated phospholipids or lacking cholesterol tend to be more readily solubilized (Fig. 3). This resistance to detergent solubilization is a characteristic feature of lipid rafts, highlighting their more ordered and tightly packed structure compared to the more fluid, non-raft regions of the membrane. (Lu & Fairn, 2018).

2.2.2.1 Role of lipid raft as BCR signaling domain

While the mechanisms by which signals from many surface receptors are transmitted to the intracellular environment are well studied and understood, the detailed processes by which BCR engagement leads to the coordination of complex intracellular responses remain incompletely understood.

The unique dynamics and molecular interactions within BCR signaling proteins continue to pose challenges to our understanding, requiring further investigation to unravel these complex processes. Studies have shown that most of the downstream signaling molecules that are activated following receptor stimulation are localized directly at the site where the receptors form signaling clusters (Honigmann & Pralle, 2016). Previous studies have shown that following surface receptor stimulation, various proteins translocate to the lipid raft domain while others are excluded. This has greatly strengthened the idea that lipid raft domains serve as assembly points for signaling molecules. While many proteins are resident to the lipid raft domains even before surface receptor activation, some are recruited to the lipid rafts upon activation (Sedwick & Altman, 2002).

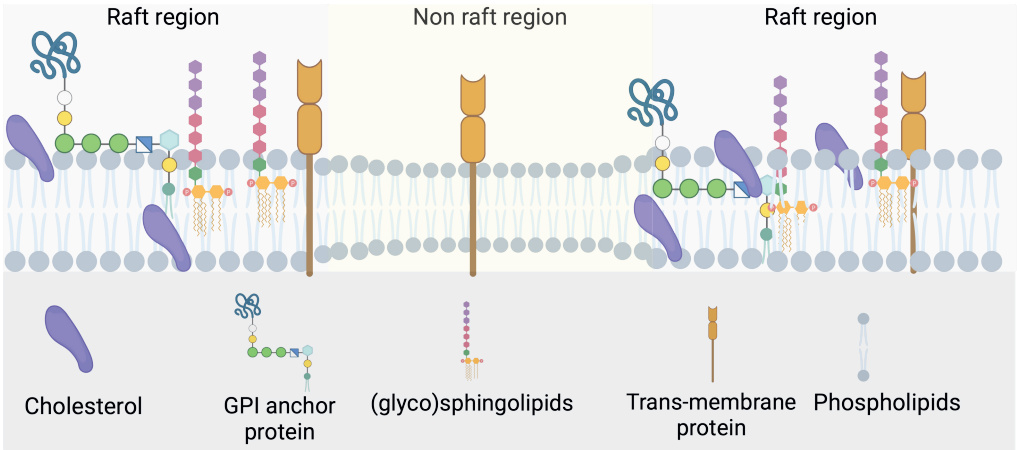

Figure 3: Schematic illustration of the lipid raft and non-raft regions on the plasma membrane. The schematics depicts example of lipid bilayer plasma membrane grouped into a region rich in cholesterols, glycerol phospholipids and GPI anchor (lipid raft region) and another region less rich in cholesterols, glycerol phospholipids and GPI anchor (non-lipid raft region).

Proteins can associate with the lipid rafts through various mechanisms, including post-translational modifications like glycophosphatidylinositol anchors (GPI) and fatty acylations, the presence of transmembrane domains that preferentially interact with lipid raft regions, and affinity for other raft associated proteins (Fig. 3) (Sedwick & Altman, 2002).

Studies have shown that cross-linking of BCR by multivalent antigen leads to the translocation of BCR from non-raft domain to the lipid raft domain of B cell plasma membrane (Pierce, 2002; Sedwick & Altman, 2002; Varshney, Yadav, & Saini, 2016). Notably, the phosphorylation of CD79A and CD79B is not required for the translocation of the B cell receptor (BCR) to the lipid raft domain; rather, this process is primarily driven by BCR cross-linking. These findings were reported by Cheng et al. (2001) in a study using a mutant immunoglobulin (Ig) with a YS→VV substitution. This mutation disrupted the association of the Ig with CD79A/CD79B, rendering it signaling incompetent. However, the mutant was still able to be internalized upon antigen binding. The YS/VV mutant was able to translocate into lipid rafts following cross-linking at 37°C. However, phosphorylation of CD79A and CD79B is required for signaling and activation of changes in the gene transcription (Cheng, Brown, Song, & Pierce, 2001; Sedwick & Altman, 2002).

Proteins like Lyn are constitutively present in the lipid rafts prior to antigen engagement while proteins such as BLNK (B-cell linker protein), PI3K and PLCγ2 (phospholipase C gamma 2) are recruited to the rafts upon BCR cross-linking. Studies have shown that lipid rafts play an important role in B cell signaling by serving as platforms that organize and concentrate signaling molecules (Sedwick & Altman, 2002). However, the identity and dynamics of proteins at the lipid raft region upon B cells activation are still not well understood (Aman & Ravichandran, 2000; Sedwick & Altman, 2002).

2.2.2.2 Fatty acylation

One of the strategies employed by cells to recruit proteins to the plasma membrane is by fatty acylation. Covalent attachment of fatty acid moieties to specific amino acid residues on proteins is a well-documented post-translational modification process that plays essential role in proteins cellular function, trafficking, stability, aggregation and structure (Deschenes & Linder, 2007; Resh, 1999). The two most common protein fatty acylation modifications are myristoylation and palmitoylation.

2.2.2.2.1 Protein myristoylation and palmitoylation

Protein N-myristoylation is a co-translational modification that involves the covalent attachment of myristic acid to the NH_2-terminal of glycine via amide linkage. This process is catalyzed by *N*-myristoyl transferase (NMT) (Deschenes & Linder, 2007; Goldston, Sharma, Paul, & Engman, 2014; Legrand & Rioux, 2010; Resh, 1999). Protein myristoylation occurs throught define sequence of events in which NMT first binds to myristoyl-coenzyme A (myristoyl-CoA), followed by association with substrate peptide. This enables transfer of the myristate moiety to the exposed N-terminal glycine residue of the substrate protein. Following acyl transfer, coenzyme A is released, after which the newly formed myristoylated peptide dissociates from the enzyme.

Myristoylation modification of proteins has been shown to aid in defining protein subcellular localization, interacting partners and membrane localization (Legrand & Rioux, 2010).

The sequence of *N*-myristoyl transferase protein substrate is typically Met–Gly–X–X–X–Ser/Thr–K. The starting Met is removed during translation by methionine aminopeptidase. The subsequent N-terminus Gly is an irreplaceable residue at N terminal of proteins that will be myristoylated. Residues at position 6 is preferably serine or threonine while

lysine or arginine is the preferred residue at position 7 and/or 8. The role of myristoylation in cells includes structural stability of proteins, membrane binding and membrane targeting. However, this attachment is not always permanent. Many N-myristoylated proteins can switch between two structural states. In one state, the myristoyl group is hidden within a hydrophobic pocket of the protein, while in the other state, the myristoyl group becomes exposed, allowing the protein to bind to the membrane.

This reversible switching mechanism, known as the myristoyl switch, is controlled by ligand binding, changes in electrostatic charge, or proteolytic cleavage, which can trigger the release or exposure of the myristoyl group and thereby influence membrane binding (Resh, 1999).

Similar to myristoylation, palmitoylation is another fatty acylation method employed for targeting proteins to the lipid rafts. Palmitoylation refers to the addition of palmitate to a cysteine residue on a protein. However, multiple long chain fatty acids can be associated with Cys residues, including oleate, stearate and arachidonate. As a result, terms such as thioacylation or S-acylation are often preferred over palmitoylation. While palmitoylation specifically refers to the attachment of a palmitic acid molecule, thioacylation and S-acylation encompass the attachment of various types of long-chain fatty acids, including oleate, stearate, and arachidonate, in addition to palmitic acid. This broader terminology is more accurate and inclusive because it highlights that different fatty acids can be associated with the Cys residue, each potentially influencing the protein's function and localization in unique ways. (Linder & Deschenes, 2007; Resh, 1999).

Usually, palmitate is linked to a Cys residue of a protein through a reversible thioester linkage (*S*-palmitoylation). However, in some cases, palmitate is linked to proteins at Cys residue at the N terminus leading to the formation of a thioester intermediate followed by rearrangement to a stable amide bond. This type of palmitoylation modification is referred to as *N*-palmitoylation (found in Hedgehog proteins). In contrast to S-palmitoylation, N-palmitoylation is irreversible (Deschenes & Linder, 2007). In other words, addition of palmitate to the N terminus Cys leads to the generation of a stable irreversible amide linkage of fatty acid.

2.3 Proximity labeling techniques

Communication between proteins underlies critical cellular processes that drive physiological functions. Therefore, any technique that provides the capability to map these interactions in time and space is important to gain better understanding of the dynamics and organization of proteins in key biological processes (Kim & Roux, 2016; Trinkle-Mulcahy, 2019). In immune cells, and particularly in B lymphocytes, protein-protein interactions play a central role in orchestrating B cell receptor (BCR) signaling. BCR activation triggers rapid and dynamic assembly of signaling complexes at the plasma membrane, followed by receptor internalization and trafficking (Hou et al., 2006; Tanaka & Baba, 2020). Many of these events occur within specialized membrane microdomains, such as lipid rafts, which act as signaling hubs enriched in kinases, adaptors, and trafficking proteins. The transient and spatially restricted nature of these interactions presents significant challenges for conventional interaction-mapping techniques (Pierce, 2002; Sedwick & Altman, 2002; Varshney, Yadav, & Saini, 2016).

Unfortunately, conventional methods such as classical AP, in which a tagged bait protein is isolated together with co-purifying interaction partners, and yeast two-hybrid are limiting in scope and physiological relevance. Although they are efficient in identifying direct physical interacting partners in vitro, they do not necessarily reflect physiological conditions of cells (Kim & Roux, 2016). These limitations are particularly relevant for studying BCR signaling, where signaling complexes assemble rapidly and transiently at the plasma membrane and within lipid raft domains. As a result, important regulatory proteins may be missed by approaches that rely on prolonged purification or non-physiological assay conditions.

In recent years, different proximity labeling techniques in living cells have become popular. These techniques are powerful tools for mapping protein-protein interactions in physiological cellular environments. The major advantages over classical AP technique or yeast-2-hybrid (Y2H) include ability to capture weak/transient protein-protein interactions, efficient detection of low-abundance proteins, applicability to both soluble and insoluble proteins, ability to withstand rigorous washes helping to reduce background contaminants, and ability to study protein-protein interactions in natural cellular settings (Kim & Roux, 2016; Mair et al., 2019; Roux et al., 2018; Trinkle-Mulcahy, 2019). The two most widely used proximity labeling techniques in live cells are BioID and APEX2. Both are biotin-based labeling methods. The major difference between these two

methods is that APEX2-mediated labeling occurs within a time frame of 1 min while BioID-mediated labeling typically takes hours (18 hrs). However, faster variants of BioID, such as TurboID and miniTurbo, have been developed to reduce labeling times to approximately 10 min, thereby offering improved temporal resolution (**Table 1**).

2.3.1 Enzyme-initiated proximity labeling: concept and mechanism

Enzyme initiated proximity labeling methods rely on the generation of small and unstable reactive molecules that have the ability to covalently interact with neighboring proteins (Kim & Roux, 2016; Rees, Li, Perrett, Lilley, & Jackson, 2015). Introducing a labeling reagent that contains biotin then leads to covalent tagging of the nearby proteins with biotin. These tagged proteins can then be enriched, after cell lysis, using a streptavidin affinity pulldown approach approach. Proteins derived after streptavidin enrichment constitute mainly those proteins in the activity sphere of the proximity labeling enzyme. As the enriched proteins predominantly represent the local molecular environment of the bait, enzyme-initiated proximity labeling provides a powerful strategy for capturing dynamic and spatially confined protein assemblies, such as those formed during BCR activation within lipid raft domains.

Table 1. Comparison of different proximity labeling techniques.

CLASS	Tags	Source	Size	Target	Labeling time	Activity region	Strength	Limitation	Ref
Biotin Ligase	BirA*	E. Coli	35kDa	Lysine	15 -24hrs	Intracellular	High specificity; established method	Long labeling time; not ideal for study fast dynamic cellular processes	1;2
	TurboID	E. Coli	35kDa		10 min		Faster labeling; exhibit greater efficiency of biotinylation than BirA* or BioID2	Can generate backgroud noise	2
	MinTurbo	E. Coli	28kDa		10 min		Smaller size and exhibit greater efficiency of biotinylation than BirA* or BioID2	Few data on its use; background noise	2
	BioID2	A. Aeolicus	27kDa		15 -24 hrs		Smaller size; lower cytotoxicity than BirA	Long labeling time; not ideal for study fast dynamic cellular processes	1;2
	BASU	B. Subtilis	28kDa		<= 1 min		Fast labeling; lesser background labeling compared to BirA*	Background labeling can be an issue	3
Peroxidase	APEX	Soybeans	28kDa	Tyrosine and electron rich amino acids	<= 1min		Rapid labeling; suited for real-time studies; lesser backgroud noise compare to BirA*	Requires hydrogen peroxide; potential toxicity	2;4
	APEX2		28kDa		<= 1 min		Improved stability and efficiency over APEX	Requires hydrogen peroxide; potential toxicity	2;4
	HRP		44kDa	Electron rich amino acids	<= 1 min	Extracellular	Well-established; widely used in biochemical assays	Limited to extracellular space	5
	Pup-IT		54kDa	Various protein targets	Variable	Intracellular and extracellular	Can target broader range of proteins	Larger size; labeling speed slower than TurboID	6

References corresponding to details in the table above: 1:(Kim et al., 2016), 2: (Branon et al., 2018), 3: (Ramanathan et al., 2018) 4: (Lam et al., 2015), 5: (Li et al., 2014), 6: (Striebel et al., 2014).

2.3.2 APEX2 technique

One of the most powerful enzyme-based proximity labeling technique is based on the enzyme ascorbate peroxidase (APEX) or its improved variant APEX2. APEX is an engineered soybean ascorbate peroxidase (APX) created by Ting group in 2013 (Rhee et al., 2013), with the initial application as a tag for electron microscopy, but it was later adapted for use as a proximity labeling enzyme in live cells (Lam et al., 2015). Similarly to horseradish peroxidase (HRP), APEX converts its substrate into radicals upon addition of H_2O_2 (Bosch et al., 2021). What makes APEX unique compared to similar enzymes like HRP is that APEX lacks disulfide bond and calcium ions that could render them susceptible to reducing environments such as the cytosol.

APX is a 28kd protein and exist as a homodimer that can actively dimerize. This makes WT-APX unsuitable for use in both electron microscope and proximity labeling as oligomerized APX can cause disruption in proteins natural localization and activities (Mandelman et al., 1998; Martell et al., 2012). To address this limitation, the Ting group

created APEX, a monomeric 28-kDa peroxidase engineered from WT-APX through three mutations: K14D, E112K and W41F. APEX2 is an improved version of APEX with additional mutations of K14D, E112K, A134P and W41F (Hung et al., 2016; Lam et al., 2015; Martell et al., 2012; Trinkle-Mulcahy, 2019). As previously mentioned, the main advantage of APEX/2 over other proximity labeling methods is fast labeling time, requiring just less than 1 min. This gave APEX/2 an advantage in studying dynamic changes in high temporal resolution (Trinkle-Mulcahy, 2019).

2.3.3 Mechanism of action of APEX2

In the presence of H_2O_2, APEX2 catalyses the oxidation of biotin-phenol (biotin teramide) that leads to the formation of a short-lived (<1 ms) biotin-phenoxyl radical (Fig 4). These radicals react with neighbouring electron-rich amino acids such as tyrosine and possibly tryptophan, cysteine, and histidine. This results into biotinylation of neighboring (in 20 nm range) proteins and RNAs (Hung et al., 2016; Kim & Roux, 2016; Lam et al., 2015; Trinkle-Mulcahy, 2019). One important property of APEX2 biotinylation over its rival tag BioID is its faster labeling efficiency. Sufficient quantities of APEX2 labeling occur within 1 min of adding H_2O_2 while BioID labeling of similar efficiency takes hours. This rapid labeling is specifically crucial for studying dynamic cellular processes where protein interactions or signaling events occur over very short timeframes. For instance, transient protein-protein interactions during BCR signaling or rapid changes in the subcellular protein localization are better captured using APEX2, as its speed allows for more accurate snapshots of these fast-moving molecular events. It is worth noting, however, that BioID has also advanced. The introduction of BioID Turbo, a faster version of BioID, has improved its efficiency. BioID Turbo reduces the labeling time to about 10 minutes, enhancing its temporal resolution and making it more competitive with APEX2 (Branon et al., 2018). Despite this, APEX2 maintains an edge due to its sub-minute speed. Another advantage of APEX is that APEX2 is smaller (27kDa) compared to BirA (35kDa) and are theoretically less likely to affect the function of the bait (Trinkle-Mulcahy, 2019).

In scenarios where real-time or very rapid changes in protein interaction networks need to be studied, such as the studying of antigen receptor signaling, the rapid biotinylation offered by APEX2 can provide a more accurate representation of these fast-evolving molecular events.

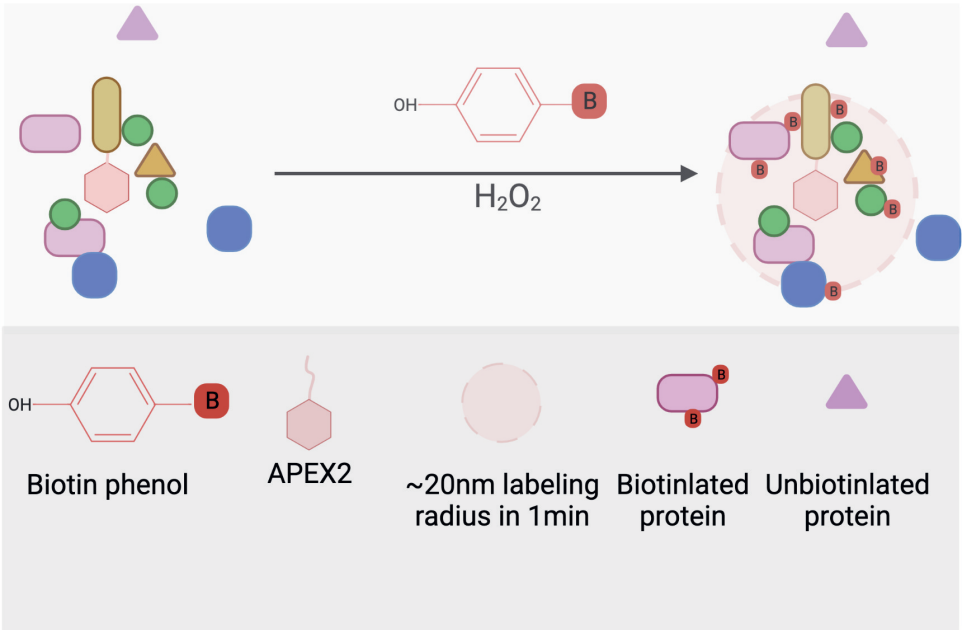

Figure 4: A schematic illustration of the APEX2 mechanism of action. The engineered ascorbate peroxidase enzyme, APEX2, catalyzes the biotinylation of nearby proteins in the presence of hydrogen peroxide and biotin-phenol. Upon activation by hydrogen peroxide, APEX2 generates a short-lived biotin-phenoxyl radical that rapidly labels proximal proteins in 20 nm range in living cell environment. Although labeling occurs within seconds, experimental protocols typically allow a 1-minute reaction window before quenching to ensure sufficient tagging.

2.4 Quantitative mass spectrometry analysis

One essential route to understanding the functions of biomolecules is by measuring their responses to various environmental cues. Quantification of these responses is important in the field of biomedicine.

Mass spectrometer is one of the common instruments used for this purpose and mass spectrometry-based quantification methods have gained popularity in the past decades because it has enabled important breakthroughs in biomedicine that has revolutionized various aspects of research, diagnosis, and treatment (Calderón-Celis et al., 2018; Rozanova et al., 2021).

In the context of immune receptor signaling, quantitative proteomics has been instrumental in dissecting B cell receptor (BCR) signaling pathways, identifying downstream effectors, and mapping activation-dependent changes in protein composition. Proteomics-based studies have been used to characterize BCR-induced phosphorylation events, signaling complexes, and receptor trafficking processes, providing system-level insight beyond what is achievable with candidate-based approaches (Oellerich et al., 2013; Satpathy et al., 2015).

Mass spectrometry is generally referred to as a technique that separates electrically charged molecules in gas phase. It is used to measure mass-to-charge ratios of ions. It has proven to be an essential technique for studying biomolecules, especially proteomics. A typical mass spectrometer is composed of three parts, namely: an ion source, the mass analyser and a detector. The ion source is where the analytes are ionized before being transferred into the mass analyser in the form of ions where it then resolves the ionized analytes in space and time according to its mass-charge-ratio (m/z). The ions separated by the mass analyser are then detected by the detector which also turns the proteins into electric signals that can be read by mass spectrometrist (Domon & Aebersold, 2006; Haag, 2016; Urban, 2016).

Traditional proteomic quantification methods are gel electrophoresis, immunochemistry, and immunoblotting (Calderón-Celis et al., 2018). However, as mentioned above, advances in mass spectrometry have revolutionized quantitative proteomic methods. With the help of mass spectrometry-based proteomic approaches researchers can identify, as well as quantify, thousands of both previously known and previously unknown proteins in a biological system (Calderón-Celis et al., 2018; Domon & Aebersold, 2006).

This capability is particularly important for studying lipid raft associated signaling, where dynamic and transient recruitment of proteins occurs in

response to receptor activation. Proteomics studies have previously been used to profile lipid raft composition and its reorganization upon immune receptor stimulation, revealing that lipid rafts act as signaling hubs enriched in kinases, adaptors, and trafficking proteins (Saeki et al., 2003). However, many of these studies have relied on biochemical fractionation or affinity purification approaches, which may miss transient or spatially restricted interactions.

Another benefit of the mass spectrometry-based proteomic approaches is the fact that many peptides are derived from a single protein, enabling robust quantification and improving reproducibility compared to immunoassays-based methods, where antibody specificity and cross-reactivity can lead into less reproducibility. Despite their limitations, immunoassay techniques are still relevant in research, clinical analysis and validation (Calderón-Celis et al., 2018).

Together, these considerations highlight the value of quantitative mass spectrometry for studying BCR signaling and lipid raft–associated proteomes. In this thesis, mass spectrometry based proteomics forms the foundation for investigating activation-dependent changes in the BCR signaling environment and for identifying novel regulators of BCR function.

There are two main approaches for mass spectrometry-based proteomic quantification methods: the label-free and the labelled quantification methods.

2.4.1 Label-free and labelled protein quantification

As mentioned above, proteomic quantification analysis can be achieved using either labelled or label-free methods.

Labelled methods involve introducing mass tags, as in the case of metabolic labeling, or introducing a specific isotopes into proteins in vivo. On the other hand, label-free methods require no labeling of either the proteins or peptides. Instead, proteins are quantified by measuring peptide intensities or counting the number of peptide fragments generated.

Table 2. Comparison of labelled and label-free LC-MS approach

Quantification methods		Labeling level	Medium used to introduce label	Max. number of conditions	Sample type
Label free	Spectral counting	n.a	n.a	n.a	All samples
	Ion Intensity				
Metabolic label	SILAC	Protein	in vivo and ex vivo	5	Cells, tissues and model organism
	^{15}N			2	Cells
Chemical label	iTRAQ	Peptide	in vitro	8	All samples
	Dimethyl labeling			3	
	TMT			18	
	ICAT	Protein		2	
Enzymatic labeling	^{18}O labeling	Preptide	in vivo	2	

Both labelled and label-free methods have their distinct advantages. The main advantage of label-free Liquid Chromatography-Mass Spectrometry (LC-MS) over labelled LC-MS approach are that label-free LC-MS approach are less expensive, require simpler workflow, and can be used to compare unlimited numbers of samples and conditions (Table 2). Conversely, labelled quantification methods offer superior accuracy and higher reproducibility.

2.4.2 Analyzing and Interpreting Proteomics Data

Making sense of proteomics data is the most important and often the most challenging aspect in proteomics data analysis workflow. Proteomics data analysis stage usually takes more time than the experiment / data collection stage. Schematics below (Fig. 5) illustrates the typical workflow for proteomics data analysis.

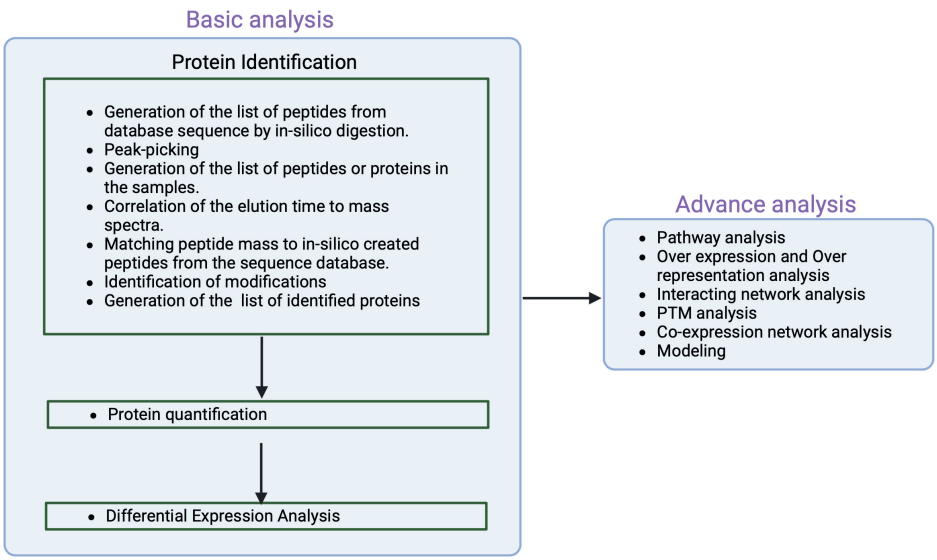

Figure 5: Illustration of a typical proteomics data analysis workflow, showcasing the sequential steps involved from protein identification to data analysis and interpretation.

2.4.2.1 Missing values and imputation

Like all real word data, proteomics data analysis is also marred with problems such as missing values, arising from stochastic peptide detection or low-abundance proteins falling below the detection threshold, and batch effects, which are systematic, non-biological sources of variation introduced during data generation that can obscure or confound true biological signals (Goh et al., 2017; Jin et al., 2021).

Missing values are a common problem in label-free quantitative proteomics. Datasets without missing values are preferred for most downstream analysis. This is because missing values can affect the quality and statistical analysis of quantitative proteomics data. As a result, it needs to be tactically handled to avoid misinterpretation.

In label-free proteomic studies, imputation of missing values is an important step prior to differential and functional analysis. It improves statistical power and coverage, thereby enabling more robust identification of biologically relevant proteins (Jin et al., 2021). In some instance rows with missing values are all filtered out, and the analysis is performed on the remaining dataset. However, this approach may not be suitable all the time as it may result into loss of important proteins.
Alternatively, a common solution that has been widely adopted is to perform data imputation. Determining the best imputation method specific for a particular proteomic dataset is essential for accuracy and proper representation of data (Jin et al., 2021). Deciding the appropriate imputation method requires good understanding of the cause of the missing values. Common source of missing values in LC-MS include low protein abundance, which can lead to some proteins or peptides being undetected by the instrument. The biological factors**,** such as total absence of a protein in a sample, can also contribute to this issue. Additionally, analytical source can play a role, including mis-cleavage of peptides during digestion, loss of sample during sample preparation, and poor ionization efficiency (Jin et al., 2021; Karpievitch et al., 2012).

Missing values present in proteomic datasets are generally characterized as missing at random (MAR) or missing not at random (MNAR) (Jin et al., 2021; Lazar et al., 2016). MAR are the type of missing values that occur due to technical limitations and stochastic fluctuations in an abundance-independent manner. Conversely, MNAR result due to low peptide abundance that falls below detection limits of the instrument, or the peptide is completely missing. This is an example of left censoring. Missing data in proteomics data are typically a mixture of both MAR and MNAR. It is believed MNAR constitute the major type of missing values in proteomic

datasets. Because it is often difficult to determine the most prominent type of missing data in a dataset, it is therefore recommended to use imputation methods capable of handling mixed types of missing values (Jin et al., 2021; Lazar et al., 2016). One way to determine the type of missing value in a proteomic dataset is to check if a protein is completely missing in a sample group or not. If a protein is completely missing in one sample group, it is probably a MNAR but if not, then it is likely MAR (Jin et al., 2021).
For MNAR methods (left-censored methods), suitable methods include lowest of detection or left-censored normal distribution. On the other hand, for MAR, imputation methods such as k-nearest neighbours (kNN), local least squares (LLS), random forest (RF), singular value decomposition (SVD) are more appropriate.

2.4.2.2 Data normalization

Normalization is an essential aspect of omics data pre-processing. It is required to remove or reduce systemic bias, including technical and biological variation, that are introduced during sample preparation and processing (Webb-Robertson et al., 2011; Willforss et al., 2019). Not knowing the origin of these biases make it difficult to determine the best normalization methods. Unfortunately, there is no one method-fits-all solution. Importantly, downstream quantitative analysis can be greatly misleading if wrong normalization method is selected. As a result, prior evaluation of the normalization strategies is critical to ensure accurate downstream quantitative analysis. Webb-Robertson et al., 2011 and Willforss et al., 2019 propose statistical frameworks for evaluating and selecting normalization methods in omics data by assessing their impact on variance structure, technical bias, and downstream biological inference. They demonstrated that normalization performance is data-dependent and should be evaluated using objective criteria rather than applied uniformly across datasets (Webb-Robertson et al., 2011; Willforss et al., 2019).

2.4.2.3 Batch correction

Another major challenge in high-throughput omics data analysis is the presence of batch effects. Batch effects typically arise during sample preparation due to factors such as different laboratory conditions, experiment times, instruments, handlers, and reagents (Goh et al., 2017; Leek et al., 2010). It is a common practice to use batch effect correction algorithms to adjust for batch effects. The most common batch effect-

correction algorithm for adjusting batch effects of known source is ComBat while SVA is common for adjusting for batch effect of unknown source (Goh et al., 2017; Johnson et al., 2007; Zhang et al., 2020).

2.4.2.4 Differential enrichment analysis

One of the main goal of mass spectrometry-based proteomics assays that rely on selective capture or labeling is to identify list of proteins that are differentially enriched between treated or disease samples and appropriate controls. This list is typically derived through differential enrichement analysis, where a specific threshold is set and the proteins that meet this threshold are considered differentially expressed.

The list delivered from this type of analysis are typically in tens to hundreds. However, some of these hits might turn out to be false positive, as subsequent validation could show that their enrichment is not driven by treatment or disease. Identifying reliable signatures/hits serves as a precursor for identifying potential biomarkers, understanding of the underlying mechanism of action, biological process or pathway. Importantly, ideal molecular signatures should be reproducible across similar disease or treatments, be stable, capable of discriminating between control and treated samples/disease, expression driven by treatment/disease and in some instance initiate phenotypic changes in cells when disrupted (Tang et al., 2020).

To mitigate false positive identification process, supplemental methods such as literature mining, transcriptional profiling and machine learning methods may be employed. One important machine learning method that can be employed is feature selection method (Pudjihartono et al., 2022).

2.5 Protein coevolution

Interaction among proteins drive and regulate many processes. Both direct and non-direct interactions of proteins are essential for the proper functioning of biological systems. Identifying protein interacting partners is a cornerstone of unraveling novel regulators. Traditional methods of determining relationship or communication between proteins mostly require experimental procedures such as Co-IP etc., that can identify physically interacting proteins. However, studies have shown that functional relationships between proteins transverse beyond physical interactions. Therefore, techniques that rely solely on identifying direct interactors may not be sufficient to solve many biological enigmas. In recent years, bioinformatics tools have been developed to aid in determining possible physical and functional relationship among proteins. One of such approach involves determining co-evolving amino acids within (intra-coevolution) or between proteins (inter-coevolution). This is based on prediction of interacting residues between or among proteins from multiple aligned orthologous sequences. Co-evolving amino acids are regarded as evidence of functional relationship within or between proteins. For example, inter-protein coevolution analysis has been used to identify specificity determining residues and predict interaction interfaces in bacterial two-component signaling systems, where mutations at coevolving sites can rewire histidine kinase response regulator specificity (Skerker et al., 2008). Protein co-evolution is influenced by its building block (amino acid sequence). This in turn is influenced by the functional and structural properties of amino acids involved (Chakrabarti & Panchenko, 2009).

Determining coevolution between two proteins requires pairing the sequences of the proteins from multiple different species (Fig. 6). Coevolution analysis using protein sequences (CAPS) is one of the most popular tools for detecting coevolution between or within proteins. CAPS predicts co-evolving amino acids by measuring the correlated evolutionary variation of these amino acids. In CAPS, evolutionary variation is measured using time corrected Blosum values for the switch between two amino acids at a particular site. This interchange, between amino acids sites is then corrected by the divergence time of the sequences. The time is calculated as an average number of substitutions / synonymous sites between the two sequences being compared. Correlation of the mean variability is calculated with Pearson coefficiency. And in the end, the significance of the correlation coefficients is determined by comparing the observed correlation coefficients with the distribution of re-sampled correlation coefficients (Fares & McNally, 2006a). Other tools for measuring coevolution includes

BIS2 (Oteri et al., 2017), ContactMap (Wang et al., 2017), DCA (Morcos et al., 2011), Evcouplings (Hopf et al., 2019) and MISTIC (Simonetti et al., 2013).

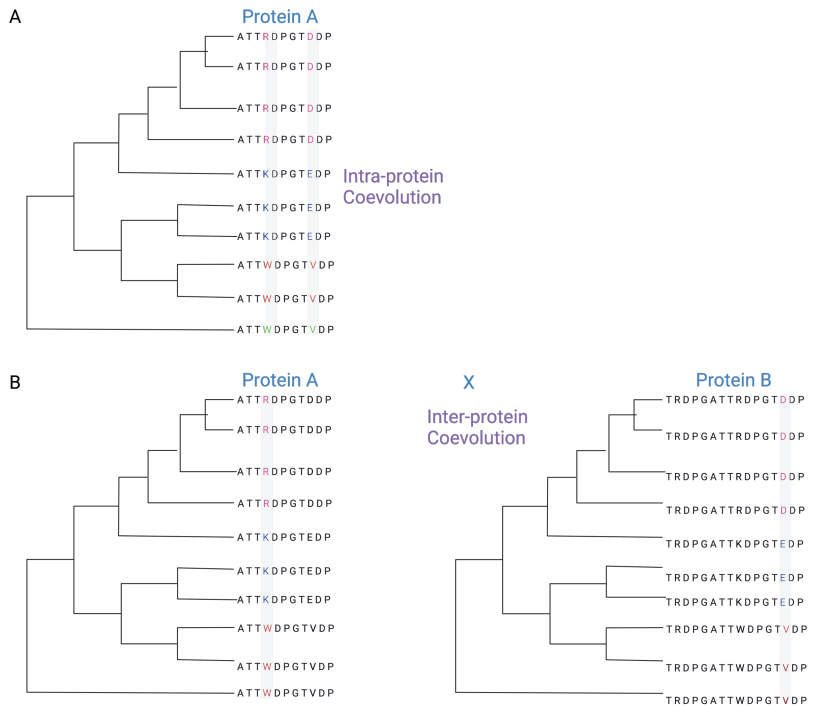

Figure 6: Intra- and inter-protein coevolution. Multiple sequence alignment of homologous sequences. a) Intra-protein coevolution; change in one amino acid within a Protein A causes change in another amino acid within same protein and b) inter-protein coevolution; change in one amino acid in Protein A causes change in another amino acid in another Protein B. Note: amino acids with same color on different location on same protein or another proteins signifies amino acids that have co-evolved.

3 THE AIMS OF THIS THESIS

The primary goal of this thesis was to identify novel regulators of B cell receptor signaling. To achieve this, we focused on methods including proximity labeling proteomics (APEX2), immunological synapse isolation and protein coevolution prediction.

The individual aims of the presented studies were as follows:

1. Using a proximity labeling method to study the dynamics of BCR signaling proteins and downstream cellular reactions by identifying proteins that are significantly recruited or excluded in the lipid raft region, the site of BCR enrichment, upon antigen activation. (study I)
2. Building a pipeline that can be used to predict coevolution between multitude of proteins, to aid the analysis of proteomics datasets. (study II)
3. Use chemical cross-linking approach to extract the B cell IS, upon surface antigen activation, for proteomic analysis. (study III)

4 MATERIALS AND METHODS

Note: Only the methods personally performed are described in detail in this section.

Design and cloning of raft-APEX2

To design and clone the Raft-APEX2, the pcDNA3-mito-APEX (Rhee et al., 2013) (Addgene plasmid # 42607), was used as a template to create and PCR amplify V5 (GKPIPNPLLGLDST) epitope-tagged APEX2 cDNA. We then cloned mCherry with N-terminal seven amino acid sequence (MGCVCSS) that encodes the acylation sequence and APEX2 were then cloned into pcDNA™4/TO plasmid with zeocin selection (Invitrogen V1020-20). This final construct was submitted to Addgene for availability, and it has been assigned Addgene plasmid # 205089.

Cells

We used the A20 mouse and Raji human B cell lines, which stably expresses a hen egg lysozyme (HEL)-specific IgM BCR (D1.3) (Aluvihare et al., 1997). A20 D1.3 cells were maintained in complete RPMI (cRPMI; RPMI 1640 with 2.05 mM L-glutamine supplemented with 10% fetal calf serum (FCS), 4 mM L-glutamine, 50 μM β-mercaptoethanol, 10 mM HEPES and 100 U/ml penicillin/streptomycin). Raji D1.3 cells were cultured in Raji cRPMI (RPMI 1640 with 2.05 mM L-glutamine supplemented with 10% FCS, 4 mM L-glutamine and 100 U/ml penicillin/streptomycin).

Generation of raft-APEX2 expressing stable cell line

The Raft-APEX2-pcDNA™4/Zeo/TO plasmid was transfected into the A20 D1.3 cell line as previously described (Šuštar et al., 2018). Briefly, 4 million cells were resuspended in 180 µl of 2S transfection buffer (15 mM HEPES, 50 mM Sodium Succinate, 180 mM Na_2HPO_4/ NaH_2PO_4 pH 7.2, 5 mM KCl and 15 mM $MgCl_2$) containing 10 µg of plasmid DNA and electroporated using AMAXA electroporation machine (program X-005, Biosystem) in 0.2 cm gap electroporation cuvettes. The cells were then transferred to 4 ml of cRPMI containing an extra 10% FCS for overnight recovery. Next, single cell sorting was performed using a Sony SH800 Cell Sorter into 96-well flat-bottom plates with 100 µl of cRPMI containing 20% FCS. Cells were left to recover in cRPMI supplemented with an extra 10% FCS for 48 hrs before adding Zeocin (600 µg/ml final concentration). Clones expressing raft-APEX2 were selected and expanded for a few weeks after sorting, and raft-APEX2 expression was verified with flow cytometry analysis for mCherry, V5, and functional biotinylation.

Proximity biotinylation

For proximity botinylation, 10 x 10^6 A20 D1.3 raft-APEX2 cells were treated with 500 µM biotin-phenol (BP) (Iris-Biotech, CAS no.: 41994-02-9) in 5 ml of cRPMI for 45 min, followed by activativation with 0 or 10 µg/ml of goat anti-mouse IgM F(ab')$_2$ fragments (Jackson ImmunoResearch 115-006-020) for 5, 10 or 15 min. Then, 1 mM H_2O_2 (Sigma-Aldrich, cat. no. H1009-100ML) was added for 1 min, followed by quenching with 2X quenching solution (20 mM sodium ascorbate, 10 mM Trolox (Sigma-Aldrich, cat. no. 238813-1G) and 20 mM sodium azide solution in PBS). The cells were repeatedly washed 4 times with 1X quenching solution. Control samples, non-biotinylated and background were prepared similarly but without anti-IgM and H_2O_2 and without BP and anti-IgM, respectively. To validate biotinylation for each experiment, we used flow cytometry, with cells fixed, permeabilised and stained with streptavidin 633 (Thermo Fischer Scientific).

Lysis

Samples were lysed with modified RIPA buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 2% Octyl glucoside (Sigma-Aldrich, 29836-26-8), 0.5% sodium deoxycholate and 1% Triton X-100, pH 7.5) with 1× protease phosphatase inhibitor mini-tablet (1 tablet/10 ml, Thermo Fisher Scientific,

cat. no. A32961). Lysate concentrations were measured using Pierce 660 nm protein assay (Thermo Fisher Scientific, cat. no. 22660), aliquoted into 360 µg of total protein/aliquot, snap-frozen and stored at -80 °C.

Streptavidin pull-down of biotinylated proteins

350 µg of whole lysate diluted in an additional 500 µl of RIPA buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate and 1% Triton X-100, pH 7.5, 1× protease phosphatase inhibitor mini-tablet) was incubated with 30 µl (0.3 mg) of streptavidin magnetic beads (Pierce, cat. no. 88817) for 1 h at room temperature on rotation. The beads were washed several times with 1 ml volumes for 5 min each, on ice: twice with RIPA buffer, twice with 1 M KCl, twice with 0.1 M Na_2CO_3, once with 4 M urea in 10 mM Tris-HCl pH 8.0, once with 50 µM biotin, 4 M urea in 10 mM Tris-HCl pH 8.0, and three times with RIPA buffer. Biotinylated proteins were eluted by boiling the beads in 30 µl of 3× SDS loading buffer (188 mM Tris-Cl (pH 6.8), 3% SDS, 30% glycerol, 0.01% bromophenol blue) supplemented with 2 mM biotin and 20 mM DTT for 10 min.

In-gel digestion

Eluted samples were run on 10% SDS-PAGE and with SimplyBlue SafeStain (ThermoFisher Scientific, cat. no. LC6065). For the digestion, the protocol adapted from Shevchenko et al. was used (Shevchenko et al., 2007). Each gel lane was cut into four pieces, washed twice with 200 µl of 0.04 M NH_4HCO_3 / 50% Acetonitrile (ACN) and dehydrated with 200 µl 100% ACN. The gel pieces were then rehydrated in 200 µl of 20 mM DTT and dehydrated again and then rehydrated with 100 µl 55 mM Iodoacetamide for 20 min in the dark at room temperature. After washing twice with 100 µl 100 mM NH_4HCO_3 and dehydrated, the gel pieces were rehydrated with 30 µl of 0.02 µg/µl of trypsin (Promega V5111) solution was added to the gel pieces for 20 min, followed by the addition of 60 µl solution containing 40 mM NH_4HCO_3 / 10% ACN to completely cover the gel pieces, and the samples were incubated at 37°C for 18 hrs. Peptides were extracted using 90 µl of ACN followed by 150 µl of 50% ACN / 5% HCOOH at 37 °C for 15 min.

Table 3: Reagents (Study I)

Reagents	Vendors	Study
Zeocin	Invitrogen V1020-20	I
Biotin-phenol	Iris-Biotech, CAS no.: 41994-02-9	I
Hydrogen Peroxide	Sigma-Aldrich, cat. no. H1009-100ML	I
Trolox	Sigma-Aldrich, cat. no. 238813-1G	I
Pierce 660 nm protein assay	Thermo Fisher Scientific, cat. no. 22660	I
Octyl glucoside	Sigma-Aldrich, 29836-26-8	I
Streptavidin 633	Thermo Fischer Scientific	I
Protease phosphatase inhibitor	Thermo Fisher Scientific, cat. no. A32961	I
Streptavidin magnetic bead	Pierce, cat. no. 88817	I
SimplyBlue SafeStain	ThermoFisher Scientific, cat. no. LC6065	I
Trypsin	Promega V5111	I
Streptavidin-HRP	Life Technologies, cat. no. S-911	I
WGA	Sigma-Aldrich, L9640	I
HEL	Sigma-Aldrich, 10837059001	I
CellTak	CellTak, Corning®, purchased from Sigma-Aldrich DLW354242	I

Table 4: Instruments (Study I)

Instruments	Vendors	Study
AMAXA electroporation machine	Biosystem	I
Sony SH800 Cell Sorter	Sony	I
nanoflow HPLC system	Easy-nLC1200, ThermoFisher Scientific	I
Orbitrap Fusion Lumos mass spectrometer	Thermo Fisher Scientific	I
AiryScan confocal microscopy	Zeiss	I
Spinning disc confocal microscope	3i	I

Table 5: Antibodies (Study I)

Antibodies	Vendors	Study
Anti-alpha-Tubulin AF488/-647	Merc Sigma-Aldrich 16-232 / 05-829-AF647	I
Anti-Golga3		I
Anti-HEL antibody	Prof Facundo Batista	I
Anti-mouse IgG1 AF488/-647	Jackson ImmunoResearch 115-545-205 / 115-605-205	I
Anti-Rabbit IgG AF555	Invitrogen A-31572	I
DAPI	Thermo FisherScientific 00495952	I
$F(ab')_2$ Fragments of donkey anti-mouse IgM	Jackson ImmunoResearch 715-546-020 / 715-606-020	I
Goat anti-mouse IgM $F(ab')_2$	Jackson ImmunoResearch 115-006-020	I
Goat anti-mouse IgM $F(ab')_2$	Jackson ImmunoResearch 115-006-020	I

Table 6 DNA constructs (Study I)

Plasmid	Source	Study
pcDNA3-mito-APEX	Alice Ting (Addgene plasmid # 42607)	I
pcDNA™4/TO	Invitrogen V1020-20	I

Table 7: Cell Lines and cell culture

	Organism	Source	Culture condition	Study
A20 D13	mouse	Prof Facundo Batista	RPMI (cRPMI; RPMI 1640 with 2.05 mM L-glutamine supplemented with 10% fetal calf serum (FCS), 4 mM L-glutamine, 50 µM β-mercaptoethanol, 10 mM HEPES and 100 U/ml penicillin/streptomycin	I
Raji B cell	human	Prof Facundo Batista	cRPMI (RPMI 1640 with 2.05 mM L-glutamine supplemented with 10% FCS, 4 mM L-glutamine and 100 U/ml penicillin/streptomycin	I

Mass spectrometry analysis

Data acquisition was conducted via LC-ESI-MS/MS using a nanoflow HPLC system (Easy-nLC1200, ThermoFisher Scientific) coupled to the Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nano-electrospray ionisation source. Peptides were initially loaded on a trapping column and subsequently separated inline on a 15 cm C18 column. Elution was performed using a linear 20 min gradient from 8 to 39% was used to elute peptides. MS data was acquired using Thermo Xcalibur 3.1 software (Thermo Fisher Scientific). A data-dependent acquisition method comprising an Orbitrap MS survey scan of mass range 300-2000 m/z followed by HCD fragmentation was used.

Protein Identification

The raw MS data were processed using MaxQuant software version 1.6.0.1 (Cox & Mann, 2008). MS/MS spectra were searched against mouse UniProt (reviewed (Swiss-Prot), September 2019 release) using Andromeda search engine (Cox et al., 2011). The Maxquant search configuration was set to digest with trypsin, allow set of 2 maximum number of missed cleavages. Carbamidomethyl was set as fixed modification, while N-terminal acetylation and methionine oxidation were set as variable modifications. Peptide and protein false discovery rate was set to 0.01. Match between runs was enabled. MaxLFQ, which enables the determination of relative intensity values of proteins and normalises proteins intensity between samples, was enabled but not used in downstream analysis (Cox et al., 2014). After MaxQuant run, 2526 proteins were identified, from which contaminants and reverse hits were removed. For further analysis, only the proteins identified with at least 2 unique peptides were considered as identified (1677 proteins). The identified proteins were then classified using both KEGG pathway analysis (Kanehisa et al., 2016) and gene ontology (GO) classification (Ashburner et al., 2000). The proteomic dataset generated in this work will be submitted to PRIDE (Vizcaíno et al., 2016).

Proteomics and differential enrichment analysis

Normalisation and differential enrichment analysis were conducted using NormalyzerDE tools in Bioconductor (Willforss et al., 2019). Quantile normalisation was selected as the best normalisation method following a comparison of various normalisation methods available NormalyzerDE

package (data not shown). Prior to normalisation and differential enrichment analysis, identified proteins with missing values ≥7 out of 18 conditions were filtered out. For the remaining proteins, missing value imputation was done using k-Nearest Neighbor (kNN) Imputation. Differential enrichment analysis was done using NormalyzerDE with statistical comparison method set to limma, logTrans set to FALSE, leastRepCount set to 1 and sigThresType set to FDR (Benjamini-Hochberg corrected p-values).

To identify potential raft-resident proteins, we applied a strategy based on work of Paek et al., 2017 (Paek et al., 2017). We first selected proteins with ≥1.5 $\log_2$ fold-change in the non-activated biotinylated sample compared to control samples (sample without H_2O_2 triggered biotinylation). Proteins that showed $\log_2$ foldchange ≥ 1 in surrogate antigen-stimulated samples compared to non-activated samples were filtered out. The list of the putative raft-resident proteins were then cross-referenced with previously published mammalian lipid raft proteins in the RaftProt database (https://raftprot.org/) (Mohamed et al., 2019).

Bioinformatics analysis

Subsequent analysis were conducted in R. Volcano plots were generated using the EnhancedVolcano package (Blighe et al., 2018). Enrichment analysis was performed using the R clusterProfiler package (Yu et al., 2012). An UpSet plot was constructed to show the intersection between conditions, and a Venn diagram was constructed to depict the overplap of proteins significantly enriched upon BCR cross-linking at different time points.

Table 8: Data analysis and Bioinformatics tools

Bioinformatic tools/softwares/packages	Function	Version	Link	Study
BASH	Unix shell and command language for scripting and automation	4.4	https://www.gnu.org/software/bash/	II
Biorender	Illustration software for creating scientific figures	Not Applicable	https://biorender.com	I
clusterProfiler	R package for functional enrichment analysis	3.16	https://bioconductor.org/packages/clusterProfiler/	I
Cytoscape	Software for network visualization and analysis	3.7	https://cytoscape.org	I & II
DAVID	Functional annotation and enrichment analysis	6.8	https://david.ncifcrf.gov	I & II
EnhancedVolcano	R package for volcano plot visualization	1.8.0	https://github.com/kevinblighe/EnhancedVolcano	I
Fiji ImageJ	Image processing and analysis software	2 2.0.	https://fiji.sc	I
Gene Ontology	Ontology for biological processes, functions, and components	Accessed 2019	http://geneontology.org	II
GitHub	Platform for version control and collaborative coding		https://github.com	II
GraphPad Prism 6	Statistical analysis and graphing software	6.07	https://www.graphpad.com	I
Inkscape	Vector graphics editor	0.92	https://inkscape.org	I & II
KEGG	Pathway database for molecular interaction and reaction networks	Accessed 2019	https://www.genome.jp/kegg/	I
MaxQuant	Software for quantitative proteomics analysis	1.6.0.1	https://www.maxquant.org	I
NormalyzerDE	R package for normalization and evaluation of omics data	1.6.0	https://github.com/ctlab/normalyzerde	I
R	Statistical computing and graphics environment	3.6.3	https://www.r-project.org	I & II
Raftprot	Database for lipid raft proteomics	v2 (accessed 201	http://raftprot.org	I
UniProt	Protein sequence and functional information database	Accessed 09/201	https://www.uniprot.org	I

Western blotting

For western blotting, Raft-APEX2 A20 D1.3 cells were starved for 20 min in serum-free media and then treated with either 10 µg/ml of goat anti-mouse IgM F(ab')$_2$ fragments (Jackson ImmunoResearch 115-006-020) for 10 min or 0.1 mM, 1 mM and 10 mM H_2O_2 for 1 min. After which cells were lysed with 2x SDS PAGE loading buffer, sonicated, and then subjected to SDS-PAGE and Western blotting, with probing using streptavidin-HRP (Life Technologies, cat. no. S-911).

Rshiny development

To enhance the usability and accessibility of AutoCoEv's complex output data, we developed a Shiny application that facilitates interactive visualization and exploration of predicted co-evolutionary relationships among proteins. The Shiny app was developed using the R Shiny framework, using packages visNetwork (Almende & Benoit, 2022) ,

ggplot2 (H. Wickham, 2016), and dplyr (R. Wickham et al., 2020) packages.

5 RESULTS AND DISCUSSION

5.1 Identification of lipid raft associated regulators of BCR signaling (Study I).

5.1.1 Results: Proximity labeling technique

Over the last decades, prior to the widespread adoption of proximity labeling techniques, AP was among the most commonly employed methods for studying protein-protein interactions (PPI). Despite its extensive use, AP has some drawbacks, particularly its ineffectiveness in capturing weak or transient interacting proteins (Qin et al., 2021). Consequently, studies using solely AP often yield incomplete interactomes, particularly in dynamic signaling systems. This limitation has hindered a comprehensive understanding of the intricate relationships and communications among proteins and biological processes they regulate. The advent of proximity labeling techniques era, such as APEX2, BioID and their variants, help overcome several of these limitations by enabling the labeling of proteins in the immediate molecular environment of a bait in living cells. In this study, we employed the APEX2 proximity labeling system to investigate BCR interactomes and signaling dynamics.

To study BCR interactomes using APEX2 proximity labeling technique, the most natural primary target (bait) would be BCR itself i.e., fusing APEX2 to the cytoplasmic side of BCR, namely Igα/Igβ signaling components. In previous studies cyan fluorescent protein (CFP) has been fused to Igα components of BCR (Sohn et al., 2006). Targeting BCR component was exactly our initial aim, however, multiple attempts to express functional APEX2 fused to Igα or Igβ were not successful. The likely cause was that APEX2 interfered with the proper folding or functional activity of these important BCR components, leading to the donwregulation of the fusion proteins (Igα-APEX2 or Igβ-APEX2). Subsequently, we explored fusing APEX2 to a non-BCR component, Lyn, a Src family tyrosine kinase that is localized in the lipid raft domain of plasma

membrane, but this approach also did not lead to succesful expression of the fusion protein in B cells.

Given these challenges, we opted to fuse APEX2 to a small amino acid sequence instead of a full-length protein. We chose amino acid sequence that could target APEX2 to the lipid raft domains (Raft-APEX2) (Fig. 6A, 6B and Pub I; fig 1A) of B cell plasma membrane. This approach was chosen based on the previous studies firmly indicating that BCR translocates to the lipid raft domains upon activation and that the rafts serve as the major signaling site for the BCR (Cheng et al., 1999; Sohn et al., 2006, 2008a). For this reason, we directed APEX2 to the lipid rafts using specific 7 amino acid lipidation sequence, MGCVCSS (Fig. 7A and B). We then transfected the Raft-APEX2 construct into A20D1.3 B cell line. The MGCVCSS sequence is part of the NH2-terminus of Lck and have previously been used to direct proteins to the lipid raft domain (Yasuda et al., 2000). This sequence contains myristoylation and palmitoylation sites, which are crucial for targeting APEX2 to the plasma membrane of B cells. Myristoylation at the glycine (Gly-2) residue facilitates initial membrane association, while palmitoylation at the 2 cysteine (Cys-3 and Cys-5) residues enhances stable localization within lipid rafts, ensuring proper localization to the site of BCR signaling.

After cloning Raft-APEX2, we sought to validate the localization of Raft-APEX2 using both microscopic and flow cytometry techniques. Outcome of these analyses demonstrated that Raft-APEX2 is, as expected, localized to the plasma membrane and enriched in the lipid raft domains (Study I, Fig. 1B and C).

We then tested if the expression of Raft-APEX2 would have any influence in the antigen-induced signaling. We compared the quantity of the surface expressed BCR and level of tyrosine phosphorylation initiated upon antigen stimulation for both Raft-APEX2 and WT A20D13. We found similar level of IgM BCR expression on the surface of both cells. Importantly, anti-phospho-tyrosine probe also showed similar level of phosphorylation in Raft-APEX2 expressing cells as the WT A20D13 (Study I, Fig. 1D and Suppl. Fig. S1E and F).

5.1.1.1 Assesment of the effects of H_2O_2 on the BCR sinalling cascade

Given that APEX2 requires H_2O_2 to initiate its biotinylation reaction, we assessed whether the concentration of H_2O_2 applied (1 mM for 1 min) could inadvertently trigger B cell activation. This concern stemmed from previous studies suggesting that H_2O_2 can activate BCR signaling through inhibition

of phosphatases (Reth, 2002; Wienands et al., 1996). To address this, we treated Raft-APEX2 cells with varying concentrations of H_2O_2 (0.1 mM, 1 mM and 10 mM) and evaluated the overall protein phosphorylation level. The results indicated that 1mM of H_2O_2 for 1 min did not induce significant BCR signaling (Study I, Fig. 1E and Suppl. Fig. S1E). This findings suggest that the conditions for APEX2-mediated biotinylation in this study are unlikely to cause unintended activation of BCR-related signaling pathway.

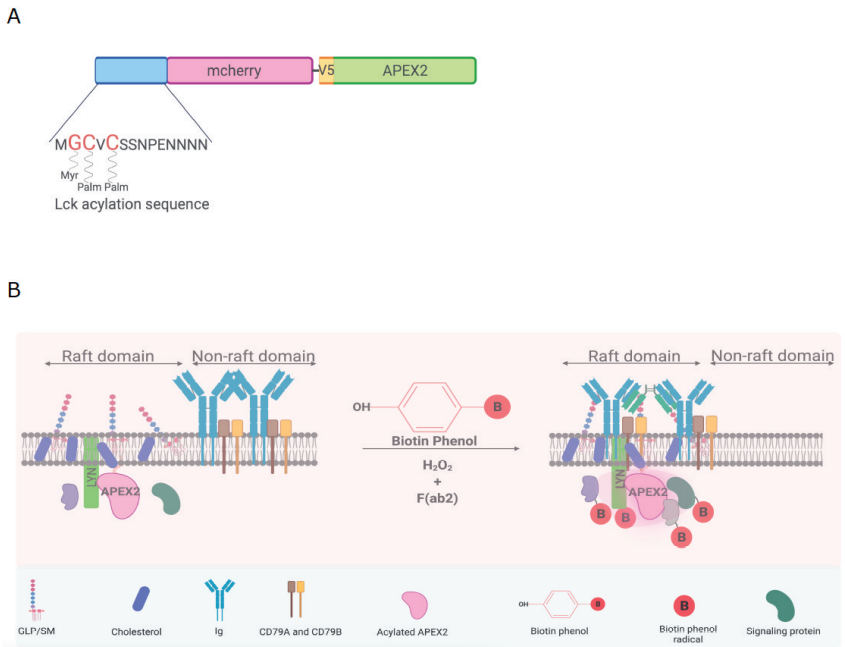

Figure 7: APEX2 targeting to the lipid raft domain of B cell plasma membrane. (a) Schematic illustration of the Raft-APEX2 construct composing of NH2-terminus of Lck (MGCVCSS sequence), mCherry, V5-tag and APEX2. (b) A schematic illustration of Raft-APEX2 initiating biotinylation reaction in the lipid raft domain upon addition of biotin-phenol and H_2O_2. Left: BCR localized in the non-raft domain of the B cell plasma membrane, positioned at in the non-raft domain away from the raft-localized APEX2. Right: BCR localized within the lipid raft domain of the B cell plasma membrane, sharing the same domain as Raft-APEX2.

5.1.1.2 Exploratory data analysis

Following mass spectrometry analysis and filtering to retain proteins identified with at least two unique peptides, a total of 1,677 proteins were quantified and used for downstream statistical analyses.

Exploratory data analysis was conducted on the data to assess sample grouping, identify outliers and evaluate the proportion and types of missing values. Missing values evaluation revealed a mixture of mixing value types. Consequently, k-nearest neighbours imputation method (Lazar et al., 2016; Troyanskaya et al., 2001) was deemed appropriate to address these missing data. Prior to imputation, a filtration step was implemented, in which rows with missing values ≥ 7 out of 18 samples were removed.

Principal component analysis (PCA) was then applied to visualize the variance among the samples in relation to the sample groups (Fig. 8). The first principal component (PCA1) accounted for variance between 15min samples (both activated and non-activated) and the rest of the samples, while the second principal component (PCA2) separated activated and non-activated samples (Fig. 8).

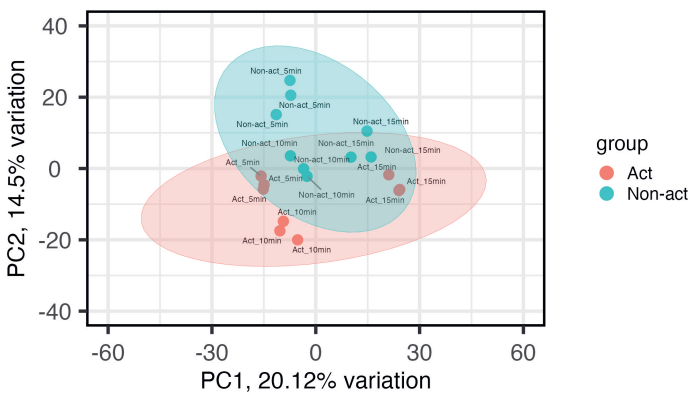

Figure 8: PCA plot illustrating the grouping of activated (Act) and non-activated (Non-act) samples during the proximity biotinylation step. For proximity biotinylation, 10 × 10^6 A20 D1.3 raft-APEX2 cells were treated with 500 µM biotin-phenol for 45 minutes, followed by activation with 0(Non-act) or 10 µg/ml (Act) of anti-IgM for 5, 10, or 15 minutes. Biotinylation was initiated with 1 mM H_2O_2 for 1 minute, and the reaction was quenched with a quenching solution (see more details in Section 3, Materials and Methods).

5.1.1.3 BCR proteome identification and dynamics

The APEX2 study was designed to identify novel proteins involved in BCR signaling, and to monitor the dynamics of proteins associated with key biological processes triggered by BCR activation.

On one hand, we focused on identifying proteins that exhibit specific enrichment or exclusion in activated or non-activated conditions, which may suggest a potential role as a novel regulator. On the other hand, we analysed the temporal changes in the recruitment of different proteins within the lipid raft region across different time points. This study provideed real-time insights into the behavior of proteins within the lipid raft vicinity during 5-, 10-, and 15-min intervals of BCR stimulation. Additionally, we predicted the resident proteins within the lipid raft membrane domains in B cells.

5.1.1.3.1 BCR components were among top differentially enriched proteins.

We reasoned that proteins significantly enriched in the Raft-APEX2 proximity upon BCR antigen activation may be functionally associated with BCR signaling. To identify such proteins, we used the LIMMA method, a univariate method statistical approach (Ritchie et al., 2015), to indentify proteins exhibiting dynamic behaviour, such as recruitment or reduction, near the lipid rafts upon BCR stimulation.
We focused to compare the raft-proximal proteins in activated versus non-activated cells at each time point. Notably, the topmost significant differentially expressed proteins between either 5-, 10-, 15min activated Raft-APEX2 samples and the non-activated samples included the key BCR components (CD79b, Ighm, Igkv and CD79a) (Study I, Fig. 5A and C). These findings confirm the translocation of the BCR to the the lipid rafts region, thereby validating the use of Raft-APEX2 as a tool for monitoring BCR signaling proteins.

5.1.1.3.2 BCR proteome identification

Raft-APEX2 provided us an opportunity to derive unprecedented information about the identity of molecules in the vicinity of lipid rafts and, upon BCR activation, to observe the dynamic recruitment of signaling components to these domains. In this study, we identified 1,677 high

confidence proteins that, according to the reported activity of APEX2, were present within 20 nm of the targeted raft region (Study I, Fig. 2C). Among these, 1,143 proteins were identified in all conditions while 48 were exclusively identified in all non-activated samples (Fig. 4A,B; Table S1C,D). Notably, only Golga3 and Kif20A were identified solely in all activated conditions (Study I, Fig. 4A), making them candidates for novel activation-specific regulators in B cells.

To validate Golga3, we performed immunofluorescence analysis in Raji human B cells and observed its widespread vesicular distribution (Study I, Fig. 6A). Upon BCR activation with $F(ab')_2$ anti-IgM, Golga3 vesicles became more prominent at the plasma membrane, with a significant increase in vesicle size and intensity (Study I, Fig. 6A, B and Suppl. Fig.S4B). Colocalization analysis revealed partial overlap with internalized antigen, as indicated by a weak Pearson's correlation coefficient (0.21) but higher Manders' overlap coefficients (M1: 0.61, M2: 0.56) (Study I, Fig. 6B), suggesting Golga3's involvement in antigen trafficking.

In contrast, Kif20A could not be validated due to the lack of suitable antibodies, preventing further functional assessment. Nonetheless, its exclusive presence in the activated proteome warrants future study, pending the availability of specific molecular tools.

5.1.1.3.3 Characterization of Lipid Raft-Associated Proteins in B cells

As previously discussed, BCR rapidly translocates to the lipid rafts upon activation, thus positioning it as a critical hub for BCR signaling and internalization. Traditionally, the identification of proteins localized to the lipid rafts has been through the isolation of detergent-resistant regions of the plasma membrane using sucrose gradient ultracentrifugation method (Saeki et al., 2003).

In this thesis, we leveraged the Raft-APEX2, specifically targeted to the lipid raft domain of B cell plasma membrane, to expand our current knowledge of lipid raft domain associated proteins. We employed a strategy based on the work of Paek et al. (2017) to identify potential raft-resident proteins. We first selected proteins with a log2 fold-change of ≥ 1.5 in the non-activated, biotinylated samples compared to control samples (i.e., samples without H_2O_2-triggered biotinylation). We then filtered out proteins showing a log2 fold-change of ≥ 1 in surrogate antigen-stimulated samples compared to non-activated samples. This approach enabled us to focus solely on proteins that are consistently associated with the lipid rafts under

basal conditions. Utilizing this strategy, we identified 346 lipid raft resident proteins that were indicated to be resident within the lipid rafts of B cells (Study I, Fig. 2C and 3A).

To validate our findings, we compared the identified lipid raft proteins with those cataloged in RaftProt database (Mohamed et al., 2019). Importantly, approximately 90% of our proposed raft resident proteins overlapped with those in RaftProt database, while the remaining 10% represented novel addition to the lipid raft proteome (Study 1, Fig. 3A).

These findings significantly contribute to our understanding of lipid raft proteome in B lymphocytes, providing new insights into the protein landscape within these critical signaling microdomains.

5.1.1.4 Monitoring of pathway dynamics upon BCR activation

5.1.1.4.1 Dynamics of BCR signaling proteins in the lipid rafts following antigen stimulation

Our analysis revealed distinct patterns in the behaviour of IgM-BCR components, including Ighm, IgK-V, Igα and Igβ across different time points following antigen stimulation.

The Ighm, and IgK-V, the heavy and light chain part of IgM-BCR, respectively, exhibited significant enrichment in the lipid raft region in all the 3 time points post stimulation (Study I, Fig. 5A and C). This observation aligns with previous report that BCRs translocate to the lipid rafts domain of B cell upon stimulation with antigens. In contrast, the cytoplasmic components of BCR, Igα and Igβ displayed significantly reduced enrichment in the vicinity of lipid rafts upon stimulation with antigens (Study I, Fig. 5A and C). This reduction may be attributed to the destabilization or dissociation of BCR components upon antigen stimulation, as previously reported (Vilen et al., 1999). It could suggest separation of the extracellular and cytoplasmic parts of BCR upon stimulation that could lead to different components being enriched in different microdomains of plasma membrane. Another possible explanation for the decreased enrichment of Igα and Igβ upon antigen stimulation could be steric obstruction of Igα and Igβ from active phenolic radical, where

receptor activation-induced protein-protein interactions prevent Igα and Igβ from being effectively biotinylated by active phenolic radical.

Notably, majority of the known BCR signaling molecules did not exhibit significant dynamic changes upon antigen stimulation (Study I, Fig. 5A and C). This finding supports the notion that post-translational modifications, such as phosphorylation that occurs following BCR engagement, are primary drivers of the signaling, rather than protein translocation.

Protein tyrosine kinases (PTK) are essential components of biological signaling pathways and BCR signaling pathway is not an exception. PTKs catalyze the transfer of γ phosphate of ATP to tyrosine residues on protein substrates, thereby, creating a binding site that serves to recruit other molecules (Hubbard & Till, 2000). Some of the PTKs identified in this analysis that are known to play roles in B cell signaling pathways include Ptk2b, Fyn, Lyn, and Btk. Among these, Fyn remained unchanged upon BCR stimulation while Ptk2b showed increased recruitment at 5 min post-stimulation but diminished at 10- and 15-min time point. Conversely, Lyn and Btk displayed diminished enrichment at all time points post-stimulation (Study I, Fig. 5A and C). The observed behavior of Lyn and Btk may be linked to their dynamic interactions with the BCR during activation. This aligns with previous findings suggesting that Lyn's proximity to the BCR is greatest shortly after activation (within the first few minutes), before this interaction declines (Sohn et al., 2008). Such transient behavior may reflect Lyn's complex role as both a positive and negative regulator of BCR signaling, as also reported in earlier studies (Xu et al., 2005).

This analysis underscores the complex dynamics of BCR components and signaling molecules within lipid rafts and highlights the importance of spatial and temporal regulation in BCR-mediated signaling.

5.1.1.4.2 Dynamics of proteins involved in endocytosis

Endocytosis is one of the biological processes that is upregulated following BCR antigen stimulation (Kuokkanen et al., 2015). It is needed in order to internalize the antigen for further processing, vital for the subsequent peptide-antigen presentation and efficient immune response (Kuokkanen et al., 2015; Kwak et al., 2019). Generally, endocytosis occurs primarily via two main mechanism: the clathrin-mediated endocytosis and the clathrin independent endocytosis. Among these, the clathrin-mediated endocytosis is the main pathway through which BCR internalize antigen (Busman-Sahay et al., 2013; Hoogeboom & Tolar, 2016; Natkanski et al., 2013; Stoddart et al.,

2005). As expected, our analysis detected clathrin-mediated endocytosis-linked proteins in the vicinity of the lipid rafts. Some of these proteins exhibited distint dynamic changes upon BCR stimulation. For instance, Clta showed significant enrichment at 5min post-activation, while Cltb remained unchanged at all time points (Fig. 9). Additionally, Ap2a1 was recruited to the lipid raft region at 5 and 10min post-stimulation and appeared diminished from the lipid raft region at the 15min time point. In contrast, Ap2b1 was enriched at both 5min and 15min of BCR stimulation (Fig. 9). Clathrin and AP-2 complex components are well-established mediators of clathrin-mediated endocytosis, a process known to be involved in BCR internalization following antigen engagement. The transient enrichment of Clta at early time points after stimulation is consistent with rapid recruitment of clathrin machinery during the initial phases of BCR internalization, whereas the stable levels of Cltb may reflect isoform-specific roles or constitutive association. Similarly, the early recruitment of Ap2a1 at 5 and 10 min post-stimulation likely reflects activation-induced assembly of the AP-2 adaptor complex at the plasma membrane, while its subsequent reduction at 15 min may indicate progression of endocytic trafficking away from the membrane. In contrast, sustained enrichment of Ap2b1 at later time points may suggest prolonged involvement in vesicle maturation or cargo sorting during BCR internalization.

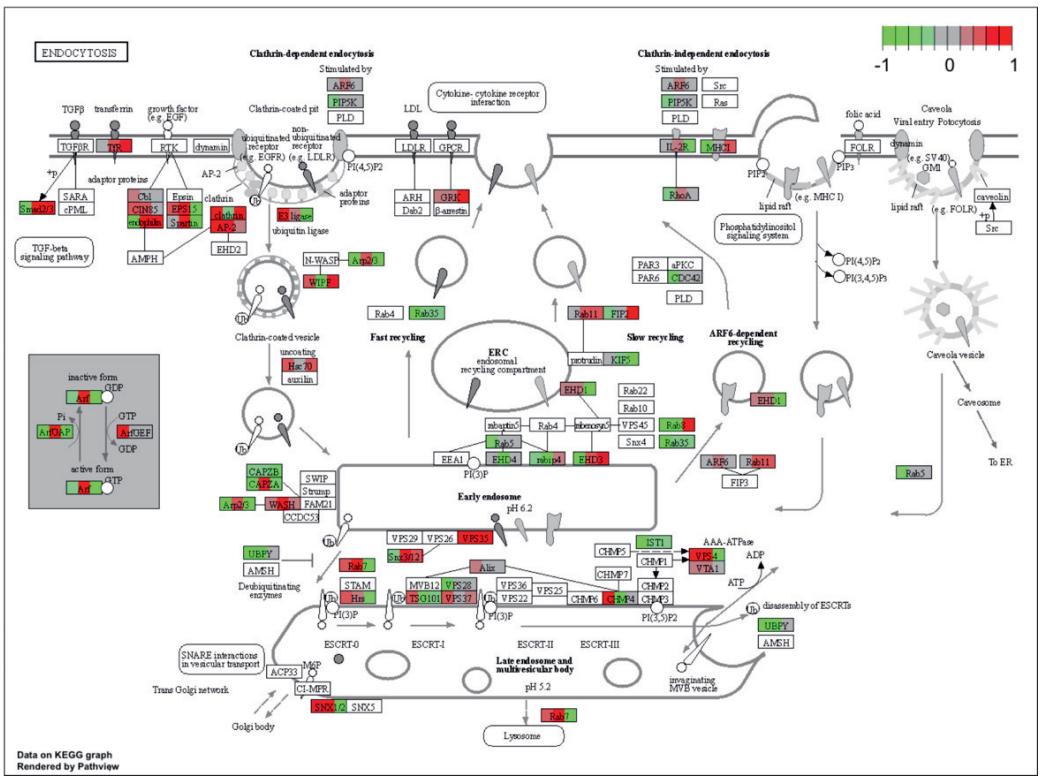

Figure 9: Temporal recruitment of proteins linked to endocytosis in lipid rafts following BCR stimulation: This visualization depicts the dynamic changes in KEGG (Kyoto Encyclopedia of Genes and Genomes) endocytosis pathway components within the lipid raft vicinity upon BCR antigen stimulation at 5, 10 and 15min timepoint. The fold changes have been normalized to a scale of -1 to +1. Each gene is represented by a rectangular block divided into three segments corresponding to 5, 10, and 15-minutes. The visualization was generated using the ClusterProfiler package in R, which enabled pathway mapping and integration of temporal proteomic data with KEGG pathway annotations.

5.1.1.4.3 The role of SUMO1 in BCR signaling and regulation.

Protein clustering analysis using an unsupervised K-means machine learning algorithm grouped Xpot, RanGAP1 (Ran GTPase-activating protein 1), and SUMO1 into the same cluster (Study I, Suppl. Fig. S5). Clustering was based on the fold changes of differentially expressed proteins identified in our proteomics analysis. RanBP2, a key component of

the nuclear pore complex, forms a functional subcomplex with RanGAP1, SUMO1, and Ubc9, SUMO-conjugating enzyme (Hampoelz et al., 2019; He et al., 2021). Notably, our results indicated that both Sumo1 and RanGAP1 became enriched at 15 min post-BCR activation (Study I, Suppl. Fig. S5). This is consistent with previous findings linking TCR to the phosphorylation-dependent SUMOylating of RanGAP1 (He et al., 2021).

Based on our unsupervised machine learning findings, which clustered RanGAP1 and SUMO1 together, we next investigated the role of SUMOylation in BCR signaling (Study I, Suppl. Fig. S5). We explored its role in B cell activation using microscopic and bioinformatics methods. Our analysis showed that SUMO1 was strongly concentrated in the nucleus before BCR stimulation, reflecting its known nuclear functions. However, at early activation stages (5–15 min), SUMO1 colocalized with IgM-BCR clusters at the cell periphery, suggesting an association between BCR and SUMO1 (Study I, Fig. 6A). By 30 minutes, this colocalization diminished as internalized antigen accumulated in processing compartments, thereby suggesting transient association between SUMO1 and BCR (Study I, Fig. 6A). Also, surface stimulation of BCR using antigen-coated beads, which mimic IS formation, resulted in punctate SUMO1 accumulation at the bead interface alongside BCRs, a pattern absent with non-activatory beads (Study I, Fig. 6B).

Additionally, we also assessed the fuctional role of SUMO1 using an inhibitor, TAK981, a selective inhibitor of the SUMO-activating enzyme (SAE) complex. TAK981 effectively eliminated SUMOylation patterns in immunoblots (Study I, Fig. 7B).

Further analysis of downstream BCR signaling under SUMOylation inhibition in both primary mouse B cells and A20 D1.3 cells, using soluble and surface-tethered antigen activation showed a significant reduction in phosphorylated AKT (pAKT) and ERK1/2 (pERK1/2) in primary B cells under both activation conditions, whereas Syk phosphorylation remained unaffected (Study I, Fig. 7B and C), suggesting that SUMOylation influences downstream signaling rather than initial activation.

These findings suggest that early BCR stimulation influences SUMOylation dynamics, which may contribute to the regulation of signaling events during BCR activation and immune responses.

5.1.2 Discussion

The APEX2 study was strategically designed with two primary objectives: the identification of novel proteins involved in BCR signaling and the examination of protein dynamics related to key biological processes, such as BCR signaling and endocytosis, under varying conditions and at different stages of BCR activation. This dual focus allowed us to uncover new insights into the molecular mechanisms underpinning BCR activation, particularly the recruitment and exclusion of proteins within lipid raft domains during different stages of antigen stimulation. A key biological insight from this study is the identification of previously unappreciated proteins and processes associated with BCR activation within lipid raft domains. In addition to confirming known components of BCR signaling and endocytosis, our analysis revealed several proteins not previously linked to BCR function, including Golga3, Kif20A, and components of the SUMOylation machinery. The activation-dependent recruitment of these proteins suggests that BCR signaling engages a broader network of membrane trafficking, cytoskeletal regulation, and post-translational modification pathways than previously appreciated.

Our results highlight the potential of APEX2 proximity labeling to reveal novel regulators of BCR signaling. With APEX2, we can overcome the limitations of traditional AP techniques, particularly their inability to capture weak or transient PPIs. Incomplete identification of interacting partners by AP techniques can impede a comprehensive understanding of complex signaling networks like those in BCR signaling. In these networks, the direct, indirect, transient and stable interactions are crucial. Therefore, the use of proximity labeling techniques, such as APEX2, represents a significant advancement. APEX2 enables the labeling of proximal proteins and biomolecules within a small spatial range (~20 nm) in real time. This is particularly advantageous for studying dynamic signaling events, such as those triggered by BCR activation.

We also highlight the challenges that could be encountered expressing APEX2 fused proteins as our attempts to express APEX2 fused to BCR components (Igα and Igβ) and Lyn were unsuccessful, suggesting that APEX2 might interfere with the proper folding or function of these proteins, leading to their downregulation in B cells. This outcome highlights a limitation of using large tags or enzymes like APEX2, which may disrupt the native structure or function of the target protein. It also reinforces the notion that proximity labeling tools, while powerful, require careful consideration of the target protein's structure and function to avoid mislocalization or altered expression. Nonetheless, our strategy of anchoring

APEX2 to the lipid raft domain using a short lipidation sequence proved effective and did not disrupt BCR expression or signaling. This validates the utility of Raft-APEX2 for studying signaling processes within the membrane microdomains.

One important consideration when using APEX2 is its reliance on H_2O_2 to initiate labeling. As H_2O_2 can modulate signaling pathways through phosphatase inhibition, we carefully tested the concentration used. Our results confirmed that the labeling conditions (1 mM H_2O_2 for one min) did not trigger unintended activation, supporting the compatibility of APEX2 with BCR signaling studies when properly controlled.

The identification of over 1,600 high-confidence proteins near the lipid raft region across different activation stages reveals the depth of the BCR-associated proteome. Notably, we identified proteins like Golga3 and Kif20A exclusively in activated conditions, implicating Golgi-derived vesicular trafficking and microtubule-based transport in BCR signaling respectively. The selective enrichment of Golga3 upon activation, together with its colocalization with internalized antigen, suggests a potential role in coordinating post-endocytic trafficking of BCR-containing vesicles. While Kif20A remains unvalidated due to lack of suitable antibodies, its known role in vesicle transport and cytoskeletal organization raises the possibility that it contributes to spatial organization of signaling compartments following BCR activation (Moon., 2024).

In addition, we observed strong enrichment of the Q-SNARE protein Vti1b, which was further investigated for its potential role in BCR signaling and vesicle trafficking (Music et al., 2022a). Although Vti1b partialy colocalized with antigen following BCR activation with soluble antigen, functional analyses in Vti1b-deficient mice revealed no impairment in BCR signaling, immunological synapse formation, or antigen processing and presentation (Music et al., 2022b). This suggests that not all proteins recruited to the BCR microenvironment directly regulate signal transduction.

In addition to novel hits, we identified 346 proteins potentially resident in the lipid raft domain, greatly expanding our knowledge of the B cell raft proteome. These proteins may serve as platforms or modulators of signaling responses during B cell activation.

Our investigation into the dynamics of proteins involved in BCR signaling and endocytosis further revealed distinct patterns of behavior for various BCR components. For instance, IgM-BCR components such as Ighm and IgK-V showed significant enrichment in the lipid raft region across all time points post-stimulation. In contrast, the cytoplasmic components Igα and Igβ exhibited reduced enrichment, suggesting possible

dissociation or steric obstruction following antigen stimulation. This highlights that BCR signaling dynamics are not solely driven by changes in protein localization but also regulatory mechanisms such as post-translational modifications. It also indicated that BCR may not always be one entity. The enrichment of clathrin-mediated and endocytosis proteins BCR signaling and endocytosis. These findings open up new possibilities for understanding the interplay between BCR signaling and other cellular pathways.

One of the most striking findings of this study is the activation-dependent recruitment of SUMOylation machinery to the BCR signaling environment.We observed significant changes in the dynamics of proteins associated with the nuclear pore complex (NPC) and nuclear transport receptors, particularly the RanBP2 subcomplex. The enrichment of proteins like Sumo1 and RanGAP1 at 15 minutes post-BCR activation suggests early reorganization of the nuclear transport machinery and SUMOylation during BCR activation, which could play a crucial role in mediating the interplay between nuclear and cytoplasmic signaling during immune responses. These observations, supported by machine learning clustering, suggest early crosstalk between nuclear transport and surface signaling. The colocalization of SUMO1 with BCR clusters, and its recruitment to the immunological synapse upon antigen-bead stimulation, indicates a possible role in trafficking or synapse organization. Inhibition of SUMOylation impaired cytoskeletal remodeling and reduced downstream signaling via AKT and ERK, implicating SUMOylation as a modulator of signal propagation beyond the initial activation step. This findings identifies SUMOylation as a previously unrecognized regulator of BCR signaling.

Overall, this study provides a comprehensive overview of the BCR-associated proteome within lipid raft domain. By revealing novel protein candidates and post-translational regulatory mechanisms, including SUMOylation. Our findings expand the current understanding of BCR signaling and also emphasize the importance of spatial and temporal regulation in BCR-mediated signaling and highlight the potential of the Raft-APEX2 approach for studying complex signaling networks in immune cells.

5.2 A High-Throughput In Silico Pipeline for Predicting Inter-Protein Coevolution (study II)

5.2.1 AutoCoEv

As part of our approaches toward unraveling novel BCR regulators, we developed a computational method, AutoCoEv, to complement our experimental techniques. Computational methods can outwit experimental methods in some instances. Such instances include the ability to simulate biological processes and analyze large datasets much faster than experimental approaches (AlQuraishi, 2019), suitability for high-throughput analysis (McLendon et al., 2008), enable predictions, cost effectiveness, accessibility, and reproducibility (Schreiber et al., 2014; Wilson et al., 2017). While computational biology offers the above benefits, it is important to note that the integration of both computational and experimental approaches is often the most powerful strategy for advancing our understanding of complex biological systems. Combining these approaches can lead to more accurate predictions, better experimental design, and a more comprehensive understanding of biological phenomena. To this end, we developed AutoCoEv, an automated analysis pipeline designed to predict inter-protein coevolution, thus supplementing our experimental approaches to unravel novel BCR regulators.

AutoCoEv was designed to help in selecting hits from large dataset such as the ones generated by proximity labeling protein mass spectrometry approaches. Although APEX2 is efficient for mapping protein interactions, its promiscuous nature makes it insufficient as a proof for viable functional interaction among the identified hits. This makes it imperative to use other techniques for hit selection. For instance, because APEX2 identifies proteins in 20 nm range to bait/protein of interest, a large fraction of the hits generated may not necessarily be directly involved in BCR signaling. One reason is that proximity labeling method is marred with non-specific labeling due to the diffusion of reactive intermediates which can result in potential false positive hits (Martell et al., 2012). Also, some of the hits could just be part of a complex found in the region of BCR and may not necessarily be involved in BCR signaling regulation. To tackle this, and to supplement differential enrichment analysis that was used in selecting hits above (study I), in Study II we developed a computational method (AutoCoEv) as a second line large-scale analysis to aid hit selection for wet lab experiments.

The underlining principle of AutoCoEv is to identify protein pairs that are suggested to interact physically or functionally. The core of AutoCoEv

was built aroud CAPS (Coevolution Analysis using Protein Sequences), a PERL-based software designed to detect co-evolution between amino acid sites. It utilizes BLOSUM-corrected amino acid distances to identify co-variation among sites while utilizing phylogenetic sequence relationships to eliminate phylogenetic and random dependencies. Additionally, CAPS leverages 3D protein structures to determine the nature of interactions between co-evolving amino acid sites (Fares & McNally, 2006b). AutoCoEv pipeline was designed to analyze large-scale protein input as oppose to CAPS that only has the ability to analyze individual pairs of proteins. While there have been previous tools developed for detecting co-evolution within or between proteins, AutoCoEv has the advantage of being able to process large protein inputs in a user friendly manner (Petrov et al., 2022).

AutoCoEv takes in list of Uniprot IDs and species as input. After that, it uses the input information to retrieve orthologues from OrthoDB. Before processing to multiple sequence alignment, AutoCoEv checks the sanity of the sequences by two methods. Firstly, it performs reciprocal BLAST against the reference organism in an instance where there is more than one homologue per species. Secondly, it uses a tool called Guidance to exclude orthologs that are very different from the rest. Performing a quality check of the sequences before multiple sequence alignment (MSA) is critical because CAPS coevolution accuracy is dependent on the quality of alignment. Importantly, AutoCoEV provides three options for MSA; MAFFT (Multiple Alignment using Fast Fourier Transform) PRANK (Probabilistic AligNment Kit) and MUSCLE (MUltiple Sequence Comparison by Log-Expectation) (Study II, Fig. 1A and B). However, with our analysis PRANK seems to provide more reliable coevolution prediction (Study II, Fig.2 and Suppl. 1A). It is also possible to select between trees generated by CAPS and those generated by an external tool, PhyML (Phylogenetic estimation using Maximum Likelihood). The coevolution analysis is done by comparing every base pair to every base pair in each protein pair, which is very time consuming when there are thousands of protein pairs to analyse. In AutoCoEv the speed is optimized by running CAPS2 on individual protein pairs via GNU/Parallel on several cores of the computer simultaneously.

To complement proximity labelling based identification of BCR-associated proteins, we applied AutoCoEv. AutoCoEv was used as an independent strategy to prioritize candidates from the lipid raft associated proteome identified in Study I, with the aim of distinguishing proteins that are merely proximal from those that may be functionally or physically coupled.

Applying AutoCoEv to 324 proteins predicted to localize to lipid raft domains revealed extensive coevolutionary relationships among these candidates. The analysis identified multiple clusters of coevolving proteins, suggestive of coordinated functional modules within the BCR signaling environment (Study II, Fig. 4A and B). Notably, approximately 25% of the predicted coevolving protein pairs overlapped with known or predicted interactions in the STRING database, providing independent support for the biological relevance of the predictions. The remaining coevolving pairs were not previously reported, indicating the ability of AutoCoEv to uncover potentially novel functional associations.

The detected coevolutionary clusters included proteins implicated in membrane trafficking and metabolic regulation, consistent with processes known to accompany BCR activation. These findings suggest that subsets of lipid raft associated proteins have evolved together, indicating that they are likely involved in the same signaling or regulatory pathways.

To facilitate intuitive exploration and interpretation of predicted coevolutionary relationships among proteins obtained from the AutoCoEv pipeline, we developed an interactive web-based application using R Shiny (Fig. 10). The app is automatically generated upon pipeline completion and provides a user- friendly interface for visualizing and inspecting interactions of predicted coevolving proteins, and filtering results based on custom thresholds (e.g., p-values). Key features include interactive network graphs of coevolving protein pairs, sortable and searchable data tables, and downloadable visual outputs. This tool significantly enhances accessibility and biological insight by enabling researchers without programming experience to analyze large-scale coevolution data effectively.

By integrating coevolutionary predictions with proximity labeling data, AutoCoEv enabled prioritization of candidates for downstream experimental validation, reducing reliance on spatial proximity alone. This combined approach strengthened confidence in selected hits and provided an additional layer of evidence for functional coupling within the BCR signaling network.

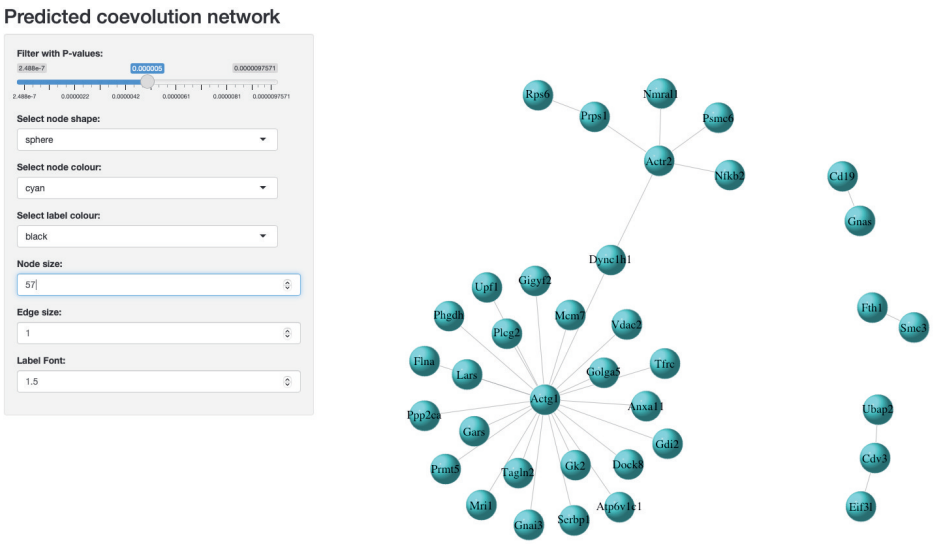

Figure 10: Screenshot of the Shiny app interface showing coevolutionary network visualization and filtering options.

5.2.2 Discussion

AutoCoEv advances the integration of computational methods with experimental approaches for the study of protein interactions. This tool offers a novel way to filter and select potential hits from large datasets, such as those generated by proximity labeling protein mass spectrometry, which, despite its strengths, can yield a substantial number of false positives due to the promiscuous nature of the technique.

The core strength of AutoCoEv lies in its ability to predict co-evolutionary relationships between proteins, offering a complementary method to refining and selecting the results obtained from experimental techniques like APEX2. The use of co-evolution as a criterion is particularly valuable because it implies a functional or physical interaction between protein pairs that has been conserved through evolutionary time. This makes AutoCoEv a tool for narrowing down the list of candidate proteins for further validation.

The application of AutoCoEv in this study revealed that a subset of lipid raft–associated proteins are not merely spatially proximal but appear to be functionally linked through shared evolutionary constraints. This finding suggests that BCR signaling operates within coordinated protein modules rather than as a collection of independently recruited components. Remarkably, about 25% of the proteins predicted to co-evolve were also

found to interact in the STRING database provides independent support for their biological relevance, while the large proportion of previously undocumented associations suggests the existence of uncharacterized functional relationships within the BCR signaling environment. However, it is important to acknowledge that AutoCoEv has not yet been fully validated, and the proportion of true interactions, whether functional or physical, remains uncertain. Additionally, the extent of background noise or false positives generated by the method is unknown. Despite these limitations, the identification of previously undocumented interactions highlights the potential of AutoCoEv to generate new hypotheses for further experimental validation and deepen our understanding of protein network dynamics.

To further test and refine AutoCoEv, its predictions could be validated through experimental approaches such as co-immunoprecipitation, FRET (Förster Resonance Energy Transfer), or crosslinking mass spectrometry. These techniques can provide direct evidence of physical interactions between the proteins identified by AutoCoEv, thereby testing the tool's predictive power. Additionally, comparative studies with other co-evolution detection tools should be conducted to benchmark AutoCoEv's performance and identify areas for improvement.

One area for potential improvement in AutoCoEv is its sequence alignment process. Although it currently supports multiple MSA options such as MAFFT, PRANK, and MUSCLE, optimizing the alignment process is crucial for accurate co-evolution detection. The choice of alignment tool can significantly impact the quality of the predictions, as demonstrated by the observed reliability of PRANK in our analysis. Enhancing AutoCoEv's ability to automatically select the best alignment strategy based on the dataset's characteristics could further improve its accuracy.

Another area for development is the computational efficiency of AutoCoEv. Given the time-intensive nature of co-evolution analysis, especially when dealing with thousands of protein pairs, optimizing the speed of CAPS2 through parallel processing is crucial. Continued refinement of the tool's efficiency will enable it to handle even larger datasets, making it more applicable to high-throughput studies.

In addittion, the development of a ShinyApp for visualization of the predicted protein-protein interactions derived from AutoCoEv analysis represents an important step towards making the tool more user-friendly. This feature not only facilitates the interpretation of complex data but also broadens the accessibility of the tool to researchers who may not have extensive computational expertise.

Overall, the integration of AutoCoEv with experimental proximity labeling demonstrates the value of combining evolutionary analysis with

spatial proteomics. In the context of BCR signaling, this approach revealed that lipid raft associated proteins form evolutionarily constrained functional assemblies, rather than independent components.

5.3 Proteomic analysis of the Immunological Synapse (Study III)

While the APEX2 approach provided us unprecedented information about BCR signaling upon soluble antigen stimulation, we also sought to investigate the BCR interactome upon stimulation with surface-bound antigens. This is imperative because previous studies have suggested that BCR stimulation by surface-bound antigens may be more common in vivo than soluble antigen stimulation (Kuokkanen et al., 2015).

5.3.1 Results: Immunological Synapse Isolation

We developed a pull-down technique to extract proteins forming the immunological synapse. We coated the surface of paramagnetic beads with surrogate antigen (anti-BCR antibodies) to imitate BCR stimulation by APCs. We then used chemical crosslinking to stabilize the immunological synapse formed between primary mouse B cells and the anti-BCR coated beads before the cell bodies were removed by sonication. The remaining footprint of the cell adhesion, or the IS, on the bead surface was extracted. Extracted proteomes were then analyzed using mass spectrometry. We used beads coated with anti-TfR as a control (Fig. 11; Fig. 12).

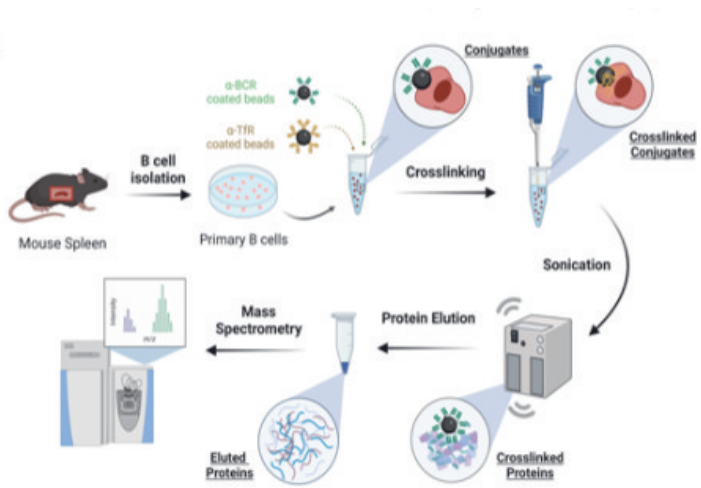

Figure 11: **Isolation of the Primary B Cell Immune Synapse**: A schematic diagram illustrating the IS isolation method. Primary mouse B cells are allowed to form immune synapses on magnetic beads, after which the IS structures are extracted for subsequent proteomic analysis. (Reproduced from Cunha et al., 2023, study III)

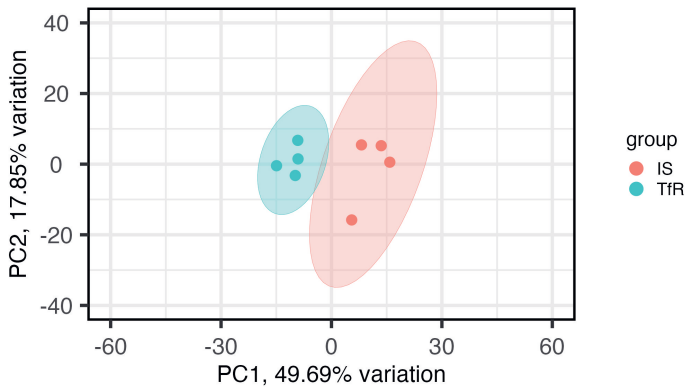

Figure 12: **Principal Component Analysis (PCA) of Proteomic Samples from B Cells Exposed to Anti-BCR (IS) and Anti-TfR Coated Beads**: The PCA plot shows a clear separation between the two groups along the first principal component (PC1), which accounts for 49.69% of the variance. The second principal component (PC2) accounts for 17.85% of the variance. Differential enrichment analysis reveals multiple proteins that are differentially expressed between the groups.

Our proteomic analysis of the isolated IS identified a total of 2,036 proteins. Following stringent filtering to eliminate contaminants and low-confidence identifications, we obtained 1,039 high-confidence proteins. Among these, 661 proteins exclusively associated with the IS, while 365 proteins common to both anti-BCR and anti-TfR.

The 661 IS exclusive proteins are of particular interest as they likely represent the most specific molecular components of the BCR-driven IS. Notably, many known BCR signalling proteins, including Cd79b, Blk, Vav1, and Cd22 etc, were exclusively detected in IS samples (Study III, Fig. 4C), providing strong validation of our isolation approach. However, Cd79a was indetified in both IS and control sample. The divergent behaviour of Cd79b to Cd79a could be linked to different microdomain protein interactions preference. Early evidence already pointed to distinct effector associations for the two cytoplasmic tails (Clark et al., 1992) offering a mechanistic basis for their differential recovery in IS versus control fractions.

To gain insight into proteins common between anti-BCR and anti-TfR samples, we conducted differential enrichment analysis on the 365 common proteins. The results reveal the enrichement of many known components of the BCR signaling pathway such as Syk, Lyn, Vim, Ap2, Dock2, and Itsn2 (Study III, Fig. 4B and C). The enrichment of BCR-specific signaling in the IS samples compared to the control TfR samples confirms the successful capture of BCR-specific signaling proteins demonstrating the method's reliability.

Among other proteins that were clearly enriched in the IS are actin cytoskeleton-related proteins, reflecting the dynamic reorganization of the cytoskeleton during BCR activation. This approach also identifies both known and novel players in BCR signaling, highlighting its potential for uncovering new regulatory mechanisms.

Network analysis reveals key protein networks such as cytoskeletal regulators, vesicle transport mediators, and the proteasome complex, all of which have established roles in B cell activation (Study III, Fig. 5B and C). Notably, we also detected some biological process, such as spliceosome and ribosomal complexes that have no previously characterized functions in antigen receptor signaling (Study III, Fig. 5B and C). These discoveries open new research directions for understanding BCR-driven immune responses.

5.3.2 Discussion

In this study, we developed and employed a robust method for isolating the immunological synapse from primary mouse B cells, allowing for an in-depth proteomic analysis. This approach enabled the identification of 1,039 high-confidence proteins, with a significant portion being unique to the IS, specifically those involved in BCR signaling. The differential enrichment analysis further confirmed the specificity of the method by revealing significant enrichment of known BCR activation proteins, such as Syk and Lyn in the IS samples.

Network analysis of differential expressed proteins identified key networks involved cytoskeletal regulation, vesicle transport, and proteasome activity, aligning with their known roles in B cell activation. These findings reinforce the importance of cytoskeletal remodeling and protein degradation in immune synapse formation and signal propagation, as well as suggests for new players involved in these activities in B cells.

Beyond previously established pathways, we also identified large protein networks, such as the spliceosome and ribosomal complexes, which have not been clearly linked to the IS. Prior studies have reported enrichment of these proteins at T cell receptor signaling sites (de Wet et al., 2011), suggesting they may play unrecognized roles in B cell activation. Their presence in IS samples raises intriguing questions about potential functions in localized mRNA processing, translation, or regulation of signaling dynamics such as BCR.

Overall, our findings in study III above expand the current understanding of BCR signaling by identifying both known and novel players in the IS. Future studies should explore the mechanistic contributions of newly identified pathways, particularly the role of RNA processing and translation in antigen receptor signaling. Understanding these interactions could provide deeper insights into immune regulation and potential therapeutic targets.

6 Conclusion

The humoral immune response is dependent on specific antibodies secreted by antigen stimulated B cells differentiated into plasma cells (Sahay et al., 2013). A comprehensive understanding of antigen-induced BCR signaling is crucial for our basic knowledge as well as harnessing the cell biological mechanisms behind these immune responses, for example, for improved vaccine development. Additionally, the cellular responses that follow BCR stimulation are complex and multifaceted, requiring coordination between various other biological processes and pathways.
To address this complexity, this thesis sought to provide new understanding into the proteins and molecular pathways involved in BCR signaling upon stimulation by both surface-bound and soluble antigens. We employed a combination of proximity labeling, cross-linking, and computational methods to study these interactions. Identification of new protein-protein interactions is crucial for understanding the molecular mechanisms underlying biological processes such as BCR signaling.
Over the past decades, the field of BCR signaling has made significant strides in understanding how antigen engagement by BCR on the plasma membrane translates to changes such as signal propagation, actin remodeling, cell polarization, activation of endocytosis processes, antigen processing, and gene expression. However, gaps remain in understanding the links between these processes and BCR antigen engagement. Achieving a holistic understanding of BCR signaling and the immune response requires understanding how the BCR signaling pathway communicates with other pathways. BCR activation triggers multiple protein-protein interactions and protein modifications such as phosphorylation. Understanding these interactions may hold the key to a deeper understanding of the connections between BCR signaling pathways and other biological processes and pathways.

This study demonstrates the utility of APEX2 as a proximity labeling tool for investigating BCR signaling, particularly when targeted to specific membrane domains like lipid rafts. The application of the Raft-APEX2

technique did not only identify novel proteins associated with BCR signaling but also provided a temporal view of their behavior during antigen stimulation. These insights contribute to a deeper comprehension of the molecular mechanisms governing BCR-mediated immune responses. Proximity labeling coupled with proteomics generated valuable information about proteins recruited or excluded from the vicinity of lipid raft domain in cells expressing Raft-APEX2 at 5-, 10-, and 15min time point of activation with soluble antigens. We obtained large-scale proteomes of lipid rafts, independent of BCR signaling, as well as BCR receptor proximity proteomes with high spatial and temporal resolution. We identified SUMOylation as a novel post-translational modification resposive to BCR signaling.

In the second study of this thesis, we developed AutoCoEv to predict potential interacting proteins from large protein datasets. AutoCoEv serves as a hit selection approach aimed at reducing the time and costs associated with validating potential false positives from the APEX2 strategy. It holds significant potential as a tool for selecting and refining candidate proteins in large-scale proteomic studies. Its ability to predict co-evolutionary relationships offers a valuable complement to experimental approaches, helping to filter out false positives and focus on functionally relevant interactions. Further testing and development of AutoCoEv, particularly in terms of sequence alignment optimization and computational efficiency, will enhance its utility and reliability, paving the way for deeper insights into complex biological systems.

In the third study of this thesis, we developed a technique to isolate and characterize the proteomic composition of the immunological synapse in primary mouse B cells, providing novel insights into the molecular mechanisms underlying BCR signaling. The findings from this study enhance our understanding of the IS's proteome and different cellular functions linked to the IS-mediated B cell activation.

Overall, the findings presented in this thesis contribute to a deeper understanding of the molecular mechanisms underlying BCR signaling and provide a foundation for future research in the field.

References

Abbas, A. K., Lichtman, A. H., & Pillai, S. (2014). *Cellular and Molecular Immunology E-Book: Cellular and Molecular Immunology E-Book*. Elsevier Health Sciences.

Almende, B. V., & Benoit, T. (2022). *visNetwork: Network Visualization using 'vis.js' Library* (Version 2.1.2) [Computer software]. https://cran.r-project.org/web/packages/visNetwork/index.htmlhmm!

AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics (Oxford, England)*, *35*(22), 4862–4865. https://doi.org/10.1093/bioinformatics/btz422

Aluvihare, V. R., Khamlichi, A. A., Williams, G. T., Adorini, L., & Neuberger, M. S. (1997). Acceleration of intracellular targeting of antigen by the B-cell antigen receptor: Importance depends on the nature of the antigen-antibody interaction. *EMBO Journal*, *16*(12), 3553–3562. https://doi.org/10.1093/emboj/16.12.3553

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., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Blighe, K., Rana, S., & Lewis, M. (2018). *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.* https://github.com/kevinblighe/EnhancedVolcano

Bosch, J. A., Chen, C.-L., & Perrimon, N. (2021). Proximity-dependent labeling methods for proteomic profiling in living cells: An update. *Wiley Interdisciplinary Reviews. Developmental Biology*, *10*(1), e392. https://doi.org/10.1002/wdev.392

Brack, C., Hirama, M., Lenhard-Schuller, R., & Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell*, *15*(1), 1–14. https://doi.org/10.1016/0092-8674(78)90078-8

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N., & Ting, A. Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology*, *36*(9), 880. https://doi.org/10.1038/nbt.4201

Busman-Sahay, K., Drake, L., Sitaram, A., Marks, M., & Drake, J. R. (2013). Cis and trans regulatory mechanisms control AP2-mediated B cell receptor endocytosis via select tyrosine-based motifs. *PloS One*, *8*(1), e54938. https://doi.org/10.1371/journal.pone.0054938

Butters, T. D., & Hughes, R. C. (1974). Solubilization and fractionation of glycoproteins and glycolipids of KB cell membranes. *The Biochemical Journal*, *140*(3), 469–478. https://doi.org/10.1042/bj1400469

Calderón-Celis, F., Encinar, J. R., & Sanz-Medel, A. (2018). Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrometry Reviews*, *37*(6), 715–737. https://doi.org/10.1002/mas.21542

Chakrabarti, S., & Panchenko, A. R. (2009). Coevolution in defining the functional specificity. *Proteins*, *75*(1), 231–240. https://doi.org/10.1002/prot.22239

Cheng, P. C., Dykstra, M. L., Mitchell, R. N., & Pierce, S. K. (1999). A role for lipid rafts in B cell antigen receptor signaling and antigen targeting. *Journal of Experimental Medicine*, *190*(11),1549–1560. https://doi.org/10.1084/jem.190.11.1549

Clark, M. R., Campbell, K. S., Kazlauskas, A., Johnson, S. A., Hertz, M., Potter, T. A., Pleiman, C., & Cambier, J. C. (1992). The B Cell Antigen Receptor Complex: Association of Ig-α and Ig-β with Distinct Cytoplasmic Effectors. *Science*, *258*(5079), 123–126. https://doi.org/10.1126/science.1439759

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics: MCP*, *13*(9), 2513–2526. https://doi.org/10.1074/mcp.M113.031591

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*, *10*(4), 1794–1805. https://doi.org/10.1021/pr101065j

de Wet, B., Zech, T., Salek, M., Acuto, O., & Harder, T. (2011). Proteomic characterization of plasma membrane-proximal T cell activation responses. *The Journal of Biological Chemistry*, *286*(6), 4072–4080. https://doi.org/10.1074/jbc.M110.165415

Domon, B., & Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. *Science*, *312*(5771), 212–217. https://doi.org/10.1126/science.1124619

Dustin, M. L. (2014). The immunological synapse. *Cancer Immunology Research*, *2*(11), 1023–1033. https://doi.org/10.1158/2326-6066.CIR-14-0161

Dustin, M. L. (2016). Immunological Synapses. In M. J. H. Ratcliffe (Ed.), *Encyclopedia of Immunobiology* (pp. 16–24). Academic Press. https://doi.org/10.1016/B978-0-12-374279-7.11005-7

Fares, M. A., & McNally, D. (2006a). CAPS: Coevolution analysis using protein sequences. *Bioinformatics*, *22*(22), 2821–2822. https://doi.org/10.1093/bioinformatics/btl493

Fares, M. A., & McNally, D. (2006b). CAPS: Coevolution analysis using protein sequences. *Bioinformatics*, *22*(22), Article 22. https://doi.org/10.1093/bioinformatics/btl493

Goh, W. W. B., Wang, W., & Wong, L. (2017). Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends in Biotechnology*, *35*(6), 498–507. https://doi.org/10.1016/j.tibtech.2017.02.012

Haag, A. M. (2016). Mass Analyzers and Mass Spectrometers. *Advances in Experimental Medicine and Biology*, *919*, 157–169. https://doi.org/10.1007/978-3-319-41448-5_7

Hämälistö, S., Batalla, F. D. V., Yuseff, M. I., & Mattila, P. K. (2024). Endolysosomal vesicles at the center of B cell activation. *The Journal of Cell Biology*, *223*(3), e202307047. https://doi.org/10.1083/jcb.202307047

Herzog, S., & Jumaa, H. (2012). Self-recognition and clonal selection: Autoreactivity drives the generation of B cells. *Current Opinion in Immunology*, *24*(2), 166–172. https://doi.org/10.1016/j.coi.2012.02.004

Honigmann, A., & Pralle, A. (2016). Compartmentalization of the Cell Membrane. *Journal of Molecular Biology*, *428*(24 Pt A), 4739–4748. https://doi.org/10.1016/j.jmb.2016.09.022

Hoogeboom, R., & Tolar, P. (2016). Molecular Mechanisms of B Cell Antigen Gathering and Endocytosis. *Current Topics in Microbiology and Immunology*, *393*, 45–63. https://doi.org/10.1007/82_2015_476

Hopf, T. A., Green, A. G., Schubert, B., Mersmann, S., Schärfe, C. P. I., Ingraham, J. B., Toth-Petroczy, A., Brock, K., Riesselman, A. J., Palmedo, P., Kang, C., Sheridan, R., Draizen, E. J., Dallago, C., Sander, C., & Marks, D. S. (2019). The

EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics*, *35*(9), 1582–1584. https://doi.org/10.1093/bioinformatics/bty862

Hou, P., Araujo, E., Zhao, T., Zhang, M., Massenburg, D., Veselits, M., Doyle, C., Dinner, A. R., & Clark, M. R. (2006). B Cell Antigen Receptor Signaling and Internalization Are Mutually Exclusive Events. *PLoS Biology*, *4*(7), e200. https://doi.org/10.1371/journal.pbio.0040200

Hubbard, S. R., & Till, J. H. (2000). Protein Tyrosine Kinase Structure and Function. *Annual Review of Biochemistry*, *69*(1), 373–398. https://doi.org/10.1146/annurev.biochem.69.1.373

Hung, V., Udeshi, N. D., Lam, S. S., Loh, K. H., Cox, K. J., Pedram, K., Carr, S. A., & Ting, A. Y. (2016). Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nature Protocols*, *11*(3), Article 3. https://doi.org/10.1038/nprot.2016.018

Jin, L., Bi, Y., Hu, C., Qu, J., Shen, S., Wang, X., & Tian, Y. (2021). A comparative study of evaluating missing value imputation methods in label-free proteomics. *Scientific Reports*, *11*, 1760. https://doi.org/10.1038/s41598-021-81279-4

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Karpievitch, Y. V., Dabney, A. R., & Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics*, *13 Suppl 16*, S5. https://doi.org/10.1186/1471-2105-13-S16-S5

Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., & Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell*, *27*(8), 1188–1196. https://doi.org/10.1091/mbc.E15-12-0844

Kuokkanen, E., Šuštar, V., & Mattila, P. K. (2015a). Molecular Control of B Cell Activation and Immunological Synapse Formation. *Traffic*, *16*(4), 311–326. https://doi.org/10.1111/tra.12257

Kuokkanen, E., Šuštar, V., & Mattila, P. K. (2015b). Molecular control of B cell activation and immunological synapse formation. *Traffic (Copenhagen, Denmark)*, *16*(4), 311–326. https://doi.org/10.1111/tra.12257

Kurosaki, T. (2011). Regulation of BCR signaling. *Molecular Immunology*, *48*(11), 1287–1291. https://doi.org/10.1016/j.molimm.2010.12.007

Kwak, K., Akkaya, M., & Pierce, S. K. (2019a). B cell signaling in context. *Nature Immunology*, *20*(8), 963–969. https://doi.org/10.1038/s41590-019-0427-9

Kwak, K., Akkaya, M., & Pierce, S. K. (2019b). B cell signaling in context. *Nature Immunology*, *20*(8), 963–969. https://doi.org/10.1038/s41590-019-0427-9

Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., & Ting, A. Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods*, *12*(1), Article 1. https://doi.org/10.1038/nmeth.3179

Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016a). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research*, *15*(4), 1116–1125. https://doi.org/10.1021/acs.jproteome.5b00981

Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016b). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research*, *15*(4), Article 4. https://doi.org/10.1021/acs.jproteome.5b00981

LeBien, T. W., & Tedder, T. F. (2008). B lymphocytes: How they develop and function. *Blood*, *112*(5), 1570–1580. https://doi.org/10.1182/blood-2008-02-078071

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, *11*(10), Article 10. https://doi.org/10.1038/nrg2825

Li, X.-W., Rees, J. S., Xue, P., Zhang, H., Hamaia, S. W., Sanderson, B., Funk, P. E., Farndale, R. W., Lilley, K. S., Perrett, S., & Jackson, A. P. (2014). New Insights into the DT40 B Cell Receptor Cluster Using a Proteomic Proximity Labeling Assay. *The Journal of Biological Chemistry*, *289*(21), 14434. https://doi.org/10.1074/jbc.M113.529578

Linder, M. E., & Deschenes, R. J. (2007). Palmitoylation: Policing protein stability and traffic. *Nature Reviews Molecular Cell Biology*, *8*(1), Article 1. https://doi.org/10.1038/nrm2084

Lu, S. M., & Fairn, G. D. (2018). Mesoscale organization of domains in the plasma membrane—Beyond the lipid raft. *Critical Reviews in Biochemistry and Molecular Biology*, *53*(2), 192–207. https://doi.org/10.1080/10409238.2018.1436515

Lutz, C., Ledermann, B., Kosco-Vilbois, M. H., Ochsenbein, A. F., Zinkernagel, R. M., Köhler, G., & Brombacher, F. (1998). IgD can largely substitute for loss of

IgM function in B cells. *Nature*, *393*(6687), 797–801. https://doi.org/10.1038/31716

Mair, A., Xu, S.-L., Branon, T. C., Ting, A. Y., & Bergmann, D. C. (2019). Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. *eLife*, *8*, e47864. https://doi.org/10.7554/eLife.47864

Mandelman, D., Schwarz, F. P., Li, H., & Poulos, T. L. (1998). The role of quaternary interactions on the stability and activity of ascorbate peroxidase. *Protein Science : A Publication of the Protein Society*, *7*(10), 2089–2098.

Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H., & Ting, A. Y. (2012). Engineered ascorbate peroxidase as a genetically-encoded reporter for electron microscopy. *Nature Biotechnology*, *30*(11), 1143–1148. https://doi.org/10.1038/nbt.2375

Mårtensson, I.-L., Almqvist, N., Grimsholm, O., & Bernardi, A. I. (2010). The pre-B cell receptor checkpoint. *FEBS Letters*, *584*(12), 2572–2579. https://doi.org/10.1016/j.febslet.2010.04.057

McLendon, R., Friedman, A., Bigner, D., Van Meir, E. G., Brat, D. J., M. Mastrogianakis, G., Olson, J. J., Mikkelsen, T., Lehman, N., Aldape, K., Alfred Yung, W. K., Bogler, O., VandenBerg, S., Berger, M., Prados, M., Muzny, D., Morgan, M., Scherer, S., Sabo, A., … National Human Genome Research Institute. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, *455*(7216), 1061–1068. https://doi.org/10.1038/nature07385

Möckl, L. (2020). The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Frontiers in Cell and Developmental Biology*, *8*, 253. https://doi.org/10.3389/fcell.2020.00253

Mohamed, A., Shah, A. D., Chen, D., & Hill, M. M. (2019). RaftProt V2: Understanding membrane microdomain function through lipid raft proteomes. *Nucleic Acids Research*, *47*(D1), D459–D463. https://doi.org/10.1093/nar/gky948

Moon, D. O. (2024). Advancing cancer therapy: The role of KIF20A as a target for inhibitor development and immunotherapy. *Cancers*, 16(17), 2958. https://doi.org/10.3390/cancers16172958

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., & Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, *108*(49), E1293–E1301. https://doi.org/10.1073/pnas.1111471108

Music, A., Tejeda-González, B., Cunha, D. M., Fischer von Mollard, G., Hernández-Pérez, S., & Mattila, P. K. (2022a). The SNARE protein Vti1b is recruited to the sites of BCR activation but is redundant for antigen internalisation, processing and presentation. *Frontiers in Cell and Developmental Biology*, *10*. https://www.frontiersin.org/articles/10.3389/fcell.2022.987148

Music, A., Tejeda-González, B., Cunha, D. M., Fischer von Mollard, G., Hernández-Pérez, S., & Mattila, P. K. (2022b). The SNARE protein Vti1b is recruited to the sites of BCR activation but is redundant for antigen internalisation, processing and presentation. *Frontiers in Cell and Developmental Biology*, *10*. https://www.frontiersin.org/articles/10.3389/fcell.2022.987148

Natkanski, E., Lee, W.-Y., Mistry, B., Casal, A., Molloy, J. E., & Tolar, P. (2013). B cells use mechanical energy to discriminate antigen affinities. *Science (New York, N.Y.)*, *340*(6140), 1587–1590. https://doi.org/10.1126/science.1237572

Oellerich, T., Bremes, V., Neumann, K., Bohnenberger, H., Dittmann, K., Hsiao, H.-H., Engelke, M., Schnyder, T., Batista, F. D., Urlaub, H., & Wienands, J. (2011). The B-cell antigen receptor signals through a preformed transducer module of SLP65 and CIN85. *The EMBO Journal*, 30(17), 3620–3634. https://doi.org/10.1038/emboj.2011.251

Oteri, F., Nadalin, F., Champeimont, R., & Carbone, A. (2017). BIS2Analyzer: A server for co-evolution analysis of conserved protein families. *Nucleic Acids Research*, *45*(Web Server issue), W307–W314. https://doi.org/10.1093/nar/gkx336

Padrón, A., Iwasaki, S., & Ingolia, N. T. (2019). Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. *Molecular Cell*, *75*(4), 875-887.e5. https://doi.org/10.1016/j.molcel.2019.07.030

Paek, J., Kalocsay, M., Staus, D. P., Wingler, L., Pascolutti, R., Paulo, J. A., Gygi, S. P., & Kruse, A. C. (2017). Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell*, *169*(2), 338-349.e11. https://doi.org/S0092-8674(17)30347-1 [pii]

Petrov, P. B., Awoniyi, L. O., Šuštar, V., Balci, M. Ö., & Mattila, P. K. (2022). AutoCoEv—A High-Throughput In Silico Pipeline for Predicting Inter-Protein Coevolution. *International Journal of Molecular Sciences*, *23*(6), Article 6. https://doi.org/10.3390/ijms23063351

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, *2*, 927312. https://doi.org/10.3389/fbinf.2022.927312

Qin, W., Cho, K. F., Cavanagh, P. E., & Ting, A. Y. (2021). Deciphering molecular interactions by proximity labeling. *Nature Methods*, *18*(2), Article 2. https://doi.org/10.1038/s41592-020-01010-5

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., Palmer, T. D., Carette, J. E., & Khavari, P. A. (2018). RNA–protein interaction detection in living cells. *Nature Methods*, *15*(3), 207. https://doi.org/10.1038/nmeth.4601

Resh, M. D. (1999). Fatty acylation of proteins: New insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica Et Biophysica Acta*, *1451*(1), 1–16. https://doi.org/10.1016/s0167-4889(99)00075-0

Reth, M. (2002). Hydrogen peroxide as second messenger in lymphocyte activation. *Nature Immunology*, *3*(12), 1129–1134. https://doi.org/10.1038/ni1202-1129

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science (New York, N.Y.)*, *339*(6125), 1328–1331. https://doi.org/10.1126/science.1230593

Rickert, R. C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nature Reviews. Immunology*, *13*(8), 578–591. https://doi.org/10.1038/nri3487

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Rozanova, S., Barkovits, K., Nikolov, M., Schmidt, C., Urlaub, H., & Marcus, K. (2021). Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods in Molecular Biology (Clifton, N.J.)*, *2228*, 85–116. https://doi.org/10.1007/978-1-0716-1024-4_8

Saeki, K., Miura, Y., Aki, D., Kurosaki, T., & Yoshimura, A. (2003). The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. *The EMBO Journal*, *22*(12), 3015–3026. https://doi.org/10.1093/emboj/cdg293

Sahay, B., Owen, J. L., Yang, T., Zadeh, M., Lightfoot, Y. L., Ge, J.-W., & Mohamadzadeh, M. (2013). Activation of B Cells by a Dendritic Cell-Targeted Oral Vaccine. *Current Pharmaceutical Biotechnology*, *14*(10), 867–877.

Satpathy, S., Wagner, S. A., Beli, P., Gupta, R., Kristiansen, T. A., Malinova, D., Francavilla, C., Tolar, P., Bishop, G. A., Hostager, B. S., & Choudhary, C. (2015). Systems-wide analysis of BCR signalosomes and downstream

phosphorylation and ubiquitylation. *Molecular Systems Biology*, 11, 808. https://doi.org/10.15252/msb.20145880

Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M., & Bateman, A. (2014). TreeFam v9: A new website, more species and orthology-on-the-fly. *Nucleic Acids Research*, *42*(D1), D922–D925. https://doi.org/10.1093/nar/gkt1055

Seda, V., & Mraz, M. (2015). B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *European Journal of Haematology*, *94*(3), 193–205. https://doi.org/10.1111/ejh.12427

Sedwick, C. E., & Altman, A. (2002). Ordered Just So: Lipid Rafts and Lymphocyte Function. *Science's STKE*, *2002*(122), re2–re2. https://doi.org/10.1126/stke.2002.122.re2

Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V., & Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, *1*(6), 2856–2860. https://doi.org/10.1038/nprot.2006.468

Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., & Marino Buslje, C. (2013). MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Research*, *41*(Web Server issue), W8-14. https://doi.org/10.1093/nar/gkt427

Simons, K., & Ikonen, E. (1997). Functional rafts in cell membranes. *Nature*, *387*(6633), Article 6633. https://doi.org/10.1038/42408

Singer, S. J., & Nicolson, G. L. (1972). The Fluid Mosaic Model of the Structure of Cell Membranes. *Science*, *175*(4023), 720–731. https://doi.org/10.1126/science.175.4023.720

Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., & Laub, M. T. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell*, 133(6), 1043–1054. https://doi.org/10.1016/j.cell.2008.04.040

Sohn, H. W., Tolar, P., Jin, T., & Pierce, S. K. (2006). Fluorescence resonance energy transfer in living cells reveals dynamic membrane changes in the initiation of B cell signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8143–8148. https://doi.org/10.1073/pnas.0509858103

Sohn, H. W., Tolar, P., & Pierce, S. K. (2008a). Membrane heterogeneities in the formation of B cell receptor–Lyn kinase microclusters and the immune synapse. *Journal of Cell Biology*, *182*(2), 367–379. https://doi.org/10.1083/jcb.200802007

Sohn, H. W., Tolar, P., & Pierce, S. K. (2008b). Membrane heterogeneities in the formation of B cell receptor-Lyn kinase microclusters and the immune synapse.

The Journal of Cell Biology, *182*(2), Article 2. https://doi.org/10.1083/jcb.200802007

Sonnino, S., & Prinetti, A. (2013). Membrane domains and the 'lipid raft' concept. *Current Medicinal Chemistry*, *20*(1), 4–21.

Stoddart, A., Jackson, A. P., & Brodsky, F. M. (2005). Plasticity of B Cell Receptor Internalization upon Conditional Depletion of Clathrin. *Molecular Biology of the Cell*, *16*(5), 2339–2348. https://doi.org/10.1091/mbc.E05-01-0025

Striebel, F., Imkamp, F., Özcelik, D., & Weber-Ban, E. (2014). Pupylation as a signal for proteasomal degradation in bacteria. *Biochimica Et Biophysica Acta*, *1843*(1), 103–113. https://doi.org/10.1016/j.bbamcr.2013.03.022

Šuštar, V., Vainio, M., & Mattila, P. K. (2018). Visualization and quantitative analysis of the actin cytoskeleton upon B cell activation. In *Methods in Molecular Biology* (Vol. 1707, pp. 243–257). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7474-0_18

Tanaka, S., & Baba, Y. (2020). B Cell Receptor Signaling. *Advances in Experimental Medicine and Biology*, *1254*, 23–36. https://doi.org/10.1007/978-981-15-3532-1_2

Tang, J., Wang, Y., Luo, Y., Fu, J., Zhang, Y., Li, Y., Xiao, Z., Lou, Y., Qiu, Y., & Zhu, F. (2020). Computational advances of tumor marker selection and sample classification in cancer proteomics. *Computational and Structural Biotechnology Journal*, *18*, 2012–2025. https://doi.org/10.1016/j.csbj.2020.07.009

Treanor, B. (2012). B-cell receptor: From resting state to activate. *Immunology*, *136*(1), 21–27. https://doi.org/10.1111/j.1365-2567.2012.03564.x

Trinkle-Mulcahy, L. (2019). Recent advances in proximity-based labeling methods for interactome mapping. *F1000Research*, *8*, F1000 Faculty Rev-135. https://doi.org/10.12688/f1000research.16903.1

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520

Urban, P. L. (2016). Quantitative mass spectrometry: An overview. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2079), 20150382. https://doi.org/10.1098/rsta.2015.0382

van Meer, G., Stelzer, E. H., Wijnaendts-van-Resandt, R. W., & Simons, K. (1987). Sorting of sphingolipids in epithelial (Madin-Darby canine kidney) cells. *The Journal of Cell Biology*, *105*(4), 1623–1635. https://doi.org/10.1083/jcb.105.4.1623

Vilen, B. J., Nakamura, T., & Cambier, J. C. (1999). Antigen-Stimulated Dissociation of BCR mIg from Ig-α/Ig-β: Implications for Receptor Desensitization. *Immunity*, *10*(2), 239–248.

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., & Hermjakob, H. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research*, *44*(D1), 447. https://doi.org/10.1093/nar/gkv1145

Wang, S., Sun, S., Li, Z., Zhang, R., & Xu, J. (2017). Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Computational Biology*, *13*(1), e1005324. https://doi.org/10.1371/journal.pcbi.1005324

Webb-Robertson, B.-J. M., Matzke, M. M., Jacobs, J. M., Pounds, J. G., & Waters, K. M. (2011). A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics*, *11*(24), 4736–4741. https://doi.org/10.1002/pmic.201100078

Wen, Y., Jing, Y., Yang, L., Kang, D., Jiang, P., Li, N., Cheng, J., Li, J., Li, X., Peng, Z., Sun, X., Miller, H., Sui, Z., Gong, Q., Ren, B., Yin, W., & Liu, C. (2019). The regulators of BCR signaling during B cell activation. *Blood Science*, *1*(2), 119–129. https://doi.org/10.1097/BS9.0000000000000026

Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Wickham, R., François, R., Henry, L., & Müller, K. (2020). *dplyr: A grammar of data manipulation (R package version 1.0.2)*. https://CRAN.R-project.org/package=dplyr

Wienands, J., Larbolette, O., & Reth, M. (1996). Evidence for a preformed transducer complex organized by the B cell antigen receptor. *Proceedings of the National Academy of Sciences*, *93*(15), 7865–7870. https://doi.org/10.1073/pnas.93.15.7865

Willforss, J., Chawade, A., & Levander, F. (2019). NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *Journal of Proteome Research*, *18*(2), 732–740. https://doi.org/10.1021/acs.jproteome.8b00523

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, *13*(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Winkler, T. H., & Mårtensson, I.-L. (2018). The Role of the Pre-B Cell Receptor in B Cell Development, Repertoire Selection, and Tolerance. *Frontiers in Immunology*, *9*, 2423. https://doi.org/10.3389/fimmu.2018.02423

Xu, Y., Harder, K. W., Huntington, N. D., Hibbs, M. L., & Tarlinton, D. M. (2005). Lyn tyrosine kinase: Accentuating the positive and the negative. *Immunity*, *22*(1), 9–18. https://doi.org/10.1016/j.immuni.2004.12.004

Yasuda, K., Kosugi, A., Hayashi, F., Saitoh, S., Nagafuku, M., Mori, Y., Ogata, M., & Hamaoka, T. (2000). Serine 6 of Lck Tyrosine Kinase: A Critical Site for Lck Myristoylation, Membrane Localization, and Function in T Lymphocytes. *The Journal of Immunology*, *165*(6), 3226–3231. https://doi.org/10.4049/jimmunol.165.6.3226

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zhang, Y., Parmigiani, G., & Johnson, W. E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, *2*(3), lqaa078. https://doi.org/10.1093/nargab/lqaa078