Large-Scale Kinetic Analyses of Protein-Protein Interactions: Advancing the Understanding of Post Translational Modifications in Biological Regulation

by

Brianne Ogata Petritis

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Joshua LaBaer, Chair
Douglas Lake
Shaopeng Wang

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Signal transduction networks comprising protein-protein interactions (PPIs) mediate homeostatic, diseased, and therapeutic cellular responses. Mapping these networks has primarily focused on identifying interactors, but less is known about the interaction affinity, rates of interaction or their regulation. To better understand the extent of the annotated human interactome, I first examined > 2500 protein interactions within the B cell receptor (BCR) signaling pathway using a current, cutting-edge bioluminescence-based platform called “NanoBRET” that is capable of analyzing transient and stable interactions in high throughput. Eighty-three percent (83%) of the detected interactions have not been previously reported, indicating that much of the BCR pathway is still unexplored. Unfortunately, NanoBRET, as with all other high throughput methods, cannot determine binding kinetics or affinities. To address this shortcoming, I developed a hybrid platform that characterizes > 400 PPIs quantitatively and simultaneously in < 1 hour by combining the high throughput and flexible nature of nucleic programmable protein arrays (NAPPA) with the quantitative abilities of surface plasmon resonance imaging (SPRi). NAPPA-SPRi was then used to study the kinetics and affinities of > 12,000 PPIs in the BCR signaling pathway, revealing unique kinetic mechanisms that are employed by proteins, phosphorylation and activation states to regulate PPIs. In one example, activation of the GTPase RAC1 with nonhydrolyzable GTP-γS minimally affected its binding affinities with phosphorylated proteins but increased, on average, its on- and off-rates by 4 orders of magnitude for one-third of its interactions. In contrast, this phenomenon occurred with virtually all unphosphorylated

proteins. The majority of the interactions (85%) were novel, sharing 40% of the same interactions as NanoBRET as well as detecting 55% more interactions than NanoBRET. In addition, I further validated four novel interactions identified by NAPPA-SPRi using SDS-PAGE migration and Western blot analyses. In one case, we have the first evidence of a direct enzyme-substrate interaction between two well-known proto-oncogenes that are abnormally regulated in > 30% of cancers, PI3K and MYC. Herein, PI3K is demonstrated to phosphorylate MYC at serine 62, a phosphosite that increases the stability of MYC. This study provides valuable insight into how PPIs, phosphorylation, and GTPase activation regulate the BCR signal transduction pathway. In addition, these methods could be applied toward understanding other signaling pathways, pathogen-host interactions, and the effect of protein mutations on protein interactions.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Joshua LaBaer for allowing me the opportunity to be trained as a graduate student in the Virginia G. Piper Center for Personalized Diagnostics. The collaborative environment of the laboratory, top-of-the-line equipment, and excellent colleagues helped ensure that my graduate studies were a fulfilling learning experience. Moreover, Dr. LaBaer's ability to lead by example, gain respect by scientists and non-scientists alike, be compassionate, juggle numerous projects and duties at work, have a strong and loving family, and still take time out of his busy schedule to be a constructive and supportive mentor – all while making it look so easy! – has inspired me, both professionally and personally. Although Dr. LaBaer has a cache of catch phrases, I feel compelled to record the most noteworthy here: "It could just be a shiny pebble" and "Is the juice worth the squeeze?"

I extend my heartfelt gratitude to my supervisory committee, Dr. Joshua LaBaer, Dr. Douglas Lake, and Dr. Shaopeng Wang, for their invaluable feedback and taking time out of their busy schedules to critically review this thesis. Their feedback only highlights the fact that science is not a one-man (or woman) pursuit!

Undoubtedly one of the best parts about science is the people. During my time in graduate school, I've had the privilege to know, work with, and learn from (in no particular order) Dr. Mitchell Magee, Justin Saul, Ian Shoemaker, Dr. Ji Qiu, Dr. Lusheng Song, Dr. Jin Park, Dr. Femina Rauf, Dr. Fernanda Festa, Dr. Xiaofang Bian, Haoyu Wang, Dr. Gokhan Demirkan, Dr. Preston Hunter, Dr. Garrick Wallstrom, Melissa Gutierrez, Seron Eaton, Jason Steel, Jennifer Van Duine, Lisa Miller, Dr. Jie

Wang, Dr. Catherine Seiler, Yan Wang, Dr. Andrea Throop, Katherine Stinchfield, Maria Hanlin, Dr. JoAnn Williams, Dr. Antonio Garcia, Laura Hawes, Karie Tepper, Vanessa Baack, Dr. Xiaobo Yu, and countless others. In addition, I admire the exceptional dedication and assistance of my undergraduate students, Benjamin Ober-Reynolds and David Haddad, both of whom were a pleasure to work with and know.

This project started as part of the Physical Sciences - Oncology Centers (PS-OC) program that was funded by the National Cancer Institute. Although the grant finished, we have continued to work toward the original vision of the project: to build – for the very first time – a dynamic, predictive model of a signal transduction network using large-scale, experimentally-produced quantitative data. This required us to develop new technology, methodology, software, and a model. It has truly been a multidisciplinary project! The contributions of the following people at Arizona State University are helping to make the PS-OC vision a reality: Ian Shoemaker, Dr. Femina Rauf, Justin Saul, Yan Wang, Dr. Jin Park, Dr. D. Mitchell Magee, Xiaobo Yu, Dr. Andrea Throop, Benjamin Ober-Reynolds, David Haddad, Marika Hopper, and Dr. Joshua LaBaer. My collaborators at Stanford University provided invaluable knowledge, manpower, and resources. I profusely thank them for their support in this endeavor; it was a pleasure to work with them: Dr. Parag Mallick, Ravali Adusumilli, Dr. Justin Carden, Dr. Wendy Fantl, Michelle Hori, and Dr. Mark Flory. I am also grateful for the help given by Dr. Calum MacAulay, who assisted in developing the first iteration of the in-house software "SPRite" to analyze the high throughput NAPPA-SPRi data as a visiting scientist from

Integrative Oncology (Canada). Throughout this thesis, I acknowledge the specific contributions of these individuals regarding this project.

Prior to graduate school, I had the fortune to work with many encouraging and inspiring mentors that set me on the path that led to this thesis, and it would be remiss of me if I didn't acknowledge the ones who were particularly influential. I am grateful that Dr. Krassen Dimitrov (Institute for Systems Biology; Seattle, WA) took a chance on an undergraduate student who hadn't performed real science outside of the classroom. His encouragement and wonderful leadership helped to break my pre-conceived stereotypes of scientists and made me embrace science as a career. Dr. Wei-Jun Qian (Pacific Northwest National Laboratory; Richland, WA) was an exceptional teacher and role model who continually challenged me to be a better scientist and ultimately motivated me to pursue my Ph.D. He also gave me a lot of advice (whether I asked for it or not) that I later learned to appreciate, including his recommendation to be a graduate student in Dr. LaBaer's laboratory.

This project would not have been possible without the funding from the National Cancer Institute through the PS-OC grant #1U54CA143907-01. I am also grateful to Arizona Science Foundation and National Science Foundation for their generous support through graduate research fellowships.

The importance of personalized and precision medicine – and the potential impact of this project – were underscored in a personal way for me when my father was diagnosed with inoperable stage IV stomach cancer in 2015. His overly optimistic attitude and tenacious struggle to survive despite the physical, emotional, and social side

effects of cancer, chemotherapy, and a poor prognosis defined what an incredible person he was. It is an honor to be his daughter.

Finally, I feel extremely blessed to have the friends and family that I do – you know who you are. Throughout it all, they have given me continual and unconditional love, support, and purpose.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

TABLES Page

LIST OF FIGURES

FIGURES Page

LIST OF ABBREVIATIONS

(in alphabetical order)

ACE: affinity capillary electrophoresis

AML: acute myeloid leukemia

AMP: adenosine monophosphate

AMR: angle of minimum reflectance

APC: antigen presenting cell

APS: SH2B adaptor protein 2

ATCC: American Type Culture Collection

ATP: adenosine triphosphate

BCR: B-cell receptor

BioGRID: Biological general repository for interaction datasets

BRET: bioluminescence resonance energy transfer

BS3: bis(sulfosuccinimidyl)suberate

CCD: charge-coupled device

CD: circular dichroism

CD: cluster of differentiation

cDNA: complementary deoxyribonucleic acid

CIP: calf intestinal alkaline phosphatase

CLL: chronic lymphocytic leukemia

CML: chronic myelogenous leukemia

CVID: common variable immune deficiency

DAG: diacylglycerol

DMSO: dimethylsulfoxide

DNA: deoxyribonucleic acid

DOE: design of experiment

dsRNA: double-stranded ribonucleic acid

EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride

EDTA: ethylenediaminetetraacetic acid

Fab: fragment antigen-binding

FDA: U.S. Food and Drug Administration

FPLC: fast purification liquid chromatography

FRET: fluorescence resonance energy transfer

GDI: guanosine nucleotide dissociation inhibitor

GEF: guanine nucleotide exchange factor

GO: Gene Ontology

GTP: guanosine triphosphate

GVHD: graft-versus-host disease

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HGNC: HUGO Gene Nomenclature Committee

HIV: human immunodeficiency virus

HPLC: high pressure liquid chromatography

HPRD: human protein reference database

HPV: human papillomavirus

HRP: horseradish peroxidase

HSA: human serum albumin

IDP: intrinsically disordered protein

Ig: immunoglobulin

ITAM: immunoreceptor tyrosine-based activation motif

ITC: isothermal calorimetry

ITD: internal tandem duplication

IVTT or IVT: *in vitro* transcription translation

k_a: association rate

k_d: dissociation rate

K_D: dissociation constant

KEGG: Kyoto encyclopedia of genes and genomes

LCK: LCK proto-oncogene, src family tyrosine kinase

LPP: lambda protein phosphatase

LT: Lysate-treated

MBP: maltose binding protein

mBU: milliBRET unit

MCL: mantle cell lymphoma

MHC: major histocompatibility complex

MS: multiple sclerosis

MST: microscale thermophoresis

MZL: marginal zone lymphoma

NAPPA: nucleic acid programmable protein array

NCI: National Cancer Institute

NHL: Non-Hodgkin lymphoma

NHS: N-hydroxysuccinimide

NMR: nuclear magnetic resonance

NP: Not phosphorylated

OFAAT: one factor at a time

PANTHER: Protein Analysis Through Evolutionary Relationships

PBS: phosphate buffered saline

PBST: PBS + 0.01% Tween-20

PCR: polymerase chain reaction

PDMS: polydimethylsiloxane

PEG: polyethylene glycol

PIP2: phosphatidylinositol (3,4)-bisphosphate

PIP3: phosphatidylinositol (3,4,5)-trisphosphate

PL: poly(L-lysine)

PMSF: phenylmethylsulfonyl fluoride

POI: protein-of-interest

PPI: protein-protein interaction

PS-OC: Physical Sciences Oncology Centers

PTM: post translational modification

RI: refractive index

RIU: refractive index units

RNA: ribonucleic acid

ROCK: rho associated coiled-coil containing protein kinase 1

ROI: region of interest

RP: research project

RU: response unit

SDS: sodium dodecyl sulfate

SDS-PAGE: SDS polyacrylamide gel electrophoresis

SH2: Src homology 2

SPR: surface plasmon resonance

SPRi: surface plasmon resonance imaging

STAT5A: signal transducer and activator of transcription 5A

TCEP: tris(2-carboxyethyl)phosphine

TEV: tobacco etch virus

TRTK: transmembrane receptor protein tyrosine kinase

VBA: Visual Basic Applications

VGP CPD: Virginia G. Piper Center for Personalized Diagnostics

X-SCID: X-linked severe combined immune deficiency

XLA: X-linked agammaglobulinemia

LIST OF GENE AND PROTEIN ABBREVIATIONS

(in alphabetical order)

See also Appendix A

AKT1: AKT serine/threonine kinase 1

AKT2: AKT serine/threonine kinase 2

AKT3: AKT serine/threonine kinase 3

BAD: BCL2 associated agonist of cell death

BAFF: B cell activating factor

BCL10: B-cell CLL/lymphoma 10

BCL2: BCL2, apoptosis regulator

BCL2A1: BCL2-related protein A1

BCL2L1: BCL2-like 1

BCR-ABL (fusion gene): B cell receptor – Abelson tyrosine protein kinase 1

BLK: B lymphoid tyrosine kinase

BSA: bovine serum albumin

BTK: Bruton agammaglobulinemia tyrosine kinase

CARD11: caspase recruitment domain family, member 11

CD-SIGN: CD209 molecule (CD209)

CD19: CD19 molecule

CD21: Complement C3d receptor 2 (CR2)

CD22: CD22 molecule

CD72: CD72 molecule

CD79A: CD79a molecule, immunoglobulin-associated alpha (IgA)

CD79B: CD79b molecule, immunoglobulin-associated beta (IgB)

CD81: CD81 molecule

CDC42: cell division cycle 42

CDKN2A: cyclin-dependent kinase inhibitor 2A

CFTR: cystic fibrosis transmembrane conductance regulator

CHP-1: calcineurin life EF-hand protein 1

DAPP1: dual adaptor of phosphotyrosine and 3-phosphoinositides (Bam32)

EGFR: epidermal growth factor response 1

EGR1: early growth response 1

ER: estrogen receptor

ETS1: ETS proto-oncogene 1, transcription factor

EZR: ezrin (VIL2)

FCGR2B: Fc fragment of IgG, receptor IIb (CD32)

FGF: fibroblast growth factor

FLT3: Fms related tyrosine kinase 3

FOS: fos proto-oncogene, AP-1 transcription factor subunit

FOXO3A: forkhead box O3

GM-CSF: granulocyte macrophage-colony stimulation factor

GRAP2: GRB2-related adaptor protein 2

GRB2: growth factor receptor-bound protein 2

GSK3B: glycogen synthase kinase 3 beta

GST: gluthione S-transferase

HPV2: human papillomavirus 2

HPV4: human papillomavirus 4

HRAS: HRas proto-oncogene, GTPase

IFITM1: interferon induced transmembrane protein 1

IKBKA: conserved helix-loop-helix ubiquitous kinase (CHUK)

IKBKB: inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta

IKBKG: inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma

INPP5D: inositol polyphosphate-5-phosphatase (SHIP1)

INPPL1: inositol polyphosphate phosphatase-like 1

JUN: Jun proto-oncogene, AP-1 transcription factor subunit

KRAS: KRas proto-oncogene, GTPase

LAT2: linker for activation of T cells family, member 2

LILRB3: leukocyte immunoglobulin-like receptor B3 (PIRB)

LIME1: Lck interacting transmembrane adaptor 1

LYN: LYN proto-oncogene, Src family tyrosine kinase

MALT1: MALT1 paracaspase

MAP2K1: mitogen-activated protein kinase kinase 1

MAP2K2: mitogen-activated protein kinase kinase 2

MAP2K3: mitogen-activated protein kinase kinase 3

MAPK1: mitogen-activated protein kinase 1 (ERK2)

MAPK3: mitogen-activated protein kinase 3 (ERK1)

MAPK8: mitogen-activated protein kinase 8 (JNK1)

MAPK9: mitogen-activated protein kinase 9 (JNK2)

MAPK12: mitogen-activated protein kinase 12 (p38 gamma)

MAPK13: mitogen-activated protein kinase 13 (p38 delta)

MAPK14: mitogen-activated protein kinase 14 (p38 alpha)

MDM2: MDM2 proto-oncogene

MYC: v-myc myelocytomatosis viral oncogene homolog

NCK1: NCK adaptor protein 1

NCKAP1L: NCK-associated protein 1-like

NFAT5: nuclear factor of activated T cells 5

NFATC1: nuclear factor of activated T cells 1

NFATC3: nuclear factor of activated T cells 3

NFATC4: nuclear factor of activated T cells 4

NFKB1: nuclear factor of kappa B subunit 1 (p105)

NFKBIA: NFKB inhibitor alpha

NFKBIB: NFKB inhibitor beta

NFKBIE: NFKB inhibitor epsilon

NOD2: nucleotide binding oligomerization domain containing 2

NRAS: neuroblastoma RAS viral oncogene homolog

PIK3AP1: phosphoinositide-3-kinase adaptor protein 1 (BCAP)

PIK3CA: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha

PIK3CB: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta

PIK3CD: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit delta

PIK3CG: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma

PIK3R1: phosphoinositide-3-kinase, regulatory subunit 1

PIK3R2: phosphoinositide-3-kinase, regulatory subunit 2

PIK3R3: phosphoinositide-3-kinase, regulatory subunit 3

PIK3R5: phosphoinositide-3-kinase, regulatory subunit 5

PLCG2: phospholipase C, gamma 2

PPP3CA: protein phosphatase 3, catalytic subunit, alpha

PPP3CB: protein phosphatase 3, catalytic subunit, beta

PPP3CC: protein phosphatase 3, catalytic subunit, gamma

PPP3R1: protein phosphatase 3, regulatory subunit B, alpha

PPP3R2: protein phosphatase 3, regulatory subunit B, beta

PRKCA: protein kinase C, alpha

PRKCB: protein kinase C, beta

PTEN: phosphatase and tensin homolog

PTPN6: protein tyrosine phosphatase, non-receptor type 6 (SHP1)

RAC1: ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)

RAC2: ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)

RAC3: ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)

RAF1: raf-1 proto-oncogene, serine/threonine kinase

RANKL: tumor necrosis factor superfamily member 11 (TNFSF11)

RAP1A: RAP1A, member of RAS oncogene family

RAP1B: RAP1B, member of RAS oncogene family

RAP2A: RAP2A, member of RAS oncogene family

RAP2C: RAP2C, member of RAS oncogene family

RasGRP3: RAS guanyl releasing protein 3

RASSF5: Ras association domain family member 5

RELA: RELA proto-oncogene, NF-KB subunit

RHOA: ras homolog family member A (RHOA)

SOS1: SOS Ras/Rac guanine nucleotide exchange factor 1

SOS2: SOS Ras/Rac guanine nucleotide exchange factor 2

SYK: spleen associated tyrosine kinase

TEC: tec protein tyrosine kinase

TP53: tumor protein TP53

VAV1: vav guanine nucleotide exchange factor 1

VAV2: vav guanine nucleotide exchange factor 2

VAV3: vav guanine nucleotide exchange factor 3

VEGF: vascular endothelial growth factor

VH3: immunoglobulin heavy constant mu (IGHM)

PREFACE

A comprehensive understanding of the dynamic and complex signaling networks within cells remains one of the grand challenges in the pursuit for precision medicine. In regards to cancer, therapy resistance and disease recurrence largely occur through multiple, yet interconnected pathways that help the diseased cell(s) to evade treatment, the immune response, and normal physiological cell-death signals. The identification of the key proteins involved in pathway crosstalk or driving disease progression and therapy resistance will no doubt aid in creating targeted and combinatorial therapy approaches that will be more effective than current treatments. For sure, the successful story of imatinib mesylate, sold under the trade name of "Gleevec," in treating chronic myelogenous leukemia (CML) by specifically targeting the BCR-ABL protein underscores the potential impact of understanding diseases at the molecular level. Combinatorial therapy has been repeatedly proven to be more effective than single-drug cancer treatment over the last five decades. For instance, the FLT3 tyrosine kinase with an internal tandem duplication mutation (FLT3-ITD) results in constitutive activation of the kinase and, subsequently, acute myeloid leukemia (AML). The small-molecule drug sorafenib in combination with chemotherapy was recently shown in a phase II clinical trial to increase the 1-year survival rate in older AML-FLT3-ITD patients than standard chemotherapy alone (Uy et al., 2015).

Given the importance of protein-protein interactions (PPIs), it is surprising that their binding kinetics and affinities have been studied only minimally. Numerous techniques, which are discussed in more detail in Chapter 1, have been developed to

study PPIs. However, many of them rely on stable interactions even though protein interactions are known to occur over a wide range of affinities (i.e., strengths) and kinetics (i.e., rates). Moreover, none of the available approaches can assess unique binding events quantitatively in a high throughput manner, thus resulting in a paucity of affinity and kinetic information. The B cell receptor (BCR) signaling pathway, for example, is considered to be one of the most well understood pathways, involving > 100 proteins and potentially > 2^{100} interactions, yet most of its interactions have been studied using classic equilibrium-based assays and only 12 protein interactions have been characterized quantitatively.

Scientists and mathematicians have proposed that models, built from large-scale binding affinity information and protein abundance data, could improve our understanding of signaling pathways and allow prediction of cellular outcomes. Such models would rely on sufficient data about the participants in the pathways, their abundance, and their interaction characteristics to be accurate. I chose to study the BCR pathway because it was already well studied, still had room for additional discovery and because my collaborators were studying other aspects of the pathway that eventually would strengthen our model. I first studied the BCR signaling pathway using an equilibrium assay, albeit a modern one with the potential to detect some transient interactions. I then developed an entirely new methodology that could detect even more interactions (including weak ones) and which would provide kinetic data on interaction rates.

In this thesis, > 2500 protein interactions of the BCR signaling pathway were first examined using a current, cutting edge technology in which transient and stable interactions can be detected in high throughput. Seventy-two known and 401 previously unreported protein interactions were identified, highlighting the fact that the BCR pathway – and the human interactome – remain largely unexplored. Just like other high throughput protein interaction methods, however, the binding rates and affinities of these interactions could not be characterized. To address this need, this thesis describes how a high throughput protein microarray platform was combined with a traditionally low throughput technique capable of studying binding events in real-time to analyze > 400 protein interactions in less than an hour. The hybrid "NAPPA-SPRi" technology then studied > 12,000 PPIs within the BCR signaling pathway under different protein activation and phosphorylation states. An initial steady state model of the B cell is currently being built from kinetic and protein abundance data obtained from NAPPA-SPRi and mass spectrometry, respectively. This project represents the *first high throughput, quantitative analyses of protein-protein interactions for any signaling pathway*.

In Chapter 1, the history of how protein-protein interactions were conceptualized is examined. Chapter 2 focuses on the current techniques to study PPIs and reviews what is known about the BCR signaling pathway. Chapter 3 contains the qualitative analyses of > 2500 protein interactions in the BCR signaling pathway as determined by a high throughput bioluminescence-based approach. Chapters 4 – 7 cover the development of the technology, methodology, and software, respectively, regarding NAPPA-SPRi

applications and analyses. Chapter 8 contains the quantitative analyses of the PPIs in the BCR signaling pathway by NAPPA-SPRi. The data are also discussed in the context of biology – for example, what is the biological purpose of increasing the on-rate for a particular PPI? Finally, in Chapter 9, a description of how the NAPPA-SPRi data can be incorporated into a steady state model of B cell response and a perspective on the potential uses and impact of NAPPA-SPRi are given

CHAPTER 1

1 PROTEIN-PROTEIN INTERACTIONS

1.1 History of protein-protein interactions: from solitary molecules to protein networks

External stimuli are transmitted through the membrane by cell surface receptors, and then propagated through the cell via protein-protein interactions to elicit specific cellular responses. Disruptions to normal signaling from aberrant proteins (e.g., from mutation or altered expression) or chemicals can therefore initiate disease (Gonzalez & Kann, 2012). However, protein interactions in homeostasis and disease were not appreciated until the mid- to late-20th century.

Prior to the 1940s, proteins were largely considered to be solitary molecules without much function. Then, following the discovery that myosin B, a protein that had been studied for nearly a century, was actually a complex of myosin and actin, physical associations between proteins were observed with increasing frequency (Braun & Gingras, 2012). Proteins' three-dimensional structures and their effect on interactions also became of interest. The first signal transduction pathway, which happened to also be a kinase cascade, was identified in 1968 during a time when phosphoproteins were believed to be "biologically inert and (…) uninteresting." The Krebs laboratory showed that protein kinase A activated phosphorylase kinase in response to increases in cyclic AMP. The activation of phosphorylase kinase via phosphorylation was proven a year later. In the 1970s and 1980s, protein interactions became widely recognized as essential for most cellular responses following studies that showed their roles in homeostasis and disease, like the cell cycle and cancer. In 1990, the src homology domain (SH2)

preferentically interacted with phosphorylated proteins, providing proof that specific domains mediate interactions with post translational modifications. It also suggested that dynamic protein interactions may occur more often than originally believed. Within a decade, low throughput technologies became commercially available that could characterize the interaction strengths and the rates at which the proteins bound and unbound, like surface plasmon resonance and isothermal titration calorimetry (see Chapters 1.3 and 4.1.2). High throughput studies using yeast-2-hybrid and affinity purification mass spectrometry constructed the first large-scale maps of the interactome in the early 2000s. These data provided insights into the structural organization of protein networks as well as assigning biological function(s) to unknown proteins unveiled by the Human Genome Project.

The importance of understanding signaling pathways was underscored in 2001 with the first U.S. Food and Drug Administration (FDA)-approved small molecule "targeted therapy" kinase inhibitor, "Gleevec," to treat chronic myelogenous leukemia (CML) (Kurzrock & Markman, 2008). Gleevec, also known as imatinib mesylate, was specifically designed to bind and block the activity of a fusion kinase, BCR-ABL, since the active form results in unchecked cell proliferation in CML patients (Figure 1).

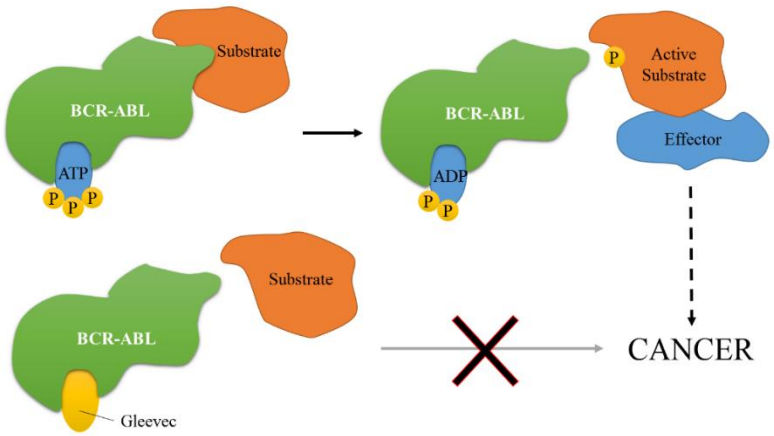


Figure 1. Mechanism of action of Gleevec (imatinib mesylate)

Gleevec increased the percentage of complete cytogenetic responses in chronic phase CML patients from 5% - 25% to 50 – 60%. Further understanding of CML's initial or eventual tumor resistance to Gleevec has led to the development of more powerful kinase inhibitors, including dasatinib, nilotinib, and bosutinib. Within 10 years of Gleevec's release, drugs targeting the epidermal growth factor receptor (EGFR), vascular endothelial growth factor (VEGF), and the proteasome became available to treat advanced non-small cell lung cancer, kidney and some gastrointestinal stromal tumors, and multiple myeloma, respectively.

1.2 Current understanding of protein-protein interactions

The human genome contains ~ 20,000 – 25,000 protein-coding genes that result in > 2 million protein species due to post-transcriptional alterations, mutations, and post translational modifications (PTMs) (Ponomarenko et al., 2016). Histone H4 alone, with combinations of its twenty PTMs, could represent > 3 million protein species with different protein interactions and functions (Phanstiel et al., 2008). The physical interactions between the proteins are specific and dynamic, resulting in signaling pathways that are often represented as independent, linear chain of events extending from the cell membrane to the

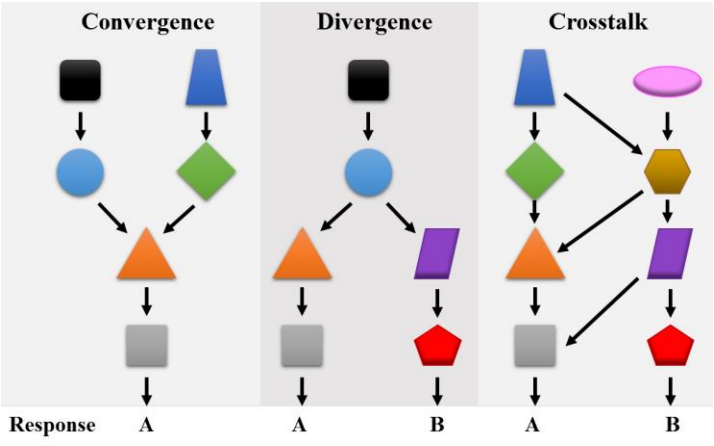


Figure 2. Signaling pathways can converge, diverge, and crosstalk with each other.

nucleus and eliciting a specific cellular phenotype(s) (e.g., proliferation, apoptosis). However, signal transduction via these protein-protein interactions (PPIs) is much more complex and interconnected where the pathways converge, diverge, and crosstalk with each other (Figure 2) (Karp & Patton, 2013). Convergence is when two or more different pathways result in the same molecular or cellular response. Divergence is when multiple signaling pathways are activated from a single stimulus. Crosstalk occurs when proteins are involved in more than one signaling pathway. Signaling cascades can be driven by a small number of proteins called "driver nodes." Thus, charting PPIs in homeostasis and disease would have a significant impact on medicine by identifying potential pharmacological targets. One of these targets, for example, could be a driver node that enables "disease crosstalk." Moreover, an in-depth understanding of how signaling pathways crosstalk with each other – and ultimately cause treatment resistance – will be fundamental in designing more effective combinational therapies based on the unique profile of the disease or patient.

Despite its importance, a complete PPI map of any species has proven to be a daunting task. The human proteome network, for instance, contains an estimated ~ 650,000 PPIs (Stumpf et al., 2008), with 49% of these estimated PPIs annotated in the Biological General Repository for Interaction Datasets (BioGRID) interaction database (Stark et al., 2006)(319,419 unique physical and genetic human PPIs; BioGRID database statistics 2018). It is probable that the coverage is much less since the PPI estimate does not take into account multiple splice variants and is based on experiments that are uninformative and inherently biased toward stable PPIs and scientific interests. For

example, equilibrium-based assays, which are commonly employed to study protein interactions, generally detect interactions with high binding affinities. Many experiments focus on studying arbitrary subsets of proteins that are known to be involved in disease, in a process that can be tautological. Moreover, very few studies fully annotate the PTMs of the proteins involved and detection methods are biased toward detecting highly abundant proteins. These reasons likely contribute to the notoriously small overlap across interaction datasets. A comparison of high throughput yeast interactions, for instance, revealed that only 14% of the detected PPIs were identified across different studies and methods (Reguly et al., 2006). Two independent large-scale yeast-2-hybrid screens using the same method had < 30% overlap (Ito et al., 2001).

Whether and how proteins physically interact with each other can be affected by numerous factors; for example, amino acid mutations, truncations, PTMs, co-regulators, intracellular location, viscosity, protein abundance, available domains, viruses, and conformation. Given the numerous aspects that can affect PPIs, it should come as no surprise that alterations to these finely-tuned protein signaling networks can lead to disease, including Huntington's disease, Von Hippel-Lindau syndrome, cystic fibrosis, Alzheimer's disease, and cervical cancer (Gonzalez & Kann, 2012). In autosomal recessive Cystic fibrosis, a mutant CFTR gene results in an amino acid deletion in the middle of the translated protein that renders it non-functional, resulting in the inability of sodium and chloride molecules to be transported across membranes (Mall & Galietta, 2015). Amino acid substitutions in the NOD2 protein result in a 2- or 40-fold risk increase for acquiring Crohn's disease depending on whether the person is heterozygous

or homozygous for the mutations, respectively (King et al., 2006). The mutations occur in regions responsible for detecting bacteria and NF-κB signaling. Estrogen (ER)-positive breast cancers are treated with a small molecule drug Tamoxifen. However, not all ER-positive breast cancer patients respond similarly to the drug. For example, the phosphorylation of ERα at serine residue 118 results in a better clinical outcome when using Tamoxifen than the unphosphorylated form (de Leeuw, Neefjes, & Michalides, 2011). On the other hand, breast cancer patients with serine phosphorylation at amino acid residue 305 of ERα do not respond to Tamoxifen. Finally, viral proteins can interact with endogenous proteins and alter homeostatic signaling networks. Human papillomavirus (HPV) increases the risk for developing cervical cancer because it encodes for E6 and E7 proteins, which bind to and inactivate the tumor suppressor protein TP53 (Yim & Park, 2005).

Bi-molecular interactions are described using the equation $A + B \Leftrightarrow AB$ where A and B represent two different proteins while AB represents the resulting complex (Goodrich & Kugel, 2007). The binding affinity, also known as the dissociation constant K_D, is generally described as the fraction of unbound proteins to bound proteins (i.e., [A][B]/[AB]) at equilibrium. Lower dissociation constants refer to protein interactions that strongly favor binding, resulting in most of the A and B proteins in the bound state at equilibrium (i.e., high binding affinity).

Interactions are also described in terms of the rates that A and B bind together (i.e., on-rate or k_a) and the rate at which AB dissociates into the two free molecules (i.e., off-rate or k_d). The association and dissociation rates at which AB forms and separates into its individual components, respectively, can be further used to determine the strength of the interaction (i.e., affinity). Thus, K_D can also be represented by the dissociation rate, k_d, divided by the association rate, k_a. Therefore, quantitative analyses of protein interactions can reveal whether the interaction is stable or transient, which has dissociations constants in the pM-nM and µM range, respectively. It is possible that protein interactions may have the same binding affinity but have different on- and off-rates as long as the changes to the binding rates are proportional (Figure 3). This raises the question about whether protein interactions are regulated at the level of binding strength (proportion of molecules bound at equilibrium; K_D), binding rates or both. To date, many more binding affinities have been collected than the on- and off-rates, which is largely due to two reasons. First, there are more available methods that can measure or estimate binding affinities than methods that can determine binding kinetics (see next section, Chapter 1.3). Second, it has been assumed that reasonable estimates of off-rates can be determined from the binding affinities because the on-rates for most proteins are believed to occur within a narrow range of 10^6

K_D (M)	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	k_a	k_d
1.15E-07	4.05E+08	4.66E+01	fast	fast
1.16E-07	1.50E+04	1.74E-03	moderate	moderate
1.15E-07	6.48E+03	7.43E-04	slow	slow

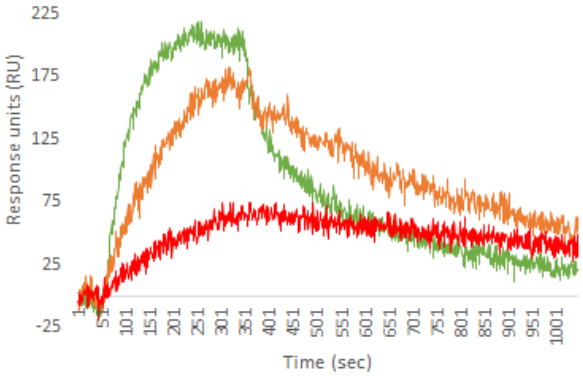


Figure 3. Same K_D, different on- and off-rates

to 10^7 $M^{-1}s^{-1}$ due to diffusion and protein size (Pollard, 2010). As such, the dissociation rate is often regarded as the main factor that determines binding affinity.

1.3 Current methods for studying protein-protein interactions

Protein-protein interactions can be predicted and studied using in silico approaches (Gonzalez & Kann, 2012; Wetie et al., 2014). The first step toward computer modeling of PPIs is to obtain structural information about the proteins-of-interest (POIs) since protein conformation can drastically affect whether and how a protein can interact with another protein (or other molecule). Protein structure can be obtained from databases using experimentally-produced data like Protein Data Bank or simulated using homology modeling (e.g., SwissModel, M4T, Modeller), de novo modeling (e.g., I-TASSER, Phre2), or threading (e.g., NovaFold, I-TASSER). The second step relies on virtual analyses of the structures and, sometimes, how the proteins behave over time and/or in various environments. The third and final step of in silico analyses is the calculation of the thermodynamics of protein complexes based on protein orientation and binding epitope(s). This information can help predict the likelihood that two proteins will interact with each other and, if so, where and when they are most likely to bind.

Numerous "wet lab" methods have been developed to study PPIs, yet very few provide quantitative (i.e., kinetic, affinity) information in a high throughput way (Gonzalez & Kann, 2012; Meyerkord & Fu, 2015; Wetie et al., 2014; Zhou, Li, & Wang, 2016). For example, commonly employed low throughput and qualitative methods include fluorescence gel retardation assay, far-Western blot, X-ray crystallography, and

Table 1. Capabilities of the most common methods for analyzing protein interactions

Method	High throughput	Binding kinetics and affinities	Transient PPIs	Stable PPIs	Concentration-independent*
2-hybrid	X				
Affinity capillary electrophoresis		X	X	X	X
Circular dichroism		X	X	X	X
Far-Western blot				X	X
Fluorescence gel retardation assay				X	X
Fluorescence polarization assay		X	X	X	X
FRET/BRET	Possible		X	X	Possible
Isothermal titration calorimetry		X	X	X	X
Mass spectrometry	X	Possible	Possible	X	
NAPPA-SPRi	X	X	X	X	X
NMR		X	X	X	X
Phage display	X			X	
Protein microarrays	X		Possible	X	X
Pull-down assays	X			X	
SPR		X	X	X	X
SPRi	Possible	X	X	X	X
X-ray crystallography				X	X

Note: This summary does not consider the use of cross-linking.

** Detection methods in in vivo systems often depends on protein concentration*

fluorescence resonance energy transfer (FRET)(Table 1). A fluorescence gel retardation assay applies potential complexes-of-interest to an SDS-PAGE gel; any resulting PPIs are identified by a shift in their gel migration when compared to the shift of single proteins. In far-Western blots, proteins separated via gel and transferred onto a membrane are probed with a labeled query protein, which is then used to determine its interaction partners. X-ray crystallography studies protein-protein structures by measuring the X-ray diffraction pattern of the crystallized complex. FRET employs chromophore-attached POIs that fluoresce at a different wavelength when they are in very close proximity to each other; FRET measures this change in fluorescence to detect interactions.

Bioluminescence resonance energy transfer (BRET) measures the change in emitted fluorescence where the resonance energy donor species is a luciferase.

High throughput qualitative methods for characterizing PPIs include phage display, mass spectrometry, 2-hybrid, protein microarrays, pull-down assays (e.g., co-immunoprecitation), and FRET\BRET. In phage display, a POI is displayed on a bacteriophage and then screened against other proteins. Mass spectrometry can identify PPIs in a couple of ways: it can ascertain which proteins are in a pull-down assay and analyze the contact areas of interacting proteins that are protected from proteolytic cleavage. 2-hybrid approaches identify PPIs by fusing one part of a transcription factor to a bait protein and the other part to a prey protein. If the proteins interact with each other, the transcription factor can bind to and initiate the transcription of a reporter gene. The 2-hybrid approach, however, has high false positive and negative rates that are estimated to be 50 – 70% and 43 – 90%, respectively (Deane, Salwinski, Xenarios, & Eisenberg, 2002; Huang & Bader, 2009). Protein interactions can be studied using protein microarrays by incubating the array with a known POI fused to a detectable tag. After washing off non-bound proteins, the tag location is determined and, since the address of each arrayed protein is known, the corresponding protein partner on the array can be identified. Pull-down assays extract a bait protein-of-interest (primarily via an antibody) along with its interacting proteins from solution. Since they isolate protein interaction complexes, the identified proteins may not interact directly but can bind through one or more bridging proteins. Without the use of chemical cross-linking (explained in more detail on page 12), pull-down assays also require the protein interactions to be stable

enough to withstand the washes prior to sample elution. In FRET and BRET, the POIs have fluorophore or luciferase tags, respectively, that will emit at a specific wavelength when a PPI occurs. More specific detail about FRET and BRET is on page 37.

Protein microarrays detect PPIs in a concentration-independent manner unlike *in vivo* methods that are biased toward detecting highly abundant proteins. Detecting interactions in a concentration-independent manner is advantageous because protein microarrays can 1) identify interactions between low abundance proteins that may significantly affect cellular responses, and 2) offer an unbiased detection of PPIs regardless of cellular state since the amount of proteins can be drastically different across conditions. While protein microarrays are primarily utilized for detecting strong PPIs (i.e., antigen-antibody interactions), transient interactions can be detected with protein microarrays in an indirect fashion. For example, protein targets of AMPylators can be determined by incubating an array with an AMPylator and N^6pATP, and then identifying the location of the N^6pATP (X. B. Yu & LaBaer, 2015). Reviews of the different types of protein microarrays in which I am first co-author include "Advancing translational research with next-generation protein microarrays" and "Advances in cell-free protein array methods" (X. Yu, Petritis, Duan, Xu, & LaBaer, 2018; X. B. Yu, Petritis, & LaBaer, 2016). I also co-authored a manuscript entitled "Multiplexed Nucleic Acid Programmable Protein Arrays" (X. B. Yu et al., 2017). This article describes a modification to Nucleic Acid Programmable Protein Arrays, or NAPPA (see also Chapter 4.1.1), where as many as five different proteins are displayed in one feature for high throughput, cost-effective biomarker screening and discovery.

Chemical cross-linking essentially "freezes" a PPI, thus allowing the detection of both stable and transient interactions by SDS-PAGE, in-gel digestion, and shotgun liquid chromatography mass spectrometry (Tang & Bruce, 2009). The reactive groups on the cross-linker act as a covalent bridge between interacting proteins. Numerous cross-linkers are available, although all of them have two or more reactive groups separated by a spacer, which may or may not be cleavable. Unfortunately, this approach has its disadvantages (Bruce, 2012). Most cross-linkers are lysine-reactive, which can be problematic for mass spectrometry analyses where the proteins are usually digested by the protease, trypsin, that also cuts at lysine residues (Holding, 2015). The identification of the protein partners during mass spectrometry analyses is challenging due to the additional mass of the cross-linker and a fragmentation spectrum that contains product ions from both peptides. Cross-linkers will covalently bind to anything within their reach, which means that proteins that are in close proximity, but not necessarily in contact with each other, will be crosslinked to each other. Finally, the binding rates or affinities *cannot* be determined with cross-linked protein complexes. The false detection rate for chemical cross-linking is unknown, but nonspecific binding of proteins to the crosslinked complexes during sample processing or to the stationary phase used for protein purification have been documented. Stringent washing during the enrichment procedure and utilization of short cross-linkers is assumed to decrease the number of non-crosslinked and nonspecific interactions that are identified, respectively.

The most common low throughput and quantitative (i.e., affinity, kinetics) methods for PPI analyses include circular dichroism (CD), surface plasmon resonance

(SPR), isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR), affinity capillary electrophoresis (ACE), and fluorescence polarization assay. CD detects changes in the far ultraviolent spectra to quantify PPIs since the changes are proportional to the number of protein-protein complexes that are formed. Interactions between unlabeled proteins can be quantitatively observed with SPR, which detects refractive index changes at the surface of a gold-coated slide that occur during interactions. ITC measures the change in temperature following the step-wise addition of a prey protein to a bait protein since thermodynamically-favored protein interactions will release heat. Using a strong magnetic field, NMR can provide structural information of protein interaction contacts by observing the proton resonance frequencies of the proteins. ACE measures changes in electrophoretic mobility that occur upon the formation of protein complexes to determine the general strength (not rate) of the interaction. Moreover, ACE requires minimal sample amounts and each binding event can be analyzed within two minutes. Affinity information can be obtained with a fluorescence polarization assay by an increase in fluorescence polarization, which occurs when a green fluorescent protein (GFP)-fused protein binds to another protein. Notably, both FRET and mass spectrometry can also be used to determine dissociation constants of PPIs, but they are not commonly used and are low throughput in regards to the number of unique protein interactions that be analyzed at one time.

There is no high throughput, quantitative method for analyzing PPIs. However, a few quantitative platforms could be adapted to large-scale studies of PPIs, such as ACE and an array format SPR technology called SPR imaging (SPRi). Unfortunately, the

throughput capability of these methods has not been appropriately tested because of their reliance on purified proteins, which is a labor- and time-intensive process (see also page 67). Herein, I describe how I combined a protein microarray technology called NAPPA that doesn't require purified proteins with SPRi to create a high throughput, quantitative platform for analyzing > 400 PPIs in less than an hour. Both stable and transient PPIs are detected in real-time, and their corresponding kinetics (i.e., k_a, k_d) and affinities (i.e., K_D) are determined. Table 1 compares the attributes of NAPPA-SPRi with the aforementioned methods for analyzing PPIs. More detail about the disadvantages of protein purification and the methodology of this "cell-free" protein microarray is provided on page 67. A detailed description and comparison of SPR and SPRi is on page 75.

1.4 Modeling signaling pathways and cells

Over the past twenty years, the human interactome has grown from a handful of protein interactions to a staggering 319,419 unique physical and genetic protein interactions (Stark et al., 2006). These data provide, for the first time, a remarkable opportunity to understand how the thousands of interconnecting molecular gears regulate homeostasis, disease, and therapeutic response. However, they also emphasize the need for a computational model that can identify proteins that are essential in signaling and predict cellular response(s) from large-scale, complex data. An accurate cell model could pinpoint attractive drug targets as well as determine the molecular events underlying disease initiation and progression.

Cell or protein signaling models are generally built in four steps (Henriques, Villaverde, Rocha, Saez-Rodriguez, & Banga, 2017; Sachs, Perez, Pe'er, Lauffenburger, & Nolan, 2005a; Saez-Rodriguez et al., 2011). First, reported protein interactions in literature and databases are curated. The modeler selects the type of information with which to work. For example, only protein interactions with directional binding may be retained for further analyses. Second, a scaffold model is constructed, which often relies on ON/OFF or AND/IF/THEN logic gates. An example of a logic gate is "IF protein A is active AND protein B is active, THEN protein C is active." The number of scaffold models can be extraordinary large. In one case in which there were 78 proteins with 112 known interactions, 10^{38} scaffold models were built. An advantage of scaffold models is that potential proteins-of-interest for perturbation experiments can be identified. Third, cells are exposed to various conditions and their responses annotated. Finally, the scaffold models are trained using these *in vitro* experiments to accurately represent the data. In other words, models are assembled, in part, by reverse engineering. Modelers may not follow all of the aforementioned steps; instead, some models have been built using only *in vitro* perturbation experiments.

The disadvantage of current cell models is that they are essentially "black boxes," providing little insight into molecular mechanisms taking place inside the cell. Another important consideration is that models are only as good as the experimental data on which they're built. Unfortunately, as Chapter 1.3 discusses, current experimental data are biased in numerous ways. Proteins associated with disease are studied much more than proteins with unknown or poorly understood functions. Detection methods are

biased toward identifying stable interactions involving highly abundant proteins. Moreover, very little kinetic and affinity information has been determined, which is primarily due to low throughput methods for analyzing the binding rates and strengths (Heinrich, Neel, & Rapoport, 2002). As such, modelers are forced to build algorithms from *qualitative*-based data, resulting in "best guess" approximations that could miss individual, yet critical binding kinetics that regulate signaling. Calculated kinetics guided by cellular responses may be misassigned to particular signaling components or diluted across multiple proteins. An example in which binding kinetics have been experimentally determined but would likely be overlooked in a signaling model is actin polymerization. GTPases CDC42 and TC10 share 70% homology to each other, but only CDC42 can stimulate actin polymerization (Ou, Matthews, Pang, & Zhou, 2017; Schreiber, Haran, & Zhou, 2009). This is because CDC42 binds to WASp 1000-fold faster than TC10. Modelers, therefore, are working with incomplete, biased networks with little understanding of the temporal regulation of signaling.

Models built from experimentally-produced kinetic and affinity data would have distinct advantages over current models. The cellular effect of converging signals could be determined since the different on-rates of competitive binders would be known. Likewise, the binding kinetics would be used to ascertain the relative effects of divergent signals. Signal duration and the relative availability of proteins to interact could be calculated. Altered binding kinetics from protein mutations could be easily incorporated. Proteins that were not previously known to be essential in homeostasis, disease, and therapeutic responses would be more accurately identified. By using experimentally-

produced binding kinetics and affinities, models would not be “black boxes” built from best guesses and approximations, but a virtual computer chip reflecting the true molecular mechanisms taking place within a cell. An accurate cell model would likely have the biggest effect on the pharmaceutical industry. Drug pipelines would be streamlined since drug targets-of-interest with predictable responses could be identified, thereby increasing the number of approved drugs for public use in less time.

CHAPTER 2

2 B CELLS AND THE B CELL RECEPTOR SIGNALING PATHWAY

2.1 Introduction

In this thesis, the protein interactions in the B cell receptor (BCR) signaling pathway were examined in detail. The BCR pathway was chosen as the focus of this study for several reasons. First, the BCR pathway has relevance in homeostasis and disease. It regulates B cell maturation, VDJ recombination, antibody production, proliferation, cell survival, somatic hypermutation, class switching, germinal center formation, and antigen presentation. Disruptions to normal BCR signaling can lead to immunodeficiencies, autoimmunity, graft-versus-host disease, and cancer. Second, my collaborators were also studying B cells and the BCR pathway using flow cytometry and mass spectrometry. Since our long-term objective is to build an accurate, predictive virtual cell model, their results could complement my data and strengthen our model. Third, as one of the more well understood signaling pathways, the BCR signaling pathway provided a backdrop in which to compare the data collected here. Finally, the current map of the protein interactions and their temporal regulation in the BCR pathway remains incomplete. Approximately 80% of the protein interactions have been determined using co-immunoprecipitation techniques, which are notorious for identifying only stable complexes. Moreover, protein partners identified with co-affinity methods may not interact directly with the target-of-interest, but through a bridging protein. The BCR signaling pathway includes > 100 proteins and possibly > 2^{100} interactions, yet only 12 interactions have been characterized quantitatively (Table 2). These data were

obtained with a literature searching algorithm written by Dr. Parag Mallick of Stanford University.

Table 2. Documented binding affinities of protein interactions in the BCR signaling pathway

Interactor A ID	Interactor A name*	Interactor B ID	Interactor B name*	KD, binding affinity (M)	Detection method**	PubMed ID(s)
uniprotkb:Q8WV28	BLNK	uniprotkb:P62993	GRB2	2.00E-05, 6.00E-06	ITC	16912232, 16912232
uniprotkb:P31994	FCGR2B	uniprotkb:Q92835	INPP5D	4.70E-07	SPR	24642916
uniprotkb:P01100	FOS	uniprotkb:P05412	JUN	4.5E-08, 1.2E-06, 7.90E-07, 2.66E-05, 1.27E-06, 6.90E-08	ITC	21199371
uniprotkb:O89100	GRAP2	uniprotkb:Q8WV28	BLNK		ITC	12773374
uniprotkb:P62993	GRB2	uniprotkb:P62993	GRB2	6.60E-07, 6.00E-07	MST	22726438
uniprotkb:P62993	GRB2	uniprotkb:Q07889	SOS1	3.90E-05, 1.17E-04, 5.60E-05, 8.20E-05	ITC	19323566
uniprotkb:Q8BTI9	PIK3CB	uniprotkb:P60766	CDC42	2.90E-06	ITC	23706742
uniprotkb:Q8BTI9	PIK3CB	uniprotkb:Q61411	HRAS	2.71E-06	ITC	23706742
uniprotkb:Q8BTI9	PIK3CB	uniprotkb:P63001	RAC1	1.47E-06	ITC	23706742
uniprotkb:Q5EBH1	RASSF5	uniprotkb:P01112	HRAS	8.00E-08, 1.23E-06, 7.50E-06, 8.00E-07, 7.50E-06, 5.70E-06, 1.91E-05, 5.00E-07, 3.50E-06	Fluorescent assay	18596699
uniprotkb:Q04207	RELA	uniprotkb:O15111	IKBKA	6.20E-11	SPR	21056038
uniprotkb:Q04207	RELA	uniprotkb:P25799	NFKB1	6.20E-11	SPR	21056038
uniprotkb:Q07889	SOS1	uniprotkb:P62993	GRB2	1.25E-04	ITC	19323566

* Only proteins analyzed with NanoBRET and NAPPA-SPRi in this thesis

** Isothermal calorimetry, ITC; surface plasmon resonance, SPR; microscale thermophoresis, MST

2.2 B cells in homeostasis

Our body's ability to protect itself from infection and disease is made possible by the immune system, a multi-layered defense strategy that includes physical, chemical, and biological barriers; signaling molecules and proteins; and white blood cells (Alberts, 2015; Murphy & Weaver, 2016). As the only cells that produce antibodies, B cells are an essential part of the adaptive (or acquired) immune response, which recognize and mark specific pathogens for destruction. B cells are also antigen presenting cells (APCs), which internalize antigens, process them into fragments that are typically 8 – 11 amino acids in length, and then present them on their surface through class I and class II MHC proteins to activate T cells.

B cells are born in the bone marrow, arising from multipotential progenitor cells. From this point on, their purpose in life is to assist in adaptive immunity as an APC and by making specific and sensitive antibodies to non-self antigens. Late pro-B-cells in the bone marrow undergo a process called VDJ recombination in which the gene segments V, D, and J of the immunoglobulin heavy chain locus are rearranged to create the unique B cell receptor on the cell surface and, later, the secreted antibodies (Alberts, 2015). High antibody diversity is the result of recombining > 8500 VDJ gene segments and its associated "junctional diversification," where nucleotides are lost or added during segment joining. Thus, $> 10^{12}$ different antibodies could be produced in the human body!

Pre-B cells are not immediately released from the bone marrow upon assembly. Instead, they are exposed to a constellation of self-antigens. This "central tolerance" test is extremely important since release of self-reactive B cells would result in autoimmune disease, in which the immune system attacks normal host cells and tissues. B cells that react to self-antigens become unresponsive or die through apoptosis. Other B cells undergo receptor editing so that they no longer bind to self-antigens.

B cells that do not bind to self-antigens in the cellular environment of the bone marrow are transported to the central sinusoids, where they then enter into circulation as IgM^+ immature B cells. They then migrate to secondary lymphoid tissues like the spleen, lymph nodes, and Peyer's patches where they are considered to be "transition" B cells. It is in the circulation and secondary lymphoid tissues where B cells bind to their specific antigen and become activated mature B cells. A subset of B cells become antibody-producing plasma cells in situ, of which 90% will undergo apoptosis after 2 – 3 days.

Some activated B cells, however, will migrate to secondary lymphoid tissues, where they will stimulate the formation of B cell islands called germinal centers.

Within the germinal center, the B cell rapidly proliferates and becomes a centroblast that undergoes somatic hypermutation where nucleotide substitutions are made in the variable region in an attempt to make a more sensitive antibody. However, many of these changes are unfavorable and will result in cell death. If the changes to the variable region are favorable, the centrocyte will receive signals in the form of released cytokines from follicular dendritic cells and T follicular helper cells to promote B cell survival, stimulate class switch recombination (e.g., IgM to IgG), and differentiation into an antibody-producing machine called a plasma cell. In immature B cells, the antigen receptors are low-affinity IgM and IgD immunoglobulins. Upon class switch recombination, B cells may also express high-affinity IgG, IgA, or IgE antibodies.

The "primary response" following the first exposure of a B cell to its specific antigen is weak and brief. The lag phase between antigen exposure and antibody production is, on average, 7 – 10 days (Institute of Medicine (U.S.). Committee to Review Adverse Effects of Vaccines. & Stratton, 2012). However, this phase can be shorter, or as long as weeks to months. The antibodies that are produced are predominately low-affinity IgM, with low levels that quickly fade.

Not all centrocytes differentiate into antibody-secreting plasma cells. Some centrocytes are transformed into quiescent, memory B cells that are essential in long-term immunity. During repeat exposure to the same antigen, the primed memory B cell will initiate a faster and more vigorous "secondary response" than a naive B-cell would. The

lag phase is 1 – 3 days, with high levels of high-affinity IgG antibodies being produced over a long period of time.

The life span of a B cell depends on its activation state and type. Circulating B cells generally live for ~ 3 days, but some are long-lived, having a half-life of 1 – 2 months. Memory B cells can live for many years, thus imparting long-lasting immunity to the host.

2.3 B cells in disease

2.3.1 Autoimmune diseases

Autoimmune diseases occur when the immune system attacks normal host cells and tissues. B cells can cause and contribute to the pathology of these diseases in several ways: 1) produce auto-reactive antibodies; 2) secrete proinflammatory molecules like TNF-alpha, interferon-gamma, and macrophage migration inhibitory factor; 3) present self-antigens for T cell activation; and 4) aid in *de novo* generation of ectopic germinal centers that are present in chronically-inflamed tissue (Hampe, 2012).

Poor checkpoint controls during central tolerance and somatic hypermutation can lead to the inappropriate production of autoantibodies. Loss of skin pigment in vitiligo is due to the destruction of pigment cells called “melanocytes” by antibodies (Gottumukkala et al., 2003). Tear and saliva glands targeted by antibodies in Sjogren’s syndrome cause dry eyes and mouth (Suresh, Malyavantham, Shen, & Ambrus, 2015). Autoantibodies in the spinal fluid of Multiple Sclerosis (MS) may contribute to the demyelination of nerve fibers within the central nervous system (Kolln et al., 2006).

Antibodies to self-antigens have also been associated with the pathology of type 1 diabetes (Taplin & Barker, 2008).

B cell involvement in autoimmune diseases can be independent of antibody production. Antigen presentation by B cells have been shown to assist in driving lupus and experimental autoimmune encephalomyelitis (MS model) in mice (Giles, Kashgarian, Koni, & Shlomchik, 2015; Pierson, Stromnes, & Goverman, 2014). B cells also promote chronic allergic lung disease in this manner (Lindell, Berlin, Schaller, & Lukacs, 2008). The B cells of rheumatoid arthritis patients expressed higher levels of an essential cytokine for bone-resorbing osteoclastogenesis, RANKL, than the B cells in healthy patients (Meednu et al., 2016). The secretion of the pro-inflammatory cytokine, granulocyte macrophage-colony stimulation factor (GM-CSF), by B cells has also been linked to driving MS and causing relapses (R. Li et al., 2015). Ectopic germinal centers may maintain the pathology of autoimmune diseases by supporting the plasma cells that secrete autoantibodies.

2.3.2 Primary immunodeficiency diseases

Primary B cell immunodeficiency diseases are characterized by insufficient antibody production as the result of too few or defective B cells. Individuals with X-linked agammaglobulinemia (XLA) have a mutated protein called Bruton's Tyrosine Kinase (BTK) that is necessary in B cell maturation (Mak, Saunders, & Jett, 2014). Since their B cells cannot mature, XLA patients have severe deficiencies in all of the antibody isotypes. While XLA is relatively rare (i.e., 1 out of 200,000 live births), Common

Variable Immune Deficiency (CVID) affects 1 out of 25,000 people (Cunningham-Rundles, 2012). B cells in CVID also fail to mature and produce normal levels of antibodies, but the cause(s) is unclear. In Hyper-IgM syndrome, the B cells are unable to undergo class switch recombination from the low-affinity IgM isotype (Davies & Thrasher, 2010). Selective deficiencies in IgM, IgG or IgA subclasses can also occur (Asano et al., 2004; Louis & Gupta, 2014; Vidarsson, Dekkers, & Rispens, 2014). In X-linked Severe Combined Immune Deficiency (X-SCID), a gene mutation for the common gamma chain abolishes B cell function since the receptors for growth factors are abnormal (Fischer, 2000). The diseases outlined here cover only some of the primary immunodeficiency diseases that arise from defects in B cell development and function.

Not surprisingly, individuals with B cell-related primary immunodeficiency diseases are prone to various infections and have a higher risk of getting cancer. Their chance of getting cancer is increased because they can have chronic inflammation as the result of infection and because their immune system, which usually monitors for and destroys neoplastic cells, is compromised. Gastrointestinal complaints are often a common symptom since antibody levels are normally high in the gut; thus, primary immunodeficiency diseases are associated with malabsorption.

Patients with severe B cell primary immunodeficiency diseases are treated with intravenous immunoglobulin replacement therapy every 1 - 4 weeks for life, depending on the route of administration (Fried & Bonilla, 2009). Prophylactic antibiotics also reduce the risk of infection by Pneumococcus, Staphylococcus, and mycobacteria.

2.3.3 Cancer

As explained in the previous section, B cell-related primary immunodeficiency diseases can increase the risk of cancer. B cells can also be cancerous, which result in "B cell lymphomas." In Hodgkin lymphoma, the cancer originates from an abnormal, giant B cell called a Reed-Sternberg cell (Kuppers & Hansmann, 2005). The cells are large, multinucleated, and have a unique morphology. Interestingly, these cells have also been detected at low levels in non-Hodgkin lymphomas and infectious mononucleosis, although their role in these diseases is unknown. Non-Hodgkin lymphoma (NHL) is the most common hematological malignancy in adults, 85% of which are caused by B cells (Table 3)(Coffey, Hodgson, & Gospodarowicz, 2003). In 2013, non-Hodgkin lymphoma was the 8th and 11th most common cancer and cause of cancer deaths worldwide, respectively (Fitzmaurice et al., 2015).

Table 3. WHO classification of B cell-related NHL and percentage of total cases

Classification	% cases
Peripheral B cell neoplasms	
Precursor B lymphoblastic leukemia/lymphoma	
Mature B cell neoplasms	
CLL/small lymphocytic lymphoma	6.7
B cell prolymphocytic leukemia	
Lymphoplasmacytic lymphoma	1.2
Splenic marginal zone lymphoma	< 1
Extranodal marginal zone B cell lymphoma of MALT	7.6
Nodal marginal zona lymphoma	1.8
Follicular lymphoma	22.1
Mantle cell lymphoma	6.0
Diffuse large B cell lymphoma	30.6
Mediastinal (thymic) large B cell lymphoma	2.4
Intravascular large B cell lymphoma	
Primary effusion lymphoma	
Burkitt lymphoma/leukemia	< 1
Hairy cell leukemia	
Plasma cell myeloma	
Solitary plasmacytoma of bone	
Uncertain malignant potential	
Lymphomatoid granulomatosis	
Post-transplant lymphoproliferative disorder, polymorphic	

2.3.4 Other B cell-related diseases

B cells have been implicated in non-autoimmune diseases, graft-versus-host disease (GVHD), and the spread of human immunodeficiency virus (HIV). Altered cytokine profiles of B cells contribute to the pathology of non-autoimmune inflammatory diseases like type 2 diabetes and periodontal disease (Nikolajczyk, 2010). In comparison to healthy patients, B cells from type 2 diabetic patients were shown to secrete elevated levels of pro-inflammatory cytokine IL-10 while also being unable to secrete the anti-inflammatory cytokine IL-10 (Jagannathan et al., 2010). A common complication following bone marrow tissue or cell transplantation is GVHD, in which donor immune cells attack host cells. This is due, in part, to a breakdown in peripheral B cell tolerance as well as abnormal processing of B cell activating factor (BAFF) that promotes B cell activation and antibody production (Sarantopoulos, Blazar, Cutler, & Ritz, 2015). Finally, HIV has been shown to bind directly with the B cell receptors, CD21, CD-SIGN, and VH3 (Haas, Zimmermann, & Oxenius, 2011). The biological consequences of these receptor-viral interactions remain to be elucidated.

2.4 B cell receptor signaling pathway

The interaction of a B cell to its specific antigen initiates a series of intracellular signaling cascades, and results in specific cellular responses at the phenotypic and genetic levels (Dal Porto et al., 2004; Justement & Siminovitch, 2000). The BCR signaling pathway begins when the immunoglobulin-based BCR at the cell membrane recognizes and binds to its antigen, resulting in BCR cluster formation into glycolipid-rich

microdomains of the plasma membrane where the Src family tyrosine kinase, LYN, is anchored via acylation (Figure 4)(Kanehisa, Furumichi, Tanabe, Sato, & Morishima, 2017).

Phosphorylation drives the B cell receptor pathway. LYN then phosphorylates the immunoreceptor tyrosine-based activation motif (ITAM) on the cytoplasmic tail of the BCR-associated heterodimer proteins, CD79A and CD79B (or Igα and Igβ, respectively). CD79A and CD79B are the primary transducing structures that couple BCR-antigen binding to intracellular effectors. Interestingly, CD79A appears to mediate phosphotyrosine kinases while CD79B activates calcium mobilization and IL-2 production.

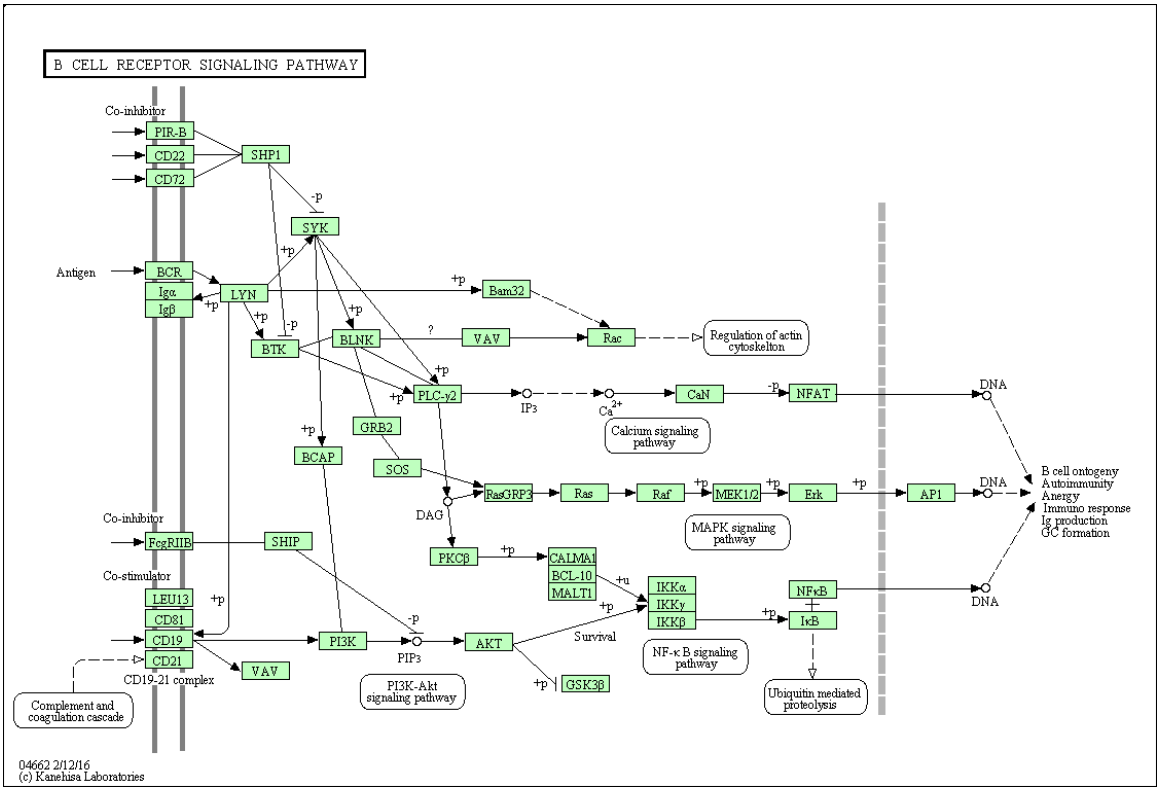


Figure 4. KEGG BCR signaling pathway (reprinted with permission)

Phosphorylation at both tyrosine residues within the ITAMs of CD79A and CD79B results in the binding of tyrosine kinase SYK through its phosphotyrosine-binding Src homology 2 (SH2) domains. SYK is then phosphorylated and activated by the nearby phosphotyrosine kinase, LYN. SYK is now capable of facilitating the initiation of several different sub-signaling pathways, which will be referred here as the PI3K/AKT, MAPK, NF-κB, RAC, NFAT pathways. Co-receptors like CD19 are essential in enhancing the BCR signal, while others like CD22 decrease it. Currently, the BCR signaling pathway is known to involve ~ 100 proteins, some of which assist in the convergence, divergence, and crosstalk of the sub-signaling pathways within it.

Phosphorylation is a post translational modification that can significantly affect PPIs and is often used to propagate signal through a cell, like the BCR signaling pathway. Phosphorylation, or the transfer of the terminal phosphate group of ATP to specific tyrosine, serine, and threonine residues of target proteins are mediated by a set of enzymes called kinases. The BCR pathway has at least 37 serine-threonine kinases and 4 tyrosine kinases (Appendix A), thereby making phosphorylation an important modification in in this pathway (Bounab, Getahun, Cambier, & Daeron, 2013). Many of the phosphatases, which are responsible for de-phosphorylating proteins, in the BCR pathway are considered to be negative regulators of signal propagation. For example, tyrosine-protein phosphatase non-receptor type 6 (PTPN6) negatively regulates BCR signaling by dephosphorylating CD79A and CD79B on their ITAM motifs as well as LYN and SYK (among others). INPP5D, INPPL1, and PTEN are other phosphatases that inhibit the BCR signaling pathway. The regulation of PPIs via phosphorylation, however,

can be very complicated. Phosphorylation can activate and inactivate proteins, which depends on the site and the presence of other phosphorylation events. The modification can also affect protein interactions. BLNK, for example, has at least 41 phosphorylated residues, some of which are known to affect the interaction partners of BLNK differently (Koretzky, Abtahian, & Silverman, 2006; Oellerich et al., 2009).

Adaptor proteins. Adaptor proteins, particularly BLNK, GRB2, NCK1, and DAPP1 (also known as Bam32), act as molecular scaffolds that bring proteins in close proximity to each other as to facilitate a PPI between them (Koretzky et al., 2006; Kurosaki & Tsukada, 2000). For example, BLNK binds to BTK and phospholipase C gamma 2 (PLCG2); BTK then phosphorylates and activates the key lipid metabolizing enzyme PLCG2.

Pathways within the BCR signaling pathway regulate a multitide of cellular responses. The PI3K/AKT, MAPK, NF-κB, RAC, NFAT pathways result in overlapping and distinct cellular responses, which are outlined in Table 4. Very briefly, the PI3K/AKT pathway is initiated when PI3K is activated via phosphorylation by receptor tyrosine kinases like CD19 (Figure 5)(Castellano & Downward, 2011). PI3K then converts phosphatidylinositol (3,4)-bisphosphate (PIP2) lipids to phosphatidylinositol (3,4,5)- trisphosphate (PIP3), which binds to and aids in the activation of a central

Table 4. Cellular responses of the signaling pathways within the BCR pathway

Pathway	Cellular responses
MAPK	proliferation, survival, apoptosis
NFAT	proliferation, differentiation, apoptosis,
NF-kB	proliferation, class switching, survival, pro-inflammatory cytokine secretion
PI3K/AKT	metabolism, growth, proliferation, survival, protein synthesis, transcription, apoptosis
RAC	proliferation, survival, differentiation, cell mobility

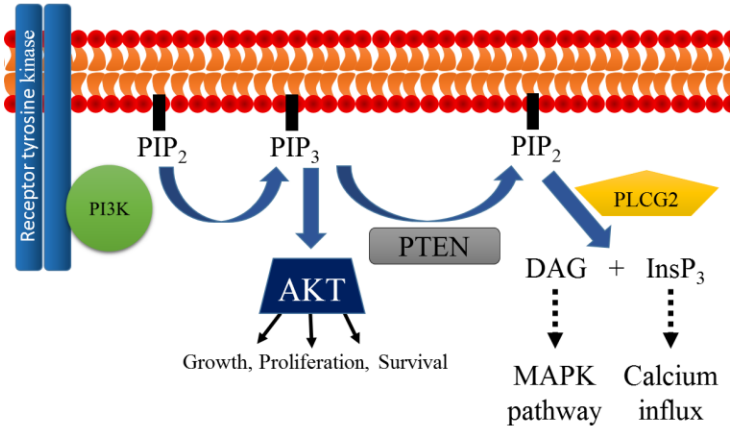


Figure 5. PTEN is a negative regulator of the PI3K/AKT pathway.

serine/threonine kinase, AKT. AKT promotes cell survival by inhibiting the ability of the pro-apoptotic protein, BAD, to heterodimerize with BCL2L1. AKT also increases nuclear localization of pro-survival transcription factors, NFAT and FOXO3A (Woyach, Johnson, & Byrd, 2012). The PI3K/AKT pathway mediates metabolism, growth, proliferation, survival, protein synthesis, transcription, and apoptosis.

Activated PLCG2 hydrolyzes PIP2 to InsP3 and diacylglycerol (DAG); it is InsP3 that stimulates an influx of calcium into the cytoplasm. A sustained rise in calcium level activates the serine/threonine phosphatases, PPP3CA, PPP3CB, PPP3CC, that then dephosphorylate the transcription factor, NFAT, to expose a nuclear localization signal (Scharenberg, Humphries, & Rawlings, 2007). NFAT translocates into the nucleus where it supports cell proliferation, differentiation, and cytokine production (Mognol, Carneiro, Robbs, Faget, & Viola, 2016; Woyach et al., 2012).

The MAPK pathway also contributes to BCR-induced survival (Woyach et al., 2012). The MAPK pathway is initiated when diacylglycerol (DAG) activates protein kinase C directly, or when VAV and GRB2 interact with RAC and SOS, respectively. These events stimulate the well-known RAF/MEK/ERK kinase cascade in which RAF's phosphorylation leads to the phosphorylation of MEK (known as proteins MAP2K1/2),

which can then phosphorylate ERK (known as proteins MAPK1/3). Phosphorylated ERK enters the nucleus where it targets specific transcription factors that facilitate cell proliferation, survival, and apoptosis.

The NF-κB pathway begins when the IκB kinase (i.e., IKBKA or IKBKB) is activated and phosphorylates inhibitors of NF-κB that complex with NF-κB homo- or hetero-dimers in the cytoplasm (Woyach et al., 2012). The phosphorylation marks the inhibitors for degradation, allowing NF-κB to translocate into the nucleus where it regulates the transcription of genes that are involved in cell proliferation, class switching, survival, and the secretion of pro-inflammatory cytokines. NF-κB consists of a protein with a transactivation domain (RELA, RELB, REL) and/or a transcriptional inhibitor (NFKB1, NFKB2). Depending on the proteins that dimerize with each other, as well as other associated proteins, the NF-κB complex can activate or inhibit transcription.

GTPases have pleiotropic roles in B cells. The activation of Rho GTPases marks the beginning of the RAC pathway. GTPases are considered to be "activated" when the enzyme is bound to GTP, and "inactivated" once the GTP is hydrolyzed to GDP (G. P. Li & Zhang, 2004). The GTPases are capable of activating or inactivating itself, although these processes are generally very slow and can be accelerated with the assistance of guanine nucleotide exchange factors or GTPase-activating proteins, respectively. GTPases modulate numerous downstream effector molecules that regulate cell mobility, differentiation, survival, and proliferation (Guo, Velu, Grimes, & Zheng, 2009; Nayak, Chang, Vaitinadin, & Cancelas, 2013; Walmsley et al., 2003). For instance, active Rho GTPases have been reported to interact with > 50 downstream proteins.

To note, B cell cancers can arise from constitutive activation or overexpression of kinases (e.g., PI3K, AKT, BTK) and GTPases within the BCR signaling pathway (Cinar et al., 2013; Rudelius et al., 2006; Vega & Ridley, 2008; Woyach et al., 2012). Some primary immunodeficiency diseases, as discussed in Chapter 2.3, stem from defects in BTK, which inhibits B cell development and production of specific types of antibodies.

Cell fate is determined by external environment, cell stage, and signals from multiple cell surface receptors. The BCR signaling pathway regulates numerous B cell phenotypes (see also Chapter 2.2), which is made possible by different external environments, developmental stages of the B cell, and signals propagated by other cell surface receptors. For instance, antigen-independent pre-BCR signaling in large pre-B cells in the bone marrow is critical in stimulating B cell development and maintaining B cell specificity (i.e., "allelic exclusion") (Martensson, Almqvist, Grimsholm, & Bernardi, 2010). The signaling mechanism, which is still incompletely understood, inhibits further VDJ recombination and produces a negative feedback loop to terminate pre-BCR expression. Antigen-dependent pre-BCR signaling causes cell apoptosis as a part of the "negative selection" process to remove B cells that are reactive to self antigens. Once in the periphery, antigen binding activates immature B cells to become short-lived plasma cells or to travel to secondary lymphoid organs where they stimulate the formation of germinal centers (Alberts, 2015). Antigen-BCR binding of mature B cells that are in germinal centers can result in somatic hypermutation, class switching, or the generation of plasma or memory B cells. An example of how signals from other receptors affect B cell response is the binding of the CD40 ligand from activated immune cells (e.g., T cells,

granulocytes) to the B cell CD40 receptor. Co-stimulation of the BCR and CD40 produce signals that, depending upon their relative levels of stimulation, induce proliferation, immunoglobulin switching, inhibition of apoptosis, or antibody secretion.

2.5 Selection of query proteins for protein interaction analyses

2.5.1 NanoBRET and NAPPA-SPRi queries

Five proteins in the B cell receptor (BCR) signaling pathway were chosen as queries to probe the entire BCR pathway protein set in both NanoBRET and NAPPA-SPRi analyses because they occur at key nodes in the BCR pathway and are important regulators of B cell response. These included an adaptor (BLNK), a tyrosine kinase (BTK), a lipid and serine/threonine kinase (PI3K), and two Rho GTPases (RAC1, RHOA). B cell-related immunodeficiencies and cancers are associated with their altered activity from mutations or overexpression. As such, they are attractive drug targets. In fact, small molecular BTK inhibitors are FDA-approved to treat certain B cell cancers and graft-versus-host disease. BLNK, BTK, and PI3K are also proximal to the membrane, mediating different signaling pathways. Therefore, their activity (or lack thereof) has a more profound effect on cell fate than proteins further downstream (Appendix A, Figure 4). The ability to test different protein types also provided insight into their unique methods of regulation. Inclusion of RAC1 and RHOA presented an opportunity to compare the binding partners and kinetic profiles of active and inactive GTPases.

BLNK. BLNK is an adaptor that binds to many proteins, thereby bringing them into close proximity to each other to interact. BLNK is essential for B cell development and in BCR signaling. Mutations in BLNK have been demonstrated to cause immunodeficiences, and downregulation of BLNK occurs in mediastinal large B cell lymphomas and Hodgkin lymphomas. Thus, BLNK appears to act as a tumor suppressor.

BTK. BTK is a non-receptor tyrosine kinase that mediates different pathways in the B cell and, as such, acts as a bottleneck. It is essential for B cell development and differentiation. BTK mutations are the cause of a severe immunodefiency disease called X-linked agammaglobulinemia (XLA) (see also Chapter 2.3.2). Increased BTK activity is observed in several B cell cancers, including diffuse large B cell lymphoma, mantle cell lymphoma (MCL), and chronic lymphocytic leukemia (CLL). Since 2013, the FDA has approved small molecule inhibitors of BTK to treat MCL, CLL, and graft-versus-host disease.

PI3K. PI3K is most well known for its lipid kinase activity, in which it phosphorylates the small signaling molecule, PIP2, to PIP3. It can also phosphorylate serine and threonine residues, most notably on itself and AKT1. It is critical for B cell metabolism, cell growth, development, and survival. Activating mutations are observed in 30% of cancers and some immunodeficiency disorders. PI3K is a heterodimer consisting of a catalytic and regulatory subunit, both of which have various isoforms. In these studies, the alpha isoforms of both the catalytic and regulatory subunits were used (i.e., PIK3CA and PIK3R1, respectively). These isoforms were chosen because, unlike the other isoforms, both PIK3CA and PIK3R1 are ubiquitously expressed. Moreover,

PIK3CA is the only catalytic isoform that is frequently mutated in cancer, while PIK3R1 is the most frequently mutated regulatory subunit in cancer (i.e., 20%) (Herrero-Gonzalez & Di Cristofano, 2011; J. J. Zhao et al., 2006).

GTPases. RAC1 and RHOA are both Rho GTPases with well-documented roles in regulating the cytoskeleton during cell growth, adhesion, and migration. They are essential for B cell development, proliferation, endocytosis, and antigen presentation. They also regulate apoptosis and survival. Increased RAC1 expression or activity are implicated in the initiation and progression of several types of cancers, including those of the lung, breast, prostate, skin, colon, but their roles in B cell-related cancers are unknown. Mutations in RHOA are associated with Burkitt's lymphoma and diffuse large B cell lymphoma. Both the inactive and active forms of RAC1 and RHOA were analyzed with GDP and GTPγS, respectively. GTPγS was used for these experiments because it is non-hydrolyzable; a hydrolyzable GTP would result in interactions representing a mixture of active and inactive states. Thus, seven queries were employed for both NanoBRET and NAPPA-SPRi analyses (Chapters 3 and 8).

2.5.2 NanoBRET queries

NanoBRET employs proteins that are produced *in vitro* without the need for purification. Therefore, obtaining functional and purified recombinant proteins was not a consideration with NanoBRET as it was with NAPPA-SPRi since any protein-of-interest could be studied with NanoBRET as long as the plasmid cDNA was available. In addition to the aforementioned queries, NanoBRET was used to analyze the protein

interactions with an additional five queries important in B cell regulation: AKT1, DAPP1, LYN, MAPK14 (i.e., p38), and SYK. AKT1 is a serine/threonine kinase that is activated downstream of PI3K in stimulated B cells. It promotes B cell growth, proliferation, survival, maturation, and survival. Increased activity of AKT1 is associated with a poorer prognosis in patients with diffuse large B cell lymphoma. DAPP1 is an adaptor protein that, via its signalosome, stimulates the RAC1/JNK pathway involved in B cell adhesion and spreading (Al-Alwan, Hou, Zhang, Makondo, & Marshall, 2010b; Ulivieri & Baldari, 2005). It is also involved in MAPK/ERK signaling, which regulates cell proliferation and survival. LYN is a tyrosine kinase proximal to the membrane that quickly becomes activated upon BCR aggregation. It then activates SYK, another tyrosine kinase, via phosphorylation. LYN is important in B cell differentiation and B cell tolerance, while SYK is essential in calcium mobilization and B cell development. Finally, MAPK14 is a serine/threonine kinase that promotes B cell proliferation and survival.

CHAPTER 3

3 QUALITATIVE ANALYSES OF THE BCR SIGNALING PATHWAY USING NANOBRET

3.1 Introduction

Mapping the human interactome has been pursued with enthusiasm following the first large-scale protein interaction studies in 2000 (Ito et al., 2001; Ito et al., 2000; Uetz et al., 2000; Walhout et al., 2000). The majority of these interactions have been determined using equilibrium-based assays (e.g., co-immunoprecipitation) that are inherently biased toward detecting stable complexes even though transient interactions are believed to occur with higher frequency. I therefore wondered how much of the protein network is still unexplored. To determine this, I decided to characterize the protein interactions of a "well understood" pathway, the B cell receptor (BCR) signaling pathway, and then compare our results to what is currently known. The identification of only a few new interactions, for example, would indicate that the BCR pathway and, as an extension, the human interactome were well annotated. Many new interactions, on the other hand, would suggest that much of the interactome remains unmapped.

To accomplish this task, I searched for a method that would meet the following criteria: high throughput, capable of detecting stable and transient interactions, high signal-to-noise ratio, easy-to-use, and amenable for *in vitro* analyses using proteins produced from plasmid cDNA using a cell-free expression system. Promega's NanoBRET™ fit four of the five criteria, with the exception that it has only been applied

in vivo. Since standard BRET has been successfully applied *in vitro*, I describe herein how NanoBRET was adapted to do the same.

My analyses focused on the protein interactions of 12 queries in the BCR pathway that are critical to various B cell phenotypes, including two GTPases that were represented in their active and inactive forms. Over 2500 interactions were analyzed, identifying 490 protein interactions, 83% of which have never been previously reported. This study suggests that only a fraction of the protein interactions in the BCR signaling pathway (and human interactome) have been characterized.

3.2 Promega NanoBRET™ technology

BRET, or "Bioluminescence resonance energy transfer," is a common method for analyzing transient and stable protein-protein interactions (Pfleger, Seeber, & Eidne, 2006; Y. Xu, Piston, & Johnson, 1999). It relies on protein "A" having a luciferase tag and protein "B" fused to a fluorescent label like the yellow fluorescent protein (YFP) (Figure 6). The proteins are mixed together along with a luciferase substrate. The

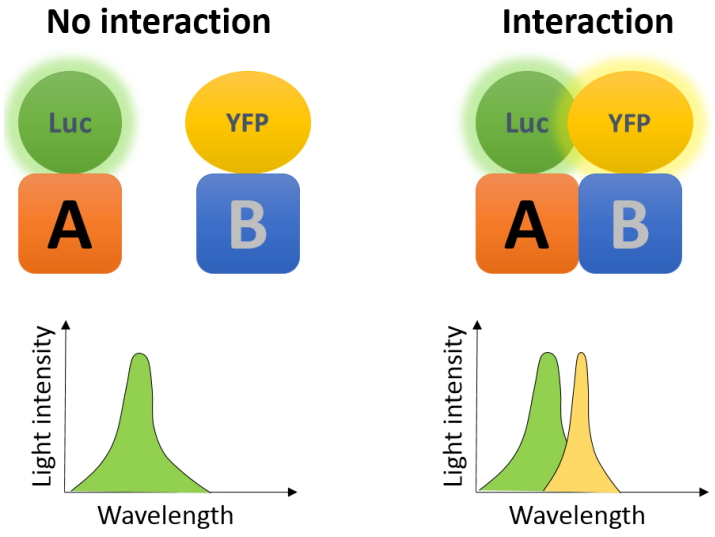


Figure 6. Schematic illustration of BRET technology.

oxidation of the substrate by the luciferase will generate light called "bioluminescence" around the luciferase tag. In the case of firefly luciferase, which is the most commonly used luciferase in BRET experiments, the substrate is a mixture of beetle D-luciferin, magnesium, ATP, and oxygen (Adams & Miller, 2014; Andreu, Zelmer, & Wiles, 2011). The oxidation of luciferin by firefly luciferase will produce AMP, carbon dioxide, oxy-luciferin, and "flash-type" light that decays over ~ 15 seconds. Other types of luciferase are derived from sea pansy (i.e., renilla), beetle, railroad worm, and copepod. If the proteins are in very close proximity to each other (< 10 nm), energy from the bioluminescence is absorbed and emitted by the fluorophore on protein "B" at a different wavelength. Protein interactions are detected when the amount of emitted light from the fluorophore (i.e., signal) is higher than that of the bioluminescence (i.e., noise).

Table 5. Advantages and disadvantages of BRET compared to other methods for analyzing PPIs

Advantages	Disadvantages
• Does not require excitation illumination	• Limited sensitivity and dynamic range
• No photobleaching	• Weaker signal than FRET
• Detects transient and stable interactions	• Influenced by fluorophore orientation
• Real-time detection	• Poor luminescence stability
• *In vitro* and *in vivo* analyses	• Donor/acceptor needs to be < 10 nm
• Fully reversible response	

Like any system, BRET has both advantages and disadvantages, which are summarized in Table 5 (Pfleger & Eidne, 2006; Xie, Soutto, Xu, Zhang, & Johnson, 2011). Its primary advantages include easy real-time measurement of PPIs *in vitro* and *in vivo*. It is often compared to a similar technology, fluorescence energy resonance transfer (FRET), in which the donor protein "A" has a fluorophore (instead of luciferase).

However, the fluorophore on protein “A" must be excited by an external light source, making analyses much more difficult. Furthermore, autofluorescence, photobleaching, and possible direct excitation of the acceptor fluorophore by the external light source (i.e., higher noise) may occur during FRET analyses. Both BRET and FRET are sensitive to protein orientation as the luciferase and fluorophores need to be < 10 nm from each other for the donor energy to be transferred to the acceptor fluorophore. BRET’s disadvantages include limited sensitivity and dynamic range, and poor luminescence stability.

Promega Corporation’s NanoBRET™ technology has a few distinct differences and advantages compared to standard BRET (Figure 7) (Hall et al., 2012; Machleidt et al., 2015). First, NanoBRET uses a different luciferase than standard BRET. Firefly or renilla luciferases that are 36 kDa and 61 kDa, respectively, are often used in BRET, although their large sizes can interfere with protein interactions and can be problematic for certain applications; for example, when inserting the reporter genes into viruses. NanoBRET, on the other hand, employs a 19 kDa luciferase subunit of the shrimp *Oplophorus gracilirostris* that Promega has dubbed “NanoLuc”®. It is fully active between pH 7 – 9 and

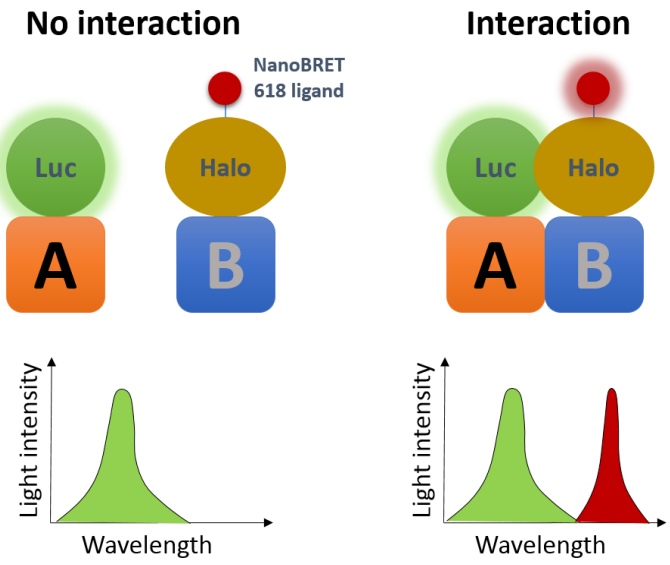


Figure 7. Schematic illustration of NanoBRET technology using a fluorophore-conjugated HaloTag ligand.

retains activity following 30 minutes at 55 ^{o}C. In comparison, firefly luciferase has a sharp decrease in activity below pH 8 and above 31 ^{o}C. Second, protein "B" is not a fluorescent protein, but is the HaloTag fusion protein that is fluorescently and covalently labeled with a fluorophore-conjugated chloroalkane ligand. Third, NanoBRET uses a novel luciferase coelenterazine analog substrate, furimazine, that produces high intensity, "glow-type" luminescence that has a signal half-life of > 2 hours. No magnesium or ATP is necessary for the reaction to occur. The combination of nanoLuc with furimazine results in a 100-fold increase in specific activity compared to that obtained with standard BRET using firefly and renilla luciferase. In addition, NanoBRET has a small overlap between the donor bioluminescence (460 nm peak) and the acceptor absorption spectra (618 nm). The large difference in wavelengths between the generated bioluminescence and emitted fluorescence during PPIs results in a sensitive NanoBRET system. In comparison, the spectral overlap in standard BRET is much higher than NanoBRET, resulting in a low signal-to-noise ratio. For example, the combination of firefly luciferase, D-luciferin, and red fluorescent protein result in a donor peak emission of 562 nm and an acceptor emission at 583 nm (Daunert & Deo, 2006). Renilla luciferase, coelenterazine, and enhanced YFP have a donor peak emission of 480 nm and an acceptor emission at 527nm.

3.3 Adapting NanoBRET^{TM} technology for *in vitro* analyses

NanoBRET was originally developed by Promega Corporation to analyze protein interactions in cells. Briefly, plasmids encoding for proteins "A" and "B" are introduced

into the cell-of-interest, and the resulting interaction is detected using a plate reader or bioluminescence microscope. In this section, the adaptation of NanoBRET technology for *in vitro* analyses of PPIs within the BCR signaling pathway is discussed.

The following filters and mirror were added to the Perkin Elmer Envision plate reader to make the instrument compatible with NanoBRET: emission filter 460/50m, emission filter 590 nm long pass, and a luminescence -/- single mirror. To ensure that the set-up was correct, a recombinant NanoLuc-HaloTag fusion protein with the HaloTag® NanoBRET™ 618 ligand from Promega Corporation was used. The negative control was the recombinant protein without the ligand. Luciferase substrate was added at 1-, 10-, and 20-fold dilutions to the protein, where the 1-fold dilution represents the dilution recommended by Promega (i.e., 500-fold). The noise, or luciferase emission, was read at 410 – 510 nm for one second. The signal, or NanoBRET 618 ligand emission, was read at $\geq$ 590 nm for 1 second. The signal-to-noise ratio was then multiplied by 1000 for the recombinant fusion protein with and without ligand, thus resulting in the milliBRET unit (mBU) (Figure 8A). The response, or corrected mBU, was obtained by subtracting the mBU of the recombinant protein without ligand (i.e., the negative control) from the mBU of the recombinant protein with ligand (Figure 8B). The Perkin Elmer Envision plate reader with the aforementioned modifications detected the NanoBRET system with

A) $$\frac{618\text{nm}_{\text{Em}}}{460\text{nm}_{\text{Em}}} = \text{BU} \times 1{,}000 = \text{mBU}$$

B) Mean mBU experimental – Mean mBU no-ligand control = Mean corrected mBU

Figure 8. NanoBRET calculations. A) PPI response where the acceptor, or signal, emission is 618 nm and the donor, or noise, emission is 460 nm. B) Mean corrected mBU calculations.

corrected mBUs > 100. This high response is not surprising since the luciferase and fluorophore are stably and covalently linked to each other; in other words, this configuration will result in the highest mBU possible. Real PPIs interactions analyzed *in vivo*, on the other hand, have much lower responses (i.e., 1.4 – 7 mBU range with 2 – 4 mBU being normal; personal communication with Dr. Thomas Machleidt of Promega Corporation).

To optimize NanoBRET for *in vitro* PPI analyses using proteins expressed in the cell-free expression system, the following parameters were tested: buffer compatibility with NanoBRET, type of buffer to block the plate from nonspecific interactions, amount of expression lysate per well, length of protein expression, query-to-target ratio, the amount of NanoBRET 618 ligand and luciferase substrate, and the incubation temperature for PPIs. Different proteins and PPIs require different buffer conditions. For example, some kinases require $MgCl_2$ while other proteins prefer Tris buffers over HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid). The NanoLuc recombinant

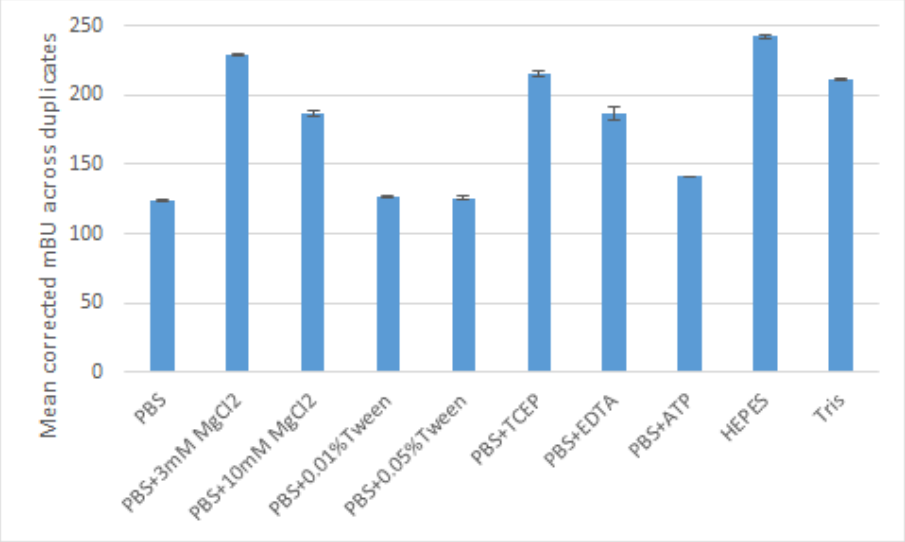


Figure 9. NanoBRET signal response across different buffers using a NanoLuc-HaloTag recombinant protein.

protein, as well as the NanoBRET 618 ligand and luciferase substrate, is routinely diluted in phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4), or PBS, for assessing instrument settings for the NanoBRET assay. Other types of buffers have not been tested by Promega Corporation, but it is an important consideration since certain reagents like phenol red can contribute to background noise. Therefore, I compared PBS, PBS + 3 mM $MgCl_2$, PBS + 10 mM $MgCl_2$, PBS + 0.01% Tween-20, PBS + 0.05% Tween-20, PBS + 1 mM tris(2-carboxyethyl)phosphine (TCEP), PBS + 5 mM ethylenediaminetetraacetic acid (EDTA), PBS + 200 µM ATP, 50 mM HEPES + 150 mM NaCl (HEPES buffer), and 50 mM Tris + 150 mM NaCl (Tris buffer). All buffers were at pH 7.4. Compared to the default PBS buffer, all of the buffers resulted in as high or higher signal response, indicating that all of the buffers tested are compatible with NanoBRET (Figure 9).

In the first step of the procedure, blocking buffer is added to the wells of the plate to inhibit nonspecific interactions between the proteins and plastic surface. 5% fat-free milk (w/v), 1% bovine serum albumin (BSA), or 1% BSA heat shock fraction (i.e., without immunoglobulin or proteases) in PBS with 0.1% Tween-20 (PBST) was applied to the wells of a half-area 96-well plate and incubated at room temperature for two hours. The plate was washed with PBST to remove excess blocking buffer. Target proteins, Fos and BTK with a C-terminal HaloTag, expressed in the *in vitro* transcription translation (IVTT) system were added to the plate and three of the six replicates were covalently bound to the NanoBRET 618 ligand. The query protein, Jun with a C-terminal NanoLuc, also expressed with IVTT, was then mixed with the target proteins, Fos or BTK, for one

hour at room temperature. Immediately after the addition of luciferase substrate, or NanoBRET NanoGlo, the plate was read with the Perkin Elmer Envision plate reader and analyzed as described above. Similar binding responses were obtained for the known Jun-Fos interaction across the different blocking buffers (Figure 10). Moreover, the blocking buffers resulted in similar background for the negative control wells containing Jun and BTK, which do not interact with each other. 1% BSA in PBST was chosen to be the blocking buffer for future experiments since the BSA heat shock fraction is significantly more expensive and the complex protein composition of milk may compete with target proteins for some queries (e.g., kinases).

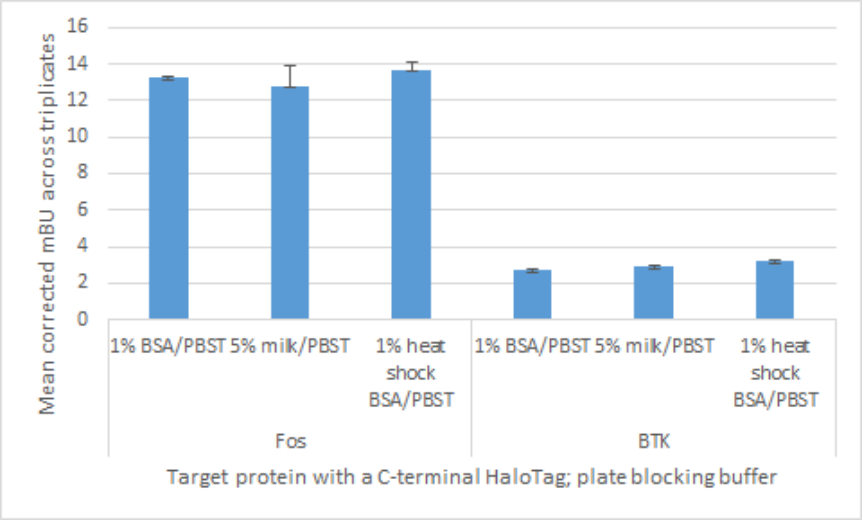
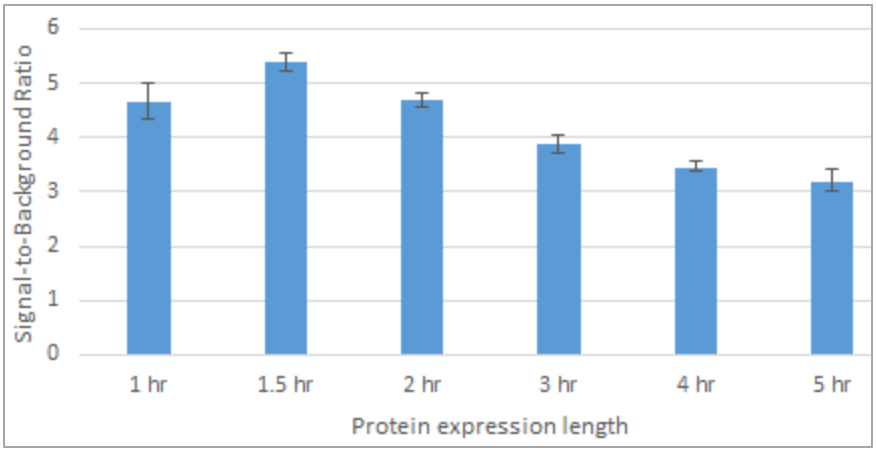


Figure 10. Effect of different blocking buffers on signal

In the second step of the procedure, target proteins are expressed in IVTT. According to the manufacturer's instructions, protein expression can occur from 90 minutes to 6 hours at 30 ^{o}C. Target and query proteins were expressed for 1, 1.5, 2, 3, 4, or 5 hours, and then processed as described above. The 1.5 hour expression resulted in the highest signal-to-background ratio, where signal represents the mBU from the Fos-

Jun interaction and background represents the mBU from the BTK-Jun non-interaction (Figure 11).



Once the target and query proteins are expressed, the target protein is diluted with a Tris- or HEPES-based buffer to 50 µL per well, the query and target proteins are added to the well, and the luciferase substrate is applied just prior to analyses. To determine the appropriate amounts of the aforementioned parameters (or factors), I used a design of experiments (DOE) approach that can easily analyze the effect of each factor on the response and to each other (see page 92 for a more detailed explanation of DOE). 30-, 36-, and 45-fold dilutions of the query and target proteins were compared. These dilutions were chosen based on previous experiments that identified this dilution range as the "sweet spot" (data not shown). In NanoBRET, the amount of NanoLuc query protein should be at or below the amount of HaloTag target protein to decrease noise. Thus, query-to-target ratios of 1:2, 3:4, and 1:1 were tested. For *in vivo* NanoBRET, the suggested dilution of the luciferase substrate is 500. In this DOE experiment, the substrate was diluted 500, 750, and 1000-fold. The target and query proteins were the same as previously mentioned, and all comparisons performed in duplicate. Based on the

signal (Jun/Fos)-to-noise (Jun/BTK) ratio, 30-fold protein dilution, 1:1 query-to-target ratio, and 500-fold dilution of the luciferase substrate had the maximum response (Figure 12). None of the factors interacted with each other (data not shown). Like the luciferase substrate in which the recommended dilution for *in vivo* PPI analyses was determined to be the best dilution for *in vitro* PPI analyses, the same was found for the NanoBRET 618 ligand dilution (i.e., 1000-fold)(data not shown).

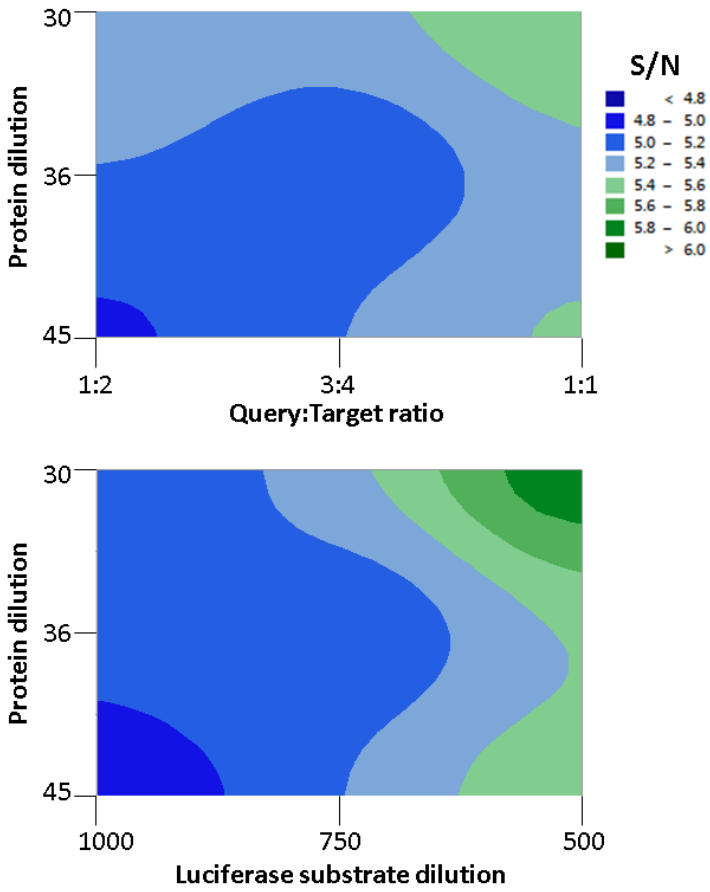


Figure 11. Contour plot of signal-to-noise (S/N) versus target dilution, query:target ratio, and luciferase substrate dilution. Blue = low S/N. Green = high S/N.

Finally, the temperature at which the query and target proteins should be incubated was investigated. The overall NanoBRET binding response is based on the accumulated light emission within the analyses window (i.e., 1 second). Thus, higher signal would result from more interactions and more stable interactions taking place during this time. Theoretically, this could be accomplished by slowing down the rate at which the proteins dissociate from each other with decreased temperature. To determine

this, the Jun query was incubated with Fos target or no target at all ("water") for 1 hour at 15 ºC, 25 ºC (i.e., room temperature), or 30 ºC, shaking at 500 rpm. For the stable Jun-Fos interaction, higher response was achieved at 15 ºC and 25 ºC (Figure 13). In the case of more transient interactions in which proteins dissociate from each other at an increased rate, even lower temperatures may be beneficial for NanoBRET analyses.

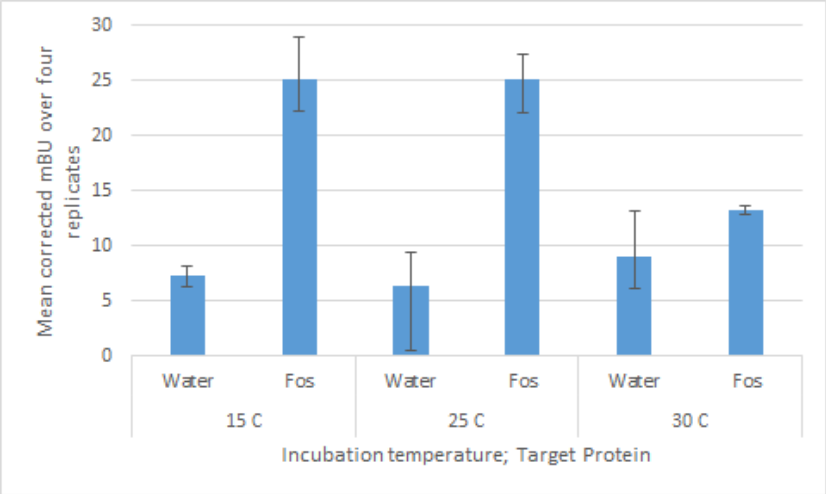


Figure 12. Incubation temperature affects NanoBRET signal.

The final protocol for *in vitro* NanoBRET analyses of PPIs using proteins expressed in cell-free lysate is provided in Chapter 3.4.2.

3.4 Analyses of PPIs in the BCR signaling pathway using NanoBRET

3.4.1 Introduction

The NanoBRET technology provides a high throughput method for analyzing both transient and stable PPIs in a qualitative manner. It was applied toward mapping the PPIs in the BCR signaling pathway to gain a more comprehensive understanding of the interactions that take place. 107 proteins in the BCR signaling pathwaywere represented separately with HaloTag at the N-terminus and C-terminus, hereafter called "target" (Table 6). 12 proteins in the BCR signaling pathway had NanoLuc as an N-terminal fusion tag; these proteins are hereafter called "query."

Table 6. Proteins in the BCR signaling pathway analyzed with NanoBRET

Target Proteins				
AKT1*	EGR1	MALT1	NFKBIA	PTEN
AKT2	ETS1	MAP2K1	NFKBIB	PTPN6
AKT3	EZR	MAP2K2	NFKBIE	RAC1*
ARHGEF7	FCGR2B	MAP2K3	NRAS	RAC2
BCL10	FOS	MAP3K3	PIK3AP1	RAC3
BCL2	GRAP2	MAPK1	PIK3CA*	RAF1
BCL2A1	GRB2	MAPK12	PIK3CB	RAP1A
BCL2L1	GSK3B	MAPK13	PIK3CD	RAP1B
BLK	HRAS	MAPK14*	PIK3CG	RAP2A
BLNK*	IFITM1	MAPK3	PIK3R1*	RAP2C
BTK*	IKBKA	MAPK8	PIK3R2	RASGRP3
CARD11	IKBKB	MAPK9	PIK3R3	RASSF5
CD19	IKBKG	MDM2	PIK3R5	RELA
CD22	INPP5D	MYC	PLCG2	RHOA*
CD72	INPPL1	NCK1	PPP3CA	SOS1
CD79A	JUN	NCKAP1L	PPP3CB	SOS2
CD79B	KRAS	NFAT5	PPP3CC	SYK*
CD81	LAT2	NFATC1	PPP3R1	TEC
CDC42	LILRB3	NFATC3	PPP3R2	TP53
CDKN2A	LIME1	NFATC4	PRKCA	VAV1
DAPP1*	LYN*	NFKB1	PRKCB	VAV2
				VAV3

* Also a query protein

3.4.2 Materials & Methods

Preparation of plasmid cDNA

Plasmid cDNAs encoding for the genes-of-interest (GOIs) were obtained from the Virginia G. Piper's Center for Personalized Diagnostics' plasmid repository, DNASU

(Tempe, AZ), and Open Biosystems (Lafayette, CO), and shuttled into vectors with a HaloTag at the N-terminus (pJFT7_nHalo), a HaloTag at the C-terminus (pJFT7_cHalo), or a NanoLuc at the N-terminus (pJFT7_nNanoLuc). Successful cloning of the GOIs was confirmed through GOI insert size analyses via DNA agarose, and GOI sequence analysis via Sangar sequencing using primers adjacent to the start and stop codons of the GOI. The Sangar sequencing was performed by the DNASU Sequencing Core at Arizona State University (Tempe, AZ). These vector backbones were created by Justin Saul at the VGP CPD and are compatible with Invitrogen's Gateway® recombination cloning technology. All target genes were represented with HaloTag at the N- and C-terminus, with the exception of genes with HaloTag only at the N-terminus (BLNK, PPP3CC) or only at the C-terminus (IFITM1, MAP2K1, PPP3R2). The target genes, AKT2, IKBKB, and PIK3R1, were represented with two different isoforms that differed significantly in size (Table 7). The query genes, AKT1, BLNK, BTK, DAPP1, LYN, MAPK14, PIK3CA/PIK3R1, GDP-bound RAC1, GTP-bound RAC1, GDP-bound RHOA, GTP-bound RHOA, and SYK, had an N-terminal NanoLuc. Note that GTP-bound RAC1 and RHOA used GTPγS (BIOLOG Life Science Institute; Germany) because GTPγS cannot be hydrolyzed.

Reagents

The 1-Step Human Coupled IVT Kit and Bond-Breaker^{TM} TCEP Solution, Neutral pH, were from Thermo Fisher Scientific (Waltham, MA). The ATP was from Cell Signaling Technology (Danvers, MA). The HaloTag® NanoBRET 618 Ligand and Nano-Glo

Substrate were from Promega Corporation (Madison, WI). Unless otherwise noted, all other materials and reagents were obtained from Sigma-Aldrich (St. Louis, MO).

NanoBRET analyses

White, high binding half-area 96-well plates (Greiner Bio-One; Austria) were blocked overnight at 4 ºC with 100 µL 1% BSA (w/v) in PBST, then washed with 100 µL PBST and 100 µL PBS. The HeLa lysate was spun at 10k x g for 2 min at 4 ºC, and the insoluble pellet was discarded. Target and query proteins were expressed for 1.5 hours at 30 ºC in the 1-Step Human Coupled IVT Kit, in which the HeLa lysate, reaction mixture, accessory proteins, and 200 ng/µL plasmid cDNA are mixed at a 5:2:1:2 ratio, respectively, such that 1 µL of target mixture or 1 µL of query mixture were added per well. The GTPase queries were GTP- or GDP-bound with 1 mM GTPγS (BIOLOG Life Science Institute; Germany) or GDP, respectively, in 50 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 5 mM EDTA, 1 mM TCEP, pH 7.4, for 1 hour at room temperature. Each query had five 96-well plates containing targets, with each set having a specific buffer. The BLNK, DAPP1, and JUN queries were analyzed in 50 mM HEPES, 150 mM NaCl, 1 mM TCEP, 0.01% Tween-20, pH 7.4. The GDP-bound RAC1, GTP-bound RAC1, GDP-RHOA, and GTP-RHOA queries were analyzed in 50 mM HEPES, 150 mM NaCl, 1 mM TCEP, 5 mM $MgCl_2$, 0.01% Tween-20, pH 7.4. The AKT1 query was analyzed in 50 mM Tris-HCl, 150 mM NaCl, 0.01% Tween-20, 250 µM ATP, 1 mM TCEP, pH 7.4. The BTK query was analyzed in 50 mM Tris-HCl, 150 mM NaCl, 0.01% Tween-20, 250 µM ATP, 1 mM TCEP, 4 mM MgCl2, pH 7.4. The LYN query was analyzed in 50 mM

HEPES, 150 mM NaCl, 1 mM TCEP, 0.01% Tween-20, 10 mM MgCl2, 250 µM ATP, pH 7.4. The MAPK14 (p38) query was analyzed in 20 mM Tris-HCl, 150 mM NaCl, 0.01% Tween-20, 250 µM ATP, 1 mM TCEP, pH 7.4. The PIK3CA/PIK3R1 (PI3K) query was analyzed in 50 mM Tris-HCl, 150 mM NaCl, 0.03% Tween-20, 250 µM ATP, 1 mM TCEP, 3 mM MgCl2, pH 7.4. The SYK query was analyzed in 60 mM HEPES, 5 mM MgCl2, 5 mM MnCl2, 1 mM TCEP, 250 µM ATP, 0.01% Tween-20, pH 7.5. Query and target proteins were diluted 30-fold each in buffer, such that 1 µL of query and 1 µL target protein were diluted to 60 µL total per well. One replicate received NanoBRET 618 ligand while the second (control) replicate received DMSO at a 1000-fold dilution. Samples were incubated at 15 ^{o}C for 1 hour. The Nano-Glo luciferase substrate was added at a 500-fold dilution and the plates immediately analyzed with a Perkin Elmer Envision plate reader equipped with an emission filter 460/50m, emission filter 590 nm long pass, and a luminescence -/- single mirror. The noise, or luciferase emission, was read at 410 – 510 nm for one second. The signal, or NanoBRET 618 ligand emission, was read at $\geq$ 590 nm for 1 second. A positive control, in triplicate, was placed on each plate, which was the Jun (query) – Fos (target) interactions with and without NanoBRET 618 ligand. The Jun-Fos interaction was used as a quality control of the sample processing as well as to determine the standard deviation within each plate. Duplicate or triplicate negative controls were included on each plate, which were the Jun (query) with no target, with and without the NanoBRET 618 ligand. All pipetting steps, with the exception of making the GTPases GDP- or GTP-bound, were performed with the Beckman Coulter Biomek FX liquid handler (Brea, CA).

NanoBRET data analyses

The corrected mBU value was determined as described in Chapter 3.3 (Figure 8). The mean (μ) and standard deviation (Δ) of the negative controls within each plate were also determined. Protein interactions were identified as those having corrected mBU values $\geq \mu + 2\Delta$, or two standard deviations higher than the mean of the replicate negative controls in their associated 96-well plate. To determine whether the detected interactions were known or novel, the data were compared with the online protein interaction databases, Biological General Repository for Interaction Datasets (BioGRID) and Human Protein Reference Database (HPRD) (Appendix C). Proteins isoforms with the same name but different sequences (i.e., AKT2, IKBKB, PIK3R1) (Appendix A) were counted as different proteins.

3.4.3 Results & Discussion

NanoBRET detected known and novel interactions. Over 2500 protein-protein interactions (PPIs) in the BCR signaling pathway were tested with NanoBRET using proteins produced with a human cell-free expression system. To determine the number of known and novel interactions detected with NanoBRET, the data were compared to the online PPI databases, Human Protein Reference Database (HPRD) and Biological

Table 7. Known and novel PPIs detected by NanoBRET

Type of PPI	AKT1	BLNK	BTK	DAPP1	LYN	MAPK14	PI3K	RAC1(GDP)	RAC1(GTP)	RHOA(GDP)	RHOA(GTP)	SYK
Known PPIs that were not detected	6	8	11	2	15	16	14	9	10	9	8	13
Known PPIs that were detected	20	8	4	1	7	2	13	5	4	4	5	8
Novel PPIs	40	50	22	22	38	12	28	39	59	32	32	35
Total PPIs detected	60	58	26	23	45	14	41	44	63	36	37	43
% known PPIs that were detected	77	50	27	33	32	11	48	36	29	31	38	38

General Repository for Interaction Datasets (BioGRID) (Prasad et al., 2009; Stark et al., 2006). NanoBRET detected ~ 40% (81/202) of known PPIs across all twelve queries, with some queries having as low as a 11% (MAPK14; 2/18) and as high as 77% overlap (AKT1; 20/26) with known PPIs (Table 7, Figure 14). As an example, PI3K, which is a heterodimer composed of the catalytic unit PIK3CA and the regulatory unit PIK3R1, had a 46% overlap with known interactions. Over half (i.e., ~60%) of previously-reported interactions were not identified with NanoBRET. Several reasons could account for this discrepancy. First, a significant portion (i.e., ~76%, Table 8) of reported PPIs were detected using pull-down (i.e., "affinity capture") methods. These types of methods feature the detection of stable protein complexes, many of which are held together by bridging protein. Therefore, many of these reported interactions may not be direct, and hence would not show up in a direct assay like this. Second, a fraction of the reported interactions may be false positives (other than inaccurately assigned direct interactions). For example, yeast-2-hybrid has a false positive rate of 25 – 40%. Third, experimental

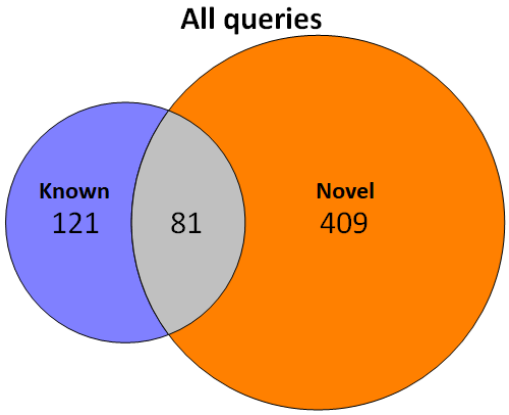


Figure 13. Venn diagram of known and novel interactions detected with NanoBRET.

conditions, such as buffers, can affect protein interactions. Fourth, NanoBRET requires the use of fusion tags, which may block binding epitopes. It is worth mentioning that even with similar experimental conditions, identical results will not be obtained. Two large-scale, independent yeast-2-hybrid screens using the same method had < 30% overlap (Ito et al., 2001). Detailed lists of the known PPIs and their associated experiments are in Appendix C.

Despite these potential limitations, NanoBRET detected 409 interactions not previously reported; these represented 83% of the total number of interactions. Although some of these novel interactions may be false positives, I detected a majority of them (64%; 167/262) using an orthogonal method, NAPPA-SPRi (see Chapter 8). The high number of new interactions is likely because NanoBRET can detect both transient and stable complexes.

The fusion tag may have interfered with some interactions. Some protein interactions were detected by NanoBRET when the HaloTag was at the N- or C-terminus, but not both (for a detailed list of protein interactions, see Appendix D). One possible reason is that the relative orientation of the luciferase to the NanoBRET 618 ligand (covalently bound to the HaloTag) is more favorable for NanoBRET analyses with particular tag locations for certain interactions since the fluorophore ligand and NanoLuc

Table 8. Most known PPIs have been detected using "pull-down" methods

	Methods using "pull-down" methods				Non "pull-down methods"										
Query*	Affinity Capture-MS	Affinity Capture-Western	Co-fractionation	Reconstituted Complex	Biochemical Activity	Co-crystal Structure	Co-localization	Co-purification	Dosage Lethality	Far Western	PCA	Phenotypic Enhancement	Protein-peptide	Synthetic Lethality	Two-hybrid
BLNK	8	18	0	8	0	1	1	0	0	0	0	0	1	0	0
BTK	3	12	0	3	6	2	0	0	1	0	0	0	0	0	2
PI3K**	16	32	2	17	4	0	4	0	0	0	1	0	0	0	9
RAC1	4	7	0	1	4	0	0	0	0	1	0	1	0	1	0
RHOA	4	4	0	2	2	0	1	2	0	0	0	0	0	0	0
Total Number	35	73	2	31	16	3	6	2	1	1	1	1	1	1	11

* These queries were used for both NanoBRET and NAPPA-SPRi analyses.

** PI3K represents the heterodimer that includes PIK3R1 and PIK3CA. The PPIs in this table are those that interact with one or both of these subunits.

Note: The experimental methods are defined by BioGRID's Experimental Evidence Codes.

need to be within 10 nm of each other to produce signal. This BRET phenomenon is well documented (Brown, Blumer, & Hepler, 2015). Please note that all query proteins had the NanoLuc fusion tag at the N-terminus; no interactions were assessed with the NanoLuc at the C-terminus. Another possible reason why a known interaction may not have been detected is that the tag may interfere with interactions, thus suggesting the location of the binding epitope on the target protein. For instance, the known LYN- SYK interaction was detected only when LYN as a target protein had the HaloTag at the N-terminus. The HaloTag at the C-terminus of LYN, near its kinase domain, may interfere with LYN's ability to phosphorylate SYK at tyrosine residue 323 (Geahlen, 2009). However, not all PPIs demonstrate this correlation. The C-terminal C2 domain of PLCG2 binds to BLNK, yet only the C-terminal tagged PLCG2 interacted with BLNK in the NanoBRET analyses (Engelke et al., 2013).

Rho GTPases. Protein targets of the GTPases, RAC1 and RHOA, in both the GDP- and GTP-bound states were analyzed. Although GTP-bound GTPases are considered to be activated and the mediator of most downstream functions, GDP-bound GTPases can also interact with other proteins, albeit to a lesser degree. In the NanoBRET analyses, RAC1 GTPase had more protein partners in its activated state, with GTP-bound RAC1 having 47% more interactions than GDP-bound RAC1 (Table 7). Some of these GDP- and GTP-based interactions overlap with each other (Figure 15). For example, RAC1 dimerized regardless of its GTP status. Indeed, this GDP/GTP-independent dimerization of Rho GTPases has already been documented (B. L. Zhang & Zheng, 1998). However, activated and inactivated RAC1 had distinct binding partners as well.

Only GDP-bound RAC1 interacted with ARHGEF7, a guanine nucleotide exchange factor (GEF) that activates RAC1 by exchanging the GDP with GTP. It also targets RAC1 to membrane ruffles and focal adhesions. ARHGEF7 was previously shown to bind to RAC1 independent of the GDP/GTP status; why ARHGEF7 only interacted with GDP-bound RAC1 in this analyses is unclear (ten Klooster, Jaffer, Chernoff, & Hordijk, 2006). Notably, ARHGEF7 interacted with GDP-bound RHOA, but not GTP-bound RHOA. It is possible that the GEF has a nucleotide preference, which affects its ability to bind and release the GTPase in the absence of competing nucleotides (GTP or GDP with the GDP- and GTP-bound GTPase queries, respectively). GRB2, an adaptor protein, bound to GDP-, but not GTP-, bound RAC1. It has been associated with RAC1 through pull-down complexes via their mutual interactions with PAK1 and SOS1; however, a direct interaction between GRB2 and RAC1 has not been detected before (Oak, Zhou, & Jarrett, 2003; Puto, Pestonjamasp, King, & Bokoch, 2003). Only GTP-bound RAC1 was able to bind to TEC and VAV1, both of which are known to be involved in the activation of RAC1 (Kline, Moore, & Clevenger, 2001). For a more detailed discussion on the GTPases, please refer to Chapter 2.4). Novel interactions were also detected, although these need to be validated and their biological roles in cellular response ascertained. A similar preference of target proteins for the

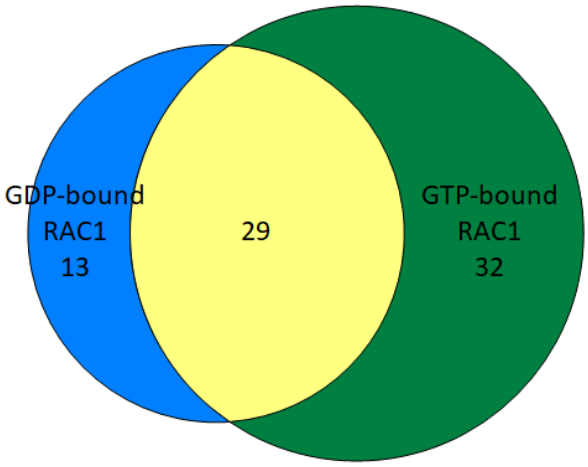


Figure 14. Venn diagram of shared PPIs between GDP- and GTP-bound RAC1 detected with NanoBRET.

GDP- or GTP-bound GTPase state was observed for RHOA. Both GDP- and GTP-bound RHOA had 36 protein partners with only 9 that interacted with RHOA regardless of activation state.

RAC1 and RHOA are well known to be involved in cytoskeleton rearrangement that is essential in cell mobility and proliferation (Heasman & Ridley, 2008). RAC1 and Rho protein family members have also been shown to promote cell survival and death (Jiang et al., 2003; Murga, Zohar, Teramoto, & Gutkind, 2002). In the NanoBRET analyses, the GTP-bound GTPases bound specifically to numerous proteins, including BTK, CD81, LAT2, LILRB3, NFATC3, PPP3CA, RAC2, RAF1, RAP1A, RAP1B, RAP2A, RASSF5, TEC, TP53, and VAV1. Moreover, they interacted with more proteins involved in the regulation of the actin cytoskeleton (e.g., Rap GTPases) than their GDP-bound counterparts, thus supporting previous reports. Proteins that specifically bound to the GDP-bound GTPases included ARHGEF7, BCL10, BCL2, CD19, DAPP1, GSK3B, HRAS, and PIK3AP1. Interestingly, both CD19 and PIK3AP1 (also known as BCAP) are involved in decreasing the threshold for antigen receptor-dependent stimulation by linking BCR signaling with PI3K activation. Upon binding an antigen, the CD19 and BCR co-localize into lipid rafts where CD19 recruits several signaling molecules, including PI3K, that can further augment signal transduction. PIK3AP1, on the other hand, promotes the translocalization of PI3K to the cell membrane where PI3K is then phosphorylated and activated. The preference of the GDP-bound GTPases for CD19 and PIK3AP1 suggests that the inactivated GTPases may help to decrease the threshold of the BCR-stimulated response.

Tyrosine and serine/threonine kinases. Several kinases were included as queries in this study. BTK, LYN, and SYK are tyrosine kinases. AKT1, MAPK14 (also known as p38), and PI3K are serine/threonine (S/T) kinases; PI3K also phosphorylates lipids. Whether their interactions with the protein substrates in the NanoBRET analyses is from the kinase phosphorylating them or from an interaction independent of their kinase activity is uncertain. AKT1 was the most promiscuous kinase, interacting with 57 unique proteins; on the other hand, MAPK14 was the least promiscuous, binding only 14 partners.

Two of the three tyrosine (Y) kinases, BTK and SYK, bound to RasGRP3 (regardless of tag location), a GEF in the MAPK signaling pathway. It is possible that RasGRP3 is phosphorylated by activated BTK and SYK following BCR stimulation, which may assist in RasGRP3's ability to activate the MAPK signaling pathway. Indeed, mass spectrometry studies have identified two tyrosine residues on RasGRP3, at tyrosine residues 45 and 179 (Cell Signaling Technology curation set 2234; 2007)(Bassani-Sternberg et al., 2016). Although the function of the phosphorylated tyrosines on RasGRP3 have yet to be elucidated, tyrosine phosphorylation of other GEFs have been shown to promote the activation of downstream GTPases (DeGeer et al., 2013; Kiyono, Kato, Kataoka, Kaziro, & Satoh, 2000; Teramoto, Salem, Robbins, Bustelo, & Gutkind, 1997).

Adaptor proteins. Two of the queries, BLNK and DAPP1, were adaptor proteins. Adaptors facilitate PPIs between other proteins by possessing a variety of protein-binding motifs that bring protein partners in close proximity to each other. BLNK

(also known as SLP-65) is a critical protein in B cell response by binding to proximal kinases to downstream effectors of various signaling pathways. Without BLNK, B cells are unable to develop properly, with a specific block in the pro- to pre-B cell transition (S. L. Xu et al., 2000). DAPP1 (also known as Bam32), on the other hand, is not crucial for B cell development, but instead plays a role in IgG3 class switching, cytoskeletal rearrangements, and calcium mobilization (Al-Alwan, Hou, Zhang, Makondo, & Marshall, 2010a; Fournier et al., 2003). BLNK was a promiscuous adaptor in this NanBRET analyses by interacting with 56 unique target proteins; DAPP1 interacted with only 22 proteins.

DAPP1 is phosphorylated by a Src family kinase (presumably LYN) upon antigen engagement with the BCR and assists in RAC1 activation, although the exact mechanism in which DAPP1 maintains GTP-bound RAC1 is unknown (Allam, Niiro, Clark, & Marshall, 2004). Allam et al. speculate that DAPP1 may interact directly with RAC1. Indeed, in this preliminary screen of PPIs within the BCR signaling pathway, GDP-bound RAC1 (and GDP-bound RHOA) interacted with DAPP1. Another possibility is that DAPP1 acts upstream to RAC1, interacting with proteins like VAV and SOS. However, these NanoBRET analyses did not detect an interaction between DAPP1 and any of the proteins in the VAV or SOS families. DAPP1 was the only query to interact with all three co-inhibitors of BCR signaling upstream to PTPN6 (also known as SHP1), which were CD22, CD72, and LILRB3 (also known as PIRB). PTPN6 is a negative regulator of cytokine signal transduction and, more recently, DAPP1 has been suggested to play a role in negatively regulating cytokine production as well (Kile, Nicola, & Alexander, 2001;

Onyilagha et al., 2015). It is possible that DAPP1's ability to control cytokine production may be the result of its interaction with CD22, CD72, and LILRB3.

Pairwise linear regression analyses to determine differences in binding interactions across queries. To understand whether there was a difference between two queries in terms of enriched biological processes and gene families in their interactions, targets were first defined using the Protein Analysis Through Evolutionary Relationships (PANTHER) and HUGO Gene Nomenclature Committee (HGNC) classification systems, respectively. The number of targets with different biological process or gene family was then determined per query and a pairwise linear regression analyses comparing different queries were performed (Appendix D). Enriched processes or gene families were defined as having more than two standardized residuals away from the predicted mean of the fitted linear regression line. Notably, this method of analyses cannot determine whether the enrichment in each group is statistically significant.

BTK interacted with significantly more proteins involved in the G-protein coupled receptor signaling pathway (Gene Ontology, GO:0007186) and intracellular signal transduction (GO:0035556) than both of the other two tyrosine kinases that were studied, with standardized residuals > 2.1 (Appendix B, D). It is interesting to note that BTK has been demonstrated to be an integral part in intracellular signal transduction and the G-protein coupled receptor signaling pathway (Qiu & Kung, 2000). LYN and SYK were not enriched in any process when compared to the other Y kinases.

PI3K interactions were enriched in the transmembrane receptor protein tyrosine kinase (TRTK) signaling pathway (GO:0007169) compared to AKT1 and MAPK14, with

standardized residuals > 2.5 (Appendix B, D). Protein targets of AKT1 and MAPK14, on the other hand, were enriched in intracellular signal transduction (GO:0035556) compared to PI3K, having standardized residuals > 4.3. In a general way, proteins in the TRTK pathway are upstream in BCR signaling compared to the proteins classified under intracellular signal transduction. For example, BTK, CD19, LYN, and SYK are all TRTK pathway proteins, while BCL2, GSK3B, MAP2K1-3, and RAC1-3 are all intracellular signal transduction proteins. These data support numerous experimental evidence that PI3K acts upstream of AKT1 and MAPK14 (p38α) (Berra, Diaz-Meco, & Moscat, 1998; Castellano & Downward, 2011; Stechschulte et al., 2014).

A comparison of the proteins that interact with S/T kinases and Y kinases reveal that S/T kinase interactions are enriched in cell surface receptor signaling pathway (GO:0007166) proteins, specifically those in the "CD molecule" protein gene family (Appendix B, D). The standardized residuals of these results were 2.4 and 3.9, respectively. CD, or cluster of differentiation, molecules are often receptors or ligands that initiate a signaling cascade. In B cells, these types of molecules act to promote or inhibit BCR signaling. These results warrant further investigation.

Compared to DAPP1, BLNK's interactions were enriched with "SH2 domain containing" proteins, as reflected by a standardized residual of 4.2 (Appendix B, D). The Src homology 2 (SH2) domain is highly conserved, ~ 100 amino acids long, present in a wide array of proteins, directs PPIs, and is important in protein-protein signaling (Gan & Roux, 2009; Schaffhausen, 1995). It recognizes, and binds to, phosphorylated tyrosine motifs in a promiscuous fashion, although the surrounding amino acid residues can affect

the overall affinities of these interactions. BLNK contains 13 potential tyrosine phosphorylation sites, eight of which have previously been determined to be phosphorylated and six as part of YXXP motifs within BLNK's C-terminal SH2 domain (Hornbeck et al., 2015; Kabak et al., 2002; Kéri & Tóth, 2003). The YXXP motif is considered to be a classic binding site for SH2 domains. DAPP1, on the other hand, has 11 potential phosphotyrosine sites, none of them in a YXXP motif, with only two of them being experimentally observed (Y139, Y195). These results reflect the biological functions of these adaptor proteins in BCR signaling: BLNK's phosphotyrosines essentially anchor a multimolecular complex proximal to BCR engagement called the "B cell signalosome" that is important in initiating *multiple* signaling cascades, whereas DAPP1 stimulates the RAC1/JNK pathway involved in B cell adhesion and spreading (see Chapter 2.4)(Al-Alwan et al., 2010b; Ulivieri & Baldari, 2005).

3.4.4 Conclusions

"NanoBRET" is a BRET-based platform that was recently developed by Promega Corporation to study PPIs in living cells, with higher sensitivity and dynamic range than standard BRET assays (Machleidt et al., 2015). Here, I adapted NanoBRET for high throughput *in vitro* studies, employing proteins produced in cell-free expression systems and 96-well plates. I compared the effect of different buffers on blocking non-specific adsorption to the plates and suspending the proteins for analyses. NanoBRET was determined to be compatible with all conditions tested, with HEPES buffer resulting in the highest signal-to-noise ratio. Additional parameters were examined, including the

length of protein expression, query-to-target ratio, the amount of NanoBRET 618 ligand and luciferase substrate. These results indicate that 1.5-hour protein expression, 30-fold protein dilution, 1:1 query-to-target ratio, 1000-fold dilution of the NanoBRET 618 ligand, and 500-fold dilution of the luciferase substrate had the maximum response. 60 µL per half-area well of a 96-well plate were applied in this experiment to ensure that the pipetting performed by an automated liquid handler was within the linear range of the instrument; however, a lower volume, depending on the liquid handler and the experimental setup, could feasibly be used (i.e., $\geq$ 30 µL).

Qualitative analyses of > 2500 potential protein interactions within the BCR signaling pathway were performed using NanoBRET between 12 query proteins and 107 unique target proteins represented separately with a fusion tag at the N- and C-terminus. This study is the first time that NanoBRET has been used to study PPIs *in vitro.* Distinct differences in binding partners between S/T kinases, Y kinases, and GDP- and GTP-bound GTPases were detected. The majority (83%) of the interactions have never been previously reported, thus indicating that much of the BCR pathway has not been mapped.

All of the queries interacted with proteins that are known to promote and inhibit BCR signaling, thus highlighting possible dual roles in signal transduction. For sure, the BCR signaling response is tightly controlled and relatively transient with proteins that can amplify and/or inhibit the BCR signal. LYN, the tyrosine kinase that is activated immediately upon BCR-antigen binding, regulates the PI3K/AKT pathway both positively and negatively (H. L. Li, Davis, Whiteman, Birnbaum, & Pure, 1999). Stimulation of the CD40 receptor can result in cell proliferation or apoptosis, depending

on the differentiated state of the B cell (Billian, Mondiere, Berard, Bella, & Defrance, 1997; Ingley, 2012). DAPP1 has also been implicated as having a dual role in inhibiting and promoting the BCR signaling pathway (Richards, Watanabe, Santos, Craxton, & Clark, 2008).

NanoBRET identified which proteins interact with each other and provides insight into the relative promiscuity and function of each query. However, further analysis to validate these results is warranted. It is important to note that all of the proteins for this study were expressed in the human Hela cell-based *in vitro* transcription translation (IVTT) system, which can modify proteins post translationally, most notably phosphorylation. The phosphorylation status of the query and target proteins was not controlled; thus, the detected PPIs may only reflect those that occur (or not occur) in HeLa cells. In addition, NanoBRET cannot measure the kinetics and affinities of the PPIs. This type of quantitative information would provide a more time-resolved picture of signal transduction, particularly for predictive algorithms or steady state models. To complement these qualitative NanoBRET data in an orthogonal fashion, I developed a high throughput and quantitative platform, NAPPA-SPRi, which is described over the next few chapters. Moreover, control over target protein phosphorylation, an important PTM in the BCR signaling pathway, is possible with this technology. By analyzing PPIs under different phosphorylation states (i.e., with and without phosphorylation), the inactive and active BCR signaling protein interaction network can be further delineated.

3.5 Acknowledgements

Abraham Lincoln is attributed to saying, “If I had five minutes to chop down a tree, I’d spend the first three sharpening my axe.” Much planning and preparation went into the final experiment, which in itself, only lasted a day. The Gateway-compatible NanoLuc and HaloTag plasmid backbones were constructed and validated by Justin Saul, with help from Promega Cooporation. Cloning, transformation, DNA purification, and DNA analysis of genes representing proteins in the BCR signaling pathway into the NanoLuc and HaloTag plasmid backbones were done, in part, by Benjamin Ober-Reynolds, David Haddad, and Dr. Andrea Throop. I extremely appreciate the help regarding programming and testing the Biomek FX robotic liquid handlers from Ian Shoemaker. Additionally, this experiment required processing a large number of 96-well plates (60!). Without Ian’s assistance during the actual run of the experiment, this NanoBRET analysis could not have been possible. I thank Dr. Thomas Machleidt from Promega Cooporation for providing valuable feedback regarding the NanoBRET platform. I also thank Dr. LaBaer who, through his discretionary funds, helped to support this experiment.

CHAPTER 4

4 DEVELOPMENT OF NAPPA-SPRI

4.1 Technology overview

4.1.1 NAPPA

Protein microarrays generally refer to a technology that displays hundreds to thousands of different proteins of known address on a solid planar or bead substrate. It enables the high throughput analysis of proteins, and has been used in basic research and translational research to study protein-protein interactions, protein-drug binding, posttranslational modifications, and antibody biomarkers of disease. More recently, bead-based protein microarrays and, to a lesser extent, planar arrays have been implemented in the clinic as *in vitro* diagnostic tools to test for serological protein and antibody biomarkers of allergies, autoimmune diseases, cancer, heart disease, infectious diseases, and Alzheimer's disease (Hartmann, Roeraade, Stoll, Templin, & Joos, 2009). Protein microarrays are primarily defined by the method in which the proteins are produced and immobilized.

Traditional protein microarrays print purified proteins expressed in *Escherichia coli* (*E. coli*), yeast and baculovirus insect cells (X. B. Yu et al., 2016). The use of nonhomologous systems to express mammalian proteins can be problematic since they may not have the appropriate chaperones for proper folding or ability to attach post translational modifications (PTMs) (Saul et al., 2014). Even if a protein were to get post translationally modified, it is unlikely that the type and location of the PTM would reflect those occurring in native systems. The purification procedure is often low throughput and

tedious regardless of the host system and requires additional protein manipulation that may negatively affect protein conformation and activity (Costa, Almeida, Castro, & Domingues, 2014). For example, a protein that can be expressed may not necessarily be able to be purified due to insoluble aggregation. Thus, it is not uncommon that individual proteins require separate optimizations, which involve protein denaturation and refolding, to increase protein recovery. Furthermore, the purification process is long and does not result in 100% purification. Protein purity, however, can be improved if fast pressure liquid chromatography (FPLC) is used. Due to the time involved in optimizing protein expression and purification, protein microarrays are not cheap. For example, the ProtoArray® Human Protein Microarrays (Thermo Fisher Scientific; Waltham, MA) that have > 9,000 full-length human proteins expressed with the baculovirus system were $1,200 per array at the time that this thesis was written (2018). This is in contrast to the Human GE 4x44K v2 cDNA microarray offered by Agilent Technologies, which cost ~$55 per 9,000 transcripts. The instability of proteins when compared to DNA also contributes to their high cost when considering shelf-life and experimental results. Spotting proteins directly onto the slide may result in non-specific adsorption and denaturation (Karlsson, Ekeroth, Elwing, & Carlsson, 2005; X. Li & Zhou, 2013). For most researchers, the biggest drawback of expressing human proteins in nonhomologous systems and/or purifying them is that the proteins may not be properly folded or functional.

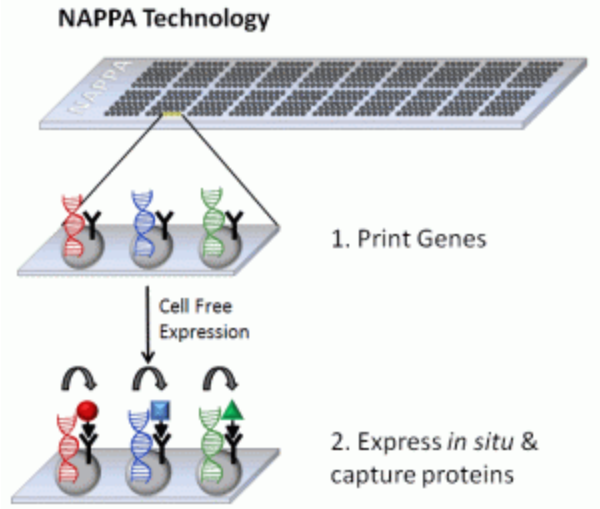


Figure 15. Preparation of GST-based NAPPA (VGP CPD website)

In 2004, another type of protein microarray based on cell-free protein expression, the nucleic acid programmable protein array (NAPPA), was introduced (Ramachandran et al., 2004; Ramachandran et al., 2008; X. Yu et al., 2018). In the NAPPA approach, plasmid cDNA encoding for any protein-of-interest (POI) is spotted onto a glass surface, which then can be transcribed and translated by an *in vitro* expression system (e.g., *E. coli*, rabbit, insect, human, wheat germ) at the time of the experiment (Figure 16). The use of a cell-free expression system has many advantages over expressing protein *in vivo*. First, proteins are produced with homologous ribosomes and chaperones, increasing the likelihood that the proteins are folded properly and are functional. Several NAPPA-based experiments studying protein-protein interactions and kinase activity have demonstrated that NAPPA proteins are functional (Ramachandran et al., 2004; Ramachandran et al., 2008; Rauf et al., 2018; X. B. Yu & LaBaer, 2015). Second, NAPPA proteins are produced in 2 hours, requiring only a plasmid backbone with the appropriate promoter and enabling same-day analyses. Third, proteins that may be toxic to cells can be synthesized. Fourth, non-natural amino acids can be incorporated. Fifth, NAPPA uses a fusion tag, which minimizes substrate-induced denaturation by distancing the protein-of-interest from the substrate surface (Karlsson et al., 2005). Sixth,

the NAPPA approach does not require the proteins to be purified in the traditional sense (e.g., FPLC) since the expressed GST-tagged proteins are captured to the slide in situ via an anti-GST antibody. Finally, additives, detergents, cofactors, and binding partners can be easily added to the cell-free expression system.

When this thesis was written, NAPPA could customarily express as many as 2,300 different proteins per standard 75 mm x 25 mm slide that are 640 µm apart (center-to-center) with a 450 µm spot diameter. Although not in general use, others have increased the number of proteins to 14,000 or more using a different type of slide and printing method (Song et al., 2017; Takulapalli et al., 2012). Additional technical work implies that NAPPA could one day reach as many as 48,000 different proteins per slide (personal communication with owner of Engineering Arts and co-developer of high throughput NAPPA, Dr. Peter Wiktor).

The plasmid cDNA employed by NAPPA is in a Gateway-compatible vector and encodes for a fusion gene encoding for the protein-of-interest and glutathione-S-transferase (GST). The Invitrogen™ Gateway™ cloning system allows the transcript inserts to be easily shuttled between Gateway-compatible vectors with the use of enzyme mixtures known as "BP Clonase II" and "LR Clonase II" that contain the restriction enzymes Int, and IHF; and Int, Xis, and IHF, respectively (Stewart, 2016). In the first step, PCR products flanked by the specific attB sequence replace the ccdB (controller of cell division or death B) gene of a donor vector and transformed into DH5α *E. coli* cells for propagation. The ccdB gene encodes for a protein that inhibits cell division, thus cells containing a vector with the "death cassette" will not grow and will not be selected for

further use. Successful shuttling of the PCR product into the donor vector is also assured through an antibiotic resistance gene encoded by the donor vector. PCR inserts in the resulting entry clone can then be transferred to any Gateway-compatible destination vector, which confers different antibiotic resistance than the entry clone. For NAPPA, genes without a stop codon are inserted into the pANT7_cGST destination vector containing a T7 transcriptional start site, ampillicin resistant gene, and a GST gene at the 3' end of the insert.

The plasmid cDNA is printed onto glass microscope slides of standard size (25 mm x 75 mm x 1 mm) functionalized with 3-aminopropyltriethoxysilane ("aminosilane"). In other words, the slides are coated with a chemical that results in free amines being exposed on the surface. In addition to DNA, the printing master mix contains bovine serum albumin (BSA), bis(sulfosuccinimidyl)suberate (BS3), dimethylsulfoxide (DMSO), and an anti-GST polyclonal antibody (Ramachandran et al., 2008). The Virginia G. Piper Center for Personalized Diagnostics (VGP CPD; Tempe, AZ) has found that while linear DNA binds well to the glass surface, it is not transcribed well in the cell-free expression systems. On the other hand, circular DNA is transcribed well, but has difficulty binding to the glass surface. The BS3 is a bifunctional amine-to-amine crosslinker that binds the amine-coated slides to the lysine residues on BSA and the anti-GST antibody to each other and, by forming a matrix, it theoretically helps trap the plasmid DNA to the surface. BSA is necessary as its removal results in the significant decrease in DNA deposition. Whether BSA "traps" DNA physically or electrostatically is unclear. DMSO is used to decrease the drying time following printing to improve spot

shape, spot size, and matrix formation. Finally, POIs are then produced and captured during a two-hour incubation using a cell-free expression system. (Genes can be shuttled into destination vectors that are compatible with the expression system-of-interest, which includes those arising from *E. coli*, rabbit, wheat germ, human, or insect.) The anti-GST antibody enables the capture of the expressed fusion protein to the slide surface. Since the 26 kDa GST fusion tag is at the C-terminus of the POI, the capture of the GST means that the POI was fully translated.

Notably, the use of a tag has both advantages and disadvantages (Costa et al., 2014; Kosobokova, Skrypnik, & Kosorukov, 2016). The primary advantage of fusion tags is that they minimize substrate-induced denaturation by distancing the protein-of-interest from the substrate surface. For example, coating glass with protein A prior to antibody addition reduces antibody denaturation (X. Li & Zhou, 2013). Non-specific protein adsorption and denaturation on positively-charged amine surfaces has also been previously documented (Karlsson et al., 2005). On the other hand, a fusion tag may negatively affect the conformation and activity of the protein. It can occlude binding epitopes, such that a protein that would otherwise interact with the POI may not be able to bind to the target protein when the tag is at one terminus. The tag at the other terminus, however, may expose the epitope for binding. Therefore, attempts to have a GST tag at the N-terminus of the POI have been made; they were unsuccessful. It is speculated that the reason why an N-terminal GST tag is incompatible with NAPPA is that the region of GST to which the capturing antibody attaches is obscured.

NAPPA often studies protein interactions at equilibrium, primarily stable antibody-antigen interactions (Diez et al., 2015; Miersch et al., 2013; J. Wang et al., 2014). After protein expression and capture, most NAPPA users subject the array to serum from patients having the disease-of-interest. After washing, a labeled anti-human antibody that can be fluorescently detected to determine the location of the antibodies and their target antigens probes the slide. The comparison of binding patterns between control (i.e., healthy) and disease samples helps to identify specific antibodies that potentially could be used as diagnostic biomarkers of disease. A similar approach tracks the immune response over time in response to infection or treatment.

Interactions other than antibody-antigen interactions detected with NAPPA (e.g., Fos binding to Jun) support the conclusion that the expressed proteins are functional and most likely folded properly (Ramachandran et al., 2008; X. B. Yu & LaBaer, 2015). Other studies performed in the lab have shown that known kinases maintain their ability to auto-phosphorylate (Rauf et al., 2018).

While NAPPA has numerous advantages, it still has its limitations. Like any equilibrium-based assay, NAPPA (in its present form) can only detect protein complexes with high binding affinities that can withstand numerous washes inherent to the procedure. Stable interactions, however, only represent a small fraction of the protein interactions that actually occur *in vivo*. Thus, NAPPA is missing interactions within signaling networks that may be fundamental to disease onset, progression, and response to treatment. The current cell-free expression system that is primarily used with NAPPA is lysate from human HeLa cells that is mixed with accessory proteins and a reaction

mixture containing various other reagents (e.g., reducing agent) to enhance transcription and translation. The expression system is capable of performing PTMs, with phosphorylation being documented the most. The type and location of these PTMs are likely similar to those seen in HeLa cells *in vivo*, but may not be relevant to the disease or interaction-of-interest. As previously acknowledged, PTMs often affect protein activity and interactions.

NAPPA has relied on the GST tag to capture the expressed protein to the slide surface, which is not advantageous for some applications. First, the GST tag is compatible with the platform only when the GST is at the C-terminus. Since some epitopes on the displayed proteins may be blocked by GST, it would be advantageous to have the displayed proteins without a fusion tag, a smaller tag, or have the proteins also represented on the array with the GST at the N-terminus. Second, the antibody-GST capturing method is not covalent.

To address the limitations of the GST tag, the laboratory has recently investigated the possibility of using a covalent capturing method using the HaloTag® fusion (Promega Corporation; Madison, WI)(J. Wang et al., 2013). The replacement of a single amino acid in a haloalkane dehalogenase enzyme originating from *Rhodococcus rhodochrous* results in a stable covalent bond between the HaloTag protein and a chloroalkane substrate (England, Luo, & Cai, 2015). Indeed, harsh denaturing conditions showed that HaloTag fusion proteins remained on the slide surface while GST fusion proteins were removed (see also Figure 23, page 91)(J. Wang et al., 2013). In addition,

proteins with HaloTag at the N- and C-terminus of the proteins-of-interest can be successfully captured to the NAPPA surface.

4.1.2 SPR and SPRi

The term "surface plasmon resonance" (SPR) refers both to a light-electron phenomenon and the sensor platform that employs it (Homola & Dostálek, 2006; Schasfoort & Tudos, 2008). The SPR phenomenon was first observed in 1902 by R. M. Wood at Johns Hopkins University, in which polarized light shone onto a metal-backed diffraction grating resulted in reflected light composed of dark and light bands. Nearly four decades later, Hugo Fano recognized that the "Wood's anomaly" was due to the incident light being converted into wave-like oscillations of the free electrons within the metal surface. In 1983, Bo Leidberg et al. exploited the SPR phenomenon for the first time as a biosensor to specifically detect the interaction of human IgG with anti-human IgG (Liedberg, Nylander, & Lundstrom, 1983). SPR has since become recognized as a sensitive, label-free approach for analyzing the interactions between DNA, drugs, peptides, and proteins in real-time. In the pharmaceutical industry, SPR identifies and studies the differential binding of drugs to their pharmacological targets (Olaru, Bala, Jaffrezic-Renault, & Aboul-Enein, 2015). This thesis, however, will focus on the detection of protein-protein interactions by SPR.

All SPR instruments have three main components: optical unit, liquid handler, and the sensor surface. The sensor surface is generally a portable chip or slide in which proteins are adhered; these captured proteins are called "ligands" or "targets." The liquid

handler is essential in injecting any reagent-of-interest in buffer across the surface, including a purified “analyte” or “query” protein. The optical unit enables the measurement of protein interactions through the use of a prism, grating, or optical waveguide. In the case of a prism, the optical unit essentially filters out light noise (i.e., s-polarized light) so that the p-polarized light can be properly focused and synchronized with the surface plasmons.

The sensor chip is generally composed of a glass substrate coated with a thin (~50 nm) layer of gold. Other metals, like silver, copper, tin, lead, mercury, cadmium, indium, and the alkali metals can also produce SPR, but have more disadvantages than gold. For example, copper results in a weaker signal. Mercury cannot be used with light in the visible range. Silver, the second most common metal used with SPR, isn’t very stable and is easily oxidized in air.

The metal is then coated with an adhesion linking layer and immobilization matrix. The adhesion linking layer is usually composed of a > 10-carbon alkane with a terminal thiol group, which stacks against each other in a self-assembling monolayer that is oriented perpendicular to the surface via the thiol during an extended incubation (i.e., 4 – 24 hours). (Whether the thiol-gold interaction is covalent is still up for debate.) The adhesion linking layer enables a bioinert, hydrophilic immobilization matrix to be attached to the surface, the purpose of which is to 1) reduce non-specific binding of the query protein to the “sticky” metal and 2) capture the target protein to the slide. Matrices can be 2D or 3D in structure, which are composed of polymers like dextran, carboxymethyldextran, poly(vinyl alcohol), poly(ethylene glycol), poly(acrylic acid), and

poly(L-lysine). The type of immobilization topcoat depends on the specific needs of the experiment. The NAPPA-SPRi platform in this thesis used a self-assembling monolayer that combined both the adhesion linking layer and immobilization matrix: HS-C_{11}-$(C_2H_4O)_6$-NH_2, where HS-C_{11} represents the adhesion linking layer and $(C_2H_4O)_6$-NH_2 represents the matrix (ProChimia Surfaces; Poland).

The target protein (usually purified) is then captured to the surface via the immobilization matrix. For example, a biotin-functionalized self-assembled monolayer will bind to proteins with a streptavidin tag. The addition of the immobilization layer and target to the slide are often done under flow.

The SPR phenomenon is made possible by the optical unit, which essentially directs the incident p-polarized light to the sensor surface so that plasmon resonance can occur. The prism-based optical unit is the most commonly employed optical set-up and is used in the SPRi platform used in my exeriments, thus only describe the "Krestchmann configuration" will be described here. Detailed explanations of

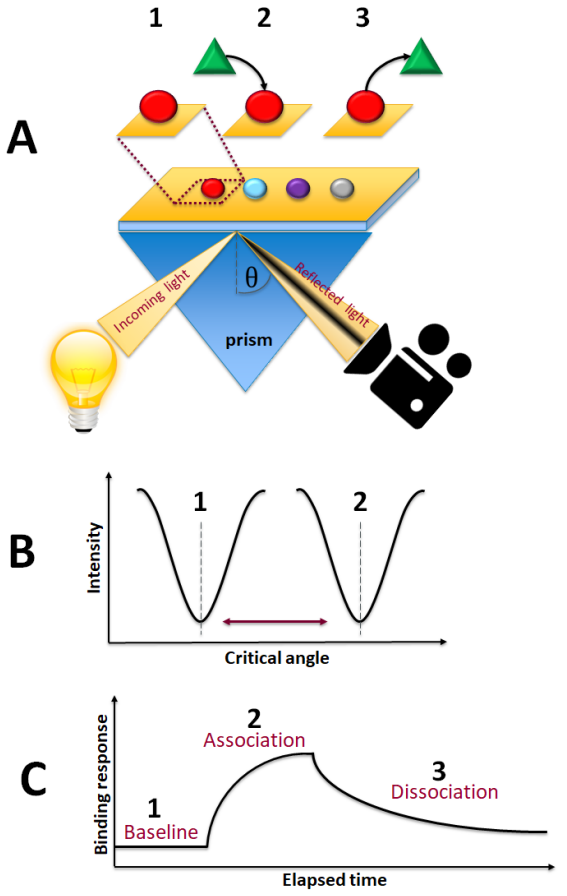


Figure 16. SPR analysis of PPI. A) Kretschmann configuration of the SPR instrument. B) PPI causes a change in AMR, or critical angle. C) Binding sensorgram depicting a PPI over the course of the experiment.

the other optical units can be found elsewhere (Schasfoort & Tudos, 2008). SPR instruments with the Krestchmann configuration require a light source to produce the incident light, a prism, a sensor surface (described above), and a device to measure the reflected light (Figure 17A). The light source can be a laser or halogen lamp that is directed toward a prism at a specific angle, resulting in photons with direction and momentum. Both the direction and momentum of these photons will change as they move from different mediums (i.e., vacuum to prism). The sensor chip is coupled to the prism through immersion oil with a similar refractive index (RI) of the slide. The key to SPR is that, at the critical incident angle or "angle of minimum reflectance" (AMR), the direction and momentum of the photons become parallel to the surface and equal to the electron plasmons of the metal, respectively. This results in total internal reflection and SPR, such that no light is reflected (Figure 17B) (i.e., the dark band observed by Wood in 1902). It is important to note that the momentum of the electron plasmons is sensitive to the RI of the medium surrounding them. Thus, an interaction between a query and target protein will change the local RI of the medium, resulting in a different AMR. SPR instruments measure the change in this angle during the association and dissociation of a protein-protein interaction (PPI).

SPR has two components that are inherently linked to one another: wave-like oscillations in the plane of the metal layer and an electric field perpendicular to the surface. The electric field, or "evanescent wave," decays exponentially from the interface surface, extending approximately one light wavelength (of the light source) into the

medium on either side of the metal layer. For most instruments, the depth of the evanescent wave is 200 – 300 nm.

There are three main steps in SPR analyses (Figure 17C). In the first step of analyses, only buffer is injected across the surface with the captured target proteins to obtain a baseline signal. In the second step known as the "association phase," a purified query protein in buffer is added, allowing the query to associate with the target on the surface. In the third and final step known as the "dissociation phase," the addition of query is ceased and only buffer is injected, allowing the flowing buffer to remove query as it dissociates from the target. SPR sensorgrams of an interaction can look like Figure 17B, where the change in AMR is depicted on the x-axis and the reflected light intensity is on the y-axis. However, the most common sensorgram format is like Figure 17C, where time is on the x-axis and the binding response (or, change in RI) is on the y-axis. The y-axis can also be interpreted as the occupancy frequency of ligand epitopes by the query (i.e., response).

The buffer for SPR analyses is often phosphate- or Tris-buffered saline with 0.01 – 0.1% Tween-20. The use of a nonionic detergent like Tween-20 decreases nonspecific hydrophobic interactions and helps to prevent bubble adsorption on the surface. High concentrations of detergent, however, may interfere with real binding events. Therefore, both buffer and detergent choice and amount can have drastic effects on the protein interactions and should be chosen carefully, taking into consideration the proteins-of-interests and the instrument set-up. The pH of Tris-buffered saline, for example, changes based on temperature. The SPRi instrument that was used in the experiments described

herein (i.e., PlexArray® HT System) can control the temperature at the sample holding area and sensor surface, but the temperature of the 1 mL sample loop is not controlled. The flux of pH in the Tris-based buffer from 4 ºC in the holding area to RT in the sample loop to 30 ºC on the sensor slide may not be desirable. Most SPR experiments use phosphate-buffered saline (PBS), but this buffer may not be appropriate for the experiment either. The phosphate groups in PBS may interfere with studying the effects of phosphorylation on protein interactions.

During the "association phase," the query and target proteins are introduced to each other. Proteins that do not interact with each other, or "non-binders," would ideally produce a sensorgram with a flat horizontal line, but this is usually not the case. Instead, the sensorgrams of non-binders look like Figure 18A, which depict differences in RI between the baseline and association steps and resemble a blockier curve than binders (Figure 18B) since the RI quickly reverts to baseline during the "dissociation step." The query protein should be buffer exchanged into the same buffer used in the baseline step; however, small variations in temperature or buffer composition will result in an altered RI. Although a

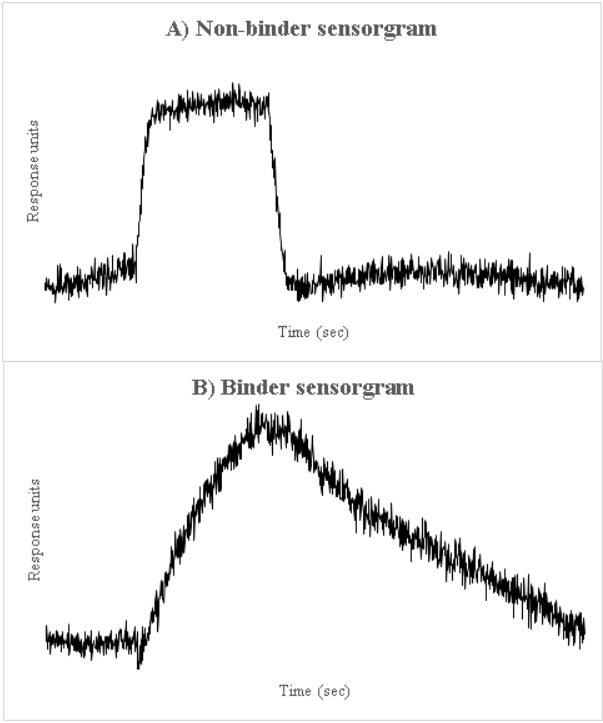


Figure 17. SPR sensorgrams reflecting A) bulk refractive index shift and B) real PPI.

region-of-interest on a slide is designated as a specific protein target, it should be remembered that there are numerous target molecules within this region (i.e., not just one protein molecule). Therefore, the y-axis reflects the sum of the target residues interacting with the query, where the amount of protein complex increases over the length of the association time (or until the complex has reached equilibrium). Instead of the blocky shape of a simple RI change in medium observed in a sensorgram of a non-binder, an interaction should have an exponential binding curve.

When the query-target on-rate is limited by the rate of flow, an issue called "mass transport" occurs (Schasfoort & Tudos, 2008). In other words, the query is binding to the target as quickly as diffusion will allow and is, therefore, dependent on the query concentration. Increasing the flow rate, decreasing the target concentration, increasing the query concentration, altering the flow chamber design above the sensor surface (to reduce stagnant areas), and decreasing the query size can all help minimize the mass transport effect. Several of these suggestions, however, may not be possible. Maximum flow rates are often limited. For example, the maximum flow on the PlexArray® HT System is 5 µL/sec; above this, the flow chamber can leak for durations > 30 sec. The volume of the sample and injection loop will also affect the flow rate. Flow chambers are often commercially optimized and produced, so alterations to their design by the user are not realistic. Reducing the query size (by only using certain domains) so that it can diffuse faster to the surface can also greatly affect results. Due to diffusion constraints, higher flow rates increase the dynamic range of the instrument.

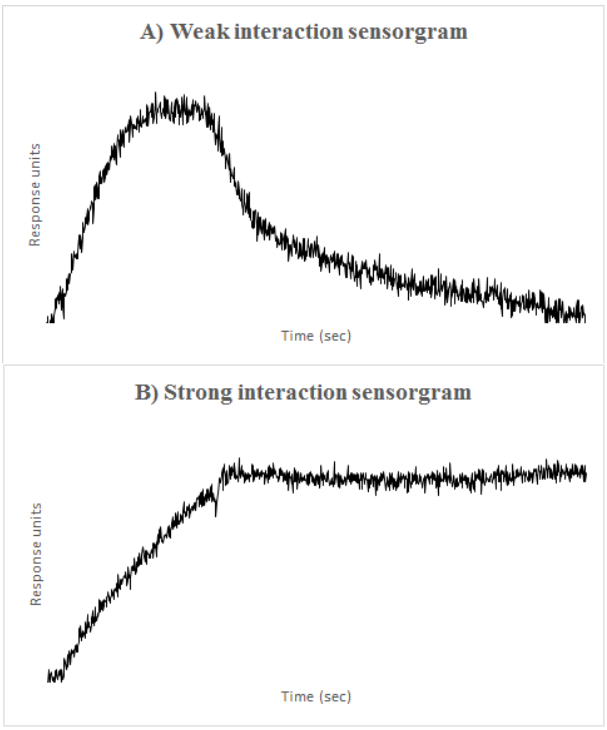


Figure 18. Sensorgrams reflecting a A) transient interaction and B) stable interaction.

With the exception of covalent protein interactions, all PPIs are in flux, coming together and apart in varying degrees of on- and off-rates. During the association step, PPIs are constantly being formed and broken under flow as query is continually transported across the surface. It is only during the "dissociation phase" in which only buffer is injected that the off-rate, or dissociation, of the PPIs can be determined. Weak or transient interactions are those in which the query protein is shuttled quickly away from the target under flow, thus resulting in a rapid drop back down to baseline as fewer queries occupy the target residues (Figure 19A). Query proteins in stable interactions, on the other hand, tenaciously bind to their targets, as evidenced by a slow return to baseline (Figure 19B).

Often, a fourth regeneration step will be used in which a high salt, acid, or base will be applied to the slide to disrupt any PPIs remaining following the dissociation step to return the sensorgram to baseline. This is especially common when working with surface-bound antibodies, which are generally stable under the regeneration conditions. Following regeneration, steps 1 – 4 can be repeated with a different query or query concentration. Regeneration has a couple of advantages: 1) removes molecules non-

specifically bound to the surface that may cause baseline issues and 2) allows the same chip to be re-used, which makes the experiment more cost effective than using one chip per query type or concentration. Regeneration is used when the immobilized target molecule, like antibodies and peptides, are stable enough to withstand the harsh regeneration treatment without detrimentally affecting their ability to bind other molecules. Thus, a regeneration step may not be appropriate when the activity and/or conformation of less stable target proteins are important to the experiment.

Because SPR measures the AMR for each ligand protein on the surface, it requires a separate light detector per ligand. This results in SPR being extremely low throughput with the most common platform used in academic settings (i.e., Biacore T100) being able to measure 4 different binding events at the same time through the use of 4 separate flow chambers and light detectors. Also, it is important that one of these analyses is used as a reference, so the interactions of interest are further decreased to three. Biacore also offers the Biacore A100, which can analyze as many as 20 PPIs at one time.

The desire for a high throughput SPR instrument led to the development of SPR imaging (SPRi). Although similar to SPR, SPRi has a few distinct differences (Wong & Olivo, 2014). One, SPR measures the AMR while SPRi measures the reflected light intensity across time at a set incident angle. Two, SPRi uses only one light detector (i.e., charge-coupled device “CCD” camera) for an entire chip. Three, since SPRi needs only one camera while SPR needs a separate detector per ligand type, SPRi can be in an array format, thus resulting in higher throughput. The MX96 instrument from IBIS

Technologies (Netherlands), for example, can analyze as many as 96 PPIs at one time (Krishnamoorthy et al., 2010). Neumann et al. demonstrated that as many as 9216 protein fragments (i.e., fragment library) could be immobilized on a single SPRi array for drug discovery and screening (Neumann, Junker, Schmidt, & Sekul, 2007). During SPRi data analyses, regions-of-interest (ROIs) – or, spotted target protein – on the captured video can be selected and the pixel intensities, or the intensity of the reflected light, of the ROIs analyzed across time. One ROI will produce a corresponding sensorgram, with time on the x-axis and the amount of reflected light on the y-axis. While SPRi has high throughput capability, its measurement of reflected light rather than the AMR results in lower sensitivity than SPR instruments. The use of a camera also results in optical vignetting, which can negatively affect the quality of data around the edges of the array.

Given the information provided above regarding SPR technology, the PPI that would result in the largest signal-to-noise ratio would be one where the 1) target protein was small (i.e., low mass), 2) target protein was very close to the surface, and 3) query protein was large. In actuality, most SPR instruments are not sensitive enough to detect binding of an analyte to queries that are less than 1 kDa.

4.2 History of NAPPA-SPRi

SPRi platforms like those from IBIS Technologies (Netherlands), GWC Technologies (Madison, WI), and Plexera LLC (Woodinville, WA) have the capacity of analyzing any type of PPI in a high throughput manner, yet most SPRi studies rely on antibody- or peptide-based interactions (Joshi, Peczuh, Kumar, & Rusling, 2014;

Stojanovic, Schasfoort, & Terstappen, 2014; Zhu et al., 2014). Even proof-of-concept demonstrations of SPRi throughput have been based on numerous replicates of stable antibody-antigen interactions rather than more biologically-relevant, non-antibody PPIs (Geertz, Shore, & Maerkl, 2012). In actuality, only one study has examined the kinetics of non-antibody, full-length proteins in high throughput. In 2016, 96 unique proteins from *Yersinia pestis* were expressed and purified from *E. coli*, biotinylated, and printed onto an SPRi chip coated with avidin using the Biacore FLEXchip and then probed with the same 96 proteins (Keasey et al., 2016). (Of note, the Biacore FLEXchip, which was reported to analyze as many as 400 binding events at one time, has been discontinued.) With the purported high throughput capabilities of various SPRi platforms, why then aren't there more studies like those of Keasey et al.? Simply put, purifying proteins is costly and labor intensive (see also page 67). It is also easier from an experimental standpoint to use peptides or antibodies as captured targets on the SPR array since they are resilient to buffer changes and harsh regeneration conditions. Antibodies also bind to their antigens with high affinity with long half-lives. Therefore, Dr. Joshua LaBaer asked in 2006 whether his high throughput DNA-based NAPPA chemistry could be made compatible with SPRi.

Dr. Sanjeeva Srivastava and Dr. Manual Fuentes first worked on making NAPPA compatible with SPRi. Their platform relied on using a C-terminal GST e-coil fusion tag in which the e-coil would bind very strongly via a coiled-coil interaction with a k-coil ligand on the surface (K_D = ~60 pM^{-1}) (De Crescenzo, Litowski, Hodges, & O'Connor-McCourt, 2003). E-coil and k-coil are short (< 40 amino acids) amphiphilic alpha-helical

peptides that form heterodimers through interhelical hydrophobic (i.e., coiled-coil) interactions (Aronsson et al., 2015). The k-coil and e-coil sequences were LKVSAL**KEK**VSAL**KEK**VSAL**KEK**VSAL**KEK**VSAL**KE** and LEVSAL**EKE**VSAL**EKE**VSAL**EKE**VSAL**EKE**VSAL**EK**, respectively. Note that the amino acids, leucine (L) and valine (V), are hydrophobic while the amino acids in bold, lysine (K) and glutamine (E), are hydrophobic.

The use of a ~4 kDa k-coil peptide instead of a massive antibody (i.e., 150 kDa) as a capturing ligand will significantly reduce the surface mass and would, theoretically, make the SPR analysis more sensitive to binding events. Furthermore, the use of a k-coil appeared to eliminate the need for BSA, which has been found to be an essential reagent for plasmid cDNA immobilization with standard NAPPA. The GST tag was maintained in order to analyze protein display levels fluorescently with an anti-GST monoclonal antibody. Briefly, the printing master mix contained only k-coil and plasmid cDNA and the protein expressed with rabbit reticulocyte lysate. (At this time, no other mammalian cell-free expression system was available.) By 2009, antibody-antigen interactions were successfully detected using the e-coil//k-coil chemistry (i.e., version 1), but the data were never published.

Dr. Lusheng Song, a collaborator in China at the National Center for NanoScience and Technology, created the second version of NAPPA-SPRi in 2012. Like standard NAPPA, he used a C-terminal GST tag and an anti-GST polyclonal antibody to bind the translated fusion protein. The antibody was not part of the printing master mix, but was instead bound to the entire bare gold surface through electrostatic interactions; no

adhesion linking layer and immobilization matrix were used. He was able to demonstrate the interactions of query antibodies to their target proteins, as well as between TP53 and MDM2. His data were not published.

Biacore SPR instruments were, at this time, the most popular; however, they relied on Biacore-specific flow chamber chips that were not amenable to printing (i.e., NAPPA). Therefore, VGP CPD worked with Plexera® Bioscience to develop a SPRi instrument that would be compatible with NAPPA, which is the PlexArray® HT System.

I worked on the first (e-coil/k-coil) and second (GST-based) versions of NAPPA-SPRi when I first joined the laboratory (separate from the individuals mentioned above), but found concerning disadvantages with both. The e-coil/k-coil chemistry prompted further optimization. First, only the k-coil and plasmid cDNA were contained in the printing master mix. In other words, the k-coil was not covalently captured to the array surface, but rather bound nonspecifically to the positive-charged slide through non-covalent, van der Waals forces. (The slides used the same amine-coating that I later used in the optimized NAPP-SPRi experiments.) Rather than adopting a vertical configuration where one end of the k-coil is captured to the surface, the k-coil molecules likely laid horizontally on the slide. It is therefore uncertain 1) whether this affected the k-coil's ability to interact specifically with the e-coil of the target protein, and 2) how much of the target protein is captured to the slide specifically via the k-coil/e-coil interaction or through van der Waals forces. To determine whether the target protein is captured specifically to the surface through the e-coil/k-coil interaction, I printed both GST-tagged proteins with and without the e-coil tag on an array and then analyzed the level of

displayed protein using a mouse anti-GST monoclonal antibody and fluorophore-conjugated secondary anti-mouse antibody. The data revealed no significant difference in the level of protein captured between GST-tagged proteins and GST-ecoil-tagged proteins (data not shown, September 2010).

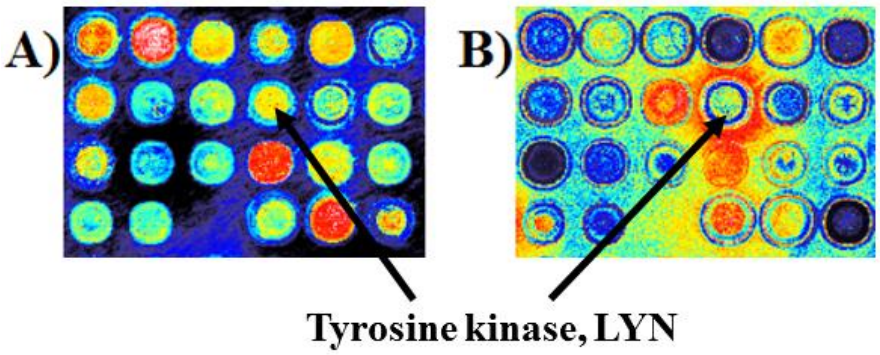


Figure 19. Proteins are denatured when they bind non-specifically to the slide. Fluorescent analyses of A) protein display using an anti-tag antibody and B) tyrosine phosphorylation using an nti-phosphotyrosine antibody. False-colored rainbow images, where blue = low antibody binding, yellow = moderate antibody binding, red = high antibody binding.

Non-specific protein adsorption and their denaturation on positively-charged amine surfaces is a known phenomenon (Karlsson et al., 2005). I have also observed non-specific target immobilization on NAPPA arrays. In fact, some antibodies bind more strongly to the area *around* the spot that has no capturing reagent (i.e., "the ring") than the spot itself (Figure 20), which may be due to the antibodies having a preference for linear epitopes or that the antibodies' epitopes are more exposed upon partial or full target denaturation. Therefore, it's important to point

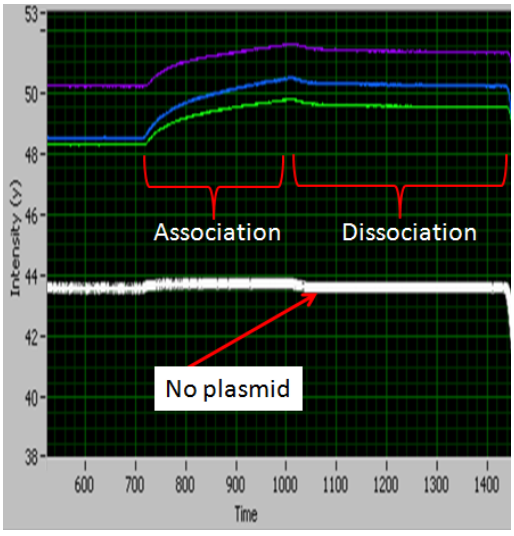


Figure 20. SPRi binding curves using a streptavidin and biotinylated k-coil coated slide surface. GST-tagged TP53, Fos, and Jun proteins probed with an anti-GST antibody.

out that while the e-coil/k-coil chemistry detects interactions between target proteins and query antibodies, there are no data supporting its ability to detect non-antibody PPIs that may rely on conformational epitopes (Figure 21). Of note, the query antibodies that were used (i.e., anti-JUN, anti-FOS, anti-TP53) are specific to linear epitopes, since synthetic peptides were used to immunize the animal hosts.

Another major issue with the k-coil/e-coil chemistry was that the amount of required plasmid cDNA was *five* times that needed for standard NAPPA. For example, standard NAPPA needs a 1.2 mg/mL concentration of plasmid cDNA while the NAPPA-SPRi printing mixture required 6 mg/mL. Indeed, it's been demonstrated that amphiphilic α-helical peptides bind to plasmid cDNA (Figure 22A)(Niidome et al., 1997). A viscous substance formed immediately upon the mixture of k-coil peptide and plasmid cDNA, which was easily seen by eye (Figure 22B). The viscous substance captured approximately 20 – 40% of the DNA, making it unavailable for printing (as determined through spectrophotometric analyses). The strong interaction between the k-coil and plasmid cDNA is likely why this approach did not require BSA to capture DNA to the slide surface.

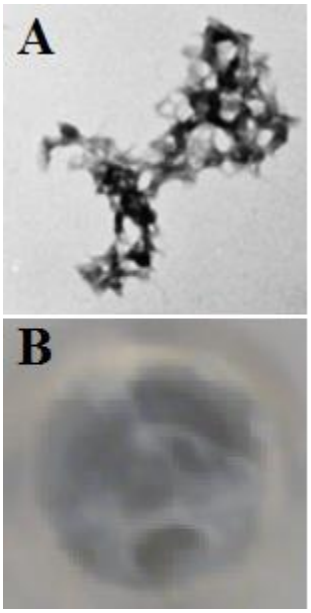


Figure 21. Amphiphilic α-helical peptide forms an insoluble aggregate with plasmid cDNA. Observed by A) Niidome et al. with electron microscopy and B) me in 2010 by eye. Reprinted with permission.

Version 2 of NAPPA-SPRi using a GST antibody to capture the GST-tagged proteins to the slide surface was also less-than-ideal. First, the GST-antibody interaction was not covalent; thus, under flow, the captured target protein theoretically should be

removed from the surface. Further, the use of a large 150 kDa anti-GST antibody would result in decreased sensitivity compared to a smaller capturing agent. In both NAPPA-SPRi versions, the target proteins were not captured covalently to the slide and only represented the target with the tag at the C-terminus. The tag may block binding epitopes and, therefore, the additional representation of the target with a tag at the N-terminus would be desirable. I decided that the disadvantages of the previous versions warranted a new approach.

In 2005, Promega Corporation (Madison, WI) reported for the first time a covalent tagging system using a modified haloalkane dehalogenase "HaloTag" enzyme that covalently binds to a small ~ 400 Da ligand (Los et al., 2005). In the unmodified version, the covalent ester interaction between the terminal chloride and aspartate residue on the enzyme is hydrolyzed by a nearby histidine at position 272 (England et al., 2015). Promega mutated the histidine to a phenylalanine, thus causing the HaloTag-ligand interaction to remain intact. The chloroalkane ligand can be altered to make the HaloTag system compatible with *in vitro* and *in vivo* platforms and numerous applications. (Please note that the term "ligand" can also refer to the protein captured to the sensor surface. To avoid confusion, the use of "ligand" in this thesis will only refer to the chloroalkane ligand that binds covalently to HaloTag.)

In parallel with my colleague, Dr. Jie Wang, who was testing HaloTag with the standard NAPPA platform, I began working with HaloTag in the context of NAPPA-SPRi in 2012. His interest in HaloTag was based on its ability to covalently capture proteins to the slide so that he could subsequently denature the proteins and probe the

array with patient serum to identify antibody biomarkers of disease (J. Wang et al., 2013). Although proteins are folded *in vivo*, they may become misfolded during disease (or occluded by the tag) and expose epitopes that would otherwise remain hidden. Dr. Wang was also interested in the idea of making the target proteins tagged at either the N- or C-terminus. In addition to these attributes, the HaloTag offered another advantage regarding SPR analyses: its small capturing ligand.

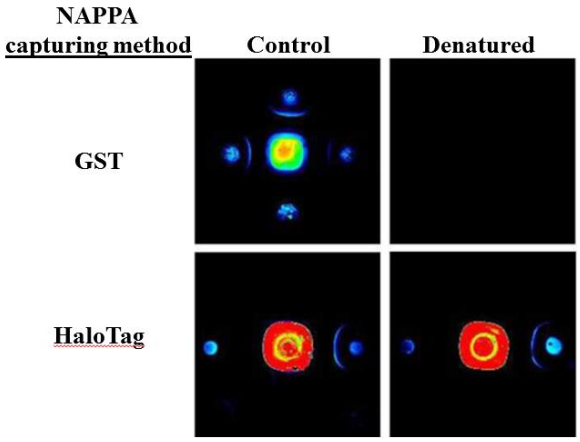


Figure 22. HaloTag NAPPA can withstand harsh denaturing conditions. Captured TP53 target protein was assessed with an anti-TP53 antibody. Both images are false-colored rainbow images, where blue = low antibody binding, yellow = moderate antibody binding, red = high antibody binding. Reprinted with permission.

The covalent and specific capture of HaloTagged target proteins to the slide surface via the HaloTag-ligand interaction was elegantly demonstrated by Dr. Wang (J. Wang et al., 2013). The target protein, TP53, was immobilized to the array using both GST- and HaloTag-based chemistries. The slides were then subjected to harsh denaturing conditions, which included incubating the array with 125 mM Tris-HCl, 2% SDS, 100 mM β-mercaptoethanol at 37 ^{o}C for 30 min with mild agitation. A comparison of captured TP53 target protein before and after denaturation revealed that TP53 continued to be captured only on the HaloTag-based NAPPA slide surface (Figure 23).

Although the VGP CPD had long conceptualized synthesizing proteins in situ for SPR analyses, they were not the first to publicly demonstrate it. In July 2012, Seefeld et al. of University of California published the article, “On-Chip Synthesis of Protein

Microarrays from DNA Microarrays Via Coupled In Vitro Transcription and Translation for Surface Plasmon Resonance Imaging Biosensor Applications," in the Journal of American Chemical Society (Seefeld, Halpern, & Corn, 2012). In it, they bound linear dsDNA to one spot on a microfluidic SPR chip while a neighboring feature had a Cu(II)-nitriloacetic acid (NTA) monolayer. His-tagged proteins were expressed in IVTT, diffused to the neighboring spot, and bound to the Cu(II)-NTA. A total of 16 features were on the array, with four features per protein. They demonstrated their platform by expressing and capturing green fluorescent protein and luciferase, then probing them with their specific antibodies.

4.3 Optimization of NAPPA-SPRi using design of experiments (DOE)

No binding signal can be obtained on the SPRi using the NAPPA chemistry used with the standard fluorescent-based protein microarrays. This incompatibility is because the "noise" is too high compared to the "signal;" in other words, the change in mass upon query binding is too little compared to the high mass of the printing master mix and target protein within the spot. The appropriate amounts of printing reagents for SPRi analyses were determined through a series of design of experiments (DOEs).

Scientists generally use the "one-factor-at-a-time" (OFAAT) approach to optimize their experiments, which means that they optimize one factor before optimizing the setting for the next factor and so on. However, the experimental landscape is like a 3D surface. Locking all parameters except one will force the scientist to only investigate a portion of that landscape, which may never include the optimal region. Additional

disadvantages of OFAAT include time, expense, and the inability to determine whether the effect of one factor will depend on the level chosen for a different factor (i.e., "factor interaction"). DOE, on the other hand, analyzes different combinations of factors across the experimental plane (Montgomery, 2008). This allows the optimum to be found and factor interactions to be identified. Data interpretation is also made easy through the use of DOE software. Although scientists, particularly those in the research arena, rarely use DOE, engineers use the DOE approach all of the time. The engineers at Honeywell International Inc. (Morris Plains, NJ) used DOE, for example, to design a Boeing airplane engine (personal communication with a Honeywell employee, Dr. Don Holcomb).

Optimization via DOE is an iterative process, requiring a cascade of experiments that are designed based on the results from the previous experiment. The questions when designing a DOE experiment are: Which and how many factors should be tested? How will the best parameters be determined quantitatively? How many experiments can be reasonably done at one time? How will the reproducibility of the experiment be determined? How many levels of each factor should be tested? And, finally, what levels of the factors should be tested?

Herein, I describe one of the last DOE experiments that I performed to optimize the NAPPA-SPRi printing chemistry. The amount of the four reagents (or factors) in the printing master mix was tested, which included the plasmid cDNA, BSA, bis(sulfosuccinimidyl) suberate amine-to-amine crosslinker (BS3), and HaloTag amine (O4) ligand. I kept the amount of DMSO constant at 2.5%. The maximum binding response of the query, an anti-TP53 D01 monoclonal antibody, to the target TP53 (with a

C-terminal HaloTag) provided a quantitative value that would determine the optimal reagent levels. Within the SPRi analyses window of 1.2 cm x 1.2 cm, 441 target proteins could be analyzed at one time (9 grids of 7 x 7).

Table 9. DOE factors and levels

Reagent	Low level -1	Center point 0	High level 1
DNA	12 µg	18 µg	24 µg
BSA	0.04 µg	7.31 µg	14.65 µg
HaloTag ligand	0.75 µg	1.25 µg	1.75 µg
BS^3 linker	5 µg	8.76 µg	12.52 µg

The reproducibility could have been done using replicate slides, replicate target spots on the array, or through the use of replicate "center points." In a standard DOE experiment, there are two levels to each factor, representing a low and high value called a "corner point." The center point is usually the average value between these two extremes. I employed a two-level factorial experiment with four center *and* four corner point replicates. Although the reproducibility could have been achieved with either type of replicates (i.e., center or corner point replicates), I had space on my sensor surface to perform both. And since the experimental landscape may not be linear across the range of values tested, having center points may be advantageous since they are able to provide more detail about the experimental plane. With 2 factor levels and four factors, there were 17 different printing mix combinations (i.e., center points + 2 levels $^{4\ factors}$). Taking into consideration the replicates, the total number of target spots was 68, or 4 replicates x (2 levels $^{4\ factors}$) + 4 center points. The levels of the corner and center points for the four factors are depicted in Table 9.

Master Mix #	DNA A	BSA B	Ligand C	BS3 D
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	-1	-1	-1	1
4	1	-1	-1	1
5	-1	-1	1	-1
6	1	-1	1	-1
7	-1	-1	1	1
8	1	-1	1	1
9	-1	1	-1	-1
10	1	1	-1	-1
11	-1	1	-1	1
12	1	1	-1	1
13	-1	1	1	-1
14	1	1	1	-1
15	-1	1	1	1
16	1	1	1	1
17	0	0	0	0
18	0	1	0	0

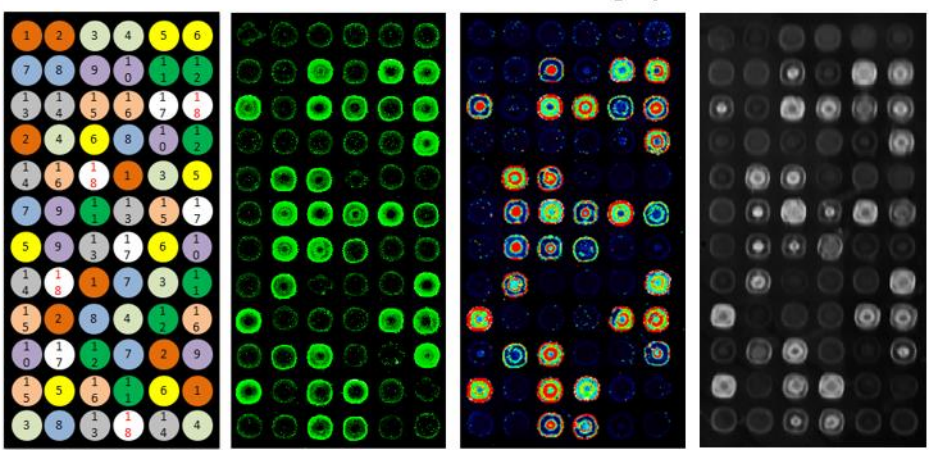


Figure 23. Print lay-out based on printing mix combinations. DNA deposition was determined via fluorescence using PicoGreen staining; green-scale false-colored image. Protein display determined via fluorescence using an anti-HaloTag antibody; rainbow false-colored image. Original SPRi image was altered to have circular spots like the other images above; spots on the SPRi usually have an oval-shape.

The lay-out of the 17 different printing mix combinations on the slide is shown in Figure 24, where the combinations are given arbitrary numbers 1 – 17. Note that the mixes were deposited in a somewhat random manner in order to minimize any possible effect of location on response. Table 9 summarizes the factor levels for each printing mix, where "+1" represents the high factor level, "-1" represents the low factor level, and "0" represents a center point.

The random print lay-out for the TP53 target protein according to the printing mix combination is shown in Figure 24. The DNA deposition, protein display, and refractive index differences (as observed by SPRi) are also shown in Figure 24. The SPRi "snapshot" image depicts the reflected light, where the brightness is proportional to the amount of mass on the surface. Figure 25 and Figure 26 show the plasmid cDNA and captured HaloTagged-TP53 target protein levels on the slide, respectively, determined via fluorescence and an anti-HaloTag antibody. Fluorescent analyses reveal that more DNA

deposition results in more protein display and brighter SPR spots, with printing mix combination #16 having highest DNA deposition and TP53 capture. SPRi analyses were still necessary to perform since fluorescent analyses may not accurately reflect binding response on the SPRi instrument.

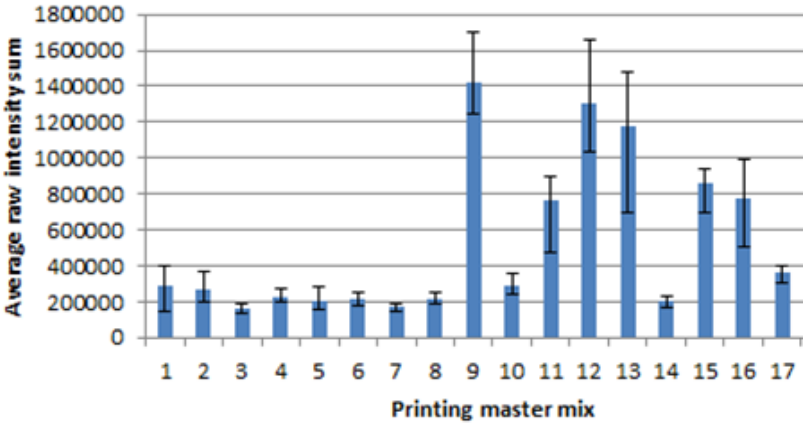


Figure 24. Deposition of TP53 plasmid cDNA on a gold SPRi slide

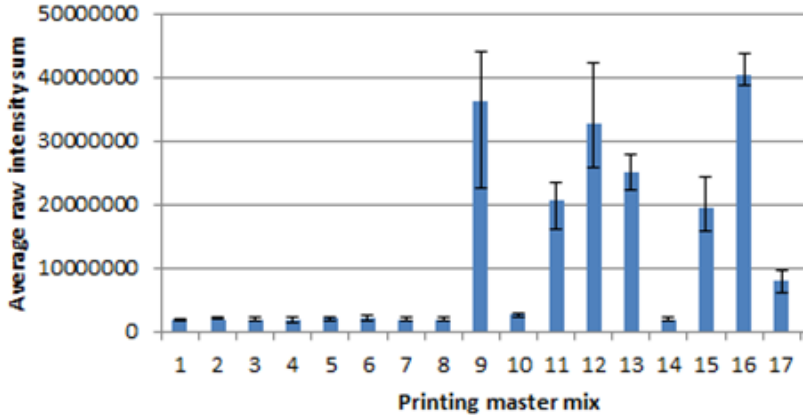


Figure 25. Protein display of HaloTagged TP53 target protein on a gold SPRi slide as determined by an anti-HaloTag polyclonal antibody

In addition to TP53, negative controls that will not bind to the anti-TP53 antibody were also included on the array, which included firefly luciferase and calcineurin life EF-hand protein 1 (CHP-1). These negative controls were laid out exactly like TP53, but in supergrids on either side of the TP53 supergrid. Unless otherwise noted, the TP53 binding sensorgrams for the following results were referenced to the binding sensorgrams of CHP-1. (Similar results were obtained without referencing as well as with referencing to firefly luciferase instead of CHP-1.)

StdOrder	RunOrder	CenterPt	Blocks	Master Mix #	A	B	C	D	Response
1	1	1	1	1	-1	-1	-1	-1	119.53
2	26	1	1	1	-1	-1	-1	-1	113.79
3	42	1	1	1	-1	-1	-1	-1	58.04
4	63	1	1	1	-1	-1	-1	-1	111.03
5	2	1	1	2	1	-1	-1	-1	133.15
6	18	1	1	2	1	-1	-1	-1	98.55
7	47	1	1	2	1	-1	-1	-1	2.05
8	56	1	1	2	1	-1	-1	-1	70.09
9	3	1	1	3	-1	-1	-1	1	154.69
10	27	1	1	3	-1	-1	-1	1	135.99
11	44	1	1	3	-1	-1	-1	1	104.55
12	64	1	1	3	-1	-1	-1	1	46.27
13	4	1	1	4	1	-1	-1	1	100.52
14	19	1	1	4	1	-1	-1	1	120.02
15	49	1	1	4	1	-1	-1	1	59.58
16	68	1	1	4	1	-1	-1	1	90.7
17	5	1	1	5	-1	-1	1	-1	100.92
18	28	1	1	5	-1	-1	1	-1	70.09
19	35	1	1	5	-1	-1	1	-1	63.8
20	59	1	1	5	-1	-1	1	-1	49.73
21	6	1	1	6	1	-1	1	-1	71.53
22	20	1	1	6	1	-1	1	-1	78.79
23	39	1	1	6	1	-1	1	-1	42.11
24	62	1	1	6	1	-1	1	-1	137.8
25	7	1	1	7	-1	-1	1	1	124.45
26	29	1	1	7	-1	-1	1	1	121.89
27	43	1	1	7	-1	-1	1	1	101.23
28	55	1	1	7	-1	-1	1	1	125.56
29	8	1	1	8	1	-1	1	1	101.53
30	21	1	1	8	1	-1	1	1	91.31
31	48	1	1	8	1	-1	1	1	11.81
32	65	1	1	8	1	-1	1	1	17.19
33	9	1	1	9	-1	1	-1	-1	257.24
34	30	1	1	9	-1	1	-1	-1	187.69
35	36	1	1	9	-1	1	-1	-1	180.25
36	57	1	1	9	-1	1	-1	-1	262.74
37	10	1	1	10	1	1	-1	-1	68.77
38	22	1	1	10	1	1	-1	-1	23.81
39	40	1	1	10	1	1	-1	-1	33.59
40	52	1	1	10	1	1	-1	-1	-9.5
41	11	1	1	11	-1	1	-1	1	264.43
42	31	1	1	11	-1	1	-1	1	216.64
43	45	1	1	11	-1	1	-1	1	268.65
44	61	1	1	11	-1	1	-1	1	238.63
45	12	1	1	12	1	1	-1	1	80.42
46	23	1	1	12	1	1	-1	1	46.62
47	50	1	1	12	1	1	-1	1	49.34
48	54	1	1	12	1	1	-1	1	44.85
49	13	1	1	13	-1	1	1	-1	211.39
50	32	1	1	13	-1	1	1	-1	215.77
51	37	1	1	13	-1	1	1	-1	204.33
52	66	1	1	13	-1	1	1	-1	201.73
53	14	1	1	14	1	1	1	-1	31.35
54	24	1	1	14	1	1	1	-1	4.34
55	41	1	1	14	1	1	1	-1	-17
56	67	1	1	14	1	1	1	-1	-6.01
57	15	1	1	15	-1	1	1	1	261.48
58	33	1	1	15	-1	1	1	1	240.44
59	46	1	1	15	-1	1	1	1	131.96
60	58	1	1	15	-1	1	1	1	140.82
61	16	1	1	16	1	1	1	1	65.09
62	25	1	1	16	1	1	1	1	79.23
63	51	1	1	16	1	1	1	1	69.7
64	60	1	1	16	1	1	1	1	39.25
65	17	0	1	17	0	0	0	0	72.2
66	34	0	1	17	0	0	0	0	60.16
67	38	0	1	17	0	0	0	0	57.75
68	53	0	1	17	0	0	0	0	21.74

Figure 26. Printing master mix combinations and their associated response on the SPRi sensorgram. A = DNA, B = BSA, C = Ligand, D = BS3. StdOrder = Standard order, organized by master mix combination. RunOrder was determined on the spot's location on the array reading from left-to-right, top-to-bottom. TP53 response was referenced to CHP-1.

The maximum response of the binding sensorgrams was chosen for each printing mix combination, as denoted in Figure 27. The point at which the response value was chosen was at the end of the association phase where the number of anti-TP53 antibody molecules bound to TP53 target proteins was at a maximum for the entire NAPPA-SPRi analyses. The data were analyzed using the Minitab® 17 software, which is software that was specifically created to design and analyze DOE data.

The normal plot is a visual way of seeing which factors or factor interactions are important. In general, the responses are first ranked from smallest to largest. Then, the

cumulative frequency of those responses is spread along the y-axis. A line is usually drawn between the 25th and 75th percentile points, which allows the user to identify which combinations may be significant. Negligible effects are normally distributed along a line where the mean is 50%. The farther away from the line, the more significant the factor. DNA and BSA appeared to have the most significant impact on response, as well as the combination of AB and ABCD (Figure 28). In other words, the interaction of factors A (DNA) and B (BSA) and between all of the factors significantly affected the SPR response. The factor with the largest impact on response is the amount of DNA.

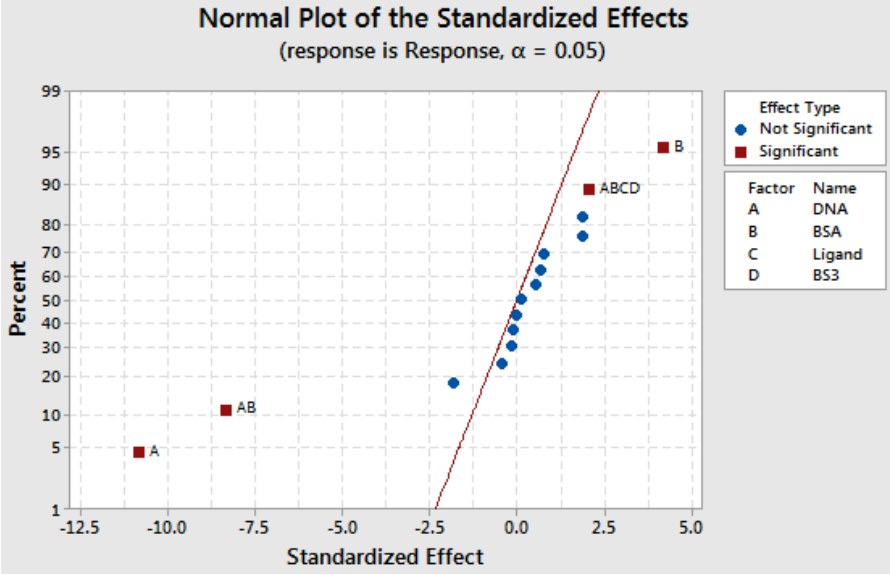


Figure 27. Normal plot of standardized effects shows that DNA, BSA, and the interaction between the two reagents are significant.

The residual plots of the referenced data looked acceptable, with expected randomness (Figure 29A). Interestingly, non-referenced TP53 response data did have reveal a nonconstant variance in observation order, showing a downward trend from the top to bottom of the slide (Figure 29B). A possible reason for this is due to the SPR lay-out where the query protein passes over the slide from top to bottom. As the antibody binds to targets with low dissociation, less and less query is available for binding across

the array. A variance-stabilizing transformation of the data to account for this phenomenon was applied, which eliminated the nonconstant variance (Figure 29C). The use of a reference also eliminated this phenomenon.

This DOE experiment supported previous observations by both others and myself in the lab that DNA and BSA interact with each other. In other words, the effect of DNA depends on the amount chosen for BSA. The interaction between these factors is depicted in Figure 30, where interactions are represented as intersecting lines; the more perpendicular, the more interaction between the factors.

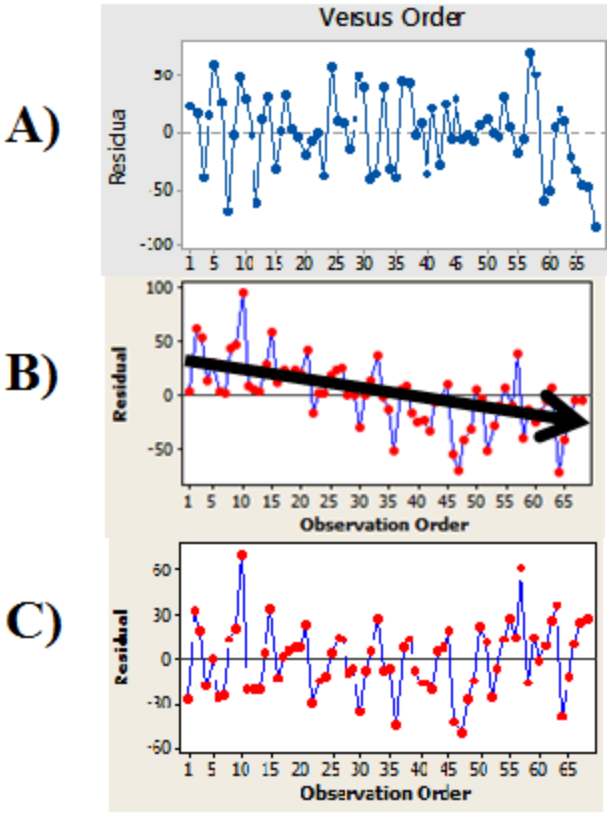


Figure 28. Residual plots of SPR response for the anti-TP53 antibody - TP53 protein interaction. Residual versus observation order for A) referenced data, B) non-referenced data, and C) non-referenced data that has undergone nonconstant variance transformation.

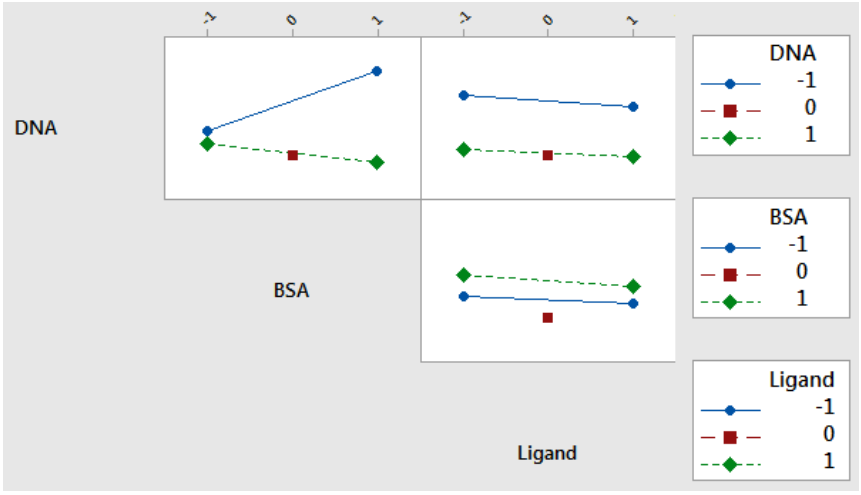


Figure 29. Interaction plot for response using average responses across replicates.

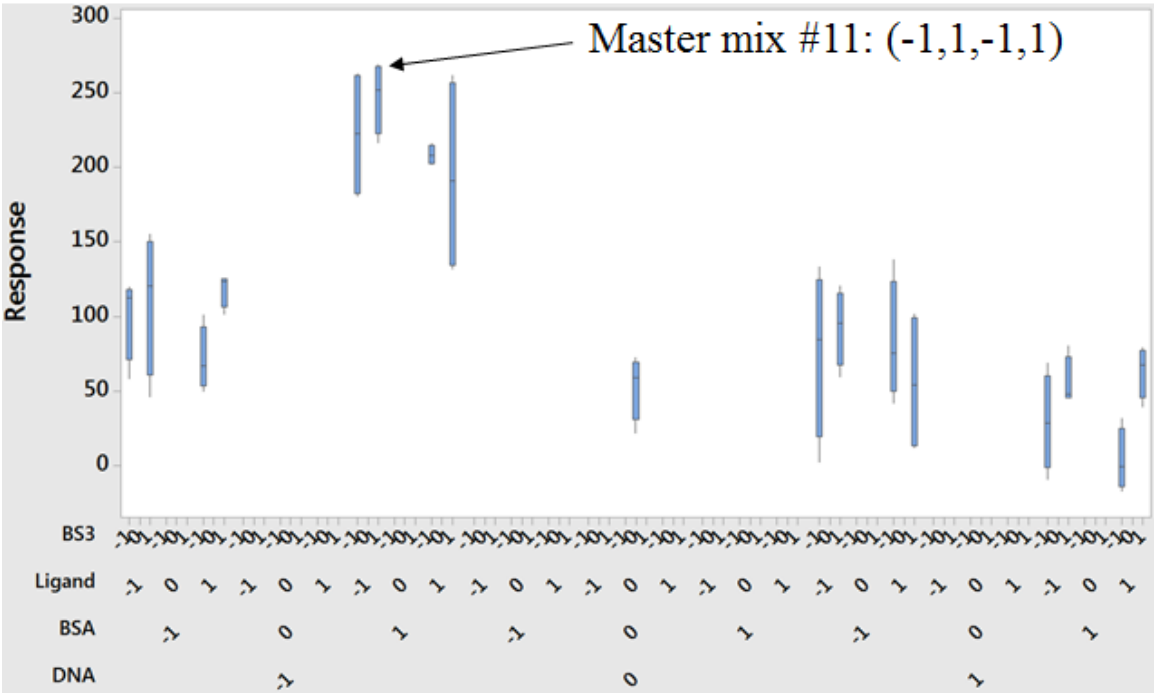


Figure 30. Boxplot of response.

A boxplot of response shows the reproducibility of each printing mix type across the four replicates. It is clear that low DNA and high BSA give the highest response. The amount of ligand didn't appear to make much difference, at least for the levels tested. The variable levels that will produce the maximum response and highest reproducibility are indicated (Figure 31).

The experiment underscores the importance of analyzing the data in an appropriate manner for the intended application. While the fluorescent analyses favored the printing mix combination #16, the printing mix that had the highest binding response (with the highest reproducibility) on the SPRi was #11.

The DOEs, through this example, decreased the amount of standard HaloTag-based NAPPA printing mix from 183 µg to 40 µg in 30 µL for optimal NAPPA-SPRi

sensitivity, representing an 78% reduction in mass (Table 10). It also altered the amount of each reagent in comparison to each other (Figure 32).

Table 10. Comparison of reagent mass for standard HaloTag-NAPPA and HaloTag-NAPPA-SPRi

Reagent	Standard NAPPA*	NAPPA-SPRi after DOE
DNA	30 µg	12 µg
BSA	110 µg	14.65 µg
HaloTag ligand	5.2 µg	0.75 µg
BS^3 linker	37.5 µg	12.52 µg
TOTAL	**~180 µg**	**~40 µg**

* In 30 µL for printing

To summarize, the DOE approach is time- and cost-effective, identifies significant factors, and determines whether factors interact with each other. In a biological setting, factor interaction may also help to elucidate function and mechanisms. The analyses and plots presented here represent only a fraction of those that are possible with DOE software.

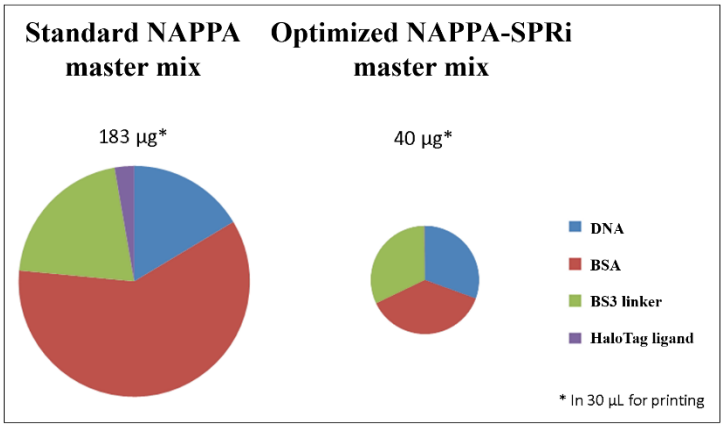


Figure 31. Pie chart comparing the reagent mass and ratio differences between standard NAPPA and NAPPA-SPRi using HaloTag-BSA chemistry.

4.4 Other parameters and methods tested

While Chapters 4.2 – 4.3 cover the history and development of NAPPA-SPRi, they only provide a small glimpse into the actual work that was required to arrive at the optimized surface chemistry. In this section, the different surface chemistries that were tested and why they weren't ideal will be covered. The investigation of some of the other parameters to improve signal response and reproducibility will also be discussed.

4.4.1 Cell-free expression systems for expressing target proteins

Until 2012, NAPPA employed rabbit reticulocyte lysate from Promega Corporation (Madison, WI) to express the target proteins because it was the only mammalian-based *in vitro* transcription translation system that was available. In December 2011, Thermo Fisher Scientific (Waltham, MA) launched a new mammalian expression system using lysate from human HeLa cells. The VGP CPD quickly adopted the new system once it was clear that the human target proteins were expressed at a much higher efficiency with HeLa cells instead of rabbit reticulocyte. As Figure 33 demonstrates, 100% of the rabbit-based reagents recommended by Promega for one reaction resulted in slightly higher expression levels than 20% of the human-based reagents recommended for one reaction by Thermo Fisher Scientific; the cost for both systems are similar. At 100% HeLa lysate, the target proteins were so highly expressed that they began to diffuse into other spots! Another advantage of using the human expression system was that it was able to express some target proteins that did not express (or express well) in the rabbit expression system. Ultimately, both regular NAPPA users and I determined that the "sweet spot" for NAPPA slides was 60% of HeLa lysate using the standard Hybriwell^{TM} seals from Grace Bio-Labs (Bend, OR).

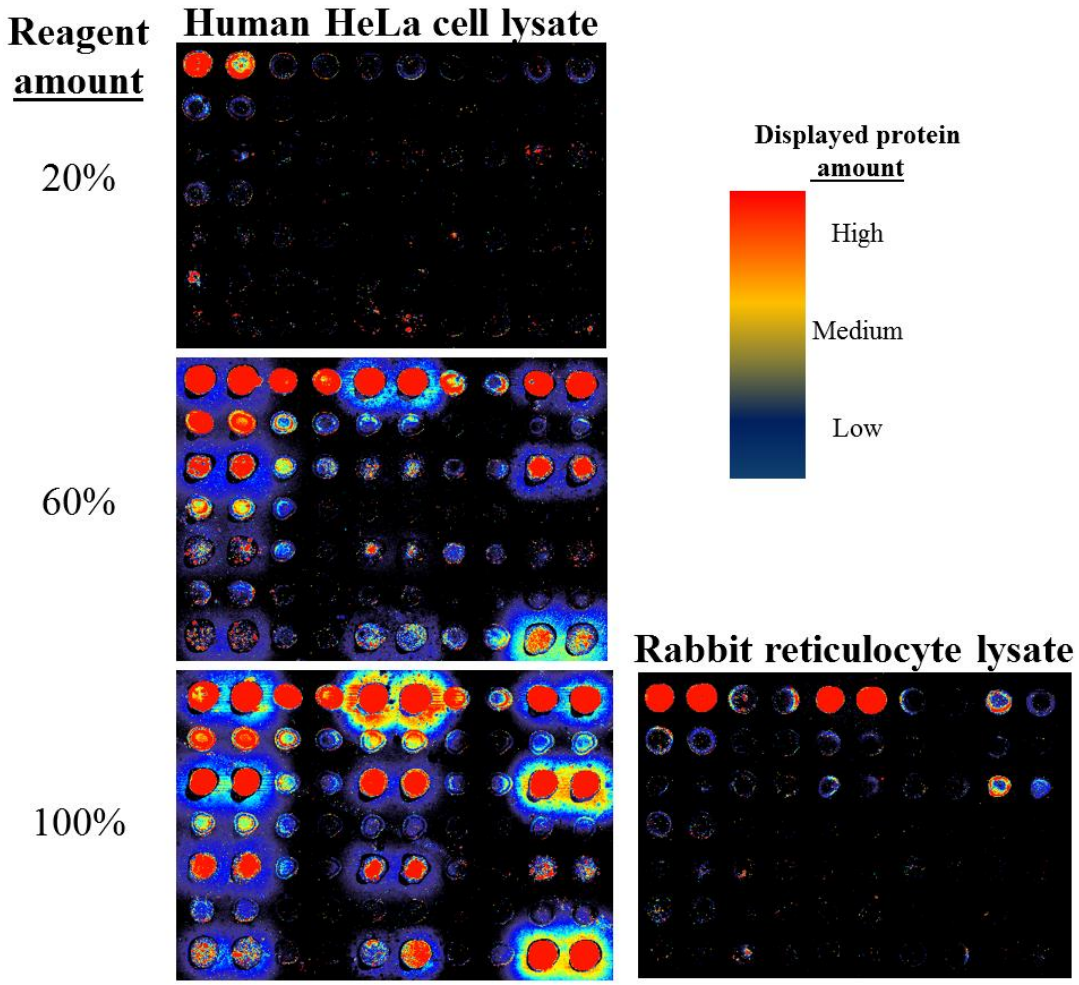


Figure 32. A comparison of the expression efficiency between rabbit reticulocyte lysate and human HeLa cell lysate. False-colored rainbow images.

4.4.2 Chamber for protein target expression

Target protein expression on standard NAPPA slides are always performed with Hybriwell^{TM} seals from Grace Bio-Labs (Bend, OR), which has a non-adhesive internal compartment that covers the entire NAPPA slide and has a ~ 150 µL capacity. These were produced specifically for NAPPA, with a chamber depth of 0.25 mm rather than the standard 0.15 mm. After expression is complete, the seal is removed and the slide is

washed to remove any non-specifically bound protein. The slide is kept wet throughout the experiment (i.e., probing with serum or antibodies) until it is dried with compressed air. The slide is finally analyzed in a microarray scanner.

I also employed the Hybriwell™ seals when I began working with the SPRi system, which had been done with the previous versions of NAPPA-SPRi. However, they presented two issues that needed to be addressed. First, the detectable area for SPRi is much smaller than that used for NAPPA (i.e., 12 mm x 12 mm for SPRi versus 21 mm x 60 mm for NAPPA); in other words, the use of the Hybriwell required much more expression lysate (and money) than was necessary. Second, the SPRi flow chamber had to be adhered to the slide prior to SPRi analyses, which would require the slide (and target proteins) to be dried. Slide drying following the removal of the Hybriwell™ seals, adherence of the SPRi flow chamber, and subsequent reconstitution of the target proteins in buffer could be performed in < 2 minutes, but was definitely far from ideal.

An initial attempt to replace the Hybriwell with the SPRi flow chamber was made to resolve both issues when rabbit reticulocyte lysate was still being employed as the cell-free expression system. The Plexera SPRi flow chamber is made of hard plastic with an internal volume of 30 µL and 0.15 mm depth and designed to distribute the query protein in a uniform manner across the sensor surface. However, the decrease in rabbit reticulocyte lysate (i.e., from 150 µL to 30 µL) resulted in a significant loss of signal, thus requiring the use of the larger volume Hybriwell (data not shown). A Hybriwell chamber with smaller surface coverage was also tested with an internal depth of 0.15

mm; this resulted in lower expression and reproducibility across the array (data not shown).

The increased expression efficiency of the human HeLa cells compared to the rabbit reticulocyte lysate invited a re-examination into employing the SPRi flow chamber for expression. Three slides were prepared for this test: 1) Hybriwell with 60% HeLa lysate, 2) Flow chamber with 60% HeLa lysate, and 3) Flow chamber with 100% HeLa lysate. Target TP53 and firefly luciferase proteins were probed with an anti-TP53 monoclonal antibody, and the TP53 binding sensorgrams were referenced to luciferase. The interaction between TP53 and anti-TP53 antibody were detected in all three methods; however, the use of the flow chamber with 100% HeLa lysate resulted in the highest signal and reproducibility (Figure 34). Please note that the Hybriwell uses 150 µL whereas the volume capacity of the flow chamber is only 30 µL; thus, 60% HeLa lysate with the Hybriwell uses 90 µL of HeLa lysate while 100% HeLa lysate with the flow chamber uses 30 µL HeLa lysate. The flow chamber, at both HeLa lysate concentrations, resulted in very good reproducibility across replicates. These data are similar to those I obtained at a previous laboratory that used Cy3/Cy5 in-house cDNA microarrays; inflexible 1 mm thick (mSeries) Lifterslips™ resulted in more consistent signal across the array compared to the standard flexible Lifterslips™ that are ~0.2 mm thick.

Based on these results, the flow chamber was adhered to the NAPPA-SPRi slide *prior* to expression and throughout the additional steps. The switch from the Hybriwell to the flow chamber decreased the overall cost of the experiment because it required less reagent. In the case of expression, the cost was cut by ~5-fold. The target proteins were also able to stay in solution throughout the duration of the experiment.

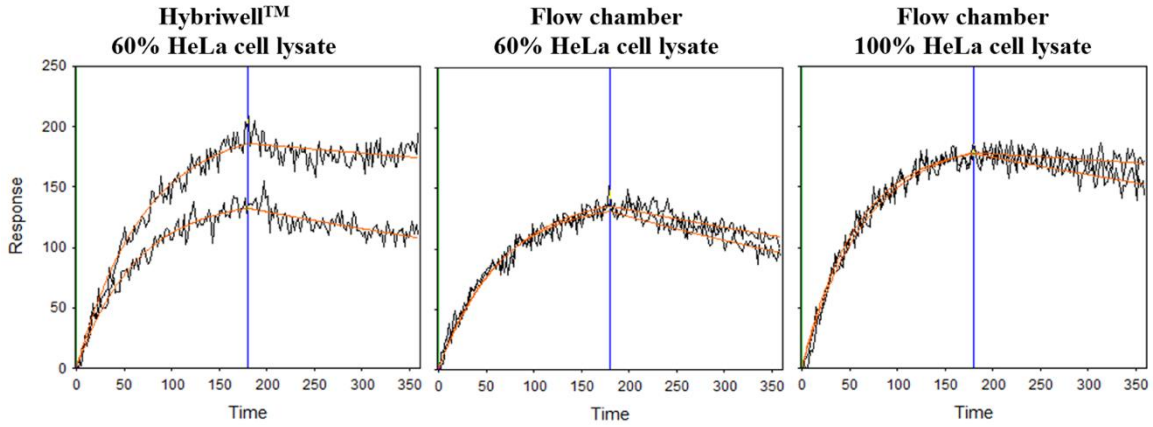


Figure 33. Binding response of TP53 target protein with an anti-TP53 antibody with different expression chambers, lysate, and lysate amount. Blue line indicates the end of the association phase and the beginning of the dissociation phase.

4.4.3 EDC-NHS surface chemistry

EDC-NHS refers to the covalent crosslinking between primary amines and carboxylic acids with the aid of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide (NHS). It is the most frequently employed crosslinking chemistry to couple target proteins to the immobilization matrix in SPR experiments. Briefly, the sensor surface is coated with an alkane-PEG self-assembled monolayer having terminal carboxyl and hydroxyl groups at a ~ 1:50 ratio. When the purified target proteins are ready to be attached to the slide, the carboxyl groups are made into an active ester with the addition of EDC; the coupling of primary

amines of target proteins to the active esters is made more efficient with NHS. The reactive NHS-ester has a half-life that is highly dependent on pH; at pH 7, the half-life is ~5 hours while at pH 8.6, the half-life is 10 minutes (Tiwari & Uzun, 2017). Regardless, the EDC-NHS reagents are used immediately upon preparation for best results (i.e., coupling).

Standard fluorescent-based NAPPA arrays do not use EDC-NHS to immobilize the capture reagent to the slide surface. Instead, the amine-to-amine bifunctional NHS crosslinker, BS3, is employed. The inactive, lyophilized form of the reagent is stored in the dark at -20 ^{o}C. Just prior to adding the BS3 to the printing mixture, the lyophilized BS3 is activated by reconstituting it in water or DMSO. Various incubation lengths and temperature have been tested; for NAPPA, an overnight incubation (~ 16 hours) at 4 ^{o}C in the dark with immediate printing the following morning is not only the most convenient approach (compared to a 1-hour incubation at room temperature), but also results in the highest plasmid cDNA deposition and protein display. In short, the BS3 crosslinks the 1) BSA lysines to each other, forming a meshwork that theoretically holds the plasmid cDNA in place, and 2) capturing reagent to the amine-coated slide surface.

The incorporation of surface chemistry similar to standard NAPPA was utilized in NAPPA-SPRi, where the slide surface was terminated with amines and the printing mixture contained BS3 crosslinker. However, I became concerned when I saw that components from the HeLa lysate expression system was binding strongly to the highly positive-charged surface as soon as it was injected over the surface. The nonspecific

interactions resulted in a wave-like image (Figure 35A) in the direction that the lysate was applied to the sample. Preliminary results with the EDC-NHS chemistry indicated that this phenomenon was significantly reduced (Figure 35B), so I investigated whether the surface chemistry should be based on EDC-NHS, rather than BS3, coupling.

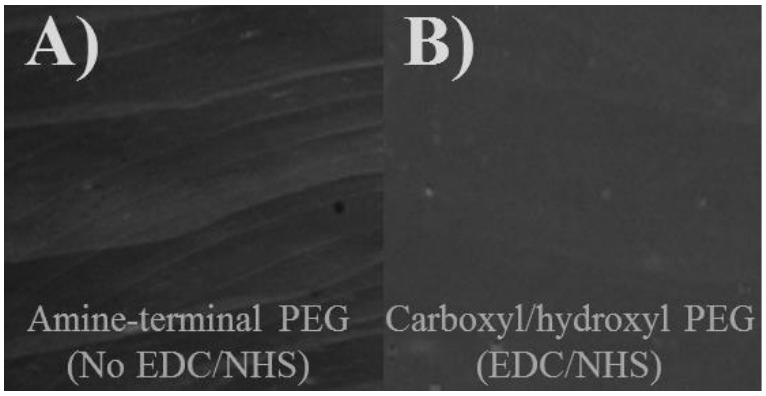


Figure 34. SPRi image of an A) amine-terminated and B) EDC-NHS chemistry sensor surface that has been incubated with the HeLa expression system. Brighter areas indicate higher mass than darker areas. Same SPRi incident angle used for both images.

In my first experiments, the sensor surface was coated with carboxyl/hydroxyl-terminated groups. The carboxyl groups were activated with EDC-NHS, and the slides were printed immediately after that with a printing mixture containing plasmid cDNA, HaloTag (04) amine ligand, BSA, and DMSO. Unfortunately, the short half-life of the reactive carboxyl group with slow print times resulted in poor protein display (data not shown).

A different approach was then tried, which was to activate the carboxyl groups and immediately coat the slide with HaloTag (04) amine ligand, followed by spotting a printing mixture without the HaloTag ligand. I compared the binding responses of TP53

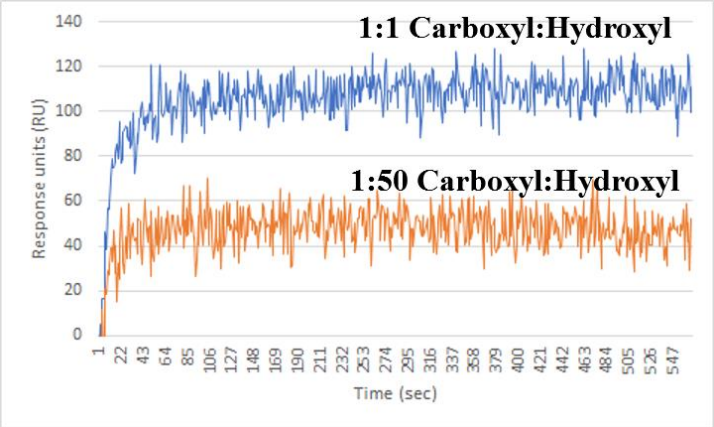


Figure 35. Binding responses of TP53 target protein with anti-TP53 antibody across different carboxyl:hydroxyl ratios. Sensorgrams were referenced to CARD11.

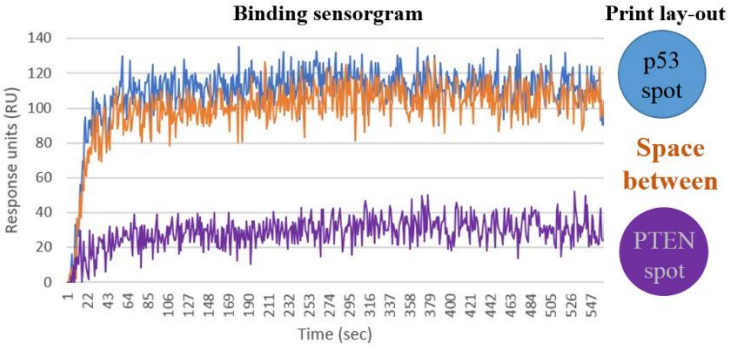


Figure 36. Binding sensorgram of an anti-TP53 antibody query interacting with an expressed, immobilized TP53 target protein using a 1:1 carboxyl:hydroxyl surface and HaloTag-BSA printing mixture. Since the HaloTag ligand coats the entire slide, diffused expressed TP53 protein can be captured specifically outside of printed spot. Referenced to CARD11.

target protein with the anti-TP53 antibody query across different carboxyl:hydroxyl ratios (i.e., 1:1, 1:2, 1:5, 1:10, 1:25, and 1:50). The responses were proportional to their carboxyl amount (Figure 36). The 1:1 ratio obtained the highest response, which was similar to that obtained with regular BS3-based chemistry. While the EDC-NHS chemistry worked for NAPPA-SPRi, the requirement to coat the slides with HaloTag ligand prior to printing resulted in (not too surprisingly) the diffused target proteins being specifically captured and displayed farther away from their printed spot than with standard BS3 chemistry. Figure 37 shows that the binding response of the target spot is the same as the response as the ring *around* the target spot. Moreover, slight binding was even observed where a non-p53 target protein, PTEN, was expressed and displayed, which would make the interpretation of PPIs on this surface impossible. The amine-terminated chemistry, on the other hand, did not have diffusion of TP53 into other

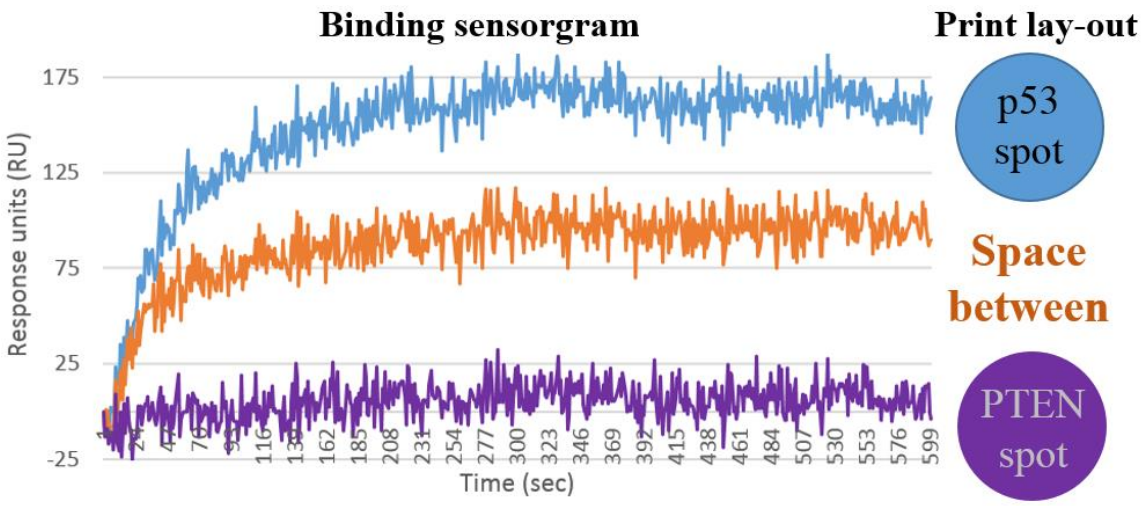


Figure 37. Binding sensorgram of an anti-TP53 antibody query interacting with an expressed, immobilized TP53 target protein using an amine-terminated surface and HaloTag-lysine printing mixture. Since the HaloTag ligand coats the entire slide, diffused expressed TP53 protein can be captured specifically outside of printed spot. Referenced to CARD11.

spots, although some non-specific capture of TP53 protein *around* the spot was observed (Figure 38); this phenomenon has also been observed with standard NAPPA and is referred to as "the ring effect." Due to the extreme diffusion with the EDC-NHS chemistry, the use of EDC-NHS as a coupling mechanism for NAPPA-SPRi was ultimately discarded.

4.4.4 Baseline instability

Baseline drift during an SPR analyses can be caused by numerous sources, including the instrument itself. Those arising from the instrument should be accounted for by the manufacturer, so baseline drift is often the result of experimental conditions, including small fluctuations in temperature, changes in flow rate, matrix effects, loss of target protein over time, and improperly matched buffers (across injections). Baseline drifts that are occur with dextran-based sensor chips can be due to changes in the depth or

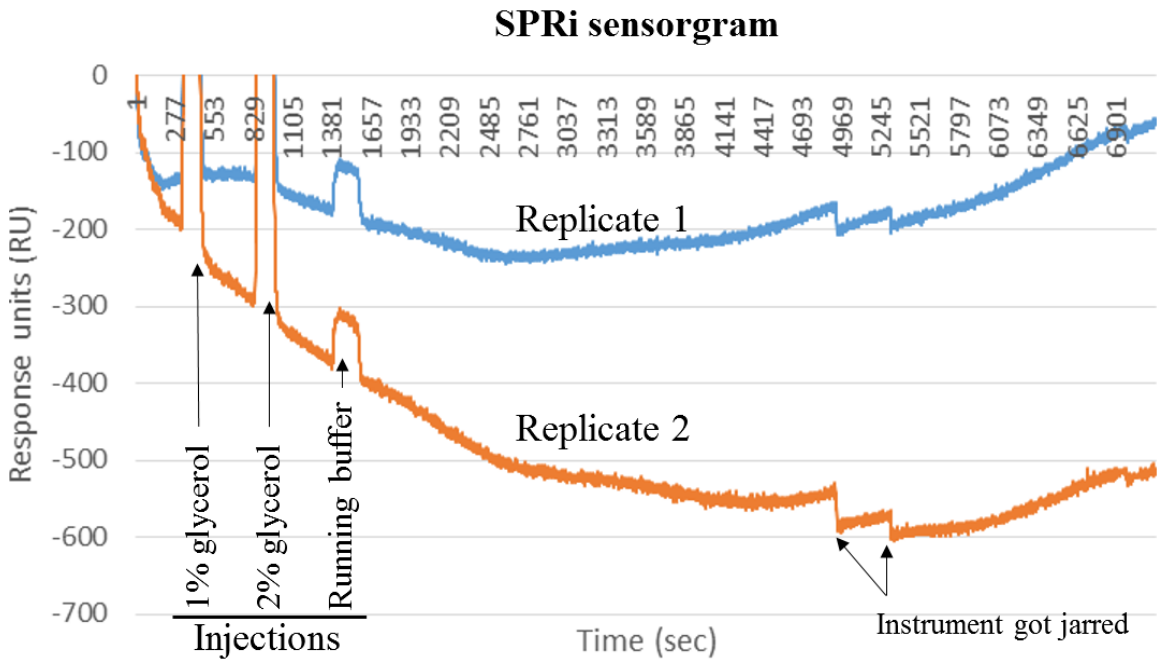


Figure 38. Baseline drift with NAPPA-SPRi across duplicate spots of expressed PRCKA target protein. Sensorgram was zeroed at time 0 sec.

extension of the dextran layer over time. Regeneration and extremely long injection times help to stabilize the surface and, therefore, the baseline drift.

Baseline drift was observed with NAPPA-SPRi regardless of the buffer that was applied. In the example shown in Figure 39, the drift was -6.4 RU / minute during part of the analyses when only buffer was applied to the slide (i.e., from a 1,000 sec segment). The drift was also not reproducible across replicate spots on the same slide. Surface regeneration, as mentioned previously, was not an option. Extremely long buffer injections (> 5 hours) per slide were not practical since multiple chips in one block or batch needed to be run after an > 8-hour slide preparation (i.e., expressing, e-phosphorylating, re-phosphorylating the target proteins). Furthermore, the non-antibody target proteins may not be stable in buffer for that long.

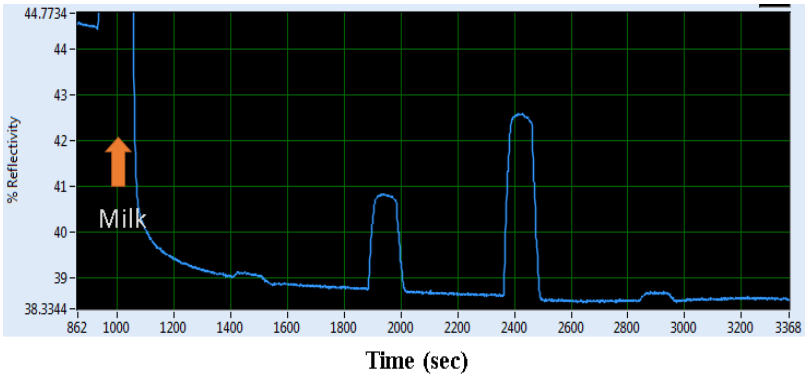


Figure 39. Addition of milk supernatant to NAPPA-SPRi sensor surface removes a lot of nonspecifically-bound mass.

The addition of 5% (w/v) milk supernatant removed nonspecifically-bound material from the slide surface (Figure 40, Figure 41). 5% BSA (w/v) also removed mass from the sensor surface, but not as much as the milk supernatant (data not shown). Milk further helped to stabilize baselines across replicates, but did not remove drift. In the example given in Figure 41, the drift was 21.7 RU / minute. Interestingly, the drift appeared to be specific for the target protein as these baselines were similar across different slides (Figure 42). Based on these observations, it appears that the baseline drift

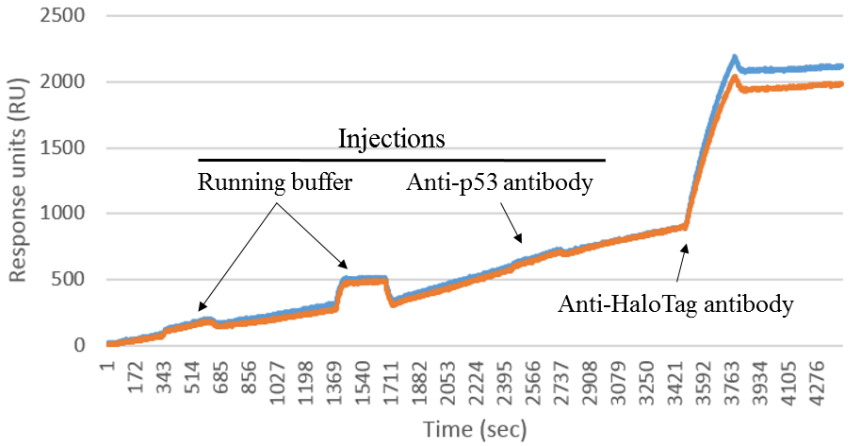


Figure 40. Baseline drift is across duplicate spots of expressed PRCKA target protein following the addition of milk supernatant.

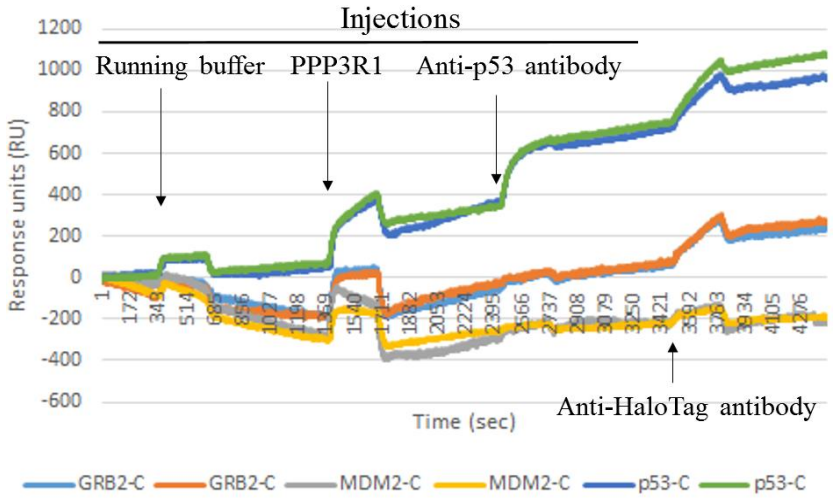


Figure 41. Baseline drift across different slides for the same expressed target protein following the addition of milk supernatant. "-C" indicates that the HaloTag is at the C-terminus of the target protein.

from NAPPA-SPRi may be due to spot-specific matrix effects in regards to the poly(L-lysine), BS3, HaloTag (O4) amine ligand, plasmid cDNA, and/or displayed protein.

Drift correction was applied on a per-spot basis during data analyses.

4.4.5 Kinetic titration

Surface regeneration with acids, bases, nonpolar water-soluble solvents, high detergents, or high salts disassociates any query protein that remains bound to the target following dissociation, thus allowing the slide to be re-used (Andersson, Areskoug, & Hardenborg, 1999; Helmerhorst, Chandler, Nussio, & Mamotte, 2012). This approach is advantageous because it reduces cost and allows the kinetics to be better approximated through multiple injections of varying query concentrations. Regeneration is often used when the target proteins are antibodies, which are very stable protein species that remain functional even in the presence of regeneration buffers. However, regeneration may not be appropriate for other applications. In the case of this experiment, in which the PPIs of

> 100 different proteins in the BCR signaling pathway are studied, most of the proteins are likely to be negatively affected by the regeneration buffer, either through partial or full denaturation.

Kinetic titration is an alternative option when regeneration is not possible (Schasfoort & Tudos, 2008). During kinetic titration, the query is added to the sensor surface as increasing concentrations in consecutive injections without regeneration (i.e., "multi-cycle kinetics"). This approach has been successfully used with 4-channel Biacore SPR instruments where each flow cell of a moderate-capacity carboxymethyldextran-derivatized CM5 sensor chip routinely captures > 10 ng of target protein (Jokiranta et al., 2001). *[The sensor chip, according to the Biacore Assay Handbook, has dimensions of 2.4 x 0.5 (l x w) with a ~ 100 nm dextran matrix coating (Murthy, Voelcker, & Jayaraman, 2006). It also well-established that 1 RU = 1 pg/mm2 for dextran-coated slides (Potyrailo & Mirsky, 2009). Jokiranta et al., for example, used CM5 sensor chips to immobilize C3b protein, which resulted in a 10,000 – 14,000 RU shift.]*

I explored the possibility of performing kinetic titrations with NAPPA-SPRi. Target proteins immobilized on the array included the HaloTag fusion target proteins, TP53 and AKT1, following cell-free expression and recombinant purified proteins, TP53 and GST, which were obtained from Sigma-Aldrich, Co. (St. Louis, MO). The recombinant proteins were diluted to 50, 100, 200, and 400 ng/µL for printing. A mouse anti-TP53 monoclonal antibody (D01 clone) was applied to the array at increasing concentrations, ranging from 2.67 E-08 to 5.33 E-07 M. Kinetic titration was a feasible

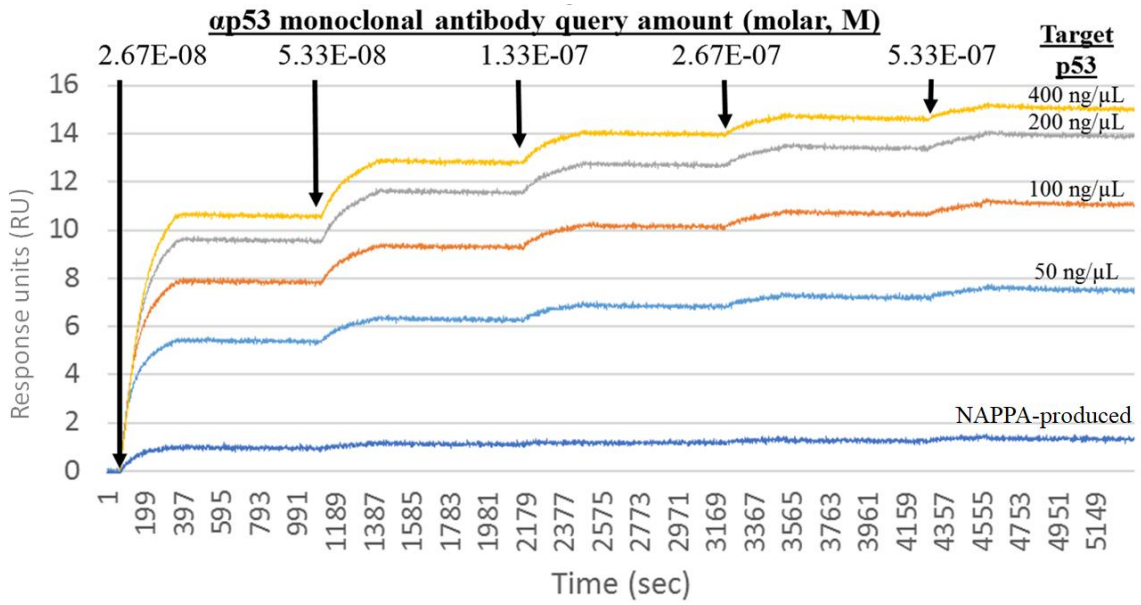


Figure 42. Kinetic titration of anti-TP53 antibody to spotted TP53 recombinant proteins and displayed TP53 expressed and captured with the NAPPA-SPRi approach.

approach for analyzing the binding kinetics between the antibody and recombinant proteins (Figure 43). However, it was not compatible with NAPPA-SPRi chemistry since the target protein on the surface is nearly saturated with minimal query concentration, indicating that amount of target protein is low.

4.4.6 3D surface chemistry

Instead of a monolayer of target protein, a 3D hydrogel surface chemistry can be employed in SPR experiments to increase the amount of target protein that is immobilized – thus resulting in higher signal (Schasfoort & Tudos, 2008). Could a 3D-based NAPPA-SPRi be possible? If so, it would increase signal and may make kinetic titration possible.

Some NAPPA users, including myself, have successfully used a 3D hydrogel slide in their fluorescent-based experiments, noticing that the slide results in a higher

signal-to-noise ratio than regular NAPPA slides coated with aminopropyltriethoxysilane due to lower nonspecific binding (personal communication with Dr. Ji Qiu and Kailash Karthikeyan of VGP CPD). These amine-reactive, thin film 3D polymer-coated slides are NEXTERION® Slide H slides from Applied Microarrays, Inc. (Tempe, AZ). They incorporate a cross-linked, multi-component polymer layer on glass that is activated with NHS to covalently immobilize amine groups. Unfortunately, the funding to optimize making this polymer layer compatible on the SPRi chips was not available.

A sensor chip with HC polycarboxylate hydrogel, NHS activated surface chemistry that is compatible with pin printing was obtained from XanTec Bioanalytics (Germany). It had a gold thickness of 43 nm and low fluorescence soda lime glass with an RI of 1.52.

Preliminary attempts to use this chip were not successful (i.e., no binding response detected). Later, Dr. Lusheng Song also tried to get a 3D-based NAPPA-SPRi version to work; he was also unsuccessful.

4.5 Optimized NAPPA-SPRi chemistry details

As mentioned previously in Chapter 4.1.1, the printing master mix must include BSA to retain the plasmid cDNA to the sensor surface. Although the mechanism is only speculative, it is believed that the BSA essentially forms a meshwork – in which the DNA is captured – through its lysine residues and the amine-terminal surface via an amine-to-amine crosslinker. BSA is 67 kDa with 618 amino acids, 59 of which are lysine residues (or < 10% of the mass). While the 90% “non-necessary” mass is not an issue

with fluorescent-based NAPPA, it is a concern with the SPR technology in which this extra mass can decrease the signal-to-noise ratio of the detected binding events. How then could these lysines be utilized while discarding the unnecessary mass from the surface?

BSA also had an additional disadvantage, which is discussed in more detail in Chapter 6.2. Very briefly, the addition of BSA in the target protein spot made analyses of phosphorylation events very difficult since BSA can also be phosphorylated.

The solution for both BSA-related issues was poly(L-lysine), a positively charged amino acid polymer of lysines. Although the polymers are large (70 – 150 kDa), 100% of the residues would theoretically be useful, thus requiring less material than BSA. Another advantage of using poly(L-lysine) is that it is unlikely to be phosphorylated. First, lysine phosphorylation is not recognized as a common post translational modification when compared to serine, threonine, and tyrosine phosphorylation. Second, the amino acid residues surrounding known phosphorylation sites are generally conserved, indicating that phosphorylation is affected by the neighboring amino acids. The use of poly(L-lysine) is discussed in more detail in Chapter 6.2.

The optimized reagent amount in the printing mixture for NAPPA-SPRi was finally determined to be: 0.4 µg/µL plasmid cDNA, 0.0003% poly(L-lysine), 250 µm HaloTag amine (O4) ligand, 291 µm BS3, and 0.007% DMSO in water. Compared to the standard GST-based NAPPA printing master mix, this represented a 92% decrease in mass (Figure 44). The substitution of poly(L-lysine) instead of BSA, along with the subsequent DOE analyses, resulted in ~50% less mass on the surface and a 90% increase in SPRi signal when compared to printing master mix #11, which was chosen following

the DOE experiment in Chapter 4.3 (Figure 45). The inter- and intra-slide reproducibility for plasmid cDNA deposition is R^2 ~ 0.93 (Appendix J). The inter- and intra-slide reproducibility for protein display is R^2 ~ 0.79 (Appendix J). The reproducibility was determined with ~ 110 unique proteins.

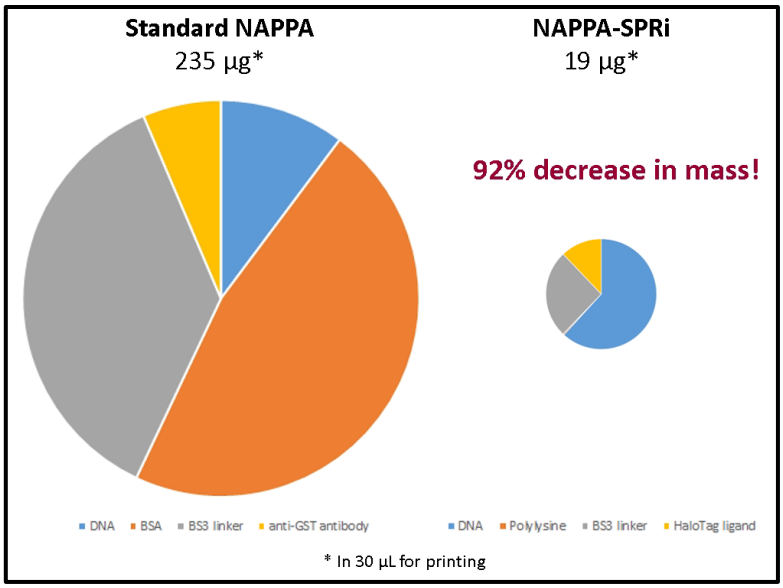


Figure 43. Mass and ratio of reagents in the optimized NAPPA-SPRi printing mixture were significantly differently than that used in standard GST-based NAPPA.

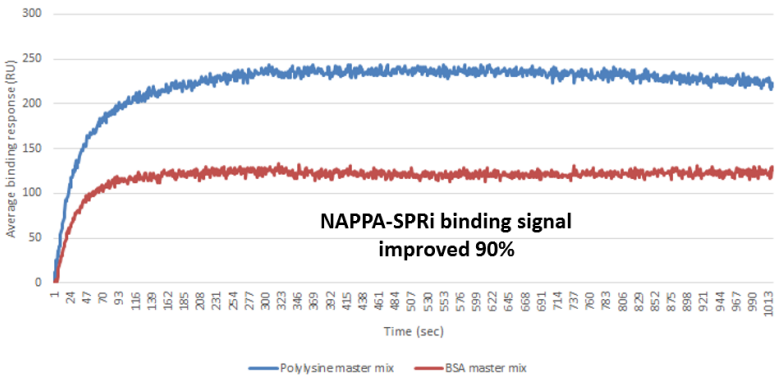


Figure 44. SPRi binding response between anti-TP53 monoclonal antibody query and C-HaloTagged TP53 target protein. Sensorgrams referenced to the non-binder, LYN target protein, with the same master mix.

This master mix combination was used in the quantitative analyses of the B cell receptor signaling pathway, which is described in more detail in Chapter 8.

4.6 Potential alterations of NAPPA-SPRi

Improvements in NAPPA-SPRi reproducibility, throughput, and protein immobilization should be explored. For example, the plasmid cDNA printing mixture

was pin-spotted onto the arrays in this iteration of NAPPA-SPRi. Other printing methods, like piezoelectric printing or microfluidics, may result in more uniform sample deposition on the array since piezo printers have better control over the amount of sample that is deposited and are not as greatly affected by changes in humidity as pin spotters.

NAPPA-SPRi throughput could also be increased. Like standard NAPPA, NAPPA-SPRi throughput is limited by the diffusion of expressed proteins during *in vitro* transcription translation since spots that are less than 640 μm apart (i.e., center-to-center) may become contaminated with proteins from neighboring features. Takulapalli et al. addressed the diffusion and issue through the use of silicon nanowells that were approximately 250 microns in diameter and 75 microns deep (Takulapalli et al., 2012). The nanowells physically blocked the diffusion of expressed proteins to other nanowells and, as such, the throughput of NAPPA increased from 2,300 to 14,000 features per slide. More recently, Karthikeyan et al. used a "contra capture" approach to capture the expressed proteins separately from the printing mixture (Karthikeyan et al., 2016). The printing mixture containing plasmid cDNA, amine-to-amine crosslinker, and BSA was deposited into microwells of a polydimethylsiloxane (PDMS) wafer. During protein expression, the wafer was sandwiched to a slide coated with capturing reagent. The expressed protein was then immobilized by the capturing slide and the PDMS wafer was discarded. A variation of these themes could be applied toward NAPPA-SPRi. Contra capture NAPPA-SPRi via the use of a nanowell wafer would increase the throughput. It would also reduce the mass on the SPR slide, thus increasing the sensitivity of the platform. Another important consideration is that DNA within each spot may not be

compatible with certain experiments; for example, transcription factors that bind to proteins *and* DNA cannot be used as queries because their mixed binding responses could not be resolved into their separate binding events (i.e., transcript factor-protein binding, transcription factor-DNA binding). Finally, it is likely that the contra capture approach would also decrease baseline drift – and therefore data analyses – because the slide surface would be less complex.

The amount of immobilized protein per spot on NAPPA-SPRi is too low for kinetic titration (see Chapter 4.4.5). Kinetic titration is a method that allows the binding kinetics of multiple query concentrations to be analyzed without the need for slide regeneration. A possible solution to this issue would be the use of a 3D surface chemistry, like a carboxymethlated dextran matrix.

4.7 Acknowledgements

I am grateful to Dr. Mitch Magee and Dr. Lusheng Song for training me on the Plexera® SPRi instrument, as well as answering my many technical questions. Dr. Shaopeng Wang was also a good source of information. I thank Ian "Instrument Whisperer" Shoemaker for his help regarding instrument issues, which always seemed to crop up at the most inopportune times. Although this is mentioned in the text above, the ability to couple NAPPA with SPRi in its current form depended on having a chip that was amenable for contact printing; Plexera Bioscience LLC worked carefully with us to help make this happen. Lastly, the development of NAPPA-SPRi could not have been

possible without the financial support of the PS-OC grant and Dr. LaBaer's discretionary funds.

CHAPTER 5

5 PURIFED QUERY PROTEINS FOR NAPPA-SPRi

5.1 Purified query proteins for accurate kinetic analysis

SPR analyses must utilize one very pure query protein per injection to simplify data analyses. A binding response between a known target protein "A" and a query solution containing only one protein "B" directly reflects an interaction between proteins "A" and "B." With query solutions containing two or more proteins, the delineation of which query (or queries) is binding to the captured target protein may not be possible. Query proteins that are not purified from the host expression system present an additional issue during data analyses. Cell lysate is markedly different than buffer alone and will result in a large bulk refractive index shift that may be outside the linear range or detection limit of the instrument. Indeed, a query-of-interest expressed in human cell lysate and diluted in buffer resulted in a refractive index shift that overwhelmed any real binding response on NAPPA-SPRi (data not shown).

Purified recombinant proteins can either be obtained in-house or through a commercial source. There are two primary reasons why researchers would prefer making the proteins instead of purchasing them through a company: flexibility and cost. In-house processing means that the user can control which proteins are purified, the host system in which they're expressed, the fusion tag, the location of fusion tag, and the final amount of material. In-house recombinant proteins can also be a cheaper option compared to commercially-produced proteins, which range from hundreds to thousands of dollars depending on the protein, purity level, and amount. However, there are two main reasons

why a recombinant protein from a company would be preferred: time and resources since protein expression and purification can require a lot of troubleshooting and optimization. The project may need the purified protein immediately and may not be able to wait for the plasmid with the gene-of-interest to be prepared and validated, the host system to be grown, the host to be transformed or transduced, and the protein to be purified and tested. The experiment may require the protein to be highly pure and functional. Whereas these proteins can be easily produced and tested in a company that routinely performs such experiments with expensive equipment like fast purification liquid chromatography (FPLC, ~ $50K for a new unit), research laboratories that perform small-scale protein purifications generally rely on more cost-effective gravity-dependent columns that have notoriously low resolution. It should be noted, however, that some SPR and non-SPR applications do not require a highly purified or functional protein. For instance, an SPR experiment may simply screen for query-target interactions without needing kinetics and affinity information.

SPR and SPRi have traditionally relied on purified recombinant proteins to be the targets as well as the queries, which has ultimately limited their throughout due to the tedious process of protein expression and purification. With NAPPA-SPRi, the need for purified target proteins is no longer a requirement; however, the need for purified proteins as queries has not yet been abrogated. For NAPPA-SPRi, the possibility of expressing and purifying the query proteins in-house was investigated with the initial intention of using $\geq$ 50 query proteins for these analyses. For the POIs in the BCR signaling pathway, these were, on average, ~ $300 / 10 µg. *In vivo E. coli* expression,

wheat germ cell-free expression, and human cell-free expression were pursued as options to produce the human proteins, which is described in more detail in the manuscript, "Development of a full-length human protein production pipeline," that arose, in part, from this work (Saul et al., 2014). It compared the expression and purification rates of 31 full-length human proteins ranging from 10 – 120 kDa in *E. coli*, the cell-free wheat germ expression system, and the cell-free human expression system.

5.2 Producing and purifying query proteins in-house

The primary advantage of expressing proteins in *E. coli* is that a large amount of protein can be made with minimal cost. The use of *E. coli* to synthesize the BCR signaling pathway proteins was first pursued using the maltose binding protein (MBP) as an N-terminal fusion tag because MBP significantly enhances the solubility of the proteins-of-interest (POIs) compared to other fusion tags, which is particularly important since inclusion bodies are often produced during recombinant protein expression in *E. coli* (Kapust & Waugh, 1999). Protein purification was performed using an amylose affinity column, which bound to the MBP tag. The MBP-protein was then eluted via competitive exchange with the addition of maltose.

E. coli is known to have difficulty expressing proteins that are larger than 50 kDa. However, the handful of BCR signaling pathway proteins that were expressed in *E. coli* appeared to be fully translated despite being > 50 kDa (Figure 46). Of note, a fraction of the POIs were not fully translated, such that only the 40 kDa MBP N-terminal tag was

expressed. The disadvantages of using *E. coli* as an expression system and the purification approach became clear with these first attempts. First, protein purification using a column resulted in very impure samples (Figure 46). In addition, the MBP-amylose interaction is rather weak (micromolar affinity), thus causing a significant amount of the POI to be lost during the washing steps (data not shown)(Terpe, 2003). Both of these issues could be minimized with the optimization of buffers and numbers of washes, although the use of amylose affinity chromatography to purify MBP-tagged proteins is known to result in insufficient purity for various types of studies (Austin, Nallamsetty, & Waugh, 2009). Second, the use of a tag, particularly a larger one like MBP, could also occlude binding sites on the query. It would be necessary to have a small tag (like His) or a specific cleavage site between the tag and POI. Finally, human proteins expressed in *E. coli* may not be folded properly since *E. coli* do not have the same chaperone proteins as human cells that facilitate proper 3D conformation. As a consequence, the activity and interactions of the POIs may be affected. Taking these considerations in mind, it was decided that eukaryotic cell-free expression systems would be a preferable alternative to synthesize the query proteins for NAPPA-SPRi.

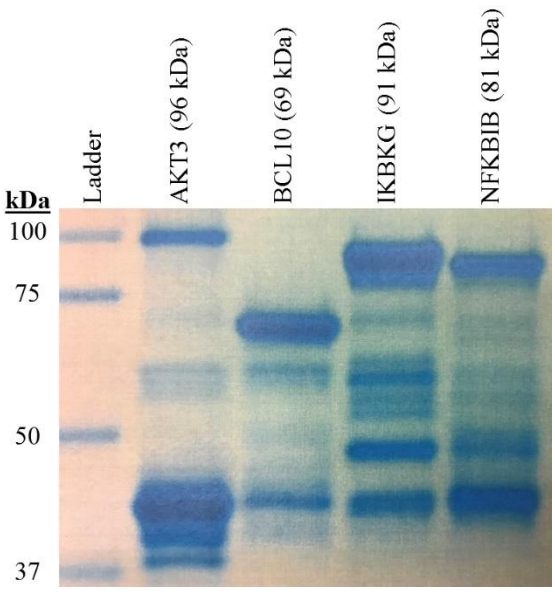


Figure 45. Coomassie gel of purified recombinant proteins with N-terminal MBP tags expressed in *E. coli*. MBP is 40 kDa.

The production of recombinant proteins using cell-free expression systems has several advantages over *in vivo* cell-based techniques (Zemella, Thoring, Hoffmeister, & Kubick, 2015). First, the only thing that is required in cell-free expression is a plasmid backbone with the appropriate promoter, which is mixed with the lysate mixture to produce proteins in < 8 hours, or more if a dialysis approach is used to increase protein yield. Thus, cell-free expression is much faster in synthesizing protein than cell-based methods because it does not require gene transfection or cell culturing. Second, proteins or labels that may be toxic to a cell host can be synthesized. Third, non-natural amino acids can be incorporated. Finally, additives, detergents, cofactors, and binding partners can be easily added to the cell-free expression system. One disadvantage of using a cell-free system is that it does not produce as much protein as a cell-based system (i.e., micrograms to milligrams in a cell-free system versus milligrams to grams of protein in a cell-based system). Another important disadvantage of a cell-free expression system is that the cost to amount of purified protein ratio is low compared to using *E. coli* cells. A comparison of the advantages and disadvantages of the different cell-free expression systems are in **Error! Reference source not found.**.

Eukaryotic cell-free expression systems that were explored for this project were derived from wheat germ and human HeLa cells. Due to its low expression efficiency, the rabbit cell-free expression system was not tested to produce the query proteins (see also page 103, Figure 33) (L. Zhao et al., 2010).

Wheat germ is the vitamin-rich sprouting section (i.e., “embryo”) of a wheat kernel, and the use of its lysate to produce proteins was first reported in 1973 when

Table 11. Comparison of different cell-free protein expression systems. Adapted from (Harbers, 2014)

System	Advantages	Disadvantages
E. coli	• Established systems Tested for many proteins •Established systems •Tested for many proteins •Low cost •Easy to make extracts •Very high translation speed •Very high protein yield •Genetic modification of strains possible •Recombinant system available (PURE)	•Many eukaryotic proteins insoluble •No post-translational modifications •Codon usage optimization is preferable
Wheat germ	•Well established system •Tested for many proteins •Best yield for eukaryotic system •Very high solubility rate •Synthesis of very large proteins demonstrated (~200 kDa range) •Cap independent translation •Commonly no codon optimization needed •Stable system allowing work with different additives	•Extract preparation time consuming •Some background phosphorylation was observed in protein MS studies •No glycosylation
Insect	•Translation of large proteins •Cap independent translation •N-glycosylation possible •Formation of disulfide bridges •Used in combination with vesicles	•New system •Lower protein yields than E. coli or wheat germ systems •Higher cost •Tested for few proteins
Rabbit reticulocyte	•Old but very well established system •Tested for many proteins •Mammalian system •Cap independent translation •Often used in research	•Treatment of animals required •Sensitive to additives •No glycosylation •Co-expression of off-target proteins •Hemoglobin concentration ~90% of protein
Human	•Some protein modification	•Low yield •Sensitive to additives •Tested for limited number of proteins

commercially available wheat germ was able to produce tobacco mosaic virus-related proteins and rabbit 9S globin (Roberts & Paterson, 1973). Since then, the cell-free wheat germ expression system has become a highly efficient protein production option, synthesizing 1.6 – 20 mg of protein depending on the different reaction formats (Harbers,

2014). Various studies have highlighted its ability to successfully synthesize human proteins compared to both *in vivo* and *in vitro E. coli* systems (Harbers, 2014).

The human cell-free expression system uses lysate from HeLa cells, a cervical cancer cell line that was cultured in 1951. The use of HeLa cells to express recombinant proteins was first suggested in 1973 when Reichman et al. showed that HeLa cell extract was able to initiate polypeptide formation using radiolabeled N-terminal methionine (Reichman & Penman, 1973). As a human expression system, it offers unique advantages over other expression systems when producing human proteins. For example, it can synthesize high molecular weight human proteins due to the natural codon usage (Zemella et al., 2015). It also possesses the chaperone machinery to fold proteins into their native conformations.

Plasmids compatible with the wheat germ and human cell-free expression system were constructed by my colleague, Justin Saul, with a HaloTag at the N-terminus or C-terminus (pJFT7_nHalo, pJFT7_cHalo). A tobacco etch virus (TEV) sequence that is specifically targeted for cleavage by TEV protease was placed between the POI and HaloTag.

Proteins in the BCR signaling pathway were successfully synthesized with *E. coli*, wheat germ, and HeLa cells (Saul et al., 2014). The eukaryotic cell-free expression systems had a higher success rate of expression than *E. coli* across the tested proteins (87% vs 73%). Of the 30 – 31 POIs that were tested, only 10 proteins were soluble in *E. coli* while at least 25 were soluble in the cell-free extracts. Expression yield was variable across the different POIs and systems. Longer incubation times are used to increase

protein yield; however, this study saw that longer incubation also resulted in increased degradation of full-length POIs.

The POIs were then purified using a ligand-conjugated resin and HaloTagged TEV protease from Promega Corporation (Madison, WI), which also bound to the resin and were removed from the purified fraction. More proteins were purified when they were expressed in the human extract, yet the yield of purified product (i.e., $\geq$ 1 µg, $\geq$ 90% purity) was the lowest of the three systems. That is, 6 proteins were purified from *E. coli* with 42% purification recovery, 10 proteins were purified from wheat germ extract with 24% purification recovery, and 13 proteins were purified from human cell extract with 15% recovery. The mean purification yield was dismal. With 100 µL (~ $125) of the human cell expression system using a dialysis format to replenish the reagents, the mean purification yield was only 3.5 µg. Furthermore, the cleaved HaloTag protein and HaloTagged TEV protease were common impurities in the samples. For a HaloTag-based NAPPA chemistry, these impurities are definitely causes for concern and would complicate SPR analyses. HaloTag can bind to unbound HaloTag ligand. The TEV protease could cleave target POIs from the slide since the TEV site is located between HaloTag and the POIs.

5.3 Purchasing purified query proteins

After this in-house effort, several considerations led us to opt to purchase the query proteins from commercial sources. An exhaustive search for purified proteins from commercial sources that could be employed as queries in the NAPPA-SPRi experiments

was performed. These were the criteria that had to be met: 1) human protein, 2) important protein in the BCR signaling pathway, 2) expressed in human or insect cells, 3) have a small tag (e.g., His), and, if possible, 4) tested for activity such that kinases were tested by their ability to phosphorylate a substrate and adaptors were tested by their ability to bind to a known interactor. The query proteins that fit all of these criteria included BLNK, BTK, and PIK3CA/PIK3R1 (PI3K). Notably, the catalytic PIK3CA protein is unstable without its heterodimer regulatory partner, PIK3R1, and is always found as a purified complex. Two major proteins in the BCR signaling pathway, SYK and DAPP1, were unattainable or not tested for functionality, respectively. DAPP1 was purchased despite not having known functionality because of its importance. Not surprisingly, it did not interact with any target proteins on the array.

Two queries produced in-house, RAC1 and RHOA, were used with NAPPA-SPRi analyses because they are important in B cell signaling and were easily and cheaply obtained. They were expressed in *Escherichia coli* by members of Dr. Kim Orth's laboratory (UT Southwestern; Dallas, TX), and have been used in various published experiments demonstrating their activity (Woolery, Yu, LaBaer, & Orth, 2014; Yarbrough et al., 2009). Moreover, the proteins were tested for functionality by Dr. Xiaobo Yu (personal communication; National Center for Protein Sciences; Bejing, China) by their ability to bind to known protein partners on NAPPA arrays.

The purified query proteins were then used to analyze protein interactions with NAPPA-SPRi. The data are in Chapter 8.

5.4 Acknowledgements

I extend my gratitude to Justin Saul, who was always eager to help me out whenever he could, and who worked tirelessly alongside me as I tried to find a way to express and purify a large amount of human protein for NAPPA-SPRi. Bejamin Ober-Reynolds assisted in these experiments. I also appreciate the useful discussions regarding cloning, bacterial DNA replication, and DNA extraction that I had with Dr. Mitch Magee and Dr. Fernanda Festa. These experiments were made possible through the PS-OC grant and Dr. LaBaer's discretionary funds.

CHAPTER 6

6 MODULATING PROTEIN PHOSPHORYLATION ON NAPPA

6.1 Introduction

Numerous cellular mechanisms are mediated by protein phosphorylation, and aberrant phosphorylation has been linked to a range of disorders, including cancer, diabetes, cardiovascular disease, inflammatory diseases, and infectious diseases (Fabbro, Cowan-Jacob, & Moebitz, 2015). Due to the role that phosphorylation plays in disease, the development of effective kinase inhibitors to treat the various disorders has been pursued with gusto. Since the first kinase inhibitor to treat chronic myeloid leukemia was approved by the FDA in 2001 (i.e., Gleevec), 27 more were approved for other types of diseases within the next fourteen years (Wu, Nielsen, & Clausen, 2015). One of these, Ibrutinib, has been approved to treat B cell lymphomas, including chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL), and marginal zone lymphoma (MZL). It is also used to treat chronic graft-versus-host disease, which has been associated with altered B cell activation and signaling (Rhoades & Gaballa, 2017). It inhibits the enzymatic activity of BTK, a kinase that is critical to B cell growth and survival, by binding covalently to a cysteine in BTK's catalytic domain (Wu et al., 2015). Notably, BTK was used as a query in my NanoBRET and NAPPA-SPRi experiments.

Phosphorylation is used as a major mechanism of signal transduction for the BCR signaling pathway and is generally considered to be a PTM of positive regulation (see also Chapter 2.4). Therefore, the study of protein interactions within the BCR signaling pathway should consider interactions with and without phosphorylation. Since the human

cell-free expression system can phosphorylate the proteins that it expresses, the de-phosphorylation of NAPPA proteins to enable PPI analyses in the absence of phosphorylation is addressed in Chapter 6.2. The phosphorylation of target proteins in a B cell-specific manner is described in Chapter 6.3.

6.2 De-phosphorylation of NAPPA proteins

6.2.1 Standard de-phosphorylation protocol

NAPPA proteins are produced with a cell-free expression system based on the HeLa cervical cancer cell line, which has phosphorylation capability that likely reflects a phosphorylation pattern specific to the cells at the time that they were collected. Thus, the inherent phosphorylation of the target proteins was unlikely to be physiologically relevant for this study of the B cell receptor signaling pathway. Herein, I describe the de-phosphorylation of target proteins, which enabled their phosphorylation in a B cell-specific manner (see Chapter 6.3) and established a baseline to which the phosphorylated data could be compared.

Target de-phosphorylation for NAPPA-SPRi analyses was first attempted following NAPPA chemistry optimization employing BSA (see also Chapter 4.3). The standard protocol that was developed in the Virginia G. Piper Center for Personalized Diagnostics (VGP CPD) to de-phosphorylate standard NAPPA arrays required that the slide be incubated twice with 18K units of lambda protein phosphatase (LPP) for 45 min at 30 ^{o}C, which resulted in 36K units or $184 of LPP per slide. However, this cost was inhibitory in this project. Furthermore, it became apparent from the data that de-

phosphorylation using the standard protocol was not complete (Figure 47). The most likely phosphorylated component in the printing mixture was BSA, which is necessary for plasmid cDNA deposition. BSA shares 76% sequence homology with human serum albumin (HSA), which is well-documented to be phosphorylated at multiple sites. The use of BSA in the printing mixture, therefore, would require more phosphatase to completely de-phosphorylate the spots than displayed proteins alone. In addition to increasing background and complicating the analyses of kinase arrays, phosphorylation profiles, and de-phosphorylation optimization, BSA also presented a mass issue for NAPPA-SPRi since > 90% of its mass was unnecessary (see also Chapter 4.5). Poly(L-lysine) (PL), a polymer of lysine residues, offered a possible alternative to BSA. In addition to providing lysines that aid in DNA retention, PL cannot be phosphorylated since it lacks serine, threonine, and tyrosine residues.

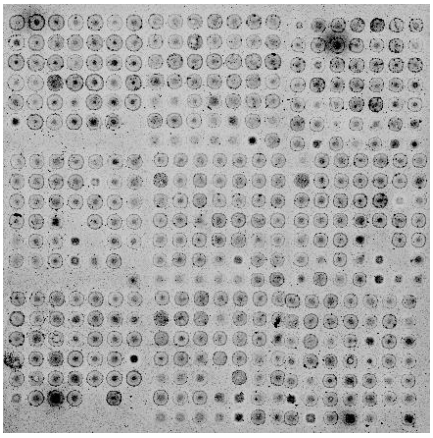


Figure 46. Fluorescent analyses of phosphorylated tyrosines following de-phosphorylation of a NAPPA array using an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody. False-colored gray-scale image where black reflects high levels of phosphorylation, while white reflects low levels of phosphorylation.

A HaloTag-based NAPPA array was prepared with BSA or PL in the printing mixture and then expressed with the human cell-free expression system. Using an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody, the level of phosphorylation for the four target proteins (i.e., GRB2, RAC3, LYN, MAPK9) was determined via fluorescence (**Error! Reference source not found.**). Although it is clear from the image that LYN is tyrosine phosphorylated, its relative intensity within the spot

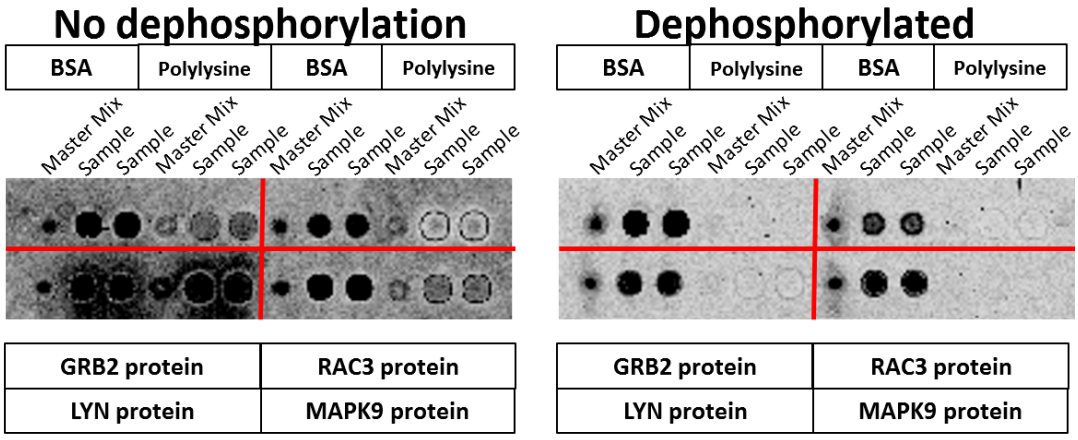


Figure 47. Tyrosine phosphorylation of target protein, LYN, before and after the addition of LPP as determined with an anti-phosphotyrosine antitbody and HRP-conjugated secondary antibody. Master mix = printing mixture without DNA or displayed protein. Gray-scale image where black reflects high phosphotyrosine level.

to other target proteins in the BSA-based printing mixture does not reflect this (Figure 49). Alternatively, LYN phosphorylation is significantly pronounced in the PL printing mixture.

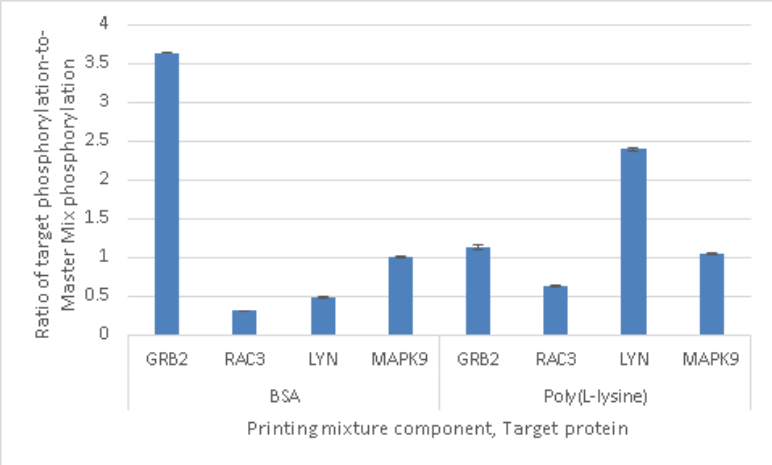


Figure 48. Tyrosine phosphorylation level of displayed proteins prior to de-phosphorylation compared to spots containing only printing master mix (no DNA or displayed protein). Error bars represent range of values across duplicate spots.

A second array was de-phosphorylated with LPP, and then probed with an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody (**Error! Reference source not found.**). De-phosphorylation of LYN in the BSA- and PL-based printing mixtures decreased the phosphotyrosine level by 31% and 84%, respectively. These results demonstrated that

BSA could be replaced by PL and, furthermore, that the use of PL resulted in lower phosphorylation background on NAPPA arrays.

The HaloTag-PL printing mixture was then optimized using a DOE approach to maximize the signal-to-noise ratio in NAPPA-SPRi analyses. Assuming an average molecular weight of 110 kDa, 0.7 µg of PL could be substituted for 14.65 µg of BSA in 30 µL of printing mixture, which helped contribute to a 90% increase in SPRi signal (see also Chapter 4.5).

6.2.2 De-phosphorylation optimization using DOE

The standard de-phosphorylation protocol of BSA-based NAPPA used 36K units of LPP, or $184 per slide, yet still resulted in incomplete de-phosphorylation. The use of PL instead of BSA in the printing mixture decreased background (see previous section), but questions remained: Could less than 36K units of LPP be used? Was only one phosphatase enough? Could the addition of other phosphatases decrease the overall cost of de-phosphorylation?

Husberg et al. compared the ability of four phosphatases to de-phosphorylate target proteins in cardiac muscle tissue. Their work demonstrated the unique and overlapping substrate specificities of two generic phosphatases (i.e., alkaline phosphatase, LPP) and two endogenous serine/threonine phosphatases in the heart (i.e., protein phosphatase 1, and protein phosphatase 2) (Husberg, Agnetti, Holewinski, Christensen, & Van Eyk, 2012). All 22 target proteins were de-phosphorylated by LPP or calf intestinal alkaline phosphatase (CIP), which supports the unique preference of LPP

Table 12. Protein Phosphatase Specificity Chart. Data from New England Biolabs

Phosphatase	Specific phosphatase activity (nmol/min/mg)*			Units required for 100% phosphatase removal
	p-Serine/Threonine	p-Tyrosine	p-NPP phosphatase activity	
LPP	26,600	2,200	1,290,000	100 - 500
CIP	1,520	26,440	128,300	10 - 50

* Phosphatase activity determined using phosphorylated myelin basic protein and p-nitrophenylphosphate (p-NPP)

for phosphorylated serine and threonine residues and CIP for tyrosine phosphorylation (Table 12).

CIP can de-phosphorylate proteins, but it is often not used in this manner in research laboratories. Instead, CIP is used to de-phosphorylate the 5' and 3' ends of DNA in cloning to prevent re-ligation of linearized plasmid DNA. Even in New England BioLab's product description, CIP is for dephosphorylating "5' and 3' ends of DNA and RNA," "cloning vector DNA to prevent recircularization during ligation," "DNA prior to end-labeling using T4 Polynucleotide," and "treatment of dNTPs in PCR reactions prior to sequencing or SNP analysis;" nowhere does it mention protein de-phosphorylation. This may be the reason why CIP was never used to de-phosphorylate NAPPA arrays.

Given the overlapping and unique substrate selectivity of LPP and CIP, I decided to perform a DOE two-level factorial experiment to optimize the de-phosphorylation of HaloTag-PL NAPPA microarrays with LPP and CIP. (To learn more about DOE, please see Chapter 4.3, page 92.) The low ("-1") and high ("1") levels of LPP, CIP, and number of incubations are depicted in **Error! Reference source not found.**. Reproducibility was determined with the use of duplicate spots. Fluorescent responses were achieved with an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody.

Table 13. De-phosphorylation DOE factors, levels, and responses as determined with an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody.

StdOrder	RunOrder	CenterPt	Blocks	Lambda PP	CIP	No. Incubations	Responses (Fluorescent Intensity)			
							SYK	BLK	BTK	MAPK14
1	3	1	1	-1	-1	-1	7489023	21242279	5687074	9981459
9	10	1	1	-1	-1	-1	7753694	20247578	7797177	8364363
11	5	1	1	-1	1	-1	5270646	12607530	3968517	6325507
3	7	1	1	-1	1	-1	4731364	12379572	5515866	5534016
10	6	1	1	1	-1	-1	4846662	11350427	6616229	5482511
2	11	1	1	1	-1	-1	4503201	11070727	5152214	5319444
4	9	1	1	1	1	-1	6877838	10085970	6874461	5429412
12	14	1	1	1	1	-1	6237153	5721424	5929799	6318374
13	8	1	1	-1	-1	1	4975975	14973518	4372892	5562295
5	15	1	1	-1	-1	1	4801562	13057346	5813109	4998380
15	1	1	1	-1	1	1	5632192	10470429	3793797	5047904
7	13	1	1	-1	1	1	5183485	9901595	4704700	5077499
14	2	1	1	1	-1	1	4005963	5106597	2500966	3678604
6	12	1	1	1	-1	1	3148167	5167750	2690539	2926834
8	4	1	1	1	1	1	6207282	14447029	4065023	6159505
16	16	1	1	1	1	1	6067050	14438219	5864541	5297188

Factor	-1	1
Lambda PP	1.2K units	12K units
CIP	30 units	300 units
No. Incubations*	1	3

* Each incubation at 30 C for 30 min

The mouse anti-phosphotyrosine monoclonal antibody (P-Tyr-100) from Cell Signaling Technology, Inc. (Danvers, MA) was used for all of the phosphotyrosine experiments herein because it detected more phosphorylation events than any of the other anti-phosphotyrosine antibodies on the NAPPA kinase arrays (antibody comparison performed by Dr. Fernanda Festa of the VGP CPD). Anti-phosphoserine and anti-phosphothreonine antibodies were also used for this de-phosphorylation optimization experiment, but did not result in sufficient signal. This is unsurprising since anti-phosphoserine and anti-phosphothreonine antibodies are sensitive to the adjacent amino acids, and generally must be chosen for the specific phosphorylation site. Anti-phosphotyrosine antibodies recognize phosphorylated tyrosines more independently of

the surrounding amino acid sequence; even so, the detection rate for this antibody was ~30% on the kinase arrays.

A phosphorylated slide and a slide de-phosphorylated with the standard protocol using LPP were also prepared at the same time as the DOE experiment. As Figure 50 shows, the standard protocol in which the slide was incubated twice with 18K units of LPP (36K units total) for 45 min at 30 ^{o}C did not result in complete de-phosphorylation despite substituting the BSA with PL in the printing mixture. LPP decreased the level of phosphotyrosine by ~ 70%.

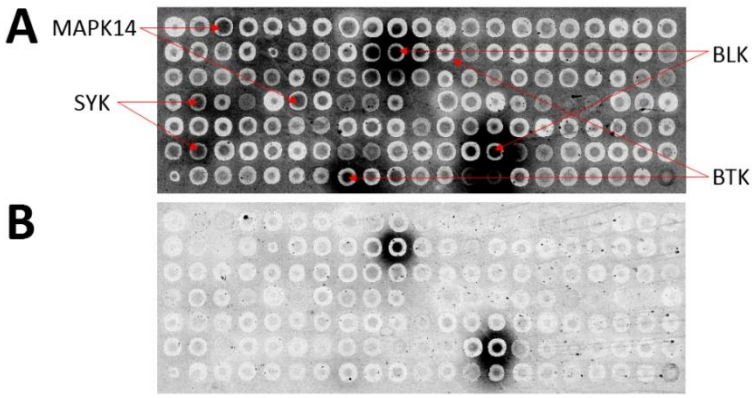


Figure 49. Fluorescent analyses of tyrosine phosphorylation of a A) phosphorylated slide and a B) slide de-phosphorylated using the standard LPP-based protocol. False-colored gray-scale image where black represents high phosphotyrosine level and white represents low phosphotyrosine level. Images analyzed with the same settings to make a direct comparison.

DOE analyses using the Minitab® 17 software indicates, through a Normal plot of standardized effects, that all factors and combinations are significant (Figure 51). As anticipated, both CIP and LPP de-phosphorylated tyrosines .

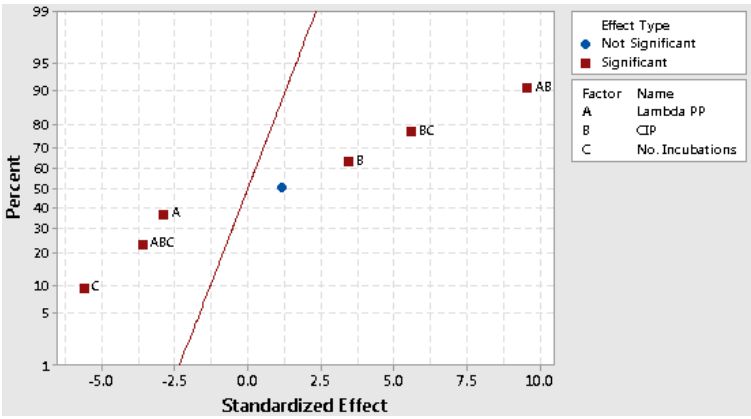


Figure 50. Normal plot of standardized effects for SYK response. Similar plots were obtained for other tyrosine phosphorylated target proteins. Alpha = 0.05

The parameter that led to the most de-phosphorylation was the highest amount of LPP (36K units), the lowest amount of CIP (300 units), and three incubations of 30 min at 30 ^{o}C. While the use of both phosphatases also de-phosphorylated the slide 50% better than the standard LPP-based de-phosphorylation.an interaction plot of the response means revealed that CIP and LPP interact with each other (Figure 52). This interaction interferes with de-phosphorylation as high levels of both phosphatases result in poorer de-phosphorylation than when CIP is added at a lower amount (Figure 53). This interaction may be the de-phosphorylation of one phosphatase by the other, which would affect phosphatase activity. In other words, LPP and CIP should not be mixed together, but rather added to the slide separately. Moreover, additional incubations with CIP did not significantly affect the level of de-phosphorylation. The number of incubations, however, was important when using LPP.

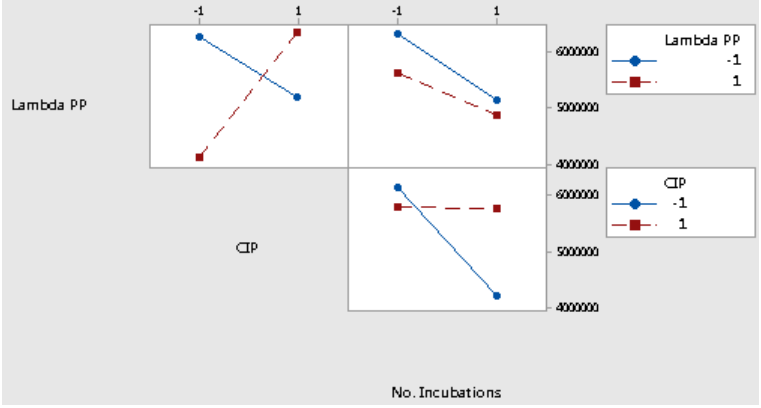


Figure 51. Interaction plot of response means for SYK. Other target proteins had similar plots.

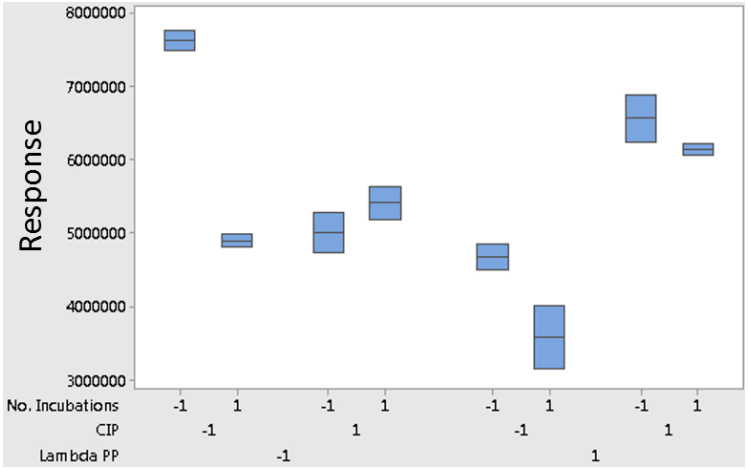


Figure 52. Box plot of response for SYK across different factors and levels. Other target proteins had similar plots.

To determine how much CIP should be used to de-phosphorylate the target proteins, 500 units of CIP (or buffer) were added for 30 min at 30 ^{o}C, then probed with an anti-phosphotyrosine or anti-phosphoserine antibody. CIP significantly de-phosphorylated tyrosine residues (Figure 54), but did not appear to affect serine phosphorylation of SYK. These results matched the known substrate specificity of CIP for phosphorylated tyrosines (Table 12).

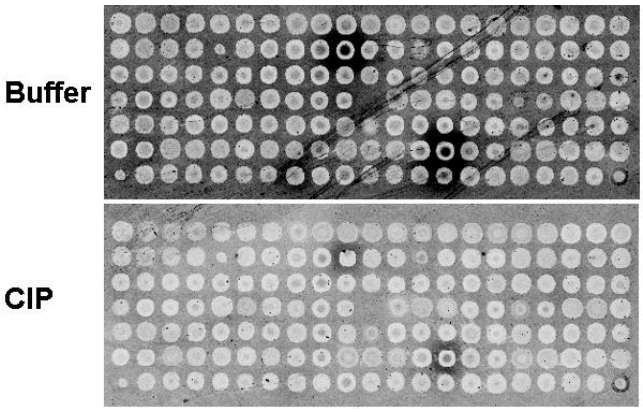


Figure 53. Fluorescent analyses of tyrosine phosphorylation of slides incubated with CIP buffer or enzyme. False-colored gray-scale image where black represents high phosphotyrosine level and white represents low phosphotyrosine level.

The DOE experiment, revealed that CIP and LPP could not be mixed together during de-phosphorylation, and that the number of incubations with LPP made a dramatic impact on the response. Therefore, NAPPA microarrays were first incubated with 500 units of CIP for 30 min at 30 ^{o}C. The arrays were then incubated 1 – 3 times with 2K, 4K, or 6K units of LPP for 30 min at 30 ^{o}C. Tyrosine and serine phosphorylation were fluorescently assessed with anti-phosphotyrosine and anti-phosphoserine antibodies, respectively, with an HRP-conjugated secondary antibody. Tyrosine de-phosphorylation was similar with 2K or 4K units of LPP across two or three incubations (Figure 55).

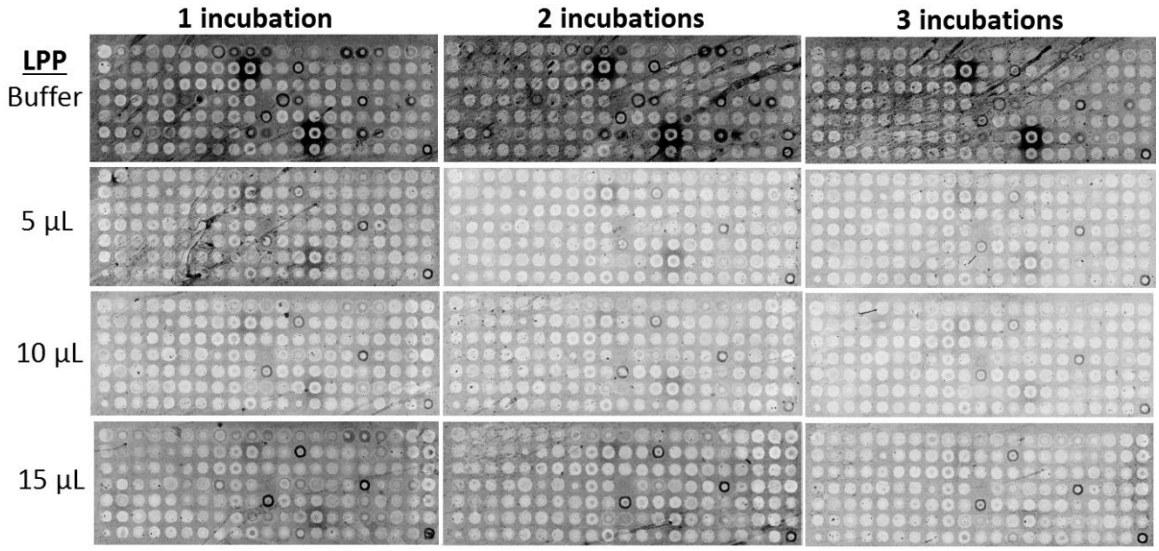


Figure 54. Fluorescent analyses of tyrosine phosphorylation of slides incubated with 500 units of CIP and then LPP buffer or enzyme. False-colored gray-scale image where black represents high phosphotyrosine level and white represents low phosphotyrosine level.

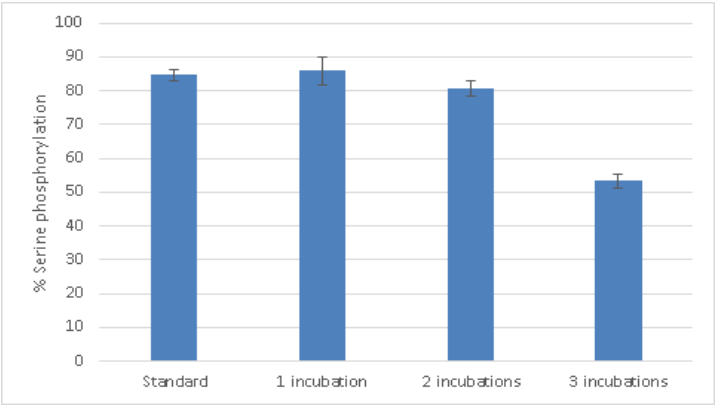


Figure 55. Percent of SYK serine phosphorylation remaining after de-phosphorylating the array with the standard protocol or new protocol using 500 units of CIP for 1 incubation and 2K units of LPP across 1 – 3 incubations. Error bars represent range across duplicate spots.

However, three incubations de-phosphorylated serine on SYK much better than two incubations (Figure 56). Taken together, the most efficient and cost effective de-phosphorylation approach was incubating the slides with 500 units of CIP for 30 min at 30 ºC, then three times with 2K units of LPP at 30 ºC for 30 min each. Compared to the standard protocol, the optimized protocol de-phosphorylated the proteins more effectively than the standard protocol by 45% (Figure 50, Figure 55), was 2 hours long instead of 1.5 hours, and was only 32% of the cost. A follow-up

experiment demonstrated that longer incubation lengths did not improve the level of de-phosphorylation (data not shown).

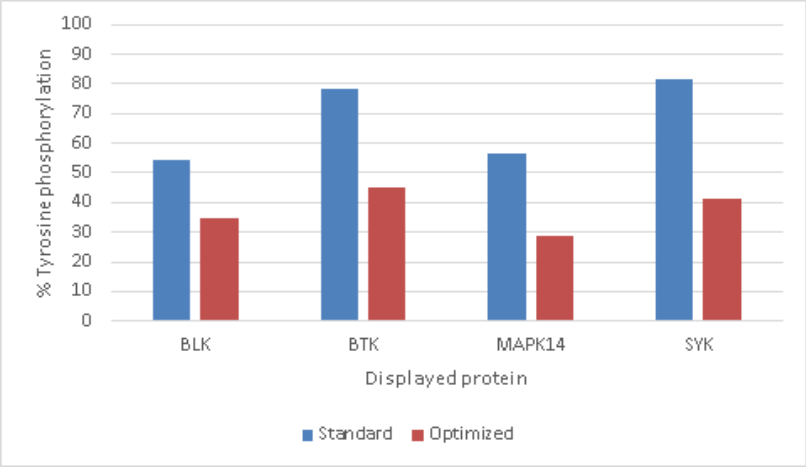


Figure 56. Percent of tyrosine phosphorylation remaining after de-phosphorylating the array with the standard protocol or optimized protocol using 500 units of CIP for 1 incubation and 2K units of LPP across 3 incubations.

6.3 Treatment of NAPPA with B cell lysate

The human cell-free expression system that is used to express the target proteins on NAPPA and NAPPA-SPRi uses lysate from HeLa cells, a cervical cancer cell line, and is well-known to contain kinase activity. While it seems reasonable to assume that the expression system phosphorylates its targets in a HeLa-specific manner, the target phosphorylation profile has never been examined. For this project, however, it was necessary to determine whether the expression system's target phosphorylation profile was unique or similar to B cells given the potential impact that phosphorylation can play in protein interactions and in the BCR signaling pathway. Briefly, the phosphorylation profile differences between the expression system and Ramos B cell lysate was performed using two HaloTag-PL NAPPA slides. The cells were rapidly proliferating (i.e., ~22 hour doubling rate), representing an activated, tonic signaling state. After expression, one slide was de-phosphorylated using the standard LPP-based de-

phosphorylation protocol and incubated in lysate from Ramos B cells that included phosphatase and protease inhibitors, 200 µM ATP, and metal additives to assist in kinase activity at 30 ºC for 1 hour.

More specifically, Ramos RA-1 cells Ramos B cells (ATCC; Manassas, VA) were grown in RPMI-1640 (ATCC; Manassas, VA) supplemented with 10% HyClone™ fetal bovine serum (GE Healthcare Life Sciences; Logan, UT). Cells were washed twice with ice-cold 1 mM Na_3VO_4 in TBS, then solubilized in 50 mM Tris-HCl (pH 7.7), 0.5% nonidet P-40, 2.5 mM EDTA, 20 mM beta-glycerophosphate, 10 mM NaF, 1 mM Na_2MoO_4, 1 mM Na_3VO_4, 0.25 µM PMSF, 1 µM pepstatin, 0.5 ug/mL leupeptin, 10 ug/mL soybean trypsin inhibitor, and 1 ug/mL microcystin-LR. Cells were spun at 4k x g for 5 min and the supernatant stored in single-use aliquots at -80 ºC such that the lysate from 20 million cells were in 1 mL of solubilization buffer. Slides were rinsed with 50 mM HEPES, 150 mM NaCl, pH 7.4. B cell lysate was buffer exchanged using a 7 kDa MWCO Zeba desalting spin column (Thermo Fisher Scientific; Waltham, MA) into kinase buffer containing 20 mM HEPES, 5 mM $MnCl_2$, 5 mM $MgCl_2$, 0.25 µM PMSF, 0.5 ug/mL leupeptin, 10 ug/mL soybean trypsin inhibitor, 20 mM beta-glycerophosphate, 10 mM NaF, 1 mM Na_2MoO_4, 1 mM Na_3VO_4, 500 µM ATP, pH 7.5. Slides were incubated with B cell lysate in kinase buffer for 1 hour at 30 ºC. Slides were rinsed with with 50 mM HEPES, 150 mM NaCl, pH 7.4. The arrays were then probed with an anti-phosphotyrosine antibody and HRP-conjugated secondary antibody to fluorescently assess the level of phosphotyrosine.

Indeed, the profiles between the HeLa expression system and Ramos B cells were different from each other (Figure 58), and indicated that the necessity of phosphorylating the target proteins with B cell lysate for this project. Note that LYN, a tyrosine kinase that is known to be tyrosine phosphorylated in B cells, is phosphorylated on the array that is incubated with B cell lysate. Moreover, VAV1, which is tyrosine-phosphorylated in activated B cells by SYK, is also phosphorylated by the B cell lysate. As mentioned previously, a large-scale screen of serine and threonine phosphorylation is not possible since anti-phosphoserine and anti-phosphothreonine antibodies do not bind to phosphorylated serines and tyrosines in a general manner and cannot be used for large-scale phosphorylation screens.

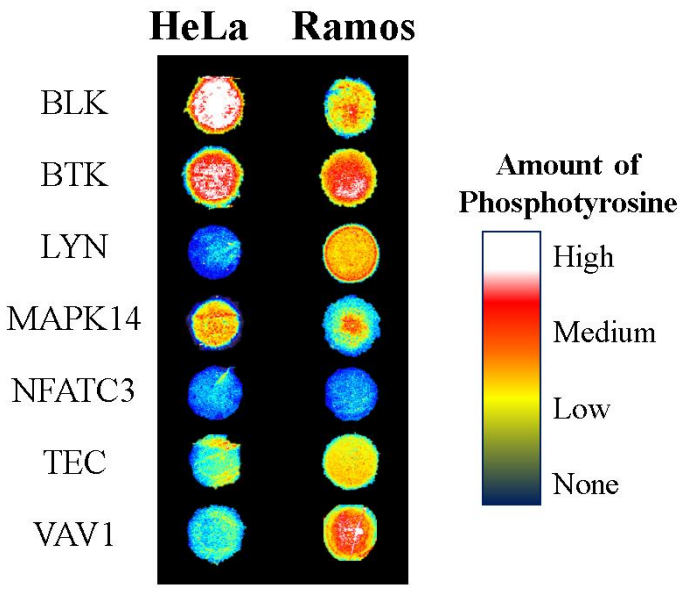


Figure 57. Tyrosine phosphorylation profile of target proteins is different between the HeLa cell-free expression system and Ramose B cell lysate on HaloTag-based NAPPA, as determined via fluorescent analyses using an anti-phosphotyrosine antibody. False-colored rainbow-scale images representing level of phosphotyrosine. Images were analyzed at the same settings.

One of the objectives of this study is to determine the effect of B cell-specific phosphorylation on protein interactions in the B cell receptor signaling pathway. It is therefore necessary to compare the interactions with targets that are un-phosphorylated and phosphorylated targets. Interactions with targets of mixed phosphorylation levels will complicate analyses and will not allow the direct comparison of the datasets. Therefore,

in Chapter 6.2, the targets were de-phosphorylated completely with phosphatases. Here, the incubation length with the B cell lysate to achieve maximum phosphorylation was determined.

The amount of phosphorylation depends on the concentration of the kinase, substrate, and ATP; the incubation length; and the enzyme kinetics (Bertics & Gill, 1985). HaloTag-PL NAPPA slides compatible with SPRi were expressed, de-phosphorylated, and incubated with B cell lysate (20 million Ramos B cells in 1 mL) in the HEPES buffer described above for 0.5, 1, 1.5. 2, 3, 4, and 5 hours at 30 ºC. The slides were the probed with an anti-phosphotyrosine monoclonal antibody and an HRP-conjugated anti-mouse secondary antibody to analyze the level of phosphorylated tyrosines via fluorescence. Figure 59 represents the various responses observed across the array, where the signal appears to plateau at 3 – 4 hours. A decrease in signal at 5 hours

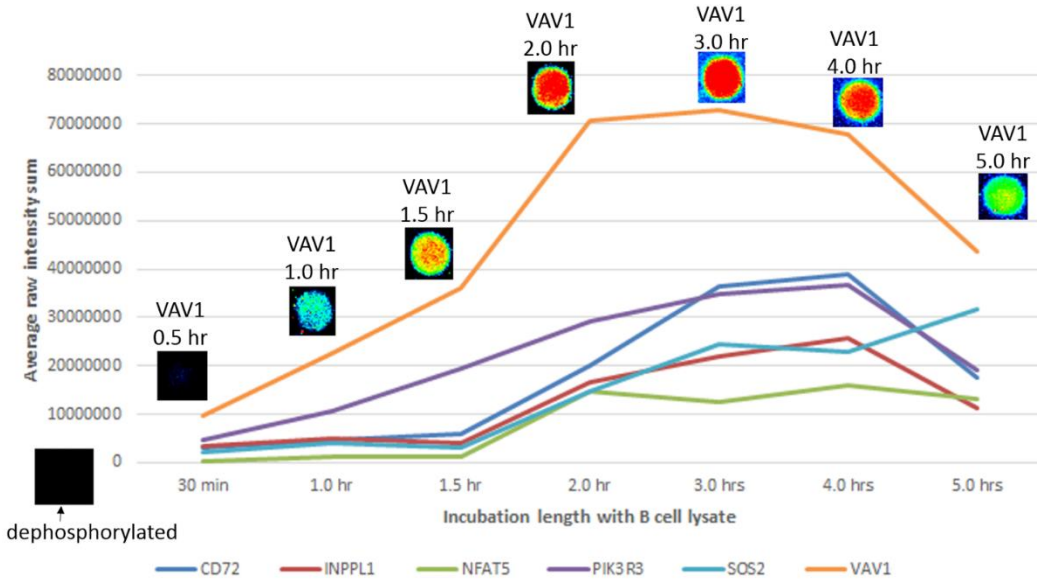


Figure 58. Fluorescent analyses of tyrosine phosphorylation of target proteins incubated with Ramos B cell lysate from 0.5 – 5.0 hours at 30 ºC. Data represents average raw intensity value of duplicate spots referenced to MAP2K2 with no phosphorylated tyrosine response on each array. Fluorescent images of VAV1 are in false-colored rainbow scale where black/blue represents low phosphorylation and red represents high phosphorylation. Images were analyzed at the same settings.

may be due to protein degradation. Based on these results, the length of B cell phosphorylation was set at 3 hours to achieve stable and maximum phosphorylation.

6.4 Acknowledgements

I thank Dr. Fernanda Festa for developing the first protocol to de-phosphorylate NAPPA slides, which I was then able to use as a "jumping off point" in my DOE experiments. Friendly and knowledgeable, Dr. Festa also answered my many questions. The de-phosphorylation and re-phosphorylation of NAPPA slides was made possible by the PS-OC grant and by Dr. LaBaer's discretionary funds.

CHAPTER 7

7 DEVELOPMENT OF HIGH THROUGHPUT SPR SOFTWARE

7.1 Standard SPR data analyses and kinetic models

The "gold standard" software for SPR data analyses is Scrubber2, a program from BioLogic Software (Australia). It can analyze as many as four binding curves at one time; a number not arbitrarily chosen since the most common SPR instrument in general research laboratories is the Biacore T100, an instrument with four channels. Scrubber2 can zero and crop the data, align the injection times, reference (e.g., subtract a non-binder curve from a binder curve), and determine the on- and off-rates of the interactions by fitting the binding curves.

Scrubber2 uses two conventional kinetic models to fit the data: Langmuir and Langmuir with mass transport. The Langmuir adsorption model is useful for simple 1:1 protein interactions; in other words, one epitope on the query interacts with one epitope on the target. Additionally, the Langmuir model is appropriate when the target protein is in a monolayer, the surface is uniform, and the proteins interact independently of neighboring residues (Edwards et al., 1995; Jonsson et al., 1991; Oshannessy, Brighamburke, Soneson, Hensley, & Brooks, 1993; Schasfoort & Tudos, 2008). The

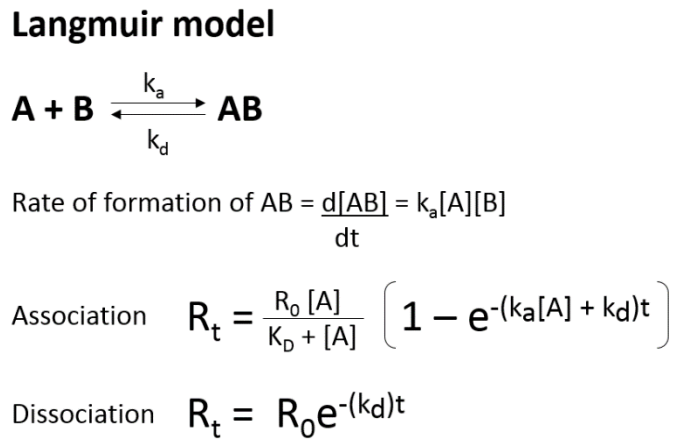


Figure 59. Equations of the Langmuir binding model assuming simple 1:1 protein interaction. A = analyte or query. B = ligand or target protein immobilized on the array. R_t = response at a specific time. R_0 = response at end of association phase.

Langmuir model equation is shown in Figure 60. It reflects a pseudo first-order kinetic reaction where the on-rate is proportional to the concentration of one reactant (i.e, query), which is initially assumed to occur on SPR platforms where the target concentration is fixed while the query concentration is in excess. As detailed in Chapter 4.1.2, mass transport occurs when the target concentration is high and the rate of diffusion is lower than the on-rate (i.e., $k_m << k_a[B]$) (OShannessy & Winzor, 1996; Schasfoort & Tudos, 2008). This results in a linear, rather than an exponential, binding signal.

The Langmuir-based models are generally sufficient for most SPR data, but there are PPIs in which a different kinetic model would better fit the data. A bivalent analyte model is appropriate when a query has two separate binding sites. For example, an antibody can bind to two identical antigens at the end of the arms of its Y-shaped structure via the variant "Fab" region. The second antibody-antigen interaction is dependent on the first interaction due to the proximity of the binding sites. During dissociation, one interaction may break while the other forms, resulting in a dissociation rate that represents the bivalent nature of the complex rather than a single interaction. (To circumvent this avidity issue of antibodies, it is recommended that antibodies are captured as targets rather than injected as queries.) Another example of why a non-Langmuir model should be implemented is when there are two target species or two different binding locations on the target, each capable of binding to the query independent from each other. The heterogeneous analyte or ligand binding models may be used in these situations. A target that changes conformation upon binding the query could be explained by the two-state conformation model. Further detail on how the

equations were derived and the background for the Langmuir and other models are provided in (Edwards et al., 1995; Oshannessy et al., 1993; Schasfoort & Tudos, 2008).

The appropriate model for the data can be obtained by knowing the PPI type *a priori* or by fitting the binding curve to various kinetic models to identify which model best fits the interaction; the Langmuir models, however, fit most data. Another popular software package is BIAevaluation software from Biacore Life Sciences, which includes the Langmuir, bivalent analyte, heterogeneous, and two-state conformation kinetic models. Like Scrubber2, this software was developed to analyze low throughput data only.

7.2 Developing in-house software to analyze Plexera SPRi data

The Plexera® HT PlexArray instrument that was used for the NAPPA-SPRi experiments has a separate Plexera Data Analysis Module software for analyzing the produced data. Unfortunately, it cannot reference the data or fit the data well. More importantly for this project, it could not handle high throughput data; the software would simply freeze indefinitely with the NAPPA-SPRi data. Low throughput, manual software packages like BIAevaluation and Scrubber2 were also not feasible options.

In collaboration with Stanford, the "SPRite" software in Python for fitting the biosensor data from NAPPA-SPRi was built. The SPRite software can do the following: 1) calibrate data; 2) alphabetize the sample names; 3) reference; 4) determine and correct for drift; 5) globally fit data using the 1:1 Langmuir kinetic model; 6) export the binding curves as PDF figures (Figure 61); 7) export a tab delimited text file that can be properly formatted

in SPRuce for Scrubber2 analyses; 8) export a tab delimited text file with calibrated curves (i.e., binder, reference, referenced binder); and 9) export a tab delimited text file with the kinetic and affinity data.

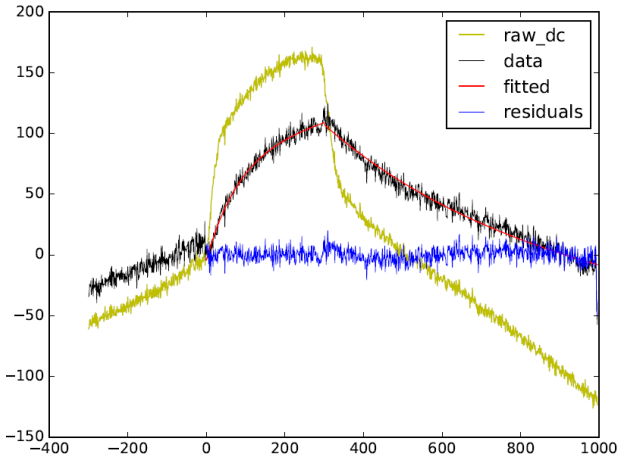


Figure 60. An example of a PDF output file of SPRite depicting the raw binding curve (light green), the referenced binding curve (black), the fitted curve (red), and the residuals between the fitted curve and referenced binding curve (blue). X-axis = time (sec). Y-axis = response units (RU).

To ensure that the binding kinetics and affinities calculated by SPRite are similar to those obtained with Scrubber2, binding curves from seven datasets analyzed with bboth software packages and then compared. More specifically, all of the binding curves from one dataset and several binding curves representing a range of binding rates and affinities from six other datasets were analyzed with SPRite and Scrubber2. These 7 datasets were chosen because they came from multiple different experiments, they had a wide range of binding kinetics and affinities (i.e., k_a = 1.2 x 10^2 to 1.33 x 10^5 $M^{-1}s^{-1}$, kd = 3.74 x 10^{-5} to 7.41 x 10^{-3} s^{-1}, K_D = 8.97 x 10^{-10} to 6.0 x 10^{-5} M), were within the linear range of the instrument, did not have mass transport, the association response had some curvature, and the binding responses followed a single exponential. A comparison of the results from SPRite and Scrubber is displayed as scatter plots for k_a, k_d, and K_D in Figure 62 where the R^2 correlations are 0.992, 0.9974, and 0.9788, respectively. These data demonstrate the accuracy of SPRite. A meta-analysis of the kinetics and affinities determined with Scrubber2 versus SPRite indicates that there were no biases across the range of k_a, k_d, and

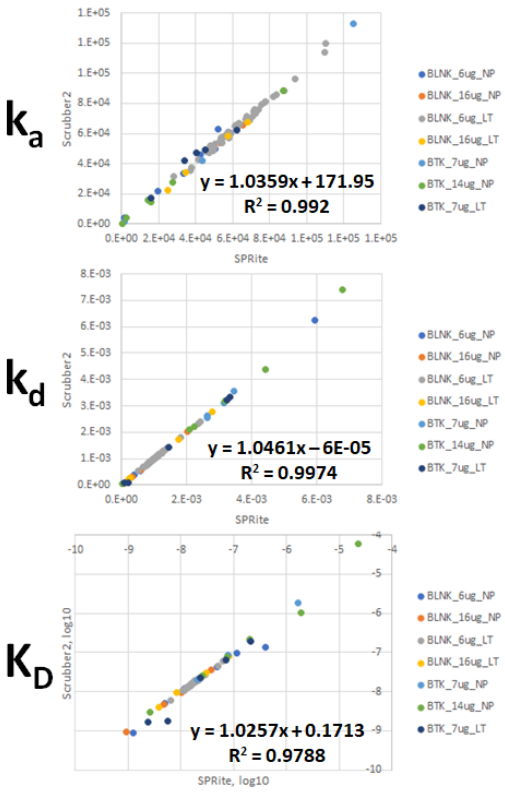


Figure 61. Correlation of k_a, k_d, and K_D values obtained with SPRite and Scrubber2 for seven datasets. NP = targets are Not Phosphorylated, LT = targets are Lysate-Treated

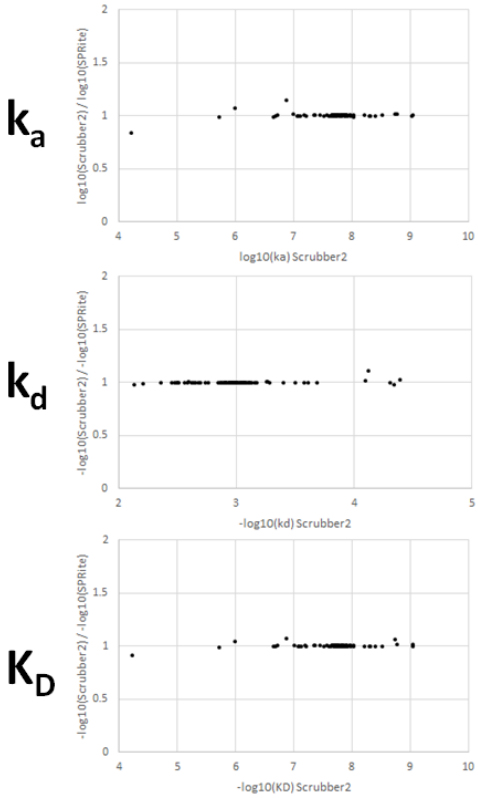


Figure 62. Meta-analysis of the k_a, k_d, and K_D values obtained with Scrubber2 and SPRite indicate that SPRite has no biases across the ranges of values. Values were obtained from the seven datasets in Figure 74.

K_D values (Figure 63). Moreover, SPRite is reproducible, calculating the same values for the same dataset across different analyses, thus resulting in an R^2 correlation of 1 (Figure 64).

SPRite allows the entry of fitting parameters for a group of samples by command line. Once this is done, it will fit ~50 curves against three different references in about 5 minutes on a standard desktop computer. Scrubber2, in comparison, requires manual adjustments from the user during the curve fitting. Thus, in the same 5 minutes, 4 curves can be processed against one reference. This represents almost a 40-fold increase in throughput for curves fit globally. If the binding curves are fit with the Langmuir model

without drift correction and global fitting in SPRite, the throughput increases to 143 curves with three different references in 5 minutes! A comparison of the options in SPRite are Scubber2 is in Figure 65.

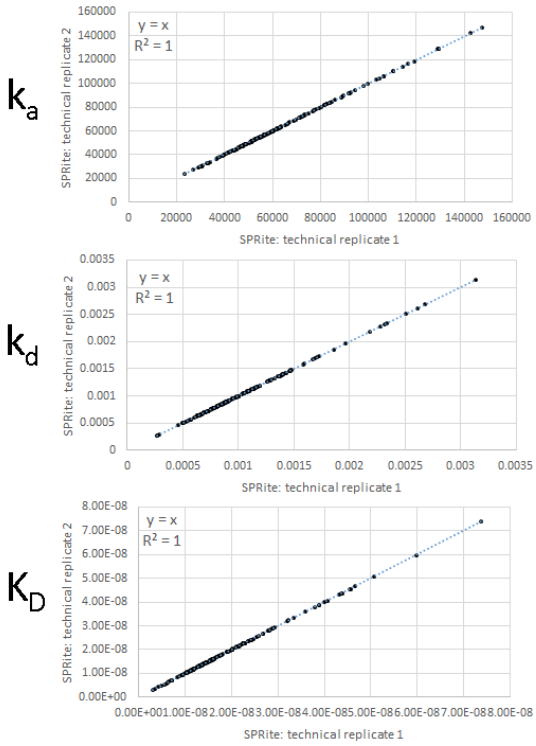


Figure 63. Technical reproducibility of the SPRite software has an R^2 correlation of 1 for k_a, k_d, and K_D

In SPRite, each binding curve is referenced separately to three non-binder curves in SPRite. Thus, instead of one generated sensorgram depicting the raw, referenced, and residual plots in PDF format, THREE PDFs are created per target protein. In other words, an array of 100 target proteins will generate a PDF file with 300 pages. In some cases, not all referenced data may be needed and a short python script, "MergePDFsOnFileNames.py," can be used to extract PDFs with a specific reference-of-interest, and subsequently collated together. The directions for running the "mergePDFsOnFileName.py" script as well as script itself can be found in Appendix H.

SPRite will be open-source, which will make it easy for the software to be added to and improved upon based on the needs of the SPR community. For example, SPRite is currently capable of analyzing curves with the 1:1 Langmuir kinetic model one dataset at a time. It could, however, be altered so that it can analyzed multiple datasets simultaneously within one command prompt window. Moreover, other kinetic models

Options	Scrubber2	SPRite
Automated		Yes
Zeroes Data	Yes	Yes
Calibrates		Yes
Alphabetizes		Yes
References	Yes	Yes
Corrects drift		Yes
Spike Correction	Yes	
Models	1:1 Langmuir 1:1 Langmuir with mass transport	1:1 Langmuir
Sensorgrams exported		Yes
Saves data		Yes
Throughput	1x	39x - 107x*
Analysis-to-analysis correlation	$R^2 = 1$	$R^2 = 1$
Open source software		Yes

* SPRite throughput depends on type of fitting performed

Figure 64. Comparison of Scrubber2 and SPRite software for analyzing SPR data

could be incorporated in SPRite, like the Langmuir model with mass transport or the heterogeneous ligand model. Spikes in the data can occur from physical anomalies during the analyses (e.g., bubble) or imperfect timing alignment (or "offset") between a binder and its reference since the target proteins will experience association and dissociation at slightly different times across the array. Scrubber2 has the ability to identify some of these spikes and remove them from the binding sensorgram; at the moment, SPRite does not (Figure 65). A generic timing "offset" file is presently created by the user (see Appendix E) and fed into SPRite. Since the offset will be different for each target protein, slide, and flow rate, a separate or an embedded script in SPRite to automatically determine the offset values for each experiment would be ideal. Some of this work has already begun.

Two files constitute SPRite, "parseSPRandFitCurves.py" and "curveFittingKineticModels.py." The command line options are first read by "parseSPRandFitCurves.py" which determines the general framework of SPRite.

Software libraries like “pandas” and “numpy” are imported into SPRite via “parseSPRandFitCurves.py” to provide standard data structures and operations for python-based scripts. It also performs calibration, defines phrases, indicates where to go in the “curveFittingKineticModels.py,” and formats the final documents. As the file name suggests, “curveFittingKineticModels.py” contains the equations for fitting the binding curves. It also drift-corrects the data. Directions on using SPRite is given in Appendix E. In Appendix F and Appendix G, the script is provided for the “parseSPRandFitCurves.py” and “curveFittingKineticModels.py” files, respectively. Please note that the appropriate python package must be installed before any of these scripts can be run. SPRite will also be available through the Mallick Lab website at mallicklab.stanford.edu.

7.3 Acknowledgements

The road to SPRite was a long, multi-year process that was often not straight and flat, but riddled with many speedbumps and potholes that required innumerable beta-tests and troubleshooting – and a strong, tenacious collaboration. I *immensely* thank Ian Shoemaker and Ravali Adusumilli (Stanford University) who wrote the SPRuce and SPRite software, respectively. SPRite would not have been possible without Ian, Ravali, Yan Wang, and Dr. Parag Mallick (Stanford University). Dr. Mallick ingeniously named the SPRite software. Funding was provided by Dr. Mallick and Dr. LaBaer.

CHAPTER 8

8 QUANTITATIVE ANALYSES OF THE BCR SIGNALING PATHWAY USING NAPPA-SPRI

8.1 Introduction

Cellular responses to external stimuli are mediated through dynamic and complex signal transduction networks that are comprised of protein-protein interactions (PPIs). Signal propagation is not only dependent on *which* proteins interact, but *how* they interact. For instance, CDC42 and WASP must interact rapidly to stimulate actin polymerization (Hemsath, Dvorsky, Fiegen, Carlier, & Ahmadian, 2005). Fast association rates and slow dissociation rates control antibody maturation and B cell selection (Foote & Milstein, 1991). Mutations that lead to faster association and dissociation rates between Ras and Raf result in more phosphorylated ERK, a downstream product of the Ras-Raf interaction, compared with the wild-type interaction (Kiel & Serrano, 2009). Despite the biological importance of kinetics and affinities, very little of this quantitative space has been explored due to the low throughput nature of current quantitative methods (see Chapter 1.3, page 8). The B cell receptor (BCR) signaling pathway, for example, is considered to be one of the best understood signaling pathways, yet only a handful of these interactions have been quantitatively characterized (Table 2**Error! Reference source not found.**).

I applied NAPPA-SPRi, a high throughput platform that is capable of analyzing > 400 protein interactions quantitatively in less than an hour, toward studying > 12,000 PPIs in the BCR signaling pathway using different query proteins (BLNK, BTK, PI3K,

RAC1, RHOA). Since kinase cascades play an important role in signal transduction in the BCR pathway, I tested these interactions under conditions where the target proteins were either dephosphorylated or treated with lysate from naturally proliferating B cells with active kinases (see Chapters 2.4, 6.2, 6.3). In addition, the GTPase query proteins were tested in different activation states (i.e., GDP- versus GTPγS-bound). Interaction kinetics, affinities, and protein partners were affected by lysate treatment of targets, GTPase query activation state, and the tag location of the target proteins.

8.2 Materials and Methods

ATP was from Cell Signaling Technology (Danvers, MA). GTPγS was obtained from BIOLOG Life Science Institute (Germany). Brij-35; Bond-Breaker TCEP Solution, Neutral pH; NuPAGE Transfer Buffer; and SuperSignal West Femto were from Thermo Fisher Scientific (Waltham, MA). All other reagents, unless otherwise noted, were obtained from Sigma-Aldrich (St. Louis, MO).

Plasmid cDNA

Plasmid cDNA was obtained from the Virginia G. Piper Center for Personalized Diagnostics' (VGP CPD) plasmid repository, DNASU (Tempe, AZ), and Open Biosystems (Lafayette, CO), and prepared as previously described using the pJFT7_nHalo_DC and pJFT7_cHalo_DC with the capturing fusion tag (i.e., HaloTag) at the N- or C-terminus (Saul et al., 2014). The list of these genes is in Appendix A with the

sequences publicly available at https://dnasu.org/DNASU/. Successful cloning of the GOIs was confirmed with Sangar sequencing at DNASU.

NAPPA-SPRi slide preparation

A 48 nm layer of gold was deposited via electron beam evaporation on low sodium optical D263 borosilicate slides with an index of refraction of 1.52 (Plexera LLC; Woodinville, WA). The slides were sonicated for 10 min in 0.1 N KOH, 100% methanol, washed three times in 100% ethanol, and then dried with compressed gas. 1 mM amine-terminated polyethylene glycol [$HS\text{-}C_{11}(C_2H_4O)_6\text{-}NH_2$] (Prochimia Surfaces; Poland) was resuspended in ethanol and applied to the slide overnight at 4 ^oC to create a self-assembled monolayer. To prevent evaporation of the ethanol during the incubation, these slides were placed on upside-down Wheaton® stainless steel 30-slide rack (Capital Scientific, Inc.; Austin, TX) within a plastic Lock & Lock food storage container (Food Storage Mall; China) with ~ 0.5 cm of 100% ethanol on the bottom. The slides were washed three times in 100% ethanol and dried with compressed gas just prior to printing.

The printing master mix included 0.0003% poly-L-lysine (Thermo Fisher Scientific; Waltham, MA), 0.3% DMSO, 250 µM BS3 (Thermo Fisher Scientific; Waltham, MA), 375 µM HaloTag® amine (O4) ligand (Promega; Madison, WI), and 0.4 mg/mL plasmid cDNA. The printing master mixture was incubated at 4 ^oC overnight, then deposited onto the prepared slides with the QArray2 spotter (Molecular Devices, LLC; Sunnyvale, CA) using solid pins. The random array layout and DNA deposition analysis are shown in Appendix J.

The development of this hybrid platform, which combines the advantages of NAPPA and SPRi, is described in more detail in Chapter 4.

Target protein expression

Slides were blocked with Tris-based SuperBlock (Thermo Fisher Scientific; Waltham, MA) to minimize non-specific binding overnight at 4 ^{o}C. They were then washed in 1x PBS three times for 2 min each, rocking. The slides were rinsed in water and dried with compressed air. SPRi flow chambers (Plexera; Woodinville, WA) with 30 µL volume were applied onto the slides followed by 1-step human coupled *in vitro* protein expression mixture according to the manufacturer's instructions (Thermo Fisher Scientific; Waltham, MA). Expression was performed for 1.5 hours at 30 ^{o}C and then 30 min at 15 ^{o}C. Slides were rinsed in 200 µL 1x PBS. Analysis of target protein expression is shown in Appendix J.

Query protein expression

Three purified queries were obtained from commercial companies. BLNK and BTK had an N-terminal His tag (Sino Biological; Beijing, China) and tested for activity through a functional ELISA and kinase assay tests, respectively. BLNK was expressed in human cells while BTK was expressed in baculovirus insect cells. PIK3CA/PIK3R1 (p110α/p85α) (Life Technologies Corporation; Carlsbad, CA) was expressed in baculovirus insect cells, tested for activity using a kinase assay, and had an N-terminal His tag on PIK3CA. The GTPases, RAC1 and RHOA, were expressed in *Escherichia coli*

by members of Dr. Kim Orth's laboratory (UT Southwestern; Dallas, TX), and have been used in various published experiments demonstrating their activity (Woolery et al., 2014; Yarbrough et al., 2009). Moreover, the proteins were tested for functionality by Dr. Xiaobo Yu (personal communication; National Center for Protein Sciences; Bejing, China) by their ability to bind to known protein partners on NAPPA arrays.

De-phosphorylation of target proteins

Calf intestinal alkaline phosphatase (CIAP) (New England BioLabs; Ipswich, MA) and lambda protein phosphatase (New England BioLabs; Ipswich, MA) were buffer exchanged into 1x NEBuffer 3 or 1x NEBuffer for PMP supplemented with 1 mM $MnCl_2$ (New England BioLabs Ipswich, MA), respectively, using 7 kDa molecular weight cut-off (MWCO) Zeba spin desalting columns (Thermo Fisher Scientific; Waltham, MA). Slides were rinsed in 200 µL 1x NEBuffer 3 and then incubated in 300 units of CIAP at 30 ºC for 30 min. Slides were rinsed in 200 µL 1x NEBuffer for PMP supplemented with 1 mM $MnCl_2$. The slides were then incubated three times with 2,000 units of lambda protein phosphatase at 30 ºC for 30 min. Slides were rinsed in the same HEPES- or Tris-based buffer that was used for SPRi analyses (Appendix J).

The de-phosphorylation of the array proteins is optimized and discussed in Chapter 6.2.

Phosphorylation of target proteins with Ramos B cell lysate

Ramos B cells (ATCC; Manassas, VA) were grown in RPMI-1640 (ATCC; Manassas, VA) supplemented with 10% HyClone™ fetal bovine serum (GE Healthcare Life Sciences; Logan, UT). Cells were washed twice with ice-cold 1 mM Na_3VO_4 in TBS, then solubilized in 50 mM Tris-HCl (pH 7.7), 0.5% nonidet P-40, 2.5 mM EDTA, 20 mM beta-glycerophosphate, 10 mM NaF, 1 mM Na_2MoO_4, 1 mM Na_3VO_4, 0.25 µM PMSF, 1 µM pepstatin, 0.5 ug/mL leupeptin, 10 ug/mL soybean trypsin inhibitor, and 1 ug/mL microcystin-LR. Cells were spun at 4k x g for 5 min and the supernatant stored in single-use aliquots at -80 ^oC such that the lysate from 20 million cells were in 1 mL of solubilization buffer. Slides were rinsed in 200 µL 50 mM HEPES, 150 mM NaCl, pH 7.4. B cell lysate was buffer exchanged using a 7 kDa MWCO Zeba desalting spin column (Thermo Fisher Scientific; Waltham, MA) into kinase buffer containing 20 mM HEPES, 5 mM $MnCl_2$, 5 mM $MgCl_2$, 0.25 µM PMSF, 0.5 ug/mL leupeptin, 10 ug/mL soybean trypsin inhibitor, 20 mM beta-glycerophosphate, 10 mM NaF, 1 mM Na_2MoO_4, 1 mM Na_3VO_4, 500 µM ATP, pH 7.5. Slides were incubated with B cell lysate in kinase buffer for 3 hours at 30 ^oC. Slides were rinsed with the same HEPES- or Tris-based buffer that was used for SPRi analyses. All slides were stored at 4 ^oC when not being analyzed.

The phosphorylation of the array proteins in a B cell-specific manner is optimized and discussed in Chapter 6.3.

Activation of GTPases, RAC1 and RHOA

Purified GTPases were incubated in 10 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM EDTA, 1 mM TCEP, pH 7.4, and 1 mM GTPγS or GDP for 1 hour at room temperature to activate or inactivate the GTPases, respectively. Samples were then buffer exchanged into 50 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, pH 7.4, which was the buffer used for SPRi analyses.

NAPPA-SPRi analyses

The Plexera HT PlexArray instrument was primed three times with filtered and degassed "running buffer" specific to the query (Appendix J). Each slide was subjected to the following runs, in consecutive order: 0.5% glycerol, 1.0% glycerol, running buffer, protein query, and 5.34 E-8 M anti-TP53 D01 monoclonal antibody. Glycerol in running buffer was injected with 100 sec association and 100 sec dissociation each and used to normalize inter- and intra-slide data where the change in refractive index is equal to 0.000565 response units (RU). Kinase query runs were performed at 5 µL/sec at 30 ºC. Running buffer, purified query protein in running buffer, and antibody in running buffer were injected with 180 sec association and 400 sec dissociation. Non-kinase query runs were performed at 3 µL/sec at RT. Running buffer, purified query protein in running buffer, and antibody in running buffer were injected with 300 sec association and 700 sec dissociation. Data were acquired in real-time with the Plexera Instrument Control software.

NAPPA-SPRi data analyses

Data analyses were performed in three steps. Regions-of-interest in the AVI video format were first identified and analyzed with the Plexera SPR Data Analysis Module software. The in-house software, SPRite, calibrated the data to standard response units (RU), formatted the data to be compatible with Scrubber2, selected the time frame(s) of interest, referenced the binding curves to non-binders, drift corrected the data, and fit the curves using Langmuir kinetic models (with and without mass transport). Finally, the curves were assessed by eye. More information about the SPRite software is in Chapter 7.2. Specific information regarding SPRi analyses is in Appendix I.

Qualitative analyses of protein interactions using NanoBRET

See Chapter 3.4. Only the protein interactions with BLNK, BTK, PI3K, GDP-bound RAC1, GTP-bound RAC1, GDP-bound RHOA, and GTP-bound RHOA are considered in this chapter.

Protein interaction validation using Phos-Tag SDS-PAGE

Recombinant human BTK with an N-terminal His tag was obtained from Sino Biological (Wayne, PA). Recombinant human ETS1, JUN, and BCL2 with an N-terminal His tag were obtained from RayBiotech (Norcross, GA), respectively. Recombinant human protein PI3K, constituting PIK3CA with an N-terminal His tag and untagged PIK3R1, was obtained from Thermo Fisher Scientific (Waltham, MA). Recombinant human MYC with an eleven-arginine tag at the C-terminus was obtained from Abcam (Cambridge, MA). MYC was first de-phosphorylated with 1200 units of lambda phosphatase for 2 hr

at 30 ºC, and then 2 mM sodium orthovanadate was added to inhibit any further phosphatase activity. Kinase and substrate were mixed together at a 3:4 ratio (w/w) in 50 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM MgCl2, 0.01% Brij-35, 1 mM ATP, and incubated at 30 ºC for 1 hr. BTK samples with substrate and ATP were de-phosphorylated with 800 units of lambda protein phosphatase (New England BioLabs; Ipswich, MA) for 2 hrs at 30 ºC. 1x Laemmli loading dye (Bio-Rad; Hercules, CA) and 10 mM Bond-Breaker TCEP Solution, Neutral pH were added to the samples before heating at 65 ºC for 10 min. Samples were added to SuperSep Phos-Tag™ (50 umol/L), 12.5% SDS-PAGE gels (Wako Pure Chemical Industries; Richmond, VA) using tris-tricine running buffer containing 50 mM Tris, 50 mM N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl] glycine (Tricine), 0.10% (w/v) SDS, 5 mM sodium bisulfite. Gels were run at 100 V for 1 hr, and then transferred overnight to a PVDF membrane at 4 ºC and 150 mA using the Bio-Rad Mini Trans-blot cell and 1x NuPAGE transfer buffer supplemented with 5 mM sodium bisulfite. Membranes were blocked with 3% BSA in PBST ("blocking buffer") for 1 hr at room temperature, and then probed with rabbit anti-human anti-c-Jun monoclonal antibody (clone 60A8; Cell Signaling Technology; Danvers, MA), anti-ETS1 monoclonal antibody (clone D808A; Cell Signaling Technology; Danvers, MA), anti-c-MYC monoclonal antibody (Abcam; Cambridge, MA), anti-c-MYC phospho S62 monoclonal antibody (Cell Signaling Technology; Danvers, MA), or mouse anti-human BCL2 monoclonal antibody (Cell Signaling Technology; Danvers, MA). The membrane was washed three times in PBST, incubated for 1 hr at room temperature with HRP-conjugated anti-rabbit IgG (Cell Signaling

Technology; Danvers, MA) or anti-mouse IgG (Jackson ImmunoResearch; West Grove, PA) at a 1:15,000 dilution in blocking buffer, and then washed again three times in PBST. Signal was visualized using SuperSignal West Femto chemiluminescent substrate using the ImageQuant LAS 4000 system (GE Healthcare Life Science; Pittsburgh, PA).

8.3 Results & Discussion

8.3.1 NAPPA-SPRi detected known and novel interactions

To evaluate how many known interactions were detected with NAPPA-SPRi, the data were compared to human, mouse, and rat PPIs curated by the online databases, BioGRID and HPRD (Appendix C) (Prasad et al., 2009; Stark et al., 2006). Of the NAPPA-SPRi PPIs that were detected, 15% were known and 85% were novel (Table 14). Across the seven queries, NAPPA-SPRi detected 72 (66%) known interactions listed in these sources while also identifying 401 previously unreported interactions (Table 14, Table 15).

Table 14. Percentage of known and novel PPIs detected by NAPPA-SPRi

	BLNK		BTK		PI3K		RAC1(GDP)		RAC1(GTP)		RHOA(GDP)		RHOA(GTP)	
Detected PPIs	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT
All	54	53	78	81	34	16	59	67	95	78	35	5	91	72
% Known*	15	17	14	12	32	31	14	13	13	15	14	20	14	14
% Novel	85	83	86	88	68	69	86	87	87	85	86	80	86	86

* Known human, mouse, or rat PPIs in the online databases, BioGRID and HPRD

Table 15. NAPPA-SPRi detected 66% of known PPIs

	BLNK			BTK			PI3K			RAC1(GDP)			RAC1(GTP)			RHOA(GDP)			RHOA(GTP)		
Type of PPI	All	NP	LT	All	NP	LT	All	NP	LT	All	NP	LT	All	NP	LT	All	NP	LT	All	NP	LT
Known PPI*	14	14	14	15	15	15	26	26	26	14	14	14	14	14	14	13	13	13	13	13	13
Known PPI detected	10	8	9	11	11	10	11	11	5	10	8	9	12	12	12	5	5	1	13	13	10
% Known PPI detected	71	57	64	73	73	67	42	42	19	71	57	64	86	86	86	38	38	8	100	100	77

* Known human, mouse, or rat PPIs in the online databases, BioGRID and HPRD

The PI3K query had one of the lowest coverage of known interactions (i.e., 42%) with NAPPA-SPRi (Table 15). In *vivo*, PI3K can exist as a heterodimer consisting of one of five different regulatory subunits and one of four different catalytic subunits (Cheung

et al., 2015). Unlike the catalytic subunits, the regulatory subunits can act as a monomer or homodimer. In this experiment, the PI3K query was a heterodimer, containing the regulatory subunit PIK3R1 and the catalytic subunit PIK3CA. Its interactions on NAPPA-SPRi were cross-referenced with those that are known for PIK3R1, PIK3CA, and the heterodimer. Most of the PI3K interactions curated by BioGRID and HPRD were for the PIK3R1 subunit, which may or may not have been also been interacting with PIK3CA or one of the other catalytic isoforms. This may explain why the known PPI coverage of PI3K was low. The GTPases, RAC1 and RHOA, identified more known proteins when they were GTP-bound (i.e., active) than when they were GDP-bound because they interacted with > 20% more proteins than the inactive GTPases.

8.3.2 Tag locations may provide helpful insight into binding sites

NAPPA-SPRi employed fusion tags to capture the expressed target protein to the slide surface. Since fusion tags may interfere with interactions by blocking binding epitopes or altering protein structure, particularly at the end of the protein where they are located, targets were represented separately on the array with the tag at the N-terminus and C-terminus. We observed 62.3% more interactions when the tag was at the C-terminus. Of the 101 target proteins represented on the array with a tag in each position, and which interacted $\geq$ 3 queries across all conditions and concentrations, 69 showed a preference of at least two-fold for a tag in one position or the other. Twenty (20) proteins had more interactions with the queries when the tag was at the N-terminus, while 49 proteins had more interactions when the tag was at the C-terminus. Some isoforms within

the same protein family behaved similarly, which may account for some interaction bias in this study toward C-terminally tagged targets. For example, most of the interactions with targets in the SHIP (INPP5D, INPPL1), p38 (MAPK12, MAPK13, MAPK14), ERK (MAPK1, MAPK3), NFATC (NFATC1, NFATC3, NFATC4), and PKC (PRKCA, PRKCB) protein families occurred when the tag was the C-terminus. Other isoforms within the same protein family behaved differently from each other. For example, AKT1 and AKT2 interacted with at least three times as many queries when the tag was at the C-terminus than at the N-terminus; AKT3 had no preference, with N-terminal tagged AKT3 binding to queries 22 times and C-terminal tagged AKT3 binding to queries 24 times. All four PI3K catalytic isoforms had different binding profiles with respect to tag location. PIK3CA had no tag preference, PIK3CB interacted with 4-fold more queries when it had a tag at the C-terminus, PIK3CD did not interact with any queries, and PIK3CG interacted with 2.3-fold more queries when it had a tag at the N-terminus. These dissimilarities may be the result of different protein structures despite having high sequence homology.

Protein interactions that are affected by the location of the tag may provide helpful insight into *where* a query may be binding. For example, an N-terminal tag is more likely to occlude binding sites toward the N-terminus. NAPPA-SPRi detected the known interaction between the adaptor proteins, GRB2 and BLNK. The interaction was altered with lysate treatment, which supports the current understanding of how GRB2 binds to BLNK in unstimulated and stimulated B cells. BLNK only bound to NP- GRB2 with a tag at the C-terminus, implying that BLNK is binding to GRB2 toward its N-

terminus (Figure 66). In fact, BLNK binds to GRB2's N-terminal SH3 domain in unstimulated B cells where GRB2 is in a homodimer. Following lysate treatment of GRB2 on NAPPA-SPRi, BLNK interacted with GRB2 regardless of the location of the fusion tag. These data suggest that GRB2 is indeed phosphorylated by the B cell lysate, resulting in a conformational or electrostatic change that favors a GRB2-BLNK interaction. Indeed, phosphorylation at tyrosine 160 of GRB2 in stimulated B cells is reported to destabilize GRB2's homodimer complex, thereby allowing BLNK to bind to the C-terminal SH3 domain (Ahmed et al., 2013; Justement & Siminovitch, 2000; S. G. Li, Couvillon, Brasher, & Van Etten, 2001; Riera et al., 2010; Wienands et al., 1998). BLNK also binds to GRB2's central SH2 domain in activated B cells. In the lysate-treated data, either the C-terminal tag does not interfere with BLNK binding to GRB2's C-terminal SH3 domain or BLNK may be binding to GRB2's central SH2 domain.

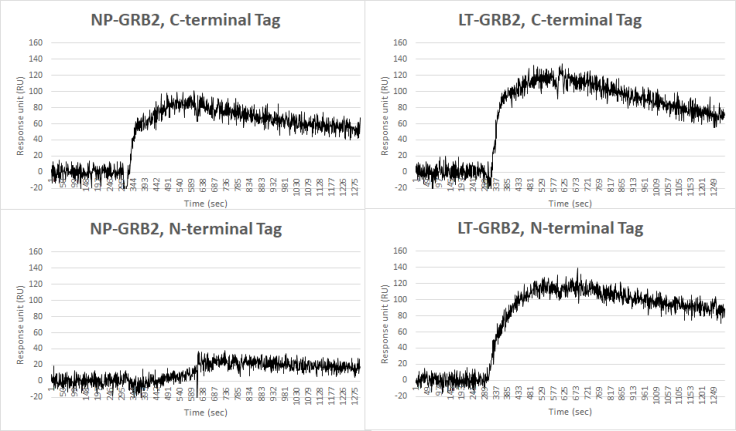


Figure 65. Binding sensorgrams of the query, BLNK, binding to NP- and LP-GRB2.

8.3.3 Phosphorylation affects binding partners

Phosphorylation is a common posttranslational modification (PTM) in which a phosphate is covalently bound to a serine, threonine, or tyrosine residue by kinases and

removed by phosphatases. It is estimated that one-third of all proteins within a cell are phosphorylated at any given time, playing important roles in intracellular signaling and metabolic control (Kitchen, Saunders, & Warwicker, 2008). Phosphorylation can affect PPIs by altering protein structure, blocking binding sites, creating new binding epitopes, or causing bulk electrostatic changes that are sensitive to the subcellular location (Nishi, Shaytan, & Panchenko, 2014; Serber & Ferrell, 2007). As such, phosphorylation has been traditionally viewed as a posttranslational modification that promotes or inhibits interactions, or activates or inactivates enzymes. Phosphosites – particularly those with serine and threonine – are often in regions that are flexible and disordered under native conditions. Since the phosphate group is dianionic at physiological pH, it can form extensive hydrogen bonds and salt bridges with neighboring residues (Nishi, Hashimoto, & Panchenko, 2011). As such, phosphorylation is generally considered to induce protein disorder-to-order conformational changes in the region around the phosphosite; order-to-disorder transitions may occur elsewhere in the protein as a consequence of phosphorylation.

Phosphorylation plays an important role in regulating the BCR pathway. For this reason, PPIs were studied using target proteins that were either de-phosphorylated (i.e., not phosphorylated; NP) or "phosphorylated" using lysate from activated B cells supplemented with kinase co-factors and phosphatase inhibitors (i.e., lysate-treated; LT). Fluorescence analyses using anti-phosphotyrosine antibodies to arrays before and after lysate treatment demonstrate that the lysate contains active kinases since proteins are phosphorylated (see Chapter 6.3, Figure 58). Although lysate treatment represents the

Table 16. Unique PPIs based on target phosphorylation

Query protein	BLNK		BTK		PI3K		RAC1(GDP)		RAC1(GTP)		RHOA(GDP)		RHOA(GTP)	
Target Phosphorylation	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT
	EGR1	CD79B	NFATC1	CDC42	AKT2*	BLK	NFKBIE	AKT2*	AKT2**		AKT2*		BCL10	GRAP2
	LYN	GSK3B	PRKCA	ETS1	CD22		TEC	BCL2A1	BCL2		AKT2**		BCL2	NFAT5
	MAPK14	IKBKA	VAV3	LILRB3	CD79A			EGR1	BLK		BCL2L1		CD79B	PIK3R3
	MAPK3	NFATC4		MAP2K2	HRAS			FOS	CD72		CARD11		CD81	RAC1
	NFKB1	SYK		RAC3	IKBKA			MAPK9	CD79A		CD72		CDC42	
	PIK3CG	VAV1		SOS2	IKBKB##			NFKB1	CDC42		ETS1		CDKN2A	
	RAC1				IKBKG			PPP3CB	FCGR2B		GRB2		DAPP1	
					INPPL1			RAC1	GRAP2		IKBKB#		EGR1	
					KRAS			SOS1	INPP5D		IKBKB##		GSK3B	
					MAP3K3			SOS2	MALT1		IKBKG		HRAS	
					MAPK12				MAPK8		INPP5D		KRAS	
					MAPK13				NFAT5		LAT2		LYN	
					MDM2				NFATC3		LILRB3		MAPK12	
					NCKAP1L				PIK3CB		MAP2K2		NFATC4	
					PLCG2				PIK3R2		MAP2K3		NFKBIA	
					PPP3CB				PIK3R5		MAP3K3		NRAS	
					RAC1				RASSF5		MAPK1		PIK3AP1	
					VAV2						MDM2		PTEN	
					VAV3						MYC		PTPN6	
											NCKAP1L		RAF1	
											NFATC3		RasGRP3	
											NFATC4		SOS1	
											NFKB1		VAV3	
											PIK3CA			
											PIK3CB			
											PIK3R1~			
											PPP3CB			
											PPP3R1			
											PTPN6			
											RHOA			

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Known interactions that are in the online databases, BioGRID or HPRD

"phosphorylation" arm of the experiment, it should be noted that certain target proteins may not be phosphorylated by the lysate and that the sites of the phosphorylation, if they do occur, are unknown. That said, as described below, I found significant differences in affinity and binding rates between the de-phosphorylated and lysate-treated proteins, thus I inferred that the main effect of the lysate treatment was to modify phosphorylation.

For the most part, treatment with lysate did not alter *which* proteins interacted; that is, 84% of the targets showed detectable binding to the same queries regardless of phosphorylation state. However, in many cases, lysate treatment had a profound effect on either binding affinity, interaction rates or both. Those interactions that occurred in one but not the other target phosphorylation state are displayed in Table 16. Notably, some query proteins bound to $\geq$ 20% more NP- than LT-targets (Table 17). These included

Table 17. Query interactions with NP- and LT-targets

	Query protein						
	BLNK	BTK	PI3K	RAC1(GDP)	RAC1(GTP)	RHOA(GDP)	RHOA(GTP)
# PPIs with NP-targets	54	78	34	59	95	35	91
# PPIs with LT-targets	53	81	16	67	78	5	72
Overlap: PPIs with NP- and LT-targets	47	75	15	57	78	5	68
% increase in PPIs with NP-targets	2%		113%		22%	600%	26%
% increase in PPIs with LT-targets		4%		14%			

PI3K, active RAC1, inactive RHOA, and active RHOA (for Venn diagrams, refer to Appendix M).

8.3.4 Phosphorylation affects binding kinetics

Protein interactions that occurred regardless of phosphorylation state were compared with each other to determine the relative effect of lysate treatment on binding kinetics and binding affinity. These results are depicted as bar plots in Figure 67, Figure 68, Figure 69, Figure 71, Figure 70, and Figure 72. On-rates, off-rates, and binding affinities for each protein interaction in the unphosphorylated dataset were set to "0" and represented as blue circles. The relative changes in log10 in binding kinetics and affinity for each protein pair following lysate treatment were represented as a connected orange circle. For example, in Figure 67, in its interaction with BLNK, VAV2 showed an increase of 0.7 log10 in its k_d (i.e., dissociation), almost no change in k_a (i.e., association), and a decrease of 0.7 log10 in the K_D (i.e., binding affinitiy).

BLNK and PI3K interactions are regulated primarily through their off-rate. Approximately half of the targets showed higher binding affinity to the adaptor protein BLNK after they were lysate-treated. In these examples, this change was associated with small increases in on-rates or decreases in off-rates (Figure 67, Appendix Q). The other

half of the BLNK interactions had lower binding affinities with LT-targets, a result largely associated with faster off-rates, with little or no change in the on-rate. Lysate treatment generally resulted in stronger binding affinities with PI3K, which were associated with slower off-rates (Figure 68).

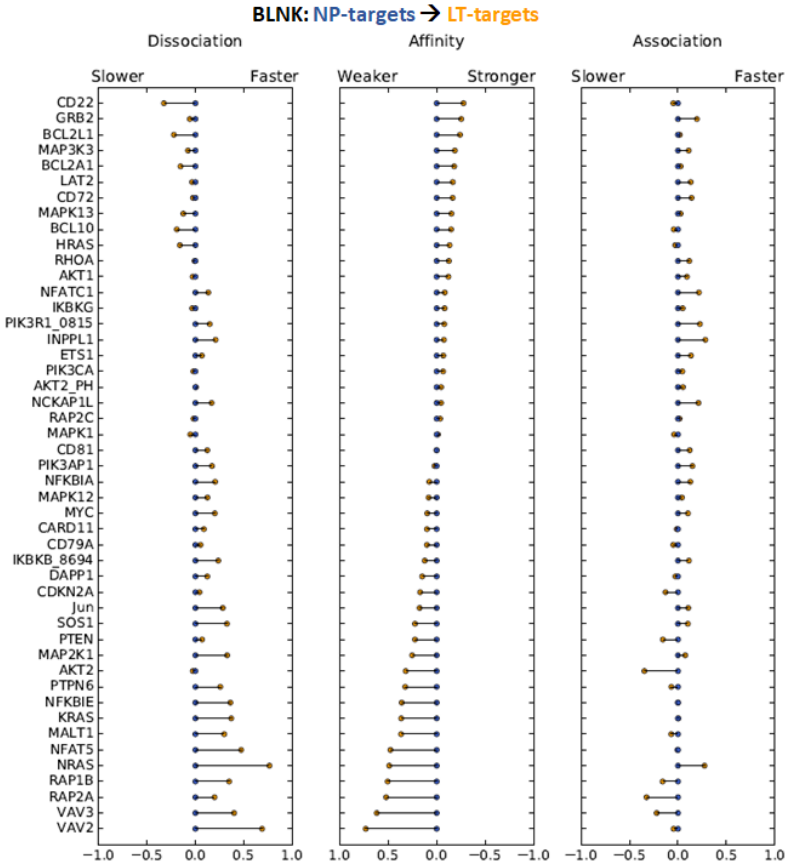


Figure 66. Bar plots showing the relative log10 change in k_d, K_D, and k_a with LT-targets compared to NP-targets for the BLNK query.

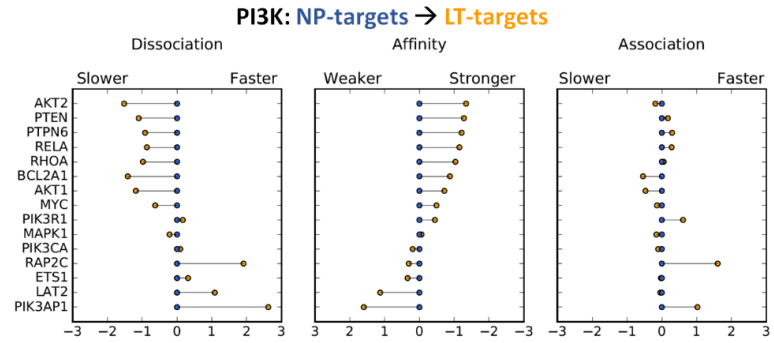


Figure 67. Bar plots showing the relative log10 change in k_d, K_D, and k_a of all of the PPIs of PI3K with NP- and LT-targets.

BTK interactions are regulated through their on- *or* off-rates. For the non-receptor tyrosine kinase, BTK, stronger affinities with LT-targets compared to NP-targets were associated with slower dissociation rates, whereas weaker affinities were associated with slower association rates (Figure 69, Appendix Q).

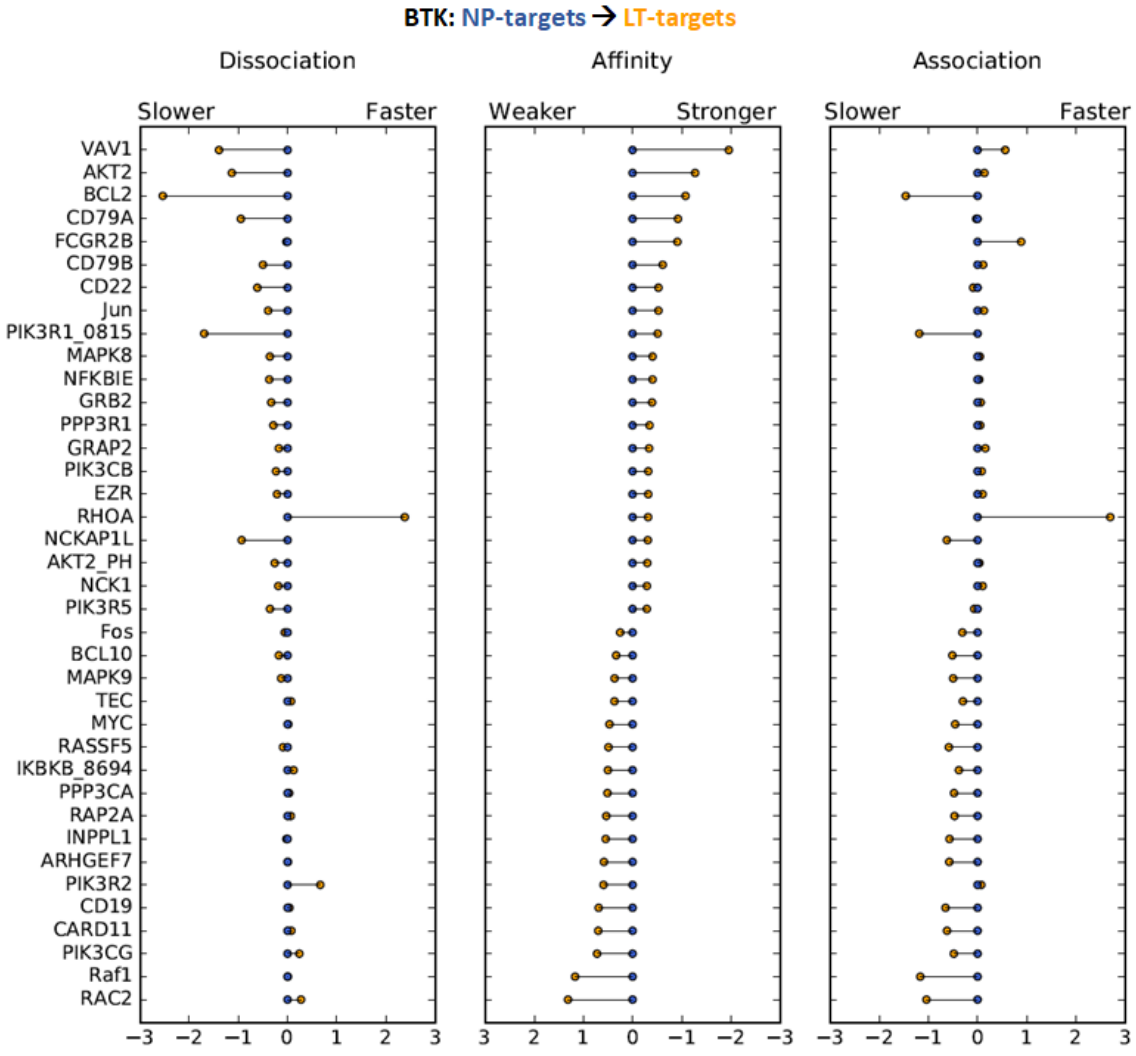


Figure 68. Bar plots showing the relative log10 change in k_d, K_D, and k_a with LT-targets compared to NP-targets for the BTK query. Bar plots showing all of the PPIs is in Appendix P.

RAC1 interactions are regulated through their on- *and* off-rates. Among the most surprising findings of this study relates to RAC1 binding to targets. GTP-bound RAC1, showed significantly faster on- and off-rates to lysate-treated targets compared to their dephosphorylated counterparts (Figure 70, Appendix Q); however, despite binding rate changes that sometimes exceeded several orders of magnitude, the overall affinity (K_D) was largely unchanged. Thus, it appears that phosphorylation of some targets results in a dramatic form of regulation of binding rates without a significant effect on the fraction of molecules bound. Such an effect has never been previously reported. The LT → NP transition increased the average on-rates and off-rates of active RAC1 by 220- and 257-fold, respectively, with only a 1.48 change in affinity. In other words, the on- and off-rates increased or decreased proportionally with relatively little change in binding affinity. In contrast, both the binding kinetics and affinities of GDP-bound ("inactive") RAC1 were minimally affected with lysate treatment (Figure 71). Overall, the on-rates, off-rates, and binding affinities of inactive RAC1 increased 2-fold, decreased 2-

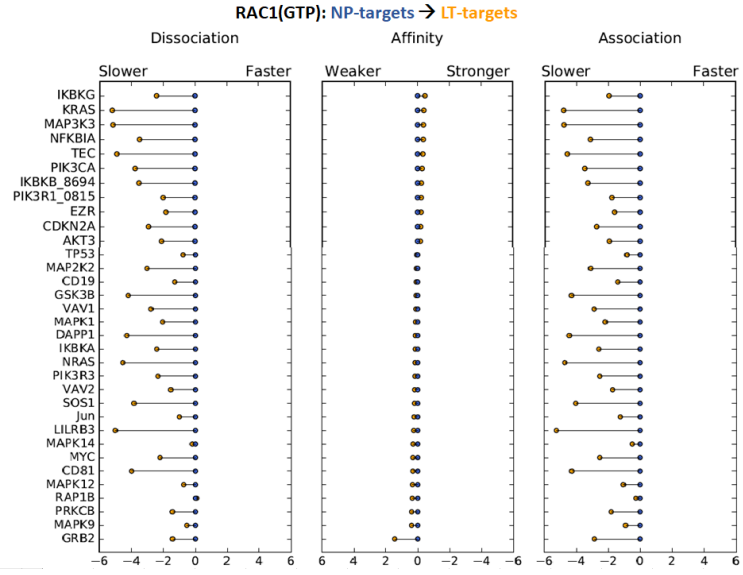


Figure 69. Bar plots showing the relative log10 change in kd, KD, and ka of some of the PPIs of the RAC1(GTP) query with NP- and LT-targets. Bar plots showing all of the PPIs is in Appendix P.

fold, and increased 1.5-fold, respectively, compared to the binding kinetics and affinities with de-phosphorylated targets.

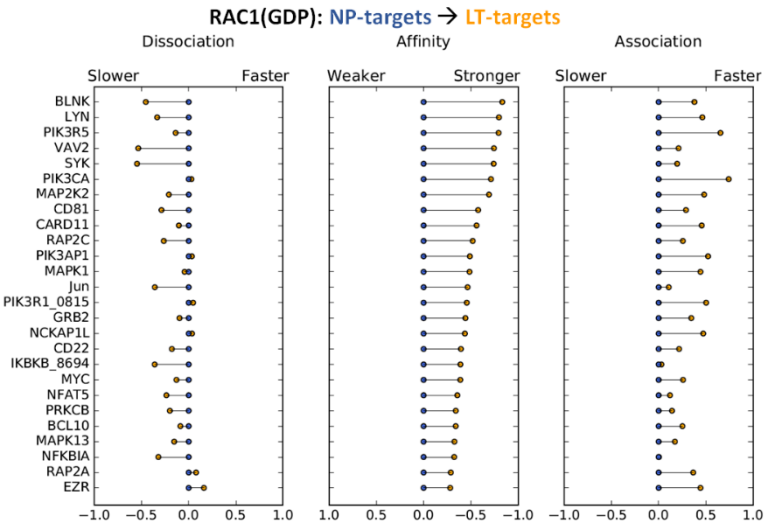


Figure 70. Bar plots showing the relative log10 change in k_d, K_D, and k_a of some of the PPIs of the RAC1(GDP) query with NP- and LT-targets. Bar plots showing all of the PPIs is in Appendix P.

RHOA interactions are regulated through minimal changes in their on- *or* off-rates. The majority of LT- targets bound to active RHOA with lower binding affinities than their NP counterparts, which was associated, in large part, to slower on-rates (Figure 72). However, roughly a quarter of the interactions resulted in stronger affinities, which were associated with slower off-rates. Inactive RHOA interacted with only five targets that were unphosphorylated and lysate-treated, with no overall differences in binding kinetics and affinities between de-phosphorylated and LT-targets (Appendix P). Taken together, the different kinetic

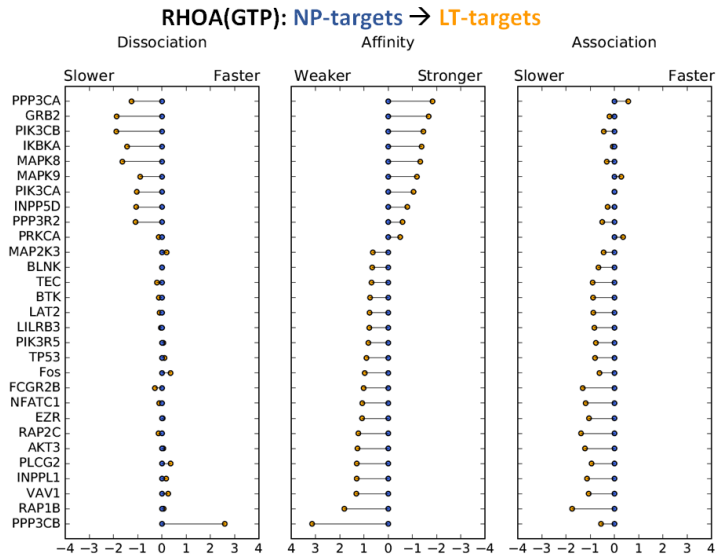


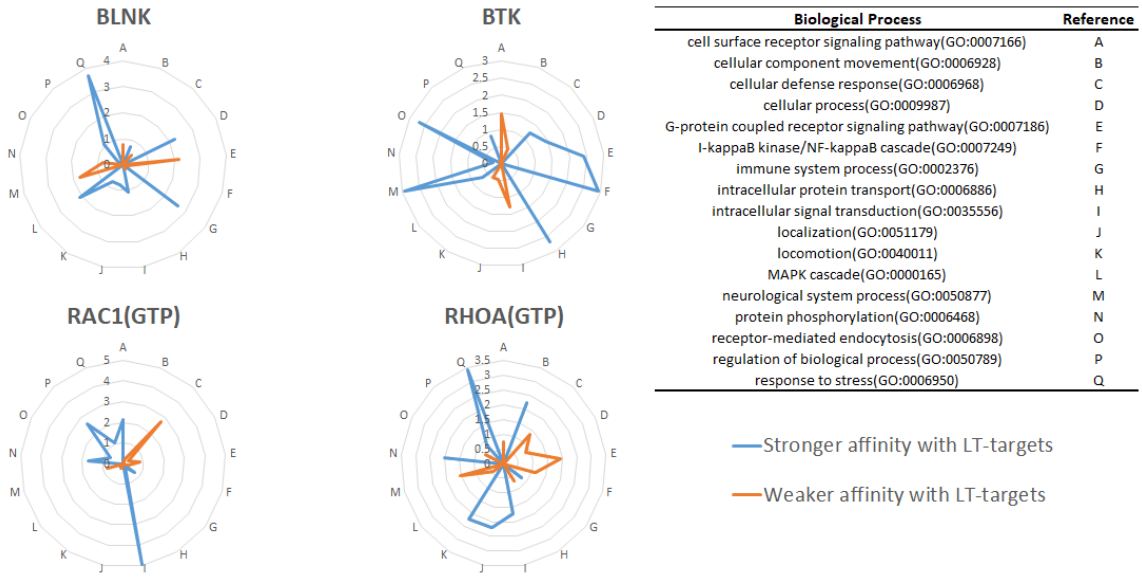
Figure 71. Bar plots showing the relative log10 change in k_d, K_D, and k_a of some of the PPIs of the RHOA(GTP) query with NP- and LT-targets. A bar plot showing all of the PPIs is in Appendix P.

profiles illustrate that proteins employ different methods of regulation in their interactions with other proteins. They also indicate that the interactions are not an artefact from the NAPPA-SPRi platform.

8.3.5 Pairwise analyses of low and high binding affinities

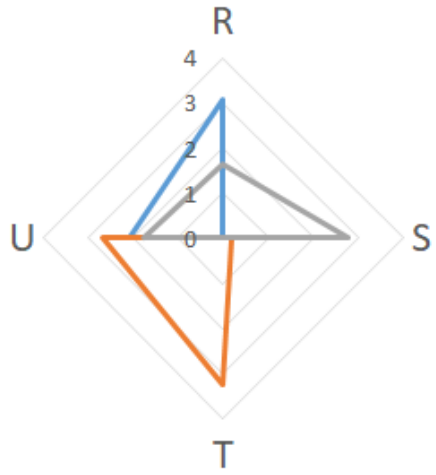
To understand whether there was a difference of interacting targets with particular biological processes and gene families that had low or high binding affinities to a specific query, targets were first defined using the Protein Analysis Through Evolutionary Relationships (PANTHER) and HUGO Gene Nomenclature Committee (HGNC) classification systems, respectively. The number of targets with different biological process or gene family were then determined per query and a pairwise linear regression analysis was performed comparing the targets with low binding affinity to those with high binding affinity to the same query (Figure 73, Figure 74). Enriched processes and gene families were defined as having more than two standardized residuals away from the predicted mean of the fitted linear regression line. Residual plots of these analyses are in Appendix N).

BLNK Targets. Target proteins of BLNK having $\geq 3 \times 10^{-9}$ M difference in binding affinities following lysate treatment were compared in terms of their biological processes and gene families using pairwise linear regression analyses. The biological process that was most enriched in targets with stronger affinities to BLNK following lysate treatment was the stress response. BLNK has already been demonstrated to be important in the stress response, mediating protein interactions for the PI3K/AKT and



Biological Process	Reference
cell surface receptor signaling pathway(GO:0007166)	A
cellular component movement(GO:0006928)	B
cellular defense response(GO:0006968)	C
cellular process(GO:0009987)	D
G-protein coupled receptor signaling pathway(GO:0007186)	E
I-kappaB kinase/NF-kappaB cascade(GO:0007249)	F
immune system process(GO:0002376)	G
intracellular protein transport(GO:0006886)	H
intracellular signal transduction(GO:0035556)	I
localization(GO:0051179)	J
locomotion(GO:0040011)	K
MAPK cascade(GO:0000165)	L
neurological system process(GO:0050877)	M
protein phosphorylation(GO:0006468)	N
receptor-mediated endocytosis(GO:0006898)	O
regulation of biological process(GO:0050789)	P
response to stress(GO:0006950)	Q

Figure 72. Radial plots of enriched PANTHER biological processes in PPIs that have stronger binding affinities following lysate treatment. Numbers represent the standard deviation away from the mean.



Gene Family	Reference
Mitogen-activated protein kinases	R
Phosphatidylinositol 3-kinase subunits	S
RAS type GTPases	T
SH2 domain containing	U

Figure 73. Radial plot of enriched HGNC gene families in PPIs that have stronger binding affinities following lysate treatment. Numbers represent the standard deviation away from the mean.

JNK signaling pathways that regulate cell survival and apoptosis, respectively (Ding et al., 2000; Han et al., 2001; Kabak et al., 2002). Stress also increases tyrosine phosphorylation of numerous proteins that are targeted by SH2 domains (Suzuki, Ohsugi, & Ono, 1996). Interestingly, the SH2 domain containing gene family was enriched in this group as well.

BTK Targets. Twenty targets of BTK with the highest binding affinities and slower dissociation constants were compared against the 29 proteins with the lowest affinities and slower association rates following lysate treatment. One of the gene families that were enriched in targets with stronger affinities following lysate treatment was RAS type GTPases. Interestingly, BTK resembles some Ras GTPase-activating proteins (GAPs) through its PH domain and BTK motif (~150 amino acids); GAPs bind to active Ras GTPases and accelerate GTP hydrolysis (Grewal, Koese, Tebar, & Enrich, 2011). It is possible that the lysate-treated Ras GTPases are activated by components in the lysate, thus resulting in a stronger interaction with BTK. Lysate treatment may have also resulted in the phosphorylation of Ras GTPase, which has previously been shown to promote GTPase activity and its association with GAPs (Bunda et al., 2014).

RHOA Targets. The role of RHOA in cytoskeleton rearrangement and cell migration is well-established (Zegers & Friedl, 2014). Targets in the biological processes of cellular component movement, localization, and locomotion were enriched in interactions with higher binding affinities to active RHOA following lysate treatment. This target group was also enriched in the PI3K subunit gene family, which acts upstream of RHOA to promote cell migration (J. M. Kim, Kim, Lee, & Jeong, 2016; A. L. Zhang, Yan, Wang, Huang, & Liu, 2017).

8.3.6 GTPase activation state affects binding partners and kinetics

GTPases regulate a variety of biological processes, including cell proliferation, survival, migration, and growth. *In vivo*, they exist in two conformational states, an

inactive GDP- form and an active GTP-bound form (Kumawat, Chakrabarty, & Kulkarni, 2017; Vetter & Wittinghofer, 2001). GTPase activation is accompanied by a structural change that occurs primarily by the switch I and II domains that bind to the γ-phosphate of GTP. The ~ 200 angstrom^2 GTP-bound switch region is a focal point of most biomolecular interactions that mediate most downstream signaling (Dvorsky & Ahmadian, 2004). I therefore sought to better understand the effect of the two conformational states on protein partners, binding kinetics, and binding affinities. Two Rho GTPases sharing 60% sequence homology and that are important regulators of the BCR signaling pathway, RAC1 and RHOA, were chosen for this study. They were inactivated and activated with GDP and non-hydrolyzable GTP (i.e., GTPγS), respectively. Since RAC1 and RHOA have intrinsic GTP hydrolysis, the use of hydrolyzable GTP would have resulted in mixed signals arising from inactive and active GTPase interactions.

Active GTPases interacted with more proteins than inactive GTPases. In this study, RAC1 and RHOA interacted with 95 and 97 targets, identifying 96% known interactions and 166 (86%) novel interactions. Both GTP-bound RAC1 and RHOA had significantly more protein interactions than their GDP-bound counterparts (Figure 75), which is consistent with the idea that active GTPases mediate most downstream

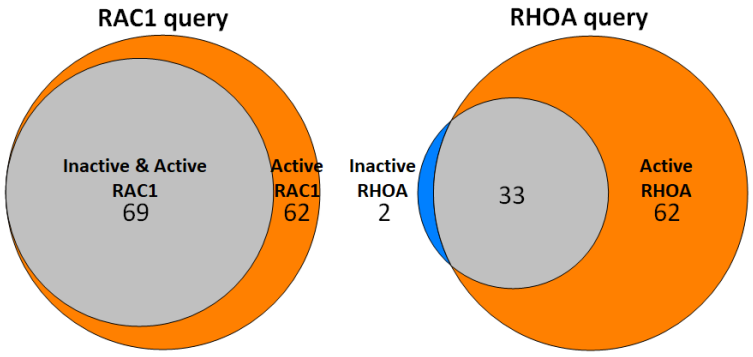


Figure 74. Active Rho GTPases, RAC1 and RHOA, interacted with more proteins than inactive Rho GTPases.

effectors. Ninety-two percent (92%) of the targets that interacted with inactive GTPases also interacted with their active forms. The primary difference between the inactive and active states of RAC1 and RHOA is that they interacted with 38% and 171% more targets, respectively, when active. Additional Venn diagrams are in Appendix M.

Some of these data are supported by a recent study by Paul et al. that examined the interaction partners of RAC1 and RHOA using affinity purification mass spectrometry (Paul et al., 2017). In short, the GTPases were expressed in *E. coli* with an N-terminal GST tag, purified with a glutathione column, loaded with GDP or GTPγS to inactivate or activate them, respectively, and then incubated in the lysate of mouse brain tissue. Mass spectrometry analyses were then performed following pull-downs in triplicate. Like the NAPPA-SPRi data, Paul et al. identified more novel interactions (82/116; 71%) than known interactions. They also observed that RAC1 and RHOA had 8- to 3.5-fold more interactions, respectively, when active. They did not, however, detect the same protein partners of the inactive and active forms. This is surprising since GDP- and GTP-bound GTPases have been demonstrated to interact with some of the same proteins (Bos, Rehmann, & Wittinghofer, 2007; Cotton et al., 2007). Other differences between our two studies are worth mentioning. NAPPA-SPRi detected 43% more interactions than the study performed by Paul et al. and, since their pool of target proteins theoretically includes all expressed proteins, NAPPA-SPRi also detected more known interactions. Ninety-six percent (96%; 26/27) of the possible known interactions of RAC1 and RHOA were detected with NAPPA-SPRi whereas Paul et al. detected 49% (34/69) of the protein interactions curated by the online BioGRID database for mice. Finally, their

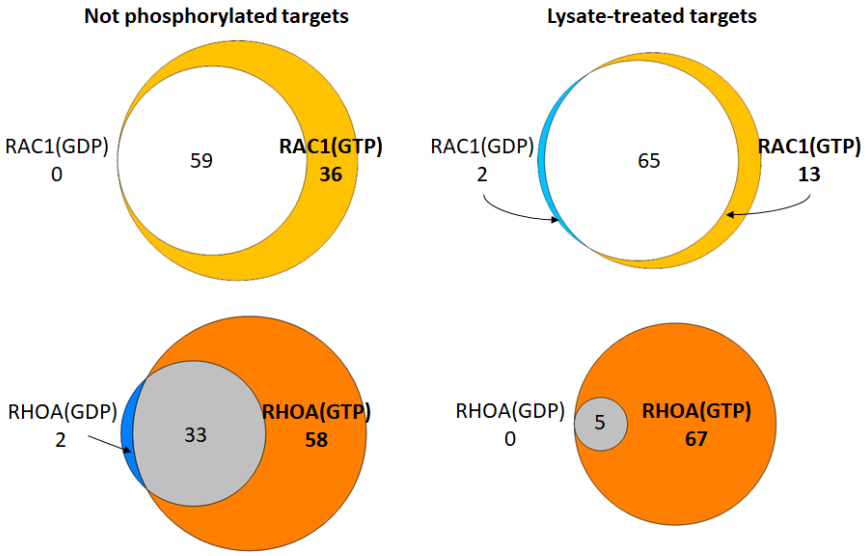


Figure 75. Venn diagram comparing the PPIs between inactive and active GTPases with (left) NP-target proteins and (right) LT-target proteins.

affinity purification method selects for stable interactions and it cannot characterize the interactions in terms of their binding kinetics of affinities.

Inactive and active GTPases interacted with NP- and LT-targets differently. As discussed in sections 8.3.4 and 8.3.5, GTPase activation affected the binding interactions with NP- and LT-targets (Figure 76, Appendix Q). Inactive RAC1 bound to 14% more LT- targets than NP- targets. However, active RAC1 and RHOA interacted with 22% and 26% more NP-targets than LT-targets. Inactive RHOA, on the other hand, interacted with 600% more NP-targets than LT-targets. This binding profile of inactive RHOA has not been reported by others, which is likely because the majority of known RHOA interactions are from studies employing active RHOA; for example, the study by Paul et al. discussed above (Paul et al., 2017). These results suggest that inactive RHOA may play a more significant role in regulating unphosphorylated proteins – perhaps in unstimulated cells where RHOA activation and protein phosphorylation are minimal –

than previously believed. Inactive and active RAC1 GTPase had significantly different kinetic profiles with NP- and LT- targets as well (Figure 70, Figure 71). The effect of lysate treatment of the target proteins affected binding with GDP-bound RAC1 by increasing the affinity, which was mostly associated with small increases in its on-rates and decreases in its off-rates.

Among the most dramatic observations in my study was the effect of lysate treatment on the target interactions with active RAC1. Lysate treatment significantly slowed both on and off rates of binding to active RAC1, but in direct proportion to each other. Thus, even though the magnitude of difference in binding to RAC1 between dephosphorylated and lysate-treated protein was often more than 220-fold, there was almost no change in the dissociation constant (i.e., binding affinity). This suggests that phosphorylation specifically regulates the interaction rates among proteins without changing the fraction of proteins bound!

Lysate treatment resulted in both lower and higher binding affinities with active RHOA, which were primarily associated with slower on-rates *or* slower off-rates, respectively (Figure 72). Overall, the changes in binding kinetics were minimal, although for ~24% of the interactions, the average on- or off-rates decreased by 17-fold and 35-fold, respectively, with a 21-fold change to the binding affinities. The binding affinities of inactive and active RAC1, on the other hand, were small (i.e., average 1.48-fold change) compared to binding kinetics that decreased, on average, > 200-fold with lysate treatment.

Similar to the effect of target de-phosphorylation, GTP-bound RAC1 interactions had faster on- and off-rates than their GDP-bound counterparts with moderate alterations to the binding affinity (Figure 77, Figure 78, Appendix Q). The increase in on- and off-rates were amplified upon RAC1 activation compared to the effect of target de-phosphorylation. More specifically, the GDP → GTP transition with LT- targets increased the average on-rates and off-rates by 25- and 64-fold, respectively, while the average binding affinity decreased 3.3-fold (not including the RAC1-VAV1 interaction) (Figure 77). However, for 31% (17/55) of RAC1's interactions with LT- targets, the on- and/or off-rates increased by > 2 orders of magnitude. That is, RAC1 activation increased the overall on- and off-rates in this group by 4.5 and 5.1 orders of magnitude, respectively, with only a decrease in binding affinity by 1.3-fold. This effect was enhanced with NP- targets where the average

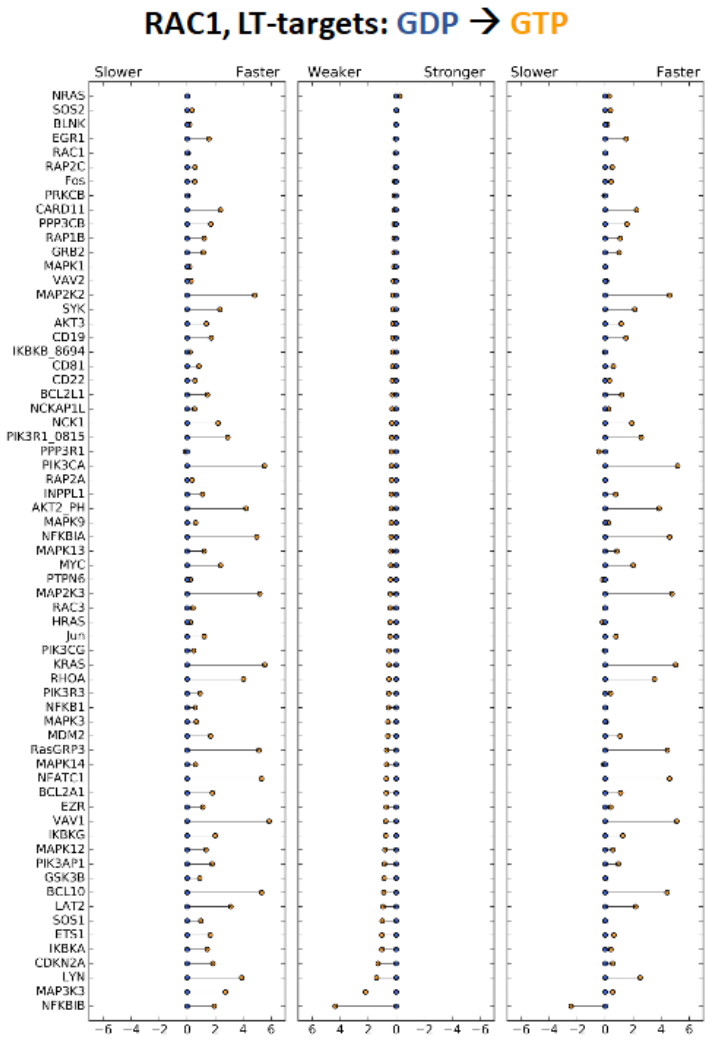


Figure 76. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs with inactive and RAC1 to LT-targets.

on- and off-rates increased by 3.9 and 4.2 orders of magnitude, respectively, with an average increase in binding affinity by 1.9-fold for 98% of the interactions (Figure 78). This novel kinetic regulation *cannot* be detected using classic equilibrium-based assays. It is also interesting that RAC1 activation significantly increased its interactions with NP- targets as well. That is, RAC1's interactions with LT- targets increased only 16% after activation while increasing 61% for NP- targets.

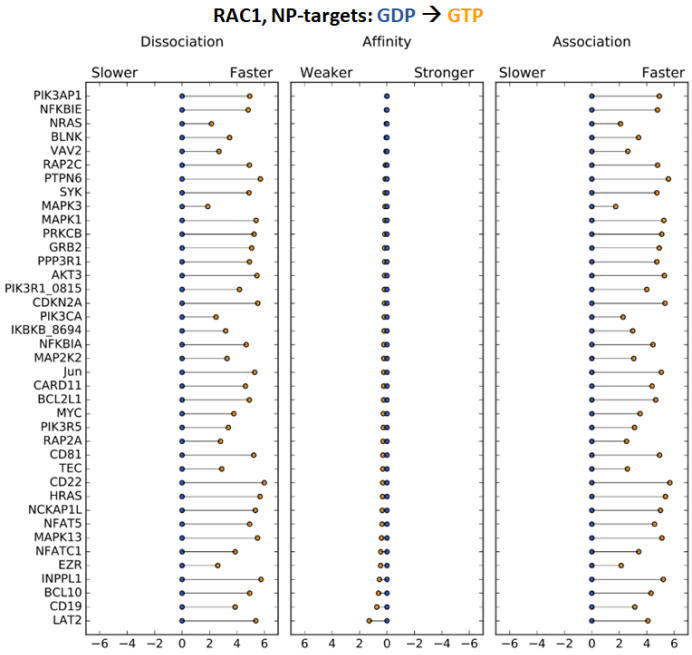


Figure 77. Bar plots showing the relative log10 change in k_d, K_D, and k_a with inactive RAC1 compared to active RAC1 to NP-targets. Bar plots showing all of the PPIs with NP- and LT-targets are in Appendix P.

Surprisingly, RHOA activation did not significantly affect the binding kinetics or affinities with NP- targets. On average, the on-rate increased 1.8-fold, the off-rate decreased by 1.25-fold, and the binding affinity decreased 1.9-fold (Figure 79). This kinetic regulation was different from the regulation following lysate treatment, which either increased the on-rates or decreased the off-rates to increase the binding affinity. Inactive and active RHOA interacted with the same five NP- and LT-targets, resulting in a small group for comparison. However, changes to binding affinities following RHOA

activation were associated with changes in the on-rate (Appendix P). Although RHOA activation did not significantly affect its binding kinetics or affinities, it increased RHOA's interactions by 160% with NP- targets (i.e., 91 vs 35) and 1340% with LT-targets (i.e., 67 vs 5).

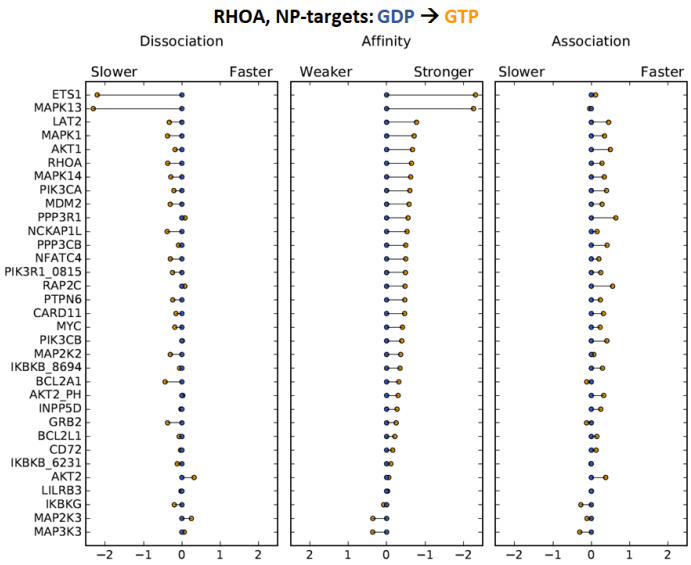


Figure 78. Bar plots showing the relative log10 change in k_d, K_D, and k_a with inactive RHOA compared to active RHOA to NP-targets. Bar plots showing all of the PPIs with LT-targets are in Appendix P.

In summary, RAC1 and RHOA are both Rho GTPases, but have different protein partners and kinetic profiles. RAC1 activation results in faster on- and off-rates with relatively little change in affinity. This effect was observed with nearly all of its interactions with unphosphorylated targets while only affecting 31% of lysate-treated targets. In regards to its biological consequences, faster association rates allow RAC1 to interact competitively with targets with much higher efficiency than inactive RAC1. It also allows RAC1 to sample more interactions in a shorter amount of time. Faster dissociation rates allow proteins like GAPs access to RAC1 to negatively regulate its signaling. RAC1 activation also allows it to interact with many more unphosphorylated targets. RHOA activation causes moderate alterations in its on- and off-rates and binding affinities. Notably, activated RHOA demonstrated binding to many more proteins than inactive RHOA, suggesting that activation expands its target range.

Disorder-to-order transitions might explain faster binding rates of active GTPases to unphosphorylated proteins. Both active RAC1 and RHOA experienced faster on-rates with NP-targets than LT-targets. The dissimilar kinetics between NP- and LT-targets suggest that the targets are being modified by the B cell lysate, most probably via phosphorylation. Phosphorylation increases the local negative charge and, at least around the phosphosite, stimulates a disorder-to-order transition via the formation of salt bridges and hydrogens between the phosphate and neighboring residues (Nishi et al., 2011; Nishi et al., 2014; Raggiaschi, Gotta, & Terstappen, 2005). This is particularly relevant for the phosphorylation of serine and threonine residues where the majority of phosphorylation occurs (i.e., 65 – 99%, depending on the source) and which are frequently found in disordered and flexible regions. Thus, these kinetic profiles may be explained by a variation of the "fly-casting" hypothesis, which was originally proposed in 2000 by Shoemaker et al. (Shoemaker, Portman, & Wolynes, 2000). The hypothesis states that unfolded, intrinsically disordered proteins (IDPs), characterized by composition biases toward polar and charged amino acids and low sequence complexity, can bind to other proteins faster because they have a larger capture radius; the bound protein can then be "reeled in" (Uversky, 2013; Wright & Dyson, 2015). Subsequent experiments indicate that IDPs with *some* pre-formed structure generally do have faster on-rates and that IDPs can be very selective about their binding partners (Mollica et al., 2016). While NP-targets are not IDPs, I speculate that the binding interfaces of NP-targets are more disordered and flexible prior to phosphorylation with lysate treatment. Notably, not all of the interactions with active RAC1 and RHOA behaved this way;

somes on-rates did not change while others actually decreased following lysate treatment. A possible explanation is that the phosphorylation either does not affect the binding interface or induces a disorder-to-order transition at the binding site(s).

8.3.7 Comparison of protein-protein interactions between NanoBRET and NAPPA-SPRi.

Protein interactions identified with NAPPA-SPRi were compared with those obtained with the qualitative, bioluminescence-based platform, NanoBRET (see Chapter 3.4.3, page 53). Forty percent (40%; 328/818) of all interactions and 41% (286/702) of novel interactions detected with NAPPA-SPRi were also detected with NanoBRET (Table 18, Figure 80) (additional Venn diagrams are in Appendix O). NAPPA-SPRi detected 473 unique interactions across all seven queries while NanoBRET only detected 305. NAPPA-SPRi likely detected 55% more interactions than NanoBRET because it can analyze interactions in real-time across a wide range of on-rates, off-rates, and binding affinities whereas NanoBRET signal is determined by the number of bound proteins at equilibrium.

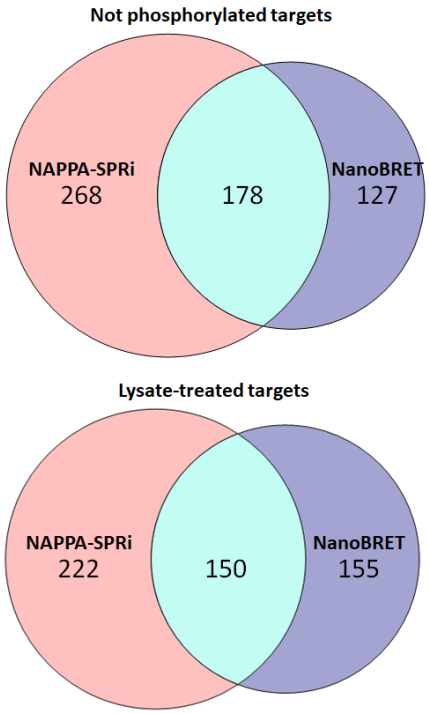


Figure 79. Venn diagram of protein interactions detected by NanoBRET and NAPPA-SPRi.

Table 18. PPIs that were observed by NanoBRET and NAPPA-SPRi

Query protein	BLNK		BTK		PI3K		RAC1(GDP)		RAC1(GTP)		RHOA(GDP)		RHOA(GTP)	
Target Phosphorylation	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT
Total count	28	28	16	19	13	6	24	29	54	44	12	2	31	22
Target proteins	AKT1	AKT1	AKT1	AKT1	AKT1	AKT1	AKT3	AKT2*	AKT1	AKT1	AKT1	AKT1	AKT2**	AKT2**
	AKT2*	AKT2*	BCL2A1	BCL2A1	BCL2A1	BCL2A1	BCL10	AKT3	AKT2*	AKT2*	CARD11	MAPK13	BCL2A1	BCL2A1
	AKT2**	AKT2**	BLNK	BLNK	CD79A	BLK	BCL2L1	BCL10	AKT2**	AKT3	ETS1		BTK	BTK
	BCL2A1	BCL2A1	CD79B	CD79B	IKBKB##	MAPK1	CD19	BCL2A1	AKT3	BCL2A1	IKBKG		CD72	CD72
	CD79A	CD79A	IFITM1	ETS1	IKBKG	MYC	CD22	BCL2L1	BCL2A1	BCL2L1	MAP2K2		CD81	IFITM1
	CD81	CD79B	LIME1	IFITM1	KRAS	PIK3R1~~	CDKN2A	CD19	BCL2L1	CARD11	MAP3K3		CDC42	IKBKA
	IKBKB##	CD81	NFATC1	LIME1	MAPK1		GRB2	CD22	BLK	CD79B	MAPK13		GSK3B	IKBKB#
	IKBKG	IKBKA	RAC2	MAP2K2	MAPK12		GSK3B	CDKN2A	CARD11	CD81	MDM2		HRAS	IKBKB##
	INPPL1	IKBKB##	RAP2A	RAC2	MYC		HRAS	EGR1	CD72	CDKN2A	NCKAP1L		IFITM1	IKBKG
	LYN	IKBKG	RAP2C	RAC3	NCKAP1L		IKBKA	GRB2	CD79A	EGR1	NFKB1		IKBKA	LAT2
	MALT1	INPPL1	RasGRP3	RAP2A	PIK3R1~~		IKBKB##	GSK3B	CD79B	ETS1	PIK3CA		IKBKB#	LILRB3
	MAP3K3	MALT1	RASSF5	RAP2C	PLCG2		JUN	HRAS	CD81	EZR	PPP3R1		IKBKB##	MAP2K2
	MAPK12	MAP3K3	RELA	RasGRP3	PPP3CB		KRAS	IKBKA	CDKN2A	IKBKA			IKBKG	MAP2K3
	MAPK13	MAPK12	TEC	RASSF5			MAP2K2	IKBKB##	EGR1	IKBKB#			KRAS	MAPK13
	MAPK14	MAPK13	VAV1	RELA			MAP2K3	JUN	ETS1	IKBKB##			LAT2	MYC
	MYC	MYC	VAV2	SOS2			MAPK1	KRAS	EZR	IKBKG			LILRB3	PPP3CA
	NFKB1	NFATC4		TEC			MAPK12	MAP2K2	GRAP2	KRAS			MAP2K2	RAP1B
	NFKBIA	NFKBIA		VAV1			MAPK13	MAP2K3	IKBKA	LAT2			MAP2K3	RAP2A
	NFKBIE	NFKBIE		VAV2			MAPK14	MAPK1	IKBKB#	LILRB3			MAPK13	RASSF5
	NRAS	NRAS					MDM2	MAPK12	IKBKB##	LYN			MYC	TEC
	PIK3CA	PIK3CA					MYC	MAPK13	IKBKG	MAP2K2			NFATC4	TP53
	PIK3CG	PIK3R1~					PIK3AP1	MAPK14	INPP5D	MAP2K3			PPP3CA	VAV1
	PIK3R1~	PTEN					PIK3R3	MDM2	KRAS	MAP3K3			PTEN	
	PTEN	RAP1B					PPP3R1	MYC	LAT2	MAPK13			RAF1	
	RAC1	RAP2A						NFKB1	LILRB3	MAPK14			RAP1B	
	RAP1B	SYK						PIK3AP1	LYN	MDM2			RAP2A	
	RAP2A	VAV1						PIK3R3	MALT1	MYC			RASSF5	
	VAV2	VAV2						PPP3R1	MAP2K2	NFATC4			TEC	
								RAC1	MAP2K3	NFKB1			TP53	
									MAP3K3	NFKBIA			VAV1	
									MAPK13	NFKBIB			VAV3	
									MAPK14	NFKBIE				
									MDM2	PIK3CA				
									MYC	PIK3CG				
									NFATC3	PPP3R1				
									NFATC4	PTEN				
									NFKB1	PTPN6				
									NFKBIA	RAC1				
									NFKBIB	RAP1B				
									NFKBIE	RAP2A				
									PIK3CA	SYK				
									PIK3CB	TEC				
									PIK3CG	TP53				
									PPP3R1	VAV1				
									PTEN					
									PTPN6					
									RAC1					
									RAP1B					
									RAP2A					
									RASSF5					
									SYK					
									TEC					
									TP53					
									VAV1					

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Known interactions that are in the online databases, BioGRID or HPRD

In both NAPPA-SPRi and NanoBRET analyses, GTP- bound RAC1 had more protein partners than GDP-bound RAC1 (38% and 43% more PPIs, respectively). NAPPA-SPRi also detected significantly more interactions with activated RHOA than inactivated RHOA (271%), although no difference between activated and inactivated RHOA was observed with NanoBRET (3%).

8.3.8 Novel interactions detected by NAPPA-SPRi and NanoBRET

BTK-ETS1 interaction. Both NAPPA-SPRi and NanoBRET detected a novel interaction between BTK and the transcription factor, ETS1, which inhibits B cell differentiation into plasma cells and decreases autoantibody tolerance (Figure 81) (Russell et al., 2015). Several studies have demonstrated that ETS1 downregulation in activated B cells is dependent upon BTK (Gutierrez, Halcomb, Coughran, Li, & Satterthwaite, 2010; Luo et al., 2014). In one study, transgenic mice were generated expressing different levels of BTK and ETS1. Increased BTK expression levels resulted in decreased expression of ETS1, and vice versa (Mayeux et al., 2015). Thus, a functional, but not physical or biochemical, relationship between BTK and ETS1 for maintaining plasma cell homeostasis has been established. This study shows that BTK and ETS1 bind to each other *in vitro*. It is also possible that the proteins physically interact with each other *in vivo* because they are both in the cytoplasm and nucleus. This interaction is discussed in more detail in the next section (Chapter 8.3.9).

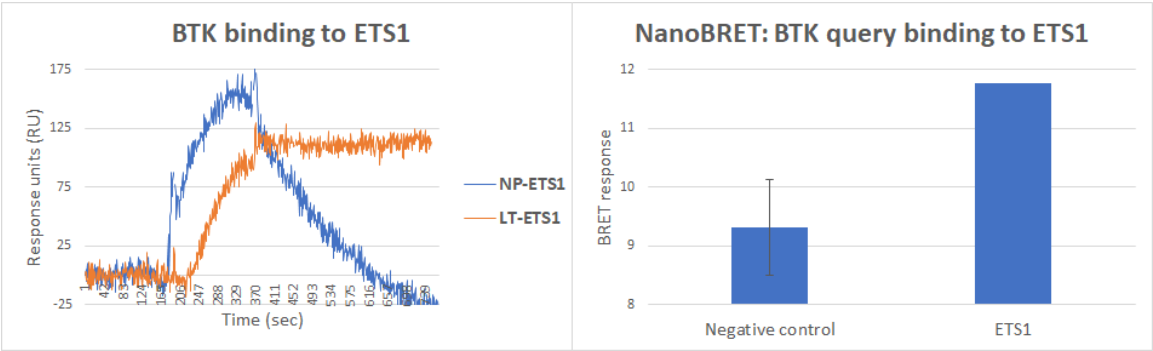


Figure 80. Novel BTK-ETS1 interaction detected by NAPPA-SPRi and NanoBRET. NAPPA-SPRi binding sensorgram (left) and NanoBRET response (right), where error bars represent the range of response across technical replicates.

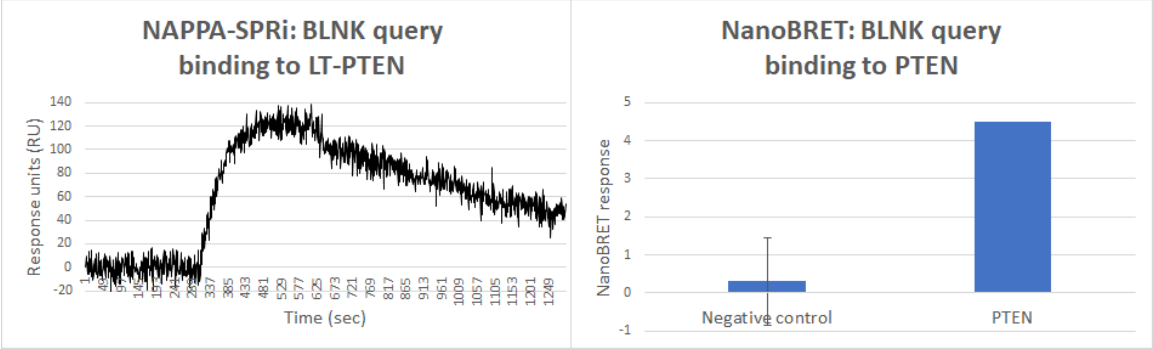


Figure 81. Novel BLNK-PTEN interaction detected by NAPPA-SPRi and NanoBRET. NAPPA-SPRi binding sensorgram (left) and NanoBRET response (right), where error bars represent the range of response across technical replicates. BLNK interacted with NP- and LT-PTEN; LT-PTEN data not shown.

BLNK-PTEN interaction. Another novel interaction that was detected by NAPPA-SPRi and NanoBRET was between the adaptor protein, BLNK, and the phosphatase and tumor suppressor, PTEN (Figure 82). PTEN is a negative regulator of PI3K signaling and, as such, is a negative regulator of the BCR signaling pathway as well. More specifically, PTEN reverses PI3K's effect by dephosphorylating the PI3K substrate, PIP3, back to PIP2 (Milella et al., 2015). PTEN also has roles that are independent from PI3K and phosphatase activity, including contributing toward centrosome stability in the nucleus. While the BLNK-PTEN interaction has not been reported previously, another phosphatase that is a negative regulator of the BCR signaling pathway and a PIP3 phosphatase, PTPN6, has been demonstrated to de-phosphorylate BLNK. De-phosphorylation of BLNK by PTPN6 modulates BLNK's ability to bind other proteins, which results in a decrease of MAPK8 (i.e., JNK) kinase activity. Therefore, it's possible that PTEN binds to BLNK in order to alter its phosphorylation and PPIs. PTEN's PI3K-independent roles in B cells have not been explored but may have relevance in B cell-related cancers that have PTEN deficiencies,

including diffuse large B cell lymphoma, mantle cell lymphoma, and primary cutaneous follicle center lymphomas (X. X. Wang, Huang, & Young, 2015).

Active RHOA-IKBKA interaction. In both NAPPA-SPRi and NanoBRET analyses, GTP-bound RHOA, but not GDP-bound RHOA, interacted with IKBKA (Figure 83). The connection between RHOA and the NFκB pathway has been well-documented (H. J. Kim, Kim, Moon, Park, & Park, 2014; Tong & Tergaonkar, 2014). Central to the NF-κB pathway is the IκB complex that is composed of two serine/threonine kinases (IKBKA, IKBKB) and one scaffold protein (IKBKG) (Woyach et al., 2012). IKBKA or IKBKB activation results in the phosphorylation – and subsequent degradation – of proteins that bind to and sequester the transcription factor NFκB in the cytoplasm. Once the inhibitor proteins are degraded, NFκB translocates into the nucleus where it regulates the transcription of genes involved in cell proliferation, class switching, survival, and the secretion of pro-inflammatory cytokines (Carlberg, 2016). RHOA, through its downstream kinases, has previously been demonstrated to increase the transcriptional activity of NFκB in the NFκB pathway (Shih, Tsui, Caldwell, & Hoffmann, 2011). These data indicate that active RHOA may have a more direct

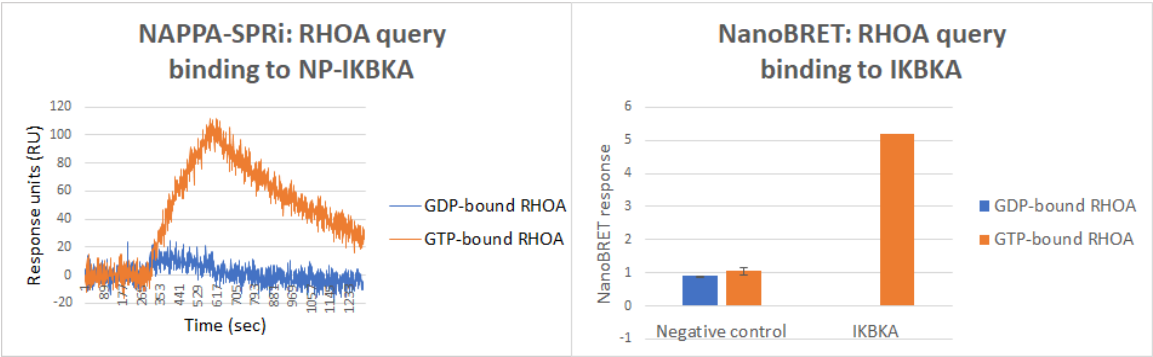


Figure 82. Novel RHOA-IKBKA interaction detected by NAPPA-SPRi and NanoBRET. NAPPA-SPRi binding sensorgram (left) and NanoBRET response (right), where error bars represent the range of response across technical replicates. NanoBRET response for GDP-bound RHOA and IKBKA is -0.00992.

involvement with the NFκB pathway by binding to IKBKA. The RHOA-IKBKA interaction may provide insight into the upstream events leading to non-canonical NFκB activation and explain why RHOA is essential in B cell development (S. M. Zhang, Zhou, Lang, & Guo, 2012)

8.3.9 Validation of novel protein-protein interactions

Analyzing protein interactions with unphosphorylated and lysate-treated "phosphorylated" targets presented a unique opportunity to look for potential novel phosphorylation events based off of distinct kinetic profiles between the two datasets. In this section, four novel interactions that appeared to involve the phosphorylation of a target by BTK or PI3K were validated. The interactions were analyzed using zinc-based Phos-Tag™ SDS-PAGE separation in which zinc attached to highly cross-linked agarose inhibits the migration of phosphorylated species more than standard SDS-PAGE. The gels were then transferred for Western blot analyses using target-specific antibodies.

BTK-JUN interaction. On the NAPPA-SPRi platform, the on-rate and binding response of the BTK query with JUN were much higher with NP-JUN than LT-JUN (Figure 84). This suggested that BTK may be phosphorylating JUN, a transcription factor that controls

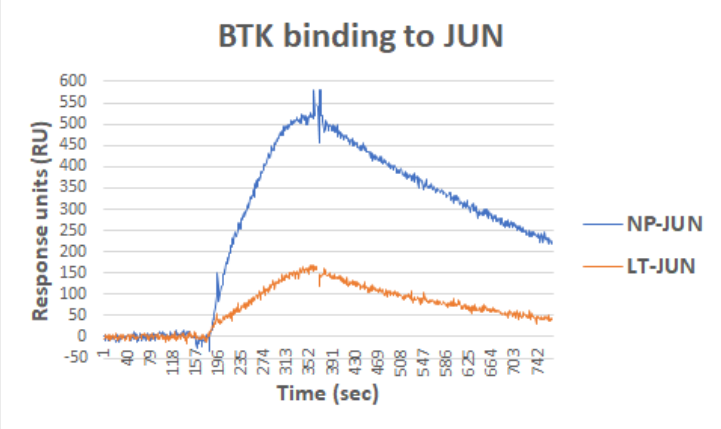


Figure 83. NAPPA-SPRi binding sensorgram of the BTK query binding JUN with a fusion tag at the C-terminus.

the expression of genes involved in cell proliferation, apoptosis, transformation, differentiation, and development, were (de Gorter, Vos, Pals, & Spaargaren, 2007). A migration shift was observed only when JUN was mixed with BTK and ATP (Figure 85). Addition of phosphatase to the mixture ensured that the migration was the result of a phosphorylation event.

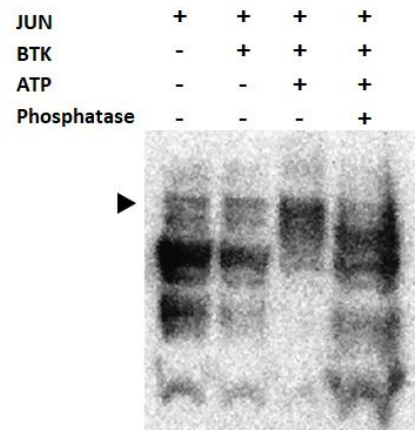


Figure 84. Western blot image showing that BTK phosphorylates JUN. Black triangle marks migrated JUN band due to phosphorylation.

In pre-B-lymphoblastic lymphoma cells, BTK knockdown inhibited JUN expression (Hiratsuka et al., 2016). Both BTK and JUN are also often overexpressed in splenic marginal zone lymphoma and Hodgkin disease (Mathas et al., 2002; Troen et al., 2004). Therefore, BTK-mediated phosphorylation may make JUN more resistant to proteases. Interestingly, the only tyrosine known to be phosphorylated on JUN, Y170, protects JUN from ubiquitination-mediated degradation (Gao, Lee, & Fang, 2006; Hornbeck et al., 2015). BTK may be phosphorylating JUN at this site.

This interaction was not detected by NanoBRET. One possible reason is that JUN may have been phosphorylated by components in the cell-free expression system *before* BTK was added. Another possibility is that the interaction between BTK and JUN was simply too transient to be detected since NanoBRET signal, as an equilibrium-based assay, is determined by the number of bound proteins at equilibrium. These data also

highlight the utility of NAPPA-SPRi in detecting novel, transient interactions in regards to post translational modifications like phosphorylation.

BTK-BCL2 interaction.

BTK also interacted with BCL2, an important anti-apoptotic protein, on the NAPPA-SPRi platform with different binding profiles when BCL2 was de-phosphorylated and lysate-treated (Figure 86). This interaction was also not detected with NanoBRET, further demonstrating the usefulness of NAPPA-SPRi in detecting phosphorylation events. BCL2 phosphorylation was validated with Phos-Tag Western blot analysis using an anti-BCL2 antibody and subsequent phosphatase treatment (Figure 87). Two migrated BCL2 bands, one heavy and one light in intensity, might reflect two different phosphorylation events.

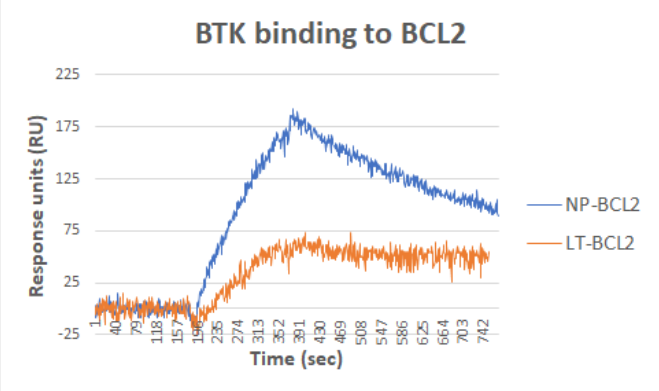


Figure 85. NAPPA-SPRi binding sensorgram of the BTK query binding JUN with a fusion tag at the N-terminus.

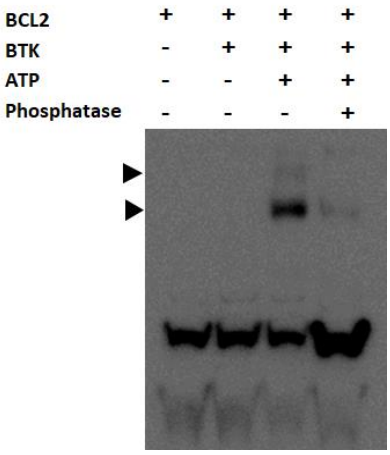


Figure 86. Western blot image that shows BTK phosphorylates BCL2. Black triangle marks migrated BCL2 band due to phosphorylation.

A functional relationship between BTK and BCL2 has been demonstrated using BTK small molecule inhibitors, in which BTK inhibition increases a cell's sensitivity to the anti-apoptotic effects of BCL2 (Deng et al., 2015). Moreover, the use of BTK and BCL2 inhibitors in combination successfully killed chronic lymphocytic leukemia (CLL)

cells *ex vivo* (Davids, 2017). Several clinical trials using BTK and BCL2 inhibitors in conjunction to treat chronic lymphocytic leukemia (CLL) are underway even though the biochemical relationship between BTK and BCL2 is not understood. The novel BTK-mediated phosphorylation of BCL2 identified in this study may inhibit BCL2's subcellular location or activity. Phosphorylation of other proteins in the BCL2 family have been demonstrated to affect their translocation to the outer mitochondrial membrane, interactions, and activity (Schinzel, Kaufmann, & Borner, 2004).

BTK-ETS1 interaction. Both NanoBRET and NAPPA-SPRi detected the interaction between BTK and ETS1 (Figure 81). Although BTK can have kinase-independent roles, the distinct binding profiles with unphosphorylated and lysate-treated ETS1 suggested that BTK may be phosphorylating ETS1, a transcription factor that is essential in B cell differentiation and tolerance (Figure 88) (Middendorp, Dingjan, Maas, Dahlenborg, & Hendriks, 2003; Russell et al., 2015; Saito et al., 2003). That is, BTK binds strongly to the unphosphorylated form and releases the lysate-treated form rapidly, which would be expected of a kinase binding its unmodified substrate, and then releasing it after the modification. As the Western blot image in Figure 88 shows, ETS1 does not experience any migration shifts in the absence of BTK or ATP. In the presence of BTK and ATP, however, a migrated band is

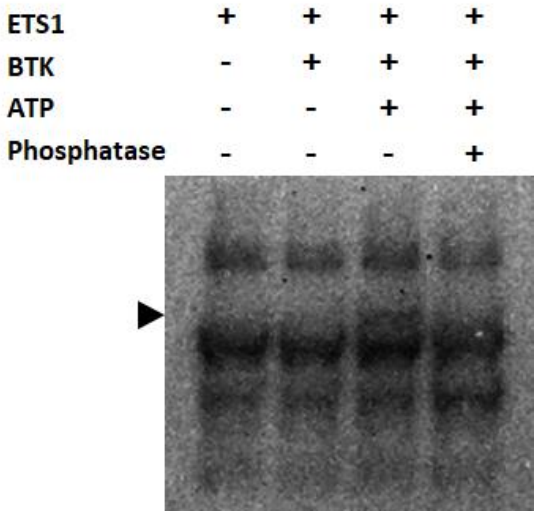


Figure 87. Western blot image showing that BTK phosphorylates ETS1. Black triangle marks migrated ETS1 band due to phosphorylation.

observed. To verify that the shift was due to phosphorylation, the BTK-ETS1-ATP sample was treated with lambda protein phosphatase.

Previous studies have shown that ETS1 phosphorylation events can inhibit or promote DNA binding, make ETS1 more protease resistant, or have no effect (Cowley & Graves, 2000; Lu et al., 2014). It's already been established that BTK downregulates ETS1 expression in activated B cells as well as B cells in diffuse large B cell lymphoma, Burkitt's lymphoma, and Hodgkin disease (Mayeux et al., 2015; Testoni, Chung, Priebe, & Bertoni, 2015). It is thus possible that BTK-mediated phosphorylation makes ETS1 more prone to degradation.

PI3K directly phosphorylates myc at serine 62. In a manner similar to those examples described above, PI3K favored binding with unphosphorylated MYC over the lysate-treated form on the NAPPA-SPRi platform (Figure 89). This interaction was detected with NanoBRET as well (Figure 90). MYC is a transcription factor that is dysregulated in 70% of all cancers with two well-studied phosphorylation sites that affect its stability: serine 62 and threonine 58. Phosphorylation of serine 62 by Erk and Src

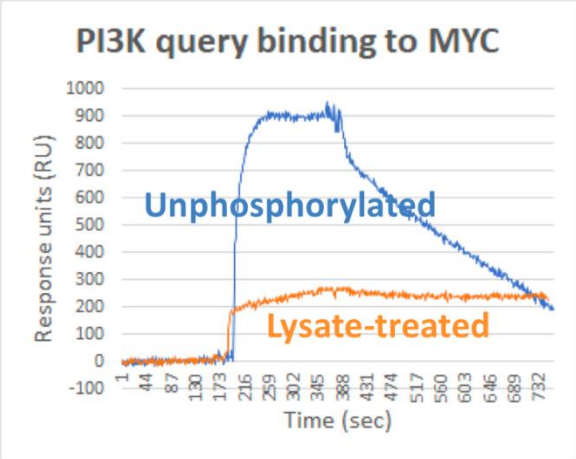


Figure 88. NAPPA-SPRi binding sensorgram of PI3K query binding to MYC.

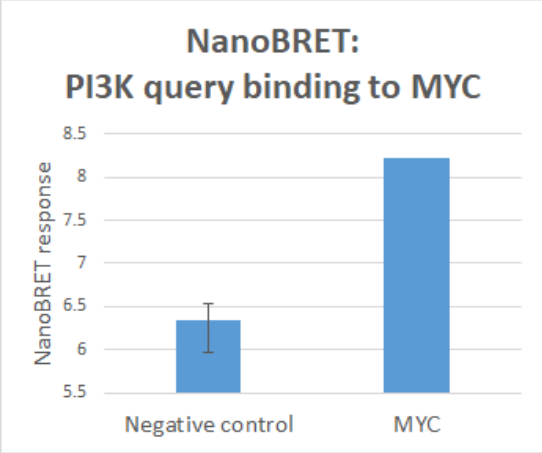


Figure 89. PI3K-MYC interaction detected by NanoBRET

family kinases increases MYC's half-life, while phosphorylation of threonine 58 by GSK3B promotes its degradation. Previous studies have shown that PI3K *indirectly* inhibits MYC's degradation by activating AKT1, a serine/threonine kinase that inhibits GSK3B. This may account for the observation that sustained PI3K activity and MYC overexpression result in cancer. However, no known direct physical or biochemical relationship between the two proteins have been previously reported. Here, PI3K-mediated phosphorylation of MYC was validated with Western blot analyses, demonstrating that PI3K also inhibits MYC's degradation by *directly* phosphorylating MYC at serine 62 (Figure 91).

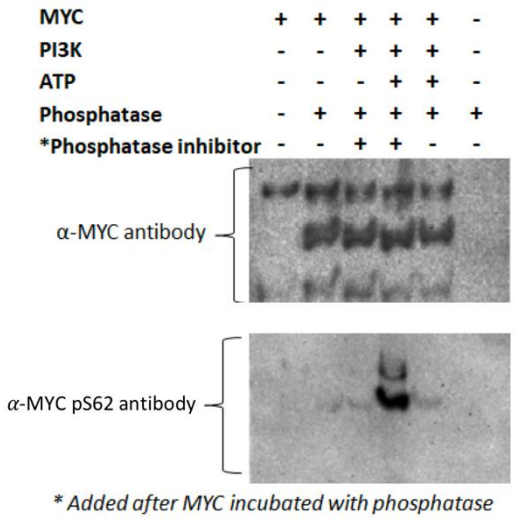


Figure 90. Western blot demonstrating that PI3K phosphorylates MYC at serine 62.

Using NAPPA-SPRi, four interactions that appeared to be novel phosphorylation events based on distinct binding responses with NP- and LT-targets were identified. These were then validated using SDS-PAGE migration and Western blot analyses. Notably, only two of these interactions, BTK-ETS1 and PI3K-MYC, were also observed with NanoBRET. The interactions, BTK-JUN, BTK-BCL2, BTK-ETS1, and PI3K-MYC, include proteins that are important in maintaining homeostasis. Therefore, their interactions have potentially real and direct applications to human health and disease. Follow-up experiments to determine whether these events occur *in vivo*, the location of

the phosphorylation site(s), and the biological consequences of these phosphorylation events are needed.

8.3.10 Identification of promiscuous proteins

Proteins have traditionally been considered to be specific in which proteins they bind. However, protein promiscuity – or the ability to bind other proteins in a nonspecific or broad manner -- in biological processes and evolutionary fitness has only recently become appreciated. Proteins hubs essential in maintaining homeostasis and often disrupted in disease, like p53, p21, BRCA1, and ubiquitin, are promiscuous by definition because they have a disproportately high connectivity in protein networks (Patil, Kinoshita, & Nakamura, 2010). Lab-directed evolution experiments in 2005 and 2011 that used error-prone PCR and gene amplification in *E. coli* demonstrated that evolution selects for promiscuous proteins and that promiscuity increases fitness (Aharoni et al., 2005; Soo, Hanson-Manful, & Patrick, 2011). Promiscuous proteins are therefore likely to be more prevalent than traditionally believed, and the identification of such proteins may help to understand the molecular mechanisms underlying homeostasis, disease, and drug resistance. In this NAPPA-SPRi study, the number of interactions for each query and target protein widely differed from each other, thus suggesting a possible advantage of using this platform to identify novel protein hubs.

This study identified target proteins that behaved promiscuously by binding to queries $\geq$ 12 times out of the 14 different query-target conditions. The eleven targets included AKT1, BCL2A1, ETS1, IKBKB (Ref Seq ID: BC108694), IKBKG, MAPK1,

MAPK13, MYC, NCKAP1L, RAP2C, and RHOA. Six of these target proteins were also shown to bind to at least 5 (out of 7) different queries in the NanoBRET analyses: AKT1, BCL2A1, IKBKB (Ref Seq ID: BC108694), IKBKG, MAPK13, and MYC.

At the other end of the promiscuity spectrum, there were some target proteins that bound to very few proteins. LIME1, NFATC3, PIK3CD, PPP3CC, and RAP1A bound to $\leq$ 2 queries out of the fourteen query-target conditions. RAP1A also displayed non-promiscuous behavior in the NanoBRET analyses, binding to two of the seven queries.

To help determine whether these differences were real or artefacts from the experimental set-up, the number of previously reported PPIs curated by the online database, BioGRID, was determined for the proteins identified as "non-promiscous" (i.e., low number of interactions) and "promiscuous" (i.e., high number of interactions) using NAPPA-SPRi (Figure 92) (Stark et al., 2006). Promiscuous proteins had an average of 177 unique human protein interactions that have been previously reported, with AKT1, IKBKB, MYC, and MAPK1 having as many as 315, 327, 618, and 249 interactions, respectively. Non-promiscuous proteins, on the other hand, had an average of 31.4 unique protein interactions, with RAP1A having the most PPIs (i.e., 87) amongst this group. Some of the proteins identified as promiscuous, including BCL2A1, IKBKG, NCKAP1L, and RAP2C, had a

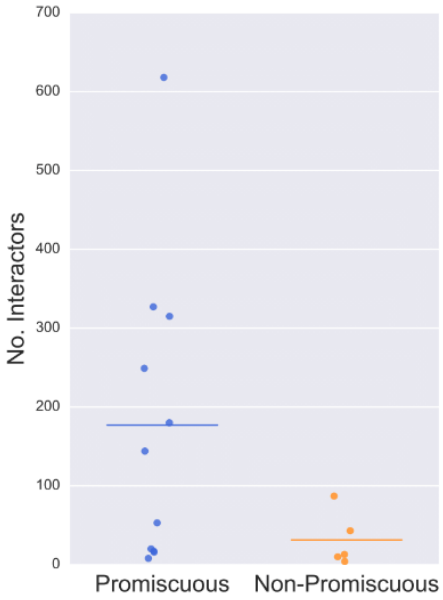


Figure 91. Number of unique human protein interactors with the target proteins identified as promiscuous and non-promiscuous with NAPPA-SPRi. Horizontal line represents the mean data point for each group.

low number ($\leq$ 20) of documented interactions in BioGRID. It is possible that these proteins bind many protein partners, but have not been studied to the same extent that AKT1, IKBKB, MYC, and MAPK1 have been. Taken together, these data suggest that the identification of novel protein hubs may be possible with NAPPA-SPRi.

No common domain, motif, or biological function among this group of proteins can explain their promiscuous behavior. However, a study of 305 enzymes by Chakraborty et al. revealed that $\geq$ 80% promiscuous proteins have > 39% polar (or > 20% charged) residues within 3 angstroms of the active site (p-value ~ 0.05) (Chakraborty & Rao, 2012). A weaker correlation was obtained regarding the features of the residues (i.e., basic, acidic, polar, charged) around the active site, such that the promiscuity was highest for charged residues and lowest for acidic residues. Thus, the promiscuity of the proteins in this study may be explained by the presence of charged residues around their binding sites rather than a specific domain or motif.

8.4 Conclusions

Here, a high throughput, quantitative method, NAPPA-SPRi, was applied toward studying protein interactions within the BCR signaling pathway. The platform detects distinct interactions, kinetics, and affinities depending on protein phosphorylation, GTPase activation state, protein isoform, and tag location. The differing kinetics indicate that the data are not the result of artefacts, but actually reflect the exquisite regulation of protein interactions to propagate signal. The vast majority of the interactions (85%) detected with NAPPA-SPRi were novel. The high overlap of novel interactions between

NAPPA-SPRi and NanoBRET as well as novel phosphorylation events validated with Western blot analyses reveal that the BCR signaling pathway – which is considered to be one of the better understood pathways – is *still* largely unmapped. Moreover, the kinetic profiles of RAC1 underscore the importance of measuring the on- and off-rates as this unique method of regulation would not be detected using methods that only measure binding affinity or simply determine which proteins bind to RAC1. Interestingly, the effect of tag location on binding may help provide information regarding binding epitopes and, theoretically, could help build structural networks by distinguishing which protein interactions are competitors or non-competitors with each other.

As an *in vitro* platform, the NAPPA-SPRi data may not accurately represent what is occurring *in vivo*. First, NAPPA-SPRi allows protein interactions to occur that would otherwise be impossible due to the proteins' *in vivo* subcellular locations. Since the subcellular location(s) of a vast majority of proteins have already been experimentally determined, however, this information could help filter out PPIs that could not, or are less likely to, occur *in vivo*. A larger concern perhaps is the high macroscopic viscosity of the cytoplasm, which affects the rotational movement and long-range diffusion of proteins, such that the concentration and location of a protein-of-interest may vary from one spot to another within a cell (Schreiber et al., 2009). The intracellular viscosity has also led scientists to estimate that the binding on-rates for most proteins to cannot exceed ~ 10^7 $M^{-1}s^{-1}$ *in vivo* (Pollard, 2010). Therefore, interactions identified in this study with extremely fast on-rates are not likely to occur *in vivo* (e.g., active RAC1 with NP-targets). Second, the immobilization of the target proteins could affect their conformation and, as a

consequence, their binding interactions. Indeed, there is evidence that kinetic values of PPIs obtained with SPR may differ from those obtained in solution (Schreiber et al., 2009). Third, NAPPA-SPRi, like many other methods that study PPIs, uses a fusion tag to capture the expressed target proteins to the slide surface. These could affect a protein's native conformation and block binding epitopes. One of these concerns was addressed by representing each target protein with a tag on the N-terminus or C-terminus, such that binding epitopes that may be blocked in one configuration could be available in the second. Fourth, the aqueous environment of NAPPA-SPRi is not well-suited for membrane proteins, which is why our study focused particularly on soluble proteins. Fifth, the buffers and cofactors were selected based on the query that was used; however, these do not accurately represent *in vivo* conditions. For example, intracellular GTP is roughly ten times higher than GDP. The non-hydrolyzable GTP (i.e., GTPγS) was also used to study the interactions of active GTPases in some of these experiments; no GTP or GDP were supplied. While exteme conditions were used in our experiments (e.g., completely unphosphorylated target proteins, GTPγS-bound GTPase queries), it is important to mention that these were necessary so that we could document the different behaviors in each state. Finally, some proteins expressed as separate spots on the array are usually found as heterodimers or complexes *in vivo*. These would include the regulatory and catalytic subunits of PI3K and calcineurin A; the heterodimer NFKB1-RELA; and the complex CARD11-BCL10-MALT1. It is possible, however, that proteins in the human cell-free expression system may bind and stabilize these monomers.

Standard SPR experiments generally use five to seven different concentrations of the query spanning as much as seven orders of magnitude to obtain absolute binding kinetics. These types of experiments are made possible through a cost-effective approach of regenerating the surface of the slide after each query concentration, which removes all residual query proteins from the slide. Regeneration is ideal for experiments in which peptides or antibodies are immobilized on the surface since these are stable to regeneration conditions that use acidic, basic, or high salt buffers. As such, regeneration may not be appropriate for some experiments as the buffers may negatively affect protein structure and, consequently, their protein interactions and binding kinetics. Regeneration was not performed in the NAPPA-SPRi experiments out of concern that the regeneration buffers were too harsh for the target proteins. An alternative approach called "kinetic titration" was explored, in which the query is added to the surface in increasing concentrations with no regeneration step. Unfortunately, NAPPA-SPRi is not compatible with kinetic titration due to the low amount of protein that is displayed (see Chapter 4.4.5). Two different query concentrations with NAPPA-SPRi were therefore tested in order to keep the experimental costs within budget. Since the standard five to seven different query concentrations were not used and some of the kinetic values obtained in this study were outside the linear detection range of the SPRi instrument (ka = 10^3 to 10^7 $M^{-1}s^{-1}$; kd = 10^{-5} to 10^{-1} s^{-1}; KD = 10^{-4} to 10^{-12} M), these results cannot be considered to be absolute kinetic values. Rather, these values are relative to each other and still represent the altered kinetics and affinities as the result of phosphorylation and protein activation states.

In these experiments, target protein phosphorylation was controlled by de-phosphorylating the proteins with phosphatases or phosphorylating them using B cell lysate from Ramos RA-1 B cells (spiked with protease and phosphatase inhibitors). Ramos is a Burkitt's lymphoma cell line that is negative for the Epstein-Barr virus. As such, its phosphorylation of the target proteins – and their protein interactions – represent a diseased state at a specific point in time. It is likely that the use of a different cell line would result in unique PPIs and kinetics.

Wild-type Rho GTPases were used in this experiment because I wanted to directly compare how GTPase activation states affected their interactions. Constitutively active or dominant negative mutant GTPases could have been used instead. However, they are structurally different than their wild-type counterparts, thereby making their kinetic analyses outside the scope of this study (Davis et al., 2013; Kumawat et al., 2017). Follow-up experiments of GTPase mutants, particularly those that are relevant in disease, would be interesting (Porter, Papaioannou, & Malliri, 2016).

This large-scale study enabled a unique perspective into the effect of protein phosphorylation on PPIs that would not be identified in low throughput experiments. The NAPPA-SPRi data show that phosphorylation does not determine whether most protein interactions occur or not occur, but rather affects their binding kinetics. Biologically, alterations to the on- and off-rates would have significant effects in signal transduction. Faster on-rates, for example, would provide an advantage to proteins that are competing for the same binding epitope. Slower off-rates would lengthen the effect of the PPI, whether it be to activate or inhibit downstream signaling. Faster off-rates would allow

regulatory proteins to turn the signal off more quickly. The differential binding kinetics across the tested queries illustrate how proteins have different ways to regulate their interactions.

Perhaps the most interesting kinetic results that were revealed in this study were those of RAC1. RAC1 activation did not change its binding affinities but increased its on- and off-rates by ~4 orders of magnitude with 31% of the LT-targets! In contrast, this phenomenon occurred with 98% unphosphorylated proteins. These data show that RAC1, a protein that regulates numerous pathways and biological outcomes, has a high competitive edge to propagate signaling while also being able to be turned off quickly. They also highlight the importance of measuring the kinetics because techniques that only calculate the binding affinity would not have been able to detect this important method of regulation.

8.5 Acknowledgements

I thank Dr. Xiaobo Yu for his feedback regarding the activation and inactivation of the GTPase queries. Benjamin Ober-Reynolds, David Haddad, Lisa Miller, Jennifer Van Duine, and Dr. Andrea Throop assisted in plasmid cDNA preparation and validation. Dr. Lusheng Song and Dr. D. Mitchell Magee provided instrument and SPR theory training. Dr. Jin Park helped create the appealing graphical representations of the kinetic data. Ian Shoemaker helped with troubleshooting computer and instrument problems. Dr. Femina Rauf assisted in the protein interaction validation studies using the Phos-Tag™ gels. Dr. LaBaer, Dr. Mallick (Stanford University), and Dr. Wendy Fantl (Stanford

University) helped with experimental design. Funding was provided by the PS-OC grant and Dr. LaBaer.

CHAPTER 9

9 CONCLUDING REMARKS

The term “interactome” to describe the interconnecting protein network was first coined by French researchers in 1999 (Ji, 2012). Since then, large-scale interaction maps have been constructed, illuminating the complexity of the human interactome and the potential to cause a paradigm shift in personalized and precision medicine by pinpointing attractive drug targets and determining the molecular events underlying disease initiation and progression. The abundance of complex information has also stimulated the development of mathematical models to understand the system behavior of signaling pathways. Unfortunately, computational models of cells and signaling pathways have thus far been built using qualitative experiments that are either inherently biased or provide little mechanistic insight. For example, proteins that are known to be involved in disease are studied more than proteins of unknown or poorly understood function, and highly abundant proteins and stable interactions are preferentially identified by current detection methods. With a paucity of kinetic and affinity data, modelers are forced to build algorithms from *qualitative*-based data, resulting in “best guess” approximations that could miss individual, yet critical binding kinetics that regulate signaling. Calculated kinetics guided by cellular responses may be misassigned to particular signaling components or diluted across multiple proteins. Finally, experiments in which proteins-of-interest are perturbed and the cellular responses observed are essentially “black boxes” in which many of the molecular processes remain obscure (Aldridge, Burke,

Lauffenburger, & Sorger, 2006; Fumia & Martins, 2013; Heydari et al., 2017; Janes & Yaffe, 2006; Kirouac et al., 2012; Sachs, Perez, Pe'er, Lauffenburger, & Nolan, 2005b).

The human interactome has been studied primarily through high throughput methods like yeast-2-hybrid and affinity purification mass spectrometry, which rely on stable protein interactions. However, transient interactions underlie important cellular processes, thus begging the question: How much of the interactome has been missed? To help answer this question, I first adapted a current, high throughput method capable of detecting transient and stable interactions *in vivo*, NanoBRET™, to analyze protein interactions *in vitro* using proteins produced from a cell-free expression system. I then applied NanoBRET toward studying > 2500 interactions in the B cell receptor signaling pathway. Although this pathway is considered to be relatively well understood compared to other pathways (e.g., Hippo/Warts/FGF), 83% of the interactions detected by NanoBRET have not been previously reported. These data indicate that the human interactome is still largely unmapped. Unfortunately, NanoBRET, just like any other high throughput method, cannot characterize protein interactions in regards to their binding kinetics and affinities.

Herein, I described the development of methods, technology, and software to determine the binding kinetics and affinities of protein interactions with and without target phosphorylation. These included:

- Modulation of protein phosphorylation on NAPPA using phosphatases or activated B cell lysate.

- A platform that quantitatively characterizes > 400 protein-protein interactions simultaneously in < 1 hour by combining the high throughput and flexible nature of nucleic programmable protein arrays (NAPPA) with the quantitative ability of surface plasmon resonance imaging (SPRi).
- Built automated "SPRite" software capable of analyzing high throughput SPR data.

I then applied NAPPA-SPRi to study the kinetics and affinities of > 12,000 protein interactions in the B cell receptor signaling pathway under different protein phosphorylation and GTPase activation states. NAPPA-SPRi detected 66% of known interactions and 401 novel interactions, 41% of which were also observed with NanoBRET. Notably, NAPPA-SPRi detected 55% more interactions than NanoBRET. NAPPA-SPRi data show that phosphorylation does not determine whether most (84%) protein interactions occur or not occur, but rather affects their binding kinetics and affinities, which appear to be uniquely modulated across proteins. Increased interactions upon RAC1 and RHOA GTPase activation align well with current understanding that active GTPases mediate most downstream pathways. RAC1 activation with nonhydrolyzable GTP-γS minimally affected its binding affinities but increased its overall on- and off-rates by ~4 orders of magnitude. This phenomenon was observed with 31% of targets treated with activated B cell lysate capable of phosphorylation and 98% of unphosphorylated targets. This underscores the importance of measuring kinetics as equilibrium assays that simply measure binding affinities would not have detected this important method of regulation.

Four novel interactions that had significantly altered binding profiles to targets before and after lysate treatment were validated as phosphorylation-mediated events using SDS-PAGE and Western blot analyses. One of these interactions included two proteins that are frequently mutated in cancer, PI3K-MYC, but were not previously known to physically associate with each other. Instead, I show that PI3K directly phosphorylates MYC at serine 62, a phosphosite that is known to increase the half-life of MYC. Follow-up experiments are necessary to determine the phosphorylation sites of the other three interactions and what, if any, the biological effects of these interactions are *in vivo*.

Target proteins are all represented equally on NAPPA-SPRi, thereby allowing interactions important in signal transduction to be identified that would otherwise be masked by interactions of highly abundant proteins *in vivo*. However, interactions *in vivo* are regulated in part by protein abundance. Thus, the abundance of proteins in the B cell receptor signaling pathway in four B cell lines, Ramos RA-1, Jeko-1, Rec-1, and Toledo, was determined with mass spectrometry analyses (data not shown). NAPPA-SPRi and mass spectrometry data are currently being incorporated into a virtual B cell model. Additional experiments will be required to determine whether the model can accurately predict proteomic – and possibly phenotypic – changes as the result of specific stimuli. It will be the *first* model of any signaling pathway built from large-scale, experimentally-produced kinetic data.

An accurate cell model has far-reaching consequences in medicine and science. It would – theoretically – be able to delineate the effects of genetic mutations on disease

pathology specific to the individual or tumor. Central protein hubs for signal transduction in homeostasis and disease would be identified. Alternate signaling pathways in treatment resistance would be known. With this knowledge in hand, drugs could be designed smarter and patients could be treated more effectively based on their unique genetic background, thereby resulting in a paradigm shift in personalized and precision medicine. Synthetic biologists interested in sustainable energy could re-engineer cyanobacteria to become highly efficient fuel producers. Tissue engineers could identify the components essential in cell-to-cell variability and signaling crosstalk. And virologists could use the generated information to develop safer and more effective vaccines for various diseases and bioterrorism incidents.

In this thesis, NAPPA-SPRi was applied toward studying protein-protein interactions in the B cell receptor signaling pathway, but it could be used to study any interactions as long as the plasmid cDNA can be constructed. Since the proteins are produced using a cell-free expression system from numerous sources (e.g., human, wheat germ, *E. coli*), NAPPA circumvents disadvantages that are inherent in expressing proteins *in vivo* (e.g., toxic proteins) or in a non-homologous system. It can be imagined that NAPPA-SPRi could be expanded to drug screening and studying other signaling pathways, host-pathogen protein-protein interactions, and the effect of protein mutations on protein interactions. The array format would also be compatible with screening antibodies or validating protein functionality. Potential substrates of kinase phosphorylation could be screened. Altered binding based on the location of the fusion tag on NAPPA-SPRi may also assist in identifying binding epitopes. Membrane proteins

are not likely to be folded correctly on NAPPA due to their native hydrophobic environment, but their hydrophilic intracellular or extracellular protein domains could be displayed instead. I demonstrate that the phosphorylation of NAPPA proteins can be altered in a B cell-specific manner using lysate from activated B cells, but the methods developed and described herein could easily be used to study the effect of phosphorylation patterns from other cell types on protein interactions. Other post translational modifications could be studied with NAPPA-SPRi as well. Proteins displayed by traditional fluorescence-based NAPPA have been citrullinated and AMPylated for autoantibody and protein interaction studies, respectively, by adding peptidyl arginine deiminase 2 and AMPylators to the array (Karthikeyan et al., 2016; X. B. Yu & LaBaer, 2015).

REFERENCES

Adams, S. T., & Miller, S. C. (2014). Beyond D-luciferin: expanding the scope of bioluminescence imaging in vivo. *Current Opinion in Chemical Biology, 21*, 112-120. doi:10.1016/j.cbpa.2014.07.003

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., & Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nature Genetics, 37*(1), 73-76. doi:10.1038/ng1482

Ahmed, Z., Lin, C. C., Suen, K. M., Melo, F. A., Levitt, J. A., Suhling, K., & Ladbury, J. E. (2013). Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity. *Journal of Cell Biology, 200*(4), 493-504. doi:10.1083/jcb.201204106

Al-Alwan, M., Hou, S., Zhang, T. T., Makondo, K., & Marshall, A. J. (2010a). Bam32/DAPP1 promotes B cell adhesion and formation of polarized conjugates with T cells. *Journal of Immunology, 184*(12), 6961-6969. doi:10.4049/jimmunol.0904176

Al-Alwan, M., Hou, S., Zhang, T. T., Makondo, K., & Marshall, A. J. (2010b). Bam32/DAPP1 Promotes B Cell Adhesion and Formation of Polarized Conjugates with T Cells. *Journal of Immunology, 184*(12), 6961-6969. doi:10.4049/jimmunol.0904176

Alberts, B. (2015). *Molecular biology of the cell* (Sixth edition. ed.). New York, NY: Garland Science, Taylor and Francis Group.

Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., & Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nature Cell Biology, 8*(11), 1195-1203. doi:10.1038/ncb1497

Allam, A., Niiro, H., Clark, E. A., & Marshall, A. J. (2004). The adaptor protein Bam32 regulates Rac1 activation and actin remodeling through a phosphorylation-dependent mechanism. *Journal of Biological Chemistry, 279*(38), 39775-39782. doi:10.1074/jbc.M403367200

Andersson, K., Areskoug, D., & Hardenborg, E. (1999). Exploring buffer space for molecular interactions. *Journal of Molecular Recognition, 12*(5), 310-315. doi:Doi 10.1002/(Sici)1099-1352(199909/10)12:5<310::Aid-Jmr470>3.0.Co;2-5

Andreu, N., Zelmer, A., & Wiles, S. (2011). Noninvasive biophotonic imaging for studies of infectious disease. *Fems Microbiology Reviews, 35*(2), 360-394. doi:10.1111/j.1574-6976.2010.00252.x

Aronsson, C., Danmark, S., Zhou, F., Oberg, P., Enander, K., Su, H. B., & Aili, D. (2015). Self-sorting heterodimeric coiled coil peptides with defined and tuneable self-assembly properties. *Scientific Reports, 5*. doi:ARTN 14063

10.1038/srep14063

Asano, T., Kaneko, H., Terada, T., Kasahara, Y., Fukao, T., Kasahara, K., & Kondo, N. (2004). Molecular analysis of B-cell differentiation in selective or partial IgA deficiency. *Clinical and Experimental Immunology, 136*(2), 284-290. doi:10.1111/j.1365-2249.2004.02440.x

Austin, B. P., Nallamsetty, S., & Waugh, D. S. (2009). Hexahistidine-tagged maltose-binding protein as a fusion partner for the production of soluble recombinant proteins in Escherichia coli. *Methods Mol Biol, 498*, 157-172. doi:10.1007/978-1-59745-196-3_11

Bassani-Sternberg, M., Braunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., . . . Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications, 7*. doi:ARTN 13404

10.1038/ncomms13404

Berra, E., Diaz-Meco, M. T., & Moscat, J. (1998). The activation of p38 and apoptosis by the inhibition of Erk is antagonized by the phosphoinositide 3-kinase Akt pathway (vol 273, pg 10792, 1998). *Journal of Biological Chemistry, 273*(26), 16630-16630.

Bertics, P. J., & Gill, G. N. (1985). Self-Phosphorylation Enhances the Protein-Tyrosine Kinase-Activity of the Epidermal Growth-Factor Receptor. *Journal of Biological Chemistry, 260*(27), 4642-4647.

Billian, G., Mondiere, P., Berard, M., Bella, C., & Defrance, T. (1997). Antigen receptor-induced apoptosis of human germinal center B cells is targeted to a centrocytic subset. *European Journal of Immunology, 27*(2), 405-414. doi:DOI 10.1002/eji.1830270210

Bos, J. L., Rehmann, H., & Wittinghofer, A. (2007). GEFs and GAPs: Critical elements in the control of small G proteins. *Cell, 129*(5), 865-877. doi:10.1016/j.cell.2007.05.018

Bounab, Y., Getahun, A., Cambier, J. C., & Daeron, M. (2013). Phosphatase regulation of immunoreceptor signaling in T cells, B cells and mast cells. *Current Opinion in Immunology, 25*(3), 313-320. doi:10.1016/j.coi.2013.04.001

Braun, P., & Gingras, A. C. (2012). History of protein-protein interactions: From egg-white to complex networks. *Proteomics, 12*(10), 1478-1498. doi:10.1002/pmic.201100563

Brown, N. E., Blumer, J. B., & Hepler, J. R. (2015). Bioluminescence resonance energy transfer to detect protein-protein interactions in live cells. *Methods Mol Biol, 1278*, 457-465. doi:10.1007/978-1-4939-2425-7_30

Bruce, J. E. (2012). In vivo protein complex topologies: Sights through a cross-linking lens. *Proteomics, 12*(10), 1565-1575. doi:10.1002/pmic.201100516

Bunda, S., Heir, P., Srikumar, T., Cook, J. D., Burrell, K., Kano, Y., . . . Ohh, M. (2014). Src promotes GTPase activity of Ras via tyrosine 32 phosphorylation. *Proc Natl Acad Sci U S A, 111*(36), E3785-3794. doi:10.1073/pnas.1406559111

Carlberg, C. (2016). *Mechanisms of gene regulation*. New York, NY: Springer Berlin Heidelberg.

Castellano, E., & Downward, J. (2011). RAS Interaction with PI3K: More Than Just Another Effector Pathway. *Genes Cancer, 2*(3), 261-274. doi:10.1177/1947601911408079

Chakraborty, S., & Rao, B. J. (2012). A Measure of the Promiscuity of Proteins and Characteristics of Residues in the Vicinity of the Catalytic Site That Regulate Promiscuity. *Plos One, 7*(2). doi:ARTN e32011

10.1371/journal.pone.0032011

Cheung, L. W. T., Walkiewicz, K. W., Besong, T., Hawke, D., Arold, S. T., & Mills, G. B. (2015). Regulation of the PI3K pathway through a p85 alpha monomer-homodimer equilibrium. *Cancer Research, 75*. doi:10.1158/1538-7445.Am2015-4715

Cinar, M., Hamedani, F., Mo, Z. C., Cinar, B., Amin, H. M., & Alkan, S. (2013). Bruton tyrosine kinase is commonly overexpressed in mantle cell lymphoma and its attenuation by Ibrutinib induces apoptosis. *Leukemia Research, 37*(10), 1271-1277. doi:10.1016/j.leukres.2013.07.028

Coffey, J., Hodgson, D. C., & Gospodarowicz, M. K. (2003). Therapy of non-Hodgkin's lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging, 30*, S28-S36. doi:10.1007/s00259-003-1157-6

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in Microbiology, 5*. doi:ARTN 63

10.3389/fmicb.2014.00063

Cotton, M., Boulay, P. L., Houndolo, T., Vitale, N., Pitcher, J. A., & Claing, A. (2007). Endogenous ARF6 interacts with Rac1 upon angiotensin II stimulation to regulate membrane ruffling and cell migration. *Molecular Biology of the Cell, 18*(2), 501-511. doi:10.1091/mbc.E06-06-0567

Cowley, D. O., & Graves, B. J. (2000). Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. *Genes & Development, 14*(3), 366-376.

Cunningham-Rundles, C. (2012). Human B cell defects in perspective. *Immunologic Research, 54*(1-3), 227-232. doi:10.1007/s12026-012-8318-2

Dal Porto, J. M., Gauld, S. B., Merrell, K. T., Mills, D., Pugh-Bernard, A. E., & Cambier, J. (2004). B cell antigen receptor signaling 101. *Molecular Immunology, 41*(6-7), 599-613. doi:10.1016/j.molimm.2004.04.008

Daunert, S., & Deo, S. K. (2006). *Photoproteins in bioanalysis*. Weinheim: Wiley-VCH.

Davids, M. S. (2017). Targeting BCL-2 in B-cell lymphomas. *Blood, 130*(9), 1081-1088. doi:10.1182/blood-2017-04-737338

Davies, E. G., & Thrasher, A. J. (2010). Update on the hyper immunoglobulin M syndromes. *British Journal of Haematology, 149*(2), 167-180. doi:10.1111/j.1365-2141.2010.08077.x

Davis, M. J., Ha, B. H., Holman, E. C., Halaban, R., Schlessinger, J., & Boggon, T. J. (2013). RAC1(P29S) is a spontaneously activating cancer-associated GTPase. *Proceedings of the National Academy of Sciences of the United States of America, 110*(3), 912-917. doi:10.1073/pnas.1220895110

De Crescenzo, G., Litowski, J. R., Hodges, R. S., & O'Connor-McCourt, M. D. (2003). Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: Effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry, 42*(6), 1754-1763. doi:10.1021/bi0268450

de Gorter, D. J. J., Vos, J. C. M., Pals, S. T., & Spaargaren, M. (2007). The B cell antigen receptor controls AP-1 and NFAT activity through Ras-mediated activation of Ral. *Journal of Immunology, 178*(3), 1405-1414. doi:DOI 10.4049/jimmunol.178.3.1405

de Leeuw, R., Neefjes, J., & Michalides, R. (2011). A role for estrogen receptor phosphorylation in the resistance to tamoxifen. *Int J Breast Cancer, 2011*, 232435. doi:10.4061/2011/232435

Deane, C. M., Salwinski, L., Xenarios, I., & Eisenberg, D. (2002). Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics, 1*(5), 349-356.

DeGeer, J., Boudeau, J., Schmidt, S., Bedford, F., Lamarche-Vane, N., & Debant, A. (2013). Tyrosine Phosphorylation of the Rho Guanine Nucleotide Exchange Factor Trio Regulates Netrin-1/DCC-Mediated Cortical Axon Outgrowth. *Molecular and Cellular Biology, 33*(4), 739-751. doi:10.1128/Mcb.01264-12

Deng, J., Isik, E., Fernandes, S. M., Brown, J. R., Letai, A., & Davids, M. S. (2015). Ibrutinib Therapy Increases BCL-2 Dependence and Enhances Sensitivity to Venetoclax in CLL. *Blood, 126*(23).

Diez, P., Gonzalez-Gonzalez, M., Lourido, L., Degano, R. M., Ibarrola, N., Casado-Vela, J., . . . Fuentes, M. (2015). NAPPA as a Real New Method for Protein Microarray Generation. *Microarrays (Basel), 4*(2), 214-227. doi:10.3390/microarrays4020214

Ding, J. Y., Takano, T., Gao, S. Y., Han, W. H., Noda, C., Yanagi, S., & Yamamura, H. (2000). Syk is required for the activation of Akt survival pathway in B cells exposed to oxidative stress. *Journal of Biological Chemistry, 275*(40), 30873-30877. doi:DOI 10.1074/jbc.M004813200

Dvorsky, R., & Ahmadian, M. R. (2004). Always look on the bright site of Rho: structural implications for a conserved intermolecular interface. *Embo Reports, 5*(12), 1130-1136. doi:10.1038/sj.embor.7400293

Edwards, P. R., Gill, A., Pollardknight, D. V., Hoare, M., Buckle, P. E., Lowe, P. A., & Leatherbarrow, R. J. (1995). Kinetics of Protein-Protein Interactions at the Surface of an Optical Biosensor. *Analytical Biochemistry, 231*(1), 210-217. doi:DOI 10.1006/abio.1995.1522

Engelke, M., Oellerich, T., Dittmann, K., Hsiao, H. H., Urlaub, H., Serve, H., . . . Wienands, J. (2013). Cutting Edge: Feed-Forward Activation of Phospholipase C gamma 2 via C2 Domain-Mediated Binding to SLP65. *Journal of Immunology, 191*(11), 5354-5358. doi:10.4049/jimmunol.1301326

England, C. G., Luo, H. M., & Cai, W. B. (2015). HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjugate Chemistry, 26*(6), 975-986. doi:10.1021/acs.bioconjchem.5b00191

Fabbro, D., Cowan-Jacob, S. W., & Moebitz, H. (2015). Ten things you should know about protein kinases: IUPHAR Review 14. *British Journal of Pharmacology, 172*(11), 2675-2700. doi:10.1111/bph.13096

Fischer, A. (2000). Severe combined immunodeficiencies (SCID). *Clinical and Experimental Immunology, 122*(2), 143-149. doi:DOI 10.1046/j.1365-2249.2000.01359.x

Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., Maclntyre, M. F., . . . Dis, G. B. (2015). The Global Burden of Cancer 2013 Global Burden of Disease Cancer Collaboration. *Jama Oncology, 1*(4), 505-527. doi:10.1001/jamaoncol.2015.0735

Foote, J., & Milstein, C. (1991). Kinetic Maturation of an Immune-Response. *Nature, 352*(6335), 530-532. doi:DOI 10.1038/352530a0

Fournier, E., Isakoff, S. J., Ko, K., Cardinale, C. J., Inghirami, G. G., Li, Z., . . . Skolnik, E. Y. (2003). The B cell SH2/PH domain-containing adaptor Bam32/DAPP1 is required for T cell-independent II antigen responses. *Current Biology, 13*(21), 1858-1866. doi:10.1016/j.cub.2003.09.034

Fried, A. J., & Bonilla, F. A. (2009). Pathogenesis, Diagnosis, and Management of Primary Antibody Deficiencies and Infections. *Clinical Microbiology Reviews, 22*(3), 396-414. doi:10.1128/Cmr.00001-09

Fumia, H. F., & Martins, M. L. (2013). Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. *Plos One, 8*(7). doi:ARTN e69008

10.1371/journal.pone.0069008

Gan, W. X., & Roux, B. (2009). Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins-Structure Function and Bioinformatics, 74*(4), 996-1007. doi:10.1002/prot.22209

Gao, B. X., Lee, S. M., & Fang, D. Y. (2006). The tyrosine kinase C-Abl protects c-Jun from ubiquitination-mediated degradation in T cells. *Journal of Biological Chemistry, 281*(40), 29711-29718. doi:10.1074/jbc.M604596200

Geahlen, R. L. (2009). Syk and pTyr'd: Signaling through the B cell antigen receptor. *Biochimica Et Biophysica Acta-Molecular Cell Research, 1793*(7), 1115-1127. doi:10.1016/j.bbamcr.2009.03.004

Geertz, M., Shore, D., & Maerkl, S. J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences of the United States of America, 109*(41), 16540-16545. doi:10.1073/pnas.1206011109

Giles, J. R., Kashgarian, M., Koni, P. A., & Shlomchik, M. J. (2015). B Cell-Specific MHC Class II Deletion Reveals Multiple Nonredundant Roles for B Cell Antigen

Presentation in Murine Lupus. *Journal of Immunology, 195*(6), 2571-2579. doi:10.4049/jimmunol.1500792

Gonzalez, M. W., & Kann, M. G. (2012). Chapter 4: Protein Interactions and Disease. *Plos Computational Biology, 8*(12). doi:ARTN e1002819

10.1371/journal.pcbi.1002819

Goodrich, J. A., & Kugel, J. F. (2007). *Binding and kinetics for molecular biologists*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Gottumukkala, R. V. S. R. K., Waterman, E. A., Herd, L. M., Gawkrodger, D. J., Watson, P. E., Weetman, A. P., & Kemp, E. H. (2003). Autoantibodies in vitiligo patients recognize multiple domains of the melanin-concentrating hormone receptor. *Journal of Investigative Dermatology, 121*(4), 765-770. doi:DOI 10.1046/j.1523-1747.2003.12494.x

Grewal, T., Koese, M., Tebar, F., & Enrich, C. (2011). Differential Regulation of RasGAPs in Cancer. *Genes Cancer, 2*(3), 288-297. doi:10.1177/1947601911407330

Guo, F. K., Velu, C. S., Grimes, H. L., & Zheng, Y. (2009). Rho GTPase Cdc42 is essential for B-lymphocyte development and activation. *Blood, 114*(14), 2909-2916. doi:10.1182/blood-2009-04-214676

Gutierrez, T., Halcomb, K. E., Coughran, A. J., Li, Q. Z., & Satterthwaite, A. B. (2010). Separate checkpoints regulate splenic plasma cell accumulation and IgG autoantibody production in Lyn-deficient mice. *European Journal of Immunology, 40*(7), 1897-1905. doi:10.1002/eji.200940043

Haas, A., Zimmermann, K., & Oxenius, A. (2011). Antigen-Dependent and -Independent Mechanisms of T and B Cell Hyperactivation during Chronic HIV-1 Infection. *Journal of Virology, 85*(23), 12102-12113. doi:10.1128/Jvi.05607-11

Hall, M. P., Unch, J., Binkowski, B. F., Valley, M. P., Butler, B. L., Wood, M. G., . . . Wood, K. V. (2012). Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *Acs Chemical Biology, 7*(11), 1848-1857. doi:10.1021/cb3002478

Hampe, C. S. (2012). B Cell in Autoimmune Diseases. *Scientifica (Cairo), 2012*. doi:10.6064/2012/215308

Han, W. H., Takano, T., He, J. S., Ding, J. Y., Gao, S. Y., Noda, C., . . . Yamamura, H. (2001). Role of BLNK in oxidative stress signaling in B cells. *Antioxidants & Redox Signaling, 3*(6), 1065-1073. doi:Doi 10.1089/152308601317203576

Harbers, M. (2014). Wheat germ systems for cell-free protein expression. *Febs Letters, 588*(17), 2762-2773. doi:10.1016/j.febslet.2014.05.061

Hartmann, M., Roeraade, J., Stoll, D., Templin, M., & Joos, T. (2009). Protein microarrays for diagnostic assays. *Analytical and Bioanalytical Chemistry, 393*(5), 1407-1416. doi:10.1007/s00216-008-2379-z

Heasman, S. J., & Ridley, A. J. (2008). Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nature Reviews Molecular Cell Biology, 9*(9), 690-701. doi:10.1038/nrm2476

Heinrich, R., Neel, B. G., & Rapoport, T. A. (2002). Mathematical models of protein kinase signal transduction. *Molecular Cell, 9*(5), 957-970. doi:Doi 10.1016/S1097-2765(02)00528-2

Helmerhorst, E., Chandler, D. J., Nussio, M., & Mamotte, C. D. (2012). Real-time and Label-free Bio-sensing of Molecular Interactions by Surface Plasmon Resonance: A Laboratory Medicine Perspective. *Clin Biochem Rev, 33*(4), 161-173.

Hemsath, L., Dvorsky, R., Fiegen, D., Carlier, M. F., & Ahmadian, M. R. (2005). An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Molecular Cell, 20*(2), 313-324. doi:10.1016/j.molcel.2005.08.036

Henriques, D., Villaverde, A. F., Rocha, M., Saez-Rodriguez, J., & Banga, J. R. (2017). Data-driven reverse engineering of signaling pathways using ensembles of dynamic models. *Plos Computational Biology, 13*(2), e1005379. doi:10.1371/journal.pcbi.1005379

Herrero-Gonzalez, S., & Di Cristofano, A. (2011). New Routes to Old Places: PIK3R1 and PIK3R2 Join PIK3CA and PTEN as Endometrial Cancer Genes. *Cancer Discovery, 1*(2), 106-107. doi:10.1158/2159-8290.Cd-11-0116

Heydari, T., Heidari, M., Mashinchian, O., Wojcik, M., Xu, K., Dalby, M. J., . . . Ejtehadi, M. R. (2017). Development of a Virtual Cell Model to Predict Cell Response to Substrate Topography. *Acs Nano, 11*(9), 9084-9092. doi:10.1021/acsnano.7b03732

Hiratsuka, T., Takei, Y., Ohmori, R., Imai, Y., Ozeki, M., Tamaki, K., . . . Tsuruyama, T. (2016). ZFP521 contributes to pre-B-cell lymphomagenesis through modulation of the pre-B-cell receptor signaling pathway. *Oncogene, 35*(25), 3227-3238. doi:10.1038/onc.2015.385

Holding, A. N. (2015). XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods, 89*, 54-63. doi:10.1016/j.ymeth.2015.06.010

Homola, J. i., & Dostálek, J. (2006). *Surface plasmon resonance based sensors*. Berlin ; New York: Springer.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research, 43*(D1), D512-D520. doi:10.1093/nar/gku1267

Huang, H., & Bader, J. S. (2009). Precision and recall estimates for two-hybrid screens. *Bioinformatics, 25*(3), 372-378. doi:10.1093/bioinformatics/btn640

Husberg, C., Agnetti, G., Holewinski, R. J., Christensen, G., & Van Eyk, J. E. (2012). Dephosphorylation of cardiac proteins in vitro - a matter of phosphatase specificity. *Proteomics, 12*(7), 973-978. doi:10.1002/pmic.201100116

Ingley, E. (2012). Functions of the Lyn tyrosine kinase in health and disease. *Cell Communication and Signaling, 10*. doi:Artn 21

10.1186/1478-811x-10-21

Institute of Medicine (U.S.). Committee to Review Adverse Effects of Vaccines., & Stratton, K. R. (2012). *Adverse effects of vaccines : evidence and causality*. Washington, D.C.: National Academies Press.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America, 98*(8), 4569-4574. doi:DOI 10.1073/pnas.061034498

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., . . . Sakaki, Y. (2000). Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proceedings of the National Academy of Sciences of the United States of America, 97*(3), 1143-1147. doi:DOI 10.1073/pnas.97.3.1143

Jagannathan, M., McDonnell, M., Liang, Y., Hasturk, H., Hetzel, J., Rubin, D., . . . Nikolajczyk, B. S. (2010). Toll-like receptors regulate B cell cytokine production in patients with diabetes. *Diabetologia, 53*(7), 1461-1471. doi:10.1007/s00125-010-1730-z

Janes, K. A., & Yaffe, M. B. (2006). Data-driven modelling of signal-transduction networks. *Nature Reviews Molecular Cell Biology, 7*(11), 820-828. doi:10.1038/nrm2041

Ji, S. (2012). *Molecular theory of the living cell : concepts, molecular mechanisms, and biomedical applications*. New York: Springer.

Jiang, K., Zhong, B., Ritchey, C., Gilvary, D. L., Hong-Geller, E., Wei, S., & Djeu, J. Y. (2003). Regulation of Akt-dependent cell survival by Syk and Rac. *Blood, 101*(1), 236-244. doi:10.1182/blood-2002-04-1251

Jokiranta, T. S., Westin, J., Nilsson, U. R., Nilsson, B., Hellwage, J., Lofas, S., . . . Meri, S. (2001). Complement C3b interactions studied with surface plasmon resonance technique. *International Immunopharmacology, 1*(3), 495-506. doi:Doi 10.1016/S1567-5769(00)00042-4

Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., . . . Malmqvist, M. (1991). Real-Time Biospecific Interaction Analysis Using Surface-Plasmon Resonance and a Sensor Chip Technology. *Biotechniques, 11*(5), 620-&.

Joshi, A. A., Peczuh, M. W., Kumar, C. V., & Rusling, J. F. (2014). Ultrasensitive carbohydrate-peptide SPR imaging microarray for diagnosing IgE mediated peanut allergy (vol 139, pg 5728, 2014). *Analyst, 139*(24), 6589-6589. doi:10.1039/c4an90090a

Justement, L. B., & Siminovitch, K. A. (2000). Signal transduction and the coordination of B lymphocyte development and function I. Transduction of BCR signals from the cell membrane to the nucleus. Introduction. *Curr Top Microbiol Immunol, 245*(1), V-X.

Kabak, S., Skaggs, B. J., Gold, M. R., Affolter, M., West, K. L., Foster, M. S., . . . Clark, M. R. (2002). The direct recruitment of BLNK to immunoglobulin alpha couples the B-Cell antigen receptor to distal signaling pathways. *Molecular and Cellular Biology, 22*(8), 2524-2535. doi:10.1128/Mcb.22.8.2524-2535.2002

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research, 45*(D1), D353-D361. doi:10.1093/nar/gkw1092

Kapust, R. B., & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science, 8*(8), 1668-1674. doi:10.1110/ps.8.8.1668

Karlsson, M., Ekeroth, J., Elwing, H., & Carlsson, U. (2005). Reduction of irreversible protein adsorption on solid surfaces by protein engineering for increased stability. *Journal of Biological Chemistry, 280*(27), 25558-25564. doi:DOI 10.1074/jbc.M503665200

Karp, G., & Patton, J. G. (2013). *Cell and molecular biology : concepts and experiments* (7th ed.). Hoboken, NJ: John Wiley.

Karthikeyan, K., Barker, K., Tang, Y. Y., Kahn, P., Wiktor, P., Brunner, A., . . . Qiu, J. (2016). A Contra Capture Protein Array Platform for Studying Post-translationally Modified (PTM) Auto-antigenomes. *Molecular & Cellular Proteomics, 15*(7), 2324-2337. doi:10.1074/mcp.M115.057661

Keasey, S. L., Natesan, M., Pugh, C., Kamata, T., Wuchty, S., & Ulrich, R. G. (2016). Cell-free Determination of Binary Complexes That Comprise Extended Protein-Protein Interaction Networks of Yersinia pestis. *Molecular & Cellular Proteomics, 15*(10), 3220-3232. doi:10.1074/mcp.M116.059337

Kéri, G. r., & Tóth, I. n. (2003). *Molecular pathomechanisms and new trends in drug research*. London ; New York: Taylor & Francis.

Kiel, C., & Serrano, L. (2009). Cell Type-Specific Importance of Ras-c-Raf Complex Association Rate Constants for MAPK Signaling. *Science Signaling, 2*(81). doi:ARTN ra38

10.1126/scisignal.2000397

Kile, B. T., Nicola, N. A., & Alexander, W. S. (2001). Negative regulators of cytokine signaling. *International Journal of Hematology, 73*(3), 292-298. doi:Doi 10.1007/Bf02981953

Kim, H. J., Kim, J. G., Moon, M. Y., Park, S. H., & Park, J. B. (2014). I kappa B Kinase gamma/Nuclear Factor-kappa B-Essential Modulator (IKK gamma/NEMO) Facilitates RhoA GTPase Activation, which, in Turn, Activates Rho-associated Kinase (ROCK) to Phosphorylate IKK beta in Response to Transforming Growth Factor (TGF)-beta 1. *Journal of Biological Chemistry, 289*(3), 1429-1440. doi:10.1074/jbc.M113.520130

Kim, J. M., Kim, M. Y., Lee, K., & Jeong, D. (2016). Distinctive and selective route of PI3K/PKC alpha-PKC delta/RhoA-Rac1 signaling in osteoclastic cell migration. *Molecular and Cellular Endocrinology, 437*(C), 261-267. doi:10.1016/j.mce.2016.08.042

King, K., Sheikh, M. F., Cuthbert, A. P., Fisher, S. A., Onnie, C. M., Mirza, M. M., . . . Mathew, C. G. (2006). Mutation, selection, and evolution of the Crohn disease susceptibility gene CARD15. *Human Mutation, 27*(1), 44-54. doi:10.1002/humu.20264

Kirouac, D. C., Saez-Rodriguez, J., Swantek, J., Burke, J. M., Lauffenburger, D. A., & Sorger, P. K. (2012). Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. *Bmc Systems Biology, 6*. doi:Artn 29

10.1186/1752-0509-6-29

Kitchen, J., Saunders, R. E., & Warwicker, J. (2008). Charge environments around phosphorylation sites in proteins. *Bmc Structural Biology, 8*. doi:Artn 19

10.1186/1472-6807-8-19

Kiyono, M., Kato, J., Kataoka, T., Kaziro, Y., & Satoh, T. (2000). Stimulation of Ras guanine nucleotide exchange activity of Ras-GRF1/CDC25(Mm) upon tyrosine phosphorylation by the Cdc42-regulated kinase ACK1. *Journal of Biological Chemistry, 275*(38), 29788-29793. doi:DOI 10.1074/jbc.M001378200

Kline, J. B., Moore, D. J., & Clevenger, C. V. (2001). Activation and association of the Tec tyrosine kinase with the human prolactin receptor: Mapping of a Tec/Vav1-receptor binding site. *Molecular Endocrinology, 15*(5), 832-841. doi:DOI 10.1210/me.15.5.832

Kolln, J., Ren, H. M., Da, R. R., Zhang, Y. P., Spillner, E., Olek, M., . . . Qin, Y. F. (2006). Triosephosphate isomerase- and glyceraldehyde-3-phosphate dehydrogenase-reactive autoantibodies in the cerebrospinal fluid of patients with multiple sclerosis. *Journal of Immunology, 177*(8), 5652-5658.

Koretzky, G. A., Abtahian, F., & Silverman, M. A. (2006). SLP76 and SLP65: complex regulation of signalling in lymphocytes and beyond. *Nat Rev Immunol, 6*(1), 67-78. doi:10.1038/nri1750

Kosobokova, E. N., Skrypnik, K. A., & Kosorukov, V. S. (2016). Overview of Fusion Tags for Recombinant Proteins. *Biochemistry-Moscow, 81*(3), 187-200. doi:10.1134/S0006297916030019

Krishnamoorthy, G., Carlen, E. T., Bomer, J. G., Wijnperle, D., deBoer, H. L., van den Berg, A., & Schasfoort, R. B. M. (2010). Electrokinetic label-free screening chip: a marriage of multiplexing and high throughput analysis using surface plasmon resonance imaging. *Lab on a Chip, 10*(8), 986-990. doi:10.1039/c000705f

Kumawat, A., Chakrabarty, S., & Kulkarni, K. (2017). Nucleotide Dependent Switching in Rho GTPase: Conformational Heterogeneity and Competing Molecular Interactions. *Scientific Reports, 7*. doi:ARTN 45829

10.1038/srep45829

Kuppers, R., & Hansmann, M. L. (2005). The Hodgkin and Reed/Sternberg cell. *International Journal of Biochemistry & Cell Biology, 37*(3), 511-517. doi:10.1016/j.biocel.2003.10.025

Kurosaki, T., & Tsukada, S. (2000). BLNK: connecting Syk and Btk to calcium signals. *Immunity, 12*(1), 1-5.

Kurzrock, R., & Markman, M. (2008). *Targeted cancer therapy*. Totowa, N.J.: Humana Press.

Li, G. P., & Zhang, X. J. C. (2004). GTP hydrolysis mechanism of Ras-like GTPases. *Journal of Molecular Biology, 340*(5), 921-932. doi:DOI 10.1016/j.jmb.2004.06.007

Li, H. L., Davis, W. W., Whiteman, E. L., Birnbaum, M. J., & Pure, E. (1999). The tyrosine kinases Syk and Lyn exert opposing effects on the activation of protein kinase Akt PKB in B lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America, 96*(12), 6890-6895. doi:DOI 10.1073/pnas.96.12.6890

Li, R., Rezk, A., Miyazaki, Y., Hilgenberg, E., Touil, H., Shen, P., . . . Team, C. B. C. M. (2015). Proinflammatory GM-CSF-producing B cells in multiple sclerosis and B cell depletion therapy. *Science Translational Medicine, 7*(310). doi:ARTN 310ra166

10.1126/scitranslmed.aab4176

Li, S. G., Couvillon, A. D., Brasher, B. B., & Van Etten, R. A. (2001). Tyrosine phosphorylation of Grb2 by Bcr/Abl and epidermal growth factor receptor: a novel regulatory mechanism for tyrosine kinase signaling. *Embo Journal, 20*(23), 6793-6804. doi:DOI 10.1093/emboj/20.23.6793

Li, X., & Zhou, Y. (2013). *Microfluidic devices for biomedical applications*. Cambridge, UK: Woodhead Publishing.

Liedberg, B., Nylander, C., & Lundstrom, I. (1983). Surface-Plasmon Resonance for Gas-Detection and Biosensing. *Sensors and Actuators, 4*(2), 299-304. doi:Doi 10.1016/0250-6874(83)85036-7

Lindell, D. M., Berlin, A. A., Schaller, M. A., & Lukacs, N. W. (2008). B Cell Antigen Presentation Promotes Th2 Responses and Immunopathology during Chronic Allergic Lung Disease. *Plos One, 3*(9). doi:ARTN e3129

10.1371/journal.pone.0003129

Los, G. V., Zimprich, C., McDougall, M. G., Karassina, N., Learish, R., Klaubert, D. H., . . . Wood, K. (2005). The HaloTag (TM): a novel technology for cellular analysis. *Journal of Neurochemistry, 94*, 15-15.

Louis, A. G., & Gupta, S. (2014). Primary Selective IgM Deficiency: An Ignored Immunodeficiency. *Clinical Reviews in Allergy & Immunology, 46*(2), 104-111. doi:10.1007/s12016-013-8375-x

Lu, G., Zhang, Q., Huang, Y., Song, J. X., Tomaino, R., Ehrenberger, T., . . . Kaelin, W. G. (2014). Phosphorylation of ETS1 by Src Family Kinases Prevents Its Recognition by the COP1 Tumor Suppressor. *Cancer Cell, 26*(2), 222-234. doi:10.1016/j.ccr.2014.06.026

Luo, W., Mayeux, J., Gutierrez, T., Russell, L., Getahun, A., Muller, J., . . . Garrett-Sinha, L. A. (2014). A Balance between B Cell Receptor and Inhibitory Receptor Signaling Controls Plasma Cell Differentiation by Maintaining Optimal Ets1 Levels. *Journal of Immunology, 193*(2), 909-920. doi:10.4049/jimmunol.1400666

Machleidt, T., Woodroofe, C. C., Schwinn, M. K., Mendez, J., Robers, M. B., Zirnmerman, K., . . . Wood, K. V. (2015). NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *Acs Chemical Biology, 10*(8), 1797-1804. doi:10.1021/acschembio.5b00143

Mak, T. W., Saunders, M. E., & Jett, B. D. (2014). *Primer to The immune response* (2nd edition / ed.). Amsterdam ; Boston: Elsevier/AP Cell, AP Cell is an imprint of Elsevier.

Mall, M. A., & Galietta, L. J. V. (2015). Targeting ion channels in cystic fibrosis. *Journal of Cystic Fibrosis, 14*(5), 561-570. doi:10.1016/j.jcf.2015.06.002

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

Mathas, S., Hinz, M., Anagnostopoulos, I., Krappmann, D., Lietz, A., Jundt, F., . . . Scheidereit, C. (2002). Aberrantly expressed c-Jun and JunB are a hallmark of Hodgkin lymphoma calls, stimulate proliferation and synergize with NF-kappa B. *Blood, 100*(11), 742a-742a.

Mayeux, J., Skaug, B., Luo, W., Russell, L. M., John, S., Saelee, P., . . . Satterthwaite, A. B. (2015). Genetic Interaction between Lyn, Ets1, and Btk in the Control of Antibody Levels. *Journal of Immunology, 195*(5), 1955-1963. doi:10.4049/jimmunol.1500165

Meednu, N., Zhang, H. W., Owen, T., Sun, W., Wang, V., Cistrone, C., . . . Anolik, J. H. (2016). Production of RANKL by Memory B Cells: A Link Between B Cells and Bone Erosion in Rheumatoid Arthritis. *Arthritis & Rheumatology, 68*(4), 805-816. doi:10.1002/art.39489

Meyerkord, C. L., & Fu, H. (2015). *Protein-protein interactions : methods and applications* (Second edition. ed.). New York: Humana Press.

Middendorp, S., Dingjan, G. M., Maas, A., Dahlenborg, K., & Hendriks, R. W. (2003). Function of Bruton's tyrosine kinase during B cell development is partially independent of its catalytic activity. *Journal of Immunology, 171*(11), 5988-5996.

Miersch, S., Bian, X. F., Wallstrom, G., Sibani, S., Logvinenko, T., Wasserfall, C. H., . . . LaBaer, J. (2013). Serological autoantibody profiling of type 1 diabetes by protein arrays. *Journal of Proteomics, 94*, 486-496. doi:10.1016/j.jprot.2013.10.018

Milella, M., Falcone, I., Conciatori, F., Incani, U. C., Del Curatolo, A., Inzerilli, N., . . . Ciuffreda, L. (2015). PTEN: multiple functions in human malignant tumors. *Frontiers in Oncology, 5*. doi:ARTN 24

10.3389/fonc.2015.00024

Mognol, G. P., Carneiro, F. R. G., Robbs, B. K., Faget, D. V., & Viola, J. P. B. (2016). Cell cycle and apoptosis regulation by NFAT transcription factors: new roles for an old player. *Cell Death & Disease, 7*. doi:ARTN e2199

10.1038/cddis.2016.97

Mollica, L., Bessa, L. M., Hanoulle, X., Jensen, M. R., Blackledge, M., & Schneider, R. (2016). Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front Mol Biosci, 3*, 52. doi:10.3389/fmolb.2016.00052

Montgomery, D. C. (2008). *Design and analysis of experiments* (7th ed.). Hoboken, NJ: Wiley.

Murga, C., Zohar, M., Teramoto, H., & Gutkind, J. S. (2002). Rac1 and RhoG promote cell survival by the activation of PI3K and Akt, independently of their ability to stimulate JNK and NF-kappa B. *Oncogene, 21*(2), 207-216.

Murphy, K., & Weaver, C. (2016). *Janeway's immunobiology* (9th edition. ed.). New York, NY: Garland Science/Taylor & Francis Group, LLC.

Murthy, B. N., Voelcker, N. H., & Jayaraman, N. (2006). Evaluation of alpha-D-mannopyranoside glycolipid micelles-lectin interactions by surface plasmon resonance method. *Glycobiology, 16*(9), 822-832. doi:10.1093/glycob/cwl014

Nayak, R. C., Chang, K. H., Vaitinadin, N. S., & Cancelas, J. A. (2013). Rho GTPases control specific cytoskeleton-dependent functions of hematopoietic stem cells. *Immunological Reviews, 256*(1), 255-268. doi:10.1111/imr.12119

Neumann, T., Junker, H. D., Schmidt, K., & Sekul, R. (2007). SPR-based fragment screening: Advantages and applications. *Current Topics in Medicinal Chemistry, 7*(16), 1630-1642. doi:Doi 10.2174/156802607782341073

Niidome, T., Ohmori, N., Ichinose, A., Wada, A., Mihara, H., Hirayama, T., & Aoyagi, H. (1997). Binding of cationic alpha-helical peptides to plasmid DNA and their gene transfer abilities into cells. *Journal of Biological Chemistry, 272*(24), 15307-15312. doi:DOI 10.1074/jbc.272.24.15307

Nikolajczyk, B. S. (2010). B cells as under-appreciated mediators of non-auto-immune inflammatory disease. *Cytokine, 50*(3), 234-242. doi:10.1016/j.cyto.2010.02.022

Nishi, H., Hashimoto, K., & Panchenko, A. R. (2011). Phosphorylation in Protein-Protein Binding: Effect on Stability and Function. *Structure, 19*(12), 1807-1815. doi:10.1016/j.str.2011.09.021

Nishi, H., Shaytan, A., & Panchenko, A. R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in Genetics, 5*. doi:ARTN 270

10.3389/fgene.2014.00270

Oak, S. A., Zhou, Y. W., & Jarrett, H. W. (2003). Skeletal muscle signaling pathway through the dystrophin glycoprotein complex and Rac1. *Journal of Biological Chemistry, 278*(41), 39287-39295. doi:10.1074/jbc.M305551200

Oellerich, T., Gronborg, M., Neumann, K., Hsiao, H. H., Urlaub, H., & Wienands, J. (2009). SLP-65 Phosphorylation Dynamics Reveals a Functional Basis for Signal Integration by Receptor-proximal Adaptor Proteins. *Molecular & Cellular Proteomics, 8*(7), 1738-1750. doi:10.1074/mcp.M800567-MCP200

Olaru, A., Bala, C., Jaffrezic-Renault, N., & Aboul-Enein, H. Y. (2015). Surface Plasmon Resonance (SPR) Biosensors in Pharmaceutical Analysis. *Critical Reviews in Analytical Chemistry, 45*(2), 97-105. doi:10.1080/10408347.2014.881250

Onyilagha, C., Jia, P., Jayachandran, N., Hou, S., Okwor, I., Kuriakose, S., . . . Uzonna, J. E. (2015). The B Cell Adaptor Molecule Bam32 Is Critically Important for Optimal Antibody Response and Resistance to Trypanosoma congolense Infection in Mice. *Plos Neglected Tropical Diseases, 9*(4). doi:ARTN e0003716

10.1371/journal.pntd.0003716

Oshannessy, D. J., Brighamburke, M., Soneson, K. K., Hensley, P., & Brooks, I. (1993). Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface-Plasmon Resonance - Use of Nonlinear Least-Squares Analysis-Methods. *Analytical Biochemistry, 212*(2), 457-468. doi:DOI 10.1006/abio.1993.1355

OShannessy, D. J., & Winzor, D. J. (1996). Interpretation of deviations from pseudo-first-order kinetic behavior in the characterization of ligand binding by biosensor

technology. *Analytical Biochemistry, 236*(2), 275-283. doi:DOI 10.1006/abio.1996.0167

Ou, L., Matthews, M., Pang, X. D., & Zhou, H. X. (2017). The dock-and-coalesce mechanism for the association of a WASP disordered region with the Cdc42 GTPase. *Febs Journal, 284*(20), 3381-3391. doi:10.1111/febs.14197

Patil, A., Kinoshita, K., & Nakamura, H. (2010). Hub Promiscuity in Protein-Protein Interaction Networks. *International Journal of Molecular Sciences, 11*(4), 1930-1943. doi:10.3390/ijms11041930

Paul, F., Zauber, H., von Berg, L., Rocks, O., Daumke, O., & Selbach, M. (2017). Quantitative GTPase Affinity Purification Identifies Rho Family Protein Interaction Partners. *Molecular & Cellular Proteomics, 16*(1), 73-85. doi:10.1074/mcp.M116.061531

Pfleger, K. D. G., & Eidne, K. A. (2006). Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nature Methods, 3*(3), 165-+. doi:DOI 10.1038/nmeth841

Pfleger, K. D. G., Seeber, R. M., & Eidne, K. A. (2006). Bioluminescence resonance energy transfer (BRET) for the real-time detection of protein-protein interactions. *Nature Protocols, 1*(1), 337-345. doi:10.1038/nprot.2006.52

Phanstiel, D., Brumbaugh, J., Berggren, W. T., Conard, K., Feng, X., Levenstein, M. E., . . . Coon, J. J. (2008). Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America, 105*(11), 4093-4098. doi:10.1073/pnas.0710515105

Pierson, E. R., Stromnes, I. M., & Goverman, J. M. (2014). B Cells Promote Induction of Experimental Autoimmune Encephalomyelitis by Facilitating Reactivation of T Cells in the Central Nervous System. *Journal of Immunology, 192*(3), 929-939. doi:10.4049/jimmunol.1302171

Pollard, T. D. (2010). A guide to simple and informative binding assays. *Molecular Biology of the Cell, 21*(23), 4061-4067. doi:10.1091/mbc.E10-08-0683

Ponomarenko, E. A., Poverennaya, E. V., Ilgisonis, E. V., Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., . . . Archakov, A. I. (2016). The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry*. doi:Artn 7436849

10.1155/2016/7436849

Porter, A. P., Papaioannou, A., & Malliri, A. (2016). Deregulation of Rho GTPases in cancer. *Small GTPases, 7*(3), 123-138. doi:10.1080/21541248.2016.1173767

Potyrailo, R. A., & Mirsky, V. M. (2009). *Combinatorial methods for chemical and biological sensors*. New York: Springer.

Prasad, T. S. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., . . . Pandey, A. (2009). Human Protein Reference Database-2009 update. *Nucleic Acids Research, 37*, D767-D772. doi:10.1093/nar/gkn892

Puto, L. A., Pestonjamasp, K., King, C. C., & Bokoch, G. M. (2003). p21-activated kinase 1 (PAK1) interacts with the Grb2 adapter protein to couple to growth factor signaling. *Journal of Biological Chemistry, 278*(11), 9388-9393. doi:10.1074/jbc.M208414200

Qiu, Y., & Kung, H. J. (2000). Signaling network of the Btk family kinases. *Oncogene, 19*(49), 5651-5661. doi:10.1038/sj.onc.1203958

Raggiaschi, R., Gotta, S., & Terstappen, G. C. (2005). Phosphoproteome analysis. *Bioscience Reports, 25*(1-2), 33-44. doi:10.1007/s10540-005-2846-0

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., . . . LaBaer, J. (2004). Self-assembling protein microarrays. *Science, 305*(5680), 86-90. doi:DOI 10.1126/science.1097639

Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., . . . LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nature Methods, 5*(6), 535-538. doi:10.1038/Nmeth.1210

Rauf, F., Festa, F., Park, J. G., Magee, M., Eaton, S., Rinaldi, C., . . . LaBaer, J. (2018). Ibrutinib inhibition of ERBB4 reduces cell growth in a WNT5A-dependent manner. *Oncogene*. doi:10.1038/s41388-017-0079-x

Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B. J., Hon, G. C., Myers, C. L., . . . Tyers, M. (2006). Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol, 5*(4), 11. doi:10.1186/jbiol36

Reichman, M., & Penman, S. (1973). Stimulation of polypeptide initiation in vitro after protein synthesis inhibition in vivo in HeLa cells. *Proc Natl Acad Sci U S A, 70*(9), 2678-2682.

Rhoades, R., & Gaballa, S. (2017). The Role of B Cell Targeting in Chronic Graft-Versus-Host Disease. *Biomedicines, 5*(4). doi:ARTN 61

10.3390/biomedicines5040061

Richards, S., Watanabe, C., Santos, L., Craxton, A., & Clark, E. A. (2008). Regulation of B-cell entry into the cell cycle. *Immunological Reviews, 224*, 183-200. doi:DOI 10.1111/j.1600-065X.2008.00652.x

Riera, L., Lasorsa, E., Ambrogio, C., Surrenti, N., Voena, C., & Chiarle, R. (2010). Involvement of Grb2 Adaptor Protein in Nucleophosmin-Anaplastic Lymphoma Kinase (Npm-Alk) Mediated Signaling and Anaplastic Large Cell Lymphoma Growth. *Haematologica-the Hematology Journal, 95*, 168-168.

Roberts, B. E., & Paterson, B. M. (1973). Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proc Natl Acad Sci U S A, 70*(8), 2330-2334.

Rudelius, M., Pittaluga, S., Nishizuka, S., Pham, T. H. T., Fend, F., Jaffe, E. S., . . . Raffeld, M. (2006). Constitutive activation of Akt contributes to the pathogenesis and survival of mantle cell lymphoma. *Blood, 108*(5), 1668-1676. doi:10.1182/blood-2006-04-015586

Russell, L., John, S., Cullen, J., Luo, W., Shlomchik, M. J., & Garrett-Sinha, L. A. (2015). Requirement for Transcription Factor Ets1 in B Cell Tolerance to Self-Antigens. *Journal of Immunology, 195*(8), 3574-3583. doi:10.4049/jimmunol.1500776

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., & Nolan, G. P. (2005a). Causal protein-signaling networks derived from multiparameter single-cell data. *Science, 308*(5721), 523-529. doi:10.1126/science.1105809

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., & Nolan, G. P. (2005b). Causal protein-signaling networks derived from multiparameter single-cell data. *Science, 308*(5721), 523-529. doi:10.1126/science.1105809

Saez-Rodriguez, J., Alexopoulos, L. G., Zhang, M., Morris, M. K., Lauffenburger, D. A., & Sorger, P. K. (2011). Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Research, 71*(16), 5400-5411. doi:10.1158/0008-5472.CAN-10-4453

Saito, K., Tolias, K. F., Saci, A., Koon, H. B., Humphries, L. A., Scharenberg, A., . . . Carpenter, C. L. (2003). BTK regulates Ptdlns-4,5-P-2 synthesis: Importance for calcium signaling and PI3K activity. *Immunity, 19*(5), 669-678. doi:Doi 10.1016/S1074-7613(03)00297-8

Sarantopoulos, S., Blazar, B. R., Cutler, C., & Ritz, J. (2015). B Cells in Chronic Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation, 21*(1), 16-23. doi:10.1016/j.bbmt.2014.10.029

Saul, J., Petritis, B., Sau, S., Rauf, F., Gaskin, M., Ober-Reynolds, B., . . . LaBaer, J. (2014). Development of a full-length human protein production pipeline. *Protein Science, 23*(8), 1123-1135. doi:10.1002/pro.2484

Schaffhausen, B. (1995). Sh2 Domain-Structure and Function. *Biochimica Et Biophysica Acta-Reviews on Cancer, 1242*(1), 61-75. doi:Doi 10.1016/0304-419x(95)00004-Y

Scharenberg, A. M., Humphries, L. A., & Rawlings, D. J. (2007). Calcium signalling and cell-fate choice in B cells. *Nature Reviews Immunology, 7*(10), 778-789. doi:10.1038/nri2172

Schasfoort, R. B. M., & Tudos, A. J. (2008). *Handbook of surface plasmon resonance*. Cambridge, UK: RSC Pub.

Schinzel, A., Kaufmann, T., & Borner, C. (2004). Bcl-2 family members: intracellular targeting, membrane-insertion, and changes in subcellular localization. *Biochimica Et Biophysica Acta-Molecular Cell Research, 1644*(2-3), 95-105. doi:10.1016/j.bbamcr.2003.09.006

Schreiber, G., Haran, G., & Zhou, H. X. (2009). Fundamental Aspects of Protein-Protein Association Kinetics. *Chemical Reviews, 109*(3), 839-860. doi:10.1021/cr800373w

Seefeld, T. H., Halpern, A. R., & Corn, R. M. (2012). On-Chip Synthesis of Protein Microarrays from DNA Microarrays via Coupled In Vitro Transcription and Translation for Surface Plasmon Resonance Imaging Biosensor Applications. *Journal of the American Chemical Society, 134*(30), 12358-12361. doi:10.1021/ja304187r

Serber, Z., & Ferrell, J. E. (2007). Tuning bulk electrostatics to regulate protein function. *Cell, 128*(3), 441-444. doi:10.1016/j.cell.2007.01.018

Shih, V. F. S., Tsui, R., Caldwell, A., & Hoffmann, A. (2011). A single NF kappa B system for both canonical and non-canonical signaling. *Cell Research, 21*(1), 86-102. doi:10.1038/cr.2010.161

Shoemaker, B. A., Portman, J. J., & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America, 97*(16), 8868-+. doi:DOI 10.1073/pnas.160259697

Song, L. S., Wallstrom, G., Yu, X. B., Hopper, M., Van Duine, J., Steel, J., . . . Achkar, J. M. (2017). Identification of Antibody Targets for Tuberculosis Serology using High-Density Nucleic Acid Programmable Protein Arrays. *Molecular & Cellular Proteomics, 16*(4), S277-S289. doi:10.1074/mcp.M116.065953

Soo, V. W. C., Hanson-Manful, P., & Patrick, W. M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America, 108*(4), 1484-1489. doi:10.1073/pnas.1012108108

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research, 34*, D535-D539. doi:10.1093/nar/gkj109

Stechschulte, L. A., Hinds, T. D., Ghanem, S. S., Shou, W. N., Najjar, S. M., & Sanchez, E. R. (2014). FKBP51 Reciprocally Regulates GR alpha and PPAR gamma Activation via the Akt-p38 Pathway. *Molecular Endocrinology, 28*(8), 1254-1264. doi:10.1210/me.2014-1023

Stewart, C. N. (2016). *Plant biotechnology and genetics : principles, techniques, and applications* (Second edition. ed.). Hoboken, New Jersey: John Wiley & Sons Inc.

Stojanovic, I., Schasfoort, R. B. M., & Terstappen, L. W. M. M. (2014). Analysis of cell surface antigens by Surface Plasmon Resonance imaging. *Biosensors & Bioelectronics, 52*, 36-43. doi:10.1016/j.bios.2013.08.027

Stumpf, M. P. H., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., & Wiuf, C. (2008). Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6959-6964. doi:10.1073/pnas.0708078105

Suresh, L., Malyavantham, K., Shen, L., & Ambrus, J. L. (2015). Investigation of novel autoantibodies in Sjogren's syndrome utilizing Sera from the Sjogren's international collaborative clinical alliance cohort. *Bmc Ophthalmology, 15*. doi:ARTN 38

10.1186/s12886-015-0023-1

Suzuki, Y., Ohsugi, K., & Ono, Y. (1996). Oxidative stress triggers tyrosine phosphorylation in B cells through a redox- and inflammatory cytokine-sensitive mechanism. *Immunology, 87*(3), 396-401. doi:DOI 10.1046/j.1365-2567.1996.431546.x

Takulapalli, B. R., Qiu, J., Magee, D. M., Kahn, P., Brunner, A., Barker, K., . . . Wiktor, P. (2012). High Density Diffusion-Free Nanowell Arrays. *Journal of Proteome Research, 11*(8), 4382-4391. doi:10.1021/pr300467q

Tang, X., & Bruce, J. E. (2009). Chemical cross-linking for protein-protein interaction studies. *Methods Mol Biol, 492*, 283-293. doi:10.1007/978-1-59745-493-3_17

Taplin, C. E., & Barker, J. M. (2008). Autoantibodies in type 1 diabetes. *Autoimmunity, 41*(1), 11-18. doi:10.1080/08916930701619169

ten Klooster, J. P., Jaffer, Z. M., Chernoff, J., & Hordijk, P. L. (2006). Targeting and activation of Rac1 are mediated by the exchange factor beta-Pix. *Journal of Cell Biology, 172*(5), 759-769. doi:10.1083/jcb.200509096

Teramoto, H., Salem, P., Robbins, K. C., Bustelo, X. R., & Gutkind, J. S. (1997). Tyrosine phosphorylation of the vav proto-oncogene product links Fc epsilon RI to the Rac1-JNK pathway. *Journal of Biological Chemistry, 272*(16), 10751-10755.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology, 60*(5), 523-533. doi:10.1007/s00253-002-1158-6

Testoni, M., Chung, E. Y. L., Priebe, V., & Bertoni, F. (2015). The transcription factor ETS1 in lymphomas: friend or foe? *Leukemia & Lymphoma, 56*(7), 1975-1980. doi:10.3109/10428194.2014.981670

Tiwari, A., & Uzun, L. (2017). *Advanced molecularly imprinting materials*. Hoboken, New Jersey

Beverly, Massachusetts: John Wiley & Sons

Scrivener Publishing LLC.

Tong, L., & Tergaonkar, V. (2014). Rho protein GTPases and their interactions with NF kappa B: crossroads of inflammation and matrix biology. *Bioscience Reports, 34*, 283-295. doi:ARTN e00115

10.1042/BSR20140021

Troen, G., Nygaard, V., Jenssen, T. K., Ikonomou, I. M., Tierens, A., Matutes, E., . . . Delabie, J. (2004). Constitutive expression of the AP-1 transcription factors c-jun, junD, junB, and c-fos and the marginal zone B-cell transcription factor notch2 in splenic marginal zone lymphoma. *Journal of Molecular Diagnostics, 6*(4), 297-307. doi:Doi 10.1016/S1525-1578(10)60525-9

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., . . . Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature, 403*(6770), 623-627. doi:10.1038/35001009

Ulivieri, C., & Baldari, C. T. (2005). The BCR signalosome: Where cell fate is decided. *Journal of Biological Regulators and Homeostatic Agents, 19*(1-2), 1-16.

Uversky, V. N. (2013). A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science, 22*(6), 693-+. doi:10.1002/pro.2261

Uy, G. L., Mandrekar, S., Laumann, K., Sanford, B., Marcucci, G., Zhao, W. Q., . . . Larson, R. A. (2015). Addition of Sorafenib to Chemotherapy Improves the Overall Survival of Older Adults with FLT3-ITD Mutated Acute Myeloid Leukemia (AML) (Alliance C11001). *Blood, 126*(23).

Vega, F. M., & Ridley, A. J. (2008). Rho GTPases in cancer cell biology. *Febs Letters, 582*(14), 2093-2101. doi:10.1016/j.febslet.2008.04.039

Vetter, I. R., & Wittinghofer, A. (2001). Signal transduction - The guanine nucleotide-binding switch in three dimensions. *Science, 294*(5545), 1299-1304. doi:DOI 10.1126/science.1062023

Vidarsson, G., Dekkers, G., & Rispens, T. (2014). IgG subclasses and allotypes: from structure to effector functions. *Frontiers in Immunology, 5*. doi:ARTN 520

10.3389/fimmu.2014.00520

Walhout, A. J. M., Sordella, R., Lu, X. W., Hartley, J. L., Temple, G. F., Brasch, M. A., . . . Vidal, M. (2000). Protein interaction mapping in C-elegans using proteins involved in vulval development. *Science, 287*(5450), 116-122. doi:DOI 10.1126/science.287.5450.116

Walmsley, M. J., Ooi, S. K. T., Reynolds, L. F., Smith, S. H., Ruf, S., Mathiot, A., . . . Tybulewicz, V. L. J. (2003). Critical roles for Rac1 and Rac2 GTPases in B cell development and signaling. *Science, 302*(5644), 459-462. doi:DOI 10.1126/science.1089709

Wang, J., Barker, K., Steel, J., Park, J., Saul, J., Festa, F., . . . Qiu, J. (2013). A versatile protein microarray platform enabling antibody profiling against denatured proteins. *Proteomics Clinical Applications, 7*(5-6), 378-383. doi:10.1002/prca.201200062

Wang, J., Figueroa, J. D., Wallstrom, G., Sampson, J., Garcia, E. M., Steel, J., . . . LaBaer, J. (2014). Autoantibody biomarker discovery in basal-like breast cancer using nucleic acid programmable protein array. *Cancer Research, 74*(19). doi:10.1158/1538-7445.Am2014-874

Wang, X. X., Huang, H. Q., & Young, K. H. (2015). The PTEN tumor suppressor gene and its role in lymphoma pathogenesis. *Aging-Us, 7*(12), 1032-1049. doi:DOI 10.18632/aging.100855

Wetie, A. G. N., Sokolowska, I., Woods, A. G., Roy, U., Deinhardt, K., & Darie, C. C. (2014). Protein-protein interactions: switch from classical methods to proteomics

and bioinformatics-based approaches. *Cellular and Molecular Life Sciences, 71*(2), 205-228. doi:10.1007/s00018-013-1333-1

Wienands, J., Schweikert, J., Wollscheid, B., Jumaa, H., Nielsen, P. J., & Reth, M. (1998). SLP-65: A new signaling component in B lymphocytes which requires expression of the antigen receptor for phosphorylation. *Journal of Experimental Medicine, 188*(4), 791-795. doi:DOI 10.1084/jem.188.4.791

Wong, C. L., & Olivo, M. (2014). Surface Plasmon Resonance Imaging Sensors: A Review. *Plasmonics, 9*(4), 809-824. doi:10.1007/s11468-013-9662-3

Woolery, A. R., Yu, X. B., LaBaer, J., & Orth, K. (2014). AMPylation of Rho GTPases Subverts Multiple Host Signaling Processes. *Journal of Biological Chemistry, 289*(47). doi:10.1074/jbc.M114.601310

Woyach, J. A., Johnson, A. J., & Byrd, J. C. (2012). The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood, 120*(6), 1175-1184. doi:10.1182/blood-2012-02-362624

Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology, 16*(1), 18-29. doi:10.1038/nrm3920

Wu, P., Nielsen, T. E., & Clausen, M. H. (2015). FDA-approved small-molecule kinase inhibitors. *Trends in Pharmacological Sciences, 36*(7), 422-439. doi:10.1016/j.tips.2015.04.005

Xie, Q., Soutto, M., Xu, X., Zhang, Y., & Johnson, C. H. (2011). Bioluminescence resonance energy transfer (BRET) imaging in plant seedlings and mammalian cells. *Methods Mol Biol, 680*, 3-28. doi:10.1007/978-1-60761-901-7_1

Xu, S. L., Tan, J. E. L., Wong, E. P. Y., Manickam, A., Ponniah, S., & Lam, K. P. (2000). B cell development and activation defects resulting in xid-like immunodeficiency in BLNK/SLP-65-deficient mice. *International Immunology, 12*(3), 397-404. doi:DOI 10.1093/intimm/12.3.397

Xu, Y., Piston, D. W., & Johnson, C. H. (1999). A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins. *Proceedings of the National Academy of Sciences of the United States of America, 96*(1), 151-156. doi:DOI 10.1073/pnas.96.1.151

Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., & Orth, K. (2009). AMPylation of Rho GTPases by Vibrio VopS Disrupts Effector Binding and Downstream Signaling. *Science, 323*(5911), 269-272. doi:10.1126/science.1166382

Yim, E. K., & Park, J. S. (2005). The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. *Cancer Res Treat, 37*(6), 319-324. doi:10.4143/crt.2005.37.6.319

Yu, X., Petritis, B., Duan, H., Xu, D., & LaBaer, J. (2018). Advances in cell-free protein array methods. *Expert Rev Proteomics, 15*(1), 1-11. doi:10.1080/14789450.2018.1415146

Yu, X. B., & LaBaer, J. (2015). High-throughput identification of proteins with AMPylation using self-assembled human protein (NAPPA) microarrays. *Nature Protocols, 10*(5), 756-767. doi:10.1038/nprot.2015.044

Yu, X. B., Petritis, B., & LaBaer, J. (2016). Advancing translational research with next-generation protein microarrays. *Proteomics, 16*(8), 1238-1250. doi:10.1002/pmic.201500374

Yu, X. B., Song, L. S., Petritis, B., Bian, X. F., Wang, H. Y., Viloria, J., . . . LaBaer, J. (2017). Multiplexed Nucleic Acid Programmable Protein Arrays. *Theranostics, 7*(16), 4057-4070. doi:10.7150/thno.20151

Zegers, M. M., & Friedl, P. (2014). Rho GTPases in collective cell migration. *Small GTPases, 5*, e28997. doi:10.4161/sgtp.28997

Zemella, A., Thoring, L., Hoffmeister, C., & Kubick, S. (2015). Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem, 16*(17), 2420-2431. doi:10.1002/cbic.201500340

Zhang, A. L., Yan, T., Wang, K., Huang, Z. H., & Liu, J. B. (2017). PI3Ka isoform-dependent activation of RhoA regulates Wnt5a-induced osteosarcoma cell migration. *Cancer Cell International, 17*. doi:ARTN 27

10.1186/s12935-017-0396-8

Zhang, B. L., & Zheng, Y. (1998). Negative regulation of rho family GTPases Cdc42 and Rac2 by homodimer formation. *Journal of Biological Chemistry, 273*(40), 25728-25733. doi:DOI 10.1074/jbc.273.40.25728

Zhang, S. M., Zhou, X., Lang, R. A., & Guo, F. K. (2012). RhoA of the Rho Family Small GTPases Is Essential for B Lymphocyte Development. *Plos One, 7*(3). doi:ARTN e33773

10.1371/journal.pone.0033773

Zhao, J. J., Cheng, H. L., Jia, S. D., Wang, L., Gjoerup, O. V., Mikami, A., & Roberts, T. M. (2006). The p110 alpha isoform of PI3K is essential for proper growth factor signaling and oncogenic transformation. *Proceedings of the National Academy of*

Sciences of the United States of America, 103(44), 16296-16300. doi:10.1073/pnas.0607899103

Zhao, L., Zhao, K. Q., Hurst, R., Slater, M. R., Acton, T. B., Swapna, G. V., . . . Montelione, G. T. (2010). Engineering of a wheat germ expression system to provide compatibility with a high throughput pET-based cloning platform. *J Struct Funct Genomics, 11*(3), 201-209. doi:10.1007/s10969-010-9093-8

Zhou, M., Li, Q., & Wang, R. X. (2016). Current Experimental Methods for Characterizing Protein-Protein Interactions. *Chemmedchem, 11*(8), 738-756. doi:10.1002/cmdc.201500495

Zhu, L., Wang, K., Cui, J., Liu, H., Bu, X. L., Ma, H. L., . . . Hu, Z. Y. (2014). Label-Free Quantitative Detection of Tumor-Derived Exosomes through Surface Plasmon Resonance Imaging. *Analytical Chemistry, 86*(17), 8857-8864. doi:10.1021/ac5023056

APPENDIX A

HUMAN PROTEINS IN THE BCR SIGNALING PATHWAY

Table 19. Detailed list of human proteins in the BCR signaling pathway (continued on next page)

Protein Symbol	Full Protein Name	Ref Seq ID	UniProt #
AKT1	v-akt murine thymoma viral oncogene homolog 1	BC000479.2	P31749
AKT2	v-akt murine thymoma viral oncogene homolog 2	BC120994	P31751
AKT2	v-akt murine thymoma viral oncogene homolog 2	BC063421	Q6P4H3
AKT3	v-akt murine thymoma viral oncogene homolog 3	AJ245709	Q9Y243
ARHGEF7	Rho guanine nucleotide exchange facgtor (GEF) 7	EU832554.1	Q14155
BCL10	B-cell CLL/lymphoma 10	NM_003921	O95999
BCL2	B-cell CLL/lymphoma 2	BC027258	P10415
BCL2A1	BCL2-related protein A1	U29680	Q16548
BCL2L1	BCL2-like 1	BC019307	Q07817
BLK	B lymphoid tyrosine kinase	BC007371	P51451
BLNK	B-cell linker	BC018906.2	Q8WV28
BTK	Bruton agammaglobulinemia tyrosine kinase	NM_000061	Q06187
CARD11	caspase recruitment domain family, member 11	BC111719	Q9BXL7
CD19	CD19 molecule	BC006338	P15391
CD22	CD22 molecule	BC109306	Q32M46
CD72	CD72 antigen	BC030227	P21854
CD79A	CD79a molecule, immunoglobulin-associated alpha	BC113733	P11912
CD79B	CD79b molecule, immunoglobulin-associated beta	BC030210.1	Q6PIS4
CD81	CD81 molecule	BC002978	P60033
CDC42	cell division cycle 42 (GTP binding protein, 25kDa)	NM_001791	P60953
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	U26727	Q8N726
DAPP1	dual adaptor of phosphotyrosine and 3-phosphoinositides (BAM32)	BC012924	Q9UN19
EGR1	early growth response 1	BC073983.1	P18146
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	X14798	P14921
EZR	villin 2 (ezrin) (VIL2)	BC013903	P15311
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	BC031992	P31994
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	BC004490	P01100
GRAP2	GRB2-related adaptor protein 2	BC025692	O75791
GRB2	growth factor receptor-bound protein 2	BC000631	P62993
GSK3B	glycogen synthase kinase 3 beta	BC000251	P49841
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	NM_005343	P01112
IFITM1	interferon induced transmembrane protein 1 (9-27)	BC000897	P13164
IKBKA	conserved helix-loop-helix ubiquitous kinase (CHUK)	NM_001278	O15111
IKBKB	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	BC006231	O14920
IKBKB	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	BC108694	Q32ND9
IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	BC000299	Q9Y6K9
INPP5D	inositol polyphosphate-5-phosphatase, 145kDa (SHIP1)	BC113580	Q92835
INPPL1	inositol polyphosphate phosphatase-like 1	BC140853	O15357
JUN	v-jun sarcoma virus 17 oncogene homolog (avian)	BC006175	P05412
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	BC013572	P01116
LAT2	linker for activation of T cells family, member 2	BC009204.2	Q9GZY6
LILRB3	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3 (PIRB)	BC112198	O75022
LIME1	Lck interacting transmembrane adaptor 1	BC017016	Q9H400
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	NM_002350	P07948
MALT1	mucosa associated lymphoid tissue lymphoma translocation gene 1	BC030143	Q9UDY8
MAP2K1	mitogen-activated protein kinase kinase 1	BC137459	A4QPA9
MAP2K2	mitogen-activated protein kinase kinase 2	BC000471	P36507
MAP2K3	mitogen-activated protein kinase kinase 3	NM_002756	P46734
MAP3K3	mitogen-activated protein kinase kinase kinase 3	AL834303	Q99759
MAPK1	mitogen-activated protein kinase 1	BC017832.1	P28482
MAPK12	mitogen-activated protein kinase 12	CR456515	P53778
MAPK13	mitogen-activated protein kinase 13	BC085196.1	O15264
MAPK14	mitogen-activated protein kinase 14	BT006933	Q16539

Protein Symbol	Full Protein Name	Ref Seq ID	UniProt #
MAPK3	mitogen-activated protein kinase 3	BC013992	P27361
MAPK8	mitogen-activated protein kinase 8	NM_002750	P45983
MAPK9	mitogen-activated protein kinase 9	NM_139069	P45984
MDM2	Mdm2, p53 E3 ubiquitin protein ligase homolog (mouse)	NM_002392.3	Q00987
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	BC000141	P01106
NCK1	NCK adaptor protein 1	BC006403.2	P16333
NCKAP1L	NCK-associated protein 1-like	BC093769	P55160
NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive	BC131509	O94916
NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	BC112243	Q2M1S3
NFATC3	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3	BC001050	Q12968
NFATC4	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4	BC053855	Q14934
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (p105)	BC051765	P19838
NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	NM_020529	P25963
NFKBIB	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta	BC015528	Q15653
NFKBIE	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon	NM_004556.2	O00221
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog	BC005219	P01111
PIK3AP1	phosphoinositide-3-kinase adaptor protein 1	NM_152309	Q6ZUJ8
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	BC113603	P42336
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta	BC114432	P42338
PIK3CD	phosphoinositide-3-kinase, catalytic, delta polypeptide	BC132919.1	O00329
PIK3CG	phosphoinositide-3-kinase, catalytic, gamma polypeptide	BC035683	P48736
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	BC030815	P27986
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	BC094795	P27986
PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta)	BC032647	Q05BV6
PIK3R3	phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma)	BC021622	Q8N381
PIK3R5	phosphoinositide-3-kinase, regulatory subunit 5	BC028212	Q8WYR1
PLCG2	phospholipase C, gamma 2 (phosphatidylinositol-specific)	BC007565	P16885
PPP3CA	protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform (calcineurin A alpha)	BC025714	Q08209
PPP3CB	protein phosphatase 3, catalytic subunit, beta isozyme	BC028049.1	P16298
PPP3CC	protein phosphatase 3, catalytic subunit, gamma isozyme	HQ448368	P48454
PPP3R1	protein phosphatase 3, regulatory subunit B, alpha	BC027913	P63098
PPP3R2	protein phosphatase 3, regulatory subunit B, beta	JF432717	Q96LZ3
PRKCA	protein kinase C, alpha	NM_002737	P17252
PRKCB	protein kinase C, beta	BC036472	P05771
PTEN	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	BC005821	P60484
PTPN6	protein tyrosine phosphatase, non-receptor type 6	BC002523	P29350
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	BC107748.1	P63000
RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	BC001485	P15153
RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	BC009605	P60763
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	BC018119	P04049
RAP1A	RAP1A, member of RAS oncogene family	BC014086.2	P62834
RAP1B	RAP1B, member of RAS oncogene family	AL080212	P61224
RAP2A	RAP2A, member of RAS oncogene family	BC070031.1	P10114
RAP2C	RAP2C, member of RAS oncogene family	BC003403.1	Q9Y3L5
RasGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	NM_170672	Q8IV61
RASSF5	Ras association (RalGDS/AF-6) domain family member 5	AL832784.1	Q8WWW0
RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	BC110830	Q2TAM5
RHOA	ras homolog family member A (RHOA)	NM_001664	P61586
SOS1	son of sevenless homolog 1 (Drosophila)	NM_005633	Q07889
SOS2	son of sevenless homolog 2 (Drosophila)	HQ258542	Q07890
SYK	spleen tyrosine kinase	BC011399	P43405
TEC	tec protein tyrosine kinase	BC143487	P42680
TP53	tumor protein p53 (TP53)	BC003596	P04637
VAV1	vav 1 oncogene	BC013361	Q96D37
VAV2	vav 2 oncogene	BC132967	P52735
VAV3	vav 3 oncogene	NM_006113	Q9UKW4

APPENDIX B

PANTHER AND HGNC ANNOTATIONS

Table 20. Unique PANTHER biological processes, part 1 (to be cross-referenced to Table 22)

Unique biological processes of target proteins*	Corresponding #
apoptotic process(GO:0006915)	1
B cell mediated immunity(GO:0019724)	2
biosynthetic process(GO:0009058)	3
blood coagulation(GO:0007596)	4
calcium-mediated signaling(GO:0019722)	5
cell adhesion(GO:0007155)	6
cell cycle(GO:0007049)	7
cell death(GO:0008219)	8
cell differentiation(GO:0030154)	9
cell proliferation(GO:0008283)	10
cell surface receptor signaling pathway(GO:0007166)	11
cell-cell adhesion(GO:0016337)	12
cellular component morphogenesis(GO:0032989)	13
cellular component movement(GO:0006928)	14
cellular component organization(GO:0016043)	15
cellular defense response(GO:0006968)	16
cellular process(GO:0009987)	17
cellular protein modification process(GO:0006464)	18
chromatin organization(GO:0006325)	19
cytoskeleton organization(GO:0007010)	20
ectoderm development(GO:0007398)	21
embryo development(GO:0009790)	22
exocytosis(GO:0006887)	23
female gamete generation(GO:0007292)	24
glycogen metabolic process(GO:0005977)	25
G-protein coupled receptor signaling pathway(GO:0007186)	26
hemopoiesis(GO:0030097)	27
I-kappaB kinase/NF-kappaB cascade(GO:0007249)	28
immune response(GO:0006955)	29
immune system process(GO:0002376)	30
induction of apoptosis(GO:0006917)	31
intracellular protein transport(GO:0006886)	32
intracellular signal transduction(GO:0035556)	33
JNK cascade(GO:0007254)	34
localization(GO:0051179)	35
locomotion(GO:0040011)	36
macrophage activation(GO:0042116)	37
MAPK cascade(GO:0000165)	38
meiosis(GO:0007126)	39
mesoderm development(GO:0007498)	40

* Obtained from PANTHER (Protein Ananalysis Through Evolutionary Relationships)

Table 21. Unique PANTHER biological processes, part 2 (to be cross-referenced to Table 23)

Unique biological processes of target proteins*	Corresponding #
mitosis(GO:0007067)	41
monosaccharide metabolic process(GO:0005996)	42
mRNA processing(GO:0006397)	43
mRNA splicing, via spliceosome(GO:0000398)	44
natural killer cell activation(GO:0030101)	45
negative regulation of apoptotic process(GO:0043066)	46
nervous system development(GO:0007399)	47
neurological system process(GO:0050877)	48
nitrogen compound metabolic process(GO:0006807)	49
nucleobase-containing compound metabolic process(GO:0006139)	50
phagocytosis(GO:0006909)	51
phosphate-containing compound metabolic process(GO:0006796)	52
phospholipid metabolic process(GO:0006644)	53
protein complex assembly(GO:0006461)	54
protein phosphorylation(GO:0006468)	55
receptor-mediated endocytosis(GO:0006898)	56
regulation of biological process(GO:0050789)	57
regulation of carbohydrate metabolic process(GO:0006109)	58
regulation of catalytic activity(GO:0050790)	59
regulation of cell cycle(GO:0051726)	60
regulation of nucleobase-containing compound metabolic process(GO:0019219)	61
regulation of phosphate metabolic process(GO:0019220)	62
regulation of sequence-specific DNA binding transcription factor activity(GO:0051090)	63
regulation of transcription from RNA polymerase II promoter(GO:0006357)	64
response to abiotic stimulus(GO:0009628)	65
response to biotic stimulus(GO:0009607)	66
response to endogenous stimulus(GO:0009719)	67
response to external stimulus(GO:0009605)	68
response to stimulus(GO:0050896)	69
response to stress(GO:0006950)	70
RNA splicing, via transesterification reactions(GO:0000375)	71
segment specification(GO:0007379)	72
signal transduction(GO:0007165)	73
single-multicellular organism process(GO:0044707)	74
synaptic transmission(GO:0007268)	75
transcription from RNA polymerase II promoter(GO:0006366)	76
transcription, DNA-dependent(GO:0006351)	77
transmembrane receptor protein serine/threonine kinase signaling pathway(GO:0007178)	78
transmembrane receptor protein tyrosine kinase signaling pathway(GO:0007169)	79

* Obtained from PANTHER (Protein Ananalysis Through Evolutionary Relationships)

Table 22. Unique PANTHER biological processes, part 1 (to be cross-referenced to Tables 20 - 21)

Gene Symbol	Biological Process(es)*																
AKT1	33	52	57	69													
AKT2	33	52	57	69													
AKT3	33	52	57	69													
ARHGEF7	2	26	34	5	16	48											
BCL2	11	33	46	70													
BCL2A1	11	33	46	70													
BCL2L1	11	33	46	70													
BLK	9	10	20	29	46	51	52	54	70	79							
BLNK	29	33	57	79													
BTK	9	10	29	52	57	70	79										
CARD11	17																
CD19	16	29	57	67	70	79											
CD22	17	69															
CD79A	9	11	27	29	57	74											
CD79B	9	11	27	29	57	74											
CD81	11	57	69														
CDC42	26	32	33	56													
CDKN2A	7																
ETS1	3	9	17	49													
EZR	13	17															
FCGR2B	2	37	45														
GRAP2	9	10	14	29	35	36	52	57	70	79							
GRB2	9	10	14	29	35	36	52	57	70	79							
GSK3B	11	21	22	25	33	40	41	47	72								
HRAS	17	32	56														
IKBKA	17	29															
IKBKB	17	29															
IKBKG	28	3	49	61	63	69	77										
INPP5D	17	32	42	53	56												
INPPL1	17	32	42	53	56												
JUN	3	8	9	10	49	60	65	66	67	68	70						
KRAS	26	28	38	6	32	48	56	75									
LILRB3	17	69															
LIME1	71	15	17	44	49												
LYN	1	4	9	10	12	14	15	23	27	29	36	47	52	57	70	74	79
MAP2K1	1	33	59	60	62	70											
MAP2K2	1	33	59	60	62	70											
MAP2K3	1	33	59	60	62	70											
MAP3K3	1	33	59	60	62	70											
MAPK1	38	7	30	70													
MAPK12	38	7	30	70													
MAPK13	38	7	30	70													
MAPK14	38	7	30	70													
MAPK3	38	7	30	70													
MAPK8	38	7	30	70													
MAPK9	38	7	30	70													

* Obtained from PANTHER (Protein Ananalysis Through Evolutionary Relationships)

Table 23. PANTHER biological processes of target proteins, part 2 (to be cross-referenced to Tables 20 - 21)

Gene Symbol	Biological Process(es)*												
MDM2	17	50											
MYC	7	64											
NFAT5	16	17	29	64	70								
NFATC1	16	64											
NFATC3	16	64											
NFATC4	16	64											
NFKBIA	28	1	30	32	53	64	70						
NFKBIB	28	1	30	32	53	64	70						
NFKBIE	64												
NRAS	26	28	38	6	32	48	56	75					
PIK3CA	3	4	14	29	33	35	36	53	55	57	70	74	
PIK3CB	3	4	14	29	33	35	36	53	55	57	70	74	
PIK3CD	3	4	14	29	33	35	36	53	55	57	70	74	
PIK3CG	3	4	14	29	33	35	36	53	55	57	70	74	
PIK3R1	3	53	62	67	79								
PIK3R2	3	53	62	67	79								
PIK3R3	3	53	62	67	79								
PPP3CA	1	25	30	33	43	39	41	55	58	61	70	76	
PPP3CB	1	25	30	33	43	39	41	55	58	61	70	76	
PPP3CC	1	25	30	33	43	39	41	55	58	61	70	76	
PPP3R1	5	18											
PPP3R2	5	18											
PRKCA	33	52	57	69									
PRKCB	33	52	57	69									
PTEN	7	53	55										
PTPN6	18												
RAC1	26	32	33	56									
RAC2	26	32	33	56									
RAC3	26	32	33	56									
RAF1	28	34	13	22	24	30	31	78					
RAP1A	26	28	38	6	32	48	56	75					
RAP1B	26	28	38	6	32	48	56	75					
RAP2A	26	28	38	6	32	48	56	75					
RAP2C	26	28	38	6	32	48	56	75					
RasGRP3	26	38	41	79									
RASSF5	1	73											
RHOA	26	32	33	56									
SOS1	26	38	41	79									
SOS2	26	38	41	79									
SYK	4	9	10	12	23	27	29	37	52	57	70	74	79
TEC	9	10	29	52	57	70	79						
TP53	1	3	18	19	33	49	60	64	65	70			
VAV1	2	26	34	5	16	48							
VAV2	2	26	34	5	16	48							
VAV3	2	26	34	5	16	48							

* Obtained from PANTHER (Protein Ananalysis Through Evolutionary Relationships)

Table 24. Associated HGNC protein domains for each protein target, part 1

Protein Symbol	Protein domain 1	Protein domain 2	Protein domain 3
AKT1	Pleckstrin homology domain containing		
AKT2	Pleckstrin homology domain containing		
AKT3	Pleckstrin homology domain containing		
ARHGEF7	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	
BCL10	Caspase recruitment domain containing	CBM complex	
BCL2	BCL2 family	Protein phosphatase 1 regulatory subunits	
BCL2A1	BCL2 family	Minor histocompatibility antigens	
BCL2L1	BCL2 family	Protein phosphatase 1 regulatory subunits	
BLK	SH2 domain containing	Src family tyrosine kinases	
BLNK	SH2 domain containing		
BTK	Pleckstrin homology domain containing	SH2 domain containing	Tec family tyrosine kinases
CARD11	Caspase recruitment domain containing	CBM complex	
CD19	CD molecules	Immunoglobulin like domain containing	Minor histocompatibility antigens
CD22	CD molecules		
CD72	CD molecules		
CD79A	CD molecules	V-set domain containing	
CD79B	CD molecules	V-set domain containing	
CD81	CD molecules		
CDC42	Rho family GTPases		
DAPP1	Pleckstrin homology domain containing	SH2 domain containing	
FCGR2B	CD molecules	Immunoglobulin like domain containing	
FOS	Basic leucine zipper proteins		
GRAP2	SH2 domain containing		
GRB2	SH2 domain containing		
HRAS	RAS type GTPases		
IFITM1	CD molecules		
INPP5D	Phosphoinositide phosphatases	SH2 domain containing	
INPPL1	Phosphoinositide phosphatases	SH2 domain containing	
JUN	Basic leucine zipper proteins		
KRAS	RAS type GTPases		
LILRB3	CD molecules		
LYN	SH2 domain containing	Src family tyrosine kinases	
MALT1	CBM complex	Immunoglobulin like domain containing	
MAP2K1	Mitogen-activated protein kinase kinases		
MAP2K2	Mitogen-activated protein kinase kinases		
MAP2K3	Mitogen-activated protein kinase kinases		
MAP3K3	Mitogen-activated protein kinase kinase kinases		
MAPK1	Mitogen-activated protein kinases		
MAPK12	Mitogen-activated protein kinases		
MAPK13	Mitogen-activated protein kinases		
MAPK14	Mitogen-activated protein kinases		
MAPK3	Mitogen-activated protein kinases		
MAPK8	Mitogen-activated protein kinases		

Note: Only the proteins with an HGNC protein domain are listed

Table 25. Associated HGNC protein domains for each protein target, part 2

Protein Symbol	Protein domain 1	Protein domain 2	Protein domain 3
MAPK9	Mitogen-activated protein kinases		
NCK1	SH2 domain containing		
NFAT5	nuclear factors of activated T-cells		
NFATC1	nuclear factors of activated T-cells		
NFATC3	nuclear factors of activated T-cells		
NFATC4	nuclear factors of activated T-cells		
NFKB1	Ankyrin repeat domain containing	NF-kappa B complex subunits	
NFKBIA	Ankyrin repeat domain containing		
NFKBIB	Ankyrin repeat domain containing		
NFKBIE	Ankyrin repeat domain containing		
NRAS	RAS type GTPases		
PIK3CA	Phosphatidylinositol 3-kinase subunits		
PIK3CB	Phosphatidylinositol 3-kinase subunits		
PIK3CD	Phosphatidylinositol 3-kinase subunits		
PIK3CG	Phosphatidylinositol 3-kinase subunits		
PIK3R1	SH2 domain containing		
PIK3R2	SH2 domain containing		
PIK3R3	SH2 domain containing		
PLCG2	SH2 domain containing		
PPP3CA	Protein phosphatase catalytic subunits		
PPP3CB	Protein phosphatase catalytic subunits		
PPP3CC	Protein phosphatase catalytic subunits		
PPP3R1	EF-hand domain containing	Protein phosphatase 3 regulatory subunits	
PPP3R2	EF-hand domain containing	Protein phosphatase 3 regulatory subunits	
PRKCA	C2 domain containing protein kinases		
PRKCB	C2 domain containing protein kinases		
PTEN	Phosphoinositide phosphatases		
PTPN6	SH2 domain containing		
RAC1	Endogenous ligands	Rho family GTPases	
RAC2	Endogenous ligands	Rho family GTPases	
RAC3	Endogenous ligands	Rho family GTPases	
RAF1	Mitogen-activated protein kinase kinase kinases		
RAP1A	RAS type GTPases		
RAP1B	RAS type GTPases		
RAP2A	RAS type GTPases		
RAP2C	RAS type GTPases		
RasGRP3	EF-hand domain containing		
RELA	NF-kappa B complex subunits		
RHOA	Rho family GTPases		
SOS1	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	
SOS2	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	
SYK	SH2 domain containing		
TEC	Pleckstrin homology domain containing	SH2 domain containing	Tec family tyrosine kinases
VAV1	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	SH2 domain containing
VAV2	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	SH2 domain containing
VAV3	Pleckstrin homology domain containing	Rho guanine nucleotide exchange factors	SH2 domain containing

Note: Only the proteins with an HGNC protein domain are listed

APPENDIX C

KNOWN PROTEIN INTERACTIONS IN BIOGRID AND HPRD

Table 26. Known protein interactions with BLNK

Protein	Experiment type*	Author	Year published	Curated by
BTK	Affinity Capture-MS	Hashimoto S	1999	BioGRID
		Oellerich T	2011	BioGRID
	Affinity Capture-Western	Yasuda T	2002	BioGRID
		Hashimoto S	1999	BioGRID
		Janda E	2011	BioGRID
		Su	1999	HPRD
		Imamura	2004	HPRD
	Co-crystal Structure	Huang	2006	HPRD
CD72	Affinity Capture-Western	Fusaki N	2000	BioGRID
CD79A	Affinity Capture-MS	Oellerich T	2011	BioGRID
	Affinity Capture-Western	Kabak S	2002	BioGRID
	Reconstituted Complex	Engels N	2001	BioGRID
		Kabak S	2002	BioGRID
GRAP2	Affinity Capture-Western	Yankee TM	2003	BioGRID
	Protein-peptide	Berry DM	2002	BioGRID
	Reconstituted Complex	Berry DM	2002	BioGRID
GRB2	Affinity Capture-MS	Oellerich T	2011	BioGRID
	Affinity Capture-Western	Fu C	1998	BioGRID
		Fusaki N	2000	BioGRID
		Oellerich T	2011	BioGRID
	Reconstituted Complex	Engels N	2001	BioGRID
		Wienands J	1998	BioGRID
LYN	Affinity Capture-MS	Oellerich T	2011	BioGRID
NCK1	Affinity Capture-Western	Fu C	1998	BioGRID
		Sauer	2001	HPRD
	Co-localization	Chen TC	2014	BioGRID
PIK3R1	Reconstituted Complex	Watanabe	2000	HPRD
PLCG2	Affinity Capture-MS	Oellerich T	2011	BioGRID
	Affinity Capture-Western	Oellerich T	2011	BioGRID
	Reconstituted Complex	Janssen	2003	HPRD
PTPN6	Affinity Capture-Western	Adachi T	2001	BioGRID
		Mizuno	2000	HPRD
SOS1	Reconstituted Complex	Watanabe	2000	HPRD
SYK	Affinity Capture-MS	Oellerich T	2011	BioGRID
	Affinity Capture-Western	Janda E	2011	BioGRID
VAV1	Affinity Capture-Western	Fu C	1998	BioGRID
VAV3	Affinity Capture-MS	Oellerich T	2011	BioGRID

** BioGRID Experimental Evidence Code*

Note: PPIs not sufficiently supported by experimental evidence were not included

Note: List contains human, mouse, or rat PPIs

Table 27. Known protein interactions with BTK

Protein	Experiment type*	Author	Year published	Curated by
BLNK	Affinity Capture-MS	Hashimoto S	1999	BioGRID
		Oellerich T	2011	BioGRID
	Affinity Capture-Western	Yasuda T	2002	BioGRID
		Hashimoto S	1999	BioGRID
		Janda E	2011	BioGRID
		Su	1999	HPRD
	Co-crystal Structure	Huang	2006	HPRD
BTK	Affinity Capture-Western	Morrogh LM	1999	BioGRID
	Biochemical Activity	Egloff AM	2001	BioGRID
		Morrogh LM	1999	BioGRID
		Park H	1996	BioGRID
	Co-crystal Structure	Mao C	2001	BioGRID
DAPP1	Biochemical Activity	Stephens	2001	HPRD
GRB2	Affinity Capture-MS	Brehme M	2009	BioGRID
	Two-hybrid	Bandyopadhyay S	2010	BioGRID
LYN	Reconstituted Complex	Cheng G	1994	BioGRID
	Affinity Capture-Western	Rawlings	1996	HPRD
MAPK1	Reconstituted Complex	Imamura Y	2004	HPRD
MYC	Dosage Lethality	Toyoshima M	2012	BioGRID
PIK3AP1	Biochemical Activity	Okada	2000	HPRD
PIK3R3	Two-hybrid	Grossmann A	2015	BioGRID
PLCG2	Affinity Capture-Western	Yasuda T	2002	BioGRID
		Guo B	2000	BioGRID
	Biochemical Activity	Watanabe	2001	HPRD
PRKCA	Biochemical Activity	Yao	1994	HPRD
PRKCB	Reconstituted Complex	Yang XL	2003	BioGRID
SYK	Affinity Capture-Western	Morrogh LM	1999	BioGRID
		Baba	2001	HPRD
TP53	Affinity Capture-Western	Jiang	2004	HPRD
VAV1	Affinity Capture-Western	Guinamard R	1997	BioGRID

** BioGRID Experimental Evidence Code*

Note: PPIs not sufficiently supported by experimental evidence were not included

Note: List contains human, mouse, or rat PPIs

Table 28. Known protein interactions with PI3K*

Protein	Experiment type*	Author	Year published	Curated by	Monomer/Heterodimer**
AKT1	Affinity Capture-Western	Mitra S	2011	BioGRID	Unclear; simply states "PI3K"
		Xiong Y	2009	BioGRID	PIK3R1 and PIK3CA antibodies
CD19	Affinity Capture-Western	Okada T	2000	BioGRID	PIK3R1 antibody
		Weng	1994	HPRD	PIK3R1
CDC42	Biochemical Activity	Tolias	1995	HPRD	PIK3CA activity
	Reconstituted Complex	Zheng	1994	HPRD	PIK3R1 antibody
EZR	Affinity Capture-Western	Gautreau A	1999	BioGRID	PIK3R1 antibody
	Reconstituted Complex	Gautreau A	1999	BioGRID	PIK3R1 domains
FOS	Affinity Capture-Western	Luo X	1997	BioGRID	PIK3R1 antibody
GRB2	Affinity Capture-MS	Bisson N	2011	BioGRID	PIK3R1 and PIK3CA probed separately
		Huttlin EL	2017	BioGRID	PIK3CA
	Affinity Capture-Western	Saleem A	1995	BioGRID	PIK3R1 antibody
		Wang J	1995	BioGRID	PIK3R1 antibody
		Gelkop S	2001	BioGRID	PIK3R1 antibody
		Jain SK	1997	BioGRID	PIK3R1 antibody
		Ren SY	2005	BioGRID	PIK3R1
		Cerboni C	1998	BioGRID	PIK3R1 antibody
		Bisson N	2011	BioGRID	PIK3R1
		Weinger JG	2008	BioGRID	PIK3R1 antibody
	Reconstituted Complex	Saleem A	1995	BioGRID	PIK3R1 antibody
		Wang J	1995	BioGRID	PIK3R1 antibody
		Weinger JG	2008	BioGRID	PIK3R1 antibody
	Two-hybrid	Wang J	1995	BioGRID	PIK3R1 antibody
		Bandyopadhyay S	2010	BioGRID	PIK3R1
HRAS	Co-localization	Chen TC	2014	BioGRID	PIK3R1 antibody
	Reconstituted Complex	Vargiu P	2004	BioGRID	PIK3CA
		Rodriguez-Viciana P	1996	BioGRID	Heterodimer
		Rodriguez-Viciana P	1997	BioGRID	Heterodimer
	Two-hybrid	Li W	2000	BioGRID	Unclear; simply states "PI3K"
INPP5D	Affinity Capture-Western	Zhang S	1999	BioGRID	PIK3R1 antibody
INPPL1	Affinity Capture-MS	Brehme M	2009	BioGRID	PIK3R1
KRAS	Two-hybrid	Li W	2000	BioGRID	Unclear; simply states "PI3K"
NCK1	Affinity Capture-Western	Gelkop S	2001	BioGRID	PIK3R1 antibody
NFKBIA	Reconstituted Complex	Beraud C	1999	BioGRID	PIK3R1 antibody
PIK3AP1	Affinity Capture-Western	Okada T	2000	BioGRID	PIK3R1 antibody
PIK3CA	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3R1 and PIK3CA probed separately
	Affinity Capture-Western	Sun M	2010	BioGRID	PIK3R1
		Fang D	2001	BioGRID	PIK3R1
		Xiong Y	2009	BioGRID	Heterodimer
		Luo	2005	HPRD	PIK3R1
	Reconstituted Complex	Woscholski R	1994	BioGRID	PIK3R1
PIK3CB	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3R1
	Affinity Capture-Western	Hu P	1994	BioGRID	PIK3R1
		Sun M	2010	BioGRID	PIK3R1
	Co-fractionation	Havugimana PC	2012	BioGRID	PIK3R1
		Wan C	2015	BioGRID	PIK3R1
	Co-localization	Chen TC	2014	BioGRID	PIK3R1 antibody
PIK3CD	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3R1
	Affinity Capture-Western	Vanhaesebroeck B	1997	BioGRID	PIK3R1
	Reconstituted Complex	Vanhaesebroeck B	1997	BioGRID	PIK3R1
PIK3R1	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3R1
		Sun M	2010	BioGRID	PIK3CA
		Fang D	2001	BioGRID	PIK3CA
		Xiong Y	2009	BioGRID	Heterodimer
	Reconstituted Complex	Woscholski R	1994	BioGRID	PIK3CA
PIK3R2	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3CA
		Huttlin EL	2015	BioGRID	PIK3R1
		Huttlin EL	2017	BioGRID	PIK3R1
	Affinity Capture-Western	Kuchay S	2013	BioGRID	PIK3CA
PIK3R3	Affinity Capture-MS	Hein MY	2015	BioGRID	PIK3CA
		Huttlin EL	2017	BioGRID	PIK3CA
		Huttlin EL	2017	BioGRID	PIK3R1
	Two-hybrid	Mothe I	1997	BioGRID	PIK3CA
		Grossmann A	2015	BioGRID	PIK3CA
PLCG2	Co-localization	Chen TC	2014	BioGRID	PIK3R1 antibody
PTPN6	Affinity Capture-Western	Cuervas	1999	HPRD	PIK3R1
RAC1	Reconstituted Complex	Zheng	1994	HPRD	PIK3R1 antibody
	Biochemistry Activity	Tolias	1995	HPRD	PIK3CA
SOS1	Co-localization	Chen TC	2014	BioGRID	PIK3R1 antibody
SYK	Affinity Capture-Western	Moon KD	2005	BioGRID	PIK3R1 antibody
		Huang ZY	2006	BioGRID	PIK3R1 antibody
	Reconstituted Complex	Moon KD	2005	BioGRID	PIK3R1 antibody
	Two-hybrid	Moon KD	2005	BioGRID	PIK3R1 antibody
VAV1	Affinity Capture-Western	Bertagnolo V	1998	BioGRID	PIK3R1 antibody
		Shigematsu H	1997	BioGRID	PIK3R1 antibody
		Weng	1994	HPRD	PIK3R1 antibody
	Reconstituted Complex	Bertagnolo V	1998	BioGRID	PIK3R1 antibody; catalytic activity suggests heterodimer
		Shigematsu H	1997	BioGRID	PIK3R1 antibody; catalytic activity suggests heterodimer
VAV3	Affinity Capture-Western	Zeng	2000	HPRD	PIK3R1

** BioGRID Experimental Evidence Code*

***Since PI3K is often found as a complex, the presence of a heterodimer cannot be discounted even if only a PI3K monomer was used to probe these PPIs.*

****Our query, PI3K, was a heterodimer containing PIK3R1 and PIK3CA.*

Note: PPIs not sufficiently supported by experimental evidence were not included

Note: List contains human, mouse, or rat PPIs

Table 29. Known protein interactions with RAC1

Protein	Experiment type*	Author	Year published	Curated by
AKT1	Biochemical Activity	Kwon	2000	HPRD
ARHGEF7	Affinity Capture-MS	Sandrock K	2010	BioGRID
	Affinity Capture-Western	Feng Q	2004	BioGRID
		Shin	2004	HPRD
GRB2	Affinity Capture-MS	Huttlin EL	2017	BioGRID
MAPK1	Affinity Capture-Western	Kim Y	2009	BioGRID
NRAS	Synthetic Lethality	Wang T	2017	BioGRID
PIK3R1	Affinity Capture-Western	Chan	2002	HPRD
	Far Western	Zheng	1994	HPRD
	Biochemical Activity	Tolias	1995	HPRD
PRKCA	Biochemical Activity	Slater	2001	HPRD
RAC1	Affinity Capture-MS	Sandrock K	2010	BioGRID
	Reconstituted Complex	Zhang B	2001	BioGRID
RAC3	Affinity Capture-MS	Huttlin EL	2017	BioGRID
SOS1	Affinity Capture-Western	Jeganathan N	2016	BioGRID
TEC	Phenotypic Enhancement	Kline JB	2001	BioGRID
VAV1	Affinity Capture-Western	Kaminuma O	2001	BioGRID
VAV2	Affinity Capture-Western	Bartolome	2006	HPRD
VAV3	Reconstituted Complex	Movilla N	1999	BioGRID

** BioGRID Experimental Evidence Code*

Note: PPIs not sufficiently supported by experimental evidence were not included

Note: List contains human, mouse, or rat PPIs

Table 30. Known protein interactions with RHOA

Protein	Experiment type*	Author	Year published	Curated by
EZR	Co-purification	Gajate C	2005	BioGRID
IKBKB	Affinity Capture-Western	Kim HJ	2014	BioGRID
IKBKG	Reconstituted Complex	Kim HJ	2014	BioGRID
MAPK8	Co-localization	Chen TC	2014	BioGRID
	Co-purification	Gajate C	2005	BioGRID
MDM2	Affinity Capture-Western	Ma J	2012	BioGRID
NFKBIA	Affinity Capture-MS	Huttlin EL	2015	BioGRID
		Huttlin EL	2017	BioGRID
PRKCA	Biochemical Activity	Slater	2001	HPRD
PTEN	Affinity Capture-Western	Chang CC	2012	BioGRID
RAP2C	Affinity Capture-MS	Huttlin EL	2017	BioGRID
RHOA	Affinity Capture-MS	Hutchins JR	2010	BioGRID
VAV1	Affinity Capture-Western	Bartolome	2006	HPRD
VAV2	Biochemical Activity	Booden	2002	HPRD
VAV3	Reconstituted Complex	Movilla N	1999	BioGRID

** BioGRID Experimental Evidence Code*

Note: PPIs not sufficiently supported by experimental evidence were not included

Note: List contains human, mouse, or rat PPIs

APPENDIX D

SUPPLEMENTAL NANOBRET DATA

Table 31. Table of PPIs detected by NanoBRET, part 1

Target	Ref Seq ID	UniProt #	Query											
			AKT1	BLNK	BTK	DAPP1	LYN	MAPK14	PI3K	RAC1-GDP	RAC1-GTP	RHOA-GDP	RHOA-GTP	SYK
AKT1	BC000479.2	P31749	N/C	N/C	C	C	C		N	C	N/C	C		
AKT2	BC063421	P31751	C	N						C	N/C			
AKT2	BC120994	Q6P4H3		N						C	N/C		C	N
AKT3	AJ245709	Q9Y243	C				C			C	C			N
ARHGEF7	EU832554.1	Q14155							N	C		C		
BCL10	NM_003921	O95999								C		C		
BCL2	BC027258	P10415				N		C		C		C		
BCL2A1	U29680	Q16548	N/C	N	C			N	C	C	N/C		N	
BCL2L1	BC019307	Q07817	N/C						N	C	N			N
BLK	BC007371	P51451					C		N		N			N
BLNK	BC018906.2	Q8WV28	N		N									N
BTK	NM_000061	Q06187							N		N		N	N
CARD11	BC111719	Q9BXL7				N			N		N	C		N
CD19	BC006338	P15391							N	C		C		
CD22	BC109306	Q32M46				N/C				C				
CD72	BC030227	P21854	C		C	N		C	C	C	C		C	
CD79A	BC113733	P11912	N/C	N		C	C		N/C	C	N/C	C	N/C	N
CD79B	BC030210.1	Q6PIS4	N/C	N	N	C	C	N	N	C	N			N
CD81	BC002978	P60033	N/C	N					C		N		N/C	N
CDC42	NM_001791	P60953		N		C		N	N			C	C	N
CDKN2A	U26727	Q8N726								C	N	C		
DAPP1	BC012924	Q9UN19								C		C		
EGR1	BC073983.1	P18146				N/C			N	C	N	C		
ETS1	X14798	P14921	C		C	N/C	C				C	C		
EZR	BC013903	P15311	N/C	N				N	N/C		N	C		N
FCGR2B	BC031992	P31994	N/C						N/C					
FOS	BC004490	P01100	C	N					N/C					N
GRAP2	BC025692	O75791		N					N		C			
GRB2	BC000631	P62993								C				
GSK3B	BC000251	P49841								C		C	C	N
HRAS	NM_005343	P01112				C				C		C	N	
IFITM1	BC000897	P13164	C		C				C				C	
IKBKA	NM_001278	O15111	C	C		C	C			C	C		C	
IKBKB	BC006231	O14920	N/C			C			N/C		C		N	N
IKBKB	BC108694	Q32ND9	N	N		N			N/C	C	N		N	N
IKBKG	BC000299	Q9Y6K9	N	N					N		N	C	N	
INPP5D	BC113580	Q92835		N					N		N			
INPPL1	BC140853	O15357		N										
JUN	BC006175	P05412								C				
KRAS	BC013572	P01116	C			C	C	C	C	C	C		C	
LAT2	BC009204.2	Q9GZY6	C			C	C				N		C	
LILRB3	BC112198	O75022	N/C	N		C			C		N		N/C	N
LIME1	BC017016	Q9H400	N		C		C	C			N	C	N	
LYN	NM_002350	P07948	N	C			C		N		C	C		N
MALT1	BC030143	Q9UDY8	N	C			N/C			C	C			
MAP2K1	BC137459	A4QPA9	C										C	
MAP2K2	BC000471	P36507	N/C		N	C	N		C	C	C	C	N	
MAP2K3	NM_002756	P46734	N/C	C			N	C		C	C		C	
MAP3K3	AL834303	Q99759	C	C	N			N			C	C		N
MAPK1	BC017832.1	P28482	N/C				C		N	C				N
MAPK12	CR456515	P53778		C					N/C	N				
MAPK13	BC085196.1	O15264	C	C			N/C			C	N/C	C	C	
MAPK14	BT006933	Q16539	C	N		N	N			C	C			

"N" and "C" indicate whether the HaloTag is at the N- or C-terminus, respectively

Only the target proteins that interacted with a query protein are shown.

Known PPIs in human and mouse are highlighted in blue. Obtained from the online PPI databases, HPRD and BioGRID.

Table 32. Table of PPIs detected by NanoBRET, part 2

Target	Ref Seq ID	UniProt #	Query											
			AKT1	BLNK	BTK	DAPP1	LYN	MAPK14	PI3K	RAC1-GDP	RAC1-GTP	RHOA-GDP	RHOA-GTP	SYK
MAPK3	BC013992	P27361												N
MAPK8	NM_002750	P45983		N										N
MDM2	NM_002392.3	Q00987		N						N	C	C		
MYC	BC000141	P01106		N/C					N/C	N	N		N	N
NCK1	BC006403.2	P16333	N						C			C		
NCKAP1L	BC093769	P55160					N		N/C			C		N
NFAT5	BC131509	O94916										C		
NFATC1	BC112243	Q2M1S3			C		N					C		
NFATC3	BC001050	Q12968	C	C			N				C		C	
NFATC4	BC053855	Q14934	C	N/C		C				C	C		C	
NFKB1	BC051765	P19838		N			N			C	C	C		
NFKBIA	NM_020529	P25963	N	N			N		N		N			N
NFKBIB	BC015528	Q15653		N			N				N			N
NFKBIE	NM_004556.2	O00221	N	N/C			N				N	C		N
NRAS	BC005219	P01111		N								C		
PIK3AP1	NM_152309	Q6ZUJ8			C		N			C		C		
PIK3CA	BC113603	P42336	C	C			N				C	C		
PIK3CB	BC114432	P42338	C	N			N			N	N			
PIK3CD	BC132919.1	O00329					N			N	N	C		
PIK3CG	BC035683	P48736		N/C				N			N/C			N
PIK3R1	BC030815	P27986	N	C					N/C					N
PIK3R1	BC094795	P27986	N	N					N/C			C		
PIK3R2	BC032647	Q05BV6		N										
PIK3R3	BC021622	Q8N381							N	C				N
PIK3R5	BC028212	Q8WYR1	C				N	N/C	C					N
PLCG2	BC007565	P16885	C	C		C	N	C	C	N/C	N/C			
PPP3CA	BC025714	Q08209	N			C					C		C	
PPP3CB	BC028049.1	P16298	N						N					N
PPP3CC	HQ448368	P48454	N	N	N	N	N				N			N
PPP3R1	BC027913	P63098	N	N					N	C	N	C		
PPP3R2	JF432717	Q96LZ3										C		
PRKCA	NM_002737	P17252		N					N					
PTEN	BC005821	P60484	C	C				C		C	C	C	C	
PTPN6	BC002523	P29350									C			
RAC1	BC107748.1	P63000	C	N/C	C		C			N/C	C			C
RAC2	BC001485	P15153	N/C	N	N/C				N		N		N	N
RAC3	BC009605	P60763	N/C	C	C		C							
RAF1	BC018119	P04049	C	C			C				N/C		C	
RAP1A	BC014086.2	P62834					N/C				N/C		C	
RAP1B	AL080212	P61224		N			N/C				C		C	C
RAP2A	BC070031.1	P10114	N	C	N		C				N/C		N	N
RAP2C	BC003403.1	Q9Y3L5	N		N		C							N
RasGRP3	NM_170672	Q8IV61			N/C									N/C
RASSF5	AL832784.1	Q8WWW0	C	N/C	C		N/C				C		C	N
RELA	BC110830	Q2TAM5	C	C	C									C
SOS2	HQ258542	Q07890			C		N/C							
SYK	BC011399	P43405		N			N/C				C			
TEC	BC143487	P42680	N	N	N		C				N/C		N	N
TP53	BC003596	P04637	N	N	N		N				N/C		N	
VAV1	BC013361	Q96D37	C	N/C	C		N				C		C	
VAV2	BC132967	P52735		N	N		C							
VAV3	NM_006113	Q9UKW4											C	N

"N" and "C" indicate whether the HaloTag is at the N- or C-terminus, respectively.

Only the target proteins that interacted with a query protein are shown.

Known PPIs in human and mouse are highlighted in blue. Obtained from the online PPI databases, HPRD and BioGRID.

Table 33. Number of PPIs per NanoBRET query within the same PANTHER Biological Process

Biological processes of target proteins	All queries	AKT1	BLNK	BTK	DAPP1	LYN	MAPK14	PI3K	RAC1(GDP)	RAC1(GTP)	RHOA(GDP)	RHOA(GTP)	SYK	Y kinases*	S/T kinases**
apoptotic process(GO:0006915)	12	11	8	5	3	8	2	3	2	10	3	6	7	23	16
B cell mediated immunity(GO:0019724)	5	2	2	2	0	2	0	2	1	1	1	2	1	5	4
biosynthetic process(GO:0009058)	12	7	8	2	1	5	1	4	4	7	5	2	3	16	12
blood coagulation(GO:0007596)	6	3	5	0	0	5	1	1	2	6	3	0	2	12	5
calcium-mediated signaling(GO:0019722)	6	2	3	2	0	2	0	2	2	2	3	2	1	6	4
cell adhesion(GO:0007155)	6	3	3	2	1	5	1	1	1	4	1	4	3	11	5
cell cycle(GO:0007049)	10	5	6	0	1	3	1	3	8	5	3	3	4	13	9
cell death(GO:0008219)	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
cell differentiation(GO:0030154)	11	5	6	3	3	7	1	6	4	9	3	3	6	19	12
cell proliferation(GO:0008283)	8	2	4	1	0	4	0	4	2	6	1	2	4	12	6
cell surface receptor signaling pathway(GO:0007166)	7	5	4	2	3	2	3	5	6	5	3	4	5	11	13
cell-cell adhesion(GO:0016337)	2	1	2	0	0	2	0	1	0	2	1	0	1	5	2
cellular component morphogenesis(GO:0032989)	2	2	2	0	0	1	1	1	0	2	1	1	1	4	4
cellular component movement(GO:0006928)	7	3	5	0	0	4	1	2	3	6	3	0	2	11	6
cellular component organization(GO:0016043)	2	2	1	1	0	2	1	1	0	2	2	1	1	4	4
cellular defense response(GO:0006968)	9	3	4	3	1	4	0	2	3	3	4	4	1	9	5
cellular process(GO:0009987)	14	7	7	2	8	3	2	6	5	10	7	6	5	15	15
cellular protein modification process(GO:0006464)	4	2	2	1	0	1	0	1	1	3	2	1	0	3	3
chromatin organization(GO:0006325)	1	1	1	1	0	1	0	0	0	1	0	1	0	2	1
cytoskeleton organization(GO:0007010)	1	0	0	0	0	1	0	1	0	1	0	0	1	2	1
ectoderm development(GO:0007398)	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0
embryo development(GO:0009790)	2	1	1	0	0	1	0	0	1	1	1	2	1	3	1
exocytosis(GO:0006887)	2	1	2	0	0	2	0	1	0	2	1	0	1	5	2
female gamete generation(GO:0007292)	1	1	1	0	0	1	0	0	0	1	0	1	0	2	1
glycogen metabolic process(GO:0005977)	4	3	1	1	2	1	0	1	1	2	1	2	3	5	4
G-protein coupled receptor signaling pathway(GO:0007186)	18	7	9	9	2	10	2	4	3	7	3	8	8	27	13
hemopoiesis(GO:0030097)	4	3	4	1	2	4	1	3	2	4	2	1	3	11	7
I-kappaB kinase/NF-kappaB cascade(GO:0007249)	10	6	7	2	1	8	1	3	1	8	2	6	5	20	10
immune response(GO:0006955)	19	10	11	3	5	10	2	9	8	15	6	6	10	31	21
immune system process(GO:0002376)	13	8	8	1	3	7	0	4	5	7	1	3	7	22	12
induction of apoptosis(GO:0006917)	1	1	1	0	0	1	0	0	0	1	0	1	0	2	1
intracellular protein transport(GO:0006886)	16	7	11	5	3	9	2	5	3	9	3	7	8	28	14
intracellular signal transduction(GO:0035556)	28	19	16	10	6	11	6	7	13	18	8	10	12	39	32
JNK cascade(GO:0007254)	5	2	3	2	0	3	0	1	1	2	1	3	1	7	3
localization(GO:0051179)	6	2	4	0	0	3	1	1	3	5	2	0	1	8	4
locomotion(GO:0040011)	7	3	5	0	0	4	1	2	3	6	3	0	2	11	6
macrophage activation(GO:0042116)	2	1	1	0	0	1	0	1	0	1	0	0	0	2	2
MAPK cascade(GO:0000165)	16	6	7	4	2	9	1	3	6	6	2	5	7	23	10
meiosis(GO:0007126)	3	3	1	1	2	1	0	1	0	2	0	1	2	4	4
mesoderm development(GO:0007498)	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0
mitosis(GO:0007067)	7	3	1	3	2	2	0	1	1	2	1	2	4	7	4
monosaccharide metabolic process(GO:0005996)	2	0	2	0	0	0	0	1	0	1	0	0	0	2	1
mRNA processing(GO:0006397)	3	3	1	1	2	1	0	1	0	2	0	1	2	4	4
mRNA splicing, via spliceosome(GO:0000398)	1	1	0	1	0	1	1	0	0	1	1	1	0	1	2
natural killer cell activation(GO:0030101)	1	1	0	0	0	0	0	1	0	0	0	0	0	0	2
negative regulation of apoptotic process(GO:0043066)	4	2	1	1	1	1	2	3	3	3	1	1	2	4	7
nervous system development(GO:0007399)	2	1	1	0	0	1	0	1	1	1	2	1	2	4	2
neurological system process(GO:0050877)	10	4	5	4	1	7	1	2	2	5	2	6	4	16	7
nitrogen compound metabolic process(GO:0006807)	5	4	2	3	1	3	1	1	1	4	3	3	0	5	6
nucleobase-containing compound metabolic process(GO:0006139)	1	0	1	0	0	0	0	0	1	1	1	0	0	1	0
phagocytosis(GO:0006909)	1	0	0	0	0	1	0	1	0	1	0	0	1	2	1
phosphate-containing compound metabolic process(GO:0006796)	13	5	8	2	1	6	0	6	5	10	2	3	6	20	11
phospholipid metabolic process(GO:0006644)	13	6	11	0	0	5	2	5	4	8	4	1	5	21	13
protein complex assembly(GO:0006461)	1	0	0	0	0	1	0	1	0	1	0	0	1	2	1
protein phosphorylation(GO:0006468)	8	6	5	1	2	4	2	1	3	7	3	2	3	12	9
receptor-mediated endocytosis(GO:0006898)	14	6	9	5	3	7	2	4	3	7	3	7	6	22	12
regulation of biological process(GO:0050789)	21	11	14	4	3	10	2	9	10	16	6	5	10	34	22
regulation of carbohydrate metabolic process(GO:0006109)	3	3	1	1	2	1	0	1	0	2	0	1	2	4	4
regulation of catalytic activity(GO:0050790)	4	4	2	2	1	2	2	0	2	3	2	3	1	5	6
regulation of cell cycle(GO:0051726)	6	5	3	3	1	3	2	0	3	4	2	4	1	7	7
regulation of nucleobase-containing compound metabolic process(GO:0019219)	4	4	2	1	2	1	0	2	0	3	1	2	2	5	6
regulation of phosphate metabolic process(GO:0019220)	8	6	5	2	1	2	2	3	3	3	3	3	3	10	11
regulation of sequence-specific DNA binding transcription factor activity(GO:0051090)	1	1	1	0	0	0	0	1	0	1	1	1	0	1	2
regulation of transcription from RNA polymerase II promoter(GO:0006357)	9	6	7	2	1	6	0	2	2	7	3	4	4	17	8
response to abiotic stimulus(GO:0009628)	2	1	1	1	0	1	0	0	1	1	0	1	0	2	1
response to biotic stimulus(GO:0009607)	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
response to endogenous stimulus(GO:0009719)	6	2	3	0	0	0	0	4	3	0	2	0	2	5	6
response to external stimulus(GO:0009605)	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
response to stimulus(GO:0050896)	10	6	7	1	3	2	0	5	5	7	2	4	4	13	11
response to stress(GO:0006950)	34	18	18	6	5	16	5	11	15	22	9	9	14	48	34
RNA splicing, via transesterification reactions(GO:0000375)	1	1	0	1	0	1	1	0	0	1	1	1	0	1	2
segment specification(GO:0007379)	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0
signal transduction(GO:0007165)	1	1	1	1	0	1	0	0	0	1	0	1	1	3	1
single-multicellular organism process(GO:0044707)	8	5	7	1	2	7	2	3	4	8	4	1	4	18	10
synaptic transmission(GO:0007268)	6	3	3	2	1	5	1	1	1	4	1	4	3	11	5
transcription from RNA polymerase II promoter(GO:0006366)	3	3	1	1	2	1	0	1	0	2	0	1	2	4	4
transcription, DNA-dependent(GO:0006351)	1	1	1	0	0	0	0	1	0	1	1	1	0	1	2
transmembrane receptor protein serine/threonine kinase signaling pathway(GO:0007178)	1	1	1	0	0	1	0	0	0	1	0	1	0	2	1
transmembrane receptor protein tyrosine kinase signaling pathway(GO:0007169)	16	5	7	4	0	5	0	8	3	6	3	2	8	20	13

** BTK, LYN, SYK*

*** AKT1, MAPK14, PI3K*

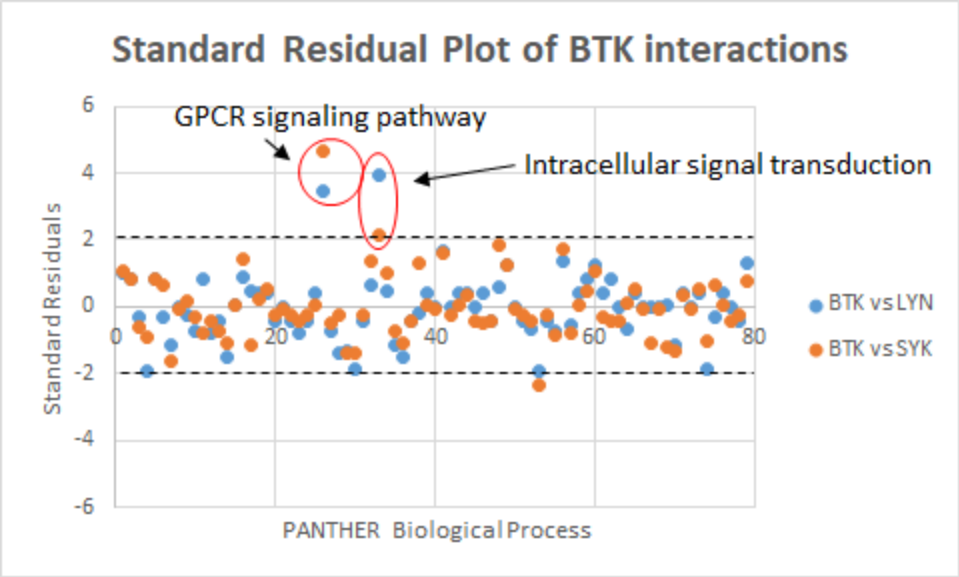


Figure 92. Enriched and under-represented PANTHER biological processes in BTK interactions compared to the other Y kinase queries

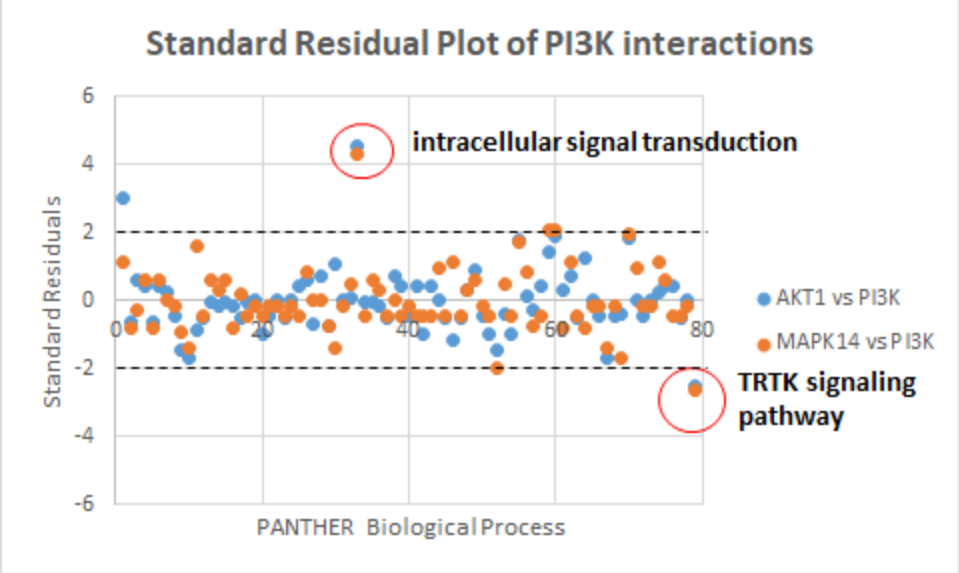


Figure 93. Enriched and under-represented PANTHER biological processes in PI3K interactions compared to the other S/T kinase queries

Table 34. Number of PPIs per NanoBRET query within the same HGNC protein domain

Biological processes of target proteins	All queries	AKT1	BLNK	BTK	DAPP1	LYN	MAPK14	PI3K	RAC1(GDP)	RAC1(GTP)	RHOA(GDP)	RHOA(GTP)	SYK	Y kinases*	S/T kinases**
Ankyrin repeat domain containing	4	2	4	0	0	4	0	1	1	4	2	0	3	11	3
Basic leucine zipper proteins	2	1	1	0	0	0	0	1	1	0	0	0	1	2	2
BCL2 family	3	2	1	1	1	0	2	2	3	2	1	1	1	2	6
C2 domain containing protein kinases	2	0	1	0	0	0	0	1	0	0	0	0	0	1	1
Caspase recruitment domain containing	2	0	0	0	1	0	0	1	1	1	2	0	1	1	1
CBM complex	3	1	1	0	1	1	0	1	2	2	2	0	1	3	2
CD molecules	9	7	4	3	5	2	2	8	5	5	2	5	4	10	17
EF-hand domain containing	3	1	1	1	0	0	0	1	1	1	2	0	1	2	2
Endogenous ligands	3	3	3	3	0	2	0	1	1	2	0	1	2	7	4
Immunoglobulin like domain containing	3	2	1	0	0	1	0	2	2	1	1	0	0	2	4
Minor histocompatibility antigens	2	1	1	1	0	0	1	2	2	1	1	1	0	1	4
Mitogen-activated protein kinase kinase kinases	2	2	2	1	0	1	1	0	0	2	1	1	1	4	3
Mitogen-activated protein kinase kinases	3	3	1	1	1	2	1	0	2	2	1	3	0	3	4
Mitogen-activated protein kinases	7	3	4	0	1	3	0	2	5	2	1	1	3	10	5
NF-kappa B complex subunits	3	1	2	1	0	1	0	0	1	1	1	0	1	4	1
nuclear factors of activated T-cells	4	2	2	1	1	2	0	0	1	2	2	2	0	4	2
Phosphatidylinositol 3-kinase subunits	4	2	3	0	0	3	1	0	2	4	2	0	1	7	3
Phosphoinositide phosphatases	3	1	3	0	0	0	1	1	1	2	1	1	0	3	3
Pleckstrin homology domain containing	13	5	6	5	1	6	0	3	6	7	3	5	5	17	8
Protein phosphatase 1 regulatory subunits	2	1	0	0	1	0	1	1	2	1	1	0	1	1	3
Protein phosphatase 3 regulatory subunits	2	1	1	0	0	0	0	1	1	1	2	0	0	1	2
Protein phosphatase catalytic subunits	3	3	1	1	2	1	0	1	0	2	0	1	2	4	4
RAS type GTPases	7	3	3	2	2	5	1	1	2	4	2	5	3	11	5
Rho family GTPases	4	3	4	3	1	2	1	2	1	2	1	2	3	9	6
Rho guanine nucleotide exchange factors	6	1	2	3	0	3	0	1	1	1	1	2	1	6	2
SH2 domain containing	21	8	12	4	1	7	1	10	4	10	4	4	8	27	19
Src family tyrosine kinases	2	1	1	0	0	2	0	2	0	2	1	0	2	5	3
Tec family tyrosine kinases	2	1	1	1	0	1	0	1	0	2	0	2	2	4	2
V-set domain containing	2	2	2	1	2	2	1	2	2	2	1	1	2	6	5

** BTK, LYN, SYK*

*** AKT1, MAPK14, PI3K*

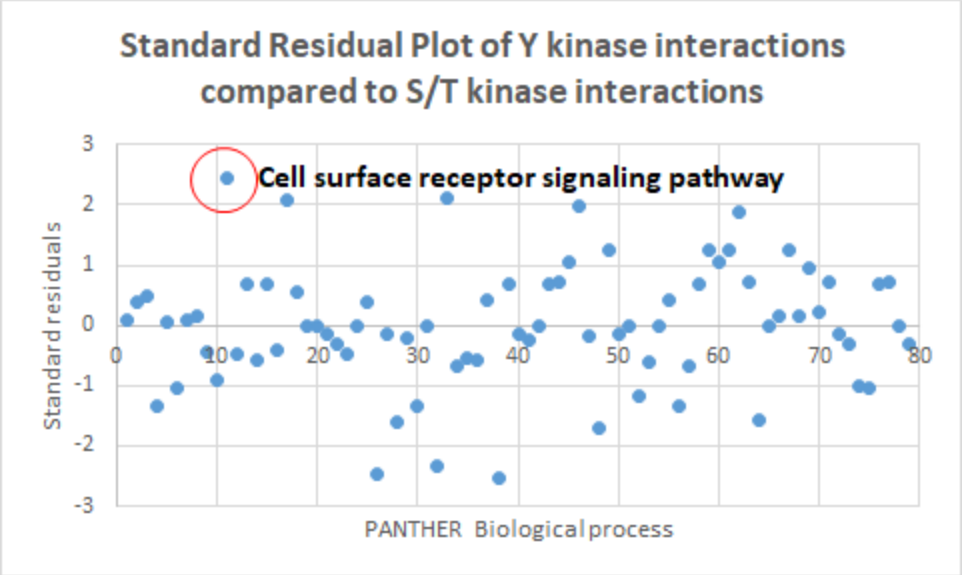


Figure 94. S/T kinases interacted with more targets in the cell surface receptor signaling pathway than Y kinases

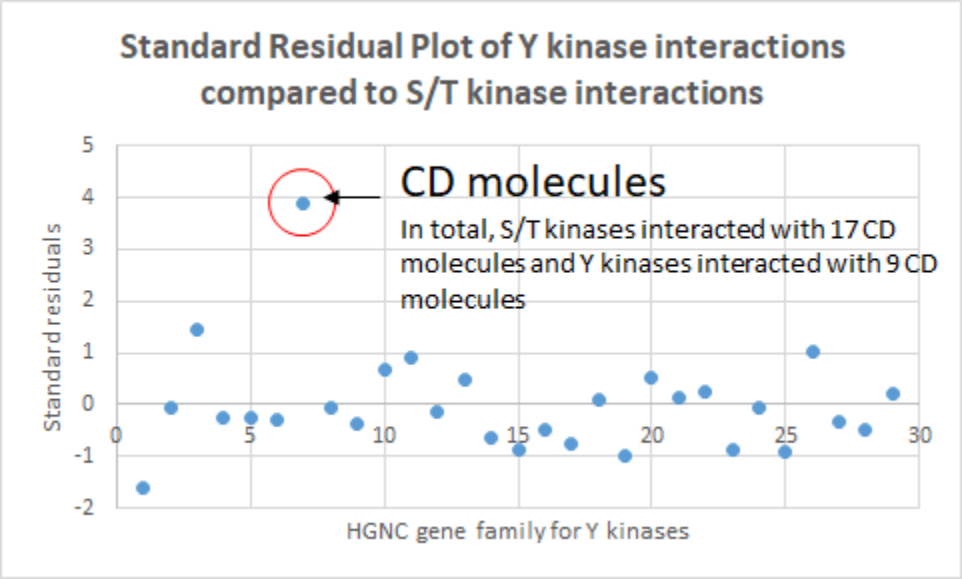


Figure 95. S/T kinases interacted with more CD molecules than Y kinases

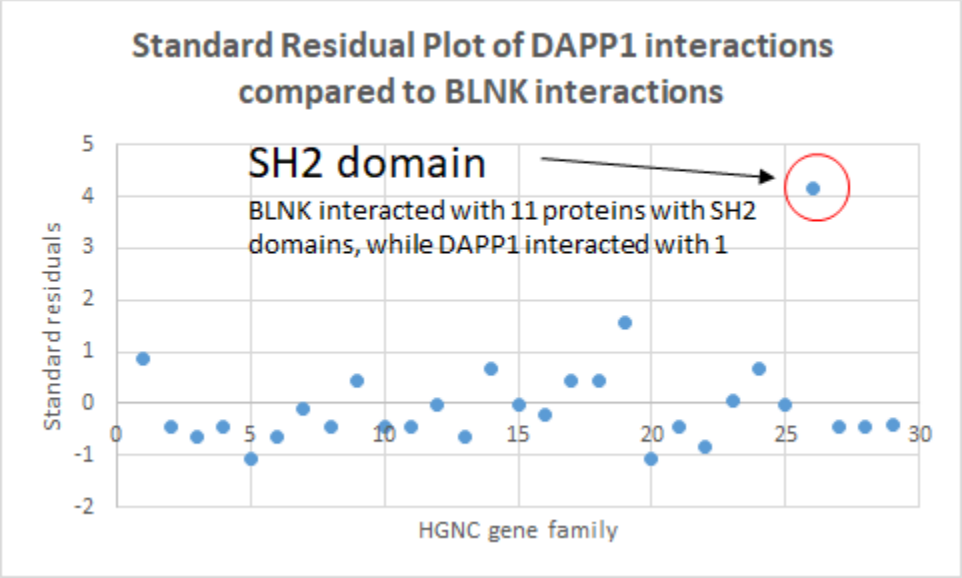


Figure 96. BLNK interacted with more proteins with SH2 domains than DAPP1

APPENDIX E

DIRECTIONS ON USING SPRITE TO ANALYZE SPR DATA

The SPRite script is a flexible software that can theoretically analyze an infinite number of protein interactions using the 1:1 Langmuir kinetic model with and without drift correction. This is in direct contrast to current software packages in which the binding curves must be fit manually in low throughput (e.g., BIAevaluation, Scrubber2). Although SPRite was developed to analyze high throughput NAPPA-SPRi data, it can be used for any SPR experiment (as long as the input file is correctly formatted). Below are directions on how to analyze SPR data with SPRite.

Step 1: Set up computer to run SPRite

2. Make a folder called "StanScripts" in your root directory e.g. c:\StanScripts
3. Download the following files from the Mallick lab website, mallicklab.stanford.edu, and place in the "StanScripts" folder
    a. curveFittingKineticModels.py
    b. OffSets.txt
    c. parseSPRFileAndFitCurves.py
    d. StanScripts.yml
4. Download and install Anaconda2
5. Open Anaconda Prompt and navigate to the "StanScripts" folder

    cd c:\StanScripts
6. Build the environment from the StanScripts.yml file

    conda env create –f StanScripts.yml –n StanScript

Step 2: Format data file

1. Analyze SPR data in Plexera® Data Analysis Module software according to Plexera® instructions
2. Export file as a tab delimited file. (An example of the file format is in Figure 97.)
    a. Column A:
        i. Row 1: Name: *Sample_Name* ID: Set: Family: Group: Block: *Position(must be a unique number)* Row: 1 Column: 1
        ii. Row 2: Relative time *(starts from 0 and continues to increase throughout the SPR analyses)*
    b. Column B:
        i. Row 2: Raw intensity *(This is the pixel intensity of the spot for each time segment)*
    c. Column C:
        i. Row 2: Satellite Intensity *(Sometimes regions around the spots are chosen by SPR users to act as a reference. The satellite intensity is the pixel intensity of these regions. Note that satellites were not used in the analyses.)*
    d. Column D:
        i. Row 2: Subtracted Intensity *(The Subtracted Intensity = Raw Intensity – Satellite Intensity)*

e. Column E: *Leave blank*

	A	B	C	D	E	F	G	H	I	J
1	Name: AKT2 ID: Set: Family: Group: Block: 1 Row: 1 Column: 1					Name: BLK ID: Set: Family: Group: Block: 2 Row: 1 Column: 1				
2	Relative Time	Raw Intensity	Satellite Intensity	Subtracted Intensity		Relative Time	Raw Intensity	Satellite Intensity	Subtracted Intensity	
3	0	0	0	0		0	0	0	0	
4	1	100.32476	0	100.32476		1	100.1549	0	100.1549	
5	2	98.427727	0	98.427727		2	97.978546	0	97.978546	
6	3	95.557419	0	95.557419		3	95.607834	0	95.607834	

Figure 97. Format of input file from the Plexera® Data Analysis Module software for SPRite analyses in text tab delimited format.

```
(stanscript) c:\Stanscripts>python parseSPRFileAndFitCurves.py -h
Usage: parseSPRFileAndFitCurves.py [options]     (Use -h or --help to see all op
tions)

Options:
  -h, --help            show this help message and exit
  -i INFILE, --infile=INFILE
                        input "Processed Prior to scrubber" csv file
                        (REQUIRED)
  -g GLYCALIB, --glycalib=GLYCALIB
                        Glycerol calibration parameter string - format:
                        units,ts1,td1,ts2,td2,dRI
                        Accepted value for units: "au, ru, riu"For example -g
                        ru,780,825,1245,1290,0.000565
  -o TOFILE, --tofile=TOFILE
                        Spot Time offset "txt" file (REQUIRED)
  -c ACONC, --aconc=ACONC
                        Analyte CONCENTRATION (REQUIRED)
  -t TSTART, --tstart=TSTART
                        Time start for SPOT-1 (REQUIRED)
  -a TASSOC, --tassoc=TASSOC
                        Association duration time (REQUIRED)
  -d TDISSOC, --tdissoc=TDISSOC
                        Dissociation duration time (REQUIRED)
  -p PREASSOC, --preassoc=PREASSOC
                        Pre Association time for drift correction (REQUIRED)
  -r REFCOUNT, --refcount=REFCOUNT
                        Number of references (REQUIRED)
  -s REFSPOTS, --refspots=REFSPOTS
                        Reference spots name -
                        values of all the reference
                        spot locations should be separated by a comma
  -m AMASS, --amass=AMASS
                        Analyte MASS (OPTIONAL)
  -f FRATE, --frate=FRATE
                        Flow Rate
  -l MSELECT, --mselect=MSELECT
                        Model select (Options: 1 , lld, lmts, mlld, mglld, hl
                        or all)
  -e DRIFTCORR, --driftcorr=DRIFTCORR
                        Drift correction method (Options: d or ad)
  -n DELIMSEP, --sep=DELIMSEP
                        Separation method for input file (Options: , or
                        or \s)
```

Figure 98. SPRite options displayed within the python terminal.

Step 3: Fit data

1. Go to computer > SYSTEM (C:) > StanScripts
2. Place correct input file of interest in “StanScripts” folder (see Figure 97).

3. Go to Start > Anaconda Prompt >
4. Type *Activate StanScript*
5. Type *cd c:\StanScripts*
6. To see all of the options like Figure 99 within SPRite, add the following command: *python parseSPRFileAndFitCurves.py –h*

   For further explanations about the options, please see "Descriptions about SPRite Options" below
7. Fill in the appropriate data in the command line
   a. python parse SPRFileAndFitCurves.py –i <input_filename> -o <OffSets.txt>-c <conc_analyte> -t <association_start_time_for_block1> -a <total_association_duration_time> -d <total_dissociation_duration_time> -r 3 -s <ref1_block_no.1,ref_block_no.2,ref_block_no.3> -m <mass_analyte> -f <flow_rate> -l <model_type> -g <Units_to_convert_file,Start_time_of_first_calibration,End_time_of_first_calibration,Start_time_of_second_calibration,End_time_of_second_calibration,Conversion_factor>
   b. An example of a command line to analyze a dataset is:

      python parseSPRFileAndFitCurves.py -i SPRdatafile.txt -o OffSets.txt -c 0.0000001166 -t 1500 -a 300 -d 700 –p 100 -r 3 –s405,236,201 -m 51465 -f 3 -l l –g ru,770,820,1240,1285,0.000565

where the inputted data file (-i) is “SPRdatafile.txt,” the concentration of the query (-c) is 0.0000001166 M, the query injection time (-t) is 1500 sec, the association length (-a) is 300 sec, the dissociation length (-d) is 700 sec, the seconds before the query injection (-p) to include in the figures are 100, the number of reference spots (-r) to use is 3, the reference spots are located (-s) in positions 405, 236, and 201, the query mass (-m) is 51465 Da, the flow rate (-f) is 3 uL/sec, the kinetic model (-l) to use is Langmuir, the data will be calibrated and converted (-g) to RU, the start time of the first glycerol injection is 770 sec, the end time of the first glycerol injection is 820 sec, the start time of the second glycerol injection is 1240 sec, the end time of the second glycerol injection is 1285 sec, and the known RIU between the first and second glycerol injections is 0.000565. Also see calibration example in Figure 100 using 0.5% and 1% glycerol where the correct command would be –g ru,115,150,580,625,0.000565.

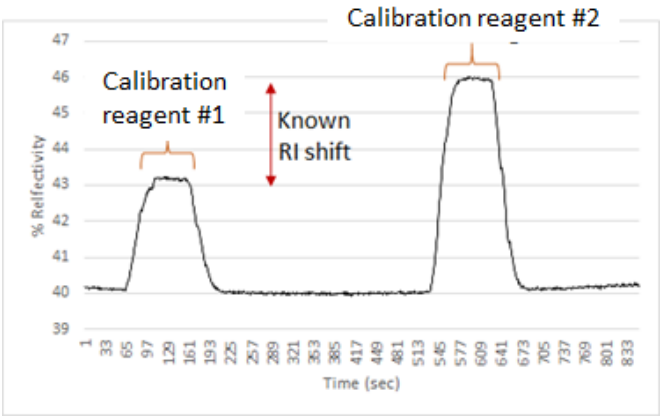


Figure 99. Plateaued responses of two calibration reagents result in a known shift in RI. The responses on the Plexera HT PlexArray instrument are in % reflectivity, or arbitrary units (AU).

8. Output will now be in the “SPRdatafile” folder within the “StanScripts” folder, which will contain:
    a. Folder denoting the model used containing separate PDFs per binding sensorgram
    b. Text file containing all of the kinetic data
    c. Text file containing the calibrated curves over time
    d. Text file that is compatible with Scrubber2
    e. PDF file of all of the sensorgrams

Descriptions about SPRite Options

1. Required input files with flags
    a. -i ,--infile: The input text tab delimited file for processing. The file must be in the correct format (see also Step 2, Figure 97, page 264)
    b. -o, --tofile: Spot Time offset "txt" file (tab-separated). In essence, this file aligns the injection time for each spot. This file is used because only the start time “-t” in which any of the target spots first observe the query is inputted into the command line in SPRite, yet the target protein spots on a SPRi slide will be

	A	B
1	Spot	Offset from t0 in s
2	1	0
3	2	0
4	3	0
5	4	0
6	5	1
7	6	1
8	7	1

Figure 100. Offset file example where column A has the spot number and column B denoted the time offset. If the start time “-t” is “100” in the command line, this offset file tells SPRite that the real start time is “100” for spots 1 – 4 and “101” for spots 5 – 7.

exposed to query protein at different times due to the flow of reagents from one end of the array to the other. The file denotes how the spots are exposed to the query from the start time "-t". The file needs to be made by the user to fit their slide format. In the case with NAPPA-SPRi, the targets close to the injection inlet will be exposed to query a few seconds before the targets close to the injection outlet. Based on observation of RI changes across the slide due to glycerol injections, the offset file was generated. The offset file should look like Figure 101. The time offset for one particular NAPPA-SPRi dataset using the Plexera flow chamber is depicted in Figure 102 in which the flow was 5 µL per second. The offset file should be changed if the flow is altered. For example, the time offset for a NAPPA-SPRi dataset with a

c. flow of 3 µL per second is different than that of 5 µL per second (Figure 102, Figure 103).

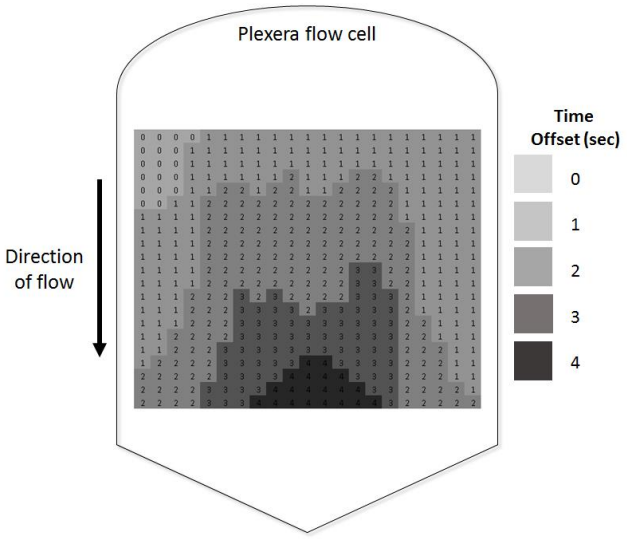


Figure 101. Time offsets for a 21 x 21 spotted array on a Plexera sensor chip using a Plexera flow cell with a dataset having 3 µL/sec flow

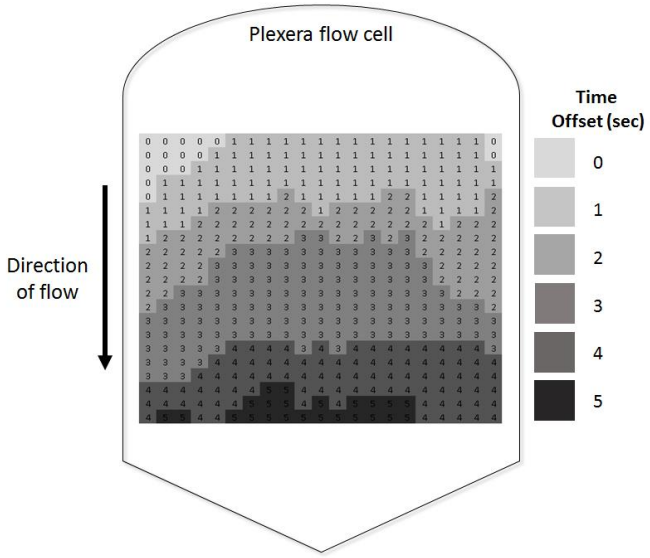


Figure 102. Time offsets for a 21 x 21 spotted array on a Plexera sensor chip using a Plexera flow cell with a dataset having 5 µL/sec flow

2. Required input parameters (strings) with flags
    a. -c, --aconc: Analyte concentration in Molarity.
    b. -t, --tstart: Assoctiation start time start for first array/spot. Calculate the start time for each array using this input and using the "SpotTOffsets.txt" file for each spot (Figure 101).
    c. -a ,--tassoc: Association duration time in seconds
    d. -d, --tdissoc: Dissociation duration time in seconds
    e. -p, --preassoc: Pre-Association time to include in figures
    f. -r, --refcount: Number of references. Currently, 3 references should be used. The response of each reference sensorgram "1, 2, or 3" will be subtracted from each target sensorgram "X" separately, such that "X" will be analyzed with "X-1," "X-2," and "X-3"

g. -s, --refspots: Reference spots name - values of all the reference spot locations should be separated by a comma (for example: 203,350,480).

3. Optional input parameters with flags
    a. -m, --amass: Mass of the analyte in Daltons.
    b. -f --frate: Flow rate as uL/sec.
    c. -l ,--mselect: Model selection for curve fitting. One or more models can be chosen at one time, each one separated using commas. Available options: l (default), lld, lmt, hl or all
        1. *l - Langmuir 1:1; default*
        2. *lmts - Langmuir mass transport (currently in progress, not ready to be used)*
        3. *lld - Langmuir linear drift; PDF files do not display drift corrected curves*
        4. *mlld – Langmuir linear drift; PDF files display drift corrected curves*
        5. *mglld – Langmuir linear drift; Association and dissociation curves are fit globally; PDF files display drift corrected curves*
        6. *hl - Heterogeneous ligand (currently in progress, not ready to be used)*
        7. *all - For now, all runs only l and lmts*

d. –g, --glycalib: Conversion of intensity units to AIU, RIU, or RU. Any type of reagent can be used to calibrate as long as the refractive index shift is known. Glycerol, PBS, and ethanol are common reagents for calibrating data. For the example given in Figure 100, the correct command to convert data into RU would be: -g ru,115,150,580,625,0.000565

   Where the first two numbers indicate the time in which the plateaued response of the first calibration reagent begins and ends, while the second two numbers indicate the time in which the plateaued response of the second calibration reagent begins and ends. The fifth number is the known RI difference between the two calibration reagents, which happen to be 0.5% and 1% glycerol in running buffer.

e. –e, --driftcorr: Drift correction method (Options: d or ad). Default =No drift correction. Options: d (dissociation), ad (association and dissociation). Leave blank if no drift correction is needed.

APPENDIX F

"PARSESPRANDFITCURVES.PY" SCRIPT

```
#!/usr/bin/env python

import re

from os.path import *

from os import getcwd, makedirs

from optparse import *

from pandas import *

from numpy import *

from scipy.optimize import curve_fit

from PyPDF2 import PdfFileReader, PdfFileMerger

import curveFittingKineticModels as CFM

import matplotlib.pyplot as plt

# Integrating SPRuce:

# Convert to RU/RIU based on glycerol injections

def glycerolCalibration(one_prot_au_df,\

                        one_prot_name,\

                glycalib_str, time_colname, riu_to_ru_factor = 1e+6):

        #riu_to_ru_factor = 1e+6 #Domain knowledge

        one_prot_converted_unit_df = DataFrame()

        if glycalib_str:

                conversion_unit, ts1, td1, ts2, td2, dRI = glycalib_str.split(",")
```

```
                ts1, td1, ts2, td2, dRI = int(ts1), int(td1), int(ts2), int(td2), float(dRI)
                tend1=ts1+td1
                tend2=ts2+td2
                ws_t_one_prot_df_ci1 = one_prot_au_df.loc[(one_prot_au_df.time>=ts1)
&\
(one_prot_au_df.time<=td1)]
                avg_au_ci1 = mean(ws_t_one_prot_df_ci1[one_prot_name])
                ws_t_one_prot_df_ci2 = one_prot_au_df.loc[(one_prot_au_df.time>=ts2)
&\

(one_prot_au_df.time<=td2)]
                avg_au_ci2 = mean(ws_t_one_prot_df_ci2[one_prot_name])
                CalFac = dRI / (avg_au_ci2 - avg_au_ci1)
                one_prot_riu_df = one_prot_au_df*CalFac
                if conversion_unit == "riu":
                                one_prot_riu_df[time_colname] =
one_prot_au_df[time_colname]
                                one_prot_converted_unit_df = one_prot_riu_df
                elif conversion_unit == "ru":
                                one_prot_ru_df = one_prot_riu_df*riu_to_ru_factor
                                one_prot_ru_df[time_colname] =
one_prot_au_df[time_colname]
```

```
                                one_prot_converted_unit_df = one_prot_ru_df
            else:
                                print "Entered incompatible conversion unit entered:",
conversion_unit,\
                                                    ". Please select one of \"riu\", \"ru\""
        else:
                        one_prot_converted_unit_df = one_prot_au_df
        return one_prot_converted_unit_df
# Get the slope and intercept values for the preassociation time RU values
# This is to drift-correct the data.
def lFitData(t, response_unit):
                                lfit = polyfit(t,\
                                response_unit,\
                                1, full=True)
        lcoef, lresid, lrank, lsing_values, lrcond = lfit
        # Get slope and intercept from the fit coefs
        response_m, response_c = lcoef
        return response_m, response_c
def calcSlopeIntercept(ws_t_response):
        t = np.array(ws_t_response.iloc[:,0].astype('int'))
        response_unit = np.array(ws_t_response.iloc[:,1])
        slope, intercept = lFitData(t, response_unit)
```

```
        return slope, intercept
def parseTimeOffsetFile(time_offset_file, delim_operator):
        time_offset_df = read_csv(time_offset_file, sep=delim_operator, engine =
'python')
        return time_offset_df
# 0 to #refs-1 need to be looped since python follows zero-indexing
def calcNewTimeForQueryRefsEachSpot(time_offset_df,\
                                        tstart_spot1,\
                                        block_name,\
                                        refn_spots_list,\
                                        refs_count):
        time_offset = time_offset_df[time_offset_df['Spot']==block_name]
        time_offset_value = int(time_offset.iloc[:,1].values)
        query_tstart = tstart_spot1+time_offset_value
        refn_tstarts_list = list()
        for ref_spot_icol in range(0, refs_count):
                ref_offset =
time_offset_df[time_offset_df['Spot']==int(refn_spots_list[ref_spot_icol])]
                ref_offset_value = int(ref_offset.iloc[:,1].values)
                ref_tstart = tstart_spot1+ref_offset_value
                refn_tstarts_list.append(str(ref_tstart))
        query_refn_tstarts_str = ",".join([str(query_tstart)]+refn_tstarts_list)
```

```
        return query_refn_tstarts_str
# Adding time and raw columns to get total column count per protein
# For each protein:
# First cell in each protein column has spot information
# Python follows zero indexing, so we split time+raw+#refs-1
# Get rid of the row with second header
# If protein name is missing in the header info. Just use "1" instead!
def splitByProteins(df, time_offset_df,\
                                                    analyte_conc,\
                                                    analyte_mass,\
                                                    flow_rate,\
                                                    tstart_spot1,\
                                                    t_assoc, t_dissoc,\
                                                    preassoc_time, refs_count,\
                                                    refn_spots_str,
time_colname):
        one_protein_col_count = refs_count+2
        refn_spots_list = refn_spots_str.split(",")
        split_cells_index = arange(0,len(df.columns),\
                                                     one_protein_col_count)
        data_by_protein_dict = dict()
        refn_spots_df = DataFrame()
```

```
        refn_spots_df_list = list()
        # Get reference spots only to fill in the columns after the reference for scrubber
format:
        refn_spots_block_list = ["Block: " + one_ref_spot_block for one_ref_spot_block
in refn_spots_list]
        for one_ref_spot_block in refn_spots_block_list:
                one_ref_spot_colname =
df.columns[df.columns.to_series().str.contains(one_ref_spot_block)][0]
                one_ref_spot_colname_loc =
df.columns.get_loc(one_ref_spot_colname)
                one_ref_spots_df = df.iloc[:,one_ref_spot_colname_loc+1]
                refn_spots_df_list.append(one_ref_spots_df)
        refn_spots_df = concat(refn_spots_df_list, axis=1)
        #refn_spots_df = refn_spots_df.convert_objects(convert_numeric = True)
        for i in split_cells_index:
                df_by_protein = df.iloc[:,i:(i+one_protein_col_count)]
                prot_array_info = df_by_protein.columns[0]
                prot_block_match = re.match(r"Name: (.*?) .*Block: (.*?) .*",
prot_array_info)
                prot_name = "NoName"
                block_name = "1"
                if prot_block_match:
```

```
                    if prot_block_match.group(1):
                        prot_name = prot_block_match.group(1)
                        prot_name = prot_name.replace(" ", "") # Might not need
this!
                    if prot_block_match.group(2):
                        block_name = prot_block_match.group(2)
                prot_block_name = prot_name + "_" + block_name
                if block_name in refn_spots_list:
                        continue
                block_name = int(block_name)
                df_by_protein = concat([df_by_protein.iloc[:,[0,1]], refn_spots_df],axis =
1)
                df_by_protein = df_by_protein[1:len(df_by_protein)]
                df_by_protein = df_by_protein.convert_objects(convert_numeric = True)
                query_refn_tstarts_str =
calcNewTimeForQueryRefsEachSpot(time_offset_df,\
                tstart_spot1,\
                block_name,\
                refn_spots_list,\
                refs_count)
```

```
                refn_spots_name_list =
[prot_block_name+"_ref"+str(one_ref_spot+1)+"_"+str(refn_spots_list[one_ref_spot])\
                        for one_ref_spot in range(0, refs_count)]
                prot_block_refn_spots_name_str =
",".join([prot_block_name]+refn_spots_name_list)
                pro_kinetics_params_info = [prot_block_refn_spots_name_str,\
                                                        str(t_assoc),\
                                                        str(t_dissoc),\
                                                        str(preassoc_time),\
                                                        str(analyte_conc),\
                                                        str(analyte_mass),\
                                                        str(flow_rate),\
                                                        query_refn_tstarts_str]
                pro_kinetics_params_info_str = "|".join(pro_kinetics_params_info)
                df_by_protein.columns = [time_colname, prot_block_name]+\
                                                                refn_spots_name_list
                data_by_protein_dict[pro_kinetics_params_info_str] = df_by_protein
        return(data_by_protein_dict)
def getProKineticsParamsAndIntensities(data_by_protein_dict):
        pro_kinetics_params_info_str = data_by_protein_dict.keys()
        for one_pro_kinetics_params_info_str in pro_kinestics_params_info_str:
```

```
                df_time_raw_refs =
data_by_protein_dict[one_pro_kinetics_params_info_str]
                prot_name, t_start, t_assoc, t_diss,\
                analyte_mass, analyte_conc, p_calib, p_driftCorr,\
                refn_spots_tstart, ref2_tstart, ref3_tstart =
one_pro_kinetics_params_info_str.split("_")
        return prot_name, t_start, t_assoc, t_diss,\
                                analyte_mass, analyte_conc, p_calib, p_driftCorr,\
                                refn_spots_tstart, ref2_tstart, ref3_tstart, df_time_raw_refs
def get_raw_ref_intensities(df_time_raw_refs):
        relative_time = df_time_raw_refs.iloc[:,0]
        time_raw_intensities = df_time_raw_refs.iloc[:,[0,1]]
        time_refn_spots_intensities = df_time_raw_refs.iloc[:,[0,2]]
        time_ref2_intensities = df_time_raw_refs.iloc[:,[0,3]]
        return relative_time, time_raw_intensities,\
                                        time_refn_spots_intensities, time_ref2_intensities
def windowSelectValues(df_time_raw_refs,\
                        preassoc_tstart,\
                        preassoc_tend,\
                         assoc_tstart,\
                         dissoc_tend,\
                         one_prot_colname, time_colname):
```

```
        ws_t_one_prot_df_preassoc =
df_time_raw_refs.loc[(df_time_raw_refs.time>=preassoc_tstart) &\
                                        (df_time_raw_refs.time<=preassoc_tend),\
                                [time_colname, one_prot_colname]]
        ws_t_one_prot_df_ad =
df_time_raw_refs.loc[(df_time_raw_refs.time>=assoc_tstart) &\
                                        (df_time_raw_refs.time<=dissoc_tend),\
                                        [time_colname, one_prot_colname]]
        ws_t_one_prot_df_pread =
df_time_raw_refs.loc[(df_time_raw_refs.time>=preassoc_tstart) &\
                                        (df_time_raw_refs.time<=dissoc_tend),\
                                                [time_colname, one_prot_colname]]
        return ws_t_one_prot_df_preassoc,\
                                ws_t_one_prot_df_ad,\
                                ws_t_one_prot_df_pread
# Zeroed time to association start time
def bcValues(window_selected_df, t_start):
        baseline_correction_var =
window_selected_df[(window_selected_df.time==t_start)]
        bc_values_df = window_selected_df-baseline_correction_var.values.squeeze()
        bc_values_df = bc_values_df.reset_index(drop=True)
        return bc_values_df
```

```
# subtract slope*(t-t0) it from the association/dissociation
# The assoc/dissoc base correction depends on given start time
def driftCorrectInputBased(ws_t_one_prot_df_pread,\
                                                ws_t_one_prot_df_preassoc,\
                                                 tstart, tend, time_colname):
        ws_t_one_prot_df_d =
ws_t_one_prot_df_pread[(ws_t_one_prot_df_pread.time>=tstart) &\
                (ws_t_one_prot_df_pread.time<=tend)]
        preassoc_prot_m, preassoc_prot_c =
calcSlopeIntercept(ws_t_one_prot_df_preassoc)
        prot_td0 =
ws_t_one_prot_df_pread.ix[ws_t_one_prot_df_pread.time==tstart,0].values
        one_prot_dc = ws_t_one_prot_df_d.iloc[:,1]-
(preassoc_prot_m*(ws_t_one_prot_df_d[time_colname]-prot_td0))
        # Copy df to different variable and apply drift correction to the new variable
        dc_t_one_prot_df_pread = ws_t_one_prot_df_pread.copy(deep=True)
        dc_t_one_prot_df_pread.ix[(dc_t_one_prot_df_pread.time>=tstart) &\
        (dc_t_one_prot_df_pread.time<=tend),1] = one_prot_dc
        return dc_t_one_prot_df_pread
# Drift correct the dissociation phase data
# Find New start time before Zeroing data-may reset the time again but later
# Get slope (m) and intercept (c) for the prior to assoc values
```

```
# Call the function to correct drift based on user input
### Options: 1. Drift correct Assoc+Dissoc 2. Dissoc ONLY 3. NO correction!
# Loop for the query and reference spots
# Note: Make a copy of dataframe, then update the dissociation phase
def dcValuesAndNewTstart(ws_t_one_prot_df_preassoc,\
                                        ws_t_one_prot_df_pread,\
                                        t_assoc,\
                                        assoc_tstart, dissoc_tend,\
                                        input_dc_method, time_colname):
        dissoc_tstart = assoc_tstart+t_assoc+1
        if input_dc_method == "d" :
                dc_t_one_prot_df_pread =
driftCorrectInputBased(ws_t_one_prot_df_pread,\
        ws_t_one_prot_df_preassoc,\
        dissoc_tstart, dissoc_tend, time_colname)
        elif input_dc_method == "ad":
                dc_t_one_prot_df_pread =
driftCorrectInputBased(ws_t_one_prot_df_pread,\
        ws_t_one_prot_df_preassoc,\
                                        assoc_tstart, dissoc_tend, time_colname)
        else:
                dc_t_one_prot_df_pread = ws_t_one_prot_df_pread
```

```
        return dc_t_one_prot_df_pread
def refQuery(input_values_query, input_values_ref):
        referenced_query = input_values_query - input_values_ref
        return referenced_query
# We need to subtract each reference from query
# But since we have a df of all values,
# Just reference the df by subtracting the entire df with the query
# Note: query column becomes zero
# Then negative correct the value and update the df to add time and set bc query values
# Finally, rearrange all the columns to form the sequence=> time,query, refd1, refd2....
def refQueryDF(bc_t_prot_allrefs_df,\
                                                prot_block_refn_spots_name_list,\
                                                refs_count, time_colname):
        query_prot_name = prot_block_refn_spots_name_list[0]
        refn_spots_name_list =
prot_block_refn_spots_name_list[1:len(prot_block_refn_spots_name_list)]
        time_only_df = bc_t_prot_allrefs_df[time_colname].T.drop_duplicates().T #
Remove multiple time columns! => Issues!
        # Hack: Just incase the time series sequence breaks at some point
        if len(time_only_df.columns) > 1 :
                time_only_df = time_only_df.iloc[:,0]
```

```
        time_refd_all_df = -(bc_t_prot_allrefs_df[refn_spots_name_list].sub(\
                                                bc_t_prot_allrefs_df[query_prot_name],\
                                                 axis=0))
        time_refd_all_df[time_colname] = time_only_df.astype(int)
        time_refd_all_df = time_refd_all_df[[time_colname]+refn_spots_name_list]
        time_query_df = concat([time_only_df, bc_t_prot_allrefs_df[query_prot_name]],
axis=1)
        return time_refd_all_df, time_query_df, query_prot_name, refn_spots_name_list
# query_tstart is assoc_start_time query_tend is dissoc_end_time
# Make a dictionary with protein nametype (query, ref1 ..) and their corresponding start
times
# Reference the query protein by subtracting query values with given reference values
def windowSelectBCRefValues(one_prot_df,\
                                                preassoc_tstart,\
                                                preassoc_tend,\
                                                t_assoc,\
                                                assoc_tstart,\
                                                dissoc_tend,\
                                                one_prot_colname,\
                                                input_dc_method, time_colname) :
        ws_t_one_prot_df_preassoc,\
```

```
        ws_t_one_prot_df_ad, ws_t_one_prot_df_pread =
windowSelectValues(one_prot_df,\
        preassoc_tstart,\
        preassoc_tend,\
        assoc_tstart,\
        dissoc_tend,\
        one_prot_colname, time_colname)
        dc_t_one_prot_df_pread = dcValuesAndNewTstart(ws_t_one_prot_df_preassoc,\
                                                      ws_t_one_prot_df_pread,\
                                                      t_assoc,\
                                                      assoc_tstart,\
                                                      dissoc_tend,\
                                                      input_dc_method,
time_colname)
        # Baseline correct/Zeroing data
        bc_t_one_prot_df = bcValues(dc_t_one_prot_df_pread, assoc_tstart)
        return bc_t_one_prot_df
def callCF(df_time_allrefs_ad,\
                                    df_time_raw,\
                                    analyte_conc,\
                                    analyte_mass,\
                                    flow_rate,\
```

```
                                        t_assoc, t_dissoc,\
                                        assoc_tstart, dissoc_tend,\
                                        plot_filename,\
                                        base_input_filename,\
                                        plots_file_path,\
                                        fit_type,\
                                        refs_spots_list,\
                                        merge_onefit_pdfs,time_colname) :
        all_coefs_fits_df_list = list()
        print "Curve fitting...."
        coefs_df = DataFrame()
        response_unit_pread = ""
        try:
                response_unit_pread, coefs_df =
CFM.fitCurveModels(df_time_allrefs_ad,\
        df_time_raw,\
        analyte_conc,\
        analyte_mass,\
        flow_rate,\
        t_assoc, t_dissoc,\
        assoc_tstart, dissoc_tend,\
        plot_filename, fit_type,\
```

```
        base_input_filename, plots_file_path,\
        refs_spots_list, merge_onefit_pdfs, time_colname)
        except ValueError:
                print "Skipping array protein. Data too noisy..."
        all_coefs_fits_df_list.append(coefs_df)
        all_coefscolumnfits_df = concat(all_coefs_fits_df_list, axis=1)
        all_fits_coefs_df =      concat(all_coefs_fits_df_list)
        return response_unit_pread, all_fits_coefs_df, all_coefscolumnfits_df
# Get the model type and populate "fit_types_list":
def getFitType(input_model_selected,\
                                                fit_types_list):
        # Basic model, in place and use with all. Make this default!
        if input_model_selected == "l" or input_model_selected == "all":
                fit_types_list.append("l")
        # In place and tested but NOT included with default-"all" option yet!
        if input_model_selected == "lld":
                fit_types_list.append("lld")
        # In place but not satisfied and NOT included with default-"all" option yet!
        if input_model_selected == "mlld":
                fit_types_list.append("mlld")
        # In place but testing and NOT included with default-"all" option yet!
        if input_model_selected == "mglld":
```

```
            fit_types_list.append("mglld")
        # NOT in place! Weird outputs
        if input_model_selected == "lmt":
            fit_types_list.append("lmt")
        # Working and validation in process!
        if input_model_selected == "lmts" or input_model_selected == "all":
            fit_types_list.append("lmts")
        # In place but NOT validated!
        if input_model_selected == "hl":
            fit_types_list.append("hl")
        return 0
def processSPRFileAndFitCurves(input_filename, time_offset_file,\
                               analyte_conc, tstart_spot1,\
                               t_assoc, t_dissoc_input,\
                               preassoc_time, refs_count,\
                               refn_spots_str, analyte_mass,\
                               flow_rate, input_model_selected,\
                               input_dc_method, delim_operator,
glycalib_str):
        # Brute force: Resetting the dissociation time by decreasing the tail by 5 seconds
        t_dissoc = t_dissoc_input - 5
        time_colname = "time"
```

```
fit_types_list = list()
getFitType(input_model_selected, fit_types_list)
# Need to set this properly
current_dirpath = getcwd()
base_input_filename = basename(input_filename)
base_input_filename_noext = splitext(base_input_filename)[0]
results_subdir = "results_"+base_input_filename_noext
results_subdir_path = join(current_dirpath, results_subdir)
# Create a subdirectory for each file under current directory (if not present):
print "Creating sub-directory: \n", results_subdir_path
if not exists(results_subdir_path):
        makedirs(results_subdir_path)
all_spots_fits_coefs_df = DataFrame()
all_spots_columnfits_coefs_df = DataFrame()
# Loop over all the fit types requested (used when "all" option is selected).
for fit_type in fit_types_list:
        plots_file_path = join(results_subdir_path, fit_type)
        print "Creating sub-directory for model: \n",\
                                fit_type, "at:", results_subdir_path
        if not exists(plots_file_path):
                makedirs(plots_file_path)
        time_offset_df = parseTimeOffsetFile(time_offset_file, delim_operator)
```

```
        # Read input processed prior to scrubber file
        input_df = read_csv(input_filename, sep=delim_operator, engine =
'python')#, skiprows=0)
        data_by_protein_dict = splitByProteins(input_df, \
                                               time_offset_df,\
                                               analyte_conc,\
                                              analyte_mass,\
                                               flow_rate,\
                                               tstart_spot1,\
                                               t_assoc, t_dissoc,\
                                              preassoc_time,\
                                               refs_count,\
                                               refn_spots_str, time_colname)
        pro_kinetics_params_info_str = sorted(data_by_protein_dict.keys())
        merge_onefit_pdfs = PdfFileMerger() # open pdf merger document!
        all_plots_merged_filename = fit_type+"_"+\
                                                  base_input_filename_noext+\
                                                      "_all_merged.pdf"
        bc_t_prot_allrefs_df_filt_list = list()
        output_prot_df_list = list()
        for one_pro_kinetics_params_info_str in pro_kinetics_params_info_str:
```

```
                    df_time_raw_refs =
data_by_protein_dict[one_pro_kinetics_params_info_str]
                    prot_block_refn_spots_name_str,\
                    t_assoc, t_dissoc, preassoc_time,\
                    analyte_conc, analyte_mass,\
                    flow_rate, query_refn_tstarts_str =
one_pro_kinetics_params_info_str.split("|")
                    t_assoc, t_dissoc, preassoc_time = int(t_assoc),\
                                                       int(t_dissoc),\
                                                    int(preassoc_time)
                    prot_block_refn_spots_name_list =
                    prot_block_refn_spots_name_str.split(",")
                    query_refn_tstarts_list = map(int, query_refn_tstarts_str.split(","))
                    prot_names_tstarts_dict =
          dict(zip(prot_block_refn_spots_name_list,\
                        query_refn_tstarts_list))
                    bc_all_t_prot_df_list = list()
                    array_protein_name = prot_block_refn_spots_name_list[0]
                    one_prot_all_list = list()
                    for one_prot_colname in prot_names_tstarts_dict.keys():
                        assoc_tstart = prot_names_tstarts_dict[one_prot_colname]
                        dissoc_tend = assoc_tstart+t_assoc+t_dissoc
```

```
                    preassoc_tstart = assoc_tstart-preassoc_time
                    preassoc_tend = assoc_tstart-1
                    one_prot_au_df = df_time_raw_refs[[time_colname,
one_prot_colname]]
                    one_prot_df = glycerolCalibration(one_prot_au_df,\
                                                      one_prot_colname,\
                                                      glycalib_str, time_colname)
                    one_prot_all_list.append(one_prot_df)
                    bc_t_one_prot_df =
windowSelectBCRefValues(one_prot_df,\
                    preassoc_tstart,\
                    preassoc_tend,\
                    t_assoc,\
                    assoc_tstart,\
                    dissoc_tend,\
                    one_prot_colname,\
                    input_dc_method, time_colname)
                    bc_all_t_prot_df_list.append(bc_t_one_prot_df)
                bc_t_prot_allrefs_df = concat(bc_all_t_prot_df_list, axis=1) # We
will have duplicates for "time" column here
                print "Protein=>"+array_protein_name+"....."
                # Reference the values (query-ref1, query-ref2...)
```

```
                time_refd_all_df, time_query_df,\
                query_prot_name, refn_spots_name_list =
refQueryDF(bc_t_prot_allrefs_df,\
                                        prot_block_refn_spots_name_list,\
                                            refs_count, time_colname)
                # Reset the association start value to zero since we zeroed the data
to start from zero
                response_unit_pread, all_fits_coefs_df, all_coefscolumnfits_df
        = callCF(time_refd_all_df,\
                                        time_query_df,\
                                        float(analyte_conc),\
                                        float(analyte_mass),\
                                        float(flow_rate),\
                                        int(t_assoc),\
                                        int(t_dissoc),\
                                        int(assoc_tstart),\
                                        int(dissoc_tend),\
                                        query_prot_name,\
                                        base_input_filename,\
                                        plots_file_path,\
                                        fit_type,\
```

```
                                        refn_spots_name_list,\
                                        merge_onefit_pdfs,time_colname)
                all_spots_fits_coefs_df =
all_spots_fits_coefs_df.append(all_fits_coefs_df)
                #### scrubber validation: START
                #### Super crude hack!
                # Filter patterns for Brianne: Ref1 with -C and Ref2 with -N for
easier filtering
                subset_col_indices = [i for i, x in
enumerate(prot_block_refn_spots_name_list) if re.search(r'-C.*ref1|-N.*ref2', x)]
                if len(subset_col_indices) < 1:
                                subset_col_indices = [1]
                time_prot_block_selrefn_spots_name_list =
[prot_block_refn_spots_name_list[i] for i in [0]+subset_col_indices]
                time_prot_block_selrefn_spots_name_list.insert(0, time_colname)
                bc_t_prot_allrefs_df_filt = bc_t_prot_allrefs_df
                bc_t_prot_allrefs_df_filt[time_colname] =
bc_t_prot_allrefs_df_filt.index
                bc_t_prot_allrefs_df_filt =
bc_t_prot_allrefs_df_filt[time_prot_block_selrefn_spots_name_list].T.drop_duplicates().
T
                bc_t_prot_allrefs_df_filt.iloc[:,1] = response_unit_pread
```

```
                    bc_t_prot_allrefs_df_filt.iloc[:,2] = 0
                    bc_t_prot_allrefs_df_filt['blank1'] = ""
                    bc_t_prot_allrefs_df_filt['blank2'] = ""
                    blank_df = DataFrame([['Relative Time','RAW
DATA','REFERENCE DATA', 'BLANK1', 'BLANK2']],\
                                columns=bc_t_prot_allrefs_df_filt.columns.values)
                    bc_t_prot_allrefs_df_filt = concat([blank_df,
bc_t_prot_allrefs_df_filt], ignore_index=True)
                    split_array_prot_info_list =
bc_t_prot_allrefs_df_filt.columns.values[2].split("_")
                    #Convert the header to scrubber format:
                    scrubber_first_header_prot_col = "Name:
"+split_array_prot_info_list[0]+"       ID:              Set:              Family:
        Group:          Block: "+split_array_prot_info_list[1]+"       Row: 1
Column: 1"
                    scrubber_first_header_all_cols = [scrubber_first_header_prot_col,
"", "", "","",""]
                    bc_t_prot_allrefs_df_filt.columns = scrubber_first_header_all_cols
                    bc_t_prot_allrefs_df_filt_list.append(bc_t_prot_allrefs_df_filt)
                    # Plexera output format?
                    one_prot_all_df = concat(one_prot_all_list, axis=1) # We will
have duplicates for "time" column here
```

```
                    one_prot_all_nodup_df =
one_prot_all_df[prot_block_refn_spots_name_list]
                    one_prot_all_nodup_df[[time_colname]] =
one_prot_all_df[time_colname].T.drop_duplicates().T
                    one_prot_all_nodup_df = one_prot_all_nodup_df[[time_colname]
+ prot_block_refn_spots_name_list]
                    blank_df = DataFrame([['Relative Time','Raw Intensity','Satellite
Intensity','Subtracted Intensity', '']],\
                                        columns=one_prot_all_nodup_df.columns.values)
                    one_prot_all_nodup_df = concat([blank_df,
one_prot_all_nodup_df], ignore_index=True)
                    one_prot_all_nodup_df.columns = scrubber_first_header_all_cols
                    output_prot_df_list.append(one_prot_all_nodup_df)
            # Combine all pdfs for one model/fit-type into one pdf for Brianne!
            # Slightly crude!
            all_pdfs = plots_file_path
            all_plots_merged_file = join(results_subdir_path,
all_plots_merged_filename)
            print "Merging all pdfs for the selected model:" + all_plots_merged_file
            merge_onefit_pdfs.write(all_plots_merged_file)
            # Get raw baselined data, reference spot data out for Brianne+Ian to cross-
check with scrubber:
```

```
            scrubber_format_df = concat(bc_t_prot_allrefs_df_filt_list, axis=1)
            output_prot_df = concat(output_prot_df_list, axis=1)
            ### Write an output file in scrubber format with only the binding curve
data for validation
            scrubber_format_file = join(results_subdir_path,\
                                                input_model_selected+\
                                                "_ScrubberFormat_"+\
                                            base_input_filename)
            scrubber_format_df.to_csv(scrubber_format_file,\
                                                index=False, sep="\t")
            ### Write an intermediate output file in plexera format? with input data
for validation across multiple platforms
            output_prot_file = join(results_subdir_path,\
                                                input_model_selected+\
                                            "_OutputData_"+\
                                                base_input_filename)
            output_prot_df.to_csv(output_prot_file,\
                                            index=False, sep="\t")
            ####Scrubber validation hack: END!
        all_spots_fits_coefs_file = join(results_subdir_path,\
```

```
                                        input_model_selected+\
                                        "_Format1coefs_"+\
                                                base_input_filename)
        print "Creating all in one file-Format1:" + all_spots_fits_coefs_file
        all_spots_fits_coefs_df.to_csv(all_spots_fits_coefs_file,\
                                                                        index=False,
sep="\t")
        print "Results located at:", all_spots_fits_coefs_file
        return 0
#
# Read command line options
#
def readAndParseCommandlineArgs():
        usage = "usage: %prog [options]        (Use -h or --help to see all options)"
        cl=OptionParser(usage=usage)
        cl.add_option('--infile', '-i', action='store',
                        help="input \"Processed Prior to scrubber\" csv file
(REQUIRED)",
                        dest="infile")
        cl.add_option('--glycalib', '-g', action='store',
                        help="Glycerol calibration parameter string - format:
units,ts1,td1,ts2,td2,dRI\
```

```
                                    Accepted value for units: \"au, ru, riu\"For example
-g ru,780,825,1245,1290,0.000565",
                                    dest="glycalib", default="")
        cl.add_option('--tofile', '-o', action='store',
                                    help="Spot Time offset \"txt\" file (REQUIRED)",
                                    dest="tofile")
        cl.add_option('--aconc', '-c', action='store',
                                    help="Analyte CONCENTRATION
(REQUIRED)",
                                    dest="aconc")
        cl.add_option('--tstart', '-t', action='store',
                                    help="Time start for SPOT-1 (REQUIRED)",
                                    dest="tstart")
        cl.add_option('--tassoc', '-a', action='store',
                                    help="Association duration time (REQUIRED)",
                                    dest="tassoc")
        cl.add_option('--tdissoc', '-d', action='store',
                                    help="Dissociation duration time (REQUIRED)",
                                    dest="tdissoc")
        cl.add_option('--preassoc', '-p', action='store',
                                    help="Pre Association time for drift correction
(REQUIRED)",
```

```
                                        dest="preassoc")
        cl.add_option('--refcount', '-r', action='store',
                                        help="Number of references (REQUIRED)",
                                        dest="refcount")
        cl.add_option('--refspots', '-s', action='store',
                                        help="Reference spots name - \
                                        values of all the reference\
                                        spot locations should be separated by a comma",
                                        dest="refspots")
        cl.add_option('--amass', '-m', action='store',
                                        help="Analyte MASS (OPTIONAL)",
                                        dest="amass")
        cl.add_option('--frate', '-f', action='store',
                                        help="Flow Rate",
                                        dest="frate")
        cl.add_option('--mselect', '-l', action='store',
                                        help="Model select (Options: l , lld, lmts, hl or
all)",
                                        dest="mselect", default="l")
        cl.add_option('--driftcorr', '-e', action='store',
                                        help="Drift correction method (Options: d or ad)",
                                        dest="driftcorr")
```

```
        cl.add_option('--sep', '-n', action='store'
                                        help="Separation method for input file (Options: ,
or \t or \s)",
                                        dest="delimsep", default="\t")
        (options, args) = cl.parse_args()
### Need to add defaults for flexibility
        #
        # Check the command line options
        if options.infile:
                if isfile(options.infile):
                        input_filename = options.infile
                        glycalib_str = str(options.glycalib)
                        time_offset_file = options.tofile
                        analyte_conc = float(options.aconc)
                        tstart_spot1 = int(options.tstart)
                        t_assoc = int(options.tassoc)
                        t_dissoc = int(options.tdissoc)
                        preassoc_time = int(options.preassoc)
                        refs_count = int(options.refcount)
                        refn_spots_str = str(options.refspots)
                        analyte_mass = options.amass
                        flow_rate = options.frate
```

```
                model_select = options.mselect
                input_dc_method = options.driftcorr
                delim_operator = str(options.delimsep)
                return(input_filename, time_offset_file,\
                                        analyte_conc, tstart_spot1,\
                                        t_assoc, t_dissoc,\
                                        preassoc_time, refs_count,\
                                        refn_spots_str, analyte_mass,\
                                        flow_rate, model_select,\
                                        input_dc_method, delim_operator,\
                                        glycalib_str)
        else:
                cl.error("Please specify an input        \"Processed Prior to scrubber\" csv
file to run the search. Use -h for more information.\n")
# Any line with "bc_ws_t_one_prot_df" variable part is only temporary for now.
# This is to check total raw biacore units with drift corrected values
def main():
        input_filename, time_offset_file,\
        analyte_conc, tstart_spot1,\
        t_assoc, t_dissoc_input,\
        preassoc_time, refs_count,\
        refn_spots_str, analyte_mass,\
```

```
        flow_rate, input_model_selected,\
        input_dc_method, delim_operator, glycalib_str =
readAndParseCommandlineArgs()
        process_results = processSPRFileAndFitCurves(input_filename, time_offset_file,\
                                                     analyte_conc,
tstart_spot1,\
        t_assoc, t_dissoc_input,\
                preassoc_time, refs_count,\
                        refn_spots_str, analyte_mass,\
                                                     flow_rate,
input_model_selected,\
                                                     input_dc_method,
delim_operator, glycalib_str)
        print process_results, "Done!"
if __name__ == "__main__":
         # stuff only to run when not called via 'import' here
         main()
#### Script command example:
##### python parseSPRFileAndFitCurves.py -i ../SPR\ analyses/Additional\
Data/BTK_Deph_TEMP.csv -o ../SPR\ analyses/20150527_SpotTOffsets.csv -c 9e-08 -t
2570 -a 180 -d 400 -s 100,100,100 -p 100 -m 77800 -r 3 -f 5 -l all -e d -n , (or -n $'\t' or -n
$'\s')
```

```
##### python parseSPRFileAndFitCurves.py -i ../SPR\
analyses/20151028ASU/20151009_RAC1_reproducibility/09102015_RAC1_GTP_4ug_
deph_3.txt \
#####                                                                -o
../SPR\ analyses/20150527_SpotTOffsets.csv -c 1.78174E-07 -m 22450 \
#####                                                                -r 3 -f
3 -l l -t 2806 -a 300 -d 700 -s 405,201,236 -p 100 -n $'\t'
```

APPENDIX G

“CURVEFITTINGKINETICMODELS.PY” SCRIPT

```
#!/usr/bin/env python
import re
from pandas import *
from numpy import *
from os.path import join, exists
from os import makedirs
from scipy.optimize import curve_fit#, leastsq
from scipy.integrate import odeint
from scipy.stats import chisquare
from collections import OrderedDict
import matplotlib.pyplot as plt
from PyPDF2 import PdfFileReader, PdfFileMerger
from parseSPRFileAndFitCurves import lFitData
def assocEqLang1(t, Req, kobs, X):
        t0 = t[0]
        r_assoc = (Req*(1-exp(-kobs*(t-t0))))+X
        return r_assoc
def dissocEqLang1(t, R0, kd, X):
        t0 = t[0]
        #t = t-t0 #=> Zeroing dissoc so they go from 0 to 699 instead of 301-1000
        r_dissoc = (R0*exp(-kd*(t-t0)))+X
        return r_dissoc
```

```
def assocEqLLD(t, Req, kobs, Y, X):
        t0 = t[0]
        t_assoc = t-t0
        r_assoc = (Req*(1-exp(-kobs*t_assoc)))+(Y*t_assoc)+X
        return r_assoc
def dissocEqLLD(t, R0, kd, Y, X):
        t0 = t[0]
        t_dissoc = t-t0
        #t = t-t0 # Zeroing dissoc so they go from 0 to 699 instead of 301-1000
        r_dissoc = (R0*exp(-kd*t_dissoc))+(Y*t_dissoc)+X
        return r_dissoc
## Langmuir 1:1 with linear drift global equation
def solveAssocDissocEqsGLLD(assoc_init, dissoc_init,\
                                                        analyte_conc, fit_type,
plot_prefix):
        t_assoc_only, ru_assoc_only, Req0, ka0, Y0_assoc, X0_assoc = assoc_init
        t_dissoc_only, ru_dissoc_only, R0, kd0, Y0_dissoc, X0_dissoc = dissoc_init
        t_assoc_dissoc = np.concatenate([t_assoc_only, t_dissoc_only])
        ru_assoc_dissoc = np.concatenate([ru_assoc_only, ru_dissoc_only])
        #ru_assoc_dissoc = ru_assoc_dissoc-ru_assoc_dissoc[0]
        assoc_slope_limit = (Req0*20)/(100*t_assoc_only[len(t_assoc_only)-1])
        ## Add R0 params for dissoc => last element from calc assoc values
```

```
        dissoc_slope_limit = (Req0*10)/(100*t_dissoc_only[0])
        assoc_dissoc_param_bounds = ((0, 0, 0,-assoc_slope_limit,-np.inf),(np.inf, np.inf,
np.inf, assoc_slope_limit, np.inf))
        assoc_dissoc_init_params = [Req0, ka0, kd0, Y0_assoc, Y0_dissoc]
        t_assoc_start = t_assoc_only[0]
        t_assoc_end = t_assoc_only[len(t_assoc_only)-1]
        t_dissoc_start = t_assoc_end+1
        t_dissoc_end = t_dissoc_only[len(t_dissoc_only)-1]
        #print t_assoc_start, t_assoc_end, t_dissoc_start, t_dissoc_end
        def assocdissocEqGLLD(t, Req, ka, kd, Y_assoc, Y_dissoc):
                t_assoc_only = t[t_assoc_start:t_assoc_end+1]
                t_assoc = t_assoc_only-t_assoc_only[0]
                t_dissoc_only = t[t_dissoc_start:t_dissoc_end+1]
                t_dissoc = t_dissoc_only-t_dissoc_only[0]
                #r_assoc = (Req*(1-exp(-kobs*t_assoc)))+(Y_assoc*t_assoc)
                r_assoc = (Req*(1-exp(-
(ka*analyte_conc+kd)*t_assoc)))+(Y_assoc*t_assoc)
                R0 = r_assoc[len(r_assoc)-1]
                r_dissoc = (R0*exp(-kd*t_dissoc))+(Y_dissoc*t_dissoc)
                r_assoc_dissoc = np.concatenate([r_assoc, r_dissoc])
                return r_assoc_dissoc
```

```
        adfit_params, adcov = curve_fit(assocdissocEqGLLD,\
                                                t_assoc_dissoc,\
                                                ru_assoc_dissoc,\
                                                assoc_dissoc_init_params,\
                                                bounds=assoc_dissoc_param_bounds,\
                                                #gtol = 1e-20,\
                                                #xtol = 1e-20, ftol = 1e-20,\
                                                max_nfev=5000000)
        Req_pred, ka_pred, kd_pred, Y_assoc_pred, Y_dissoc_pred = adfit_params
        assoc_dissoc_values_pred = assocdissocEqGLLD(t_assoc_dissoc, Req_pred,
ka_pred, kd_pred, Y_assoc_pred, Y_dissoc_pred)
        #assoc_dissoc_values_pred = assoc_dissoc_values_pred -
assoc_dissoc_values_pred[0]
        # Reverse calculating from the equation => kobs = kd+(ka*analyte_conc); Req =
[A]*Rmax/([A]+KD)
        kobs_pred = (ka_pred*analyte_conc)+kd_pred
        kD_pred = kd_pred/ka_pred
#       Rmax_pred = (Req_pred/analyte_conc)*(analyte_conc+kD_pred)
        glld_coefs_df = DataFrame([{'ka'                    : ka_pred,\
                                    'kd'                    : kd_pred,\
                                    'kD'                    : kD_pred,\
                                    'kobs'                  : kobs_pred,\
```

```
                                        'Y_assoc_pred' : Y_assoc_pred,\
                                        'Y_dissoc_pred': Y_dissoc_pred,\
                                        'Rmax'                          : R0}])
        assoc_values_pred = assoc_dissoc_values_pred[t_assoc_start:t_assoc_end+1]
        dissoc_values_pred = assoc_dissoc_values_pred[t_dissoc_start:t_dissoc_end+1]
        #print len(assoc_values_pred), len(dissoc_values_pred), len(t_assoc_only),
len(t_dissoc_only), len(ru_assoc_only), len(ru_dissoc_only)
        return ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, glld_coefs_df,\
                                adfit_params
def assocEqMT(Rt, t, Rmax,\
                                                ka_analyte_conc,\
                                                kobs, kakt_ratio):
        dRdt = (((ka_analyte_conc*Rmax)-(kobs*Rt))/(1+(kakt_ratio*(Rmax-Rt))))
        return dRdt
def dissocEqMT(Rt, t, Rmax, kd, kakt_ratio):
        dRdt = ((-kd*Rt)/(1+(kakt_ratio*(Rmax-Rt))))
        return dRdt
def assocEqHL(t, C1, C2, kobs1, kobs2, X):
        r_assoc = C1*(1-exp(-kobs1*t))+C2*(1-exp(-kobs2*t))+X
        return r_assoc
```

```
def dissocEqHL(t, D1, D2, kd1, kd2, X):
        t0 = t[0]
        t = t-t0 # Zeroing dissoc so they go from 0 to 699 instead of 301-1000
        r_dissoc = (D1*exp(-kd1*t))+(D2*exp(-kd2*t))+X
        return r_dissoc
# Calculate Residuals + Residual sum of squares/deviance
# Calculate Residual degree of freedom + Residuals Stdev
# Residual variance => reduced chisquare
def calcResidualsSsqSD(input_response, pred_response, params_count):
        resids = input_response-pred_response
        input_response_var = var(input_response)
        resid_ssq = sum(resids**2)
        resid_dof = len(input_response)-params_count
        resid_var = resid_ssq/resid_dof
        resid_sd = sqrt(resid_var)
        chisq_value, chisq_p_value = calcChisq(input_response,\
                                               pred_response,\
                                               params_count)
        return resids, resid_ssq, resid_sd, chisq_value
# Calculate chisquare and p-values
# delta degrees of freedom (ddof): p-k-1
# p = #parameters, k = #data points
```

```
def calcChisq(input_response, pred_response, params_count):
        chisq_ddof = len(input_response)-params_count-1
        chisq_value, p_value = chisquare(f_obs=input_response,\
        f_exp=pred_response,\
        ddof=chisq_ddof)
        #print "Chi>", chisq_value, p_value
        return chisq_value, p_value
# Hard coding most of the params for now (other than Rmax)
def getAssocDissocParams(t, response_unit, \
                                                analyte_conc, t_assoc,\
                                               t_dissoc, t_assoc_start,\
                                               t_dissoc_end, fit_type):
        # set common assoc-dissoc params
        ka0, kd0, kt0 = 1e5, 1e-3, 1e8
        ka10, kd10, ka20, kd20 = 1e5, 1e-3, 1e-4, 0.1
        analyte_surf_conc0 = 0
        X0 = 0
        Y0 = 0 #0.01 suggested by biaeval documentation
        # t_assoc_end = t_assoc time (since   t starts from zero)
        t_assoc_end = t_assoc_start+t_assoc
        t_dissoc_start = t_assoc_end+1
        # Adding one to extract the values because python is zero indexing!
```

```
        t_assoc_only = t[t_assoc_start:t_assoc_end+1]
        ru_assoc_only = response_unit[t_assoc_start:t_assoc_end+1]
        t_dissoc_only = t[t_dissoc_start:t_dissoc_end+1]
        ru_dissoc_only = response_unit[t_dissoc_start:t_dissoc_end+1]
        # Common association phase calcs
        # Median of last 10 seconds of association phase:
        ru_assoc_end = len(ru_assoc_only)-1 #t_assoc_end+1?
        Rmax0 = float64(median(ru_assoc_only[(ru_assoc_end-10):ru_assoc_end]))
#float(max(response_unit))
        #Rmax0 = float64(ru_assoc_only[ru_assoc_end] - ru_assoc_only[0])
        #Rmax0 = float(max(response_unit))
        kobs0 = kd0+(ka0*analyte_conc)
        ### set assoc and dissoc params for each model
        # Langmuir 1:1
        if fit_type == "l":
                assoc_all_init = [t_assoc_only, ru_assoc_only,\
                                                        Rmax0, kobs0, X0]
                dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                                         Rmax0, kd0, X0]
        # Langmuir 1:1 with linear drift
        elif fit_type == "lld":
                # Works but might not use it!
```

```
        assoc_all_init = [t_assoc_only, ru_assoc_only,\
                                        Rmax0, kobs0, Y0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                        Rmax0, kd0, Y0, X0]
# Langmuir 1:1 with linear drift multimodel correcting for drift
elif fit_type == "mlld":
        # Work in progress - troubleshooting
        assoc_all_init = [t_assoc_only, ru_assoc_only,\
                                        Rmax0, kobs0, Y0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                        Rmax0, kd0, Y0, X0]
# Langmuir 1:1 with linear drift multimodel correcting for drift
elif fit_type == "mglld":
        # Work in progress - troubleshooting
        assoc_all_init = [t_assoc_only, ru_assoc_only,\
                                        Rmax0, ka0, Y0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                        Rmax0, kd0, Y0, X0]
# Langmuir 1:1 with mass transport
elif fit_type == "lmt":
        # Need to fix this! Might not need it!
        kakt_ratio0 = ka0/kt0
```

```
        ka_analyte_conc0 = ka0*analyte_conc#*rmax
        assoc_all_init = [t_assoc_only, ru_assoc_only,\
                                        Rmax0, ka_analyte_conc0, kd0,\
                                        kobs0, kakt_ratio0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                    Rmax0, kd0, kakt_ratio0, X0]
# Langmuir 1:1 with mass transport by surface conc. prediction:
elif fit_type == "lmts":
        assoc_all_init = [t_assoc_only, ru_assoc_only,\
        Rmax0, analyte_surf_conc0,\
        kt0, ka0, kd0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
        kd0, X0]
# Heterogeneous ligand
elif fit_type == "hl":
        # NOT validated yet and might not use it!
        # Setting all second params to 10% of first. Need to fix this?
        #==>TESTING!
        kobs10 = kd10+(ka10*analyte_conc)
        kobs20 = kd20+(ka20*analyte_conc)
        #<==TESTING!
        Rmax10 = Rmax0
```

```
            Rmax20 = Rmax10*(10/100)
            assoc_all_init = [t_assoc_only, ru_assoc_only,\
            Rmax10, kobs10, Rmax20, kobs20, X0]
            dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
             kd10, kd20, X0]
        return assoc_all_init, dissoc_noD_init
#########################################################################
########
############ Experimental hidden params START. Don't these models from CLI
### FUNCTIONAL MODELS:
## Langmuir 1:1 with R0 (R0) NOT constant constant during dissociation!
def solveAssocDissocEqsLang1R0AsParam(assoc_all_init, dissoc_noD_init,\
                                      analyte_conc, fit_type, plot_prefix):
        t_assoc_only, ru_assoc_only, Req0, kobs0, X0 = assoc_all_init
        afit_params, acov = curve_fit(assocEqLang1, t_assoc_only,\
                                      ru_assoc_only,\
                                      p0=[Req0, kobs0, X0],\
                                      maxfev=2000000)
        Req_pred, kobs_pred, X_pred = afit_params
        assoc_values_pred = assocEqLang1(t_assoc_only, Req_pred,\
                                       kobs_pred, X_pred)
        ## Add R0 for dissoc => last element from calc assoc values?
```

```
        t_dissoc_only, ru_dissoc_only, R0, kd0, X0 = dissoc_noD_init
        # Reset R0 since we are trying to use it as a parameter in this model and not
constant!
        #R0 = assoc_values_pred[len(assoc_values_pred)-1]
        dfit_params, dcov = curve_fit(dissocEqLang1, t_dissoc_only,\
                                                    ru_dissoc_only,\
                                                    p0=[R0, kd0, X0],\
                                                    maxfev=5000000)
        Rmax_pred, kd_pred, X_pred = dfit_params
        dissoc_values_pred = dissocEqLang1(t_dissoc_only, Rmax_pred, kd_pred,
X_pred)
        # Reverse calculating from the equation => kobs = kd+(ka*analyte_conc)
        ka_pred = (kobs_pred-kd_pred)/analyte_conc
        kD_pred = kd_pred/ka_pred
        l_coefs_df = DataFrame([{'ka': ka_pred,\
                                    'kd': kd_pred,\
                                     'kD': kD_pred, 'kobs' : kobs_pred}])
        return ru_assoc_only, assoc_values_pred,\
                                    ru_dissoc_only, dissoc_values_pred,\
                                    t_assoc_only, t_dissoc_only, l_coefs_df,\
                                    afit_params, dfit_params
## Langmuir 1:1 with linear drift NOT constant constant during dissociation!
```

```
def solveAssocDissocEqsLLDR0AsParam(assoc_init, dissoc_init,\
                                    analyte_conc, fit_type, plot_prefix):
    t_assoc_only, ru_assoc_only, Req0, kobs0, Y0, X0 = assoc_init
    afit_params, acov = curve_fit(assocEqLLD, t_assoc_only,\
                                    ru_assoc_only,\
                                    p0=[Req0, kobs0, Y0, X0],\
                                    maxfev=1000000)
    Req_pred, kobs_pred, Y_assoc_pred, X_assoc_pred = afit_params
    assoc_values_pred = assocEqLLD(t_assoc_only, Req_pred,\
                                    kobs_pred, Y_assoc_pred, X_assoc_pred)
    ## Add R0 params for dissoc => last element from calc assoc values
    t_dissoc_only, ru_dissoc_only, R0, kd0, Y0, X0 = dissoc_init
    #R0 = assoc_values_pred[len(assoc_values_pred)-1]
    dfit_params, dcov = curve_fit(dissocEqLLD, t_dissoc_only,\
                                    ru_dissoc_only,\
                                    p0=[R0, kd0, Y0, X0],\
                                    maxfev=1000000)
    Rmax_pred, kd_pred, Y_dissoc_pred, X_dissoc_pred = dfit_params
    dissoc_values_pred = dissocEqLLD(t_dissoc_only, Rmax_pred,\
                                    kd_pred, Y_dissoc_pred, X_dissoc_pred)
    # Reverse calculating from the equation => kobs = kd+(ka*analyte_conc)
    ka_pred = (kobs_pred-kd_pred)/analyte_conc
```

```
        kD_pred = kd_pred/ka_pred
        lld_coefs_df = DataFrame([{'ka': ka_pred,\
                                   'kd': kd_pred,\
                                   'kD' : kD_pred, 'kobs' : kobs_pred}])
        return ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, lld_coefs_df,\
                                afit_params, dfit_params
############ Experimental hidden params END ############
#########################################################################
########
## Langmuir 1:1 with R0 constant during dissociation!
def solveAssocDissocEqsLang1(assoc_init, dissoc_init,\
                                analyte_conc, fit_type, plot_prefix):
        t_assoc_only, ru_assoc_only, Req0, kobs0, X0 = assoc_init
        afit_params, acov = curve_fit(assocEqLang1, t_assoc_only,\
                                ru_assoc_only,\
                                p0=[Req0, kobs0, X0],\
                                maxfev=2000000)
        Req_pred, kobs_pred, X_pred = afit_params
        assoc_values_pred = assocEqLang1(t_assoc_only, Req_pred,\
                                kobs_pred, X_pred)
```

```
## Add R0 for dissoc => last element from calc assoc values?
t_dissoc_only, ru_dissoc_only, R0, kd0, X0 = dissoc_init
#R0 = ru_dissoc_only[0]
#R0 = assoc_values_pred[len(assoc_values_pred)-1]
def dissocEqLang1_R0constant(t, kd, X):
        t0 = t[0]
        r_dissoc = (R0*exp(-kd*(t-t0)))+X
        return r_dissoc
dfit_params, dcov = curve_fit(dissocEqLang1_R0constant,\
                        t_dissoc_only,\
                        ru_dissoc_only,\
                        p0=[kd0, X0],\
                        maxfev=5000000)
kd_pred, X_pred = dfit_params
dissoc_values_pred = dissocEqLang1_R0constant(t_dissoc_only,\
                        kd_pred, X_pred)
# Reverse calculating from the equation => kobs = kd+(ka*analyte_conc)
ka_pred = (kobs_pred-kd_pred)/analyte_conc
kD_pred = kd_pred/ka_pred
Rmax_pred = (Req_pred/analyte_conc)*(analyte_conc+kD_pred)
l_coefs_df = DataFrame([{'ka'                   : ka_pred,\
                        'kd'            : kd_pred,\
```

```
                                'kD'                : kD_pred,\
                                'Rmax'  : Req0, 'kobs' : kobs_pred}])#Rmax_pred}])
        return ru_assoc_only, assoc_values_pred,\
                                        ru_dissoc_only, dissoc_values_pred,\
                                        t_assoc_only, t_dissoc_only, l_coefs_df,\
                                        afit_params, dfit_params
## Langmuir 1:1 with linear drift
def solveAssocDissocEqsLLD(assoc_init, dissoc_init,\
                                        analyte_conc, fit_type, plot_prefix):
        t_assoc_only, ru_assoc_only, Req0, kobs0, Y0_assoc, X0_assoc = assoc_init
        t_dissoc_only, ru_dissoc_only, R0, kd0, Y0_dissoc, X0_dissoc = dissoc_init
        assoc_slope_limit = (R0*20)/(100*t_assoc_only[len(t_assoc_only)-1])
        assoc_param_bounds = ((0,0,-assoc_slope_limit,-
np.inf),(np.inf,np.inf,assoc_slope_limit,np.inf))
        assoc_init_params = [Req0, kobs0, Y0_assoc, X0_assoc]
        afit_params, acov = curve_fit(assocEqLLD,\
                                        t_assoc_only,\
                                        ru_assoc_only,\
                                        assoc_init_params,\
                                        bounds=assoc_param_bounds,\
                                        #xtol = 1e-20, ftol = 1e-20,\
                                        max_nfev=1000000)
```

```
Req_pred, kobs_pred, Y_assoc_pred, X_assoc_pred = afit_params
assoc_values_pred = assocEqLLD(t_assoc_only, Req_pred,\
                                        kobs_pred, Y_assoc_pred,\
                                        X_assoc_pred)
## Add R0 params for dissoc => last element from calc assoc values
dissoc_slope_limit = (R0*10)/(100*t_dissoc_only[0])
dissoc_param_bounds = ((0,-np.inf,-np.inf),(np.inf,np.inf,np.inf))
dissoc_init_params = [kd0, Y0_dissoc, X0_dissoc]
def dissocEqLLD_R0constant(t, kd, Y, X):
        t0 = t[0]
        t_dissoc = t-t0
        r_dissoc = (R0*exp(-kd*t_dissoc))+(Y*t_dissoc)+X
        return r_dissoc
dfit_params, dcov = curve_fit(dissocEqLLD_R0constant,\
                        t_dissoc_only,\
                        ru_dissoc_only,\
                        dissoc_init_params,\
                        bounds=dissoc_param_bounds,\
                        #gtol = 1e-20,\
                        #xtol = 1e-20, ftol = 1e-20,\
                        max_nfev=5000000)
kd_pred, Y_dissoc_pred, X_dissoc_pred = dfit_params
```

```
        dissoc_values_pred = dissocEqLLD_R0constant(t_dissoc_only,\
                                                    kd_pred, Y_dissoc_pred,\
                                                    X_dissoc_pred)
        # Reverse calculating from the equation => kobs = kd+(ka*analyte_conc); Req =
[A]*Rmax/([A]+KD)
        ka_pred = (kobs_pred-kd_pred)/analyte_conc
        kD_pred = kd_pred/ka_pred
#       Rmax_pred = (Req_pred/analyte_conc)*(analyte_conc+kD_pred)
        lld_coefs_df = DataFrame([{'ka'                                    : ka_pred,\
                                   'kd'                                    : kd_pred,\
                                   'kD'                                    : kD_pred,\
                                   'kobs'                          : kobs_pred,\
                                   'Y_assoc_pred' : Y_assoc_pred,\
                                   'Y_dissoc_pred': Y_dissoc_pred,\
                                   'Rmax'                          : R0}])#Rmax_pred}])
        return ru_assoc_only, assoc_values_pred,\
                                    ru_dissoc_only, dissoc_values_pred,\
                                    t_assoc_only, t_dissoc_only, lld_coefs_df,\
                                    afit_params, dfit_params
## Langmuir 1:1 with linear drift multimodel correcting for drift
def solveAssocDissocEqsMLLD(assoc_init, dissoc_init,\
                                    t_preassoc, ru_preassoc,\
```

```
                                    analyte_conc, fit_type, plot_prefix):
        lfit = polyfit(t_preassoc,\
                                                        ru_preassoc,\
                                                        1)
        preassoc_drift = polyval(lfit, t_preassoc)
        ru_drift_corr_preassoc_only = (ru_preassoc - preassoc_drift)# + preassoc_drift[0]
        ru_assoc_only, assoc_values_pred,\
        ru_dissoc_only, dissoc_values_pred,\
        t_assoc_only, t_dissoc_only, lld_coefs_df,\
        lld_afit_params, lld_dfit_params = solveAssocDissocEqsLLD(assoc_init,\
                                        dissoc_init,\
                                        analyte_conc,\
                                        fit_type,\
                                        plot_prefix)
        Req_pred, kobs_pred, Y_assoc_pred, X_assoc_pred = lld_afit_params
        kd_pred, Y_dissoc_pred, X_dissoc_pred = lld_dfit_params
        #Rmax_pred = (Req_pred*(analyte_conc+kD_pred))/analyte_conc
        assoc_drift = Y_assoc_pred*t_assoc_only
        dissoc_drift = Y_dissoc_pred*t_dissoc_only
        # Correct association + dissociation drifts by subtracting predicted drifts from the
referenced data
        ru_drift_corr_assoc_only = (ru_assoc_only - assoc_drift) - X_assoc_pred
```

```
        ru_drift_corr_assoc_only = (ru_drift_corr_assoc_only -
ru_drift_corr_assoc_only[0]) +
ru_drift_corr_preassoc_only[len(ru_drift_corr_preassoc_only)-1]
        #ru_drift_corr_assoc_only = ru_assoc_only-ru_assoc_only[0]-assoc_drift -
X_assoc_pred+ru_drift_corr_preassoc_only[len(ru_drift_corr_preassoc_only)-1]
        ru_drift_corr_dissoc_only = (ru_dissoc_only - dissoc_drift) - X_dissoc_pred
        ru_drift_corr_dissoc_only = (ru_drift_corr_dissoc_only -
ru_drift_corr_dissoc_only[0]) + ru_drift_corr_assoc_only[len(ru_drift_corr_assoc_only)-
1]
        #ru_drift_corr_dissoc_only=ru_dissoc_only-ru_dissoc_only[0]-
dissoc_drift+ru_drift_corr_assoc_only[len(ru_drift_corr_assoc_only)-1]
        #print ru_drift_corr_assoc_only[0], ru_drift_corr_dissoc_only[0]
        t_assoc_only, ru_assoc_only, Req0, kobs0, Y0, X0 = assoc_init
        t_dissoc_only, ru_dissoc_only, R0, kd0, Y0, X0 = dissoc_init
        #print lld_afit_params, lld_dfit_params
        ru_drift_corr_assoc_end = len(ru_drift_corr_assoc_only)-1
        Rmax_drift_corr =
float64(median(ru_drift_corr_assoc_only[(ru_drift_corr_assoc_end-
10):ru_drift_corr_assoc_end]))
        mlld_assoc_init = [t_assoc_only, ru_drift_corr_assoc_only, Req_pred, kobs_pred,
Y_assoc_pred, X0]
```

```
        mlld_dissoc_init = [t_dissoc_only, ru_drift_corr_dissoc_only, R0, kd_pred,
Y_dissoc_pred, X0]
        #print type(t_assoc_only), type(ru_drift_corr_assoc_only), type(t_dissoc_only),
type(ru_drift_corr_dissoc_only)
        ru_mlld_assoc_only, mlld_assoc_values_pred,\
        ru_mlld_dissoc_only, mlld_dissoc_values_pred,\
        t_assoc_only, t_dissoc_only, mlld_coefs_df,\
        mlld_afit_params, mlld_dfit_params =
solveAssocDissocEqsLLD(mlld_assoc_init,\
                                                mlld_dissoc_init,\
                                                analyte_conc,\
                                                fit_type, plot_prefix)
        #print mlld_afit_params, mlld_dfit_params
        return ru_drift_corr_assoc_only, mlld_assoc_values_pred,\
                                ru_drift_corr_dissoc_only, mlld_dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, mlld_coefs_df,\
                                mlld_afit_params, mlld_dfit_params,
ru_drift_corr_preassoc_only
## Langmuir 1:1 with linear drift GLOBAL multimodel correcting for drift
def solveAssocDissocEqsMGLLD(assoc_init, dissoc_init,\
                                        t_preassoc, ru_preassoc,\
                                        analyte_conc, fit_type, plot_prefix):
```

```
lfit = polyfit(t_preassoc,\
                                                ru_preassoc,\
                                                1)
preassoc_drift = polyval(lfit, t_preassoc)
ru_drift_corr_preassoc_only = (ru_preassoc - preassoc_drift)# + preassoc_drift[0]
t_assoc_only, ru_assoc_only, Req0, ka0, Y0_assoc, X0_assoc = assoc_init
t_dissoc_only, ru_dissoc_only, R0, kd0, Y0_dissoc, X0_dissoc = dissoc_init
kobs0 = (ka0*analyte_conc)+kd0
assoc_init = [t_assoc_only, ru_assoc_only, Req0, kobs0, Y0_assoc, X0_assoc]
t_assoc_start = t_assoc_only[0]
t_assoc_end = t_assoc_only[len(t_assoc_only)-1]
t_dissoc_start = t_assoc_end+1
t_dissoc_end = t_dissoc_only[len(t_dissoc_only)-1]
t_assoc_dissoc = np.concatenate([t_assoc_only, t_dissoc_only])
ru_assoc_dissoc = np.concatenate([ru_assoc_only, ru_dissoc_only])
def assocdissocEqLLD_global(t, Req, ka, kd, Y_assoc, Y_dissoc):
        t_assoc_only = t[t_assoc_start:t_assoc_end+1]
        t_assoc = t_assoc_only-t_assoc_only[0]
        t_dissoc_only = t[t_dissoc_start:t_dissoc_end+1]
        t_dissoc = t_dissoc_only-t_dissoc_only[0]
        r_assoc = (Req*(1-exp(-(ka*analyte_conc+kd)*t_assoc))) +
(Y_assoc*t_assoc)
```

```
                R0 = r_assoc[len(r_assoc)-1]
                r_dissoc = (R0*exp(-kd*t_dissoc)) + (Y_dissoc*t_dissoc)
                r_assoc_dissoc = np.concatenate([r_assoc, r_dissoc])
                return r_assoc_dissoc
        assoc_slope_limit = (R0*20)/(100*t_assoc_only[len(t_assoc_only)-1])
        #assoc_param_bounds = ((0,0,-assoc_slope_limit,-
np.inf),(np.inf,np.inf,assoc_slope_limit,np.inf))
        lld_global_assoc_dissoc_init_params = [Req0, ka0, kd0, Y0_assoc, Y0_dissoc]
        lld_global_assoc_dissoc_param_bounds = ((0, 0, 0, -assoc_slope_limit, -
np.inf),(np.inf, np.inf, np.inf, assoc_slope_limit,np.inf))
        #print max(ru_assoc_dissoc), max(ru_drift_corr_assoc_only)
        lld_global_adfit_params, lld_global_adcov =
curve_fit(assocdissocEqLLD_global,\
                                        t_assoc_dissoc,\
                                        ru_assoc_dissoc,\
                                        lld_global_assoc_dissoc_init_params,\
                                        bounds=lld_global_assoc_dissoc_param_bounds,\
                                        #gtol = 1e-20,\
                                        #xtol = 1e-20, ftol = 1e-20,\
                                        max_nfev=5000000)
        Req_pred, ka_pred, kd_pred, Y_assoc_pred, Y_dissoc_pred =
lld_global_adfit_params
```

```
        #print lld_global_adfit_params
        #ru_assoc_only, assoc_values_pred,\
        #ru_dissoc_only, dissoc_values_pred, t_assoc_only, t_dissoc_only,
glld_coefs_df,\
        #glld_adfit_params = solveAssocDissocEqsGLLD(assoc_init,\
        #
                                        dissoc_init,\
        #
                                        analyte_conc,\
        #
                                        fit_type,\
        #
                                        plot_prefix)
        #Req_pred, ka_pred, kd_pred, Y_assoc_pred, Y_dissoc_pred = glld_adfit_params
        #ru_assoc_only, assoc_values_pred,\
        #ru_dissoc_only, dissoc_values_pred,\
        #t_assoc_only, t_dissoc_only, lld_coefs_df,\
        #lld_afit_params, lld_dfit_params = solveAssocDissocEqsLLD(assoc_init,\
        #
                                        dissoc_init,\
        #
                                        analyte_conc,\
```

```
        #
                                        fit_type,\
        #
                                        plot_prefix)
        #Req_pred, kobs_pred, Y_assoc_pred, X_assoc_pred = lld_afit_params
        #kd_pred, Y_dissoc_pred, X_dissoc_pred = lld_dfit_params
        assoc_drift = Y_assoc_pred*t_assoc_only
        dissoc_drift = Y_dissoc_pred*t_dissoc_only
        ru_drift_corr_assoc_only = (ru_assoc_only - assoc_drift)
        ru_drift_corr_assoc_only = (ru_drift_corr_assoc_only +
ru_drift_corr_assoc_only[0]) #+
ru_drift_corr_preassoc_only[len(ru_drift_corr_preassoc_only)-1]
        #print ru_assoc_only
        #print ru_drift_corr_assoc_only
        ru_drift_corr_dissoc_only = (ru_dissoc_only - dissoc_drift)
        ru_drift_corr_dissoc_only = (ru_drift_corr_dissoc_only -
ru_drift_corr_dissoc_only[0]) + ru_drift_corr_assoc_only[len(ru_drift_corr_assoc_only)-
1]
        ##print ru_dissoc_only
        #print ru_drift_corr_dissoc_only
        #ru_drift_corr_assoc_only = ru_drift_corr_assoc_only -
ru_drift_corr_assoc_only[0]
```

```
        #print ru_drift_corr_assoc_only[0:10], ru_drift_corr_dissoc_only[0:10]
        #ru_drift_corr_dissoc_only = ru_drift_corr_dissoc_only -
ru_drift_corr_assoc_only[0]
        ru_drift_corr_assoc_end = len(ru_drift_corr_assoc_only)-1
        Rmax_drift_corr =
float64(median(ru_drift_corr_assoc_only[(ru_drift_corr_assoc_end-
10):ru_drift_corr_assoc_end]))
        #kobs_pred = (ka_pred*analyte_conc)+kd_pred
        #ka_pred = (kobs_pred-kd_pred)/analyte_conc
        #print ka_pred
        #print glld_adfit_params
        #print lld_afit_params, lld_dfit_params
        def dissocEqLang1_R0constant(t, kd):
                t0 = t[0]
                r_dissoc = (Rmax_drift_corr*exp(-kd*(t-t0)))
                return r_dissoc
        l1_dissoc_param_bounds = ((0),(np.inf))
        l1_dfit_params, l1_dcov = curve_fit(dissocEqLang1_R0constant,\
                                        t_dissoc_only,\
                                        ru_drift_corr_dissoc_only,\
                                        bounds=l1_dissoc_param_bounds,\
                                        p0=[kd0],\
```

```
                                        max_nfev=5000000)
        kd_l1_pred = l1_dfit_params
        #print l1_dfit_params
        kobs_l1_pred = (ka_pred*analyte_conc)+kd_l1_pred
        def assocdissocEqL_kdfixed(t, Req, ka):
                #t_assoc_only = t[t_assoc_start:t_assoc_end+1]
                #t_assoc = t_assoc_only-t_assoc_only[0]
                #t_dissoc_only = t[t_dissoc_start:t_dissoc_end+1]
                #t_dissoc = t_dissoc_only-t_dissoc_only[0]
                r_assoc = (Req*(1-exp(-(ka*analyte_conc+kd_l1_pred)*t)))
                #r_assoc = (Req*(1-exp(-
(ka*analyte_conc+kd_l1_pred)*t_assoc)))+X_assoc
                #r_dissoc = (Req*exp(-kd_l1_pred*t_dissoc))+X_dissoc
                #r_assoc_dissoc = np.concatenate([r_assoc, r_dissoc])
                #return r_assoc_dissoc
                return r_assoc
        lkdf_assoc_dissoc_param_bounds = ((0, 0),(np.inf, np.inf))
        lkdf_assoc_dissoc_init_params = [Rmax_drift_corr, ka0]
        lkdf_adfit_params, lkdf_adcov = curve_fit(assocdissocEqL_kdfixed,\
                                t_assoc_only,\
                                ru_drift_corr_assoc_only,\
                                lkdf_assoc_dissoc_init_params,\
```

```
                                bounds=lkdf_assoc_dissoc_param_bounds,\
                                #gtol = 1e-20,\
                                #xtol = 1e-20, ftol = 1e-20,\
                                max_nfev=5000000)
        Req_lkdf_pred, ka_lkdf_pred = lkdf_adfit_params
        #print lkdf_adfit_params
        def assocdissocEqL_kdRelaxed(t, Req, ka, kd):
                t_assoc_only = t[t_assoc_start:t_assoc_end+1]
                t_assoc = t_assoc_only-t_assoc_only[0]
                t_dissoc_only = t[t_dissoc_start:t_dissoc_end+1]
                t_dissoc = t_dissoc_only-t_dissoc_only[0]
                r_assoc = (Req*(1-exp(-(ka*analyte_conc+kd)*t_assoc)))
                R0 = r_assoc[len(r_assoc)-1]
                r_dissoc = (R0*exp(-kd*t_dissoc))
                r_assoc_dissoc = np.concatenate([r_assoc, r_dissoc])
                return r_assoc_dissoc
        lkdr_assoc_dissoc_init_params = [Rmax_drift_corr, ka_lkdf_pred, kd_l1_pred]
        lkdr_assoc_dissoc_param_bounds = ((0, 0, 0),(np.inf, np.inf, np.inf))
        #print max(ru_assoc_dissoc), max(ru_drift_corr_assoc_only)
        ru_drift_corr_assoc_dissoc = np.concatenate([ru_drift_corr_assoc_only,
ru_drift_corr_dissoc_only])
```

```
        lkdr_adfit_params, lkdr_adcov = curve_fit(assocdissocEqL_kdRelaxed,\
                                        t_assoc_dissoc,\
                                        ru_drift_corr_assoc_dissoc,\
                                        lkdr_assoc_dissoc_init_params,\
                                        bounds=lkdr_assoc_dissoc_param_bounds,\
                                        #gtol = 1e-20,\
                                        #xtol = 1e-20, ftol = 1e-20,\
                                        max_nfev=5000000)
        Req_lkdr_pred, ka_lkdr_pred, kd_lkdr_pred = lkdr_adfit_params
        #print lkdr_adfit_params
        mlld_assoc_dissoc_values_pred = assocdissocEqL_kdRelaxed(t_assoc_dissoc,
Req_lkdr_pred, ka_lkdr_pred, kd_lkdr_pred)
        # Temp for output purposes:
        mlld_adfit_params = lkdr_adfit_params
        t_assoc_only = t_assoc_dissoc[t_assoc_start:t_assoc_end+1]
        mlld_assoc_values_pred =
mlld_assoc_dissoc_values_pred[t_assoc_start:t_assoc_end+1]
        ru_mlld_assoc_only = ru_drift_corr_assoc_dissoc[t_assoc_start:t_assoc_end+1]
        t_dissoc_only = t_assoc_dissoc[t_dissoc_start:t_dissoc_end+1]
        mlld_dissoc_values_pred =
mlld_assoc_dissoc_values_pred[t_dissoc_start:t_dissoc_end+1]
```

```
        ru_mlld_dissoc_only =
ru_drift_corr_assoc_dissoc[t_dissoc_start:t_dissoc_end+1]
        kD_lkdr_pred = kd_lkdr_pred/ka_lkdr_pred
        kobs_lkdr_pred = (ka_lkdr_pred*analyte_conc)+kd_lkdr_pred
        mlld_coefs_df = DataFrame([{'ka'                                    :
ka_lkdr_pred,\
                'kd'                                    : kd_lkdr_pred,\
                'kD'                                    : kD_lkdr_pred,\
                'kobs'                          : kobs_lkdr_pred,\
                'Y_assoc_pred' : Y_assoc_pred,\
                'Y_dissoc_pred': Y_dissoc_pred,\
                'Rmax'                          : Req_lkdr_pred}])
        #mlld_assoc_init = [t_assoc_only, ru_drift_corr_assoc_only, Rmax_drift_corr,
ka0, 0, 0]
        #mlld_dissoc_init = [t_dissoc_only, ru_drift_corr_dissoc_only, Rmax_drift_corr,
kd0, 0, 0]
        #ru_mlld_assoc_only, mlld_assoc_values_pred,\
        #ru_mlld_dissoc_only, mlld_dissoc_values_pred, t_assoc_only, t_dissoc_only,
mlld_coefs_df,\
        #mlld_adfit_params    = solveAssocDissocEqsGLLD(mlld_assoc_init,\
        #
                                                mlld_dissoc_init,\
```

```
        #
                                                analyte_conc,\
        #
                                                fit_type, plot_prefix)
        #print mlld_adfit_params
        return ru_mlld_assoc_only, mlld_assoc_values_pred,\
                                ru_mlld_dissoc_only, mlld_dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, mlld_coefs_df,\
                                mlld_adfit_params, mlld_adfit_params,
ru_drift_corr_preassoc_only
## Langmuir 1:1 with mass transfer using
## analyte surface concentration
def solveAssocDissocEqsAsurfLMT(assoc_all_init,\
                                dissoc_noD_init,\
                                analyte_conc,\
                                fit_type,\
                                plot_prefix):
        #print assoc_all_init
        ## Asurf calculation
        t_assoc_only, ru_assoc_only,\
         Rmax0, analyte_surf_conc0,\
         kt0, kd0, ka0, X0 = assoc_all_init
```

```
        t_dissoc_only, ru_dissoc_only,\
        kd0, X0 = dissoc_noD_init
        def calcAnalyteSurfConcMT(R, t, Rmax,
                                  analyte_surf_conc,\
                                  kt, X):
dAsdt = (kt*(analyte_conc-analyte_surf_conc))-X
            #dAsdt = (kt_analyte_conc-kt_analyte_surf_conc))-
(ka*analyte_surf_conc*(Rmax-R))+(kd*R)
            return dAsdt
        def solveAnalyteSurfConcMT(t, Rmax, analyte_surf_conc,\
                                                    kt, ka, kd):
            X = (ka*analyte_surf_conc*(Rmax-
ru_assoc_only[0]))+(kd*ru_assoc_only[0])
            analyte_surf_conc_pred = odeint(calcAnalyteSurfConcMT,\
                                  ru_assoc_only[0], t,\
                                  args=(Rmax,\
                                  analyte_surf_conc,\
                                  kt, X))
            return analyte_surf_conc_pred[:,0].ravel()
        asurf_params, asurf_cov = curve_fit(solveAnalyteSurfConcMT,\
                                  t_assoc_only, \
                                  ru_assoc_only,\
```

```
                                        p0=[Rmax0,\
                                        analyte_surf_conc0,\
                                        kt0, ka0,\
                                        kd0])
        #print "As---->", asurf_params
        Rmax_pred, analyte_surf_conc_pred,\
        kt_pred, ka_pred, kd_pred = asurf_params
        asurf_pred = solveAnalyteSurfConcMT(t_assoc_only,\
                                        analyte_surf_conc_pred,\
                                        kt_pred, ka_pred,\
                                        kd_pred, Rmax_pred)
        #print "Asurf Pred=>", asurf_pred[len(asurf_pred)-1]
        kobs_pred = (ka_pred*analyte_surf_conc_pred)+kd_pred
        kobs0= (ka0*analyte_surf_conc_pred)+kd0
        assoc_init = [t_assoc_only, ru_assoc_only,\
                                                Rmax0, kobs0, X0]
        dissoc_noD_init = [t_dissoc_only, ru_dissoc_only,\
                                                kd_pred, Rmax0, X0]
        ru_assoc_only, assoc_values_pred,\
        ru_dissoc_only, dissoc_values_pred,\
        t_assoc_only, t_dissoc_only, l_coefs_df =
solveAssocDissocEqsLang1(assoc_init,\
```

```
                                                    dissoc_noD_init,\
                                                     analyte_surf_conc_pred,\
                                                    fit_type,\
                                                    plot_prefix)
        lmts_coef_df_init =
DataFrame([{'analyte_surf_conc_pred':analyte_surf_conc_pred,\
                                                    'kt':kt_pred}])
        lmts_coef_df = concat([lmts_coef_df_init, l_coefs_df], axis=1)
        return ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, lmts_coef_df,\
                                afit_params, dfit_params
## Langmuir 1:1 with mass transfer
def solveAssocDissocEqsLMT(assoc_all_init, dissoc_noD_init,\
                                    analyte_conc, fit_type,\
                                    plot_prefix):
        ## Assoc phase-->
        t_assoc_only, ru_assoc_only,\
         Rmax_assoc0, ka_analyte_conc0, kd0,\
         kobs0, kakt_ratio_assoc0, X0 = assoc_all_init
        #print "Init Assoc=> ", Rmax, ka_analyte_conc_rmax0, kd0,\
```

```
        #
kobs0, kakt_ratio_assoc0, X0
        # Solve ODE for association phase:
        def solveAssocEqsLMT(t, Rmax, ka_analyte_conc,\
                                                                    kobs,
kakt_ratio_assoc):
                Rt_pred = odeint(assocEqMT, ru_assoc_only[0], t,\
                                                                args=(Rmax,
ka_analyte_conc,\
                kobs, kakt_ratio_assoc))
                return Rt_pred[:,0].ravel()
        afit_params, acov = curve_fit(solveAssocEqsLMT,\
                                t_assoc_only,\
                                ru_assoc_only,\
                                p0=[Rmax_assoc0,\
                                ka_analyte_conc0,\
                                kobs0, kakt_ratio_assoc0],\
                                maxfev=1000000)
        print "LMT ASSOC PARAMS=>",afit_params
        Rmax_assoc_pred,\
        ka_analyte_conc_pred,\
        kobs_pred, kakt_ratio_assoc_pred = afit_params
```

```
assoc_values_pred = solveAssocEqsLMT(t_assoc_only,\
                                     Rmax_assoc_pred,\
                                     ka_analyte_conc_pred,\
                                     kobs_pred,\
                                     kakt_ratio_assoc_pred)
## Dissoc phase-->
t_dissoc_only, ru_dissoc_only,\
 Rmax_dissoc0, kd0, kakt_ratio_dissoc0, X0 = dissoc_noD_init
#print "Init Dissoc=>", Rmax, kd0, kakt_ratio_dissoc0, X0
# Solve ODE for dissociation phase:
def solveDissocEqsLMT(t, Rmax, kd, kakt_ratio_dissoc):
        Rt_pred = odeint(dissocEqMT,\
                         ru_dissoc_only[0],\
                         t,\
                         args=(Rmax, kd, kakt_ratio_dissoc))
        return Rt_pred[:,0].ravel()
dfit_params, dcov = curve_fit(solveDissocEqsLMT,\
                         t_dissoc_only,\
                         ru_dissoc_only,\
                         p0=[Rmax_dissoc0, kd0,\
                         kakt_ratio_dissoc0],\
                         maxfev=1000000)
```

```
print "LMT DISSOC PARAMS=>", dfit_params
Rmax_dissoc_pred, kd_pred,\
kakt_ratio_dissoc_pred = dfit_params
dissoc_values_pred = solveDissocEqsLMT(t_dissoc_only,\
                                       Rmax_dissoc_pred,\
                                       kd_pred,\
                                       kakt_ratio_dissoc_pred)
# Return the required coefficients
ka_pred = (kobs_pred-kd_pred)/analyte_conc
kt_assoc = ka_pred/kakt_ratio_assoc_pred
ka_from_rmax_conc_pred = ka_analyte_conc_pred/(analyte_conc)
kt_from_rmax_conc_pred = ka_from_rmax_conc_pred/kakt_ratio_assoc_pred
kt_dissoc = ka_pred/kakt_ratio_dissoc_pred
kD_pred = kd_pred/ka_pred
lmt_coefs_df = DataFrame([{'ka': ka_pred,\
                    #'ka_from_rmax_conc_pred' : ka_from_rmax_conc_pred,\
                    #'kt_from_rmax_conc_pred' : kt_from_rmax_conc_pred,\
                    #'rmax_assoc_pred' : Rmax_assoc_pred,\
                    #'rmax_dissoc_pred' : Rmax_dissoc_pred,\
                    'kd': kd_pred,\
                    'kD': kD_pred,\
                    'kt_dissoc': kt_dissoc,\
```

```
                    'kt_assoc': kt_assoc}])
    return ru_assoc_only, assoc_values_pred,\
                                    ru_dissoc_only, dissoc_values_pred,\
                                    t_assoc_only, t_dissoc_only, lmt_coefs_df,\
                                    afit_params, dfit_params
## Heterogeneous Ligand
def solveAssocDissocEqsHL(assoc_all_init,\
    dissoc_noD_init,\
    analyte_conc,\
    fit_type, plot_prefix):
    t_assoc_only, ru_assoc_only,\
    C10, kobs10, C20, kobs20, X0 = assoc_all_init
    afit_params, acov = curve_fit(assocEqHL,\
                            t_assoc_only,\
                            ru_assoc_only,\
                            p0=[C10, kobs10, C20, kobs20, X0],\
                            maxfev=1000000)
    C1_pred, kobs1_pred, C2_pred, kobs2_pred, X_pred = afit_params
    assoc_values_pred = assocEqHL(t_assoc_only,\
                                    C1_pred, kobs1_pred,\
                                    C2_pred, kobs2_pred,\
                                    X_pred)
```

```
## Add R0 for dissoc => last element from calc assoc values
t_dissoc_only, ru_dissoc_only, kd10, kd20, X0 = dissoc_noD_init
D10 = assoc_values_pred[len(assoc_values_pred)-1]
D20 = D10*(30/100)
dfit_params, dcov = curve_fit(dissocEqHL, t_dissoc_only,\
                              ru_dissoc_only,\
                              p0=[D10, kd10, D20, kd20, X0],\
                              maxfev=1000000)
D1_pred, kd1_pred, D2_pred, kd2_pred, X_pred = dfit_params
dissoc_values_pred = dissocEqHL(t_dissoc_only,\
                                        D1_pred, kd1_pred,\
                                        D2_pred, kd2_pred,\
                                        X_pred)
# Reverse calculating from the equation => kobs = kd+(ka*analyte_conc)
ka1_pred = (kobs1_pred-kd1_pred)/analyte_conc
kD1_pred = kd1_pred/ka1_pred
ka2_pred = (kobs2_pred-kd2_pred)/analyte_conc
kD2_pred = kd2_pred/ka2_pred
hl_coefs_df = DataFrame([{'ka1': ka1_pred, 'kd1': kd1_pred,\
        'kD1' : kD1_pred,\
        'ka2': ka2_pred, 'kd2': kd2_pred,\
        'kD2' : kD2_pred}])
```

```
    return ru_assoc_only, assoc_values_pred,\
                                        ru_dissoc_only, dissoc_values_pred,\
                                        t_assoc_only, t_dissoc_only, hl_coefs_df,\
                                        afit_params, dfit_params
def combineAndPlot(response_unit_pread, t_pread,\
    t_assoc_dissoc,\
    assoc_dissoc_values_pred,\
    assoc_dissoc_values_resids,\
    raw_time,\
    raw_response,\
    prot_block_name,
    plot_file_path):
    plot_filenamepath = join(plot_file_path, prot_block_name + '.pdf')
    plt.plot(raw_time, raw_response, 'y-',label='raw_dc')
    plt.plot(t_pread, response_unit_pread, 'k-',\
                                        label='data', linewidth=0.5)
    plt.plot(t_assoc_dissoc, assoc_dissoc_values_pred, 'r-',\
                                        label='fitted')
    plt.plot(t_assoc_dissoc, assoc_dissoc_values_resids, 'b-',\
                                        label='residuals', linewidth=0.5)
    plt.title(prot_block_name)
    plt.legend(loc='best')
```

```
        plt.savefig(plot_filenamepath)
        plt.close()
        return plot_filenamepath
def fitCurveModels(df_time_refdall,\
                                                df_time_raw,\
                                                analyte_conc,\
                                                analyte_mass,\
                                                flow_rate,\
                                                t_assoc, t_dissoc,\
                                                assoc_tstart,
dissoc_tend,\
                                                prot_block_name,
fit_type,\
                                                base_input_filename,
                                                plot_file_path,\
                                                refn_spots_name_list,\
                                                 merger, time_colname):
        # Reset the association start value to zero since we zeroed the data to start from
zero
        # Zero Association start time since the dataframe is already zeroed to association
start time
        assoc_tstart_new = assoc_tstart-assoc_tstart
```

```
        dissoc_tend_new = dissoc_tend-assoc_tstart
        raw_response = np.array(df_time_raw.iloc[:,1])
        raw_time = np.array(df_time_raw.iloc[:,0])
        df_time_refdall_preassoc = df_time_refdall.loc[(df_time_refdall[time_colname] <
assoc_tstart_new)]
        df_time_refdall_ad = df_time_refdall.loc[(df_time_refdall[time_colname] >=
assoc_tstart_new) & (df_time_refdall[time_colname] <= dissoc_tend_new)]
        refs_count = len(refn_spots_name_list)
        t_pread = np.array(df_time_refdall[time_colname])
        t_preassoc = np.array(df_time_refdall_preassoc[time_colname])
        t = np.array(df_time_refdall_ad[time_colname])
        one_spot_all_coefs_df = DataFrame()
        coefs_df = DataFrame()
        all_plot_filenames_by_fittype = list()
        term_refd_response_unit_pread = ""
        for icol in range(0, refs_count):
                response_icol = icol+
                ref_spot_name = refn_spots_name_list[icol]
                print "Reference#", response_icol, "=>", ref_spot_name
                plot_prefix = str(ref_spot_name) + '_' + fit_type
                response_unit_preassoc =
np.array(df_time_refdall_preassoc.iloc[:,response_icol])
```

```
response_unit_pread = np.array(df_time_refdall.iloc[:,response_icol])
response_unit = np.array(df_time_refdall_ad.iloc[:,response_icol])
### get initial assoc + dissoc specific params
assoc_all_init,\
dissoc_noD_init = getAssocDissocParams(t,\
                                       response_unit,\
                                       analyte_conc,\
                                       t_assoc, t_dissoc,\
                                       assoc_tstart_new, dissoc_tend_new,\
                                       fit_type)
spot_info_df = DataFrame([{'array_protein_spot' : prot_block_name,\
                           'reference_no' : response_icol,\
                           'reference_spot' : ref_spot_name,\
                           'input_filename' : base_input_filename,\
                           'flow_rate' : flow_rate,\
                           'analyte_conc': analyte_conc,\
                           'analyte_mass': analyte_mass,\
                           'assoc_tstart': assoc_tstart,\
                           'assoc_time' : t_assoc,\
                           'dissoc_time': t_dissoc,\
                           'type_of_fit' : fit_type}])
try:
```

```
                    if fit_type == "l":
                     #print "Fitting langmuir 1:1 ...."
                              ru_assoc_only, assoc_values_pred,\
                             ru_dissoc_only, dissoc_values_pred,\
                             t_assoc_only, t_dissoc_only, coefs_df,\
                             afit_params, dfit_params =
solveAssocDissocEqsLang1(assoc_all_init,\
                             dissoc_noD_init,\
                             analyte_conc,\
                             fit_type, plot_prefix)
                    elif fit_type == "lld":
                     #print "Fitting langmuir 1:1 with drift...."
                              ru_assoc_only, assoc_values_pred,\
                             ru_dissoc_only, dissoc_values_pred,\
                             t_assoc_only, t_dissoc_only, coefs_df,\
                             afit_params, dfit_params =
solveAssocDissocEqsLLD(assoc_all_init,\
                             dissoc_noD_init,\
                             analyte_conc,\
                             fit_type, plot_prefix)
                    elif fit_type == "mlld":
                     #print "Fitting langmuir 1:1 with drift...."
```

```
                                ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, coefs_df,\
                                afit_params, dfit_params, ru_drift_corr_preassoc_only =
solveAssocDissocEqsMLLD(assoc_all_init,\
                                dissoc_noD_init,\
                                t_preassoc, response_unit_preassoc,\
                                analyte_conc,\
                                fit_type, plot_prefix)
                                response_unit_pread =
np.concatenate([ru_drift_corr_preassoc_only, ru_assoc_only, ru_dissoc_only])
                        elif fit_type == "mglld":
                         #print "Fitting langmuir 1:1 with drift...."
                                ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, coefs_df,\
                                afit_params, dfit_params, ru_drift_corr_preassoc_only =
solveAssocDissocEqsMGLLD(assoc_all_init,\
                                dissoc_noD_init,\
                                t_preassoc, response_unit_preassoc,\
                                analyte_conc,\
                                fit_type, plot_prefix)
```

```
                    response_unit_pread =
np.concatenate([ru_drift_corr_preassoc_only, ru_assoc_only, ru_dissoc_only])
                elif fit_type == "lmt":
                 #### Might not need to use this!
                 #print "Fitting langmuir 1:1 with mass transfer...."
                    ru_assoc_only, assoc_values_pred,\
                    ru_dissoc_only, dissoc_values_pred,\
                    t_assoc_only, t_dissoc_only, coefs_df,\
                    afit_params, dfit_params =
solveAssocDissocEqsLMT(assoc_all_init,\
                    dissoc_noD_init,\
                    analyte_conc,\
                    fit_type,\
                    plot_prefix)
                elif fit_type == "lmts":
                 #print "Fitting langmuir 1:1 with mass transfer (Area surface
conc.)...."
                    ru_assoc_only, assoc_values_pred,\
                    ru_dissoc_only, dissoc_values_pred,\
                    t_assoc_only, t_dissoc_only, coefs_df,\
```

```
                                afit_params, dfit_params =
solveAssocDissocEqsAsurfLMT(assoc_all_init,\
                                dissoc_noD_init,\
                                analyte_conc,\
                                fit_type,\
                                plot_prefix)
                        elif fit_type == "hl":
                         #rint "Fitting Heterogeneous ligand...."
                                ru_assoc_only, assoc_values_pred,\
                                ru_dissoc_only, dissoc_values_pred,\
                                t_assoc_only, t_dissoc_only, coefs_df,\
                                afit_params, dfit_params =
solveAssocDissocEqsHL(assoc_all_init,\
                                dissoc_noD_init,\
                                analyte_conc,\
                                fit_type, plot_prefix)
                # Calculate assoc and dissoc residuals
                # resids, resid_ssq, resid_sd
                # Add 1 to dissoc params since one value is missing till we fit assoc phase
                        assoc_params_count = len(assoc_all_init)
                        dissoc_params_count = len(dissoc_noD_init)
                        assoc_resids, assoc_resid_ssq,\
```

```
                    assoc_resid_sd, assoc_chisq =
calcResidualsSsqSD(ru_assoc_only,\
                                    assoc_values_pred,\
                                    assoc_params_count)
                    dissoc_resids, dissoc_resid_ssq,\
                    dissoc_resid_sd, dissoc_chisq =
calcResidualsSsqSD(ru_dissoc_only,\
                    dissoc_values_pred,\
                    dissoc_params_count)
                    t_assoc_dissoc = np.concatenate([t_assoc_only, t_dissoc_only])
                    assoc_dissoc_values_pred = np.concatenate([assoc_values_pred,\
                    dissoc_values_pred])
                    assoc_dissoc_values_resids = np.concatenate([assoc_resids,\
                    dissoc_resids])
                    add_coefs_df = DataFrame([{'assoc_resid_sd' : assoc_resid_sd,\
                    'dissoc_resid_sd' : dissoc_resid_sd,\
                    'fit_quality' : 'SUCCESS'}])
                    coefs_df = concat([spot_info_df, add_coefs_df, coefs_df], axis=1)
                    plot_filenamepath = combineAndPlot(response_unit_pread,
t_pread,\
                                    t_assoc_dissoc,\
                                    assoc_dissoc_values_pred,\
```

```
                                    assoc_dissoc_values_resids,\
                                    raw_time,\
                                    raw_response,\
                                    plot_prefix, plot_file_path)
                merger.append(PdfFileReader(plot_filenamepath, "rb"))
            except ValueError:
                print "Skipping ", ref_spot_name, ". Data might be too noisy..."
                add_coefs_df = DataFrame([{'fit_quality' : 'NO FIT'}])
                coefs_df = concat([spot_info_df, add_coefs_df], axis=1)
            one_spot_all_coefs_df = one_spot_all_coefs_df.append(coefs_df)
            if re.search(r'-C', prot_block_name) and response_icol == 1:
                term_refd_response_unit_pread = response_unit_pread
            elif re.search(r'-N', prot_block_name) and response_icol == 2:
                term_refd_response_unit_pread = response_unit_pread
    return term_refd_response_unit_pread, one_spot_all_coefs_df
```

APPENDIX H

“MERGEPDFSONFILENAME.PY” DIRECTIONS AND SCRIPT

DIRECTIONS

Below are the directions for running the “mergePDFsOnFileName.py” script:

1. Make folder in the C: drive labeled “MergePDFs”
2. Place script “mergePDFsOnFileNames.py” in folder
3. Copy separate PDFs generated for one SPRi analyses into the folder
4. Open Anaconda Prompt
    a. *cd c:\MergePDFs*
    b. *activate stanscript*
    c. *python mergePDFsOnFileNames.py -p c:\MergePDFs*
5. When script is complete, delete the separate PDFs and move the merged files into the appropriate folder
6. Repeat steps 3, 4b, and 5 until all PDF datasets that you want to make are completed

SCRIPT

```
import re
from os.path import *
from os import listdir
from PyPDF2 import PdfFileReader, PdfFileMerger
from optparse import *
def mergePDFsInFolder(input_folder_path,\
                      merged_ref1ref2_pdfname="Ref1_Ref2_merged.pdf",
```

```
                merged_ref3_pdfname="Ref3_merged.pdf"):
    pdf_merger_ref1ref2 = PdfFileMerger()
    pdf_merger_ref3 = PdfFileMerger()
    for one_file in sorted(listdir(input_folder_path)):
        if one_file.endswith(".pdf"):
            #print one_file
            ref1c_ref2n_pattern = re.match(".*-C.*ref1.*|.*-N.*ref2.*", one_file)
            ref3_pattern = re.match(".*ref3.*", one_file)
            if(ref1c_ref2n_pattern and ref1c_ref2n_pattern.group(0)):
                #print ref1c_ref2n_pattern.group(0)
                #print one_file
                pdf_merger_ref1ref2.append(PdfFileReader(join(input_folder_path, one_file),
"rb"))
            else:
                pass
            if(ref3_pattern and ref3_pattern.group(0)):
                #print ref3_pattern.group(0)
                #print one_file
                pdf_merger_ref3.append(PdfFileReader(join(input_folder_path, one_file),
"rb"))
            else:
                pass
```

```
        else:
            pass
    pdf_merger_ref1ref2.write(join(input_folder_path, merged_ref1ref2_pdfname))
    pdf_merger_ref3.write(join(input_folder_path, merged_ref3_pdfname))
    return(merged_ref1ref2_pdfname, merged_ref3_pdfname)
# Read input folder path, output filename and pass to rquired function
def main():
  input_folder_path,\
  merged_ref1ref2_pdfname,\
  merged_ref3_pdfname = readAndParseCommandlineArgs()
  merged_pdf_names = mergePDFsInFolder(input_folder_path,\
                                merged_ref1ref2_pdfname,\
                                merged_ref3_pdfname)
  print merged_pdf_names, "Done!"
#
# Read command line options
#
def readAndParseCommandlineArgs():
  usage = "usage: %prog [options]   (Use -h or --help to see all options)"
  cl=OptionParser(usage=usage)
  cl.add_option('--pdfpath', '-p', action='store',
              help="FULL PATH to input subdirectory",
```

```
            dest="pdfpath")
  cl.add_option('--outpdf1', '-1', action='store',
            help="Merged PDF filename (NO PATH)",
            default="Ref1_Ref2_merged.pdf",
            dest="outpdf1")
  cl.add_option('--outpdf2', '-2', action='store',
            help="Merged PDF filename (NO PATH)",
            default="Ref3_merged.pdf",
            dest="outpdf2")
  (options, args) = cl.parse_args()
### Need to add defaults for flexibility
  #
  # Check the command line options
  if options.pdfpath:
    if isdir(options.pdfpath):
      input_folder_path = options.pdfpath
      merged_ref1ref2_pdfname = options.outpdf1
      merged_ref3_pdfname = options.outpdf2
      return(input_folder_path, merged_ref1ref2_pdfname, merged_ref3_pdfname)
  else:
    cl.error("Please enter a filepath containing the PDFs\n")
if __name__ == "__main__":
```

```
# stuff only to run when not called via 'import' here

main()
```

APPENDIX I

SUPPLEMENTAL NAPPA-SPRI DATA ANALYSES INFORMATION

NAPPA-SPRi data analyses

Guidelines used for determining protein interactions with NAPPA-SPRi

1.) Affinity had to be within the detection limits of the instrument: 1E-05 to 1E-14

2.) The curve had to look real by eye, while taking into account what the raw curve looked like.

3.) Had to be represented by at least two interactions across the four replicates (two duplicates for two concentrations).

Rule exceptions: 1) Interactions that were seen at least twice (by eye), but there was something wrong with all but one curve. For example, the kinetics were outside the range of the instrument or something weird happened during the dissociation phase that screwed up the numbers -- like a bubble. 2) PI3K had very few interactions. Therefore, some interactors identified with PI3K were only identified by one binding curve. 3) Some interactions that were observed in the RHOA dataset once were also selected if the response was high (> 100 RU after referencing).

Additional notes on NAPPA-SPRi analyses

The kinetic data (k_a, k_d) was averaged within duplicates, then the K_D was determined from these data. If the query interacted with the target protein with the fusion tag at the N- and C-terminus, then the interaction with the strongest affinity was selected to represent the interaction.

BLNK, PI3K, RAC1, and RHOA data were referenced to LIME1. Array proteins with an N-terminal HaloTag were referenced to N-terminally tagged LIME1, and array

proteins with a C-terminal HaloTag were referenced to C-terminally tagged LIME1. BTK data were similarly referenced to N- or C-terminally tagged LUC2 (i.e., luciferase). Spot numbers refer to the spot location used from the Plexera Data Module software files.

All data was analyzed with SPRite "MGLLD" parameters except for the following: 1) RAC1 curves that had mass transport were processed with SPRite, but the kinetics were determined with Scrubber.

All data was analyzed with the lower-bound and upper-bound limits for drift correction as -/+ 20% except for the following: BTK and RHOA datasets. In these datasets, the lower-bound limit for drift correction was changed to 0% to +20%. In the curveFittingKineticModels.py script (Appendix G), the line "lld_global_assoc_dissoc_param_bounds = ((0, 0, 0, -assoc_slope_limit, -np.inf),(np.inf, np.inf. np.inf, assoc_slope_limit,np.inf))" under the "def assocdissocEqLLD_global" sub routine was changed to "lld_global_assoc_dissoc_param_bounds = ((0, 0, 0, 0, -np.inf),(np.inf, np.inf. np.inf, assoc_slope_limit,np.inf))". The original bounds resulted in binding curves with strange-looking dissociation curves (e.g., curves that were 0 or negative k_d).

APPENDIX J

NAPPA-SPRI LAY-OUT, PLASMID DNA DEPOSITION, AND PROTEIN DISPLAY

Lay-out of plasmid cDNA on SPR slide

Top of slide (where sample is first injected)

AKT2-N[1]	BLK-C	Fos-N	IKBKB-C[3]	MAPK8-N	MDM2-C	PIK3R1-N[6]	AKT2-N[2]	BCL2L1-C	FCGR2B-N	IKBKA-C	MAPK3-N	MAPK14-C	PIK3R1-N[5]	AKT1-N	BCL2A1-C	EZR-N	IFITM1-C	MAPK1-N	MAPK13-C	PIK3CG-N
PPP3CA-C	ARHGEF7-N	CARD11-C	GSK3B-N	INPP5D-C	MAPK13-N	NCKAP1L-C	PLCG2-C	ARHA-N	BTK-C	GRB2-N	IKBKG-C	MAPK12-N	NCK1-C	PIK3R5-C	AKT3-N	BLNK-C	GRAP2-N	IKBKB-C[4]	MAPK9-N	MYC-C
PIK3R5-N	PPP3R2-C	RasGRP3-N	RAP2C-C	SOS1-N	RELA-C	TEC-N	PIK3R3-N	PPP3R1-C	BCL2-N	CD22-C	IKBKA-N	Jun-C	MDM2-N	PIK3R2-N	PPP3CB-C	BCL10-N	CD19-C	HRAS-N	INPPL1-C	MAPK14-N
SYK-C	VAV2-N	VAV1-C		AKT2-N[1]	Fos-N	MAPK8-N	NFATC1-C	PPP3CA-N	PRKCB-C	VAV2-C	AKT2-N[2]	FCGR2B-N	MAPK3-N	NFAT5-C	PLCG2-N	PRKCA-C		AKT1-N	EZR-N	MAPK1-N
PIK3R1-N[6]	ARHGEF7-N	GSK3B-N	MAPK13-N	PIK3R5-N	RasGRP3-N	SOS1-N	PIK3R1-N[5]	ARHA-N	GRB2-N	MAPK12-N	PIK3R3-N	BCL2-N	IKBKA-N	PIK3CG-N	AKT3-N	GRAP2-N	MAPK9-N	PIK3R2-N	BCL10-N	HRAS-N
TEC-N	VAV2-N	BLK-C	IKBKB-C[3]	MDM2-C	PPP3CA-C	CARD11-C	MDM2-N	PPP3CA-N	BCL2L1-C	IKBKA-C	MAPK14-C	PLCG2-C	BTK-C	MAPK14-N	PLCG2-N	BCL2A1-C	IFITM1-C	MAPK13-C	PIK3R5-C	BLNK-C
INPP5D-C	NCKAP1L-C	PPP3R2-C	RAP2C-C	RELA-C	SYK-C	VAV1-C	IKBKG-C	NCK1-C	PPP3R1-C	CD22-C	Jun-C	NFATC1-C	PRKCB-C	IKBKB-C[4]	MYC-C	PPP3CB-C	CD19-C	INPPL1-C	NFAT5-C	PRKCA-C
AKT2-C[1]	CD81-N	FCGR2B-C	Luc2-N	MAP3K3-C	NFKBIE-N	PIK3CD-C	AKT2-C[2]	CD79B-N	EZR-C	LIME1-N	MAP2K3-C	NFKBIB-N	PIK3CB-C	AKT1-C	CD79A-N	ETS1-C	LILRB3-N	MAP2K2-C	NFKBIA-N	PIK3CA-C
Raf1-N	ARHGEF7-C	DAPP1-N	GRB2-C	MAP2K2-N	MAPK8-C	PIK3CA-N	RAC3-N	ARHA-C	CDKN2A-N	GRAP2-C	MALT1-N	MAPK3-C	PIK3AP1-N	RAC2-N	AKT3-C	CDC42-N	Fos-C	LYN-N	MAPK1-C	NRAS-N
PIK3R1-C[6]	RAP2A-N	RAP2A-C	RELA-N	RASSF5-C	SYK-N	SOS2-C	PIK3R1-C[5]	RAP1B-N	BCL2-C	ETS1-N	HRAS-C	MAP3K3-N	MAPK12-C	PIK3CG-C	RAP1A-N	BCL10-C	EGR1-N	GSK3B-C	MAP2K3-N	MAPK9-C
VAV1-N	TP53-C	RAP1B-C		AKT2-C[1]	FCGR2B-C	MAP3K3-C	PIK3CD-N	PIK3R3-C	MSP1	VAV2-C	AKT2-C[2]	EZR-C	MAP2K3-C	PIK3CB-N	PIK3R2-C	RAP2C-N		AKT1-C	ETS1-C	MAP2K2-C
PIK3CD-C	ARHGEF7-C	GRB2-C	MAPK8-C	PIK3R1-C[6]	RAP2A-C	RASSF5-C	PIK3CB-C	ARHA-C	GRAP2-C	MAPK3-C	PIK3R1-C[5]	BCL2-C	HRAS-C	PIK3CA-C	AKT3-C	Fos-C	MAPK1-C	PIK3CG-C	BCL10-C	GSK3B-C
SOS2-C	TP53-C	CD81-N	Luc2-N	NFKBIE-N	Raf1-N	DAPP1-N	MAPK12-C	PIK3R3-C	CD79B-N	LIME1-N	NFKBIB-N	RAC3-N	CDKN2A-N	MAPK9-C	PIK3R2-C	CD79A-N	LILRB3-N	NFKBIA-N	RAC2-N	CDC42-N
MAP2K2-N	PIK3CA-N	RAP2A-N	RELA-N	SYK-N	VAV1-N	RAP1B-C	MALT1-N	PIK3AP1-N	RAP1B-N	ETS1-N	MAP3K3-N	PIK3CD-N	MSP1	LYN-N	NRAS-N	RAP1A-N	EGR1-N	MAP2K3-N	PIK3CB-N	RAP2C-N
BLK-N	CD79B-C	IKBKG-N	LILRB3-C	NCKAP1L-N	NFKB1-C	PPP3R1-N	BCL2L1-N	CD79A-C	IKBKB-N[4]	LAT2-C	NCK1-N	NFATC4-C	PPP3CC-N	BCL2A1-N	CD72-C	IKBKB-N[3]	KRAS-C	MYC-N	NFATC3-C	PPP3CB-N
RAC1-C	CD19-N	CDKN2A-C	Jun-N	LYN-C	NFATC3-N	NFKBIE-C	PTPN6-C	CARD11-N	CDC42-C	INPPL1-N	Luc2-C	NFATC1-N	NFKBIB-C	PTEN-C	BTK-N	CD81-C	INPP5D-N	LIME1-C	NFAT5-N	NFKBIA-C
PTEN-N	Raf1-C	RASSF5-N	RasGRP3-C	SOS2-N	SOS1-C	TP53-N	PRKCB-N	RAC3-C	CD72-N	EGR1-C	LAT2-N	MAP2K1-C	NFKB1-N	PRKCA-N	RAC2-C	CD22-N	DAPP1-C	KRAS-N	MALT1-C	NFATC4-N
TEC-C	VAV3-N	VAV3-C		BLK-N	IKBKG-N	NCKAP1L-N	PIK3AP1-C	RAC1-N	pDONR221		BCL2L1-N	IKBKB-N[4]	NCK1-N	NRAS-C	PTPN6-N	RAP1A-C		BCL2A1-N	IKBKB-N[3]	MYC-N
PPP3R1-N	CD19-N	Jun-N	NFATC3-N	PTEN-N	RASSF5-N	SOS2-N	PPP3CC-N	CARD11-N	INPPL1-N	NFATC1-N	PRKCB-N	CD72-N	LAT2-N	PPP3CB-N	BTK-N	INPP5D-N	NFAT5-N	PRKCA-N	CD22-N	KRAS-N
TP53-N	VAV3-N	CD79B-C	LILRB3-C	NFKB1-C	RAC1-C	CDKN2A-C	NFKB1-N	RAC1-N	CD79A-C	LAT2-C	NFATC4-C	PTPN6-C	CDC42-C	NFATC4-N	PTPN6-N	CD72-C	KRAS-C	NFATC3-C	PTEN-C	CD81-C
LYN-C	NFKBIE-C	Raf1-C	RasGRP3-C	SOS1-C	TEC-C	VAV3-C	Luc2-C	NFKBIB-C	RAC3-C	EGR1-C	MAP2K1-C	PIK3AP1-C	pDONR221	LIME1-C	NFKBIA-C	RAC2-C	DAPP1-C	MALT1-C	NRAS-C	RAP1A-C

[1] BC063421
[2] BC120994
[3] BC006231
[4] BC108694
[5] BC030815
[6] BC094795

Figure 103. Lay-out of plasmid cDNA and expressed proteins on SPR slide. The plasmid cDNAs encoding for genes-of-interest were deposited on the array in a random manner using a pin spotter.

Plasmid cDNA deposition on SPR slide

The quality of the printing onto the SPR slide was assessed using a fluorescent nucleic stain, PicoGreen® (Thermo Fisher Scientific; Waltham, MA). First, the slides were blocked with Tris-based SuperBlock (Thermo Fisher Scientific; Waltham, MA) to minimize non-specific binding overnight at 4 ^{o}C. Then, PicoGreen® diluted in SuperBlock at 1:500 was applied to the slide, incubated in the dark for 10 min, washed three times in 1x PBS, rinsed in water, and dried under compressed air. Fluorescence was determined using the PowerScanner Micorarray^{TM} from Tecan Group Ltd. (Switzerland).

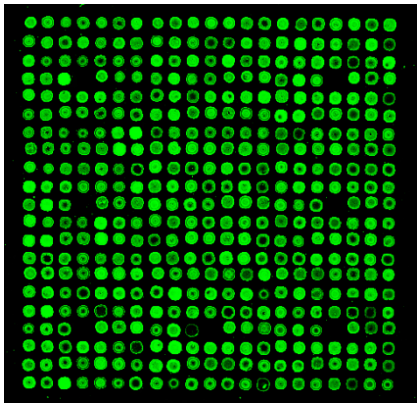


Figure 104. False-colored image of DNA deposition using PicoGreen

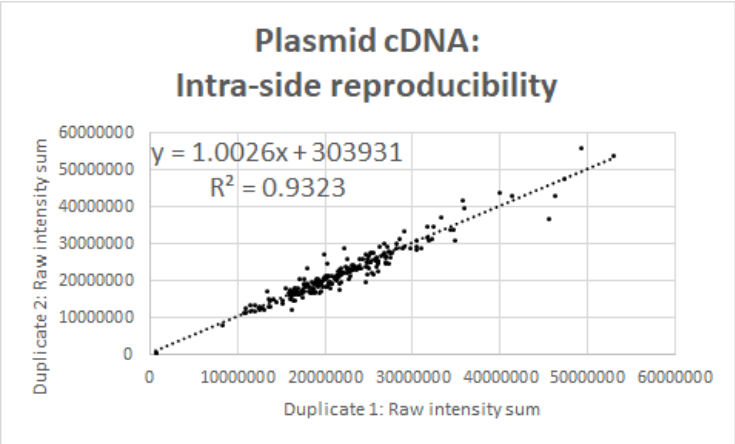


Figure 105. Reproducibility of plasmid cDNA deposition across duplicates

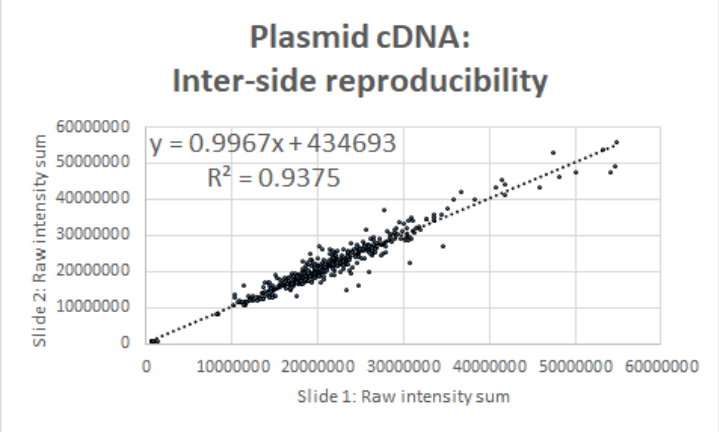


Figure 106. Reproducibility of plasmid cDNA deposition across slides

Protein expression on SPR slide

The expression and subsequent capture of target proteins onto the SPR slides were assessed fluorescently. First, the slides were blocked with Tris-based SuperBlock (Thermo Fisher Scientific; Waltham, MA) to minimize non-specific binding overnight at 4 ^{o}C. They were then washed in 1x PBS three times for 2 min each, rocking. The slides were rinsed in water and dried with compressed air. SPRi flow chambers (Plexera;

Woodinville, WA) with 30 µL volume were applied onto the slides followed by 1-step human coupled *in vitro* protein expression mixture according to the manufacturer's instructions (Thermo Fisher Scientific; Waltham, MA). Expression was performed for 1.5 hours at 30 ºC and then 30 min at 15 ºC. To remove the flow cells, the slides were placed at -80 ºC for 30 sec. Slides were rinsed in 200 µL 1x PBS and blocked for 1 hour at RT with 5% milk in 1x PBST ("blocking buffer"). The slides were incubated in rabbit anti-HaloTag polyclonal antibody (Promega; Madison, WI) diluted 1:250 in blocking buffer for 1 hour at RT, rocking. After washing the slides three times in blocking buffer, the slides were incubated in Alexa Fluor 555 goat anti-rabbit IgG (Thermo Fisher Scientific; Waltham, MA) diluted 1:500 in blocking buffer for 1 hour at RT, rocking. The slides were then washed three times in 1x PBS, rinsed in water, and dried under compressed air. Fluorescence was determined using the PowerScanner Micorarray™ from Tecan Group Ltd. (Switzerland).

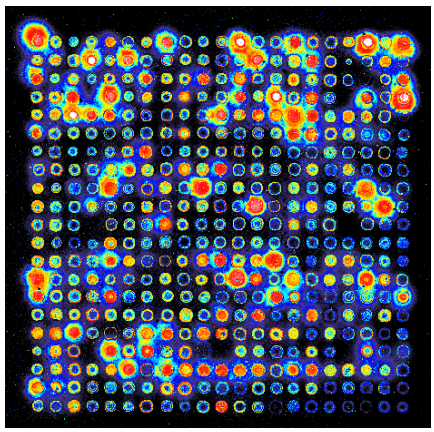


Figure 107. Target protein expression on an SPR slide as determined with an anti-HaloTag antibody. False-colored rainbow image where black represents low protein expression and red represents high protein expression. Note that the HaloTag binds preferentially to proteins with an N-terminal HaloTag.

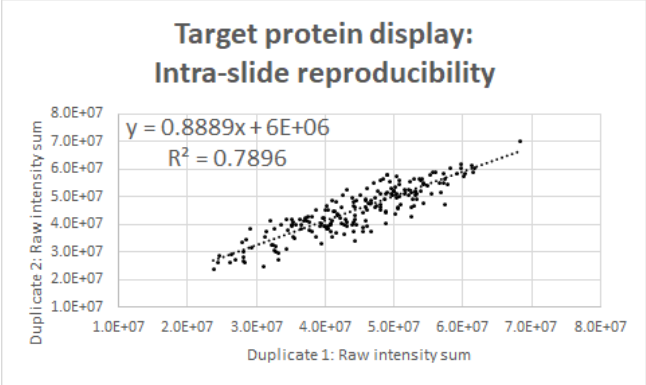


Figure 108. Reproducibility of displayed protein across duplicates.

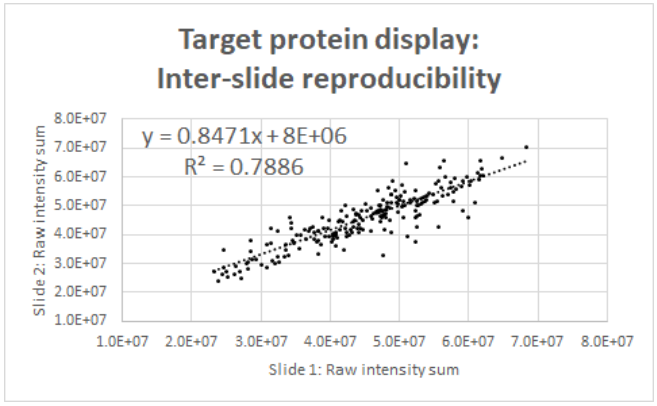


Figure 109. Reproducibility of displayed protein across different slides.

Table 35. Buffer conditions used for NAPPA-SPRi

Query	Flow (uL/sec)	Assocation (sec)	Dissociation (sec)	Temp °C	SPR running buffer (bulk)	Additives in sample
BLNK	3	300	700	25	50 mM HEPES, 150 mM NaCl, 0.05% Tween-20, pH 7.4	1 mM TCEP
BTK	5	180	400	30	50 mM Tris-HCl, 150 mM NaCl, 4 mM MgCl2, 0.01% Tween, pH 7.5	500 uM ATP, 1 mM TCEP
PIK3CA/PIK3R1	5	180	400	30	50 mM HEPES, 100 mM NaCl, 3 mM MgCl2, 0.03% Tween, pH 7.5	1 mM TCEP, 250 uM ATP
RAC1(GDP)	3	300	700	25	50 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 0.05% Tween-20, pH 7.4	1 mM TCEP, 1 mM GDP
RAC1(GTP)	3	300	700	25	50 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 0.05% Tween-20, pH 7.4	1 mM TCEP, 1 mM GTPγS
RHOA(GDP)	3	300	700	25	50 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 0.05% Tween-20, pH 7.4	1 mM TCEP, 1 mM GDP
RHOA(GTP)	3	300	700	25	50 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 0.05% Tween-20, pH 7.4	1 mM TCEP, 1 mM GTPγS

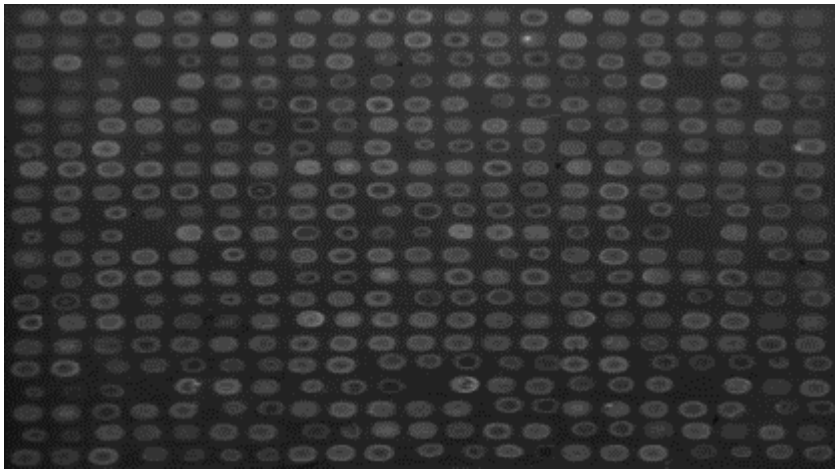


Figure 110. NAPPA-SPRi image on the Plexera SPRi biosensor.

APPENDIX K

PROTEIN INTERACTIONS DETECTED BY NAPPA-SPRI

Table 36. PPIs detected by NAPPA-SPRi, part 1

		BLNK		BTK		PI3K		RAC1(GDP)		RAC1(GTP)		RHOA(GDP)		RHOA(GTP)	
Protein	Isoform	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT
AKT1		C/N	C	N	N	C	C/N			C	C	C	C	C	C
AKT2	BC063421	C	C	C/N	C/N	N			C	C	C	C		C	C
AKT2	BC120994	C	C	C/N	C/N	C	C			C		C		C	C
AKT3				C/N	N			C/N	C/N	C/N	C/N			C	C
ARHGEF7				N	C/N									N	N
BCL10		N	N	C/N	C/N			N	N	C/N	N			C/N	
BCL2				C/N	C/N					C				N	
BCL2A1		C/N	C/N	C	C	C	C		C	C/N	C	C	C	C	C
BCL2L1		N	C	C	C			C	C	C	C	C		C	C
BLK				C	C		N			C/N				C	C
BLNK				C	C			C	C	C	C			C	C
BTK				C	C									C	C
CARD11		C	C	C	C			C	C	C/N	C	C		C	C
CD19				C	C			C	C	C	C			C	C
CD22		C	C	C	C	C		C	C	C/N	C/N			C	C
CD72		C	C							C/N		C		C	C
CD79A		C/N	C	N	N	C				C					
CD79B			C	N	N					C	C			C	
CD81		N	N					N	N	C/N	C/N			C/N	
CDC42					N					N				N	
CDKN2A		N	C/N					N	N	C/N	C/N			N	
DAPP1		C	N							C/N	N			N	
EGR1		C							N	C/N	C/N			N	
ETS1		N	N		C	N	N	N	N	C/N	C/N	C		C/N	C
EZR				C/N	C/N			N	N	N	N			C/N	C
FCGR2B				C/N	C/N					C				C	C
Fos				N	N				C	C/N	C			C	C
GRAP2				C/N	C/N					C					C
GRB2		C/N	C/N	N	C/N			N	C/N	C/N	C/N	C		C/N	C/N
GSK3B			C	N	N			C	C	C	C			C/N	
HRAS		C/N	C/N	C/N	N	C/N		C/N	C/N	C/N	C/N			C/N	
IFITM1				C	C									C	C
IKBKA			N	C/N	C/N	N		N	N	N	N			C/N	C/N
IKBKB	BC006231			C	C					C/N	C	C		C/N	C
IKBKB	BC108694	C	C	C	C	C		C	C	C	C	C		C	C
IKBKG		C	C	C	C	C		C	C	C/N	C/N	C		C	C
INPP5D				C	C					C		C		C	C
INPPL1		C	C	C	N	C		C	C	C	C			C	C
Jun		C/N	C/N	C	C			C/N	C/N	C/N	C/N			C/N	C
KRAS		N	N			N		N	N	N	N			C/N	
LAT2		C/N	C/N			N	N	N	C/N	N	C/N	C		C/N	C
LILRB3					N					C	C	C		C	C
LIME1				N	N										
LYN		N						N	N	C/N	C/N			N	
MALT1		C	C	N	N					C					
MAP2K1		C	C							C	C				
MAP2K2					C			C/N	C/N	C/N	C/N	C		C/N	C
MAP2K3				C	C			C/N	C/N	C/N	C/N	C		C/N	C
MAP3K3		N	N			N		N	N	C/N	C/N	C		C/N	C

"N" and "C" indicate whether the HaloTag is at the N- or C-terminus, respectively.
Only the target proteins that interacted with a query protein are shown.
Known PPIs in human and and mouse are highlighted in blue. Obtained from the online PPI databases, HPRD and BioGRID.

Table 37. PPIs detected by NAPPA-SPRi, part 2

		BLNK		BTK		PI3K		RAC1(GDP)		RAC1(GTP)		RHOA(GDP)		RHOA(GTP)	
Protein	Isoform	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT	NP	LT
MAPK1		C	C	C/N	N	C	C	C	C	C	C	C		C	C
MAPK12		C	C	C/N	N	C		C	C	C	C			C	
MAPK13		C	C	C/N	C/N	C		C	C	C	C	C	C	C/N	C
MAPK14		C		C/N	C/N			C	C	C	C	C	C	C/N	C/N
MAPK3		N		C/N	N			C	C	C	C			C	C
MAPK8				C/N	C/N					C/N				C	C
MAPK9				N	N				C	C	C			C	C
MDM2				C/N	C/N	C		N	N	C/N	C/N	C		C/N	C
MYC		C/N	C/N	C	C	C	C	C/N	C/N	C/N	C/N	C		C	C
NCK1				C	C			C/N	C/N	C/N	C/N			C	C
NCKAP1L		C	C	C	C	C		C	C	C	C	C		C	C
NFAT5		N	N	C	C			C/N	C/N	C/N					C
NFATC1		C	C	C				C	C	C	C			C	C
NFATC3										C		C			
NFATC4			C							C	C	C		C	
NFKB1		N							N	C/N	C/N	C			
NFKBIA		C/N	N					N	N	C/N	C/N			C/N	
NFKBIB								N	N	N	N				
NFKBIE		C	C	N	N			C		C	C			C	C
NRAS		N	N					N	N	C/N	N			C/N	
PIK3AP1		N	N			N	N	N	N	N	N			C	
PIK3CA		C/N	C/N			N	N	C/N	C/N	C/N	C/N	C/N		C/N	C/N
PIK3CB				C	C					N		C		C	C
PIK3CD															
PIK3CG		N		N	N			N	N	C/N	N			C/N	C
PIK3R1	BC030815	N	N	C/N	C/N			C/N	C/N	C/N	C/N	N		N	C/N
PIK3R1	BC094795			N	N	N	N								
PIK3R2				N	N					N					
PIK3R3				C/N	C/N			C	C	C	C				C/N
PIK3R5				C/N	C/N			C	C	C				C/N	C
PLCG2				C/N	C/N	C								C/N	C
PPP3CA				C/N	C/N									C/N	C/N
PPP3CB				C	C	N			C	C	C	C		C	C
PPP3R1				C	C			C	C	C	C	C		C	C
PPP3R2				C	C									C	C
PRKCA				C										C	C
PRKCB				C	C			C	C	C	C			C	C
PTEN		C	C			C	C			C/N	C			C	
PTPN6		C	C			C	C	C	C	C/N	C	C		C	
RAC1		N				N			N	C/N	C/N				C
RAC2				N	N										
RAC3					N			C	C	C	C				
Raf1				N	N									C	
RAP1A															
RAP1B		C	C	C	C/N			C	C	C	C			C	C
RAP2A		C	C	C	C			C	C	C	C			C/N	C
RAP2C		N	N	C	C	C/N	N	C/N	C/N	C/N	C/N	C	C	C	C
RasGRP3				N	N			N	N	C/N	N			N	
RASSF5				C	C					C/N				C	C
RELA				C	C	C	C			C/N	C/N			C/N	C
RHOA		C	C	C/N	N	C	C	C	C	C	C	C		C	C
SOS1		C/N	N	N	N				N	C/N	N			N	
SOS2					C				N	C/N	C/N			C	C
SYK			C	C	C			C	C	C/N	C/N			C	C
TEC				C/N	C/N			N		C/N	N			N	N
TP53										C/N	C/N			C	C
VAV1			N	C	C			N	N	C/N	N			C	C
VAV2		N	N	C	C	C		N	N	C/N	N			C/N	C/N
VAV3		C	C	C		C				C/N	C			C	

"N" and "C" indicate whether the HaloTag is at the N- or C-terminus, respectively.
Only the target proteins that interacted with a query protein are shown.
Known PPIs in human and and mouse are highlighted in blue. Obtained from the online PPI databases, HPRD and BioGRID.

APPENDIX L

TABLES OF NAPPA-SPRI KINETIC DATA

Table 38. BLNK and BTK queries: K_D, k_a, and k_d data, part 1

	BLNK_NP			BLNK_LT			BTK_NP			BTK_LT		
Protein	KD	ka	kd	KD	ka	kd	KD	ka	kd	KD	ka	kd
AKT1	7.56E-09	6.15E+04	4.65E-04	1.73E-08	6.15E+04	1.07E-03	5.11E-08	5.35E+04	2.73E-03	1.19E-07	2.69E+04	3.21E-03
AKT2*	2.77E-08	5.81E+04	1.61E-03	1.90E-08	8.02E+04	1.52E-03	3.56E-08	7.28E+04	2.59E-03	1.89E-08	1.12E+05	2.11E-03
AKT2**	2.90E-08	6.64E+04	1.93E-03	2.68E-08	6.89E+04	1.84E-03	1.78E-08	1.07E+05	1.90E-03	2.61E-08	1.17E+05	3.04E-03
AKT3							4.07E-08	5.85E+04	2.38E-03	1.99E-07	1.31E+04	2.60E-03
ARHGEF7							1.37E-07	1.70E+04	2.33E-03	6.61E-08	2.15E+04	1.42E-03
BCL10	2.97E-08	5.77E+04	1.71E-03	1.71E-08	6.03E+04	1.03E-03	8.91E-08	3.57E+04	3.19E-03	3.14E-07	9.61E+03	3.01E-03
BCL2							2.09E-08	1.33E+05	2.77E-03	1.75E-09	4.60E+03	8.06E-06
BCL2A1	1.85E-08	4.13E+04	7.64E-04	1.56E-08	7.91E+04	1.23E-03	6.66E-08	5.94E+04	3.96E-03	5.56E-08	4.60E+04	2.56E-03
BCL2L1	1.62E-08	4.44E+04	7.19E-04	1.46E-08	7.22E+04	1.06E-03	3.66E-08	7.71E+04	2.82E-03	2.02E-08	7.44E+04	1.50E-03
BLK							5.30E-08	6.09E+04	3.23E-03	4.54E-08	6.75E+04	3.06E-03
BLNK							7.51E-08	3.23E+04	2.42E-03	1.10E-07	1.88E+04	2.07E-03
BTK							7.94E-08	2.36E+04	1.88E-03	9.24E-08	1.02E+04	9.41E-04
CARD11	3.34E-08	3.78E+04	1.26E-03	2.17E-08	4.88E+04	1.06E-03	1.80E-07	1.55E+04	2.78E-03	2.21E-08	1.19E+05	2.62E-03
CD19							1.48E-07	4.92E+04	7.28E-03	4.41E-08	4.01E+04	1.77E-03
CD22	1.25E-08	8.18E+04	1.02E-03	1.21E-08	7.46E+04	9.06E-04	2.65E-08	1.03E+05	2.72E-03	2.82E-08	1.26E+05	3.56E-03
CD72	3.02E-08	6.19E+04	1.87E-03	1.60E-08	5.54E+04	8.85E-04						
CD79A	1.83E-08	6.38E+04	1.17E-03	2.22E-08	7.01E+04	1.56E-03	2.42E-08	9.11E+04	2.21E-03	1.28E-07	3.00E+04	3.83E-03
CD79B				4.26E-08	5.59E+04	2.38E-03	2.65E-08	8.90E+04	2.36E-03	5.71E-08	2.74E+04	1.56E-03
CD81	1.98E-08	4.46E+04	8.86E-04	1.99E-08	5.89E+04	1.17E-03						
CDC42										8.23E-07	1.91E+03	1.57E-03
CDKN2A	1.89E-08	4.21E+04	7.96E-04	2.37E-08	5.33E+04	1.26E-03						
DAPP1	1.76E-08	5.10E+04	9.00E-04	2.34E-08	6.61E+04	1.55E-03						
EGR1	2.28E-09	6.46E+04	1.47E-04									
ETS1	1.10E-08	4.36E+04	4.77E-04	1.83E-08	5.50E+04	1.01E-03				1.52E-07	1.36E+04	2.06E-03
EZR							3.28E-08	8.02E+04	2.63E-03	5.74E-08	4.65E+04	2.67E-03
FCGR2B							2.45E-08	8.19E+04	2.01E-03	4.38E-08	4.02E+04	1.76E-03
Fos							8.79E-08	3.05E+04	2.68E-03	8.76E-08	3.39E+04	2.96E-03
GRAP2							3.16E-08	6.57E+04	2.08E-03	9.33E-08	2.30E+04	2.15E-03
GRB2	1.93E-08	5.35E+04	1.03E-03	1.37E-08	4.84E+04	6.64E-04	2.96E-08	6.15E+04	1.82E-03	1.65E-08	1.47E+05	2.42E-03
GSK3B				2.40E-08	5.67E+04	1.36E-03	5.21E-08	4.72E+04	2.46E-03	9.08E-08	2.40E+04	2.18E-03
HRAS	1.90E-08	4.83E+04	9.17E-04	1.71E-08	5.43E+04	9.29E-04	5.49E-08	5.42E+04	2.98E-03	1.70E-08	3.54E+03	6.01E-05
IFITM1							8.03E-08	4.61E+04	3.70E-03	7.83E-08	3.49E+04	2.73E-03
IKBKA				2.00E-08	5.82E+04	1.16E-03	3.96E-08	5.99E+04	2.37E-03	4.39E-08	5.30E+04	2.33E-03
IKBKB#							4.70E-08	5.85E+04	2.75E-03	5.13E-08	5.16E+04	2.64E-03
IKBKB##	4.19E-08	4.19E+04	1.76E-03	2.34E-08	6.56E+04	1.54E-03	9.75E-08	3.96E+04	3.86E-03	3.83E-08	4.41E+04	1.69E-03
IKBKG	2.40E-08	4.75E+04	1.14E-03	1.80E-08	6.22E+04	1.12E-03	4.41E-08	6.03E+04	2.66E-03	2.00E-08	6.83E+04	1.37E-03
INPP5D							4.65E-08	6.55E+04	3.05E-03	3.62E-08	7.61E+04	2.76E-03
INPPL1	2.48E-08	5.86E+04	1.45E-03	1.83E-08	5.50E+04	1.01E-03	1.23E-07	5.81E+04	7.13E-03	6.03E-08	1.38E+04	8.31E-04
Jun	7.86E-09	3.49E+04	2.74E-04	2.42E-08	6.56E+04	1.59E-03	2.66E-08	8.64E+04	2.29E-03	8.41E-08	3.61E+04	3.04E-03
KRAS	1.37E-08	5.19E+04	7.12E-04	3.19E-08	4.42E+04	1.41E-03						
LAT2	1.34E-08	5.16E+04	6.90E-04	1.59E-08	6.91E+04	1.10E-03						
LILRB3										5.74E-07	2.54E+03	1.46E-03
LIME1							5.33E-08	4.63E+04	2.47E-03	2.66E-07	1.11E+04	2.94E-03
LYN	1.61E-09	3.73E+04	6.01E-05									
MALT1	2.74E-08	6.55E+04	1.79E-03	3.45E-08	5.86E+04	2.03E-03	8.17E-08	2.91E+04	2.38E-03	1.20E-06	1.99E+03	2.40E-03
MAP2K1	7.90E-09	1.27E+05	1.01E-03	2.62E-08	6.04E+04	1.58E-03						
MAP2K2										9.06E-08	3.23E+04	2.92E-03
MAP2K3							3.94E-08	8.76E+04	3.45E-03	3.58E-08	6.94E+04	2.49E-03
MAP3K3	1.79E-08	3.86E+04	6.93E-04	1.48E-08	6.35E+04	9.41E-04						

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Table 39. BLNK and BTK queries: K_D, k_a, and k_d data, part 2

	BLNK_NP			BLNK_LT			BTK_NP			BTK_LT		
Protein	KD	ka	kd	KD	ka	kd	KD	ka	kd	KD	ka	kd
MAPK1	2.81E-08	5.08E+04	1.43E-03	1.98E-08	5.40E+04	1.07E-03	6.36E-08	9.00E+04	5.73E-03	6.94E-08	3.76E+04	2.61E-03
MAPK12	2.45E-08	5.85E+04	1.43E-03	2.10E-08	6.48E+04	1.36E-03	4.24E-08	6.18E+04	2.62E-03	1.62E-07	1.65E+04	2.66E-03
MAPK13	2.42E-08	5.94E+04	1.44E-03	1.60E-08	6.31E+04	1.01E-03	3.90E-08	4.64E+04	1.81E-03	6.83E-08	4.39E+04	3.00E-03
MAPK14	1.60E-08	9.26E+04	1.48E-03				4.62E-08	6.65E+04	3.08E-03	3.06E-08	7.43E+04	2.28E-03
MAPK3	1.39E-08	5.34E+04	7.40E-04				3.68E-08	7.11E+04	2.62E-03	1.26E-07	2.42E+04	3.04E-03
MAPK8							2.93E-08	7.46E+04	2.19E-03	9.48E-08	2.48E+04	2.35E-03
MAPK9							9.09E-08	1.85E+04	1.68E-03	1.09E-08	1.73E+04	1.89E-04
MDM2							4.93E-08	4.61E+04	2.27E-03	4.02E-08	5.61E+04	2.25E-03
MYC	1.19E-08	4.57E+04	5.43E-04	2.12E-08	5.43E+04	1.15E-03	9.50E-08	4.99E+04	4.74E-03	4.56E-08	6.05E+04	2.76E-03
NCK1							3.61E-08	7.14E+04	2.58E-03	3.10E-08	7.47E+04	2.31E-03
NCKAP1L	2.58E-08	4.88E+04	1.26E-03	1.96E-08	6.03E+04	1.18E-03	3.33E-08	7.20E+04	2.40E-03	2.76E-08	7.73E+04	2.14E-03
NFAT5	1.68E-08	5.84E+04	9.82E-04	3.51E-08	2.63E+04	9.24E-04	4.35E-08	7.60E+04	3.30E-03	1.34E-07	1.98E+04	2.66E-03
NFATC1	2.59E-08	5.28E+04	1.37E-03	1.77E-08	7.15E+04	1.27E-03	2.61E-08	1.13E+05	2.94E-03			
NFATC4				2.02E-08	6.75E+04	1.36E-03						
NFKB1	7.21E-09	4.69E+04	3.38E-04									
NFKBIA	1.39E-08	6.08E+04	8.45E-04	2.06E-08	4.52E+04	9.31E-04						
NFKBIE	5.78E-09	4.62E+04	2.67E-04	3.12E-08	4.17E+04	1.30E-03	2.95E-08	7.43E+04	2.19E-03	6.05E-07	6.79E+03	4.11E-03
NRAS	4.28E-08	1.91E+04	8.20E-04	3.58E-08	3.22E+04	1.15E-03						
PIK3AP1	1.34E-08	3.75E+04	5.02E-04	2.01E-08	4.79E+04	9.60E-04						
PIK3CA	1.80E-08	3.80E+04	6.83E-04	1.89E-08	5.35E+04	1.01E-03						
PIK3CB							3.24E-08	5.80E+04	1.88E-03	4.04E-08	4.23E+04	1.71E-03
PIK3CG	1.53E-08	4.73E+04	7.27E-04				2.11E-07	1.80E+04	3.79E-03	6.29E-08	2.42E+04	1.53E-03
PIK3R1~	2.20E-08	4.29E+04	9.45E-04	1.82E-08	4.80E+04	8.75E-04	2.90E-08	8.57E+04	2.49E-03	1.97E-08	1.26E+05	2.49E-03
PIK3R1~~							5.94E-08	5.40E+04	3.20E-03	4.90E-08	5.90E+04	2.89E-03
PIK3R2							1.45E-07	2.31E+04	3.35E-03	1.65E-07	1.11E+04	1.83E-03
PIK3R3							4.54E-08	5.07E+04	2.30E-03	2.43E-09	6.99E+04	1.70E-04
PIK3R5							3.62E-08	6.87E+04	2.48E-03	5.36E-08	4.67E+04	2.50E-03
PLCG2							8.04E-08	4.57E+04	3.68E-03	8.04E-08	2.58E+04	2.08E-03
PPP3CA							1.04E-07	3.78E+04	3.92E-03	8.17E-08	2.89E+04	2.37E-03
PPP3CB							7.21E-08	3.41E+04	2.46E-03	9.78E-08	2.38E+04	2.32E-03
PPP3R1							3.10E-08	6.92E+04	2.14E-03	3.95E-08	3.76E+04	1.48E-03
PPP3R2							3.73E-08	7.54E+04	2.81E-03	2.86E-08	7.67E+04	2.20E-03
PRKCA							3.29E-07	7.39E+03	2.43E-03			
PRKCB							4.63E-08	7.50E+04	3.47E-03	5.12E-08	4.16E+04	2.13E-03
PTEN	1.55E-08	9.05E+04	1.40E-03	2.59E-08	6.32E+04	1.64E-03						
PTPN6	2.14E-08	5.89E+04	1.26E-03	2.70E-08	5.71E+04	1.54E-03						
RAC1	6.85E-09	5.14E+04	3.52E-04									
RAC2							2.94E-06	3.75E+00	1.10E-05	1.42E-06	1.86E+03	2.65E-03
RAC3										1.28E-08	2.35E+04	3.01E-04
Raf1							6.23E-07	4.06E+03	2.53E-03	1.52E-07	5.25E+03	7.99E-04
RAP1B	1.24E-08	7.19E+04	8.92E-04	3.97E-08	5.00E+04	1.99E-03	6.26E-08	2.74E+04	1.71E-03	3.20E-08	3.47E+04	1.11E-03
RAP2A	2.00E-08	5.58E+04	1.11E-03	4.22E-08	4.77E+04	2.01E-03	1.05E-07	2.16E+04	2.26E-03	1.16E-09	7.89E+04	9.14E-05
RAP2C	1.40E-08	5.66E+04	7.93E-04	1.97E-08	5.33E+04	1.05E-03	6.21E-08	7.36E+04	4.57E-03	3.19E-08	6.28E+04	2.01E-03
RasGRP3							7.94E-08	4.11E+04	3.26E-03	7.97E-08	4.27E+04	3.40E-03
RASSF5							9.52E-08	3.68E+04	3.50E-03	3.75E-08	3.97E+04	1.49E-03
RELA							4.72E-08	7.21E+04	3.41E-03	4.10E-08	3.88E+04	1.59E-03
RHOA	2.00E-08	4.64E+04	9.30E-04	1.71E-08	6.32E+04	1.08E-03	3.31E-08	6.40E+04	2.11E-03	5.09E-08	3.96E+04	2.02E-03
SOS1	5.87E-09	5.26E+04	3.09E-04	2.44E-08	3.17E+04	7.73E-04	5.45E-08	5.36E+04	2.92E-03	1.27E-07	1.71E+04	2.16E-03
SOS2										1.54E-07	1.22E+04	1.88E-03
SYK				2.84E-08	5.86E+04	1.66E-03	4.61E-08	7.25E+04	3.34E-03	1.85E-08	8.39E+04	1.55E-03
TEC							9.25E-08	2.89E+04	2.68E-03	4.28E-08	4.15E+04	1.78E-03
VAV1				5.30E-08	3.93E+04	2.08E-03	3.80E-09	1.41E+04	5.35E-05	1.48E-08	1.67E+04	2.47E-04
VAV2	1.63E-08	4.23E+04	6.89E-04	4.88E-08	4.19E+04	2.04E-03	3.98E-08	7.82E+04	3.11E-03	2.00E-08	8.52E+04	1.70E-03
VAV3	1.92E-08	4.83E+04	9.28E-04	4.46E-08	4.87E+04	2.17E-03	5.55E-08	7.35E+04	4.08E-03			

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Table 40. PI3K query: K_D, k_a, and k_d data

	PI3K_NP			PI3K_LT		
Protein	**KD**	**ka**	**kd**	**KD**	**ka**	**kd**
AKT1	6.81E-09	5.86E+05	3.99E-03	1.31E-09	1.99E+05	2.61E-04
AKT2*	5.20E-09	1.98E+05	1.03E-03			
AKT2**	1.24E-08	4.54E+05	5.63E-03	5.69E-10	2.99E+05	1.70E-04
BCL2A1	7.63E-09	5.76E+05	4.40E-03	1.02E-09	1.66E+05	1.69E-04
BLK				1.03E-08	2.40E+05	2.47E-03
CD22	7.97E-09	5.90E+05	4.70E-03			
CD79A	1.89E-09	4.04E+05	7.65E-04			
ETS1	1.32E-08	7.88E+04	1.04E-03	2.88E-08	7.43E+04	2.14E-03
HRAS	1.38E-08	3.32E+05	4.57E-03			
IKBKA	4.58E-09	1.98E+05	9.09E-04			
IKBKB##	1.62E-08	3.35E+05	5.44E-03			
IKBKG	9.07E-09	4.93E+05	4.47E-03			
INPPL1	1.59E-08	2.92E+05	4.64E-03			
KRAS	3.59E-08	5.37E+04	1.93E-03			
LAT2	1.48E-09	1.14E+05	1.69E-04	1.98E-08	1.03E+05	2.04E-03
MAP3K3	1.72E-09	8.77E+04	1.51E-04			
MAPK1	6.38E-09	5.81E+05	3.71E-03	5.58E-09	4.06E+05	2.26E-03
MAPK12	9.95E-09	4.20E+05	4.18E-03			
MAPK13	1.03E-08	6.30E+05	6.51E-03			
MDM2	8.70E-09	6.25E+05	5.44E-03			
MYC	1.31E-08	4.74E+05	6.19E-03	4.24E-09	3.44E+05	1.46E-03
NCKAP1L	1.01E-08	5.01E+05	5.03E-03			
PIK3AP1	4.15E-10	1.94E+05	8.06E-05	1.64E-08	2.05E+06	3.37E-02
PIK3CA	7.45E-09	1.50E+05	1.12E-03	1.16E-08	1.18E+05	1.38E-03
PIK3R1~~	2.43E-09	3.67E+05	8.89E-04	8.69E-10	1.49E+06	1.30E-03
PLCG2	1.14E-08	6.60E+05	7.54E-03			
PPP3CB	1.95E-09	2.30E+05	4.49E-04			
PTEN	1.35E-08	2.63E+05	3.55E-03	7.13E-10	3.93E+05	2.80E-04
PTPN6	8.85E-09	2.98E+05	2.64E-03	5.44E-10	5.88E+05	3.20E-04
RAC1	4.20E-09	8.52E+04	3.58E-04			
RAP2C	8.11E-09	1.08E+05	8.74E-04	1.64E-08	4.33E+06	7.11E-02
RELA	1.79E-08	3.51E+05	6.27E-03	1.27E-09	6.70E+05	8.54E-04
RHOA	7.18E-09	3.32E+05	2.38E-03	6.66E-10	3.75E+05	2.50E-04
VAV2	1.00E-08	4.55E+05	4.57E-03			
VAV3	4.12E-09	2.32E+05	9.55E-04			

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Table 41. RAC1 query: K_D, k_a, and k_d data

	RAC1(GDP)_NP			RAC1(GDP)_LT			RAC1(GTP)_NP			RAC1(GTP)_LT		
Protein	KD	ka	kd	KD	ka	kd	KD	ka	kd	KD	ka	kd
AKT1							1.12E-07	7.87E+08	8.84E+01	7.82E-08	5.10E+05	3.99E-02
AKT2*				3.73E-08	4.13E+04	1.54E-03	1.21E-07	6.46E+08	7.82E+01	6.23E-08	1.62E+09	1.01E+02
AKT2**							1.12E-07	4.16E+07	4.68E+00			
AKT3	5.99E-08	2.17E+04	1.30E-03	2.44E-08	2.32E+04	5.67E-04	1.03E-07	5.22E+08	5.39E+01	8.49E-08	2.43E+04	2.06E-03
BCL10	5.15E-08	2.38E+04	1.22E-03	6.09E-08	3.02E+04	1.84E-03	5.92E-08	5.98E+07	3.54E+00	1.40E-07	1.21E+09	1.69E+02
BCL2							1.13E-07	3.13E+04	3.55E-03			
BCL2A1				5.26E-08	1.95E+04	1.03E-03	1.28E-07	4.43E+08	5.67E+01	7.39E-08	2.27E+05	1.68E-02
BCL2L1	8.34E-08	1.89E+04	1.58E-03	3.01E-08	4.19E+04	1.26E-03	1.16E-07	3.32E+09	3.86E+02	4.69E-08	5.05E+04	2.37E-03
BLK							4.93E-08	2.65E+09	1.31E+02			
BLNK	4.38E-09	1.40E+04	6.14E-05	4.18E-09	2.18E+04	9.12E-05	8.55E-08	1.69E+08	1.44E+01	4.41E-08	5.53E+04	2.44E-03
CARD11	3.50E-08	1.76E+04	6.17E-04	1.06E-08	3.18E+04	3.36E-04	6.84E-08	1.50E+09	1.02E+02	1.11E-07	1.32E+05	1.47E-02
CD19	6.49E-08	1.82E+04	1.18E-03	2.65E-08	3.31E+04	8.78E-04	7.06E-08	1.09E+09	7.68E+01	1.02E-07	3.87E+04	3.94E-03
CD22	4.46E-08	2.51E+04	1.12E-03	2.93E-08	3.16E+04	9.26E-04	6.79E-08	2.50E+08	1.70E+01	9.51E-08	3.37E+09	3.20E+02
CD72							1.22E-07	3.19E+06	3.90E-01			
CD79A							3.63E-07	3.57E+08	1.30E+02			
CD79B							1.30E-07	2.23E+09	2.89E+02	6.00E-08	5.77E+04	3.46E-03
CD81	3.49E-08	1.63E+04	5.69E-04	2.85E-08	2.59E+04	7.36E-04	6.14E-08	7.29E+08	4.48E+01	6.77E-08	1.75E+05	1.18E-02
CDC42							1.18E-07	4.63E+09	5.48E+02			
CDKN2A	7.71E-08	1.81E+04	1.40E-03	7.88E-08	2.50E+04	1.97E-03	1.16E-07	3.44E+09	4.00E+02	1.02E-07	6.59E+04	6.73E-03
DAPP1							1.04E-07	4.17E+08	4.35E+01	1.29E-07	3.22E+05	4.14E-02
EGR1				4.99E-09	2.62E+04	1.31E-04	8.05E-08	3.99E+06	3.21E-01	1.27E-07	8.17E+06	1.03E+00
ETS1	9.49E-08	1.85E+04	1.76E-03	6.16E-08	1.99E+04	1.22E-03	7.92E-08	2.39E+09	1.89E+02	3.38E-08	3.82E+04	1.29E-03
EZR	5.78E-08	6.70E+04	3.88E-03	5.42E-08	3.08E+04	1.67E-03	7.55E-08	4.10E+09	3.09E+02	6.18E-08	9.56E+05	5.91E-02
FCGR2B							7.45E-08	6.49E+08	4.83E+01			
Fos				1.02E-08	3.33E+04	3.41E-04	1.03E-07	3.42E+09	3.53E+02	7.59E-08	3.55E+04	2.69E-03
GRAP2							7.88E-08	3.66E+04	2.88E-03			
GRB2	4.26E-08	1.66E+04	7.08E-04	1.72E-08	2.74E+04	4.70E-04	8.65E-08	7.95E+02	6.88E+02	8.71E-08	3.42E+05	2.98E-02
GSK3B	1.12E-07	1.07E+04	1.20E-03	6.04E-08	4.35E+04	2.63E-03	7.56E-08	6.16E+06	4.65E-01	1.06E-07	3.92E+04	4.16E-03
HRAS	5.98E-08	2.80E+04	1.68E-03	4.22E-08	4.44E+04	1.87E-03	8.22E-08	1.52E+09	1.25E+02	1.10E-07	2.59E+09	2.86E+02
IKBKA	1.38E-07	2.00E+04	2.75E-03	6.55E-08	2.01E+04	1.32E-03	7.24E-08	1.78E+07	1.29E+00	9.16E-08	7.05E+05	6.45E-02
IKBKB#							8.18E-08	4.56E+09	3.73E+02	5.65E-08	4.42E+04	2.50E-03
IKBKB##	4.53E-08	2.63E+04	1.19E-03	2.68E-08	4.16E+04	1.11E-03	8.30E-08	3.37E+07	2.80E+00	7.07E-08	4.03E+04	2.85E-03
IKBKG	9.38E-08	1.53E+04	1.43E-03	5.70E-08	2.73E+04	1.56E-03	7.82E-08	1.24E+09	9.66E+01	7.65E-08	2.36E+04	1.80E-03
INPP5D							5.74E-08	1.38E+06	7.93E-02			
INPPL1	6.69E-08	1.83E+04	1.22E-03	3.72E-08	3.05E+04	1.13E-03	7.56E-08	2.22E+06	1.67E-01	7.62E-08	2.44E+06	1.86E-01
Jun	4.12E-08	1.87E+04	7.68E-04	4.23E-08	2.38E+04	1.01E-03	7.48E-08	5.99E+07	4.48E+00	8.89E-08	8.74E+06	7.77E-01
KRAS	6.42E-08	2.18E+04	1.40E-03	4.53E-08	2.79E+04	1.26E-03	6.37E-08	4.77E+07	3.04E+00	6.20E-08	4.83E+06	3.00E-01
LAT2	1.40E-07	1.74E+04	2.43E-03	6.14E-08	2.30E+04	1.41E-03	1.42E-07	1.44E+09	2.04E+02	1.18E-07	4.00E+04	4.71E-03
LILRB3							8.63E-08	1.47E+07	1.27E+00	9.67E-08	4.01E+04	3.88E-03
LYN	1.34E-08	1.80E+04	2.41E-04	8.84E-08	9.33E+03	8.24E-04	1.06E-07	2.34E+09	2.48E+02	8.16E-08	2.19E+04	1.79E-03
MALT1							1.29E-07	4.98E+08	6.41E+01			
MAP2K1							8.78E-08	1.61E+07	1.41E+00	2.51E-06	2.10E+04	5.26E-02
MAP2K2	3.40E-08	1.28E+04	4.36E-04	2.03E-08	2.24E+04	4.56E-04	7.06E-08	2.94E+09	2.08E+02	1.05E-07	5.60E+04	5.89E-03
MAP2K3	1.15E-07	1.76E+04	2.03E-03	3.96E-08	2.25E+04	8.90E-04	9.62E-08	1.32E+07	1.27E+00	5.66E-08	2.21E+05	1.25E-02
MAP3K3	1.89E-07	1.40E+04	2.63E-03	8.91E-08	2.08E+04	1.85E-03	1.10E-07	3.79E+06	4.15E-01	1.02E-07	6.50E+04	6.66E-03
MAPK1	4.01E-08	2.70E+04	1.08E-03	1.83E-08	3.74E+04	6.85E-04	8.90E-08	2.65E+09	2.36E+02	1.15E-07	1.30E+05	1.49E-02
MAPK12	1.10E-07	1.63E+04	1.78E-03	5.73E-08	2.76E+04	1.58E-03	1.01E-07	2.70E+09	2.72E+02	1.07E-07	4.07E+05	4.35E-02
MAPK13	5.21E-08	2.45E+04	1.28E-03	3.77E-08	3.96E+04	1.49E-03	6.99E-08	9.21E+09	6.43E+02	1.45E-07	4.66E+05	6.76E-02
MAPK14	7.07E-08	1.51E+04	1.07E-03	5.09E-08	3.66E+04	1.86E-03	6.54E-08	2.77E+06	1.81E-01	7.31E-08	3.80E+04	2.78E-03
MAPK3	8.99E-08	8.95E+03	8.04E-04	4.63E-08	2.08E+04	9.64E-04	1.42E-07	1.66E+06	2.36E-01	8.50E-08	4.09E+04	3.48E-03
MAPK8							7.79E-08	2.06E+09	1.61E+02			
MAPK9				3.74E-08	3.02E+04	1.13E-03	8.21E-08	1.64E+09	1.35E+02	8.33E-08	1.68E+05	1.40E-02
MDM2	3.04E-07	5.08E+03	1.55E-03	4.90E-08	2.30E+04	1.12E-03	9.99E-08	4.11E+09	4.11E+02	1.59E-07	7.45E+07	1.19E+01
MYC	4.54E-08	2.20E+04	9.96E-04	3.78E-08	2.79E+04	1.05E-03	7.75E-08	2.50E+09	1.94E+02	8.43E-08	1.97E+08	1.66E+01
NCK1	8.74E-08	1.24E+04	1.09E-03	3.05E-08	3.96E+04	1.21E-03	7.39E-08	7.62E+07	5.63E+00	3.22E-08	5.62E+04	1.81E-03
NCKAP1L	4.37E-08	2.13E+04	9.31E-04	3.02E-08	3.40E+04	1.03E-03	6.40E-08	1.72E+09	1.10E+02	1.45E-07	9.30E+08	1.35E+02
NFAT5	4.76E-08	8.11E+03	3.86E-04	3.17E-08	1.88E+04	5.98E-04	6.55E-08	4.41E+05	2.89E-02			
NFATC1	9.79E-08	1.17E+04	1.15E-03	5.12E-08	3.26E+04	1.67E-03	1.16E-07	4.83E+06	5.59E-01	2.91E-07	5.81E+05	1.69E-01
NFATC3							1.36E-07	2.53E+04	3.44E-03			
NFATC4							9.85E-08	5.29E+06	5.20E-01	1.72E-07	3.04E+05	5.21E-02
NFKB1				4.62E-08	1.58E+04	7.30E-04	1.91E-07	8.01E+06	1.53E+00	1.29E-07	9.26E+04	1.19E-02
NFKBIA	5.51E-08	5.38E+04	2.96E-03	3.75E-08	1.26E+05	4.73E-03	9.19E-08	1.52E+09	1.40E+02	8.03E-08	4.50E+04	3.61E-03
NFKBIB	1.41E-07	8.80E+03	1.24E-03	1.24E-07	1.47E+04	1.82E-03	6.87E-08	5.18E+06	3.56E-01	1.42E-07	1.52E+04	2.16E-03
NFKBIE	2.09E-07	8.46E+03	1.77E-03				3.33E-07	7.94E+06	2.64E+00	1.14E-07	8.78E+04	1.00E-02
NRAS	1.89E-08	1.98E+04	3.73E-04	5.36E-12	2.50E+08	1.34E-03	1.04E-07	2.65E+07	2.75E+00	1.23E-07	9.22E+05	1.13E-01
PIK3AP1	4.01E-08	1.42E+04	5.70E-04	5.97E-08	2.24E+04	1.34E-03	1.66E-07	2.86E+08	4.74E+01	1.44E-07	2.23E+06	3.21E-01
PIK3CA	3.01E-08	2.72E+04	8.20E-04	3.31E-08	2.99E+04	9.91E-04	6.80E-08	1.00E+09	6.82E+01	5.61E-08	3.86E+06	2.16E-01
PIK3CB							1.26E-07	1.19E+09	1.49E+02			
PIK3CG	1.16E-07	9.89E+03	1.15E-03	4.25E-08	2.93E+04	1.25E-03	9.50E-08	6.93E+08	6.58E+01	7.30E-08	4.02E+05	2.93E-02
PIK3R1~	4.23E-08	1.97E+04	8.34E-04	3.16E-08	3.38E+04	1.07E-03	8.37E-08	1.03E+07	8.60E-01	7.17E-08	5.87E+04	4.21E-03
PIK3R2							3.68E-08	3.96E+05	1.46E-02			
PIK3R3	2.37E-07	1.09E+04	2.59E-03	4.58E-08	6.04E+04	2.77E-03	1.09E-07	7.19E+06	7.84E-01	1.24E-07	3.87E+04	4.81E-03
PIK3R5	1.66E-08	1.55E+04	2.58E-04	4.40E-09	3.02E+04	1.33E-04	4.85E-08	2.10E+06	1.02E-01			
PPP3CB				1.12E-08	2.26E+04	2.54E-04	1.34E-07	3.60E+04	4.83E-03	3.83E-08	5.44E+04	2.08E-03
PPP3R1	9.98E-08	1.24E+04	1.24E-03	3.26E-08	3.44E+04	1.12E-03	1.43E-07	1.51E+09	2.17E+02	1.05E-07	3.04E+04	3.18E-03
PRKCB	5.05E-08	1.24E+04	6.25E-04	1.03E-08	3.75E+04	3.86E-04	9.62E-08	4.08E+06	3.93E-01	1.01E-07	3.61E+04	3.66E-03
PTEN							2.06E-07	2.77E+04	5.72E-03	8.33E-08	4.05E+04	3.37E-03
PTPN6	8.27E-08	1.85E+04	1.53E-03	3.78E-08	3.30E+04	1.25E-03	1.39E-07	2.08E+07	2.89E+00	1.84E-07	2.62E+04	4.83E-03
RAC1				6.51E-09	1.87E+04	1.22E-04	5.16E-08	4.34E+06	2.24E-01	1.29E-07	6.38E+04	8.23E-03
RAC3	1.26E-07	1.32E+04	1.66E-03	4.07E-08	4.40E+04	1.79E-03	1.23E-07	2.55E+06	3.15E-01	1.03E-07	3.00E+04	3.10E-03
RAP1B	9.49E-08	1.98E+04	1.88E-03	1.40E-08	4.73E+04	6.60E-04	9.09E-08	3.65E+09	3.32E+02	9.65E-08	4.09E+05	3.95E-02
RAP2A	5.54E-08	1.90E+04	1.05E-03	3.39E-08	3.19E+04	1.08E-03	8.48E-08	4.04E+09	3.42E+02	1.71E-07	1.26E+09	2.15E+02
RAP2C	3.90E-08	2.61E+04	1.02E-03	7.08E-09	4.09E+04	2.89E-04	9.63E-08	3.31E+09	3.19E+02	6.19E-08	6.06E+06	3.75E-01
RasGRP3	2.77E-07	1.20E+04	3.33E-03	4.99E-08	1.96E+04	9.75E-04	9.42E-08	7.23E+07	6.81E+00	1.09E-07	2.91E+09	3.18E+02
RASSF5							8.59E-08	9.65E+07	8.29E+00			
RELA							7.54E-08	3.07E+09	2.32E+02	1.38E-07	1.65E+04	2.27E-03
RHOA	1.65E-07	1.15E+04	1.90E-03	4.55E-08	3.29E+04	1.50E-03	1.36E-07	1.84E+08	2.51E+01	1.00E-07	3.34E+04	3.35E-03
SOS1				6.14E-08	2.08E+04	1.28E-03	9.53E-08	2.44E+09	2.32E+02	1.09E-07	7.92E+04	8.67E-03
SOS2				1.19E-09	2.40E+04	2.86E-05	8.59E-08	9.42E+05	8.10E-02	1.88E-07	8.19E+04	1.54E-02
SYK	2.84E-08	1.65E+04	4.68E-04	2.32E-08	3.82E+04	8.88E-04	8.10E-08	4.29E+07	3.47E+00	1.77E-07	1.02E+09	1.81E+02
TEC	5.24E-08	1.45E+04	7.61E-04				7.82E-08	7.83E+08	6.13E+01	6.63E-08	1.60E+05	1.06E-02
TP53							8.01E-08	5.99E+08	4.80E+01	1.22E-07	1.76E+06	2.14E-01
VAV1	3.48E-07	7.90E+03	2.75E-03	5.55E-08	2.30E+04	1.28E-03	6.68E-08	2.72E+08	1.81E+01	9.71E-08	6.95E+05	6.74E-02
VAV2	2.57E-08	2.48E+04	6.38E-04	1.89E-08	2.47E+04	4.66E-04	9.07E-08	3.96E+09	3.59E+02	1.06E-07	3.13E+09	3.31E+02
VAV3							7.50E-08	4.00E+06	3.00E-01	1.27E-07	7.11E+09	9.06E+02

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

Table 42. RHOA query: K_D, k_a, and k_d data

	RHOA(GDP)_NP			RHOA(GDP)_LT			RHOA(GTP)_NP			RHOA(GTP)_LT		
Protein	KD	ka	kd	KD	ka	kd	KD	ka	kd	KD	ka	kd
AKT1	1.49E-07	6.90E+03	1.03E-03	9.29E-07	1.21E+03	1.12E-03	7.96E-08	1.22E+04	9.68E-04	1.14E-07	4.47E+03	5.11E-04
AKT2*	3.80E-07	5.93E+03	2.25E-03				9.45E-08	1.47E+04	1.39E-03	1.34E-07	7.99E+03	1.07E-03
AKT2**	4.61E-07	5.11E+03	2.36E-03				8.85E-08	1.12E+04	9.95E-04	3.43E-07	3.96E+03	1.36E-03
AKT3							4.03E-07	6.27E+03	2.52E-03	2.38E-06	8.40E+02	2.00E-03
ARHGEF7							1.26E-07	1.73E+04	2.18E-03	2.90E-07	7.09E+03	2.06E-03
BCL10							2.73E-07	5.31E+03	1.45E-03			
BCL2							1.10E-07	1.56E+04	1.72E-03			
BCL2A1	3.77E-07	3.84E+03	1.45E-03	3.97E-07	2.35E+03	9.35E-04	1.25E-07	1.37E+04	1.72E-03	2.94E-07	4.36E+03	1.28E-03
BCL2L1	4.06E-07	5.23E+03	2.12E-03				1.64E-07	1.32E+04	2.17E-03	4.33E-07	3.72E+03	1.61E-03
BLK							1.80E-07	1.23E+04	2.22E-03	2.66E-07	6.63E+03	1.76E-03
BLNK							2.22E-07	1.10E+04	2.45E-03	7.48E-07	1.79E+03	1.34E-03
BTK							2.48E-07	1.08E+04	2.67E-03	8.12E-07	1.12E+03	9.13E-04
CARD11	3.38E-07	7.46E+03	2.52E-03				1.14E-07	1.28E+04	1.46E-03	2.49E-07	4.31E+03	1.07E-03
CD19							2.86E-07	7.78E+03	2.22E-03	3.70E-07	4.90E+03	1.81E-03
CD22							1.90E-07	1.51E+04	2.88E-03	6.28E-07	2.64E+03	1.66E-03
CD72	4.34E-07	4.70E+03	2.04E-03				7.24E-08	1.33E+04	9.59E-04	3.32E-07	2.89E+03	9.60E-04
CD79B							1.02E-07	1.59E+04	1.61E-03			
CD81							1.45E-07	1.27E+04	1.84E-03			
CDC42							9.69E-08	1.18E+04	1.14E-03			
CDKN2A							4.04E-07	3.02E+03	1.22E-03			
DAPP1							1.42E-07	1.93E+04	2.75E-03			
EGR1							1.20E-07	8.86E+03	1.06E-03			
ETS1	6.49E-08	1.44E+04	9.34E-04				1.48E-07	1.11E+04	1.65E-03	4.96E-07	2.26E+03	1.12E-03
EZR							2.09E-07	1.03E+04	2.15E-03	2.17E-06	5.03E+02	1.09E-03
FCGR2B							7.78E-08	1.50E+04	1.16E-03	1.62E-06	1.28E+03	2.08E-03
Fos							2.01E-07	1.01E+04	2.02E-03	1.22E-06	1.49E+03	1.81E-03
GRAP2										4.54E-07	4.68E+03	2.12E-03
GRB2	3.95E-07	8.15E+03	3.21E-03				1.26E-07	1.28E+04	1.61E-03	5.28E-09	1.10E+04	5.78E-05
GSK3B							1.16E-07	1.86E+04	2.17E-03			
HRAS							3.74E-07	6.16E+03	2.30E-03			
IFITM1							2.34E-07	1.02E+04	2.38E-03	4.40E-07	1.67E+03	7.36E-04
IKBKA							2.67E-07	7.83E+03	2.09E-03	5.72E-09	4.87E+03	2.78E-05
IKBKB#	4.54E-07	7.67E+03	3.49E-03				1.07E-07	1.67E+04	1.79E-03	2.81E-07	6.81E+03	1.91E-03
IKBKB##	3.75E-07	4.82E+03	1.81E-03				1.04E-07	2.09E+04	2.18E-03	5.10E-07	2.63E+03	1.34E-03
IKBKG	4.97E-07	6.62E+03	3.29E-03				1.29E-07	1.26E+04	1.63E-03	3.60E-07	3.11E+03	1.12E-03
INPP5D	3.82E-07	8.23E+03	3.15E-03				2.96E-07	8.05E+03	2.38E-03	4.84E-08	4.20E+03	2.03E-04
INPPL1							2.77E-07	1.01E+04	2.80E-03	4.76E-06	4.20E+02	2.00E-03
Jun							1.54E-07	1.25E+04	1.93E-03	1.19E-07	2.51E+03	2.99E-04
KRAS							1.44E-07	1.56E+04	2.24E-03			
LAT2	1.14E-07	7.09E+03	8.10E-04				6.21E-10	6.41E+03	3.98E-06	8.70E-07	1.79E+03	1.55E-03
LILRB3	4.72E-07	6.61E+03	3.12E-03				1.07E-07	1.25E+04	1.33E-03	9.95E-07	3.01E+03	3.00E-03
LYN							2.92E-07	8.33E+03	2.43E-03			
MAP2K2	3.69E-07	6.99E+03	2.58E-03				2.57E-07	9.27E+03	2.38E-03	1.35E-07	5.65E+03	7.62E-04
MAP2K3	5.29E-07	4.79E+03	2.54E-03				1.13E-07	1.50E+04	1.69E-03	6.36E-07	1.96E+03	1.25E-03
MAP3K3	7.13E-07	4.77E+03	3.40E-03				4.01E-07	3.60E+03	1.44E-03	3.12E-07	4.50E+03	1.40E-03
MAPK1	1.42E-07	6.35E+03	9.02E-04				1.25E-07	1.50E+04	1.87E-03	9.92E-08	4.75E+03	4.71E-04
MAPK12							3.21E-07	7.50E+03	2.40E-03			
MAPK13	8.02E-08	1.55E+04	1.24E-03	5.18E-08	2.07E+04	1.07E-03	1.85E-07	7.71E+03	1.42E-03	2.14E-07	5.21E+03	1.12E-03
MAPK14	1.70E-07	1.14E+04	1.94E-03	2.52E-07	4.01E+03	1.01E-03	1.04E-07	1.59E+04	1.65E-03	2.85E-07	3.87E+03	1.11E-03
MAPK3							1.27E-07	1.23E+04	1.56E-03	4.47E-07	3.83E+03	1.71E-03
MAPK8							8.91E-07	1.65E+03	1.47E-03	1.32E-08	6.10E+03	8.07E-05
MAPK9							3.13E-07	5.30E+03	1.66E-03	2.84E-08	5.31E+03	1.51E-04
MDM2	2.37E-07	6.92E+03	1.64E-03				1.06E-07	1.35E+04	1.43E-03	2.38E-07	6.51E+03	1.55E-03
MYC	3.39E-07	5.72E+03	1.94E-03				1.70E-07	1.20E+04	2.04E-03	2.28E-07	3.37E+03	7.68E-04
NCK1							1.63E-07	9.76E+03	1.59E-03	1.69E-07	5.72E+03	9.67E-04
NCKAP1L	2.59E-07	7.91E+03	2.05E-03				1.00E-07	1.34E+04	1.34E-03	1.68E-07	5.41E+03	9.11E-04
NFAT5										4.07E-07	2.54E+03	1.04E-03
NFATC1							1.39E-07	1.19E+04	1.66E-03	1.68E-06	1.05E+03	1.77E-03
NFATC3	3.31E-07	5.70E+03	1.89E-03									
NFATC4	3.20E-07	4.71E+03	1.51E-03				1.55E-09	6.04E+03	9.39E-06			
NFKB1	3.16E-07	6.37E+03	2.01E-03									
NFKBIA							2.20E-07	8.23E+03	1.81E-03			
NFKBIE							6.69E-08	1.63E+04	1.09E-03	2.62E-07	3.64E+03	9.56E-04
NRAS							2.03E-07	7.59E+03	1.54E-03			
PIK3AP1							8.88E-07	1.36E+03	1.20E-03			
PIK3CA	2.36E-07	1.17E+04	2.77E-03				1.14E-07	8.92E+03	1.01E-03	2.93E-08	2.79E+03	8.17E-05
PIK3CB	3.58E-07	7.25E+03	2.59E-03				1.16E-07	1.28E+04	1.48E-03	5.53E-09	6.13E+03	3.39E-05
PIK3CG							3.50E-07	7.67E+03	2.69E-03	1.45E-07	1.17E+04	1.70E-03
PIK3R1~	3.25E-07	1.14E+04	3.71E-03				9.55E-08	1.61E+04	1.53E-03	3.77E-07	3.32E+03	1.25E-03
PIK3R3										8.18E-08	3.74E+03	3.05E-04
PIK3R5							1.51E-07	1.32E+04	1.99E-03	9.99E-07	2.26E+03	2.26E-03
PLCG2							2.22E-07	3.43E+03	7.60E-04	4.42E-06	3.88E+02	1.72E-03
PPP3CA							1.37E-07	2.02E+04	2.78E-03	4.87E-09	7.34E+03	3.58E-05
PPP3CB	2.88E-07	1.01E+04	2.89E-03				1.24E-07	1.16E+04	1.44E-03	8.07E-06	2.08E+02	1.68E-03
PPP3R1	2.41E-07	7.50E+03	1.81E-03				2.28E-07	7.53E+03	1.71E-03	1.72E-07	3.38E+03	5.80E-04
PPP3R2							1.65E-07	1.08E+04	1.77E-03	8.22E-08	4.83E+03	3.97E-04
PRKCA							1.47E-06	1.98E+03	2.90E-03	9.62E-08	3.78E+03	3.64E-04
PRKCB							1.56E-07	8.49E+03	1.33E-03	3.80E-07	3.04E+03	1.16E-03
PTEN							8.05E-07	3.26E+03	2.63E-03			
PTPN6	3.34E-07	5.93E+03	1.98E-03				3.98E-07	3.16E+03	1.26E-03			
RAC1										5.08E-07	3.59E+03	1.82E-03
Raf1							2.32E-07	6.83E+03	1.59E-03			
RAP1B							5.56E-07	3.57E+03	1.98E-03	6.57E-06	2.31E+02	1.52E-03
RAP2A							1.65E-07	8.60E+03	1.42E-03	2.45E-07	3.40E+03	8.34E-04
RAP2C	3.28E-07	7.05E+03	2.32E-03	4.87E-07	5.50E+03	2.68E-03	1.12E-07	1.45E+04	1.63E-03	2.25E-06	1.07E+03	2.41E-03
RasGRP3							1.34E-07	1.09E+04	1.47E-03			
RASSF5							6.26E-07	2.40E+03	1.50E-03	2.01E-07	5.44E+03	1.10E-03
RELA							1.85E-07	1.13E+04	2.09E-03	2.70E-07	2.97E+03	8.00E-04
RHOA	1.63E-07	8.31E+03	1.35E-03				5.17E-08	2.13E+04	1.10E-03	4.11E-07	3.37E+03	1.39E-03
SOS1							1.77E-07	1.12E+04	1.99E-03			
SOS2							1.03E-07	1.21E+04	1.26E-03	2.69E-07	3.15E+03	8.47E-04
SYK							1.03E-07	1.39E+04	1.42E-03	3.09E-07	4.67E+03	1.44E-03
TEC							1.85E-07	1.05E+04	1.95E-03	8.00E-07	3.76E+03	3.01E-03
TP53							9.92E-07	2.30E+03	2.28E-03	1.14E-06	1.61E+03	1.84E-03
VAV1							2.65E-07	7.91E+03	2.10E-03	4.93E-06	4.80E+02	2.37E-03
VAV2							2.28E-07	6.96E+03	1.59E-03	3.67E-07	2.86E+03	1.05E-03
VAV3							6.95E-09	3.20E+04	2.23E-04			

Reference Sequence ID = *BC063421, **BC120994 , #BC006231, ##BC108694, ~BC030815, ~~BC094795

NP = target proteins are Not Phosphorylated. LT = target proteins are Lysate-Treated.

APPENDIX M

NAPPA-SPRI: VENN DIAGRAMS OF NP- AND LT-TARGET INTERACTIONS

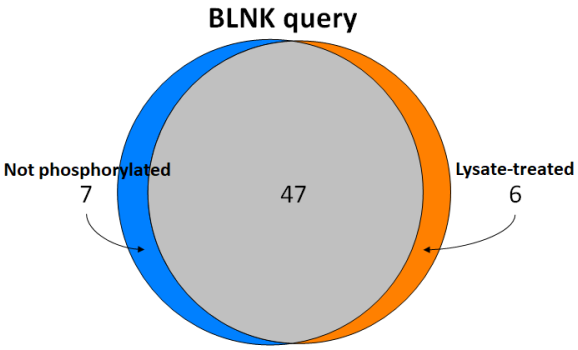


Figure 111. Venn diagram comparing the PPIs between BLNK and NP- and LT-targets

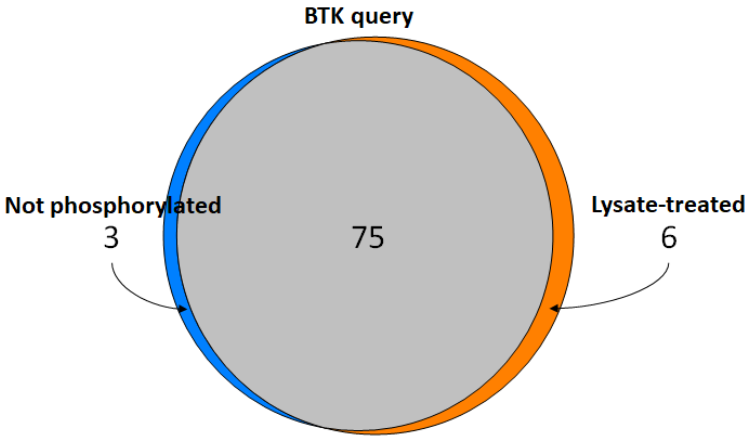


Figure 112. Venn diagram comparing the PPIs between BTK and NP- and LT-targets

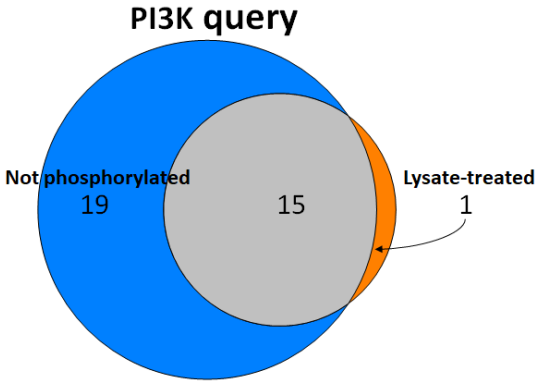


Figure 113. Venn diagram comparing the PPIs between PI3K and NP- and LT-targets

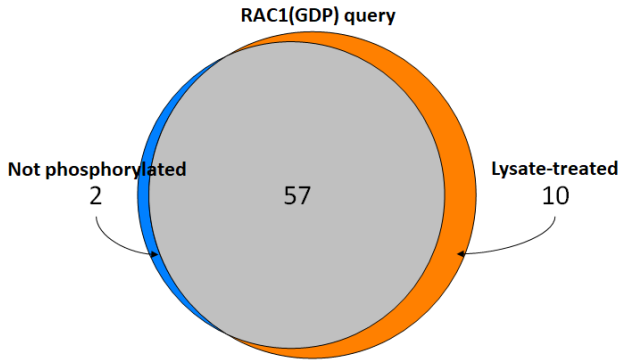


Figure 114. Venn diagram comparing the PPIs between GDP-bound RAC1 and NP- and LT-targets

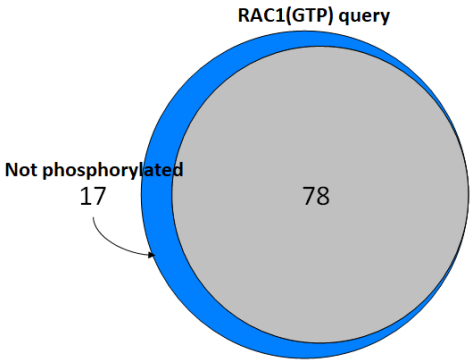


Figure 115. Venn diagram comparing the PPIs between GTP-bound RAC1 and NP- and LT-targets

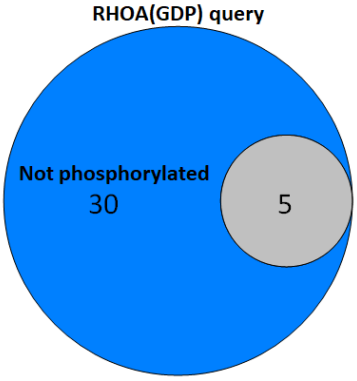


Figure 116. Venn diagram comparing the PPIs between GDP-bound RHOA and NP- and LT-targets

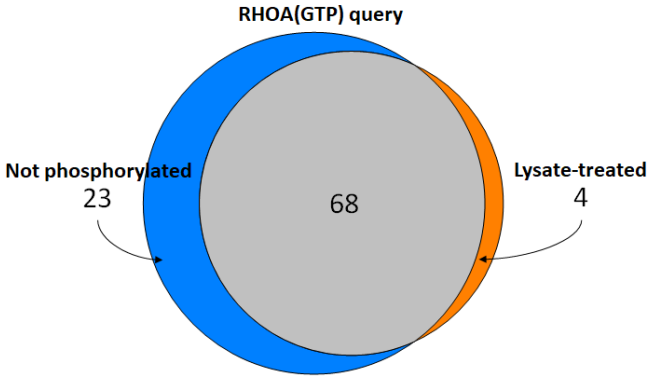


Figure 117. Venn diagram comparing the PPIs between GTP-bound RHOA and NP- and LT-targets

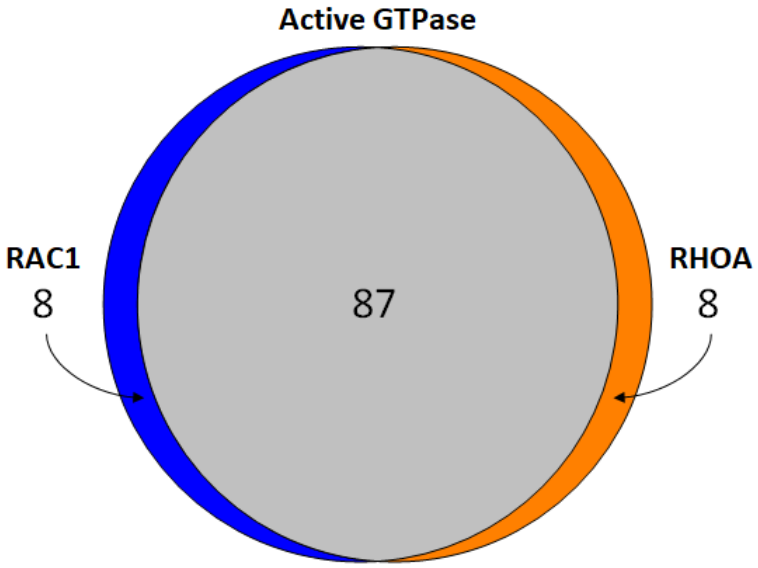


Figure 118. Venn diagram comparing the PPIs with active RAC1 and RHOA

APPENDIX N

NAPPA-SPRI: STANDARDIZED RESIDUAL PLOTS

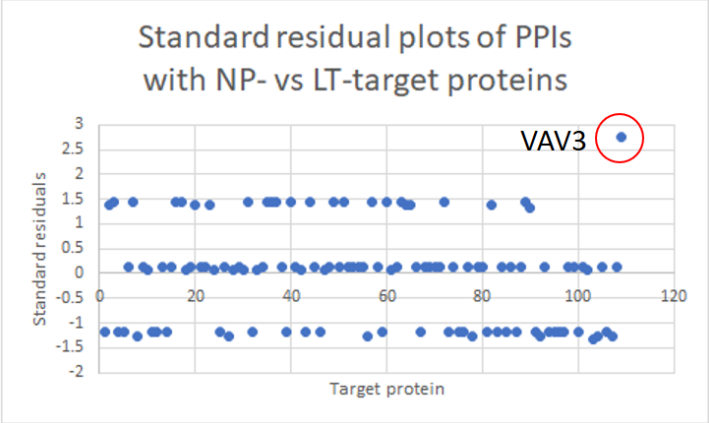


Figure 119. Residual plot comparing the binding of NP- and LT-targets. Queries preferentially bound to NP-VAV3 compared to LT-VAV3.

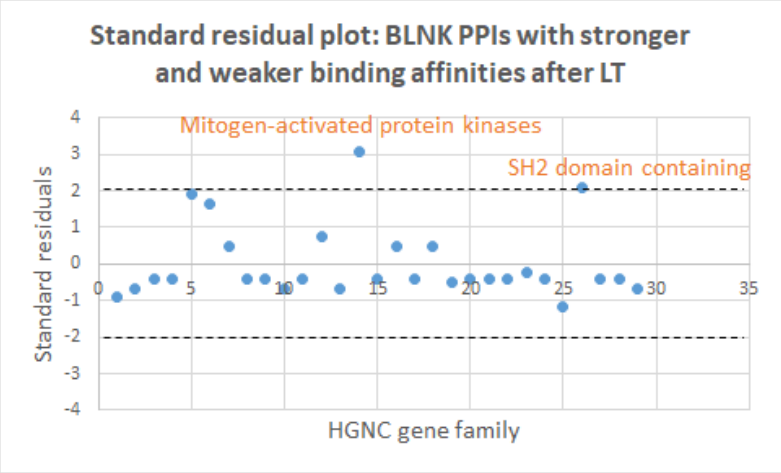


Figure 120. Residual plot comparing the HGNC gene families that interacted with BLNK with stronger and weaker binding affinities following lysate treatment.

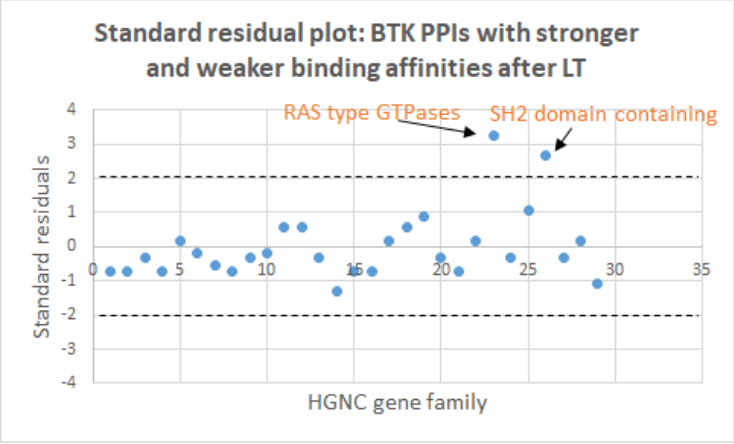


Figure 121. Residual plot comparing the HGNC gene families that interacted with BTK with stronger and weaker binding affinities following lysate treatment.

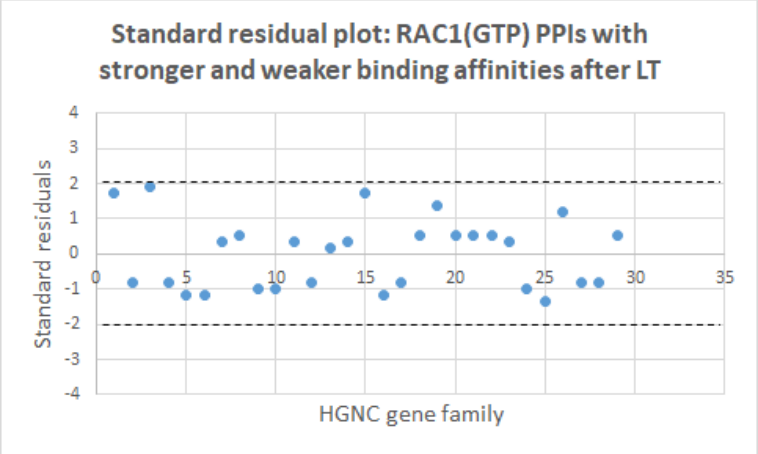


Figure 122. Residual plot comparing the HGNC gene families that interacted with RAC1(GTP) with stronger and weaker binding affinities following lysate treatment.

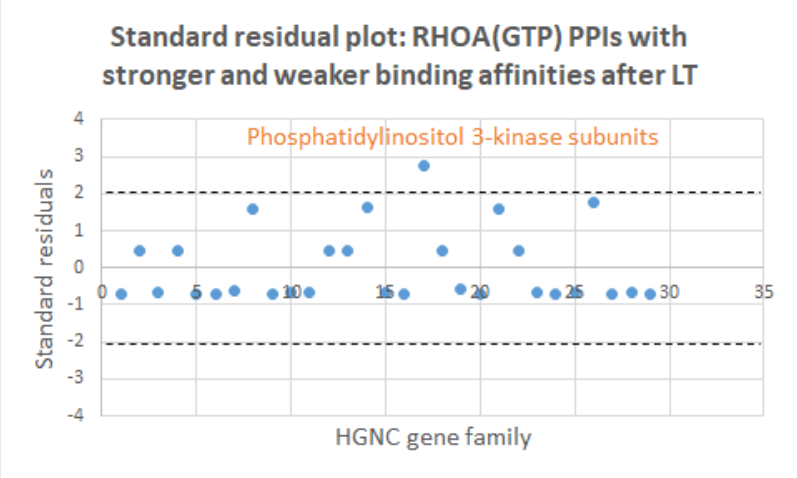


Figure 123. Residual plot comparing the HGNC gene families that interacted with RHOA(GTP) with stronger and weaker binding affinities following lysate treatment.

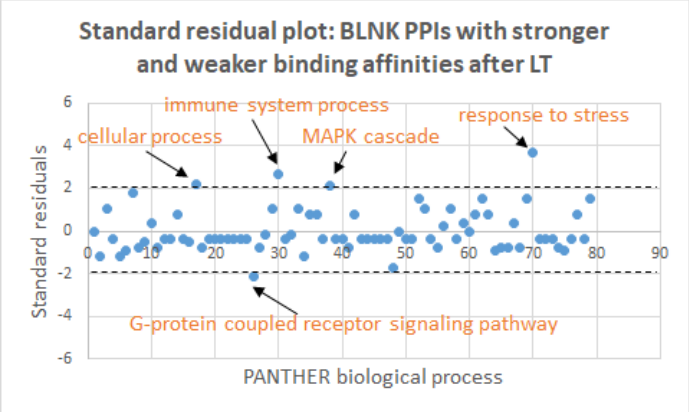


Figure 124. Residual plot comparing the PANTHER biological processes that interacted with BLNK with stronger and weaker binding affinities following lysate treatment.

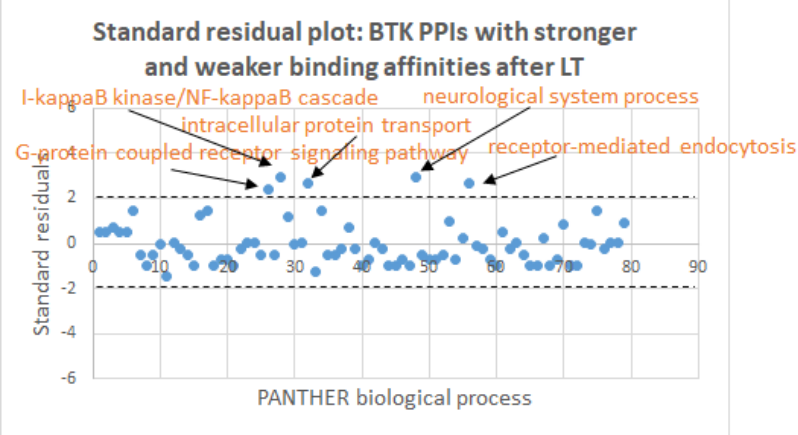


Figure 125. Residual plot comparing the PANTHER biological processes that interacted with BTK with stronger and weaker binding affinities following lysate treatment.

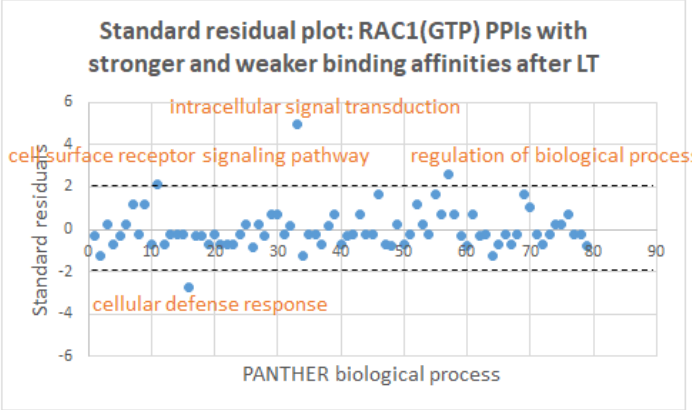


Figure 126. Residual plot comparing the PANTHER biological processes that interacted with RAC1(GTP) with stronger and weaker binding affinities following lysate treatment.

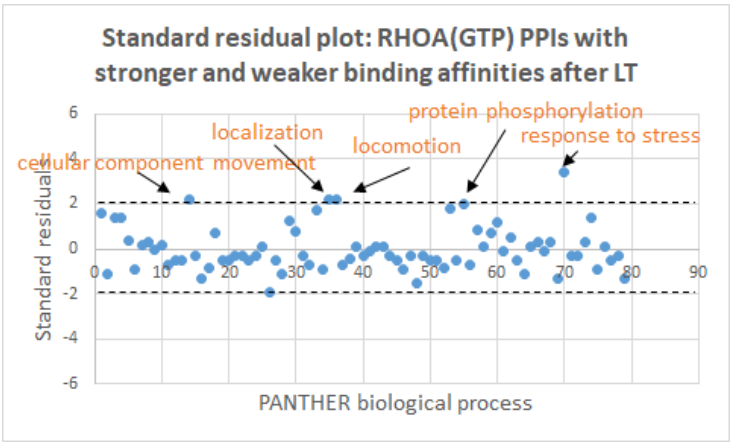


Figure 127. Residual plot comparing the PANTHER biological processes that interacted with RHOA(GTP) with stronger and weaker binding affinities following lysate treatment.

APPENDIX O

VENN DIAGRAMS OF NAPPA-SPRI AND NANOBRET INTERACTIONS

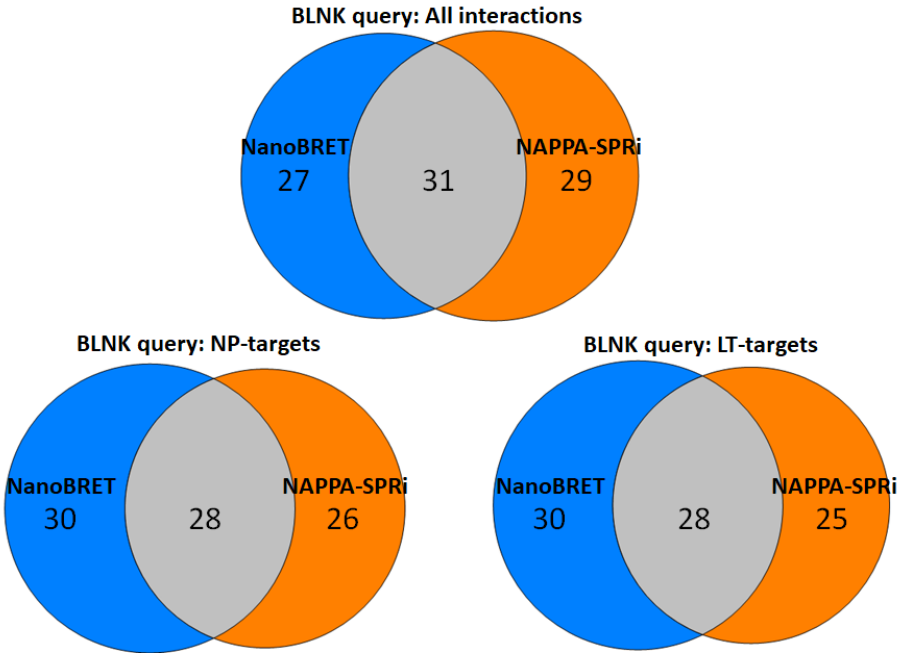


Figure 128. Venn diagram comparing BLNK's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

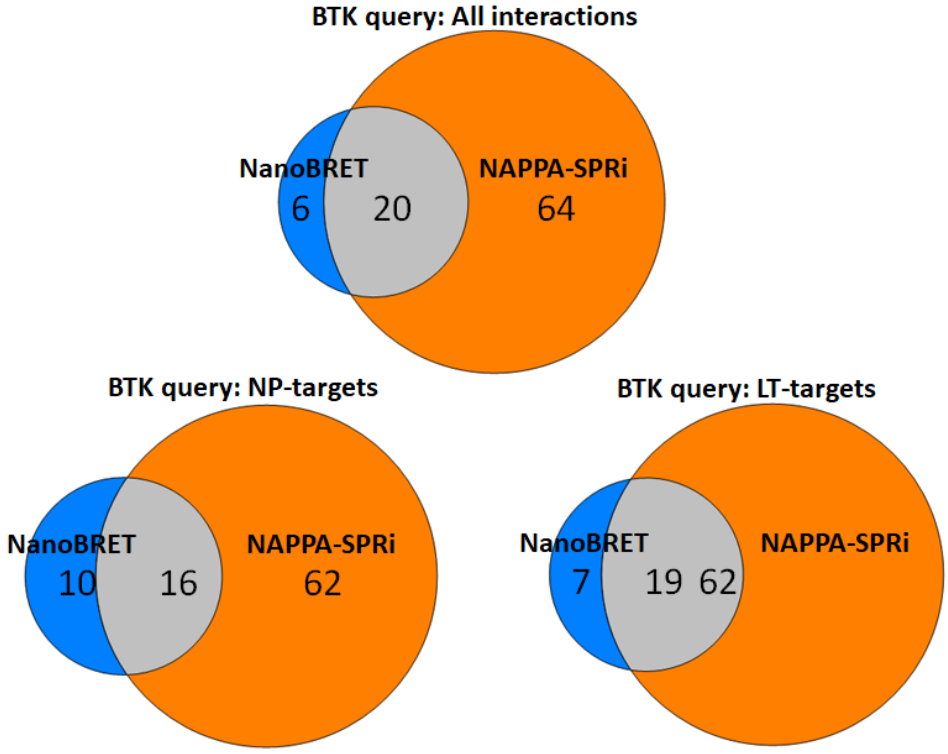


Figure 129. Venn diagram comparing BTK's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

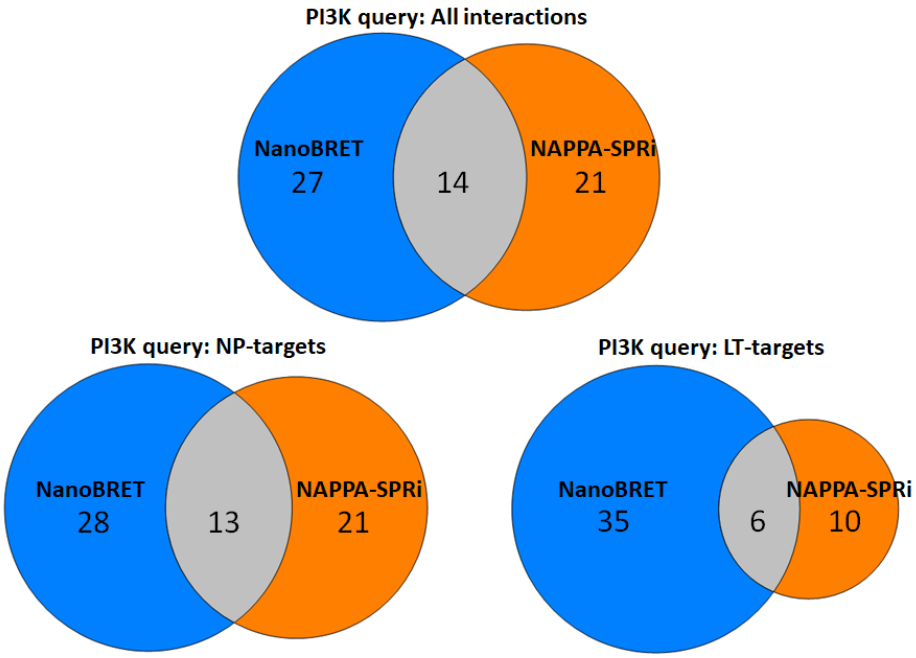


Figure 130. Venn diagram comparing PI3K's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

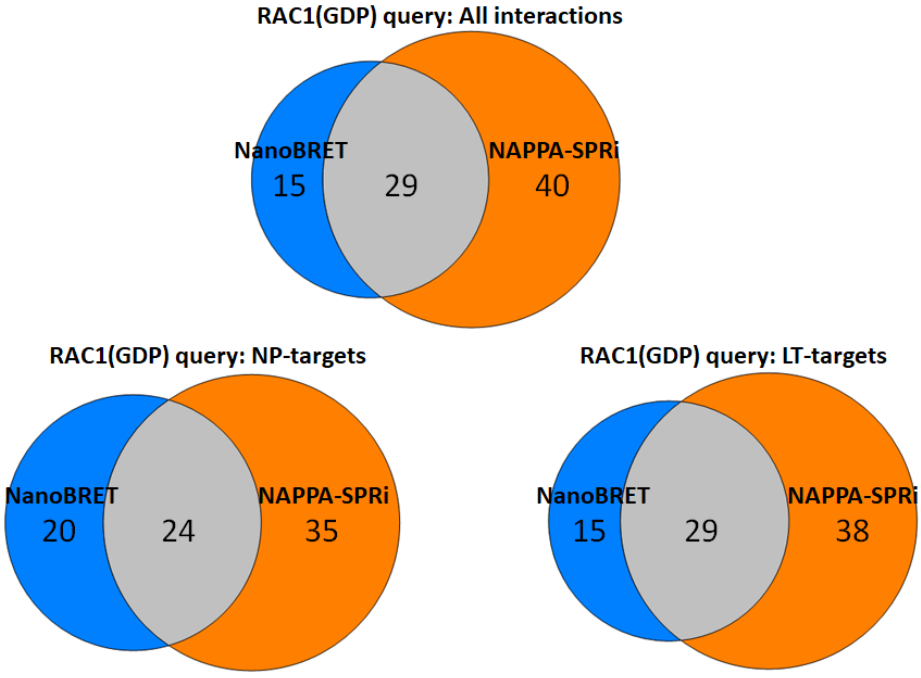


Figure 131. Venn diagram comparing GDP-bound RAC1's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

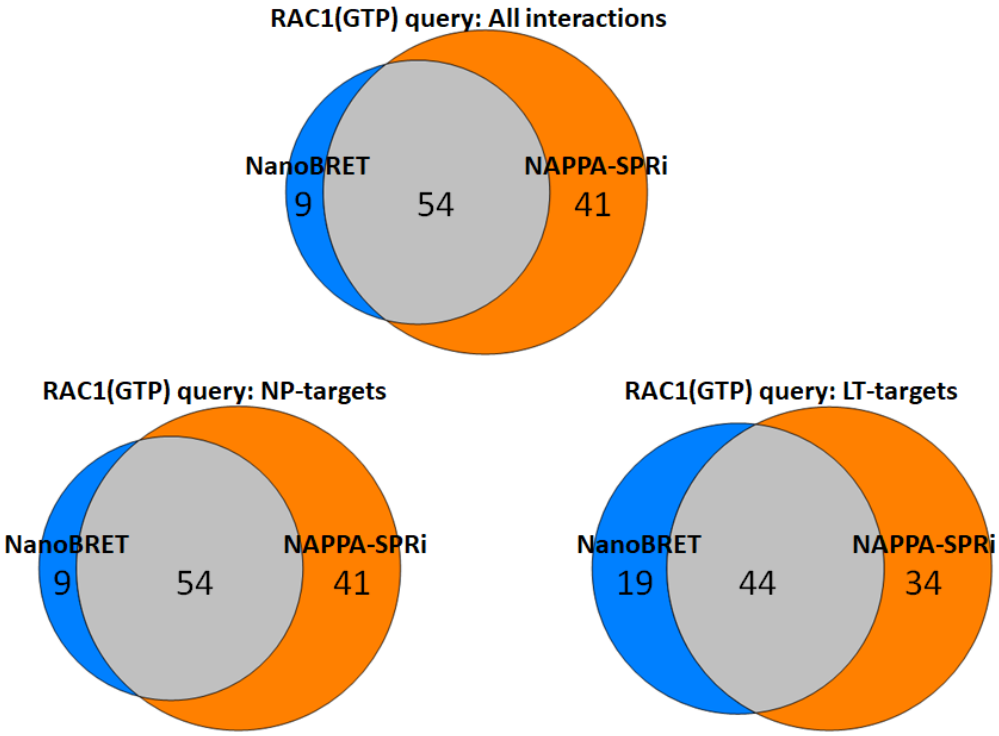


Figure 132. Venn diagram comparing GTP-bound RAC1's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

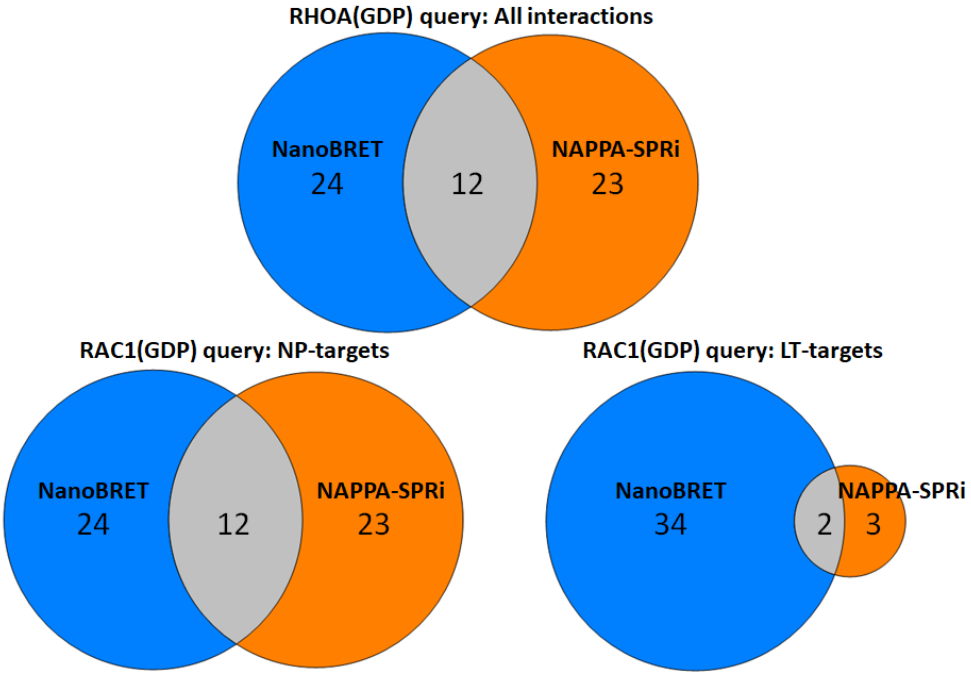


Figure 133. Venn diagram comparing GDP-bound RHOA's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

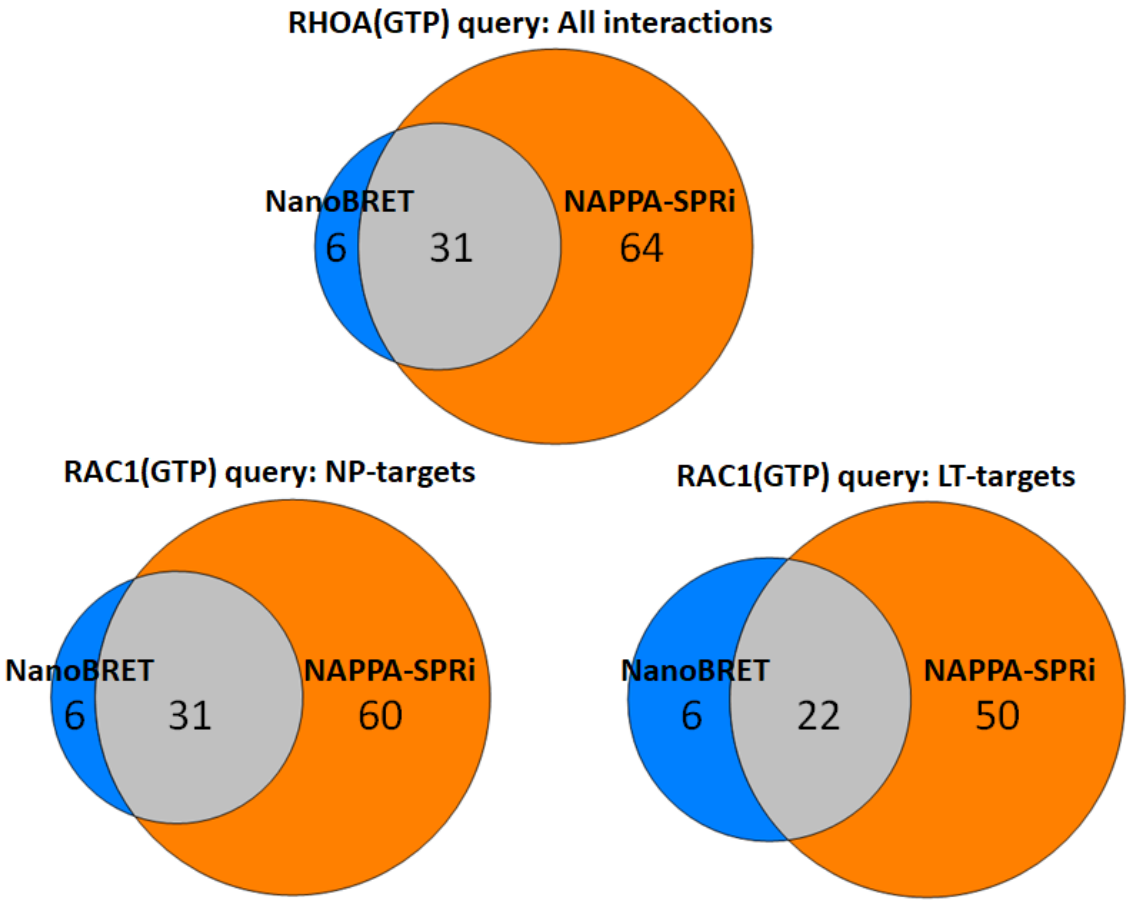


Figure 134. Venn diagram comparing GTP-bound RHOA's PPIs identified similarly and uniquely with the NanoBRET and NAPPA-SPRi platform

APPENDIX P

BAR PLOTS OF RELATIVE BINDING KINETICS

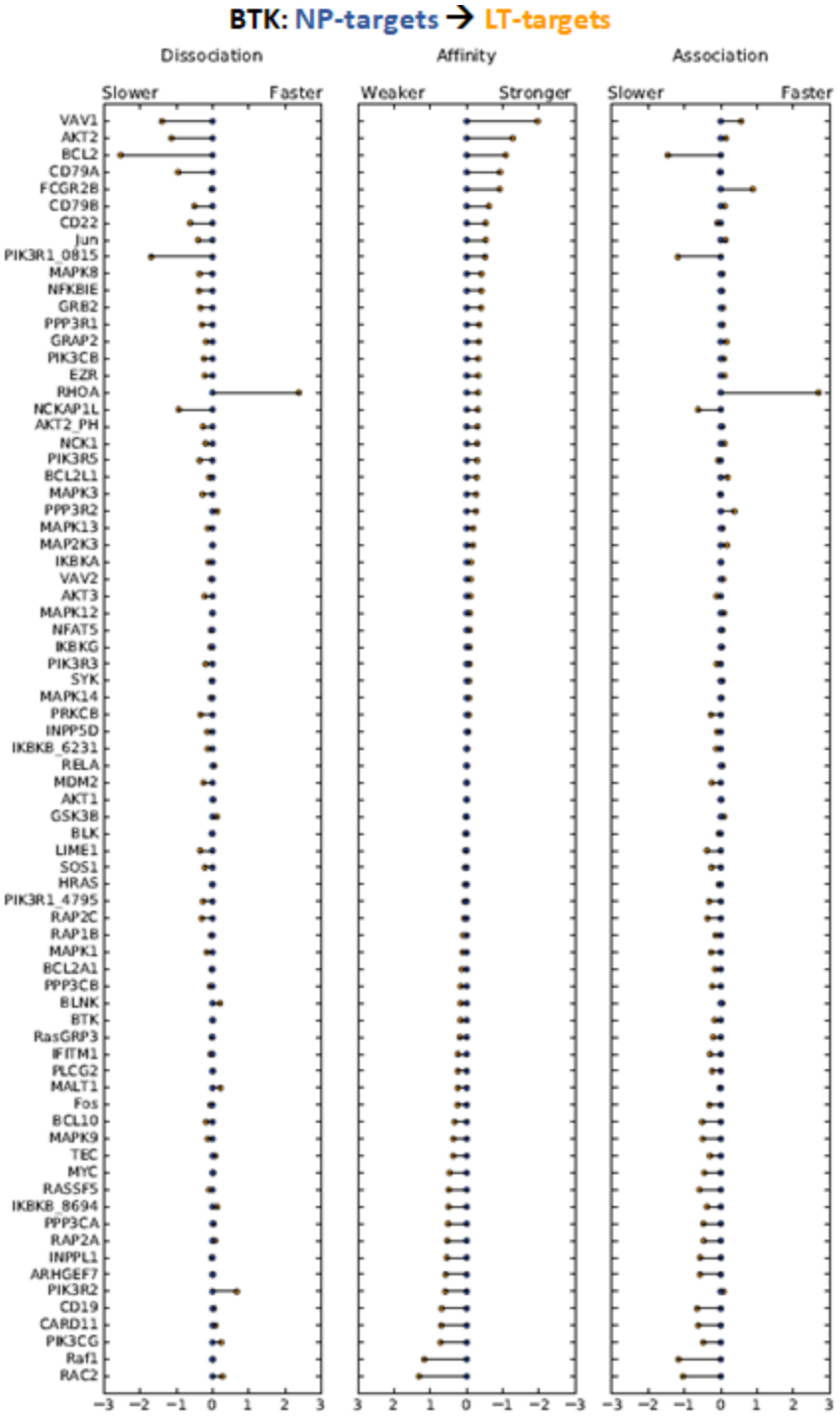


Figure 135. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs between NP- and LT-targets and the BTK query.

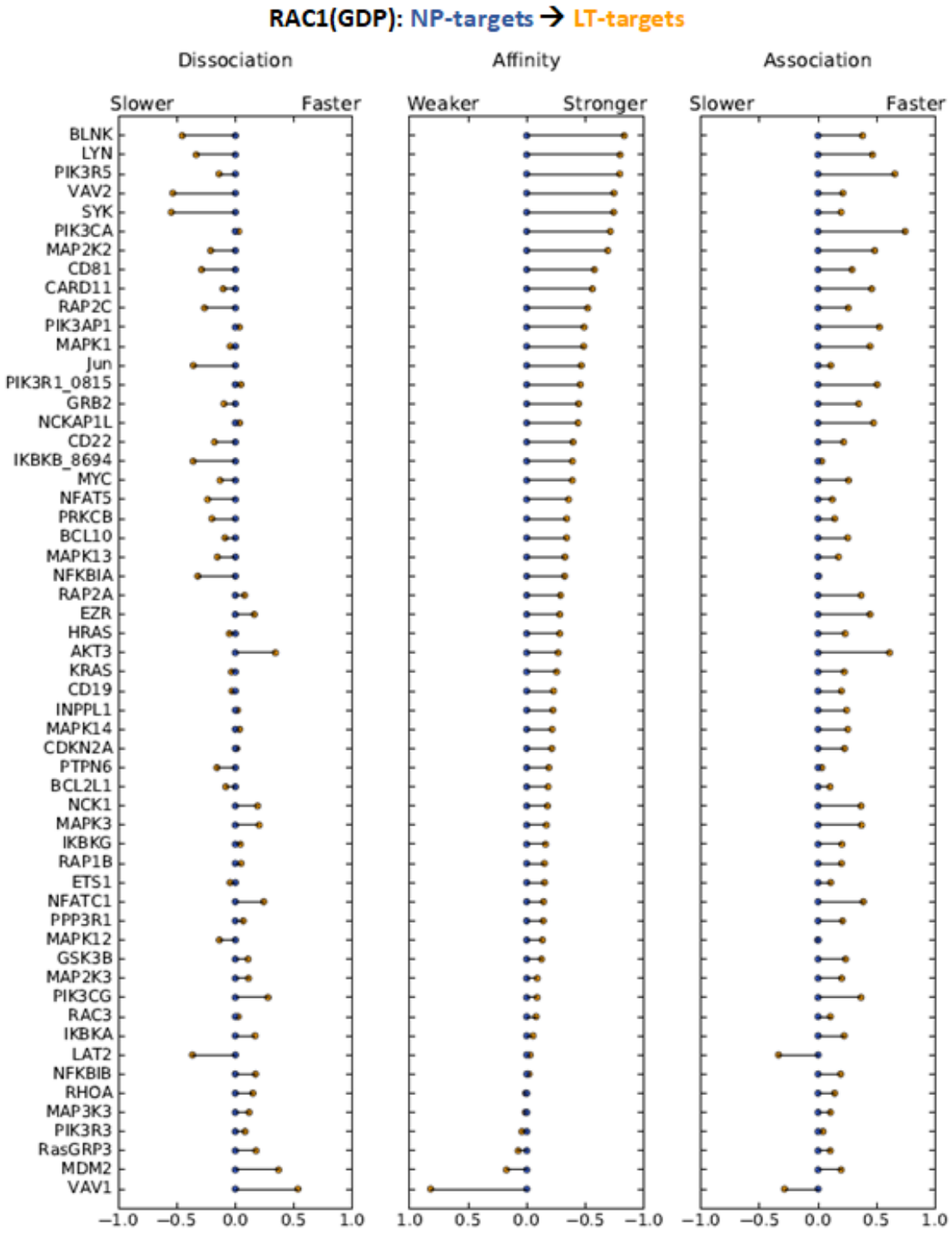


Figure 136. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs between NP- and LT-targets and the RAC1(GDP) query.

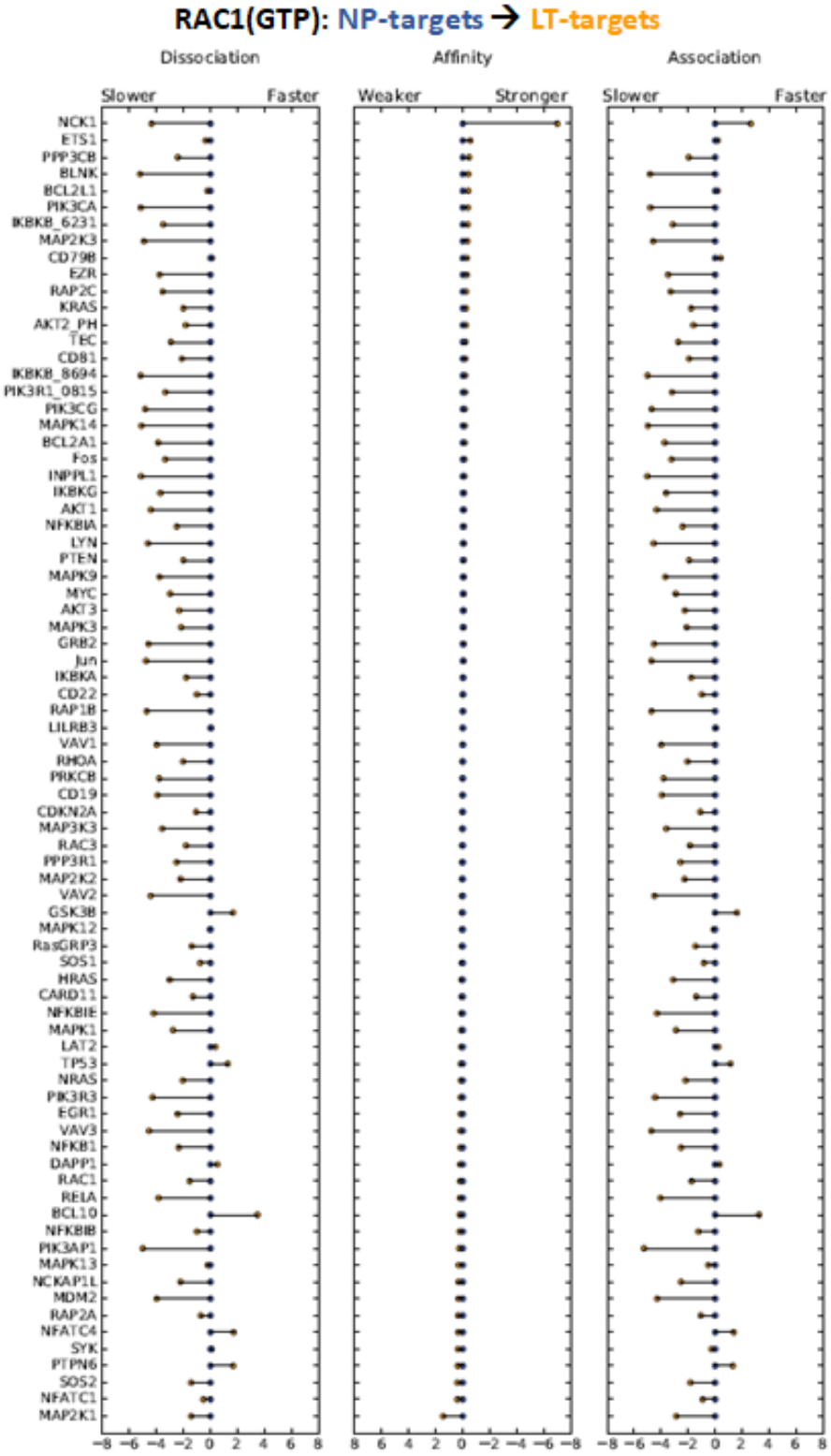


Figure 137. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs between NP- and LT-targets and the RAC1(GTP) query.

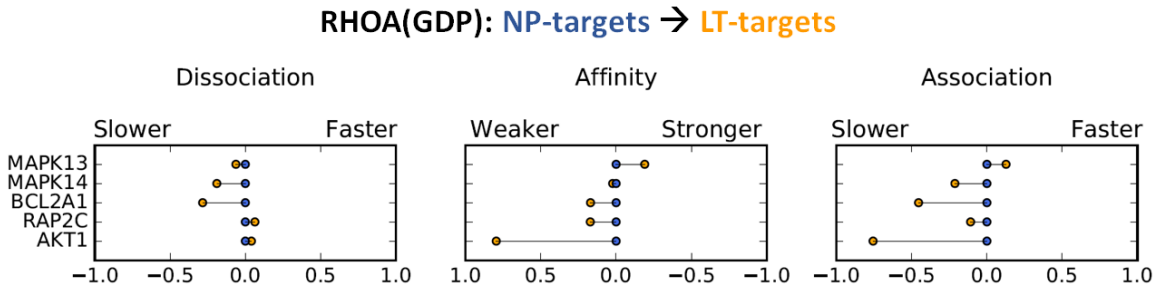


Figure 138. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs between NP- and LT-targets and the RHOA(GDP) query.

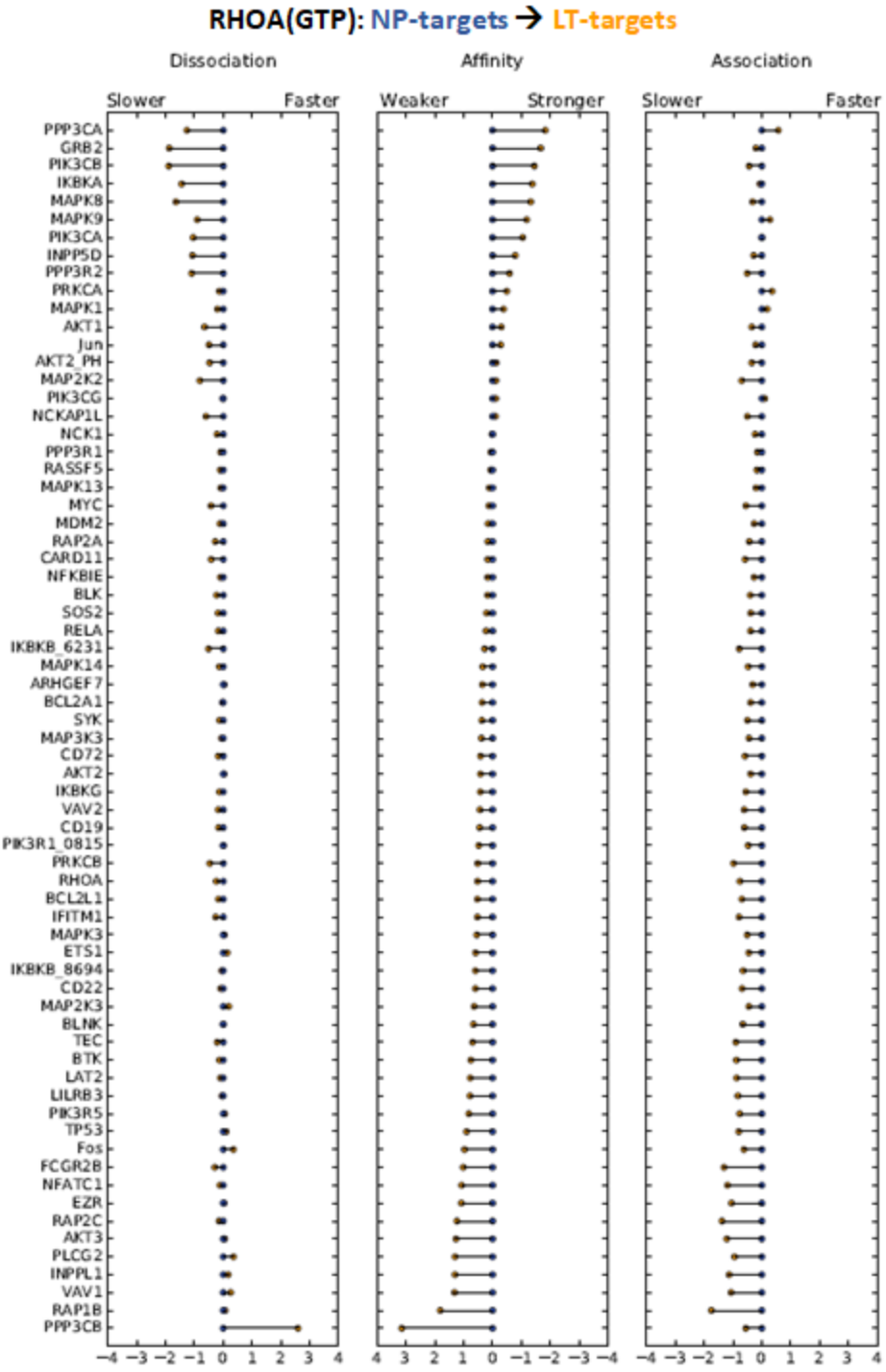


Figure 139. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs between NP- and LT-targets and the RHOA(GTP) query.

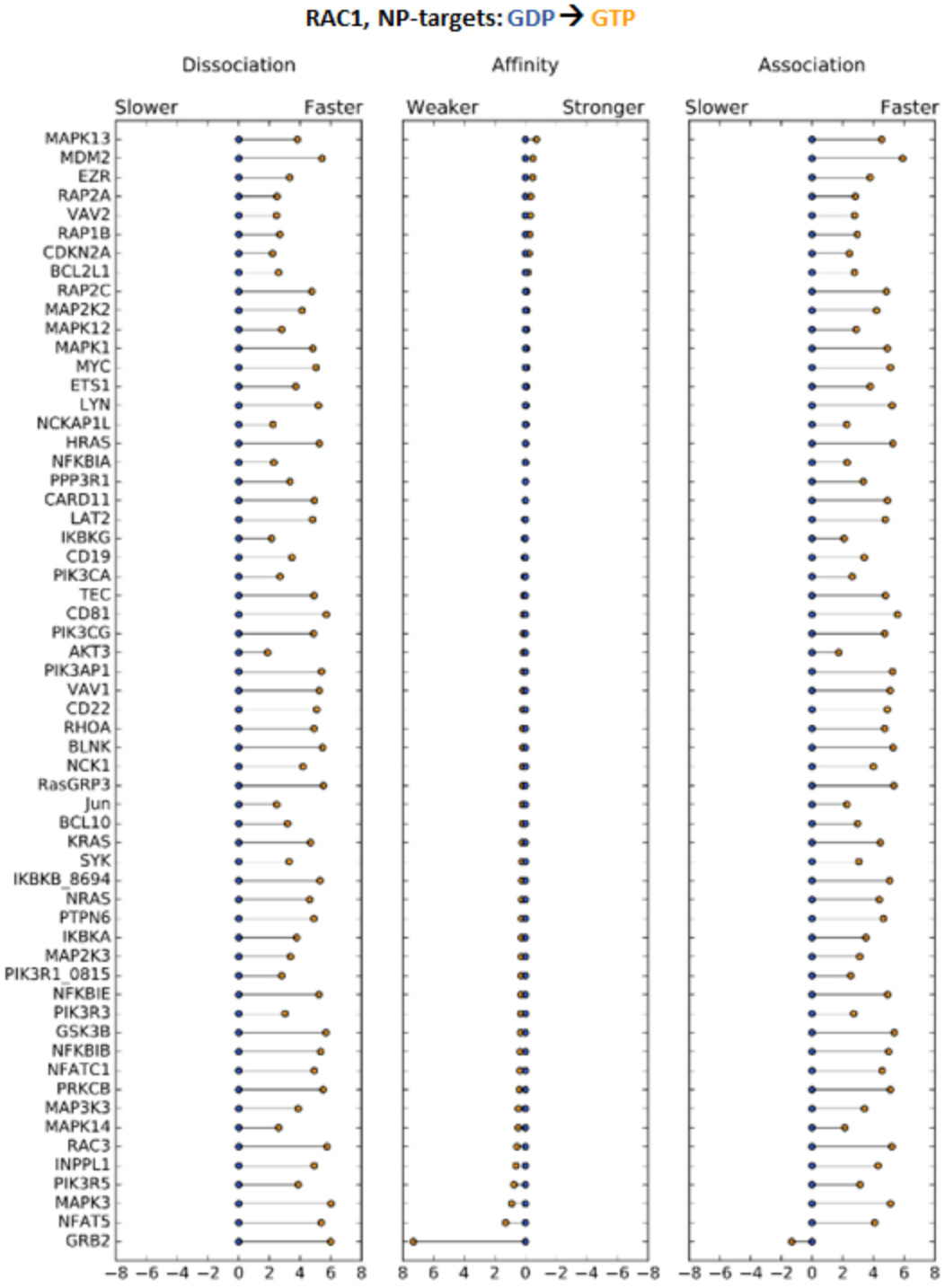


Figure 140. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs with inactive and active RAC1 to NP-targets.

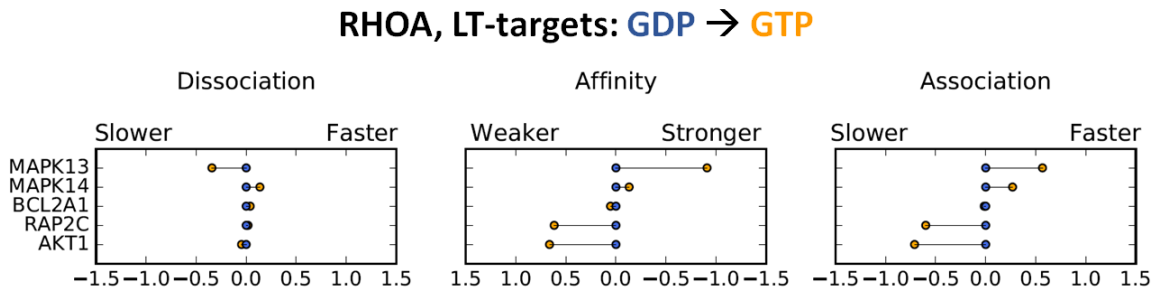


Figure 141. Bar plots showing the relative log10 change in kd, KD, and ka of all PPIs with inactive and active RHOA to NP-targets.

APPENDIX Q

NAPPA-SPRI BINDING SENSORGRAMS

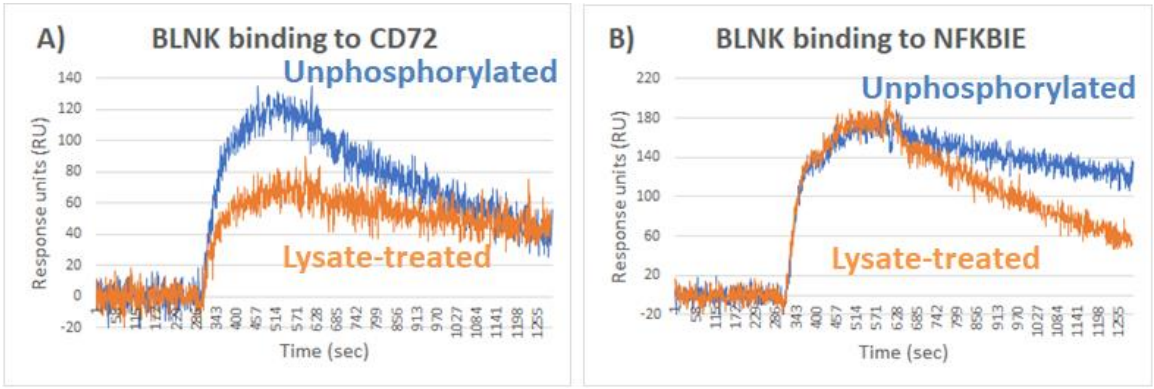


Figure 142. BLNK's interactions are largely regulated by their off-rates

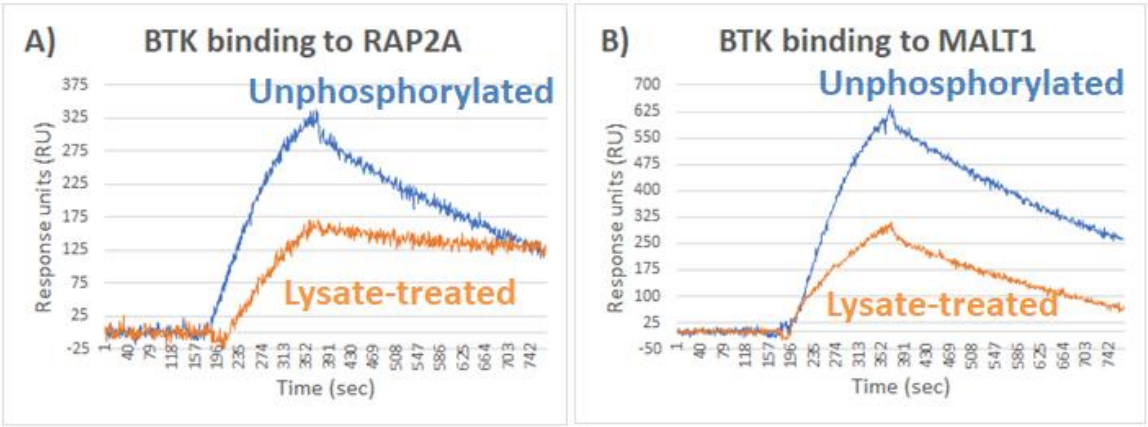


Figure 143. BTK's interactions are largely regulated by their A) off-rates or B) on-rates.

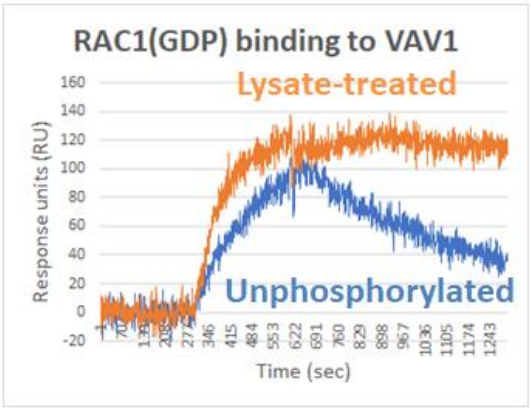


Figure 144. Inactive RAC1's interactions are largely regulated by their on- AND off-rates.

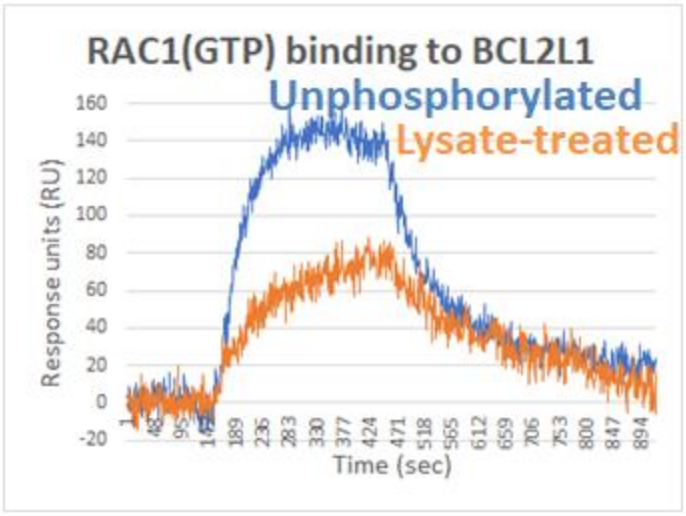


Figure 145. Active RAC1's interactions are largely regulated by their on- AND off-rates.

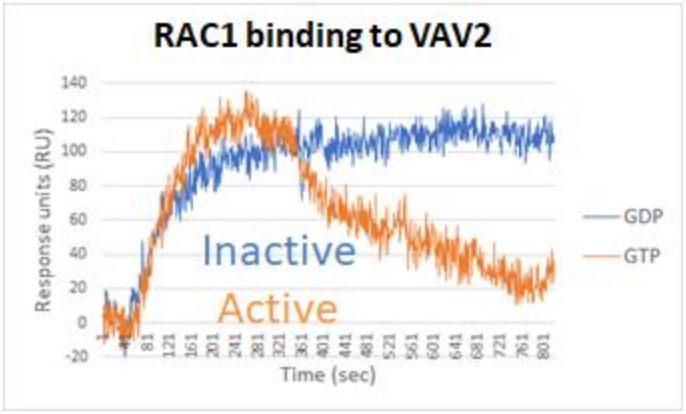


Figure 146. RAC1 activation increases its on- and off-rates with little change in binding affinities with LT-targets

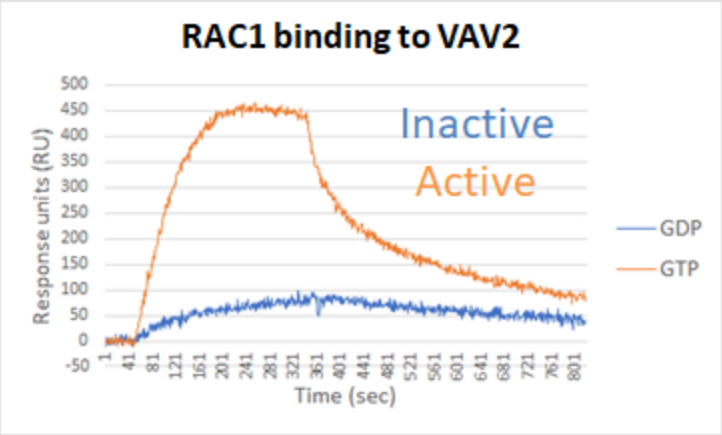


Figure 147. RAC1 activation significantly increases its on- and off-rates with little change in binding affinities with NP-targets

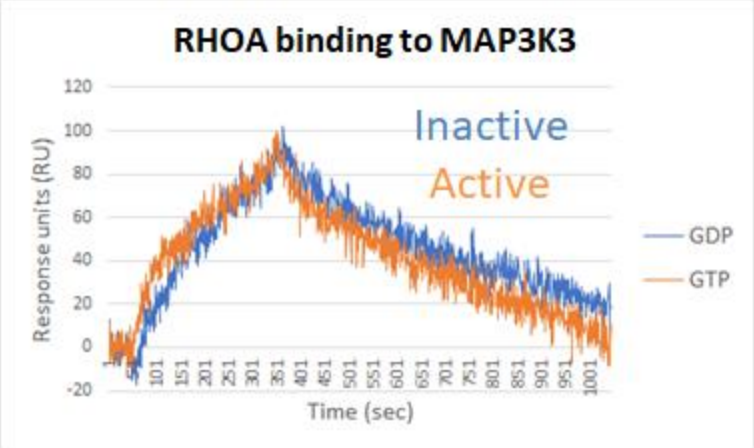


Figure 148. RAC1 activation minimally affects binding kinetics and affinities