Proteomic and Chemoproteomic Strategies to Interrogate Post-translational Modifications

Aaron Josef Maurais

A dissertation

submitted to the Faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

October 2021

Proteomic and Chemoproteomic Strategies to Interrogate Post-translational Modifications

Aaron Josef Maurais

Advisor: Eranthie Weerapana, Ph.D.

Abstract

Protein activity is modulated by hundreds of post-translational modifications (PTMs). This thesis will describe the development and application of proteomic methods to study three chemically distinct PTMs. In the first project we describe the development of a proteomics platform to identify cysteine oxidation sites on interactors of the NADPH oxidase complex in response to EGF activation. The NADPH oxidases (Nox) are the source of H_2O_2 which acts as a secondary messenger during EGFR activation. Known targets of Nox H_2O_2 include phosphatases PTP1B and PTEN. Oxidation of the active site of PTP1B and PTEN temporarily inactivates their phosphatase activity which allows for EGF signal propagation. The platform involves combining TurboID with OxICAT to identify proteins which are oxidized by Nox2 in a spatially and temporally controlled manner. In the second project, our goal is to identify proteins which recognize two Met oxidation sites in actin known to play a role in regulating the transition between F and G actin. We utilized a peptide based photo-crosslinking approach to identify PFKL and HSP70s HSPA8 and HSPA1B as putative "readers" of oxidized or unoxidized methionine in actin respectively. Finally, protein citrullination is a enzyme catalyzed PTM where the guanidinium on arginine is converted into a urea by a family of enzymes called protein arginine deiminases. Aberrant citrullination is linked to many human diseases including rheumatoid arthritis. Therefore, proteomic methods to characterize citrullination can provide insights into disease pathophysiology. We describe the identification of novel protein targets of with a chemoselective biotin phenyl glyoxal probe, and the development of a label

free proteomic method to identify sites of citrullination.

For my parents,

Gerald and Elli

Acknowledgments

First of all, I would like to thank my advisor, Eranthie Weerapana. Your thoughtful and patient guidance have helped me become a better scientist. I am especially grateful that you gave me a lot of freedom to pursue areas of research that interested me. In the last year of my undergraduate degree, I wasn't even sure if I wanted to attend graduate school. When you visited Merrimack and gave a seminar on your research, you inspired me to apply to BC so I could work in your lab. I am very glad that I made that decision. I would also like to thank my committee members Jianmin Gao and Abhishek Chatterjee. Thank you for all the help and guidance over the years.

I would also like to thank my collaborators in Professor Paul Thompson's lab at Umass Medical School. Ronak, Li, and Son, I am glad that we were able to work together on so many on so many fruitful projects. Thank you also to Scott Shaffer at the proteomics facility at Umass Medical School for your assistance in putting the data together for the ionFinder project.

To all the past and current of the Weerapana lab: Lisa, Kyle, Masahiro, Tyler, Dan, Jenny, Julia, Eleni, Ben, Qianni, loud Sarah, Emma, quiet Sarah, and Angela. I would especially like to thank Dan for being a consistent source of help and support. You always go above and beyond when I ask for your help troubleshooting an experiment and for that I am very grateful. Ari, I am glad that we were able to spend so much time working together. Even though we have very different dispositions and personalities, I thoroughly enjoyed our scientific and not so scientific discussions. I would also like to thank you for indulging me in debates about which government agencies should be abolished.

Finally, I would like to thank family for all your love and support. Mama and Papa, thank you for instilling in me the importance of doing well in school, even

though I wasn't always the best student. I could never have accomplished what I have without your guidance and patience. Paul, Rita, and Carrie, you are the best younger siblings anyone could ask for. I always enjoy the time we spend together.

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

4.2 Identification of readers of reduced and oxidized methionine. **(A)** Workflow used to identify proteins which interact selectively with oxidized or reduced methionine. 50 µM probes **1** or **2** are crosslinked in isotopically heavy or light MCF7 lysates. After UV crosslinking, equal amounts of heavy and light lysate are combined. Proteins which are bound to the probe are enriched on streptavidin agarose beads, trypsinized, and subjected to LC-MS/MS analysis. **(B)** 2D plot showing SILAC $R_{L/H}$ for each protein in the "forward" (x-axis) and "reversed" (y-axis) experiments. Proteins which selectively interact preferentially with probe **1** are in the upper right portion of the plot and preferential interactors of probe **2** are in the bottom left portion.

4.3 Detection of endogenous PFKL and HSPA8 in MCF7 lysates enriched with probe **1** or **2**. MCF7 lysates were crosslinked with 75 µM of the indicated probe. An aliquot of the labeled lysate was taken and run as the "input" fraction. The labeled proteins were pulled down on streptavidin-agarose beads. The bound proteins were eluted by heating in gel loading dye and 50 µM biotin to obtain the "enriched" fraction. Each fraction was analyzed by western blot antibodies for HSPA8 and PFKL and streptavidin-HRP to detect probe labeled proteins.

4.4 PFKL activity assay. **(A)** Coupled enzyme assay used to measure PFKL activity. **(B)** Michaelis–Menten plot of recombinant PFKL activity at varying concentrations of fructose 6-phosphate (F6P).

4.5 Crosslinking with recombinant PFKL and HSPA8. 12 µM of the indicated probe was combined with 1 µM purified **(A)** HSPA8 and **(B)** PFKL *in vitro*. Each lane is a biological replicate performed with a new preparation of protein. **(C)** Mean densitometry measurements of 3 biological replicates shown in **A** and **B**. Band intensities from the streptavidin blot were normalized to the intensity of the HA or PFKL band and then normalized intensity of the most intense replicate, such that the intensity of the most intense replicate is set to 1. Replicates were statistically analyzed by Student's t test ($^*p < 0.005$). Error bars represent the standard deviation of the 3 measurements.

4.6 Mical1 NADPH consumption in the presence of probe **1**. The consumption of NADPH by the catalytic domain of human Mical1 was monitored by the A340 in the presence of 0 to 100 µM probe **1**, as well as in the presence of 0 and 100 µM probe **1** absent Mical1.

Appendix I

List of Tables

List of Equations

Chapter 2

Chapter 4

Chapter 5

List of Schemes

Chapter 2

List of Abbreviations

Standard 3-letter and 1-letter codes are used for the 20 natural amino acids.

ABPP	activity-based protein profiling.
ACN	acetonitrile.
ACPA	anti-citrullinated protein antibody.
Akt	protein kinase B.
ALDO	aldolase.
AMP	adenosine monophosphate.
ATP	adenosine triphosphate.
AUC	area under the curve.
BSA	bovine serum albumin.
BTD	benzothiazine dioxide.
CaM	calmodulin.
CaMKII	Ca^{2+}/CaM-dependent protein kinase II.
cBio	photocaged biotin.
CBK	caged bromomethyl ketone.
CDI	carbonyldiimidazole.
cDNA	complementary DNA.
CH	calponin homology.
CID	collision-induced dissociation.
CIK4	caged iodomethyl ketone.

Cit	citrulline.
CuAAC	copper-catalyzed azide-alkyne cycloaddition.
DAB	denaturing alkylation buffer.
DAPI	4',6-diamidino-2-phenylindole.
DBIA	desthiobiotin iodoacetamide.
DCM	dichloromethane.
DDA	data dependent acquisition.
DMAP	4-dimethylaminopyridine.
DMF	dimethylformamide.
DMSO	dimethyl sulfoxide.
EBX	ethynyl benziodoxolone.
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide.
EGF	epidermal growth factor.
EGFR	epidermal growth factor receptor.
EIC	extracted ion chromatogram.
ESI^+	positive electrospray ionization.
F1,6BP	fructose-1,6-diphosphate.
F6P	fructose 6-phosphate.
FBS	fetal bovine serum.
FMO	flavin-containing monooxygenase.
FPR	false positive rate.
fRMsr	free Met-R-SO reductase.
GAPD	glyceraldehyde 3-phosphate.
GEF	guanine exchange factor.
GPCR	G-protein-coupled receptor.
GPDH	glycerol-3-phosphate dehydrogenase.
GPx	glutathione peroxidase.

Grb2	growth receptor bound protein.
HCD	high-energy collisional dissociation.
HIS	heat inactivated serum.
HOBt	hydroxybenzotriazole.
HRP	horseradish peroxidase.
IA	iodoacetamide.
IAA	iodoacetamide-alkyne.
IAH	IA-alkyne heavy.
IAL	IA-alkyne light.
iBAQ	intensity-based absolute quantification.
ICAT	isotope-coded affinity tags.
IMAC	immobilized metal affinity chromatography.
IMPDH2	inosine-5′-monophosphate dehydratase 2.
isoTOP-ABPP	isotopic tandem orthogonal proteolysis - activity-based protein profiling.
ITGB4	integrin beta-4.
iTRAQ	isobaric tags for relative and absolute quantization.
KDE	kernel density estimate.
LFQ	label free quantification.
MAPK	Ras/mitogen activated kinase.
MCS	multiple cloning site.
MeOH	methanol.
MOI	multiplicity of infection.
MS	mass-spectrometry.
Msr	methionine sulfoxide reductase.
MsrA	methionine sulfoxide reductase A.
MsrB	methionine sulfoxide reductase B.

MTRP	multiplexed thiol reactivity profiling.
NADH	nicotinamide adenine dinucleotide (reduced).
NADPH	nicotinamide adenine dinucleotide phosphate.
NEM	N-ethylmaleimide.
NET	neutrophil extracellular trap.
NHS	N-hydroxysuccinimide.
NL	neutral loss.
*o*NB	ortho-nitrobenzyl.
OxICAT	oxidative isotope-coded affinity tags.
PAD	protein arginine deiminase.
PBS	phosphate buffered saline.
PEG	polyethylene glycol.
PFK	phosphofructokinase.
PG	phenyl glyoxal.
PI3K	phosphoinositide 3-kinase.
PIP3	phosphatidylinositol 3,4,5-trisphosphate.
PLCγI	phospholipase C, gamma 1.
PPI	protein-protein interaction.
Prx	peroxiredoxin.
PSA	penicillin-streptomycin-amphotericin B.
PSM	peptide-spectrum match.
PTEN	phosphate and tension homolog.
PTM	post-translational modification.
PTP	protein tyrosine phosphatase.
RA	rheumatoid arthritis.
ReDiMe	reductive dimethylation.
$R_{L/H}$	light to heavy ratio.

RNS	reactive nitrogen species.
ROC	receiver operating curve.
ROS	reactive oxygen species.
RT	retention time.
RTK	receptor tyrosine kinase.
SDS	sodium dodecyl sulfate.
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis.
Sec	selenocysteine.
SECIS	selenocysteine insertion sequence).
SERPIN	serine protease inhibitor.
SF	synovial fluid.
SH2	src homology 2.
SILAC	stable isotope labeling by amino acids in cell culture.
SLC-ABPP	streamlined cysteine activity-based protein profiling.
SOD	superoxide dismutase.
SRM	selected reaction monitoring.
Srx	sulfiredoxin.
TBS-T	Tris buffered saline with tween 20.
TCA	trichloroacetic acid.
TCEP	tris(2-carboxyethyl)phosphine.
TEAB	triethylammonium bicarbonate.
TFA	trifluoroacetic acid.
THF	tetrahydrofuran.
TIS	triisopropylsilane.
TKD	tyrosine kinase domain.
TMT	tandem mass tags.
TNR	true negative rate.

TPI	triosephosphate isomerase.
TPR	true positive rate.
UniProt	universal protein resource.

Chapter 1

Cysteine Oxidation in Cell Signaling

A portion of this chapter has been published in:

Maurais, A. J. & Weerapana, E. Reactive-cysteine profiling for drug discovery. *Curr Opin Chem Biol* **2019**, *50*, 29–36.

1.1 Introduction to reactive oxygen species and cysteine

Reactive oxygen species (ROS) are produced as a byproduct of normal cellular processes in all organisms which undergo aerobic respiration. Although high levels of ROS can lead to oxidative stress, and contribute to DNA, protein and lipid damage, at low levels it plays a crucial role in cell signaling.[2] The primary ROS species are superoxide (O_2^-), hydrogen peroxide (H_2O_2), and the hydroxyl radical ($^\bullet OH$), each of which have distinct characteristics and half lives. The majority or ROS in eukaryotes is produced as a byproduct of mitochondrial respiration. Components of the electron chain reduce oxygen (O_2) to water during adenosine triphosphate (ATP) synthesis, however partial reduction of O_2 can result in the production of O_2^-.[3] The other major source of ROS are the nicotinamide adenine dinucleotide phosphate (NADPH) oxidases (Nox) which oxidize NADPH and generate O_2^-.[4] In both cases, O_2^- is rapidly converted to H_2O_2 by superoxide dismutases (SODs) 1, 2 and 3 (Figure 1.1). The overproduction of H_2O_2 can result in the formation of hydroxyl radicals via Fenton chemistry catalyzed by divalent metal cations, such as Fe^{2+}. Hydroxyl radicals are highly reactive and indiscriminately oxidize proteins, lipids, and DNA. The production of hydroxyl radicals is generally associated exclusively with oxidative stress.

1.1.1 Role of cysteine in oxidative signaling

Of the 20 proteogenic amino acids, cysteine is unique in its elevated nucleophilicity and redox sensitivity. Despite its low abundance, cysteine is highly conserved at functionally important sites.[6,7] The high nucleophilicity and redox sensitivity of the cysteine thiolate facilitates key roles in several aspects of protein function:[8] (1) active-site nucleophiles in catalysis, or resolving residues in cellular redox buffering systems;[9] (2) protein structure stabilization through disulfide bonds, and metal coordination; and, (3) regulation of protein function through post-translational modifica-

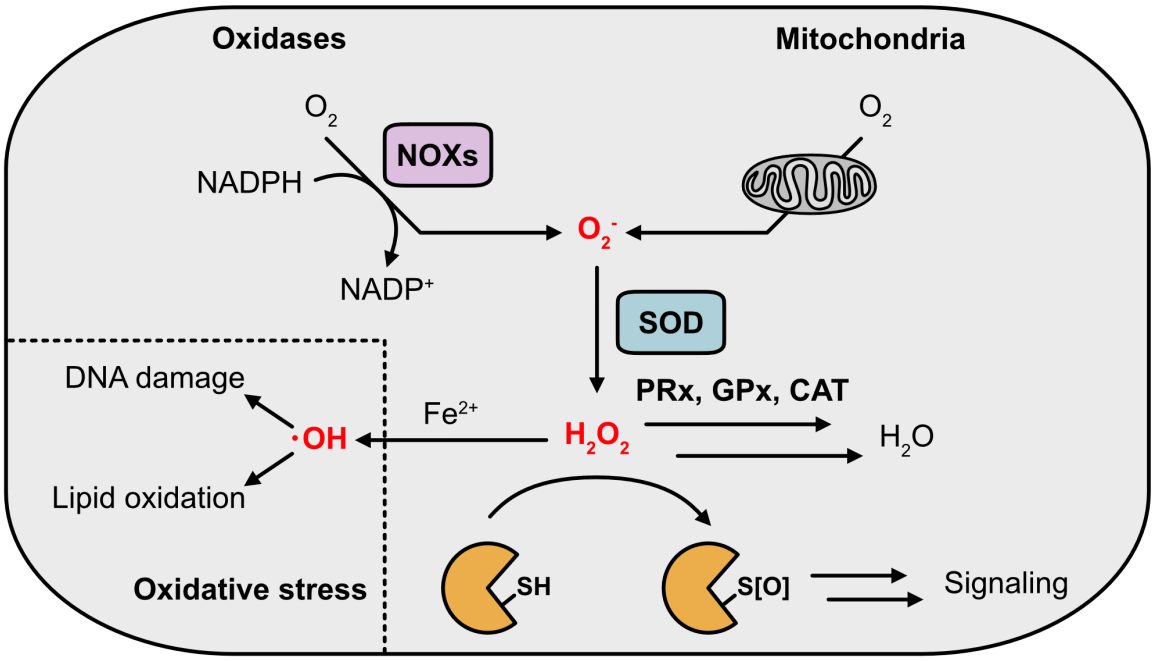


Figure 1.1: Superoxide (O_2^-) is produced as either a byproduct of oxidation by NADPH oxidase (Nox), or by incomplete reduction of oxygen during aerobic respiration in the mitochondria. O_2^- is rapidly converted into H_2O_2 by superoxide dismutases (SODs), where it can go on to oxidize redox sensitive cysteine residues, or be reduced to H_2O by redox buffering enzymes, such as peroxiredoxins (PRx), glutathione peroxidase (GPx), or catalase (CAT). Extremely high levels of H_2O_2, can result in the formation of hydroxyl radicals ($^{\bullet}OH$) via Fenton chemistry catalyzed by Fe^{2+}, which can lead to lipid oxidation and DNA damage. Figure adapted from Schieber *et. al.*[2]

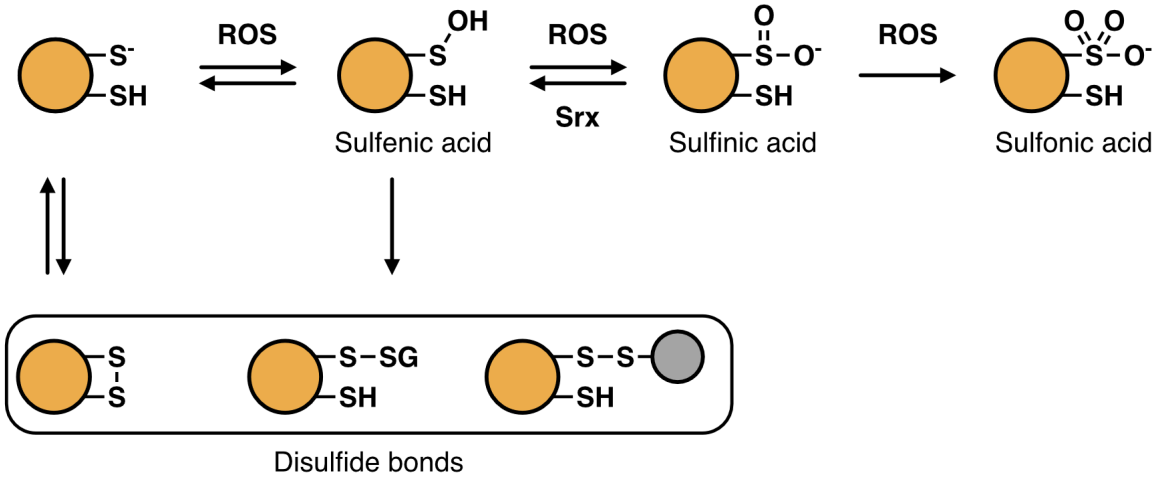


Figure 1.2: Examples of reversible and irreversible cysteine oxidative PTMs. Figure adapted from Abo *et. al.*[5]

tions (PTMs), such as oxidation, nitrosation, and glutathionylation.[10] Diverse protein classes, including proteases, oxidoreductases, kinases, and acyltransferases, contain reactive and functional cysteine residues.[8] Cysteine is subject to a wide array of distinct PTM including as S-sulfenylation, S-sulfinylation, and S-sulfonylation, as well as inter and intramolecular disulfide formation (Figure 1.2). S-sulfenylation (R–SOH) is readily reversible under cytosolic redox potentials by glutathionone, thioredoxin, and peroxiredoxin.[5] S-sulfinylation ($R–SO_2H$) is reversed by sulfiredoxin (Srx),[11] while cysteine S-sulfonylation ($R–SO_3H$) is considered to be an irreversible modification.

In addition to the SODs, cells utilize redox buffering systems including the catalase, glutaredoxin / glutathione, thioredoxen / thioredoxin reductase, and the glutathione /glutathione reductase systems to control cellular levels of ROS.[2] Catalase is thought to be exclusively localized to the peroxisome, while glutathione peroxidase (GPx) and peroxiredoxin (Prx) are widely distributed throughout the cell. GPxs are much less abundant than Prxs which account for between 0.1–0.8 % of total soluble protein[12] Therefore, Prx family members play a more significant role in regulating H_2O_2 levels in the context of signaling. Kinetic experiments with human Prx1 and 3 revealed that the second order rate constant of the oxidation of the Prx thiolate to the sulfenic acid ($C_p–S^- \longrightarrow C_p–SOH$) by H_2O_2 was between 1×10^6 and $1 \times 10^8\,M^{-1}\,s^{-1}$, which is several orders of magnitude faster than the thiol groups of other proteins.[13,14] These data suggest that the cysteine residues on Prx are at a competitive disadvantage compared to those on proteins targeted by H_2O_2 during cell signaling, and in order for H_2O_2 to rise to sufficient levels for signaling transduction, the basel level of Prx activity must be overcome.

1.2 EGFR pathway

The epidermal growth factor receptor (EGFR) is the prototypical member of the HER family of receptor tyrosine kinases (RTKs) which regulate many cellular pro-

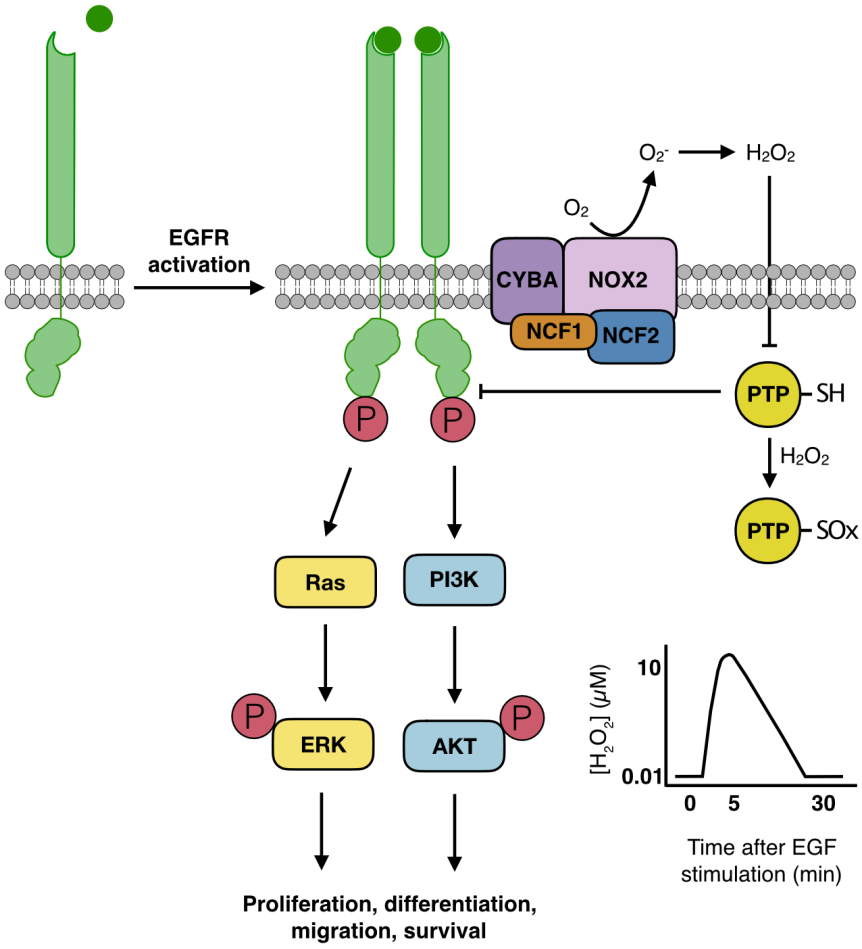


Figure 1.3: EGFR signaling pathway. Upon stimulation with EGF, EGFR monomeric receptors dimerize and autophosphorylate tyrosine residues in the cytoplasmic domain. Phospo-tyrosines (red circles) act as binding sites for scaffolding proteins for down stream kinases. Two major EGFR effectors, the Ras/ERK and PI3K/AKT pathways are shown. EGFR activation also triggers the assembly of the NADPH oxidase complex (Nox) at the cell membrane. H_2O_2 transiently inactivates phosphatases thereby amplifying growth factor signaling. Figure adapted from Truong *et. al.*[9]

cesses including growth, proliferation, and differentiation. As such, overexpression, and mutations of EGFR and many of its downstream signaling targets are found in a vast majority of cancers.[15] EGFR is a single pass transmembrane protein, comprised of an extracellular ligand binding domain, a transmembrane domain, and an intracellular kinase domain. Activation of RTKs by their specific extracellular ligands leads to the formation of receptor dimers, trans-autophosphorylation, and propagation of activation signals via intracellular mediators (Figure 1.3). The primary EGFR ligand in humans is epidermal growth factor (EGF), a 6 kDa protein which is found ubiquitously in various tissues.[16]

The two most important effectors activated by the binding of EGF to EGFR are the Ras/mitogen activated kinase (MAPK) and the phosphoinositide 3-kinase (PI3K) / protein kinase B (Akt) pathways. Activation of the Ras/MAPK cascade is initiated by the recruitment of growth receptor bound protein (Grb2) to bind to phosphotyrosine (pY) 1068 and pY1086 on EGFR via its src homology 2 (SH2) domain.[16] Grb2 acts as adaptor protein to recruit the RAS guanine exchange factor (GEF) SOS and form the SH2-Grb2-SOS complex. Once associated with the receptor, SOS mediates the guanine nucleotide exchange of Ras, a monomeric GTPase which subsequently activates the Raf-MEK-Erk1/2 kinase cascade. EGFR can also transmit signals through the PI3K/Akt pathway. Upon EGFR stimulation, PI3K generates phosphatidylinositol 3,4,5-trisphosphate (PIP3), which serves as a membrane anchor for activated Akt.[9]

1.2.1 Synergy between ROS production and protein phosphorylation during cell signaling

The connection between hydrogen peroxide (H_2O_2) and EGFR was first suggested when it was observed that cells stimulated with exogenous H_2O_2 showed increased incorporation of [^{32}P]phosphate. Although classically thought of as a toxic byproduct

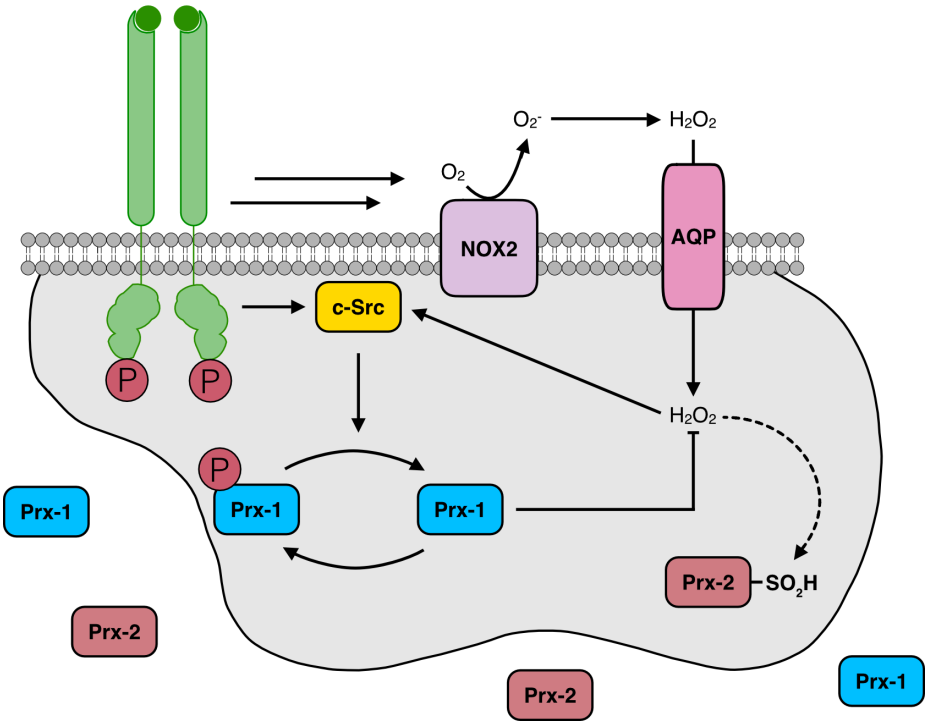


Figure 1.4: A model for local H_2O_2 accumulation. Figure adapted from Woo *et. al.*[17]

of cellular respiration, it was concurrently demonstrated in the late 1990s that H_2O_2 can act as a secondary messenger in cell signaling.[9] A landmark study by Rhee *et al* showed that EGF stimulation increased intracellular levels of H_2O_2 and Tyr phosphorylation on phospholipase C, gamma 1 (PLCγI) and suggested that inactivation of cysteine dependent protein tyrosine phosphatases (PTPs) by H_2O_2 may be required for EGF signaling.[18]

The NADPH oxidases (Nox) are the source of H_2O_2 which acts as a secondary messenger during EGFR activation. Five Nox isoforms and their dual oxidase counterparts (Duox) generate superoxide via electron transfer from NADPH to molecular oxygen. Nox2 directly associates with EGFR upon activation and has been show to sulfeynate several cysteine residues on the receptor and amplify its activity.[19] The hypothesis put forward by Rhee *et al*, that oxidation of PTPs is required for EGF signaling, was confirmed when later studies showed that H_2O_2 produced by Nox temporally inactivates PTP1B[17] (which acts on pY residues on EGFR) and phosphate

and tension homolog (PTEN),[20,21] by oxidation of their active site cysteine residues during EGF stimulation. PTEN is a lipid phosphatase which reverses phosphorylation of PIP3 and is commonly inactivated in cancer.[16] In order for H_2O_2 to act as a signaling messenger, its concentration must rise above basal levels of between 20 nM and 2 µM.[22] Thus, during EGFR activation, in local areas of the cytosol adjacent to the cell membrane, levels of H_2O_2 are allowed to rapidly rise resulting in inactivation of PTPs, thereby amplifying signal propagation via kinase cascades (Figure 1.4).

1.3 Proteomic methods to characterize cysteine PTMs

Because of the central role cysteine plays in redox signaling and protein function, a wide array of proteomic techniques have been developed for characterizing cysteine PTMs. Herein, we classify these methods into two general categories. The first section describes indirect methods which target the free cysteine thiolate. The second section describes methods to which use chemoselective methods to target specific PTMs.

1.3.1 Cysteine targeted chemical-proteomic methods

1.3.1.1 OxICAT

The oxidative isotope-coded affinity tags (OxICAT)[23] approach is based off earlier isotope-coded affinity tags (ICAT) probes.[24] ICAT reagents have a cysteine reactive iodoacetamide (IA) group, a linker with 8 hydrogen or deuterium atoms which allow for the differentiation between heavy and light labeled species by mass-spectrometry (MS), and a biotin enrichment handle (Figure 1.5). In an OxICAT experiment samples are labeled sequentially, first with the light ICAT probe, reduced, then labeled with the heavy probe. Labeled proteins are isolated by streptavidin enrichment, trypsinized, and cysteine containing peptides are recovered. The tagged peptides are then analyzed by LC-MS/MS, where the heavy and light peptide signal intensities

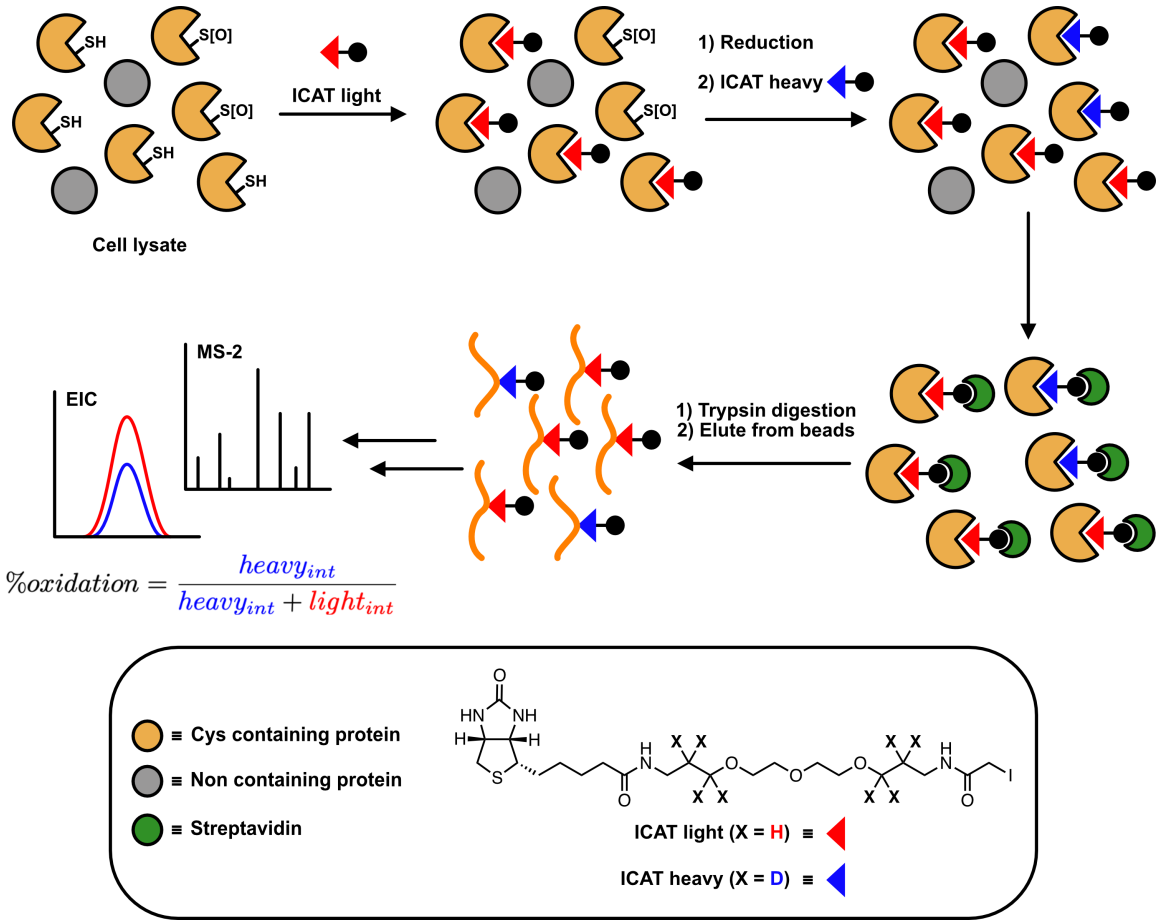


Figure 1.5: OxICAT workflow. Cell lysates are sequentially labeled with light then heavy isotope-coded affinity tags (ICAT) reagent. Labeled proteins are enriched and trypsin digested. Labeled peptides are then recovered and analyzed by LC-MS/MS.

are used to determine the stoichiometry of reduced vs. oxidized cysteines (Figure 1.5). A modified version of the OxICAT workflow utilizing a cleavable biotin can also be used to facilitate a more efficient recovery of labeled peptides from streptavidin beads.[25,26] Another variant, used a bifunctional IA probe coupled to a non-hydrolyzable phosphate which enables the purification of labeled peptides utilizing existing immobilized metal affinity chromatography (IMAC) techniques already used in phosphoproteomics.[27] Variants of OxICAT has been used to quantify the extent of cysteine oxidation in both purified proteins,[25] and in a wide array of model organisms in including yeast,[28] *Drosophila*,[29] mammalian cells,[26] as well as several mouse tissues.[27]

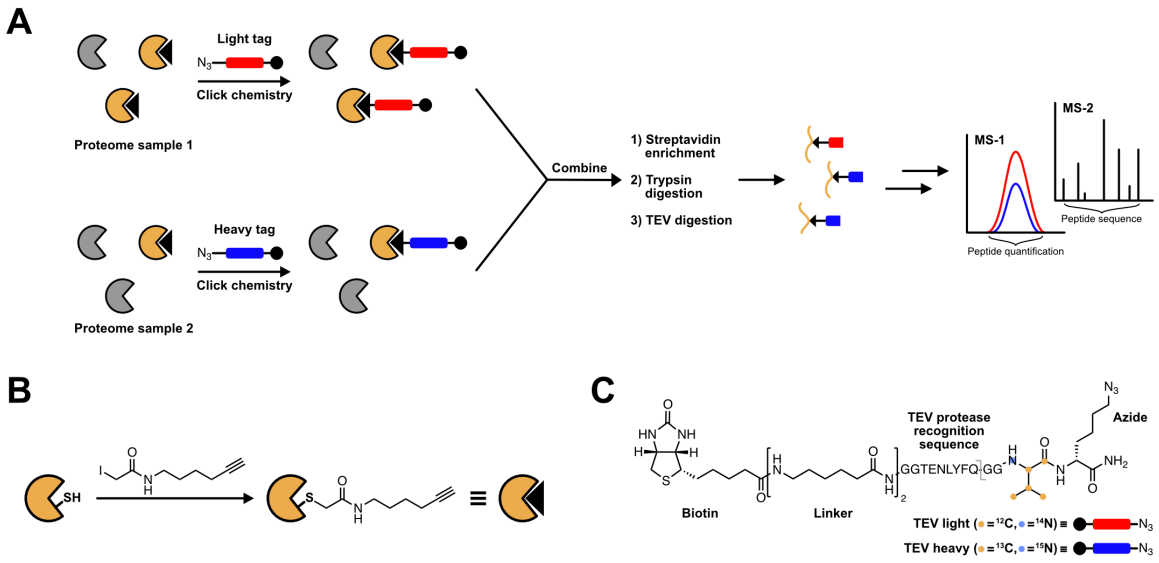


Figure 1.6: **(A)** General isoTOP-ABPP workflow. Reactive cysteine residues on two proteome samples are labeled with iodoacetamide-alkyne (IAA), followed by CuAAC with an isotopically heavy or light biotin-azide cleavable linker. The two lysates are combined, biotinylated proteins are enriched on streptavidin-agarose beads, and subjected to an on-bead trypsin digestion. The IA-labeled peptides are released and analyzed by LC/LC-MS/MS. Heavy and light peptide pairs are quantified by their extracted MS1 peaks. **(B)** IAA structure and cysteine-labeling scheme. **(C)** The protease cleavable biotin-azide tag for isoTOP-ABPP.

1.3.1.2 Activity based protein profiling

One disadvantage of OxICAT is that the detection of oxidative modified as well as highly reactive cysteines may be diminished by other highly abundant cysteine containing proteins.[10] Instead of measuring the percent of oxidation within a single sample, an alternative strategy is to target the analysis to highly reactive cysteines and compare reactivity or site occupancy across two separate samples. Isotopic tandem orthogonal proteolysis - activity-based protein profiling (isoTOP-ABPP) is a derivative of activity-based protein profiling (ABPP), a pioneering technology for interrogating protein activity directly in complex biological systems. In general, ABPP probes contain three elements: (1) a reactive warhead for covalently labeling target proteins; (2) a reporter tag for affinity purification or fluorescence detection; and, (3) a linker to minimize steric hindrance between the reporter and reactive groups.[30,31]

Early ABPP methods utilized reactive warheads targeting a specific enzyme family, such as the fluorophosphonate probe for the serine hydrolases.[32] In latter iterations, more reactive and promiscuous electrophiles were utilized,[33] culminating in the use of an iodoacetamide-alkyne (IAA) probe for modification of reactive cysteines in the proteome. The isoTOP-ABPP platform couples the IAA probe with an isotopically tagged cleavable linker enabling the selective enrichment, release, and MS relative quantification of IA-labeled peptides from two samples. The isoTOP-ABPP platform involves the following steps: (1) treatment of lysates with IAA to label reactive cysteine thiolates; (2) conjugation of IA-labeled cysteines in control and experimental samples to isotopically differentiated cleavable azide-biotin tags using copper-catalyzed azide-alkyne cycloaddition (CuAAC);[34] (3) enrichment of IA-labeled proteins on streptavidin beads, followed by on-bead tryptic digestion, and linker cleavage to release IA-labeled peptides; and, (4) analysis of the resulting isotopically heavy and light peptide pairs using LC/LC-MS/MS to quantify reactivity differences in two samples using light:heavy isotopic ratios[6] (Figure 1.6).

Since the initial development of the isoTOP-ABPP platform, various iterations to the initial workflow have been reported, including variations to: (1) the cysteine-reactive electrophile; (2) the cleavable biotin-azide tag; and, (3) the mode of heavy isotope incorporation. These recent advances are summarized below.

IAA is widely used as the reactive warhead in ABPP,[6,35,36] however other electrophiles have also been incorporated into the isoTOP-ABPP workflow. A photocaged bromomethyl ketone (CBK),[37] and iodomethyl ketone (CIK4)[38] were shown to have lower cytotoxicity compared to IA-alkyne and used to profile reactive cysteines in live cells with high spatial and temporal control (Figure 1.7A,B). CBK was used to monitor changes in cysteine reactivity in A431 cells in response to the EGF stimulated release of reactive oxygen species.[37] However, one challenge of live cell labeling followed by reactivity profiling is that the starting protein concentrations of the

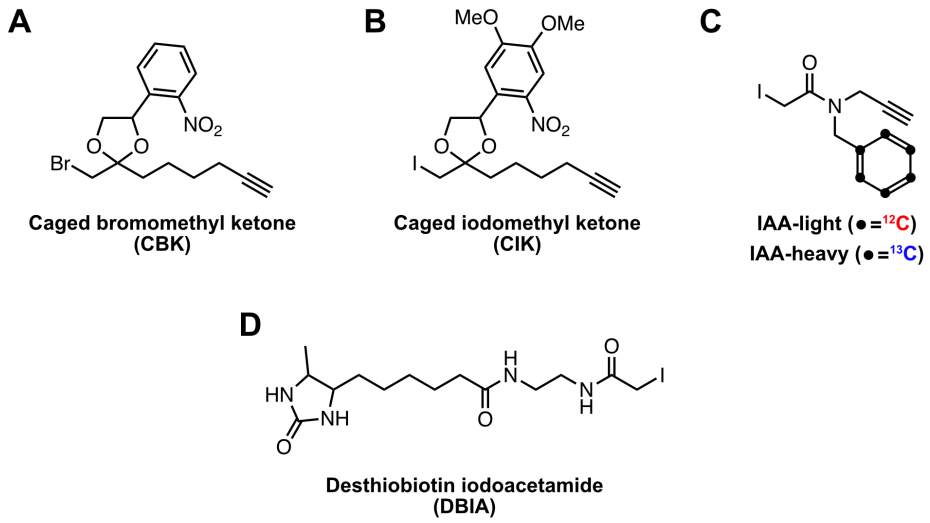


Figure 1.7: Structures of **(A)** caged bromomethyl ketone (CBK), **(B)** caged iodomethyl ketone (CIK4), **(C)** iodoacetamide alkyne heavy and light and, **(D)** desthiobiotin iodoacetamide (DBIA).

heavy and light samples must be equalized during the process of culturing the cells. Normalizing the protein concentration after lysis will result in changes in the stoichiometry of the cysteine reactive probes used to label the cells. Alternatives to halo acetamide electrophiles have also been developed, including aryl halides such as p-chloronitrobenzene,[39] and hypervalent iodine regents, such as ethynyl benziodoxolone (EBX).[40]

The initial isoTOP-ABPP platform utilized isotopically labeled, protease-cleavable biotin-azide tags for CuAAC-mediated conjugation to probe-labeled proteins. Due to the advent of a wide-variety of cleavable chemistries, isotopically tagged variants of chemically cleavable[41] and photocleavable[42,43] biotin-azide tags have been generated and shown to be compatible with the isoTOP-ABPP workflow.

Lastly, approaches to incorporate isotopic labels into the isoTOP-ABPP workflow have been explored, including: (1) stable isotope labeling by amino acids in cell culture (SILAC) to incorporate isotopes into the proteomes under evaluation; (2) isotopically tagged cysteine-reactive probes; (3) isotopically light and heavy biotin-azide linkers; and, (4) post-digest peptide labeling using reductive dimethylation

(ReDiMe) or isobaric tags (iTRAQ, TMT). Incorporation of isotopic labels into the cysteine-reactive probe was accomplished through synthesis of isotopically differentiated iodoacetamide-alkyne probes containing a $^{12}C_6$ or $^{13}C_6$ benzyl moiety, termed IA-alkyne light (IAL) or IA-alkyne heavy (IAH), respectively[25] (Figure 1.7C). IAL and IAH can be obtained through a more facile synthesis, and using less expensive starting materials compared to the isotopic biotin-azide tags.[25] IAL and IAH also support the profiling of reversible cysteine modifications within the same sample in a work flow similar to the OxICAT method.[23] Combining TOP-ABPP with reductive dimethylation (ReDiMe) results in a method called rdTOP-ABPP.[42] rdTOP-ABiPP was shown to be comparable to several commercially available linkers for site of identification, and has the added benefit of supporting triplex quantitative experiments.[42] Multiplexed thiol reactivity profiling (MTRP) cysteine profiling with isobaric tags for relative and absolute quantization (iTRAQ) labeling allows for quantitative site of labeling to be performed with up to eight samples in parallel.[43] Although MTRP reagents are more expensive than those used for rdTOP-ABPP, MTRP requires less sample input when highly multiplexed comparisons are required for an experiment. Another multiplexed cysteine profiling method, termed streamlined cysteine activity-based protein profiling (SLC-ABPP) uses desthiobiotin iodoacetamide (DBIA) which allows for a streamlined one-pot sample preparation[44] (Figure 1.7). The SLC-ABPP method further improves the depth of proteome coverage by employing real time searching[45] to MS reduce analysis time to 3 hours instead of a longer multidimensional fraction approach traditionally used for isoTOP-ABPP.[6,46]

1.3.2 Cysteine PTM targeted chemical-proteomic methods

Cysteine targeted approaches rely on the intrinsic reactivity of cysteine and therefore do not differentiate between specific PTMs such as sulfenic acids, sulfinic acids, and disulfides.[5] A complementary strategy is to target the analysis to specific cys-

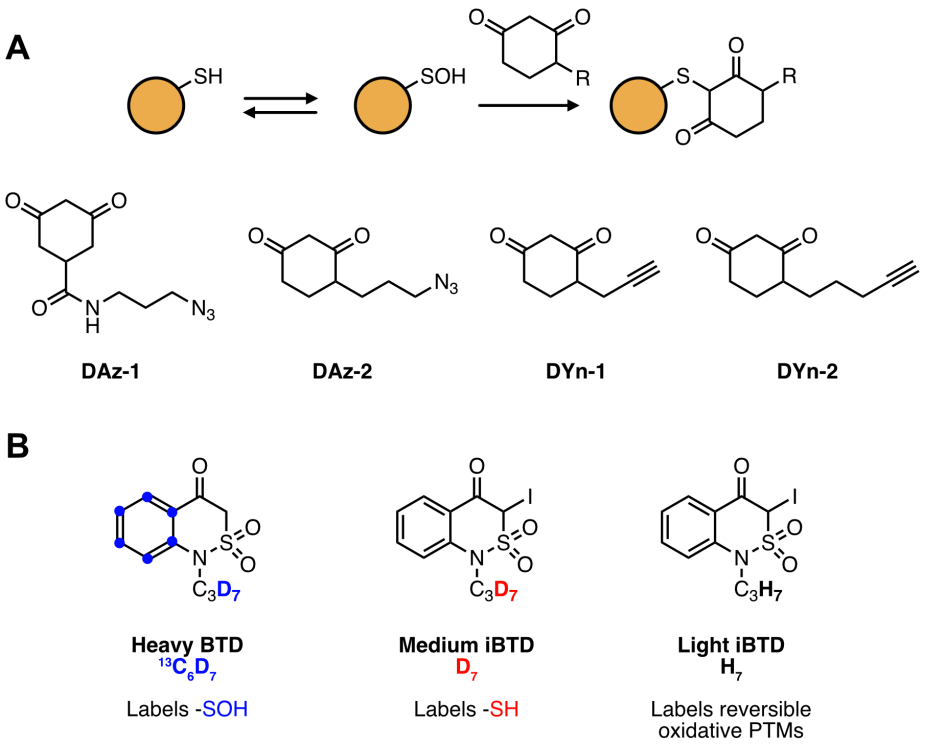


Figure 1.8: Structures of sulfenic acid probes. **(A)** Reaction between sulfenic acid and dimedone, and structures of dimedone probes. **(B)** Structures of benzothiazine dioxide (BTD) and iodo BTD (iBTD) probes. Figure adapted from Abo *et. al.*[5]

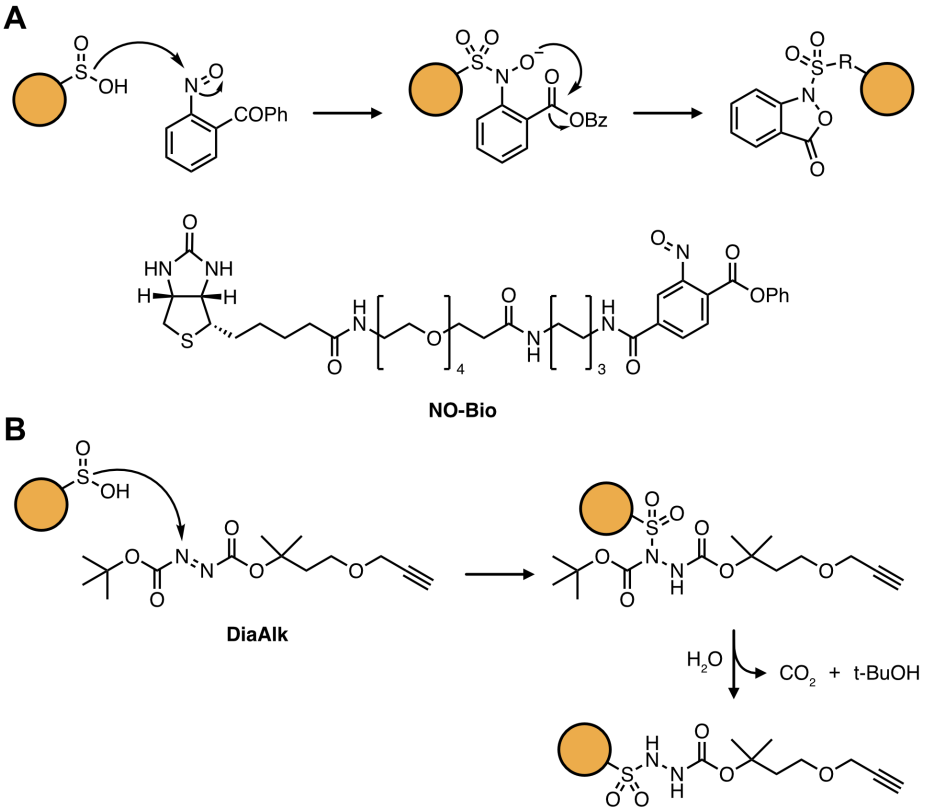


Figure 1.9: Structures of sulfinic acid probes. **(A)** 2-nitroso benzoic acid probes for sulfinic acid. **(B)** Reaction between diazine probe and sulfinic acid. Figure adapted from Abo *et. al.*[5]

teine PTMs. One strategy for selective profiling of sulfenylation is to selectively reduce sulfenic acids with arsenite, then label with a biotin maleamide probe,[47] in a workflow analogous to the biotin switch assay.[48] Newer methods use chemoselective probes to directly react with sulfenic acids. The enolate form of dimedone will undergo nucleophilic attack to the sulfenic acids to form a covalent adduct.[5] Azide functionalized dimedone derivatives **DAz-1** and **DAz-2** which are compatible with bioconjugation, were assessed for live cell labeling and proteomic applications[49] (Figure 1.8A). **Dyn-2** was shown to have superior cell permeability and reactivity and was used to identify 175 candidate sites of sulfenation.[49] In later iterations of dimedone probes, the azide was replaced with an alkyne to develop probes with better *in vitro* stability[50] (Figure 1.8A). Application of **DYn-2** in A431 cells revealed an EGF dependent sulfenation of Cys797 on EGFR.[19]

In addition to identifying sites of sulfenation, another important goal is to measure the extent of modification. 2-iodo dimedone will selectively label the free cysteine thiol, thereby enabling the percent of sulfenylation on a peptide to be quantified by sequential labeling with D_6-dimedone and 2-iodo dimedone.[51] A similar approach approach developed by Albertolle *et al.* uses benzothiazine dioxide (BTD) and iodo-BTD probes to simultaneously quantify relative levels of sulfenic acids, free thiols, and reversable oxidative PTMs, however the lack of an enrichment handle limited the number of sulfenylated residues which were identified to $\sim$ 600 sites[52] (Figure 1.8B). Finally, Shi *et. al* used triphenylphosphonium ylide probes with enrichment handles to quantify the percent of $\sim$ 2000 sites in the mitochondria.[53]

S-sulfinylation was originally regarded an irreversible artifictual oxidation event, but more recent studies have suggested that the modification is reversible,[11] but also important in redox signaling.[54] Prx 1-4 are believed to be regulated by S-sulfinylation as well as reduction by Srx[55] 2-nitroso benzoic acid derivatives will undergo an electrophilic reaction to selectively capture sulfinic acids[56] (Figure 1.9A). A significant

drawback of **NO-Bio** probes is that they exhibit complex fragmentation patterns which are incompatible with identification of the site of sulfination by MS. To facilitate the detection sites of S-sulfinylation, Akter *et al.* developed a diazene probe (**DiaAlk**) which shows superior sensitivity and suports the site site-centric identification and quantification of sulfinic acids.[57]

1.3.3 Summary

Above we have summarized indirect and modification targeted methods to study cysteine PTMs. There is an nuanced difference between what the quantitative information given by each method represents. Cysteine targeted methods give information about the site occupancy and stoichiometry of labeling. However information about the chemical nature of the specific modification(s) is lost. In addition, some modifications occur at such low levels that they are difficult to discern in indirect labeling experiments. While it is true that changes in redox state may not be as apparent in data from indirect approaches compared to PTM targeted approaches, it is also important to note that many solvent exposed cysteine residues are non-specifically oxidized to some extent.[23,29,58,59] For example, a transition from 0.01 % to 0.5 percent oxidation would represent a 50-fold change, however such a small overall level of oxidation is unlikely to have significant biological consequences.[27] Therefore, PTM directed methods are more informative about the nature of the modification, but loose the context of the overall extent of modification of the protein.

1.4 Proximity dependent labeling methods

As mentioned in section 1.2.1, ROS produced during cell signaling is localized to specific sub cellular regions. Cell fractionation workflows can be combined with cysteine profiling to enrich subsets of proteins confined to a particular organelle,[26,60] however such approaches are not amenable to profiling targets of ROS during signaling

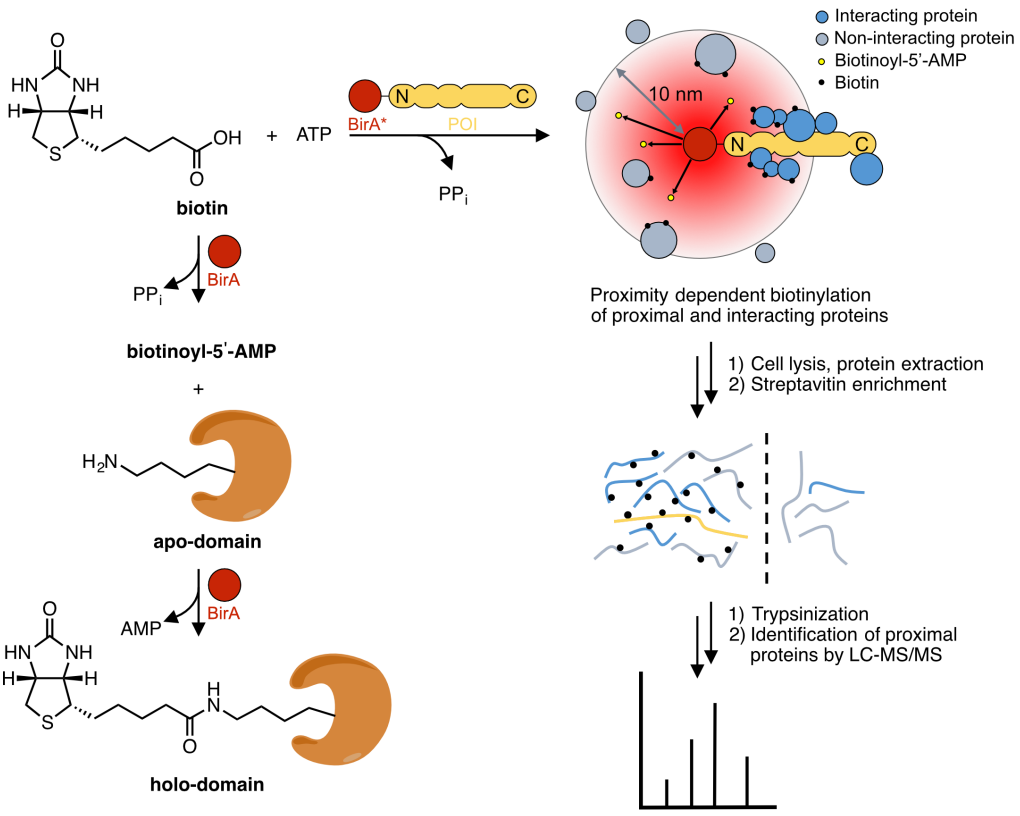


Figure 1.10: BirA hydrolyzes ATP to form a reactive biotinoyl-5′-AMP intermediate. Wild-type BirA remains bound to biotinoyl-5′-AMP until it is transferred to the apo acetyl-CoA carboxylase subunit. The R118G BirA* mutant still hydrolyzes ATP, but does not retain the affinity for biotinoyl-5′-AMP after it forms. Instead reactive biotinoyl-5′-AMP will be released allowing it to react with interacting proteins within a ~10 nm radius. In a BioID experiment, biotinyated proteins can be enriched, digested and identified by LC-MS/MS. Figure adapted from Varnaitė *et. al.*[61]

because ROS production mostly occurs in the cytosol. In the following sections, we will provide an overview of two proximity ligation strategies, BioID and APEX which are which provide a potential platform to spatially control profiling of ROS targets during cell signaling.

1.4.1 BioID

BioID is a proximity dependent labeling technique which offers significant advantages over traditional tools used to identify protein-protein interactions (PPIs). Specifically, because the labeling process occurs in live cells, transient or weak PPIs

which would otherwise not survive harsh cell lysis conditions can still be identified. BioID uses a mutant form of the *E. coli* biotin ligase (BirA*) to promiscuously biotinylate interacting proteins.[61,62] The wild-type BirA is a biotin ligase which transfers biotin to a subunit on the *E. coli* acetyl-CoA carboxylase.[61,63] The transfer of biotin to acetyl-CoA carboxylase is a two step process. First BirA hydrolyzes one molecule of ATP and transfers it to biotin to form a highly reactive biotinoyl-5′-AMP (adenosine monophosphate) intermediate[63] which is then transferred to the ε-amine on a lysine on the protein target to form an amide bond accompanied by the release of AMP.[64] Wild-type BirA binds to biotinoyl-5′-AMP with a high affinity until it is transferred to a lysine residue on the apo domain of acetyl-CoA carboxylase. The R118G BirA mutant (BirA*) retains the ability to hydrolyze ATP but has a reduced affinity for the intermediate, and therefore releases activated biotin in a reactive cloud where it labels interacting proteins in a proximity dependent manner (Figure 1.10). The labeling radius of BirA* was estimated to be ~10 nm based off the known structure of the nuclear pore complex.[65] In a BioID experiment, a bait protein is expressed fused to BirA* and 50 µM exogenous biotin is added to initiate labeling. After labeling is complete, cells are lysed and biotinylated proteins are isolated and identified by LC-MS/MS. For the first generation biotin ligases, the labeling process takes between 3 and 24 hours.[62,66] Therefore labeled proteins represent a history of interactions over the course of several hours, rather than an instantaneous snapshot.

To improve on the temporal resolution of the original BirA*, Branon *et. al* used yeast display directed evolution of the *E. coli* BirA to engineer a faster biotin ligase, named TurboID.[67] TurboID exhibits a 6 fold increase in activity relative to BirA* and is able to achieve the same level of labeling in 10 minutes compared to the same level of labeling achieved in 18 hours with BirA *. TurboID has a total of 14 mutations Interestingly, most of the mutations are located distal to the catalytic site and are distributed through the structure of the protein. Due to the wide distribution of

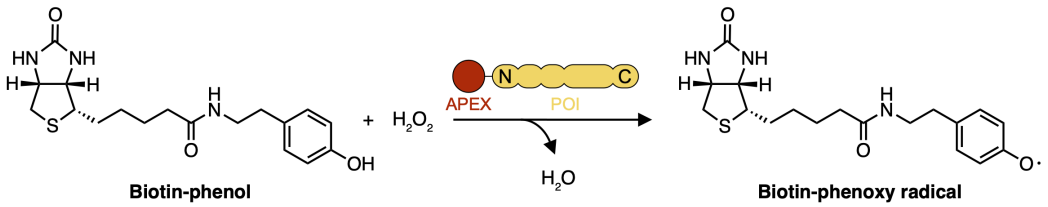


Figure 1.11: In APEX, a bait protein of interest is expressed fused to a peroxidase, which in the presence of H_2O_2, oxidizes biotin pheonol to the phenoxyl radical. The activated biotin forms a covalent adduct with electron rich amino acids, such as tyrosine, cysteine, and histidine on proximal proteins.

mutations in TurboID, there is no clear mechanistic explanation for how they effect activity.[67,68] Due to its increases catalytic efficiency TurboID is also compatible with *in vivo* applications. Branon and coworkers also demonstrated the use of TurboID in *Drosophila* and *C. elegans*.[67] Others have used TurboID in mice,[69] Zebrafish,[70] and *Arabidopsis*.[71]

1.4.2 APEX

Another proximity labeling method, termed APEX, uses an engineered ascorbate peroxidase to activate biotin-pheonol by oxidation to form a phenoxyl radical. Biotin-phenoxide then reacts with tryosine, cysteine, and histidine on proximal proteins to forma a covalent adduct (Figure 1.11). Although horseradish peroxidase (HRP) can also be used for similar applications extracellularly, it is inactive when expressed in the cytosol, presumably due to the presence of four structural disulfide bonds which depend on binding to Ca^{2+} for the enzyme to function.[72] APEX was initially used as a genetically encoded electron microscopy contrast tag[73] and was later used to map the proteome of the mitochondrial matrix,[74] and inter membrane space.[75] APEX is also compatible with *in vivo* applications in *Drosophila* where it was used with tissue specific and subcellular resolution.[76] A notable difference between APEX and first generation biotin ligases is that cells are first incubated with biotin phenol and labeling is activated with the addition of 1 mM H_2O_2. The labeling process takes

as little as 1 minute before the reaction is quenched. Therefore APEX provides a short snapshot of protein interactions compared to the longer labeling time required by BioID. Importantly, the shorter labeling time allowed APEX to be used in a temporally controlled experiment to profile interactors of G-protein-coupled receptors (GPCRs).[77,78] However, the requirement of 1 mM H_2O_2 is likely to interfere with endogenous production of H_2O_2 used to regulate cell signaling pathways.

1.4.3 Design of proximity labeling experiments

Normally, bait proteins are ectopically expressed via lentiviral transduction,[79] by insertion into a recombinant site[80] in Flp-In cell lines, or transiently expressed for *in vivo* applications.[69] However, Vandemoortele *et. al* used CRISPR/Cas9 to append BirA* to the C-terminus of the endogenous p53 tumor suppressor gene in HCT116 cells.[81] To achieve the most biologically relevant relevant results from a proximity labeling experiment, the native activity of the bait protein should be preserved as much as possible. Indeed, excessive protein overexpression can have a deleterious effect on localization and activity.[82,83] A significant advantage of their approach is that the expression of the bait protein is under the control of the endogenous promoter which limits artifacts caused by constitutive overexpression of a bait protein. Even if a complex and labor intensive genome editing method is not used to express a bait protein, the proper localization and retention of native activity should still be verified.

Even if the bait retains the behavior of the native protein, negative controls are essential to account for non-specific interactions *in cellulo* as well as contaminants which are pulled down during streptavidin enrichment.[84] Depending on experimental design, several types of negative controls are necessary to discrimination in true vs. false interactions. The first type of control could be the cell line in the absence of labeling reagents (omission of supplemental biotin for BioID or H_2O_2 or biotin-phenol for APEX), which can help to identify endogenously biotinylated proteins as well as

non-specific binders to the matrix used for affinity purification.[68] Another type of control involves expression of BirA* or APEX without the bait fusion accounts for non-specific interactions with the bait protein.[68] Finally in experiments where the same bait construct is being used to compare interactors in the presence of some stimulant, the a sample with the omission of the stimulant can serve as a negative control. For example, Zheng *et. al* compared interactors of the same protein arginine deiminase (PAD) 2 BioID construct (+/-) of 5 µM ionomycin to identify proteins involved in the nuclear import of PAD2.[85] For measuring the relative abundance or proteins in control vs. bait samples, a wide array of quantification methods have been used including SILAC,[75,85,86] label free quantification (LFQ),[69,81] label free selected reaction monitoring (SRM),[78] tandem mass tags (TMT),[77] and spectral counting.[65] However maintaining consistent expression and labeling levels between controls which are different from the bait construct are necessary for the quantitative information to be meaningful.

1.5 Conclusion

In summary, we have provided an overview of redox regulation of cell signaling, proteomic methods to characterize cysteine oxidative PTMs, and proximity biotinylation methods compatible with live cells. The next chapter will describe the development of a proteomic platform combining TurboID with OxICAT to identify localized sites of cysteine oxidation during EGFR activation.

References

[1] Maurais, A. J.; Weerapana, E. Reactive-cysteine profiling for drug discovery. *Curr Opin Chem Biol* **2019**, *50*, 29–36.

[2] Schieber, M.; Chandel, N. S. ROS function in redox signaling and oxidative stress. *Curr Biol* **2014**, *24*, R453–62.

[3] Paulsen, C. E.; Carroll, K. S. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chem Rev* **2013**, *113*, 4633–79.

[4] Bedard, K.; Krause, K. H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev* **2007**, *87*, 245–313.

[5] Abo, M.; Weerapana, E. Chemical Probes for Redox Signaling and Oxidative Stress. *Antioxid Redox Signal* **2019**, *30*, 1369–1386.

[6] Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790–5.

[7] Marino, S. M.; Gladyshev, V. N. Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *J. Mol. Biol.* **2010**, *404*, 902–916.

[8] Giles, N. M.; Giles, G. I.; Jacob, C. Multiple roles of cysteine in biocatalysis. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 1–4.

[9] Truong, T. H.; Carroll, K. S. Redox regulation of epidermal growth factor receptor signaling through cysteine oxidation. *Biochemistry* **2012**, *51*, 9954–65.

[10] Yang, J.; Carroll, K. S.; Liebler, D. C. The Expanding Landscape of the Thiol Redox Proteome. *Mol. Cell. Proteomics* **2016**, *15*, 1–11.

[11] Biteau, B.; Labarre, J.; Toledano, M. B. ATP-dependent reduction of cysteine-sulphinic acid by S. cerevisiae sulphiredoxin. *Nature* **2003**, *425*, 980–4.

[12] Rhee, S. G.; Woo, H. A. Multiple functions of 2-Cys peroxiredoxins, I and II, and their regulations via post-translational modifications. *Free Radic Biol Med* **2020**, *152*, 107–115.

[13] Peskin, A. V.; Low, F. M.; Paton, L. N.; Maghzal, G. J.; Hampton, M. B.; Winterbourn, C. C. The high reactivity of peroxiredoxin 2 with H_2O_2 is not reflected in its reaction with other oxidants and thiol reagents. *J Biol Chem* **2007**, *282*, 11885–92.

[14] Cox, A. G.; Peskin, A. V.; Paton, L. N.; Winterbourn, C. C.; Hampton, M. B. Redox potential and peroxide reactivity of human peroxiredoxin 3. *Biochemistry* **2009**, *48*, 6495–501.

[15] Lemmon, M. A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2010**, *141*, 1117–34.

[16] Wee, P.; Wang, Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)* **2017**, *9*, DOI: 10.3390/cancers9050052.

[17] Woo, H. A.; Yim, S. H.; Shin, D. H.; Kang, D.; Yu, D. Y.; Rhee, S. G. Inactivation of peroxiredoxin I by phosphorylation allows localized H_2O_2 accumulation for cell signaling. *Cell* **2010**, *140*, 517–28.

[18] Bae, Y. S.; Kang, S. W.; Seo, M. S.; Baines, I. C.; Tekle, E.; Chock, P. B.; Rhee, S. G. Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Role in EGF receptor-mediated tyrosine phosphorylation. *J Biol Chem* **1997**, *272*, 217–21.

[19] Paulsen, C. E.; Truong, T. H.; Garcia, F. J.; Homann, A.; Gupta, V.; Leonard, S. E.; Carroll, K. S. Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nat Chem Biol* **2011**, *8*, 57–64.

[20] Lee, S. R.; Yang, K. S.; Kwon, J.; Lee, C.; Jeong, W.; Rhee, S. G. Reversible inactivation of the tumor suppressor PTEN by H2O2. *J Biol Chem* **2002**, *277*, 20336–42.

[21] Park, H. S.; Lee, S. H.; Park, D.; Lee, J. S.; Ryu, S. H.; Lee, W. J.; Rhee, S. G.; Bae, Y. S. Sequential activation of phosphatidylinositol 3-kinase, βPix, Rac1, and Nox1 in growth factor-induced production of H_2O_2. *Mol Cell Biol* **2004**, *24*, 4384–94.

[22] Stone, J. R.; Yang, S. Hydrogen peroxide: a signaling messenger. *Antioxid Redox Signal* **2006**, *8*, 243–70.

[23] Leichert, L. I.; Gehrke, F.; Gudiseva, H. V.; Blackwell, T.; Ilbert, M.; Walker, A. K.; Strahler, J. R.; Andrews, P. C.; Jakob, U. Quantifying changes in the thiol redox proteome upon oxidative stress in vivo. *Proc Natl Acad Sci U S A* **2008**, *105*, 8197–202.

[24] Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999**, *17*, 994–9.

[25] Abo, M.; Li, C.; Weerapana, E. Isotopically-Labeled Iodoacetamide-Alkyne Probes for Quantitative Cysteine-Reactivity Profiling. *Mol Pharm* **2018**, *15*, 743–749.

[26] Bechtel, T. J.; Li, C.; Kisty, E. A.; Maurais, A. J.; Weerapana, E. Profiling Cysteine Reactivity and Oxidation in the Endoplasmic Reticulum. *ACS Chem Biol* **2020**, *15*, 543–553.

[27] Xiao, H. et al. A Quantitative Tissue-Specific Landscape of Protein Redox Regulation during Aging. *Cell* **2020**, *180*, 968–983 e24.

[28] Topf, U.; Suppanz, I.; Samluk, L.; Wrobel, L.; Boser, A.; Sakowska, P.; Knapp, B.; Pietrzyk, M. K.; Chacinska, A.; Warscheid, B. Quantitative proteomics identifies redox switches for global translation modulation by mitochondrially produced reactive oxygen species. *Nat Commun* **2018**, *9*, 324.

[29] Menger, K. E.; James, A. M.; Cocheme, H. M.; Harbour, M. E.; Chouchani, E. T.; Ding, S.; Fearnley, I. M.; Partridge, L.; Murphy, M. P. Fasting, but Not Aging, Dramatically Alters the Redox Status of Cysteine Residues on Proteins in Drosophila melanogaster. *Cell Rep* **2015**, *13*, 1285.

[30] Weerapana, E.; Speers, A. E.; Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP) - A general method for mapping sites of probe modification in proteomes. *Nat. Protoc.* **2007**, *2*, 1414–1425.

[31] Chen, X.; Wong, Y. K.; Wang, J.; Zhang, J.; Lee, Y. M.; Shen, H. M.; Lin, Q.; Hua, Z. C. Target identification with quantitative activity based protein profiling (ABPP). *Proteomics* **2017**, *17*, 3–4.

[32] Liu, Y.; Patricelli, M. P.; Cravatt, B. F. Activity-based protein profiling : The serine hydrolases. *Proc. Natl. Acad. Sci.* **1999**, *96*, 14694–14699.

[33] Weerapana, E.; Simon, G. M.; Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* **2008**, *4*, 405–407.

[34] Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **2001**, *40*, 2004–2021.

[35] Pace, N. J.; Weerapana, E. A competitive chemical-proteomic platform to identify zinc-binding cysteines. *ACS Chem. Biol.* **2014**, *9*, 258–265.

[36] Wang, C.; Weerapana, E.; Blewett, M. M.; Cravatt, B. F. A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. *Nat. Methods* **2014**, *11*, 79–85.

[37] Abo, M.; Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J Am Chem Soc* **2015**, *137*, 7087–90.

[38] Abo, M.; Bak, D. W.; Weerapana, E. Optimization of Caged Electrophiles for Improved Monitoring of Cysteine Reactivity in Living Cells. *Chembiochem* **2017**, *18*, 81–84.

[39] Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E. Investigating the proteome reactivity and selectivity of aryl halides. *J. Am. Chem. Soc.* **2014**, *136*, 3330–3333.

[40] Abegg, D.; Frei, R.; Cerato, L.; Prasad hari, D.; Wang, C.; Waser, J.; Adibekian, A. Proteome-Wide Profiling of Targets of Cysteine reactive Small Molecules by Using Ethynyl Benziodoxolone Reagents. *Angew. Chemie - Int. Ed.* **2015**, *54*, 10852–10857.

[41] Qian, Y.; Weerapana, E. In *Activity-Based Proteomics: Methods and Protocols*, Overkleeft, H. S., Florea, B. I., Eds.; Springer New York: New York, NY, 2017, pp 11–22.

[42] Yang, F.; Gao, J.; Che, J.; Jia, G.; Wang, C. A Dimethyl-Labeling-Based Strategy for Site-Specifically Quantitative Chemical Proteomics. *Anal Chem* **2018**, *90*, 9576–9582.

[43] Tian, C.; Sun, R.; Liu, K.; Fu, L.; Liu, X.; Zhou, W.; Yang, Y.; Yang, J. Multiplexed Thiol Reactivity Profiling for Target Discovery of Electrophilic Natural Products. *Cell Chem. Biol.* **2017**, *24*, 1416–1427.e5.

[44] Kulkarni, V. V.; Anand, A.; Herr, J. B.; Miranda, C.; Vogel, M. C.; Maday, S. Synaptic activity controls autophagic vacuole motility and function in dendrites. *J Cell Biol* **2021**, *220*, DOI: 10.1083/jcb.202002084.

[45] Schweppe, D. K.; Eng, J. K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E. L.; Erickson, B. K.; Paulo, J. A.; Gygi, S. P. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J Proteome Res* **2020**, *19*, 2026–2034.

[46] Washburn, M. P.; Wolters, D.; Yates J. R., 3. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19*, 242–7.

[47] Saurin, A. T.; Neubert, H.; Brennan, J. P.; Eaton, P. Widespread sulfenic acid formation in tissues in response to hydrogen peroxide. *Proc Natl Acad Sci U S A* **2004**, *101*, 17982–7.

[48] Forrester, M. T.; Foster, M. W.; Benhar, M.; Stamler, J. S. Detection of protein S-nitrosylation with the biotin-switch technique. *Free Radic Biol Med* **2009**, *46*, 119–26.

[49] Leonard, S. E.; Reddie, K. G.; Carroll, K. S. Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem Biol* **2009**, *4*, 783–99.

[50] Yang, J.; Gupta, V.; Carroll, K. S.; Liebler, D. C. Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nat Commun* **2014**, *5*, 4776.

[51] Seo, Y. H.; Carroll, K. S. Quantification of protein sulfenic acid modifications using isotope-coded dimedone and iododimedone. *Angew Chem Int Ed Engl* **2011**, *50*, 1342–5.

[52] Albertolle, M. E.; Glass, S. M.; Trefts, E.; Guengerich, F. P. Isotopic tagging of oxidized and reduced cysteines (iTORC) for detecting and quantifying sulfenic acids, disulfides, and free thiols in cells. *J Biol Chem* **2019**, *294*, 6522–6530.

[53] Shi, Y.; Fu, L.; Yang, J.; Carroll, K. S. Wittig reagents for chemoselective sulfenic acid ligation enables global site stoichiometry analysis and redox-controlled mitochondrial targeting. *Nat Chem* **2021**, DOI: 10.1038/s41557-021-00767-2.

[54] Lo Conte, M.; Carroll, K. S. The redox biochemistry of protein sulfenylation and sulfinylation. *J Biol Chem* **2013**, *288*, 26480–8.

[55] Woo, H. A.; Jeong, W.; Chang, T. S.; Park, K. J.; Park, S. J.; Yang, J. S.; Rhee, S. G. Reduction of cysteine sulfinic acid by sulfiredoxin is specific to 2-cys peroxiredoxins. *J Biol Chem* **2005**, *280*, 3125–8.

[56] Lo Conte, M.; Carroll, K. S. Chemoselective ligation of sulfinic acids with aryl-nitroso compounds. *Angew Chem Int Ed Engl* **2012**, *51*, 6502–5.

[57] Akter, S.; Fu, L.; Jung, Y.; Conte, M. L.; Lawson, J. R.; Lowther, W. T.; Sun, R.; Liu, K.; Yang, J.; Carroll, K. S. Chemical proteomics reveals new targets of cysteine sulfinic acid reductase. *Nat Chem Biol* **2018**, *14*, 995–1004.

[58] Requejo, R.; Hurd, T. R.; Costa, N. J.; Murphy, M. P. Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. *FEBS J* **2010**, *277*, 1465–80.

[59] Xiao, H.; Chatterjee, A.; Choi, S. H.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew Chem Int Ed Engl* **2013**, *52*, 14080–3.

[60] Bak, D. W.; Pizzagalli, M. D.; Weerapana, E. Identifying Functional Cysteine Residues in the Mitochondria. *ACS Chem Biol* **2017**, *12*, 947–957.

[61] Varnaitė, R.; MacNeill, S. A. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics* **2016**, *16*, 2503–2518.

[62] Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196*, 801–10.

[63] Lane, M. D.; Rominger, K. L.; Young, D. L.; Lynen, F. The Enzymatic Synthesis of Holotranscarboxylase from Apotranscarboxylase and (+)-Biotin. Ii. Investigation of the Reaction Mechanism. *J Biol Chem* **1964**, *239*, 2865–71.

[64] Chapman-Smith, A.; Cronan J. E., J. The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem Sci* **1999**, *24*, 359–63.

[65] Kim, D. I.; Birendra, K. C.; Zhu, W.; Motamedchaboki, K.; Doye, V.; Roux, K. J. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A* **2014**, *111*, E2453–61.

[66] Kim, D. I.; Jensen, S. C.; Noble, K. A.; Kc, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **2016**, *27*, 1188–96.

[67] Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **2018**, *36*, 880–887.

[68] Samavarchi-Tehrani, P.; Samson, R.; Gingras, A. C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Mol Cell Proteomics* **2020**, *19*, 757–773.

[69] Wei, W.; Riley, N. M.; Yang, A. C.; Kim, J. T.; Terrell, S. M.; Li, V. L.; Garcia-Contreras, M.; Bertozzi, C. R.; Long, J. Z. Cell type-selective secretome profiling in vivo. *Nat Chem Biol* **2021**, *17*, 326–334.

[70] Rosenthal, S. M.; Misra, T.; Abdouni, H.; Branon, T. C.; Ting, A. Y.; Scott, I. C.; Gingras, A. C. A Toolbox for Efficient Proximity-Dependent Biotinylation in Zebrafish Embryos. *Mol Cell Proteomics* **2021**, *20*, 100128.

[71] Mair, A.; Xu, S. L.; Branon, T. C.; Ting, A. Y.; Bergmann, D. C. Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. *Elife* **2019**, *8*, DOI: 10.7554/eLife.47864.

[72] Hopkins, C.; Gibson, A.; Stinchcombe, J.; Futter, C. Chimeric molecules employing horseradish peroxidase as reporter enzyme for protein localization in the electron microscope. *Methods Enzymol* **2000**, *327*, 35–45.

[73] Martell, J. D.; Deerinck, T. J.; Sancak, Y.; Poulos, T. L.; Mootha, V. K.; Sosinsky, G. E.; Ellisman, M. H.; Ting, A. Y. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* **2012**, *30*, 1143–8.

[74] Rhee, H. W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, *339*, 1328–1331.

[75] Hung, V.; Zou, P.; Rhee, H. W.; Udeshi, N. D.; Cracan, V.; Svinkina, T.; Carr, S. A.; Mootha, V. K.; Ting, A. Y. Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell* **2014**, *55*, 332–41.

[76] Chen, C. L.; Hu, Y.; Udeshi, N. D.; Lau, T. Y.; Wirtz-Peitz, F.; He, L.; Ting, A. Y.; Carr, S. A.; Perrimon, N. Proteomic mapping in live Drosophila tissues using an engineered ascorbate peroxidase. *Proc Natl Acad Sci U S A* **2015**, *112*, 12093–8.

[77] Paek, J.; Kalocsay, M.; Staus, D. P.; Wingler, L.; Pascolutti, R.; Paulo, J. A.; Gygi, S. P.; Kruse, A. C. Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **2017**, *169*, 338–349 e11.

[78] Lobingier, B. T.; Huttenhain, R.; Eichel, K.; Miller, K. B.; Ting, A. Y.; von Zastrow, M.; Krogan, N. J. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **2017**, *169*, 350–360 e12.

[79] Kim, D. I.; Roux, K. J. Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends Cell Biol* **2016**, *26*, 804–817.

[80] Lambert, J. P.; Tucholska, M.; Go, C.; Knight, J. D.; Gingras, A. C. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics* **2015**, *118*, 81–94.

[81] Vandemoortele, G.; De Sutter, D.; Moliere, A.; Pauwels, J.; Gevaert, K.; Eyckerman, S. A Well-Controlled BioID Design for Endogenous Bait Proteins. *J Proteome Res* **2019**, *18*, 95–106.

[82] Zhang, Y. B.; Howitt, J.; McCorkle, S.; Lawrence, P.; Springer, K.; Freimuth, P. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr Purif* **2004**, *36*, 207–16.

[83] Kim, T. K.; Eberwine, J. H. Mammalian cell transfection: the present and the future. *Anal Bioanal Chem* **2010**, *397*, 3173–8.

[84] Gingras, A. C.; Abe, K. T.; Raught, B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr Opin Chem Biol* **2019**, *48*, 44–54.

[85] Zheng, L.; Nagar, M.; Maurais, A. J.; Slade, D. J.; Parelkar, S. S.; Coonrod, S. A.; Weerapana, E.; Thompson, P. R. Calcium Regulates the Nuclear Localization of Protein Arginine Deiminase 2. *Biochemistry* **2019**, *58*, 3042–3056.

[86] Dong, J. M.; Tay, F. P.; Swa, H. L.; Gunaratne, J.; Leung, T.; Burke, B.; Manser, E. Proximity biotinylation provides insight into the molecular composition of focal adhesions at the nanometer scale. *Sci Signal* **2016**, *9*, rs4.

Chapter 2

Spatially Controlled Profiling of Redox Signaling During EGFR Activation

2.1 Introduction

Reactive oxygen species (ROS) are generated by all aerobic organisms as a byproduct of cellular respiration and other cellular processes. In addition to spontaneous production of ROS, H_2O_2 has been shown to play an important role in cell signaling.[1,2] The epidermal growth factor receptor (EGFR) pathway is an important and widely studied cell signaling pathway responsible for regulating growth, proliferation, and differentiation.[3] In addition to an increase in protein phosphorylation, activation of EGFR is also accompanied by a rapid increase in cellular ROS levels.[2] Cells maintain tight control over both H_2O_2 levels through redox buffering systems including catalase, peroxiredoxins (Prxs), and glutathione peroxidases (GPxs), and phosphorylation through protein phosphatases.[4] Upon engagement of EGFR, cells must overcome basal levels of phosphatase activity to allow signal propagation to occur. Protein tyrosine phosphatases (PTPs) are inactivated by oxidation by H_2O_2[5,6] and Src[7] and EGFR[2] kinases are activated by it. In addition, the lipid phosphatase PTEN is known to be reversibly inactivated by oxidation.[8] Conversely, phosphorylation of the redox buffering enzymes PrdxI[4,9] and PrdxII[9–11] allows localized H_2O_2 levels to rise as high as 10 µM during signaling.[1] Thus, the simultaneous inactivation of both phosphatases by oxidation and Prxs by phosphorylation act as a positive feedback loop necessary for sustained growth factor signaling.

The primary source of H_2O_2 used in cell signaling is the Nox family of nicotinamide adenine dinucleotide phosphate (NADPH) oxidases. The protopical NADPH oxidase is Nox2 which exists as a multi subunit transmembrane complex comprised of two multi-pass transmembrane subunits (CYBA and NOX2), and several cytosolic subunits (NCF1, NCF2, NCF4).[12] Nox enzymes function by transferring electrons from cytosolic NADPH to oxygen to form superoxide (O_2^-). The O_2^- is rapidly dismutated by SOD3 to H_2O_2.[12] The extracellular H_2O_2 then re-enters the cell either

through passive diffusion or through specialized aquaporin channels.[13] Two important features of ROS generated by Nox2 during signaling is that it is relatively short lived, the concentration of H_2O_2 peaks after 5 minutes and returns to base line after 20 minutes,[2] and it is localized to cytosolic regions adjacent to the cell membrane where the signal is propagating from.[14]

While the targets of EGFR dependent ROS have been studied on a global scale,[2,15] no study has yet achieved the spatial resolution required to identify local targets of ROS generated by Nox2 upon EGF stimulation. To that end, the goal of this study is to capture EGF-dependent protein-protein interactions (PPIs) with Nox2. The technique used to capture such interactions has two main requirements: (1) it must be capable of capturing transient PPIs in live cells, and (2) it must have high temporal resolution because EGFR mediated kinase activity lasts for ∼6 minutes.[16] Proximity-dependent biotinylation (BioID) represents an intriguing tool because it supports the profiling of protein interactions in live cells.[17–19] However, one significant limitation of first and second generation biotin ligases is that the labeling process takes place over the course of ∼24 hours.[17,18] In this study we explore methods of improving the temporal resolution of BioID towards, the development of a proteomics platform which supports redox proteomic profiling of EGF dependent interactors of Nox2.

2.2 Results and discussion

2.2.1 Use of photocaged biotins to increase the temporal control of BioID

2.2.1.1 First generation caged biotins

A possible strategy to improve the temporal resolution of BioID is to use biotin analogues which contain a photolabile protecting group. After being saturated with photocaged biotin, cells can be briefly irradiated with UV light to rapidly increase the intracellular concentration of free biotin and initiate biotin labeling. For a proof

of concept system, a previously reported BioID construct, histone H3A,[20] was fused to BirA* and stably expressed in 293T cells. A time course experiment with 0.1 µM supplemental biotin was performed and saturation of labeling was reached after ~24 hours (Figure 2.1). There are 2 possible rate limiting steps which could lead to the slow labeling process; the first is the diffusion of supplemental biotin into the cell, the second is the slow kinetic activity of the biotin ligase. Under the hypothesis that the rate limiting step is the diffusion of biotin into cells, we synthesized a panel of photocaged biotin (cBio) analogs (Figure 2.2A) in which the carboxylic acid in biotin is blocked by a ortho-nitrobenzyl (*o*NB) caging group. *o*NB derivatives are widely used for caging biomolecules such as neurotransmitters, and chemical probes due to their high uncaging efficiency and biocompatibility.[15,21–23] In a BioID workflow utilizing cBio, cells can be saturated with cBio without any background labeling, next the photolabile protecting group can then be rapidly removed by irradiation with UV light to initiate proximity-dependent labeling in a controlled manner. **cBio1-3** show UV and time dependent uncaging *in vitro* (Figure 2.2C). However, when added to cells expressing a BirA* fusion protein they show biotinylation even in the absence of UV irradiation (Figure 2.2B). The caging group on **cBio1-3** is attached via an ester bond so the most likely explanation for the UV independent uncaging is the activity of cellular esterases.

2.2.1.2 Second Generation Caged Biotins

To address the problem of esterase mediated uncaging, we synthesized a second generation of caged biotins using two different strategies. The use of nitroindoline groups have been widely used to cage neurotransmitters.[24–27] **cBio4** has a 4-methoxy-7-nitroindoline caging group attached by an amide bond. In addition the 5-nitroindoline isomer (**NPA1**) is generated as a byproduct in the synthesis of **cBio4** (Figure 2.3A, Scheme 2.2). **NPA1** is not photoactive and was included as a negative

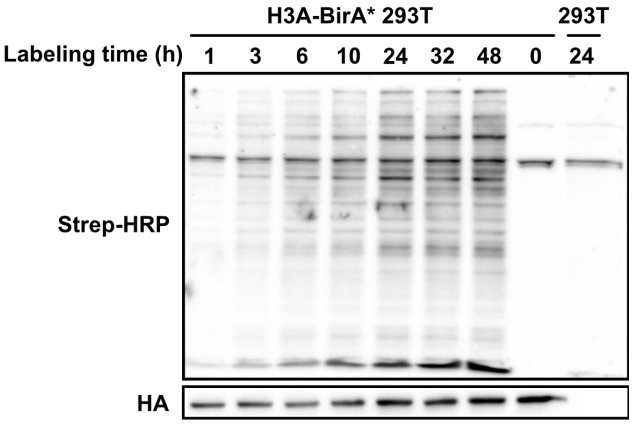


Figure 2.1: Timecourse experiment with BirA*. HEK 293T stably expressing histone H3A fused to BirA*-HA were treated with 0.1 µM biotin from 0 to 48 hours. A 293T cell line not expressing a biotin ligase was included as a negative control.

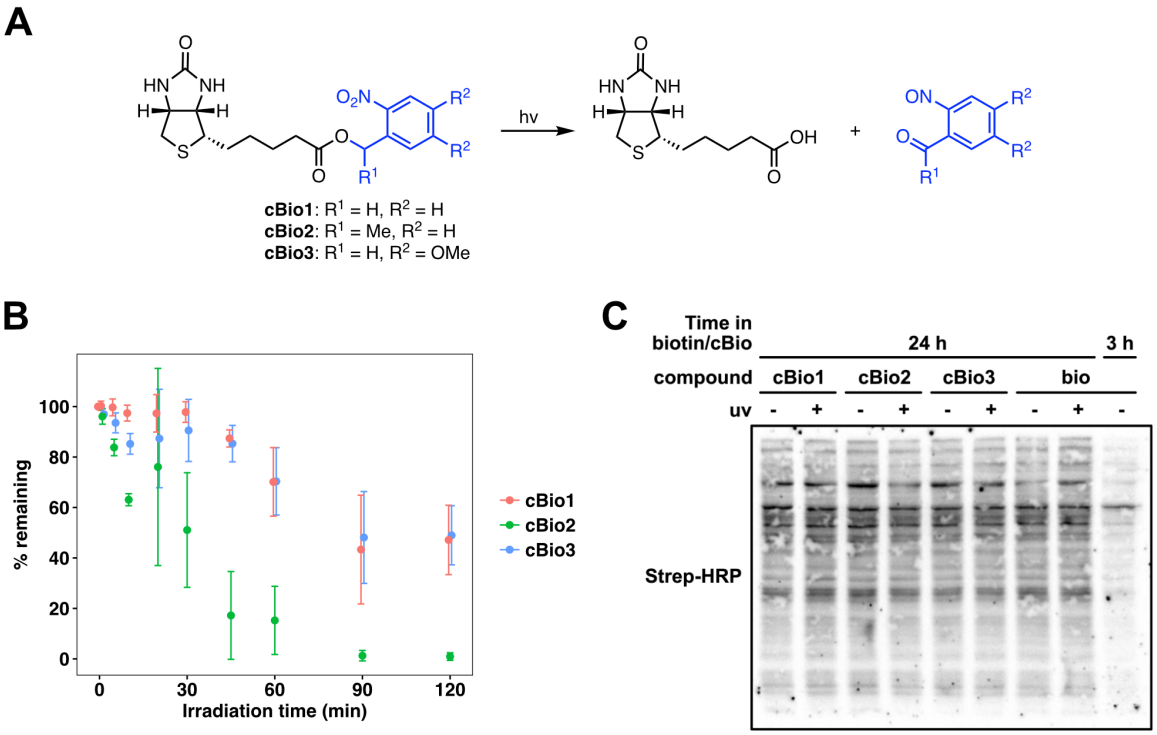


Figure 2.2: Characterization of first generation caged biotins. **(A)** Uncaging reaction of **cBio1-3**. **(B)** An aqueous solution of 400 µM **cBio1-3** were irradiated with UV light for the the indicated time. The percentage of caged compound remaining was quantified by LC/MS. Error bars represent the standard deviation of 3 measurements. **(C)** 0.1 µM **cBio1-3** were added to HEK 293T stably expressing H3A-BirA*-HA and incubated at 37 °C and 5 % CO_2 for 24 hours. Cells were then irradiated with UV light for 5 minutes, incubated at 37 °C for an additional 10 minutes then harvested.

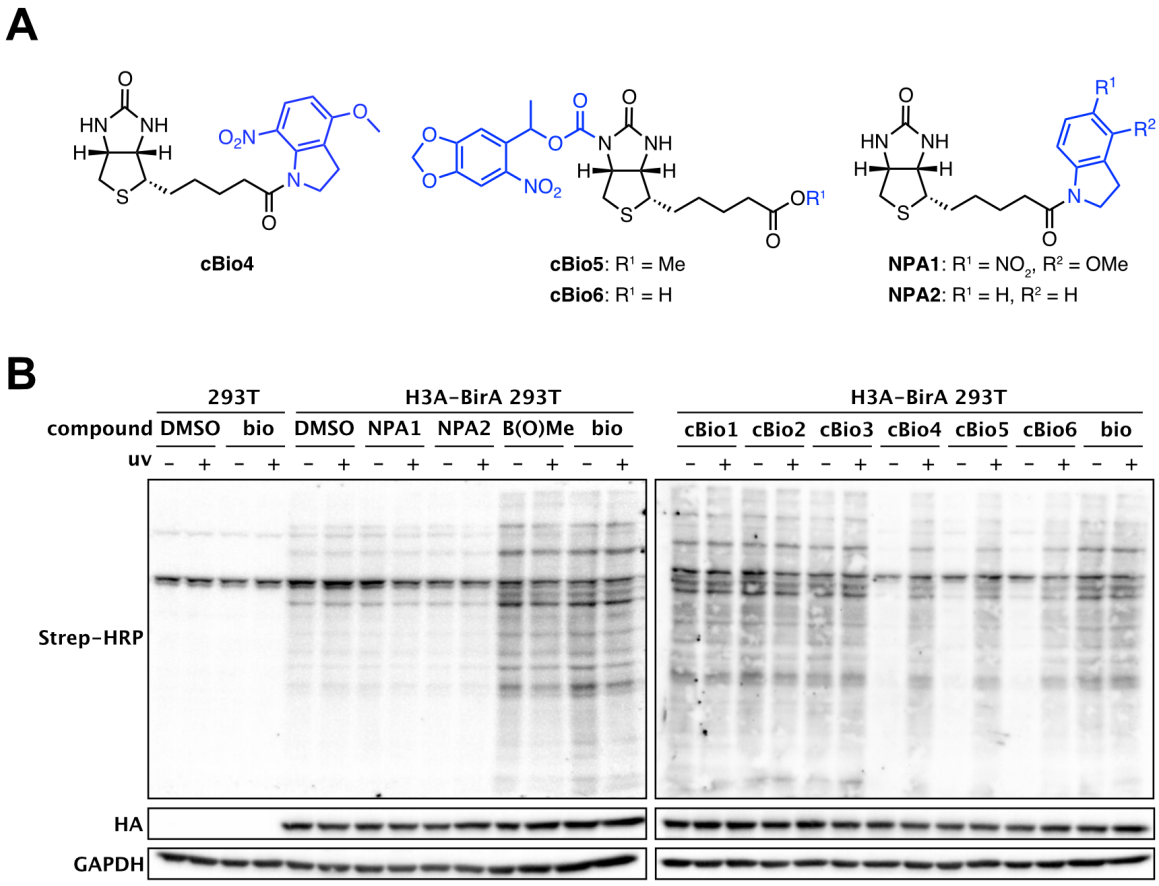


Figure 2.3: Second generation caged biotins allow for the control of biotin labeling in live cells. **(A)** Structures of **cBio4-6** and non photoactive (NPA) biotins 1 and 2. **(B)** 0.1 µM of various cBio analogues were added to H3A-BirA*-HA 293T cells and incubated for 24 hours. Cells were then irradiated with UV light for 5 min and incubated at 37 °C for an additional 24 hours. Labeled proteins were then detected by biotin blot.

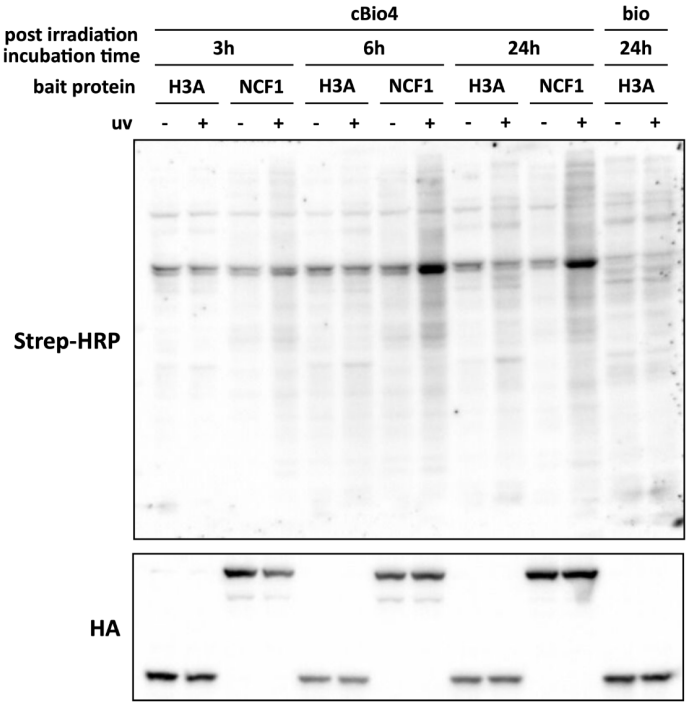


Figure 2.4: 293T cells stably expressing H3A or NCF1 fused to BirA*-HA were incubated with 0.1 µM **cBio4** for 24 hours, then irradiated with UV light for 10 minutes, and harvested at the indicated time points.

control. The use of a photocaging group to control the interaction between biotin and avidin has been previously reported.[28,29] **cBio5-6** have a *o*NB caging group attached to the biotin uredo ring. The carboxylic acid on **cBio5** has a methyl ester protecting group and is an intermediate in the synthesis of **cBio6** (Figure 2.3A, Scheme 2.3). We included the methyl ester intermediate to test whether it would show improved cell permeability. The entire panel of biotin analogues were tested in a BioID cell line. As expected **NPA1-2** showed a low level of labeling comparable level of labeling to DMSO. **cBio1-3** again showed UV independent uncaging comparable to both biotin and biotin methyl ester. **cBio4-6** showed UV dependent uncaging with UV irradiated cells showing a comparable level of labeling to unmodified biotin (Figure 2.3B). However, when a timecourse experiment was performed with **cBio4**, cells still needed a full 24 hours after UV irradiation to achieve a saturation of biotin labeling (Figure 2.4). Unfortunately, the initial hypothesis, that the rate limiting step in BioID was

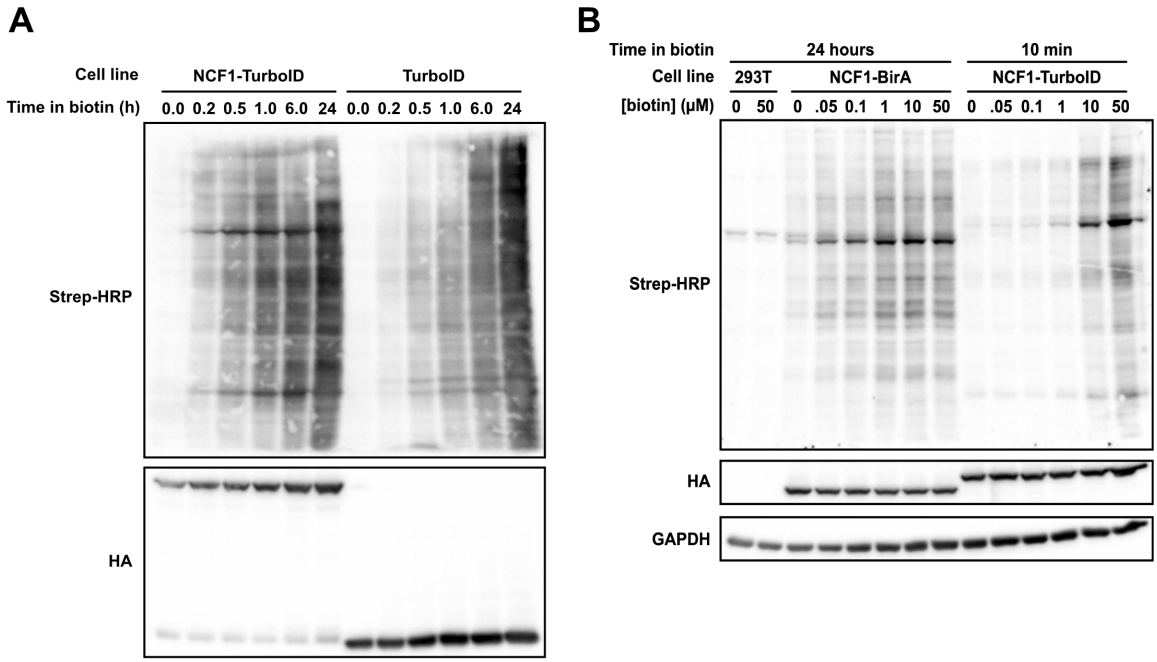


Figure 2.5: Time and concentration course experiments with TurboID. **(A)** 293T cells stably expressing NCF1 fused to TurboID-HA and the TurboID-HA control were treated with 50 µM biotin and harvested after 10, or 30 minutes, or 1, 2, or 24 hours. **(B)** 293T cells expressing NCF1 fused to either BirA*-HA or TurboID-HA were treated with 0-50 µM biotin for 24 hours for BirA*, or 10 minutes for TurboID.

the diffusion of biotin into the cell was incorrect. In order to improve the rate of biotinylation a biotin ligase with faster catalytic activity is necessary.

2.2.2 Generation of A431 cell lines stably expressing TurboID

A biotin ligase with higher catalytic efficiency, named TurboID has recently been developed through directed evolution.[30] TurboID shows saturation of labeling as fast as within 10 minutes.[30] Two TurboID constructs were expressed in HEK 293T cells; one with the Nox cytosolic subunit NCF1, fused to TuboID-HA, and another with just the free TurboID-HA. Both constructs showed robust biotinylation within 10 minutes and labeling intensity increased over the course of 24 hours (Figure 2.5A). TurboID shows a similar level of biotinylation in as fast as 10 minutes, compared to 24 hours for the second generation BirA*[18] (Figure 2.5B). Interestingly although TurboID was much faster, it was only able to achieve a comparable level of labeling

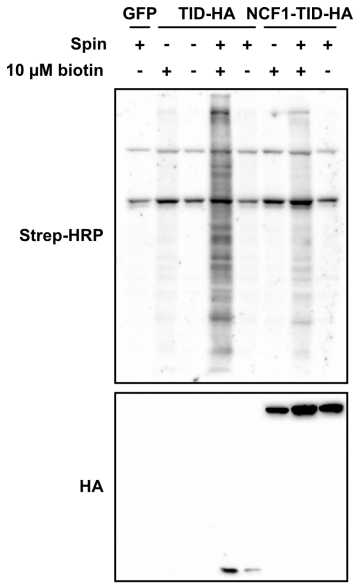


Figure 2.6: Biotin blot showing biotin labeling and HA western blot showing fusion protein expression in A431 cells transduced with or without spinoculation.

to BirA* at a mininum concentration of 10 µM supplemental biotin, whereas BirA* showed biotinylation activity at as low as 50 nM (Figure 2.5B). The wild type *E. coli* BirA was used as the starting point for the directed evolution of TurboID[30] which requires 10–50 µM supplemental biotin for proximity biotinylation.[17,31] However the second generation BirA* was based off the the *Aquifex aeolicus* BirA which is active at a much lower biotin concentration.[18] The *A. aeolicus* BirA even shows labeling in the absence of suplemental biotin with some constructs which necessitates culturing cells in biotin depleted media prior to initiating biotin labeling.[18]

Next, we sought to generate an A431 cell line expressing TurboID fused to multiple Nox subumits. A431 cells are an epidermoid carcinoma line which naturally overexpresses EGFR, and are widely used as a model cell line for studies of the EGFR pathway.[2,4,32] A431 cells are a suitable stable transduction host,[32] but do not express recombinant proteins as readily as 293T cells. To improve the transduction and expression efficiency, we transduced A431 cells with a modified spinoculation protocol.[33] Spinoculation involves centrifuging cells at $800 \times g$ after adding adding virus particles to increase the concentration of virus at the surface of target cells, resulting in

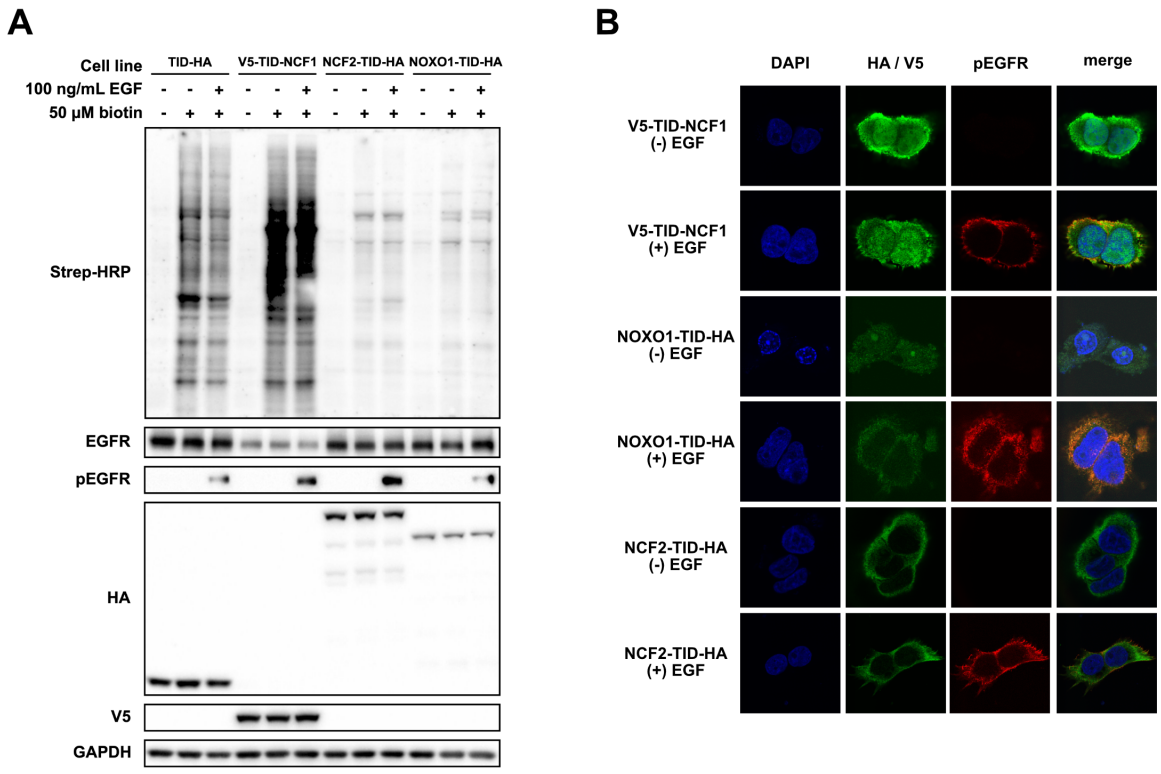


Figure 2.7: Stimulation of TurboID cell lines with EGF. **(A)** A431 cells were incubated for 10 minutes in 50 µM biotin and the presence of absence of 100 ng mL^{-1} EGF. Whole cell lysates were analyzed by western blot with streptavidin-HRP and antibodies for EGFR, phospo-EGFR, HA, V5, and GAPDH. **(B)** A431 cells were incubated for 5 minutes in the presence of absence of 100 ng mL^{-1} EGF. Cells were then fixed, and subjected to immunofluorescence staining with antibodies against the HA or V5 tags on TurboID (green) and phosphorylated EGFR (red); nuclei were stained with DAPI (blue).

more cells being infected.[33] A431 cells which were infected by spinoculation showed significantly higher levels of fusion protein expression and biotin labeling (Figure 2.6).

Four TurboID constructs were generated and expressed in A431 cells: (1) a TurboID control with an N-terminal HA tag, (2) the Nox2 organizer subunit NCF1 with an N-terminal TurboID-HA fusion, (3) the Nox2 activator subunit NCF2 with a C-terminal V5-TurboID fusion, and (4) the Nox1 organizer subunit NOXO1 with an N-terminal TurboID-HA fusion. Each cell line shows varying levels of biotinylation and all show EGF dependent phosphorylation of EGFR (Figure 2.7A). Immunofluorescence analysis shows that each construct co-localizes with phosphorylated EGFR upon stimulation with 100 ng mL^{-1} EGF (Figure 2.7B).

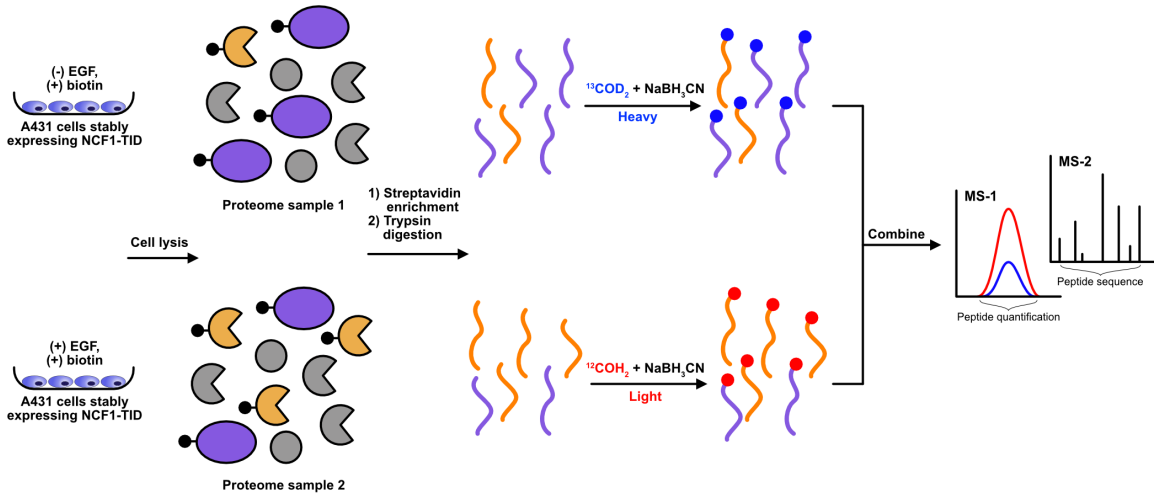


Figure 2.8: Workflow to identify protein targets of Nox2 ROS. A431 cells stably expressing Nox2 TurboID constructs are treated with 50 µM biotin and (+/-) 100 ng mL^{-1} EGF for 15 minutes at 37 °C then harvested. Whole cell lysates are streptavidin enriched, trypsinized on bead, and dimethylated with isotopically heavy and light formaldehyde. Proteins which were labeled by TurboID are identified by LC-MS/MS.

To identify proteins that selectively interact with Nox2 during EGF stimulation, the cell line for each construct was treated with 50 µM biotin in the presence or absence of 100 ng mL^{-1} EGF. The biotinylated proteins are enriched on streptavidin-agarose beads, digested with trypsin, and subjected to reductive dimethylation (ReDiMe).[34] The digest from the cells treated with EGF was dimethylated with heavy formaldehyde and the no EGF control with light formaldehyde (Figure 2.8). The light to heavy ratio ($R_{L/H}$) represents the abundance ratio between proteins associating with the bait protein in the presence or absence of EGF. Thus, a high $R_{L/H}$ indicates that the labeled protein selectively interacts with the bait during EGF stimulation.

We performed 2 technical replicates of the experiment descried in Figure 2.8. The $R_{L/H}$ in replicate 2 is plotted vs. the ratio in replicate 1 (Figure 2.9, Table 2A.1). Proteins with elevated $R_{L/H}$ in both replicates are in the upper right corner of each panel. A total of 26 proteins had $R_{L/H} > 2$ which were consistent across both replicates. Only 1 protein fell into this category in the TID-HA dateset while 11, 9, and 5 were in the V5-TID-NCF1, NCF2-TID-HA, and NOXO-TID-HA datasets

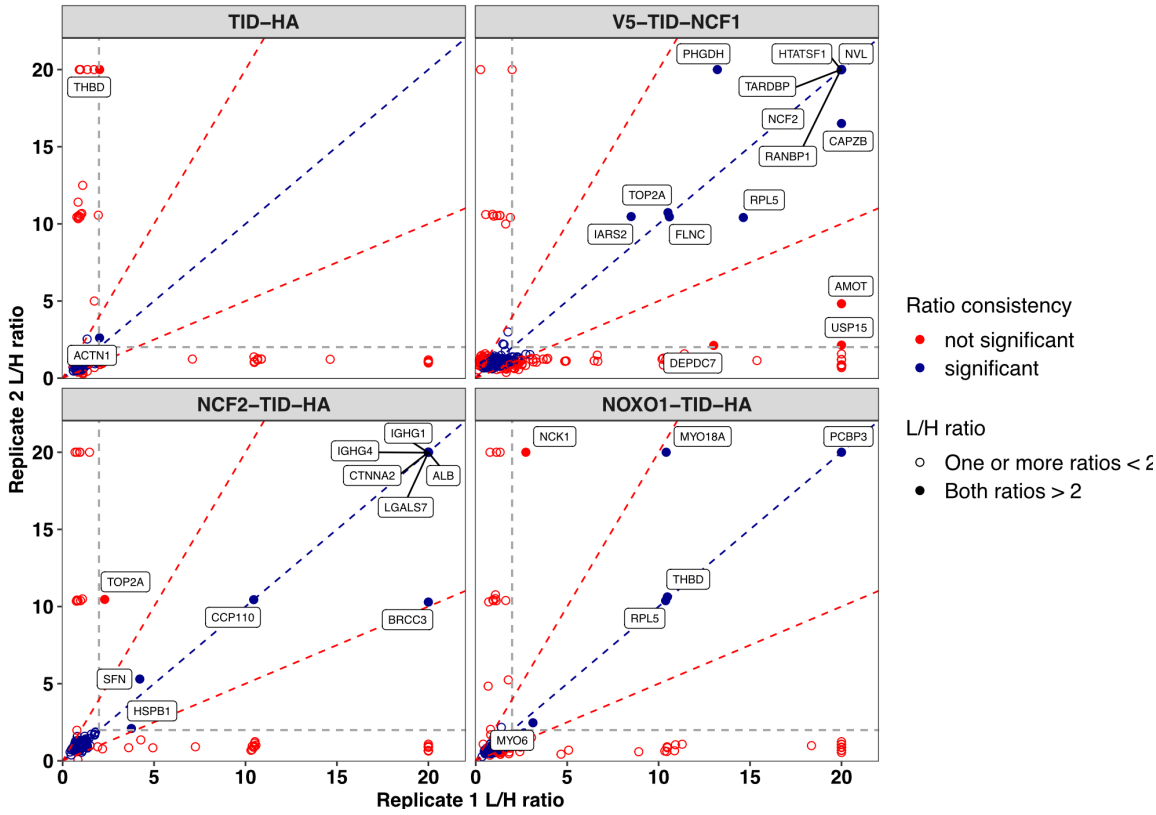


Figure 2.9: Identification of protein targets of Nox2 ROS. Protein $R_{L/H}$ in the second technical replicate (y-axis) are plotted vs. the first technical replicate (x-axis). Each panel contains data from a different bait protein. The red dotted lines have slopes of 2 and 0.5 and the 2 vertical and horozintal grey lines are at a $R_{L/H}$ of 2 in each replicate. Proteins are considered to be reproducible EGF dependent interactors with the bait protein if they have a $R_{L/H} > 2$ in both replicates and are between the 2 red dotted lines.

respectively. Encouragingly, NCF2 had a $R_{L/H}$ of 20 in both V5-TID-NCF1 replicates which indicates that the NCF1 fusion protein retains the activity of the wild type. A number of proteins with involvement in cell cycle and mitogen signaling were also identified. Most notibly, BRCC3 is a deubiquitinase and a subunit in the BRCA1/2 complex, a well characterized tumor suppressor gene.[35,36] In addition, SFN is a 14-3-3 scaffolding protein which binds to phosphoserine and phosphothreonine residues on receptor tyrosine kinases (RTKs)[37] and has been show to stabilize EGFR and MET by recruiting deubiquitinases to prevent their degradation.[38] A ubiquitin ligase, TRIM25 known to regulate SFN was also identified in multiple datasets,[39] but it did not have an elevated $R_{L/H}$.

2.2.3 Mass spectrometry method development

2.2.3.1 Dual enrichment strategy

With the A431 TurboID cell lines in hand, we next sought to identify cysteine residues which are targeted by Nox2 ROS. The oxidative isotope-coded affinity tags (OxICAT) method uses cysteine reactive probes with differential isotopic tags to quantify the stoichiometry of cysteine labeling within a sample.[41–43] However, because the TurboID and OxICAT sample preparations both use a biotin-streptavidin enrichment step, it is necessary to modify the workflow when combining both approaches to allow for both the enrichment of proteins labeled by TurboID and cysteine containing peptides. To maximize the coverage of cysteine residues we first tested a dual enrichment workflow in which the first enrichment is performed for proteins labeled by the TurboID bait construct and the second enrichment is performed for cysteine containing peptides (Figure 2.10). A test sample preparation was performed with 2 and 4 mg of lysate from cells expressing TurboID treated with 50 µM supplemental biotin following the workflow in figure 2.10. The percent oxidation of a total of 6 and 376 cysteines in the 2 and 4 mg samples respectively were quantified, with only 4

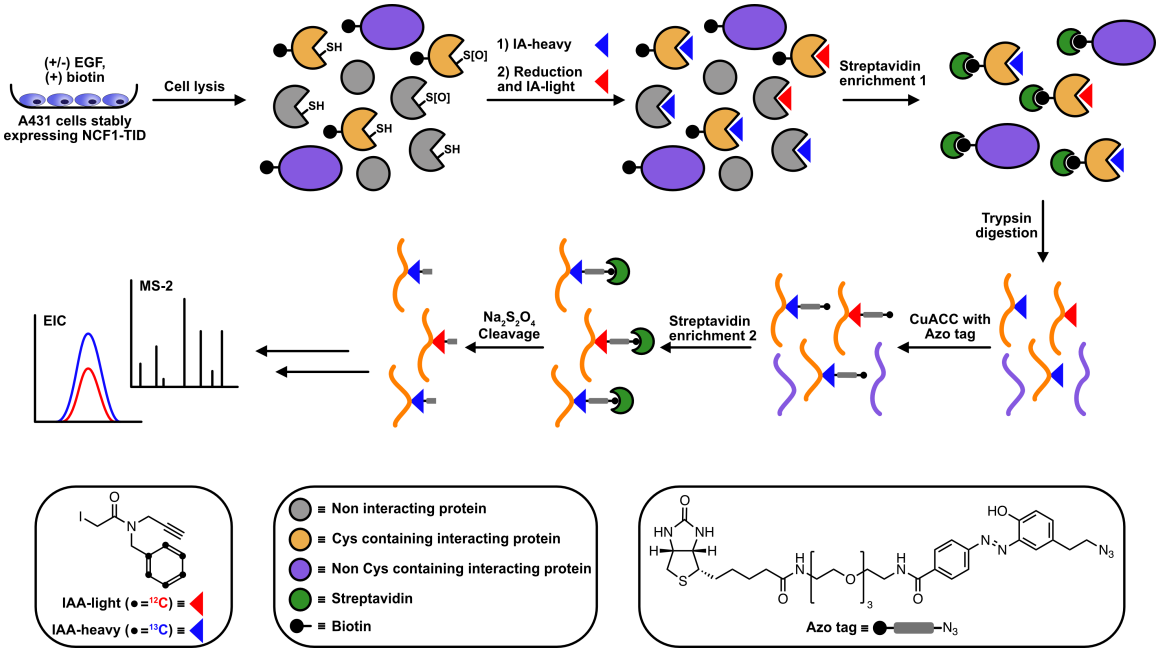


Figure 2.10: Dual enrichment workflow. A431 cells expressing a Nox TurboID construct are stimulated with EGF then lysed under denaturing conditions in the presence of 10 mM iodoacetamide-alkyne (IAA) light to cap free cysteine thiols, oxidized cysteines are reduced with TCEP then caped with 10 mM IAA-heavy. The first of 2 streptavidin pull downs is performed to enrich proteins biotinylated by the TurboID construct, then digested with trypsin. Copper-catalyzed azide-alkyne cycloaddition (CuAAC) is performed on the tryptic digest with a chemically cleavable biotin azide tag, followed by a second streptavidin enrichment and chemical cleavage of the biotin tag to obtain cysteine containing peptides.

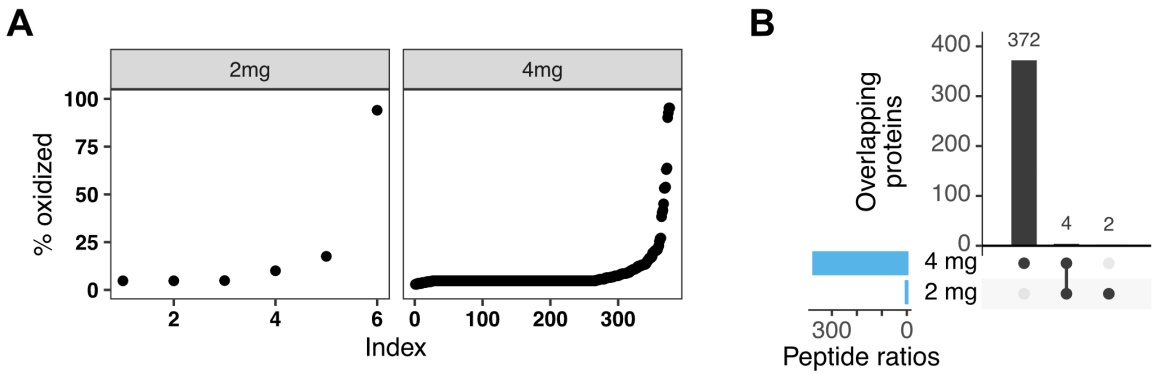


Figure 2.11: Cysteines identified by the dual enrichment workflow. **(A)** Percent oxidation of cysteine residues identified in the 2 and 4 mg samples. **(B)** UpSet[40] depicting the number of quantified cysteines which were unique and overlapping between each sample.

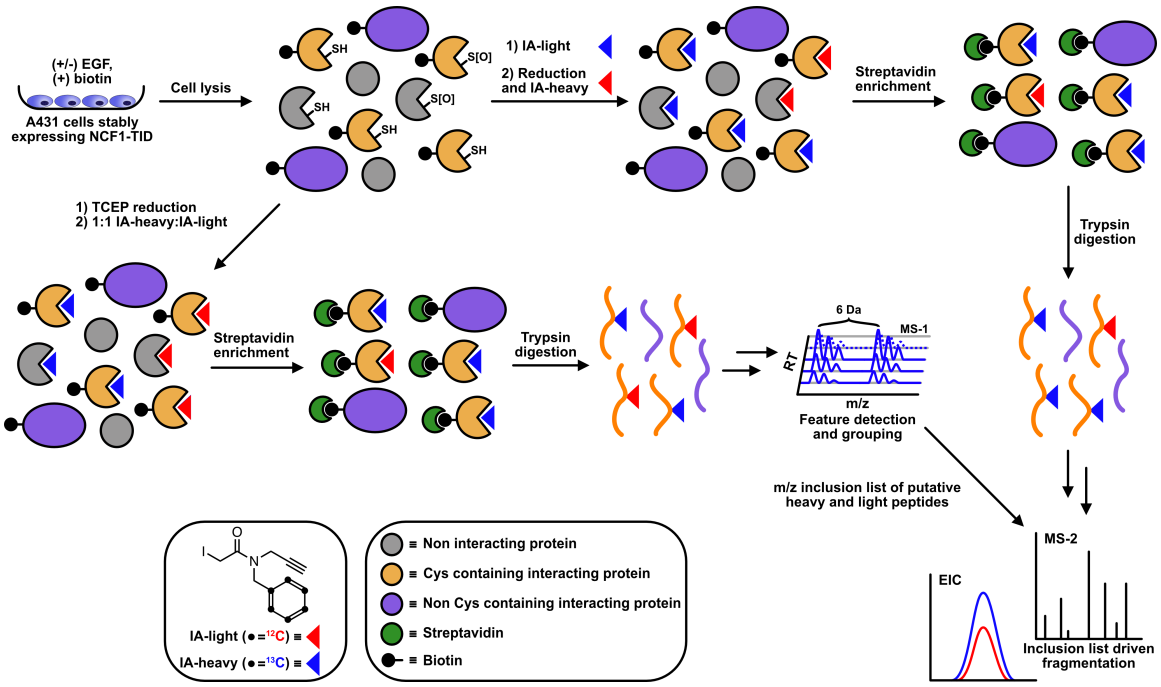


Figure 2.12: Isotopically linked feature detection and targeted fragmentation workflow. A431 cells expressing a Nox TurboID construct are stimulated with EGF, and split into 2 fractions. The first is labeled with a 1:1 ratio of IAA-light and heavy, enriched on streptavidin beads, trypsinized and analyzed by LC-MS/MS. A list of putative cysteine containing peptides is generated by performing feature detection and linking isotopically related peptides on the MS1 level data. The list of isotopically linked features is used to generate a mass list to target the MS2 fragmentation to peptides of interest in a second sample.

cysteines being identified in both samples (Figure 2.11A,B). Thus, a dual enrichment workflow using TurboID and OxICAT is feasible, but it requires 4 or more mg of cell lysate per sample.

2.2.3.2 Isotopically linked feature detection and targeted fragmentation

A significant disadvantage of a dual enrichment workflow is that ∼2 mg of both IAA-light and heavy is required per sample. Because IAA must be prepared in house using costly starting materials,[43] we sought an alternative approach which supports the use of commercially available reagents. One technique to address the problem of dual enrichment is to eliminate one of the enrichment steps all together. In such a workflow, the first enrichment for proteins biotinylated by the bait construct would be

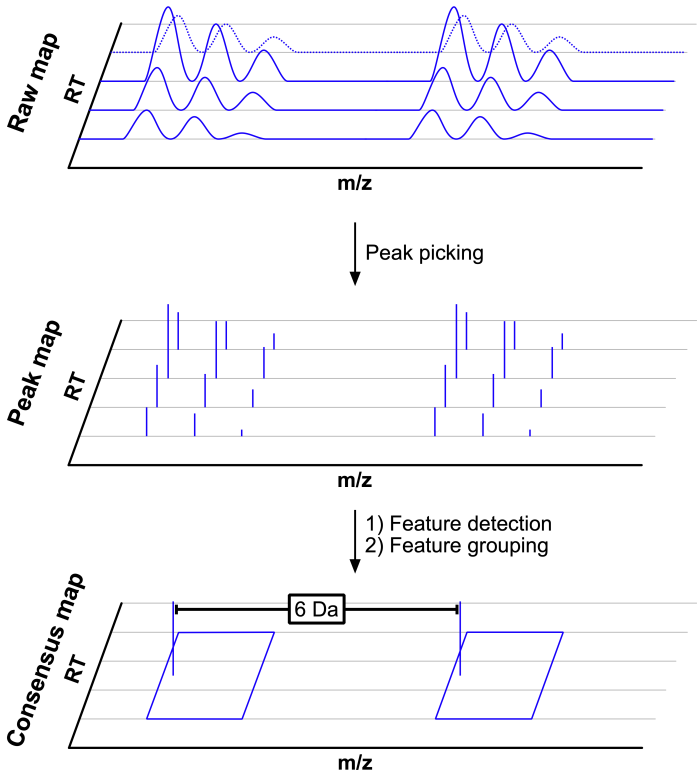


Figure 2.13: Steps MS1 data reduction. Profile MS1 peaks are centroided, then peaks from the same peptide are grouped by the m/z and RT dimensions to form features. Co-eluting features which are 6 Da apart in the m/z dimension are grouped to form consensus features.

retained and cysteine containing peptides would be targeted for MS2 fragmentation using an inclusion list. The inclusion list containing putative cysteine containing peptides would be obtained by analyzing a portion of the sample of interest which has been labeled with a 1:1 ratio of a heavy and light cysteine reactive reagent. Then a list of masses of peptides which co-elute and have a mass difference equivalent to that of the tag would be constructed by analyzing the MS1 level data (Figure 2.12). Because MS1 level data is comprehensive, all cysteine containing peptides should be visible at that level, even if a MS2 scan was not obtained for a given peptide. The MS2 coverage of cysteine containing peptides could be maximized by analyzing a second portion of the same sample with a mass list to target fragmentation of those peptides.

In the analysis of mass spectrometry data, feature detection is a data reduction

Parent file	Sequest PSMs	Consensus features	PSMs mapped	Single id	Multiple id identical	Multiple id divergent	TPR
IA HL 1-1 01	0	140	0	0	0	0	NA
IA HL 1-1 02	363	1099	182	69	108	5	50.14
IA HL 1-1 03	1216	2634	904	278	609	17	74.34
IA HL 1-1 04	1221	2583	884	283	561	40	72.40
IA HL 1-1 05	191	153	19	9	10	0	9.95
Total	**2991**	**6609**	**1989**	**639**	**1288**	**62**	**66.50**

Table 2.1: True positive rate (TPR) for MS1 features compared to Sequest identifications. The number of peptide-spectrum matches (PSMs), features, and PSMs matches to features in each of the 5 sample fractions are shown. PSMs which are mapped to features can be broken down into 3 categories: (1) in "single id" matches, only a single PSM was mapped to a feature, (2) in "multiple ID identical" matches, multiple PSMs with the same sequence are mapped to a feature, and (3) in "multiple ID divergent" matches, multiple PSMs with differing sequences are mapped to the same feature.

technique where isotopic peaks belonging to the same peptide are linked in the m/z dimension and isotopic clusters eluting in adjacent scans are linked in the retention time (RT) dimension.[44] There are various open source software tools which implement feature detection algorithms, including the Trans-Proteomic Pipeline,[45] MaxQuant,[46] and OpenMS.[44,47] Because the individual tools in OpenMS are modular, highly extensible, and available as both stand alone command line tools as well as a C++ or Python library, we chose to use it to perform the custom feature detection step.

To evaluate the utility of mass list targeted fragmentation of cysteine containing peptides, we treated A431 cells expressing a Nox TurboID construct with 50 µM biotin in the presence or absence of 100 ng mL^{-1} EGF. Two replicates of a sample where half the sample is labeled with 100 µM IAA heavy and half with 100 µM IAA light. In the evaluation of this method we used IAA heavy and light, but any cysteine reactive reagent with a stable isotope label, such as N-ethylmaleimide (NEM) H_5 and D_5, would also be supported because the alkyne on IAA is not required. Biotinylated proteins were enriched on streptavidin beads, digested with trypsin. The entire tryptic digest from the first replicate was analyzed by LC-MS/MS. MS1 peaks were centroided, then the OpenMS centroid feature finder[44,47] tool was used to group

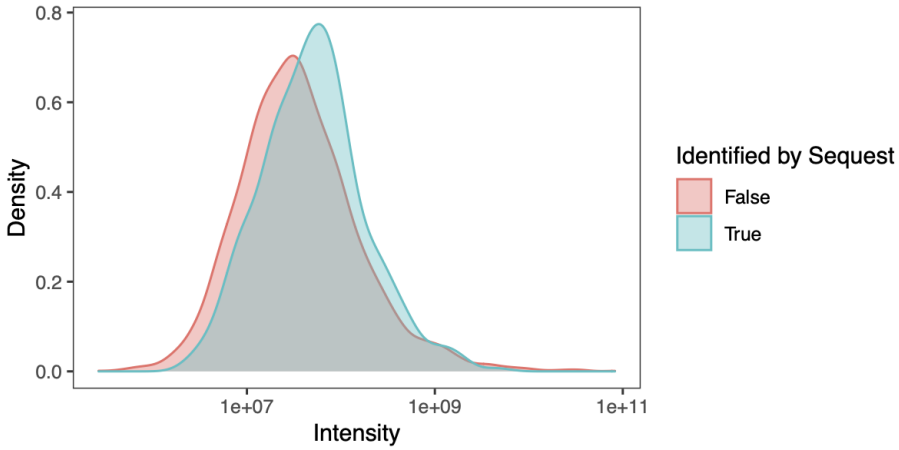


Figure 2.14: Kernel density estimate (KDE) plot showing the distribution of feature intensities for peptide features which were matches or not matches to PSMs identified by Sequest.

peptide features into a feature map data structure. Co-eluting features in the RT dimension which are 6 Da apart in the m/z dimension are grouped together and stored in a consensus map data structure (Figure 2.13). The MS2 level data in the first sample were also analyzed with Sequest[48] to determine how peptides in the MS1 level data were also identified by database searching. The number of PSMs identified by Sequest in each on MudPIT fraction[49] are shown in Table 2.1. Overall 66.50 % of peptides identified by Sequest were also identified in the feature detection data. The most likely reason that many features were not also identified by Sequest is that the Sequest IDs were biased towards higher intensity features (Figure 2.14), because those are the most likely to be sampled for MS2 fragmentation. The benefit of using a mass list based off putative cysteine containing peptides is that fragmentation would also be targeted to lower abundance peptides.

Next, the second sample replicate was run with a mass list comprised of features identified in the first replicate which met the requisite m/z and RT criteria. Overall, 368 cysteine containing peptides were identified which were found in both samples, 74 additional peptides were unique to the sample run with a mass list, and 148 additional peptides were unique to the no mass list run (Figure 2.15). Surprisingly,

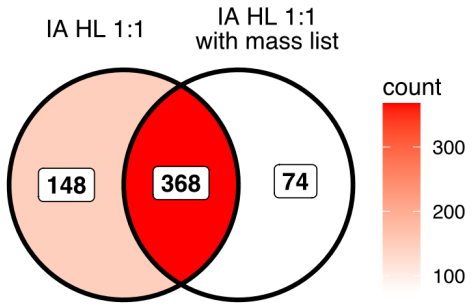


Figure 2.15: Numbers of cysteine containing peptides identified with and without a mass list targeting heavy and light peptide pairs.

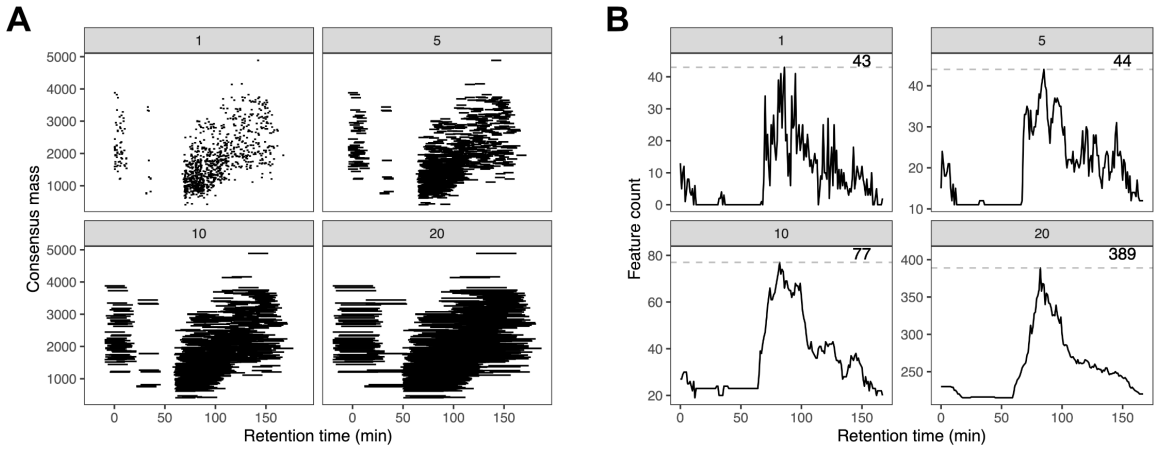


Figure 2.16: Representative retention time (RT) and m/z distribution of consensus features. **(A)** Mass of consensus features identified by with a RT window of 1, 5, 10, and 20 minutes. **(B)** Number of consensus features at a given RT with varying RT windows for consensus features.

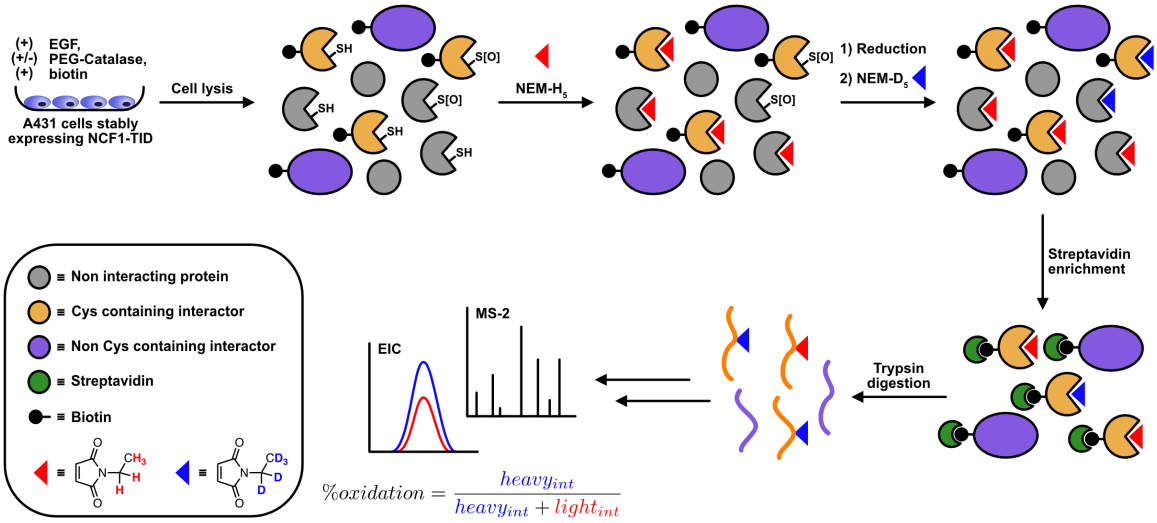


Figure 2.17: TurboID OxICAT workflow. Nox TurboID A431 cells are stimulated with 100 ng mL^{-1} EGF and 1000 U mL^{-1} PEG-catalase. Lysates are labeled sequentially with 10 mM NEM-H_5 and NEM-D_5. A streptavidin pull down is then performed to enrich for proteins which are biotinylated by TurboID and digested with trypsin on bead. The full tryptic digest is then analyzed by LC-MS/MS.

more cysteines were quantified in the sample run without a mass list. To investigate why using a mass list did not increase cysteine coverage, we examined the m/z and RT distribution of features with a range of RT tolerances around each feature. Most features eluted between 60 and 160 minutes in the representative MudPIT fraction we examined (Figure 2.16A). The maximum number of features were 43, 44, 77, and 389 with feature retention time windows of 1, 5, 10, and 20 minutes respectively (Figure 2.16B). Even if the mass list were to be broken down in to several segments over the course of the sample acquisition, the high complexity and number of target masses is likely beyond what is possible in a data dependent acquisition (DDA) experiment.

2.2.4 Turbo-OxICAT

To identify cysteine residues which are targeted by Nox2 ROS we adopted a single enrichment strategy where proteins labeled by TurboID are pulled down by streptavidin enrichment and the entire tryptic digest is analyzed by LC-MS/MS (Figure 2.17). Non cysteine containing peptides are then removed during the data analysis

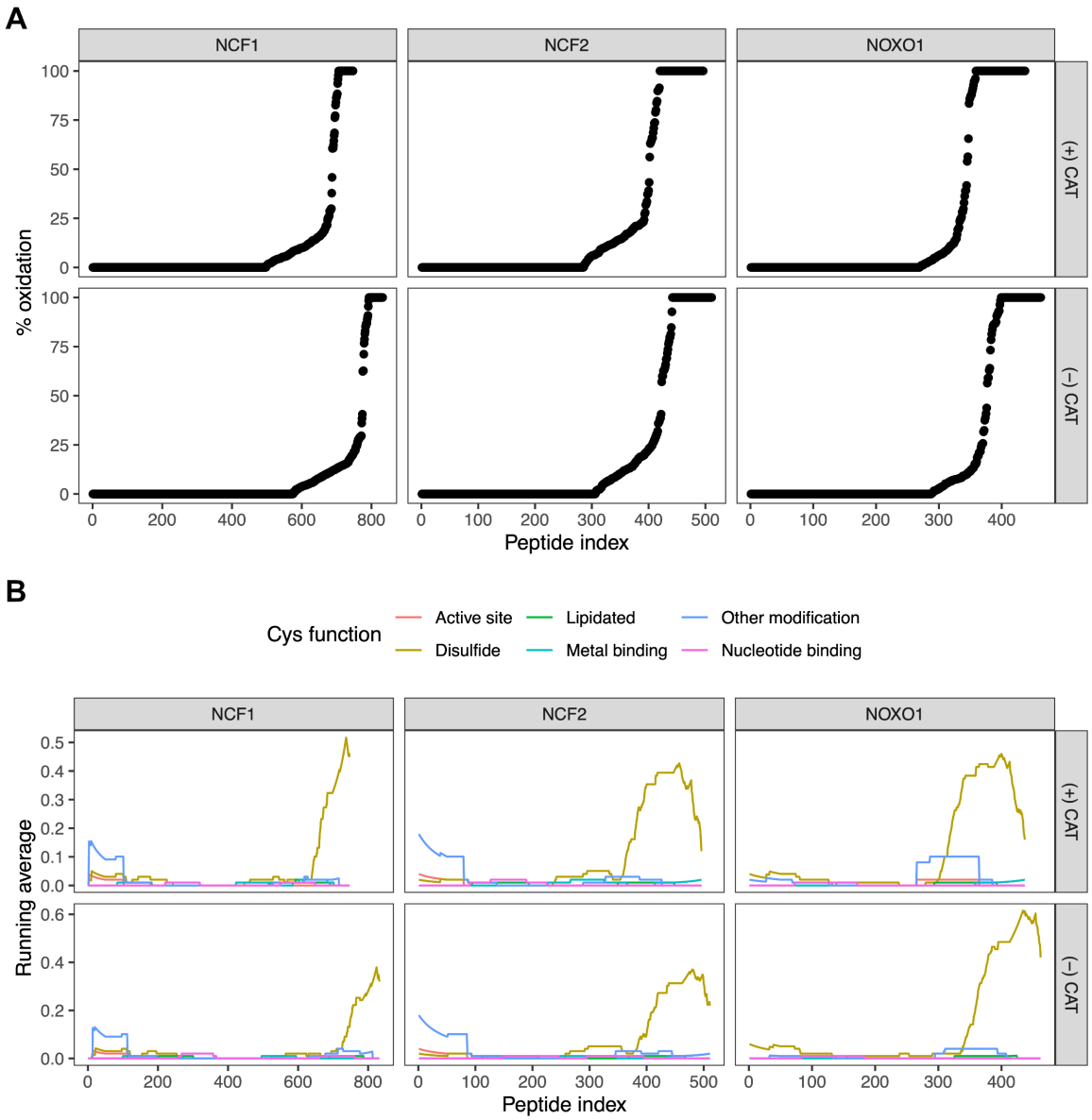


Figure 2.18: **(A)** Percent oxidation of proteins pulled down with NCF1, NCF2, and NOXO1. The cell line expressing each bait protein was treated with 100 ng mL^{-1} EGF, 50 µM biotin, and (+/−) 1000 U mL^{-1} PEG-catalase. **(B)** Moving average (window of 50) of cysteine functional annotations as percent of total peptides.

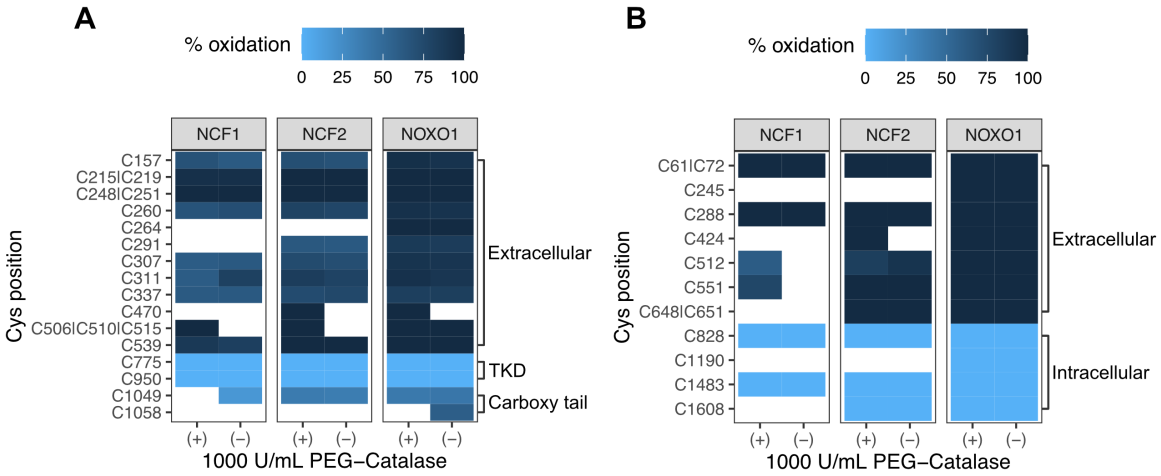


Figure 2.19: Percent oxidation of cysteines identified in each bait protein data set in **(A)** EGFR and **(B)** integrin beta-4 (ITGB4) with the locations of extracellular and intracellular domains shown.

process. A challenge with the workflow is that there could be a change in both the binding partners of the TurboID bait as well as the level of oxidation on those proteins. In an attempt to reduce the sample variability all cells were treated with EGF and supplemental biotin and cells were either treated or not treated with 1000 U mL^{-1} polyethylene glycol (PEG)-catalase. Pre-treatment of cells with PEG-catalase can be used to attenuate EGFR activation by eliminating extracellular H_2O_2.[50,51] Our hypothesis is that the bait protein binding partners of the (+/−) PEG-catalase datasets should remain relatively consistent, but the oxidation profile of those proteins should be altered.

The NCF1, NCF2, and NOXO1 TurboID A431 cell lines were treated with 50 µM biotin, 100 ng mL^{-1} EGF, and (+/−) 1000 U mL^{-1} PEG-catalase in technical triplicate. The level of oxidation on between 500 and 800 peptides were quantified in each of the 6 samples. No trends were apparent in the number of peptides or overall levels of oxidation in each sample (Figure 2.18A, Table 2A.2). One important difference between OxICAT and activity or reactivity based methods is that all cysteines, not just reactive ones, will be labeled by OxICAT reagents. Therefore, we would expect to see more structural disulfide bonds with high oxidation percentages. We used a

custom python script[52] to query the universal protein resource (UniProt) database and retrieve functional annotation information for each cysteine we identified. The running average percentage of proteins which fall into several functional categories are shown in Figure 2.18B. For all samples, annotated disulfides comprised up to half of half of all cysteines with an oxidation percentage over 75 %. For NCF1 and NCF2, several nitrosylated, ADP-ribosylated, and succinylated cysteines on GAPDH were found with lower oxidation percentages.

EGFR and ITGB4 are both homologous, single pass transmembrane proteins which have transmit signals across the cell membrane. 17 and 11 distinct cysteine containing peptides were identified on EGFR and ITGB4 respectively, and their membrane topology and oxidation profile provides an additional means of validating the OxICAT method because the cytosol is a reducing environment and the extracellular space is oxidizing. A stark transition in percent oxidation is apparent between the extracellular and intracellular transition in both proteins (Figure 2.19). There is also a slight increase increase between the carboxyl tail in EGFR relative to fully reduced cysteines in the tyrosine kinase domain (TKD) (Figure 2.19A). Cysteines were essentially fully oxidized or reduced in ITGB4 depending on what side of the cell membrane they are located (Figure 2.19B).

Ultimately, the goal of this study is to identify sites on proteins which are selectively oxidized in EGF stimulated cells. To identify proteins with a higher level of oxidation in the (-) PEG-catalase sample, we plotted the percent oxidation in the (-) PEG-catalase sample vs. the (+) PEG-catalase sample (Figure 2.20, Table 2A.2). EGFR C311 is located in EGFR extracellular domain and showed 60 vs. 82 % oxidation in the NCF1 (+) vs. (-) PEG-catalase treatment. Another residue in EGFR, C797 is already an established site of regulatory sulfenylation,[2] however we did not observe it in our data. β-integrins are involved in transmitting signals across the cell membrane via the activation of tyrosine kinases.[53] A peptide containing both C568

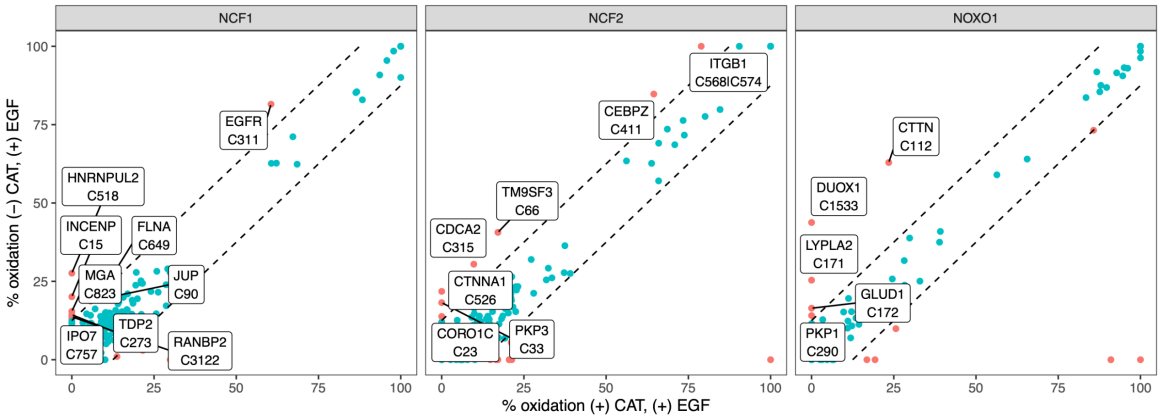


Figure 2.20: Percent oxidation of cysteines identified in the (-) catalase, (+) EGF sample vs. cysteines identified in the (+) catalase, (+) EGF sample. Peptides in blue have a percent oxidation which is across both cell treatments. Peptides in red are considered to be selectively oxidized in one or the other treatment.

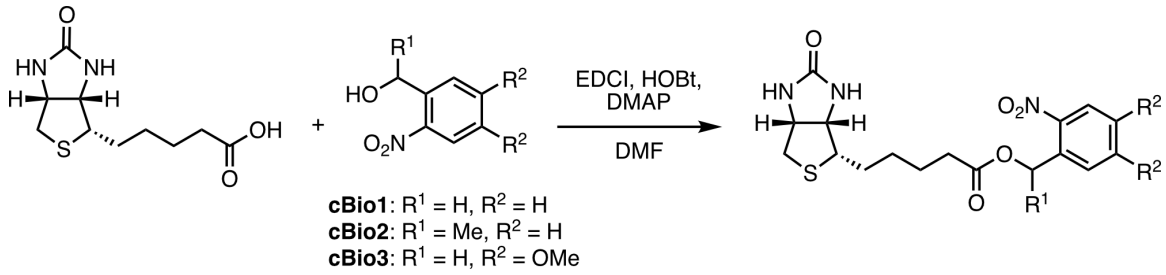
and C574 ITGB1 79 vs. 100 % oxidation in the NCF2 (+) vs. (-) PEG-catalase sample. CTTN is a substrate of Src kinase and a known oncogene which contributes to oncogenesis by inhibiting the ubiquitin of EGFR.[54] CTTN C112 showed 23 vs. 63 % oxidation in the NOXO1 (+) vs. (-) PEG-catalase sample.

2.3 Conclusion

In summary, we describe the development of a proteomics platform combining TurboID and OxICAT to identify localized sites of cysteine oxidation during EGFR activation. Proteins we identified sites on a number of proteins responsible for regulating cell growth and proliferation. Future experiments will include validating oxidation sites on EGFR, CTTN, and ITGB1 to determine how they alter activity.

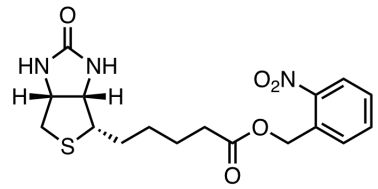
2.4 Materials and methods

2.4.1 Synthesis of cBio1-3



Scheme 2.1: Synthesis of **cBio1-3**.

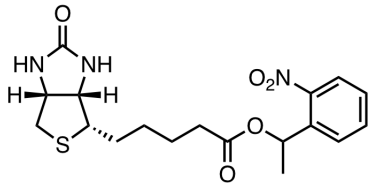
2.4.1.1 cBio1



D-Biotin (1 eq, 488.6 mg, 2 mmol) was dissolved in dry DMF (10 mL) by gentle heating. EDCI (1.1 eq, 341.6 mg), DMAP (1 eq, 244.3 mg) and HOBt (cat.) were added the the reaction was stirred for 30 minutes at room temperature. (2-nitrophenyl)methanol (1.1 eq, 336.9 mg) dissolved in dry DMF (3 mL) was added dropwise and the reaction was left overnight stirring at room temperature. The reaction was diluted with 100 mL ethyl acetate and washed with 1 M HCl (3×75 mL). The organic layer was then washed with saturated $NaHCO_3$ and dried with $MgSO_4$. The crude product was then purified by flash chromatography using a solvent system containing $2-4$ % MeOH in DCM to afford **cBio1** as a orange solid (327.5 mg, 31.3 % yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.09 (dd, J = 8.1, 1.3 Hz, 1H), 7.70 – 7.62 (m, 1H), 7.57 (d, J = 7.7 Hz, 1H), 7.51 (t, J = 7.7 Hz, 1H), 7.26 (s, 3H), 5.50 (s, 2H), 4.56 – 4.49 (m, 1H), 4.33 (dd, J = 7.8, 4.6 Hz, 1H), 3.17 (dd, J = 12.0, 7.3 Hz, 1H),

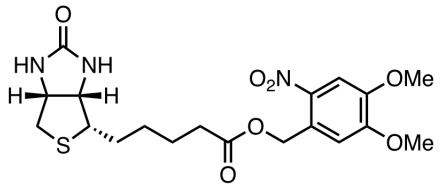
2.93 (dd, J = 12.8, 5.1 Hz, 1H), 2.73 (d, J = 12.8 Hz, 1H), 2.44 (t, J = 7.4 Hz, 2H), 1.70 (dt, J = 16.7, 8.3 Hz, 4H), 1.52 – 1.44 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 172.92, 163.33, 146.92, 133.70, 131.95, 129.38, 128.88, 125.04, 62.98, 61.87, 60.07, 55.32, 40.52, 33.74, 28.28, 28.21, 24.71. HRMS for $C_{17}H_{21}N_3O_5S + H^+$: *m/z* calcd. 380.1280; obsd. 380.1264.

2.4.1.2 cBio2



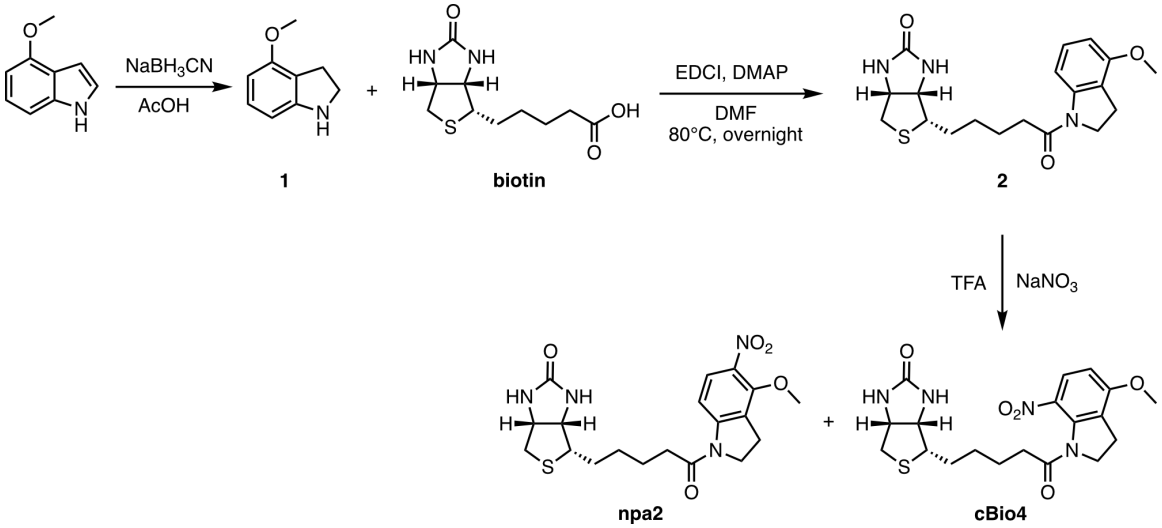
D-Biotin (1 eq, 488.6 mg, 2 mmol) was dissolved in dry DMF (10 mL) by gentle heating. EDCI (1.1 eq, 341.6 mg), DMAP (1 eq, 244.3 mg) and HOBt (cat.) were added the the reaction was stirred for 30 minutes at room temperature. 1-(2-nitrophenyl)ethan-1-ol (1.1 eq, 367.8 mg) dissolved in dry DMF (3 mL) was added dropwise and the reaction was left overnight stirring at room temperature. The reaction was diluted with 100 mL ethyl acetate and washed with 1 M HCl (3 × 75 mL). The organic layer was then washed with saturated $NaHCO_3$ and dried with $MgSO_4$. The crude product was then purified by flash chromatography using a solvent system containing 2 – 4 % MeOH in DCM to afford **cBio2** as a yellow solid (316.4 mg, 40.2 % yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ: (s, 1H), 6.98 (s, 1H), 6.15 (s, 1H), 5.46 (s, 2H), 4.51 – 4.45 (m, 1H), 4.29 (dd, J = 7.8, 4.6 Hz, 1H), 3.95 (d, J = 21.3 Hz, 6H), 3.13 (ddd, J = 8.5, 6.4, 4.5 Hz, 1H), 2.88 (dd, J = 12.8, 5.0 Hz, 1H), 2.70 (d, J = 12.8 Hz, 1H), 2.43 (td, J = 7.5, 1.1 Hz, 2H), 1.77 – 1.60 (m, 4H), 1.44 (qd, J = 8.6, 6.5 Hz, 2H). ^{13}C NMR (151 MHz, $CDCl_3$) δ: 175.17, 166.49, 150.44, 136.25, 131.03, 129.89, 129.87, 127.02, 70.64, 62.77, 58.12, 58.09, 43.17, 30.98, 30.90, 27.33, 27.31, 24.54. HRMS for $C_{18}H_{23}N_3O_5S + H^+$: *m/z* calcd. 394.1424; obsd. 394.1437.

2.4.1.3 cBio3



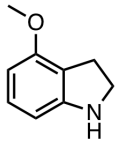
D-Biotin (1 eq, 488.6 mg, 2 mmol) was dissolved in dry DMF (10 mL) by gentle heating. EDCI (1.1 eq, 341.6 mg), DMAP (1 eq, 244.3 mg) and HOBt (cat.) were added the the reaction was stirred for 30 minutes at room temperature. (4,5-dimethoxy-2-nitrophenyl)methanol (1.1 eq, 470.4 mg) dissolved in dry DMF (3 mL) was added dropwise and the reaction was left overnight stirring at room temperature. The reaction was diluted with 100 mL ethyl acetate and washed with 1 M HCl (3×75 mL). The organic layer was then washed with saturated $NaHCO_3$ and dried with $MgSO_4$. The crude product was then purified by flash chromatography using a solvent system containing $2-4$ % MeOH in DCM to afford **cBio3** as a pale yellow solid (128.5 mg, 14.6 % yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ: 7.92 – 7.88 (m, 1H), 7.66 – 7.59 (m, 2H), 7.41 (ddt, J = 8.5, 6.0, 2.3 Hz, 1H), 6.29 (qd, J = 6.5, 3.5 Hz, 1H), 6.11 – 6.08 (m, 1H), 5.61 (s, 1H), 4.46 (dd, J = 7.9, 5.0 Hz, 1H), 4.25 (ddd, J = 7.8, 4.6, 3.0 Hz, 1H), 3.10 (dddd, J = 10.9, 8.5, 6.3, 4.6 Hz, 1H), 2.86 (ddd, J = 12.8, 5.1, 3.7 Hz, 1H), 2.68 (d, J = 12.8 Hz, 1H), 2.39 – 2.27 (m, 2H), 1.72 – 1.56 (m, 6H), 1.44 – 1.32 (m, 2H), 1.32 – 1.21 (m, 1H). ^{13}C NMR (151 MHz, $CDCl_3$) δ: 175.70, 166.48, 150.98, 142.76, 129.49, 113.54, 110.94, 64.59, 62.78, 59.16, 59.08, 58.17, 36.51, 31.01, 30.88, 27.48. HRMS for $C_{19}H_{25}N_3O_7S + H^+$: *m/z* calcd. 440.1491, obsd. 440.1491.

2.4.2 Synthesis of cBio4



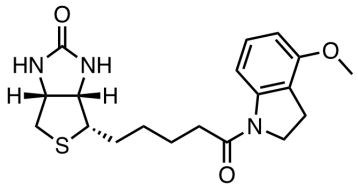
Scheme 2.2: Synthesis of **cBio4**.

2.4.2.1 4-methoxyindoline (1)



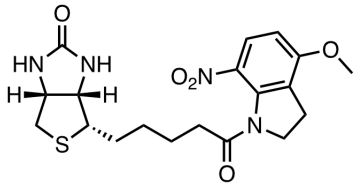
To a solution of 4-methoxy-1H-indole (1.0 eq, 1 g, 6.8 mmol) in acetic acid (10 mL, 0.68 M), sodium cyanoborohydride (1.1 eq, 472 mg, 7.5 mmol) was added slowly over the course of 10 minutes. The reaction was sired at room temperature for 3 hours, then adjusted to pH 10 with 0.5 N NaOH. The reaction was extracted with ethyl acetate (3×40 mL) and the organic layer washed with brine, then dried over Na_2SO_4. The crude product was purified by flash chromatography with $5 - 10$ % ethyl acetate in hexane to afford **1** as a brown solid (0.861 g, 84.9 % yield). ^{1}H NMR (600 MHz, Chloroform-d) δ: 7.00 (t, 1H), 6.33 (d, J = 7.8 Hz, 1H), 6.30 (d, J = 8.2 Hz, 1H), 3.82 (s, 3H), 3.75 (s, 1H), 3.57 (t, J = 8.5, 1.0 Hz, 2H), 3.00 (t, J = 8.5 Hz, 2H). HRMS for $C_9H_{11}NO + H^+$: m/z calcd. 150.0919; obsd. 150.0921.

2.4.2.2 Compound 2



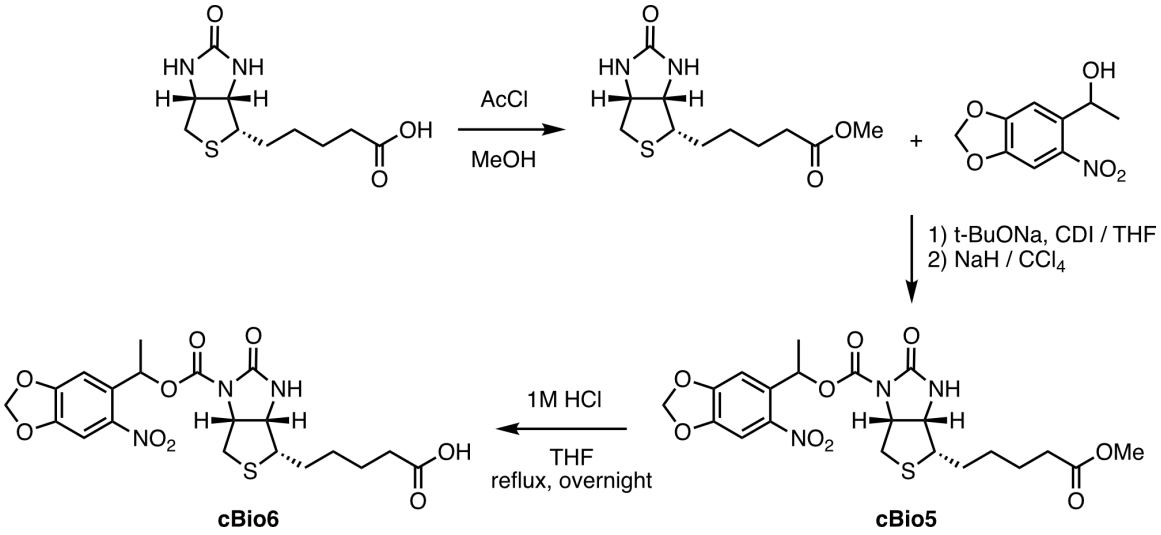
D-Biotin (2 eq, 16.52 mmol, 4.037 g) was dissolved in dry DMF by gentle heating (25 mL). Compound **2** (1 eq, 8.26 mmol, 1.23 g), EDCI (2 eq, 2.56 g) DMAP (2 eq, 2.02 g) were added and the reaction was heated to 80 °C overnight. The reaction mixture was diluted with 5 % MeOH in DCM (100 mL) and extracted with 1 M HCL (3×100 mL) and saturated $NaHCO_3$ (3×100 mL). The organic layer was dried with Na_2SO_4 and purified by flash chromatography using a gradient of $5 - 10$ % MeOH in DCM to to give **2** as a beige solid (2.47 g, 79.6 % yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ: 12.53 (s, 2H), 7.17 (t, J = 8.2 Hz, 1H), 6.70 (s, 1H), 4.60 (dd, J = 8.0, 4.7 Hz, 1H), 4.41 (dd, J = 8.0, 4.5 Hz, 1H), 4.07 (t, J = 8.4 Hz, 2H), 3.82 (s, 3H), 3.22 (td, J = 7.4, 4.4 Hz, 1H), 3.10 (t, J = 8.4 Hz, 2H), 2.94 (dd, J = 13.1, 4.9 Hz, 1H), 2.77 (d, J = 13.1 Hz, 1H), 2.48 (t, J = 7.4 Hz, 2H), 1.83 – 1.61 (m, 3H), 1.58 – 1.45 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 172.38, 164.94, 155.74, 143.45, 129.08, 118.90, 110.45, 106.72, 62.51, 60.82, 55.33, 55.01, 48.73, 40.28, 35.31, 28.15, 27.90, 24.92, 24.33. HRMS for $C_{19}H_{25}N_3O_3S + H^+$: *m/z* calcd. 376.16166; obsd. 376.1705.

2.4.2.3 cBio4



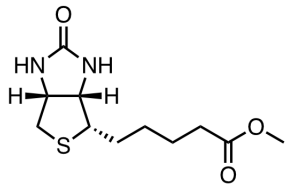
Compound **2** (1 eq, 3.09 mmol) was dissolved in TFA (10 mL) and $NaNO_3$ (0.9 eq, 289.2 mg) was added. The reaction was stirred at room temperature for 1 hour then the reaction mixture was added to saturated $NaHCO_3$ and extracted with 5 % MeOH in DCM. The crude product was purified by flash chromatography give a mixture of the 7 and 5 nitro regioisomers of **cBio4**. The isomers were separated by HPLC with a gradient of increasing ACN 0.1 % TFA (solvent) in water- 0.1 % TFA (solvent A) to give cBio4 as a bright yellow sold (217.5 mg, 16.7 % yield). ^{1}H NMR (400 MHz, Acetone-D_6) δ: 7.72 (d, J = 9.0 Hz, 1H), 6.87 (d, J = 9.0 Hz, 1H), 4.60 – 4.53 (m, 1H), 4.44 – 4.39 (m, 1H), 4.35 (t, J = 8.2 Hz, 2H), 3.96 (s, 3H), 3.83 (s, 0H), 3.32 – 3.24 (m, 1H), 3.11 (t, J = 8.0 Hz, 2H), 2.97 (dd, J = 12.6, 5.0 Hz, 1H), 2.74 (d, J = 12.6 Hz, 1H), 2.55 (t, J = 7.4 Hz, 2H), 1.80 (ddd, J = 12.6, 9.5, 6.5 Hz, 1H), 1.68 (ddt, J = 25.6, 14.3, 7.7 Hz, 3H), 1.54 (p, J = 6.8 Hz, 2H). HRMS for $C_{19}H_{24}N_4O_5S$ + H^+: *m/z* calcd. 421.1467, obsd. 421.1545.

2.4.3 Synthesis of cBio5-6



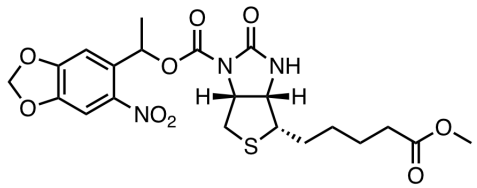
Scheme 2.3: Synthesis of **cBio5-6**.

2.4.3.1 Biotin methyl ester (biotin-OMe)[28,29]



Acetyl chloride (1.6 eq, 13.3 mmol, 946.5 mL) was added dropwise to ice cold MeOH and stirred for 5 minutes to generate a methanolic solution of hydrogen chloride. The acetyl chloride / MeOH solution was added to a suspension of D-biotin (1 eq, 8.19 mmol, 2 g) in MeOH (10 mL) and stirred at room temperature for 3 hours. The reaction was then concentrated *in vacuo* and partitioned between 5 % MeOH / DCM and saturated $NaHCO_3$. The aqueous layer was extracted again with 5 % MeOH / DCM and the combined organic layers were dried with Na_2SO_4 and concentrated *in vacuo* to afford biotin-OMe as a white solid (1.9 g, 90 % yield). ^{1}H NMR (600 MHz, Chloroform-d) δ: 5.86 (s, 1H), 5.47 (s, 1H), 4.56 – 4.47 (m, 1H), 4.31 (ddd, J = 7.8, 4.6, 1.5 Hz, 1H), 3.66 (s, 3H), 3.15 (ddd, J = 8.4, 6.4, 4.6 Hz, 1H), 2.91 (dd, J = 12.8, 5.0 Hz, 1H), 2.74 (d, J = 12.8 Hz, 1H), 2.33 (t, J = 7.5 Hz, 2H), 1.76 – 1.62 (m, 4H), 1.45 (dtd, J = 14.0, 6.8, 3.4 Hz, 2H). HRMS for $C_{11}H_{18}N_2O_3S + H^+$ ester: *m/z* calcd. 258.3360; obsd. 258.3312.

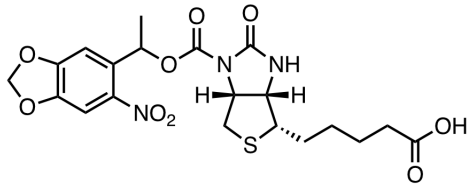
2.4.3.2 cBio5[28,29]



To a solution of methylnitropiperonyl alcohol (1 eq, 2.37 mmol, 500 mg) in dry THF, 1M sodium *t*-butoxide (236 mL) was added and stirred for 10 minutes at room tem-

perature. CDI (1.5 eq, 575.9 mg) was added and the reaction was stirred overnight at room temperature under N_2. The solvent was removed under reduced pressure and the residue was suspended in carbon tetrachloride then filtered through Celite. The filtrate was concentrated *in vacuo* and resuspended in chloroform **Biotin-OMe** (0.6 eq, 367 mg) and sodium hydride (63 mg, 60 % dispersion in mineral oil) were added and the mixture was refulxed under N_2 for 3 days. The reaction was quenched with 3 drops of acetic acid. The reaction mixture was concentrated *in vacuo* and purified by flash chromatography (3 % acetone, 3 % MeOH in DCM) to afford **cBio5** as a pale yellow solid (422 mg, 65.9 % yield). ^{1}H NMR (400 MHz, Acetone-D_6) δ: 7.53 (d, J = 2.8 Hz, 1H), 7.48 (s, 1H), 6.85 (d, J = 5.9 Hz, 1H), 6.42 (p, J = 6.2 Hz, 1H), 6.24 (p, J = 1.0 Hz, 2H), 4.95 (ddd, J = 7.7, 4.9, 1.5 Hz, 1H), 4.88 (ddd, J = 7.8, 4.8, 1.5 Hz, 1H), 4.39 – 4.31 (m, 1H), 3.61 (d, J = 0.8 Hz, 3H), 3.33 (ddt, J = 9.1, 6.6, 3.4 Hz, 1H), 3.05 (ddd, J = 10.2, 7.1, 3.2 Hz, 2H), 2.32 (t, J = 7.4 Hz, 3H), 1.89 – 1.75 (m, 1H), 1.70 – 1.57 (m, 7H), 1.55 – 1.39 (m, 2H). ^{13}C NMR (101 MHz, Acetone-D_6) δ: 174.13, 153.80, 148.54, 136.04, 107.60, 104.65, 71.03, 63.59, 58.70, 56.23, 56.21, 39.21, 39.09, 34.19, 25.69, 22.74, 22.69. HRMS for $C_{21}H_{25}N_3O_9S$, $[2M + Na^+]$: m/z calcd. 1013.2522, obsd. 1013.2521.

2.4.3.3 cBio6[28,29]



cBio5 (0.667 mmol, 329.9 mg) was dissolved in a 1:1 solution of 1 N HCl in THF and heated to 70 °C under refulx overnight. The reaction was extracted with 1 N HCl, 5 % MeOH in DCM, and dried over $NaSO_4$. The crude product was purified by HPLC with a gradient of increasing ACN 0.1 % TFA (solvent) in water- 0.1 %

TFA (solvent A) to give cBio6 as a yellow sold (171.5 mg, 52 % yield). ^{1}H NMR (600 MHz, acetone-D_6) δ: 7.53 (d, J = 4.2 Hz, 1H), 7.48 (d, J = 1.8 Hz, 1H), 6.46 – 6.39 (m, 1H), 6.27 – 6.23 (m, 2H), 4.96 (dd, J = 7.5, 5.5 Hz, 1H), 4.40 – 4.32 (m, 1H), 3.34 (ddd, J = 8.4, 6.1, 4.2 Hz, 1H), 3.13 – 3.00 (m, 2H), 2.31 (t, J = 7.4 Hz, 2H), 2.10 – 2.05 (m, 2H), 1.83 (ddt, J = 12.8, 9.5, 6.4 Hz, 1H), 1.69 – 1.60 (m, 6H), 1.50 (dqd, J = 11.8, 6.4, 3.0 Hz, 2H). HRMS for $C_{20}H_{23}N_3O_9S$, $[2M + Na^+]$: m/z calcd. 985.22077, obsd. 985.22923.

2.4.4 Molecular biology

2.4.4.1 Constructs

The biotin ligase, TurboID was subcloned from V5-TurboID-NES in pcDNA3.1 (Addgene plasmid 107169)[30] into a pDonor221 Gateway destination vector via a BP recombination. Three different TurboID constructs were generated, with additional elements added by primer overhang: (1) TurboID with a 5′ multiple cloning site (MCS) and a 3′ HA tag and stop codon (Figure 2A.1 2A.2), (2) TurboID with a 5′ start codon and V5 tag and a 3′ MCS (Figure 2A.3, 2A.4), and (3) TurboID with a 5′ start codon and 3′ MCS (Figure 2A.5). Human NCF1 (NP_000256.4), were cloned into V5-TurboID-MCS pDonor221 (Figure 2A.6) and human NCF2 (NP_000424.2), and NOXO1 (NP_001254650.1) were cloned into MCS-TurboID-HA pDonor221 (Figure 2A.7, 2A.8). A LR recombination was performed to move the payload of the pDonor vector into the Gateway destination vector, pLenti CMV Puro DEST (Addgene plasmid 17452).[55]

2.4.4.2 Cloning primers

attB1_V5_for (53-mer):

5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGCCACCATGGGCAAG
CCCATC-3′

TurboID_MCS_attB2_rev (81-mer):

5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCGGATCCGAATTCCACTG
TCTCGAGCGGCCGCCTTTTCGGCAGACCGCAGAC-3′

attB1-MCS-TurboID_for (90-mer):

5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCTCGAGCGGCCGCTCC
GGATTCGAATTCGGATCCAAAGACAATACTGTGCCTCTGAAGC-3′

attB2_HA_rev (83-mer):

5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATGCGTAATCCGGTA
CATCGTAAGGGTAAGCCTTTTCGGCAGACCGCAGAC-3′

attB1-start-TurboID_for (65-mer):

5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTGCCACCATGGGCAAAGAC
AATACTGTGCCTCTGAAG-3′

NCF1_for_2 (37-mer):

5′-TTCAGTGCGGCCGCTGGCATGGGGGACACCTTCATCC-3′

NCF1_rev_6 (34-mer):

5′-TTCAGTGAATTCTCAGACGGCAGACGCCAGCTTC-3′

NCF2_rev_1 (34-mer):

5′-TTCAGTGGATCCGACTTCTCTCCGAGTGCTTTCC-3′

NCF2_for_1 (32-mer):

5′-TTCAGTGCGGCCGCGCCACCATGTCCCTGGTG-3′

NOXO1_rev (31-mer):

5′-TTCAGTGGATCCCTGCTCCGTCGTGGGGTGC-3′

NOXO1_for (30-mer):

5′-TTCAGTGCGGCCGCGCCACCATGGCAGGCC-3′

2.4.4.3 Lentivirus concentration and spinoculation

For each TurboID construct, HEK293T cells were split into 4×10 cm plates and grown to ∼70 % confluency in DMEM supplemented with 10 % heat inactivated serum (HIS). Cells were co-transfected with 2 µg each of VSV-G (Addgene plasmid 8454), psPAX (Addgene plasmid 12260), and the TurboID construct in a pLenti plasmid using Lipofectamine 2000 (Fisher Scientific). After 24 hours cells were re-fed with fresh DMEM containing HIS. 24 hours later, the supernatant was harvested and filtered through a 0.45 µm PES syringe filter. For every 3 volumes of viral supernatant, 1 volume of virus concentrator solution (1.2 M NaCl, 40 % (w/v) PEG-8000, 100 mM phosphate, pH 7.2) was added and the solution was incubated overnight at 4 °C with agitation. The viral pellet was collected by centrifugation ($1,600 \times g$, 60 min, 4 °C), resuspend in serum free DMEM, and stored at −80 °C until use.

7.5×10^3 A431 cells per well were seeded into a 6 well plate in DMEM supplemented with HIS and incubated at 37 °C for 24 hours. Concentrated virus was thawed at 37 °C and polybrene (3 µL of 10 mg mL^{-1} stock) was added to the virus solution. The virus solution was added to the appropriate target cell well and the 6 well plate was centrifuged at $800 \times g$ for 30 min at 25 °C. Cells were incubated with virus for a further 24 hours at 37 °C. After 24 hours, the media was replaced and cells were selected with 0.5 µg mL^{-1} puromycin. The selection was considered complete when the (-) virus control cells were completely dead.

2.4.4.4 Cell culture and biotin labeling

A431 TurboID cell lines were grown in DMEM supplemented with 10 % dialyzed fetal bovine serum (FBS) (Gibco) and 1 % penicillin-streptomycin-amphotericin B (PSA) (Gibco) at 37 °C and 5 % CO_2. In the final passage prior to biotin labeling, cells were grown in biotin depleted media. To prepare biotin depleted media, 1 mL of a streptavidin agarose slurry (Thermo Fisher) was washed with PBS (3×5

mL) and added to 500 mL complete DMEM. The media was rotated overnight at 4 °C, then sterile filtered through a 0.2 µm cellulose membrane. Cells were grown to ∼70 % confluency and stimulated with 50 µM biotin (from 100 mM stock in DMSO), 100 ng mL^{-1} EGF (Sigma), and (+/−) 1000 U mL^{-1} PEG-catalase (Sigma) for 15 minutes at 37 °C. Cells were harvested by scraping and stored at −80 °C until use.

2.4.4.5 Antibodies

Streptavidin-HRP Conjugate (EMD Millipore); EGFR (D38B1), Phospho-EGFR (Tyr1068) (D7A5), HA-Tag (C29F4), GAPDH, α-rabbit-HRP, α-mouse-HRP, α-mouse IgG Alexa Fluor 488, α-rabbit IgG Alexa Flour 594, (Cell Signaling); V5-Tag (Invitrogen) antibodies were used for western blotting and/or immunofluorescence assays.

2.4.4.6 Immunofluorescence

A431 cells expressing a TurboID construct were trypsinized and plated onto 35 mm glass bottom dishes (MatTek Corporation, P35G-1.5-10-C) and after 48 hours were treated with or without 100 ng mL^{-1} EGF as above. Cells were then washed with 5 mL PBS and fixed with 4 % paraformaldehyde in PBS for 15 minutes at room temperature. Plates were washed (3 × 5 min) with PBS and blocked with 5 % goat serum, 0.3 % Tween-20 in PBS. The cells were incubated with a rabbit α-Phospho-EGFR (Tyr1068) at a 1:800 dilution and a mouse α-HA-tag or α-V5-tag antibody at a 1:400 dilution in 1 % BSA, 0.3 % Tween-20 in PBS overnight at 4 °C. Then, cells were washed (3 × 5 min) with PBS, then incubated with an α-rabbit-A549 and α-mouse-488 antibody diluted 1:500 in 1 % BSA, 0.3 % Tween-20 in PBS for 1 hour at room temperature. Cells were washed (3 × 5 min) with PBS and mounted with Prolong Gold Antifade reagent with DAPI (Cell Signaling Technologies). Images were acquired on a Leica (Wetzlar, Germany) TCS SP5 scanning confocal microscope using

a Plan-Apochromat 63 × /1.40 numerical aperture lens.

2.4.4.7 Western blotting

Samples were separated on a 10 % SDS-PAGE gel then transferred to nitrocellulose membranes at 15 V for 18 hours (EGFR and pEGFR) or 75 V for 100 minutes (all others). The membranes were blocked with 5 % non-fat powdered milk in TBS-T for 1 hour at room temperature. The membrane was probed with primary antibody (1:1000) in 5 % powdered milk / TBS-T overnight at 4 °C. The blot was washed (3 × 5 min) with TBS-T then incubated with a α-rabbit-HRP conjugated secondary antibody (1:2000) for 2 hours at room temperature. For biotin-blots, after electrotransfer, membranes were blocked overnight with 5 % non-fat dry milk in TBS-T at room temperature. The membrane was then washed (3 × 5 min) and probed with streptavidin-HRP (1:1000) in TBS-T for 1 hour at room temperature. After probing with secondary antibody or streptavidin-HRP, membranes were washed 3×5 min with TBS-T and treated with chemiluminescence reagents (Thermo Fisher) and developed on a Bio-Rad ChemDoc MP imaging system.

2.4.5 Proteomics experiments

2.4.5.1 Identification of protein targets of Nox2 ROS and ReDiMe labeling

Sample preparation

Cells were lysed by sonication and lysates were diluted to 2 mg mL^{-1} in PBS. A total of 2 mg of protein was used for each of the heavy and light samples for a total of 4 mg protein per prep. Proteins were precipitated by the addition of 10 % TCA (100 µL of 100 % (w/v) TCA in PBS) and frozen at −80 °C overnight. After thawing, proteins were pelleted by centrifugation (15,000 × g, 15 min) and the supernatant discarded. Pellets were resuspended in 500 µL ice cold acetone by sonication and

pelleted by centrifugation ($15,000 \times g$, 15 min, 4 °C). After removing the supernatant and allowing the pellet to air dry, pellets were resuspended in 1.2 % SDS in PBS and solubilized by heating at 65 °C and sonication. Once dissolved, samples were diluted to a final concentration of 0.2 % SDS by the addition of 5 mL PBS and incubated with a streptavidin agarose slurry (170 µL) overnight at 4 °C, then 3 hours at room temperature. The beads were washed with PBS (3×5 mL), and water (3×5 mL). The beads were pelleted by centrifugation ($1400 \times g$, 3 min) between washes. The washed beads were resuspended in 500 µL 6 M urea / TEAB and 10 mM DTT and incubated at 65 °C for 15 min. 20 mM iodoacetamide was then added and the samples were incubated at 37 °C for 30 min. After reduction and alkylation the beads were pelleted by centrifugation ($1400 \times g$, 3 min), and resuspended in 200 µL 2 M urea / TEAB 1 mM $CaCl_2$, and 2 µg trypsin (Promega) and incubated overnight at 37 °C. After tryptic digest, reductive dimethylation[34,56] was performed by the addition of 8 µL of 20 % light ($^{12}COH_2$) (+ EGF) or heavy ($^{13}COD_2$) (− EGF) formaldehyde and 40 µL of 0.6 M sodium cyanoborohydride at room temperature for 2 hours. The reaction was quenched by the addition of 8 µL ammonium hydroxide. The light (+ EGF) and heavy (− EGF) tryptic peptide samples were then combined, desalted on a Sep-Pak, and dried by speed-vac. Peptide samples were stored at −20 °C until LC-MS/MS analysis.

Liquid chromatography mass spectrometry

Mass spectrometry was performed on an LTQ Orbitrap Discovery (Thermo Fisher) coupled to an Agilent 1200 series HPLC. Peptide pellets were resuspended in 200 µL 1 % formic acid and pressure loaded onto a 250 µm fused silica desalting column, packed with 4 cm Aqua C18 reverse phase resin (Phenomenex). Peptides were eluted onto a 100 µm fused silica biphasic column, packed with 4 cm strong cation exchange resin (Watman) and 10 cm C18 resin using a using a five-step multidimensional LC/LC-MS/MS protocol (MudPIT).[49] Each of the five runs begins with a salt

push (0 %, 25 %, 50 %, 80 %, and 100 % buffer C), followed by a gradient 0–100 % buffer B in buffer A (buffer A: 95 % water, 5 % ACN, 0.1 % formic acid; buffer B: 20 % water, 80 % ACN, 0.1 % formic acid; buffer C 95 % water, 5 % ACN, 500 mM ammonium acetate). The flowrate was $\sim$0.25 µL min^{-1} and the spray voltage was set to 2.75 kV. One survey MS1 scan (400-1800 *m/z*) was followed by 8 data dependent scans of the n$^{\text{th}}$ most intense ion.

Data analysis

Peptides were assigned to MS2 spectra using the SEQUEST[48] algorithm, searching against a human reverse-concatenated non-redundant protein database. Data sets were searched independently for peptides containing isotopically light and heavy dimethyl modifications. Static modifications for specified for cysteine (+57.0215Da, iodoacetamide alkylation), and N-terminal and lysine dimethylation (+28.031 30 Da, light), (+34.063 12 Da, heavy). MS2 matches were assembled by protein and filtered to achieve a false discovery rate of 5 % using DTASelect 2.0.[57] Light/heavy intensity ratios ($R_{L/H}$) were calculated using the CIMAGE quantification package as described previously.[49,58] CIMAGE reports peptide $R_{L/H}$ for co-eluting light and heavy peptides. The peptides are grouped by protein and a representative $R_{L/H}$ for each protein is calculated from the median $R_{L/H}$ for each peptide identified for that protein.

2.4.5.2 TurboID-OxICAT

Sample preparation

Methods for OxICAT sample preparation were adapted from Topf *et al.*[42] and Bechtel *et al.*.[59] A431 cell pellets were resuspended in 500 µL denaturing alkylation buffer (DAB) (6 M urea, 200 mM Tris-HCl, 10 mM EDTA, 0.5 % SDS, pH 8.5) and 10 mM NEM-H_5 and lysed using an ultrasonic tip sonicator. After lysis, cells were labeled for 2 hours at 37 °C. Lysates were diluted 3-fold with dH_2O and precipitated with the addition of 5 sample volumes of ice cold acetone and incubated at −20 °C for

2 hours. Precipitated proteins were pelleted by centrifugation ($4,500 \times g$, 4 °C), pellets were washed with ice cold acetone and centrifuged for a further 10 min ($4,500 \times g$, 4 °C). Pellets were resolubilized in DAB and protein concentrations were determined using the DC Assay Kit. 4 mg protein was brought to a final volume of 400 µL in DAB. Samples were treated with 2.5 mM TCEP and incubated at 37 °C for 5 min. Following reduction, samples were diluted with 600 µL DAB. 10 mM NEM-D_5 was added and samples were incubated at 37 °C for 2 hours. Samples were diluted 3-fold with dH_2O and proteins were precipitated with 5 sample volumes of ice-cold acetone and incubated overnight at −20 °C. Pellets were washed with ice-cold acetone, centrifuged for 10 min ($4,500 \times g$, 4 °C), and allowed to air dry. Pellets were resuspended in 1.2 % SDS in PBS and solubilized by heating at 65 °C and sonication. Once dissolved, samples were diluted to a final concentration of 0.2 % SDS by the addition of 5 mL PBS and incubated with a streptavidin agarose slurry (170 µL) overnight at 4 °C, then 3 hours at room temperature. The beads were washed with PBS (3×5 mL), and water (3×5 mL). The beads were pelleted by centrifugation ($1400 \times g$, 3 min) between washes. After the final wash, beads were resuspended in 200 µL 2 M urea / PBS 1 mM $CaCl_2$, and 2 µg trypsin (Promega) and incubated overnight at 37 °C. The tryptic digest was collected, and the beads were washed 2×75 µL with water and the washes combined with the tryptic digest for a final sample volume of ∼350 µL. Samples were then desalted on a Sep-Pak, and dried by speed-vac. Peptide samples were stored at −20 °C until LC-MS/MS analysis.

Mass spectroscopy

Mass spectrometry was performed on an Orbitrap Exploris 240 (Thermo Fisher) coupled to UltiMate 3000 RSLCnano LC system. Samples were loaded onto a Pepmap 100 C18 trap (Thermo Fisher) and eluted onto a Pepmap 100 C18 analytical column with a gradient 0–30 % buffer B in buffer A (buffer A: 0.1 % formic acid in H_2O, buffer B: 0.1 % formic acid in ACN). Under positive ionization mode, one survey MS1

scan (350-1800 m/z) was followed by 20 data dependent scans of the n^{th} most intense ion. The orbitrap mass resolution was set to $60,000$; isolation window 2.0 m/z units; default charge 2^+; maximum ion time, 50 ms; dynamic exclusion, 20.0 s.

Data analysis

The LC-MS/MS raw data were processed using Thermo Proteome Discoverer (PD) 2.4 (Thermo Fisher). The data was searched searched against the Human SwissProt database with the addition of the TurboID biotin ligase sequence. The PSM search was performed using Sequest HT. The search parameters included searching specific protease cleavage sites with 2 maximum missed cleavages. Protein N-terminal acetylation, methionine oxidation, and cysteine modification with NEM-H_5 and NEM-D_5 were set as dynamic modifications. NEM-H_5 and NEM-D_5 were used as modifications for the the light and heavy quantification channels respectively. The normalized peptide intensity reported by PD for the heavy and light quantification channels was used to calculate the percent oxidation by the formula:

$$\%oxidation = \frac{heavy_{int}}{heavy_{int} + light_{int}} \tag{2.1}$$

References

[1] Stone, J. R.; Yang, S. Hydrogen peroxide: a signaling messenger. *Antioxid Redox Signal* **2006**, *8*, 243–70.

[2] Paulsen, C. E.; Truong, T. H.; Garcia, F. J.; Homann, A.; Gupta, V.; Leonard, S. E.; Carroll, K. S. Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nat Chem Biol* **2011**, *8*, 57–64.

[3] Wee, P.; Wang, Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)* **2017**, *9*, DOI: 10.3390/cancers9050052.

[4] Woo, H. A.; Yim, S. H.; Shin, D. H.; Kang, D.; Yu, D. Y.; Rhee, S. G. Inactivation of peroxiredoxin I by phosphorylation allows localized H_2O_2 accumulation for cell signaling. *Cell* **2010**, *140*, 517–28.

[5] Van Montfort, R. L.; Congreve, M.; Tisi, D.; Carr, R.; Jhoti, H. Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature* **2003**, *423*, 773–7.

[6] Salmeen, A.; Andersen, J. N.; Myers, M. P.; Meng, T. C.; Hinks, J. A.; Tonks, N. K.; Barford, D. Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature* **2003**, *423*, 769–73.

[7] Chiarugi, P. Src redox regulation: there is more than meets the eye. *Mol Cells* **2008**, *26*, 329–37.

[8] Rhee, S. G.; Kang, S. W.; Jeong, W.; Chang, T. S.; Yang, K. S.; Woo, H. A. Intracellular messenger function of hydrogen peroxide and its regulation by peroxiredoxins. *Curr Opin Cell Biol* **2005**, *17*, 183–9.

[9] Rhee, S. G.; Woo, H. A. Multiple functions of 2-Cys peroxiredoxins, I and II, and their regulations via post-translational modifications. *Free Radic Biol Med* **2020**, *152*, 107–115.

[10] Qu, D. et al. Role of Cdk5-mediated phosphorylation of Prx2 in MPTP toxicity and Parkinson's disease. *Neuron* **2007**, *55*, 37–52.

[11] Rashidian, J.; Rousseaux, M. W.; Venderova, K.; Qu, D.; Callaghan, S. M.; Phillips, M.; Bland, R. J.; During, M. J.; Mao, Z.; Slack, R. S.; Park, D. S. Essential role of cytoplasmic cdk5 and Prx2 in multiple ischemic injury models, in vivo. *J Neurosci* **2009**, *29*, 12497–505.

[12] Bedard, K.; Krause, K. H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev* **2007**, *87*, 245–313.

[13] Thiagarajah, J. R.; Chang, J.; Goettel, J. A.; Verkman, A. S.; Lencer, W. I. Aquaporin-3 mediates hydrogen peroxide-dependent responses to environmental stress in colonic epithelia. *Proc Natl Acad Sci U S A* **2017**, *114*, 568–573.

[14] Oakley, F. D.; Abbott, D.; Li, Q.; Engelhardt, J. F. Signaling components of redox active endosomes: the redoxosomes. *Antioxid Redox Signal* **2009**, *11*, 1313–33.

[15] Abo, M.; Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J Am Chem Soc* **2015**, *137*, 7087–90.

[16] Murakoshi, H.; Iino, R.; Kobayashi, T.; Fujiwara, T.; Ohshima, C.; Yoshimura, A.; Kusumi, A. Single-molecule imaging analysis of Ras activation in living cells. *Proc Natl Acad Sci U S A* **2004**, *101*, 7317–22.

[17] Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196*, 801–10.

[18] Kim, D. I.; Jensen, S. C.; Noble, K. A.; Kc, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **2016**, *27*, 1188–96.

[19] Kim, D. I.; Roux, K. J. Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends Cell Biol* **2016**, *26*, 804–817.

[20] Lambert, J. P.; Tucholska, M.; Go, C.; Knight, J. D.; Gingras, A. C. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics* **2015**, *118*, 81–94.

[21] Finch, E. A.; Augustine, G. J. Local calcium signalling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites. *Nature* **1998**, *396*, 753–6.

[22] Lima, S. Q.; Miesenböck, G. Remote Control of Behavior through Genetically Targeted Photostimulation of Neurons. *Cell* **2005**, *121*, 141–152.

[23] Peeler, J. C.; Falco, J. A.; Kelemen, R. E.; Abo, M.; Chartier, B. V.; Edinger, L. C.; Chen, J.; Chatterjee, A.; Weerapana, E. Generation of Recombinant Mammalian Selenoproteins through Genetic Code Expansion with Photocaged Selenocysteine. *ACS Chem Biol* **2020**, *15*, 1535–1540.

[24] Papageorgiou, G.; Corrie, J. E. Effects of Aromatic Substituents on the Photocleavage of 1-Acyl-7-nitroindolines. *Tetrahedron* **2000**, *56*, 8197–8205.

[25] Wang, D.; Yu, Z.; Yan, J.; Xue, F.; Ren, G.; Jiang, C.; Wang, W.; Piao, Y.; Yang, X. Photolysis of Caged-GABA Rapidly Terminates Seizures In Vivo: Concentration and Light Intensity Dependence. *Front Neurol* **2017**, *8*, 215.

[26] Grushevskyi, E. O.; Kukaj, T.; Schmauder, R.; Bock, A.; Zabel, U.; Schwabe, T.; Benndorf, K.; Lohse, M. J. Stepwise activation of a class C GPCR begins with millisecond dimer rearrangement. *Proc Natl Acad Sci U S A* **2019**, *116*, 10150–10155.

[27] Kulkarni, V. V.; Anand, A.; Herr, J. B.; Miranda, C.; Vogel, M. C.; Maday, S. Synaptic activity controls autophagic vacuole motility and function in dendrites. *J Cell Biol* **2021**, *220*, DOI: 10.1083/jcb.202002084.

[28] Sundberg, S. A.; Barrett, R. W.; Pirrung, M.; Lu, A. L.; Kiangsoontra, B.; Holmes, C. P. Spatially-Addressable Immobilization of Macromolecules on Solid Supports. *J Am Chem Soc* **1995**, *117*, 12050–12057.

[29] Terai, T.; Maki, E.; Sugiyama, S.; Takahashi, Y.; Matsumura, H.; Mori, Y.; Nagano, T. Rational development of caged-biotin protein-labeling agents and some applications in live cells. *Chem Biol* **2011**, *18*, 1261–72.

[30] Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **2018**, *36*, 880–887.

[31] Roux, K. J.; Kim, D. I.; Burke, B.; May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* **2018**, *91*, 19.23.1–19.23.15.

[32] Engelman, J. A.; Janne, P. A.; Mermel, C.; Pearlberg, J.; Mukohara, T.; Fleet, C.; Cichowski, K.; Johnson, B. E.; Cantley, L. C. ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. *Proc Natl Acad Sci U S A* **2005**, *102*, 3788–93.

[33] ODoherty, U.; Swiggard, W. J.; Malim, M. H. Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J Virol* **2000**, *74*, 10074–80.

[34] Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **2009**, *4*, 484–94.

[35] Dong, Y.; Hakimi, M. A.; Chen, X.; Kumaraswamy, E.; Cooch, N. S.; Godwin, A. K.; Shiekhattar, R. Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in DNA repair. *Mol Cell* **2003**, *12*, 1087–99.

[36] Boudreau, H. E.; Broustas, C. G.; Gokhale, P. C.; Kumar, D.; Mewani, R. R.; Rone, J. D.; Haddad, B. R.; Kasid, U. Expression of BRCC3, a novel cell cycle regulated molecule, is associated with increased phospho-ERK and cell proliferation. *Int J Mol Med* **2007**, *19*, 29–39.

[37] Yaffe, M. B.; Rittinger, K.; Volinia, S.; Caron, P. R.; Aitken, A.; Leffers, H.; Gamblin, S. J.; Smerdon, S. J.; Cantley, L. C. The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* **1997**, *91*, 961–71.

[38] Kim, Y.; Shiba-Ishii, A.; Nakagawa, T.; Iemura, S. I.; Natsume, T.; Nakano, N.; Matsuoka, R.; Sakashita, S.; Lee, S.; Kawaguchi, A.; Sato, Y.; Noguchi, M. Stratifin regulates stabilization of receptor tyrosine kinases via interaction with ubiquitin-specific protease 8 in lung adenocarcinoma. *Oncogene* **2018**, *37*, 5387–5402.

[39] Qin, X.; Qiu, F.; Zou, Z. TRIM25 is associated with cisplatin resistance in non-small-cell lung carcinoma A549 cell line via downregulation of 14-3-3sigma. *Biochem Biophys Res Commun* **2017**, *493*, 568–572.

[40] Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuillemot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* **2014**, *20*, 1983–92.

[41] Leichert, L. I.; Gehrke, F.; Gudiseva, H. V.; Blackwell, T.; Ilbert, M.; Walker, A. K.; Strahler, J. R.; Andrews, P. C.; Jakob, U. Quantifying changes in the thiol redox proteome upon oxidative stress in vivo. *Proc Natl Acad Sci U S A* **2008**, *105*, 8197–202.

[42] Topf, U.; Suppanz, I.; Samluk, L.; Wrobel, L.; Boser, A.; Sakowska, P.; Knapp, B.; Pietrzyk, M. K.; Chacinska, A.; Warscheid, B. Quantitative proteomics identifies redox switches for global translation modulation by mitochondrially produced reactive oxygen species. *Nat Commun* **2018**, *9*, 324.

[43] Abo, M.; Li, C.; Weerapana, E. Isotopically-Labeled Iodoacetamide-Alkyne Probes for Quantitative Cysteine-Reactivity Profiling. *Mol Pharm* **2018**, *15*, 743–749.

[44] Weisser, H.; Nahnsen, S.; Grossmann, J.; Nilse, L.; Quandt, A.; Brauer, H.; Sturm, M.; Kenar, E.; Kohlbacher, O.; Aebersold, R.; Malmstrom, L. An automated pipeline for high-throughput label-free quantitative proteomics. *J Proteome Res* **2013**, *12*, 1628–44.

[45] Keller, A.; Eng, J.; Zhang, N.; Li, X. J.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **2005**, *1*, 2005 0017.

[46] Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26*, 1367–72.

[47] Rost, H. L. et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **2016**, *13*, 741–8.

[48] Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994**, *5*, 976–89.

[49] Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790–5.

[50] Cheng, J. C.; Klausen, C.; Leung, P. C. Hydrogen peroxide mediates EGF-induced down-regulation of E-cadherin expression via p38 MAPK and snail in human ovarian cancer cells. *Mol Endocrinol* **2010**, *24*, 1569–80.

[51] Rafikova, O.; Rafikov, R.; Kangath, A.; Qu, N.; Aggarwal, S.; Sharma, S.; Desai, J.; Fields, T.; Ludewig, B.; Yuan, J. X.; Jonigk, D.; Black, S. M. Redox regulation of epidermal growth factor receptor signaling during the development of pulmonary hypertension. *Free Radic Biol Med* **2016**, *95*, 96–111.

[52] Maurais, A. J.; Bak, D. W. `cimage_annotation`., Computer Program, 2021.

[53] Takada, Y.; Ye, X.; Simon, S. The integrins. *Genome Biol* **2007**, *8*, 215.

[54] Zhang, X.; Liu, K.; Zhang, T.; Wang, Z.; Qin, X.; Jing, X.; Wu, H.; Ji, X.; He, Y.; Zhao, R. Cortactin promotes colorectal cancer cell proliferation by activating the EGFR-MAPK pathway. *Oncotarget* **2017**, *8*, 1541–1554.

[55] Campeau, E.; Ruhl, V. E.; Rodier, F.; Smith, C. L.; Rahmberg, B. L.; Fuss, J. O.; Campisi, J.; Yaswen, P.; Cooper, P. K.; Kaufman, P. D. A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS One* **2009**, *4*, e6529.

[56] Yang, F.; Gao, J.; Che, J.; Jia, G.; Wang, C. A Dimethyl-Labeling-Based Strategy for Site-Specifically Quantitative Chemical Proteomics. *Anal Chem* **2018**, *90*, 9576–9582.

[57] Tabb, D. L.; McDonald, W. H.; Yates J. R., 3. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **2002**, *1*, 21–6.

[58] Gao, J.; Liu, Y.; Yang, F.; Chen, X.; Cravatt, B. F.; Wang, C. CIMAGE2.0: An Expanded Tool for Quantitative Analysis of Activity-Based Protein Profiling (ABPP) Data. *J Proteome Res* **2021**, DOI: 10.1021/acs.jproteome.1c00455.

[59] Bechtel, T. J.; Li, C.; Kisty, E. A.; Maurais, A. J.; Weerapana, E. Profiling Cysteine Reactivity and Oxidation in the Endoplasmic Reticulum. *ACS Chem Biol* **2020**, *15*, 543–553.

Chapter 3

Regulation of Protein Function by Targeted Methionine Oxidation

3.1 Introduction

The other sulfur containing amino acid, methionine is also subject to oxidative modifications. Natural production of reactive oxygen species (ROS) through cellular respiration or enzymatic processes is well known to play a role in oxidative stress as well as cell signaling.[1] While the role of reversible cysteine oxidation is well established and widely studied,[2] the oxidative modification of other residues are less well characterized. Although methionine is known to be oxidized *in vivo* both under stress and normal conditions, the lower reactivity of the thioether on methionine relative to that of the thiolate on cysteine represents a paradox.[3] This chapter will discuss the targeted oxidation and reduction of protein methionine residues, and the role that methionine oxidation plays is regulating function.

3.2 Sources of methionine oxidation

It is generally accepted that oxidants such as H_2O_2 or hypochlorous acid, damage protein and free methionine in a nonspecific manner, and damaged proteins are repaired by methionine sulfoxide reductases (Msrs) (Figure 3.1, 3.3). In this model, oxidation is a chemical step and the reduction is an enzymatic step. The model is influenced by a similar paradigm in cysteine mediated redox signaling. The high intrinsic nucleophilicity of the cysteine thiolate mediates reactions with H_2O_2 and,[5] reactive nitrogen species (RNS).[6] Moreover the microenvironment proximal to a given cysteine can further increase reactivity towards specific oxidants, which in effect transforms the chemical step into a catalytically-driven reaction.[7,8] However a similar model where methionine is oxidized randomly in a chemical process, then repaired by an enzymatic one is not supported by experimental evidence. The thioether on the methionine side chain is much less nucleophilic than cysteine, and the thioether has no acid base equilibrium at physiological pH. In the reaction between oxidants such

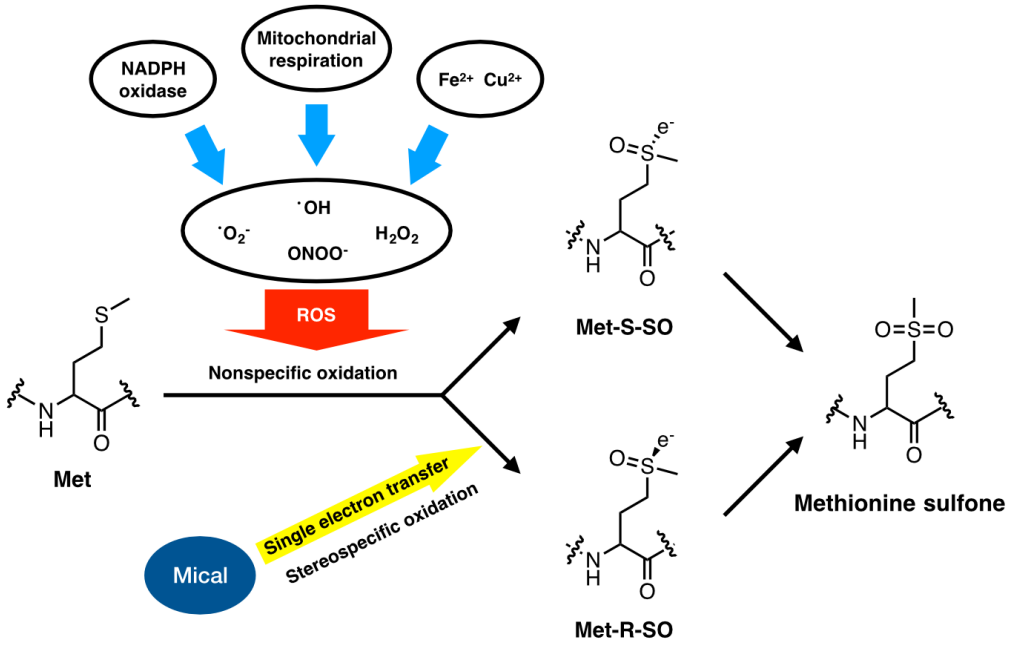


Figure 3.1: Biological sources of methionine oxidation. Nonspecific oxidation by reactive oxygen species (ROS) produced by NADPH oxidases, components of the electron transport chain, or by fenten chemistry will result in a mixture of the methionine R and S diastereomers. Direct oxidation by Micals will result in the formation of exclusively the R isomer. Further oxidation to methionine sulfone is irreversible. Figure adapted from Kaya *et. al.*[4]

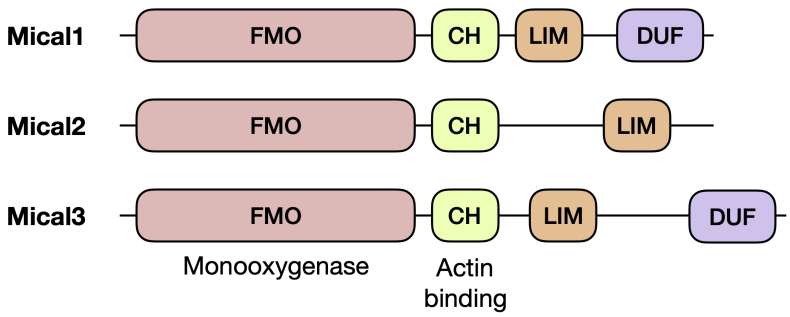


Figure 3.2: Conserved domain topology of mammalian Micals. Mamals have 3 Mical genes comprised of flavin-containing monooxygenase (FMO) domains, actin binding calponin homology (CH) domains, and LIM, and DUF3585 domains. Figure adapted from Manta *et. al.*[3]

as H_2O_2 during oxidative stress, the sulfide on methionine acts as a nucleophile resulting in the formation of methionine sulfoxide (MetO). Due to the additional lone pair of electrons on sulfur, sulfoxides adopt a trigonal pyramidal geometry. The energy barrier to pyramidal inversion of sulfoxide is between 34–43 $\mathrm{kcal\,mol^{-1}}$, which is sufficiently high for methionine sulfoxide to exist as R and S diastereomers under physiological conditions.[9,10] After an oxidation with H_2O_2, MetO would exist as a racemic mixture of the Met-(S)-O and Met-(R)-O diastereomers. However the activity of Msrs is steriospecific, which suggests that the chirality of MetO is biologically significant.[10] In addition, methionine reacts with H_2O_2 with a rate constant of $\sim 1 \times 10^{-2}\,\mathrm{M^{-1}\,s^{-1}}$,[11] which is 3 to 4 orders of magnitude slower than peroxiredoxins or glutathione peroxidase.[12–14] Although a small fraction of methionines may be oxidized in cells, significant oxidation by chemical oxidants is unlikely to occur even under conditions of oxidative stress.[3]

A more likely scenario is that both the oxidation and reduction of methionine are enzyme catalyzed. Indeed, the Mical family of flavin-containing monooxygenase (FMOs) have recently gained attention as redox modulators of actin polymerization.[15–17] Humans have 3 Mical family members each with a FMO domain, calponin homology (CH) actin binding domain, a LIM domain and a C-terminal domain of unknown

function (3.2). An interaction between *Drosophila* Mical and Plexin which is a transmembrane signaling proteins involved in cytoskeleton remodeling, was first discovered by Terman *et. al.*[18] It was later shown that Mical catalyzes the steriospecific oxidation of two methionine residues in actin which are conversed across both vertebrates and invertebrates.[15] Unlike diffuse production of ROS during cysteine mediated redox signaling, Mical must be in close contact with actin for the oxidation to occur.[16] In addition actin treated with H_2O_2 did not mimic the effect of oxidation by Mical,[15,16] and treatment with reduction and ROS scavengers did not effect the activity of Mical.[16,19] Lee *et. al* later determined that mammalian Mical1 and 2 also exhibit the same regulatory redox activity against human actin.[17]

3.3 Methionine sulfoxide reductases

There are three classes of enzymes methionine sulfoxide reductase A (MsrA), methionine sulfoxide reductase B (MsrB), and free Met-R-SO reductase (fRMsr), which catalyze the reduction of MetO back to Met (Table 3.1). Each enzyme class exhibits stereospecific activity on their respective substrates (3.3). MsrA was first discovered over 40 years ago, and charcterized with respect to its activity towards *E. coli* ribosomal protein L12.[20] MsrA is selective for the Met-*S*-SO stereomer and is active towards both protein MetO and free MetO. The enzyme has such a broad substrate specificity that it is also active against small molecules containing *S*-methylsulfinyl groups.[10,21] The implications of MsrA activity on drug metabolism are that if a compound is pharmacologically active with a methylsulfinyl group it should be administered as

Msr	**Substrate**	**Subcelluar localization**
MsrA	Protein and free Met-*S*-SO	Cytosol, nucleus, and mitichrondria
MsrB1	Protein Met-*R*-SO	Cytosol and nucleus
MsrB2	Protein Met-*R*-SO	Mitochondria
MsrB3	Protein Met-*R*-SO	ER and mitochondria
fRMsr	Free Met-*R*-SO	Cytosol and nucleus (only single celled eukaryotes)

Table 3.1: Stereospecificity and subcellular distribution of methionine sulfoxide reductases.

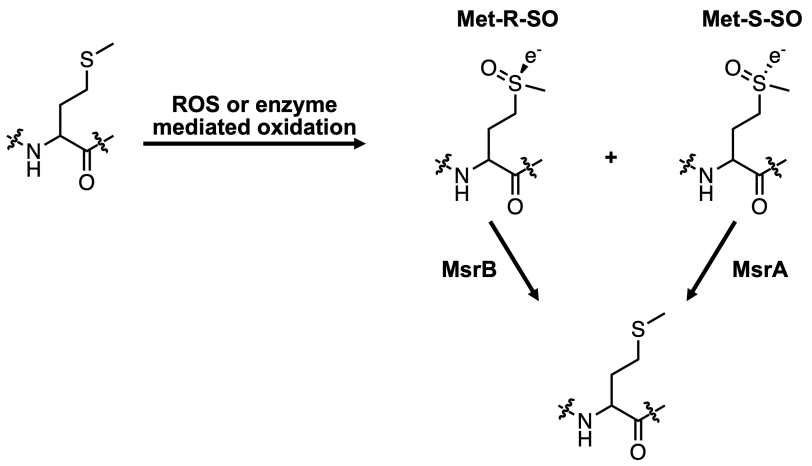


Figure 3.3: Activity of mammalian methionine sulfoxide reductase (Msr). MsrA and MsrB steriospecifically reduce Met-*R*-SO and Met-*R*-SO respectively.

the *R*-sulfoxide, whereas if the thioether analog is active, it should be administered as a pro-drug in the form of *S*-sulfoxide.[21] In addition to broad reductase activity, MsrA also plays a regulatory role as a bidirectional stereospecific Met oxidase in and reductase calmodulin (CaM).[22,23]

Mammals express three MsrB isoforms each of which have different subcellular localization. MsrB1 is localized to to the cytosol and nucleus, MsrB2 to the mitochondria, and MsrB3 to the endoplasmic reticulum (Table 3.1). Unlike MsrA, MsrB only is active towards protein Met-*R*-SO.[4] Mammalian MsrB1 is a selenoprotein that contains a selenocysteine (Sec) as the active site residue. The insertion of Sec is mediated by a selenocysteine insertion sequence) (SECIS) in the 3′-untranslated region of the MsrB1 mRNA transcript in response to a UGA stop codon.[24] Due to its higher pK_a, higher polarizability, and the ease to which it can be reared when overoxidized, Sec is generally more catalytically active than than an analogous Cys residue.[25] The third type of, Msr, fRMsr is the most recently discovered and is only found in single cell eukaryotes.[26] It is not currently clear how or whether mammals are able to reduce free Met-*R*-SO.

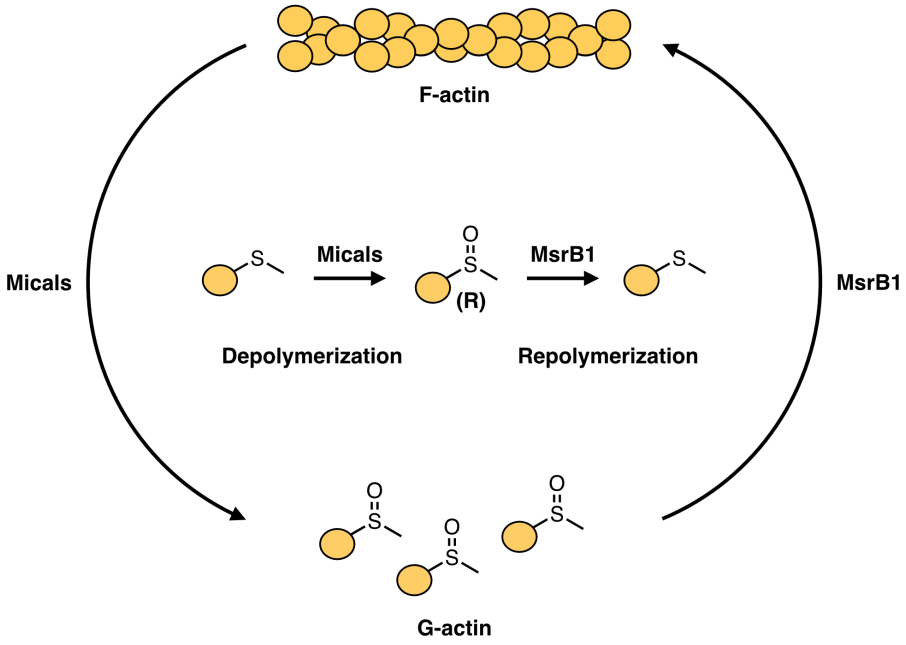


Figure 3.4: Actin polymerization is redox regulated. Figure adapted from Lee *et. al.*[17]

3.4 Regulatory roles of methionine oxidation

There are several well studied examples of oxidation of methionine affecting protein function. One of the most well studied examples is the regulation of CaM and Ca^{2+}/CaM-dependent protein kinase II (CaMKII).[22,23] Methionine oxidation has also been shown to negitivally regulate cytoskeleton organization impacting actin,[27–29] myosin,[30] and cofilin[31] function. Followup studies on the Mical regulation of actin showed that the oxidation is reversed by MsrB1, but not MsrA, indicating that Mical selectivally generates the Met-*R*-SO[17,19] (Figure 3.4). In addition to destabilizing actin filaments, the oxidation events were shown to enhance the binding of the actin severing protein cofilin.[32] Given the important role that cofilin has been shown to play in promoting actin filament disassembly and remodeling, through the recognition of oxidized methionine, in the next chapter we will discuss the use of a peptide base photocrosslinking approach to to identify proteins which selectively recognize oxidized and reduced forms of methionine in actin.

References

[1] Schieber, M.; Chandel, N. S. ROS function in redox signaling and oxidative stress. *Curr Biol* **2014**, *24*, R453–62.

[2] Truong, T. H.; Carroll, K. S. Redox regulation of epidermal growth factor receptor signaling through cysteine oxidation. *Biochemistry* **2012**, *51*, 9954–65.

[3] Manta, B.; Gladyshev, V. N. Regulated methionine oxidation by monooxygenases. *Free Radic Biol Med* **2017**, *109*, 141–155.

[4] Kaya, A.; Lee, B. C.; Gladyshev, V. N. Regulation of protein function by reversible methionine oxidation and the role of selenoprotein MsrB1. *Antioxid Redox Signal* **2015**, *23*, 814–22.

[5] Stone, J. R.; Yang, S. Hydrogen peroxide: a signaling messenger. *Antioxid Redox Signal* **2006**, *8*, 243–70.

[6] Zhou, Y.; Wynia-Smith, S. L.; Couvertier, S. M.; Kalous, K. S.; Marletta, M. A.; Smith, B. C.; Weerapana, E. Chemoproteomic Strategy to Quantitatively Monitor Transnitrosation Uncovers Functionally Relevant S-Nitrosation Sites on Cathepsin D and HADH2. *Cell Chem Biol* **2016**, *23*, 727–37.

[7] Hall, A.; Parsonage, D.; Poole, L. B.; Karplus, P. A. Structural evidence that peroxiredoxin catalytic power is based on transition-state stabilization. *J Mol Biol* **2010**, *402*, 194–209.

[8] Ferrer-Sueta, G.; Manta, B.; Botti, H.; Radi, R.; Trujillo, M.; Denicola, A. Factors affecting protein thiol reactivity and specificity in peroxide reduction. *Chem Res Toxicol* **2011**, *24*, 434–50.

[9] Bentley, R. Role of sulfur chirality in the chemical processes of biology. *Chem Soc Rev* **2005**, *34*, 609–24.

[10] Lee, B. C.; Gladyshev, V. N. The biological significance of methionine sulfoxide stereochemistry. *Free Radic Biol Med* **2011**, *50*, 221–7.

[11] Liang, X.; Kaya, A.; Zhang, Y.; Le, D. T.; Hua, D.; Gladyshev, V. N. Characterization of methionine oxidation and methionine sulfoxide reduction using methionine-rich cysteine-free proteins. *BMC Biochem* **2012**, *13*, 21.

[12] Cox, A. G.; Peskin, A. V.; Paton, L. N.; Winterbourn, C. C.; Hampton, M. B. Redox potential and peroxide reactivity of human peroxiredoxin 3. *Biochemistry* **2009**, *48*, 6495–501.

[13] Peskin, A. V.; Low, F. M.; Paton, L. N.; Maghzal, G. J.; Hampton, M. B.; Winterbourn, C. C. The high reactivity of peroxiredoxin 2 with H_2O_2 is not reflected in its reaction with other oxidants and thiol reagents. *J Biol Chem* **2007**, *282*, 11885–92.

[14] Flohe, L.; Toppo, S.; Cozza, G.; Ursini, F. A comparison of thiol peroxidase mechanisms. *Antioxid Redox Signal* **2011**, *15*, 763–80.

[15] Hung, R. J.; Yazdani, U.; Yoon, J.; Wu, H.; Yang, T.; Gupta, N.; Huang, Z.; van Berkel, W. J.; Terman, J. R. Mical links semaphorins to F-actin disassembly. *Nature* **2010**, *463*, 823–7.

[16] Hung, R. J.; Pak, C. W.; Terman, J. R. Direct redox regulation of F-actin assembly and disassembly by Mical. *Science* **2011**, *334*, 1710–3.

[17] Lee, B. C.; Peterfi, Z.; Hoffmann, F. W.; Moore, R. E.; Kaya, A.; Avanesov, A.; Tarrago, L.; Zhou, Y.; Weerapana, E.; Fomenko, D. E.; Hoffmann, P. R.; Gladyshev, V. N. MsrB1 and MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. *Mol Cell* **2013**, *51*, 397–404.

[18] Terman, J. R.; Mao, T.; Pasterkamp, R. J.; Yu, H. H.; Kolodkin, A. L. MICALs, a family of conserved flavoprotein oxidoreductases, function in plexin-mediated axonal repulsion. *Cell* **2002**, *109*, 887–900.

[19] Hung, R. J.; Spaeth, C. S.; Yesilyurt, H. G.; Terman, J. R. SelR reverses Mical-mediated oxidation of actin to regulate F-actin dynamics. *Nat Cell Biol* **2013**, *15*, 1445–54.

[20] Caldwell, P.; Luk, D. C.; Weissbach, H.; Brot, N. Oxidation of the methionine residues of Escherichia coli ribosomal protein L12 decreases the protein's biological activity. *Proc Natl Acad Sci U S A* **1978**, *75*, 5349–52.

[21] Lee, B. C.; Fomenko, D. E.; Gladyshev, V. N. Selective reduction of methylsulfinyl-containing compounds by mammalian MsrA suggests a strategy for improved drug efficacy. *ACS Chem Biol* **2011**, *6*, 1029–35.

[22] Erickson, J. R. et al. A dynamic pathway for calcium-independent activation of CaMKII by methionine oxidation. *Cell* **2008**, *133*, 462–74.

[23] Lim, J. C.; Kim, G.; Levine, R. L. Stereospecific oxidation of calmodulin by methionine sulfoxide reductase A. *Free Radic Biol Med* **2013**, *61*, 257–64.

[24] Fomenko, D. E.; Novoselov, S. V.; Natarajan, S. K.; Lee, B. C.; Koc, A.; Carlson, B. A.; Lee, T. H.; Kim, H. Y.; Hatfield, D. L.; Gladyshev, V. N. MsrB1 (methionine-R-sulfoxide reductase 1) knock-out mice: roles of MsrB1 in redox regulation and identification of a novel selenoprotein form. *J Biol Chem* **2009**, *284*, 5986–93.

[25] Kim, H. Y.; Gladyshev, V. N. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol* **2005**, *3*, e375.

[26] Le, D. T.; Lee, B. C.; Marino, S. M.; Zhang, Y.; Fomenko, D. E.; Kaya, A.; Hacioglu, E.; Kwak, G. H.; Koc, A.; Kim, H. Y.; Gladyshev, V. N. Functional analysis of free methionine-R-sulfoxide reductase from Saccharomyces cerevisiae. *J Biol Chem* **2009**, *284*, 4354–64.

[27] Dalle-Donne, I.; Rossi, R.; Giustarini, D.; Gagliano, N.; Di Simplicio, P.; Colombo, R.; Milzani, A. Methionine oxidation as a major cause of the functional impairment of oxidized actin. *Free Radic Biol Med* **2002**, *32*, 927–37.

[28] Fedorova, M.; Kuleva, N.; Hoffmann, R. Identification of cysteine, methionine and tryptophan residues of actin oxidized in vivo during oxidative stress. *J Proteome Res* **2010**, *9*, 1598–609.

[29] Ghesquiere, B.; Jonckheere, V.; Colaert, N.; Van Durme, J.; Timmerman, E.; Goethals, M.; Schymkowitz, J.; Rousseau, F.; Vandekerckhove, J.; Gevaert, K. Redox proteomics of protein-bound methionine oxidation. *Mol Cell Proteomics* **2011**, *10*, M110 006866.

[30] Klein, J. C.; Moen, R. J.; Smith, E. A.; Titus, M. A.; Thomas, D. D. Structural and functional impact of site-directed methionine oxidation in myosin. *Biochemistry* **2011**, *50*, 10318–27.

[31] Luo, S.; Uehara, H.; Shacter, E. Taurine chloramine-induced inactivation of cofilin protein through methionine oxidation. *Free Radic Biol Med* **2014**, *75*, 84–94.

[32] Grintsevich, E. E.; Yesilyurt, H. G.; Rich, S. K.; Hung, R. J.; Terman, J. R.; Reisler, E. F-actin dismantling through a redox-driven synergy between Mical and cofilin. *Nat Cell Biol* **2016**, *18*, 876–85.

Chapter 4

HSPA8 and PFKL Act As Readers of Distinct Methionine Oxidation States

4.1 Introduction

Oxidative post-translational modifications (PTMs) play an important role in regulating protein function. Cells produce reactive oxygen species (ROS) as a byproduct of metabolism or in a controlled manner during oxidative signaling. Cysteine is traditionally thought of as the primary ROS target due to its highly reactive thiol side chain. However, the other sulfur containing amino acid methionine (Met) can also undergo reversible oxidation to form Met sulfoxide. Methionine sulfoxide is reduced by Met sulfoxide reductases A (MsrA) and B (MsrB).[1] Spontaneous Met oxidation generates a mixture of diastereomers, Met-S-sulfoxide and Met-R-sulfoxide, which are reduced stereospecifically by MsrA and MsrB respectively.[1] Met oxidation typically inhibits protein function, and Msrs act to restore it. Several examples of targeted, reversible Met oxidation playing a regulatory role on protein function have been reported.[2] Met oxidation has been shown to regulate ROS sensing transcription factors in *E. Coli*,[3] activation of CaMKII[4] and the activity of calmodulin.[5,6]

Of particular interest is the role Met oxidation has been shown to play in regulating actin dynamics. Actin filaments provide mechanical support, are the driving force for cell movement, and are critical for trafficking intracellular vesicles.[7] Eukaryotic actin can exist in either a soluble form (G-actin) or filamentous form (F-actin). Mical monooxygenases stereospecifically oxidize actin M44 and M47 to Met-R-sulfoxide and promote actin filament disassembly.[8,9] MsrB1 reduced both residues and facilitates filament assembly.[10] We are interested in identifying proteins which recognize a specific Met oxidation state in actin and play a role in regulating the transition between F and G actin.

Significant progress in proteomics has been made in detecting PTMs, however identifying the specific "readers" of distinct PTMs remains a challenge. Existing methods for studying protein-protein interactions (PPIs) which involve the cellular

expression of a bait protein, provide no easily accessible methods of controlling the PTM state of said protein in the cellular milieu. In the case of stable, non-reversible PTMs, genetic code expansion approaches,[11] could be a solution. However, in the case of Met sulfoxide, there is no way of guaranteeing that once the protein is expressed, the modification is not reversed by Msr activity. Likewise, oxidation by Micals or nonspecific ROS could oxidize an unmodified bait protein. To circumvent challenges associated with controlling Met oxidation in whole actin either in live cells or lysates, we utilized a peptide based photo-crosslinking approach.[12] We identified PFKL and HSP70s HSPA8 and HSPA1B as putative "readers" of oxidized or reduced methionine in actin respectively.

4.2 Results and Discussion

4.2.1 Design of probes

Our probe design was inspired by similar photo-cross-linking based workflows which have been used to identify readers of trimethylated,[12] and succinylated lysine.[13] The probe has three essential structural features. (1) 2 oxidized or reduced methionine residues; (2) a benzophenone photo-cross-linker, which upon UV irradiation forms a covalent adduct with proximal C–H bonds, enabling the capture of transient probe–protein interactions; and (3) biotin tag, which facilitates detection (by biotin blot), or affinity enrichment by biotin-streptavidin enrichment. Two peptides spanning the sequence from H42 to K52 in human actin B (ACTB), with M46 and M49 reduced (probe **1**) or oxidized (probe **2**) were used to identify proteins which selectively interact the oxidized or reduced ACTB (Figure 4.1A). Both probes showed a concentration dependent labeling of MCF7 lysates including a strongly labeled protein at $\sim$75 kDa in the probe **1** samples (Figure 4.1B).

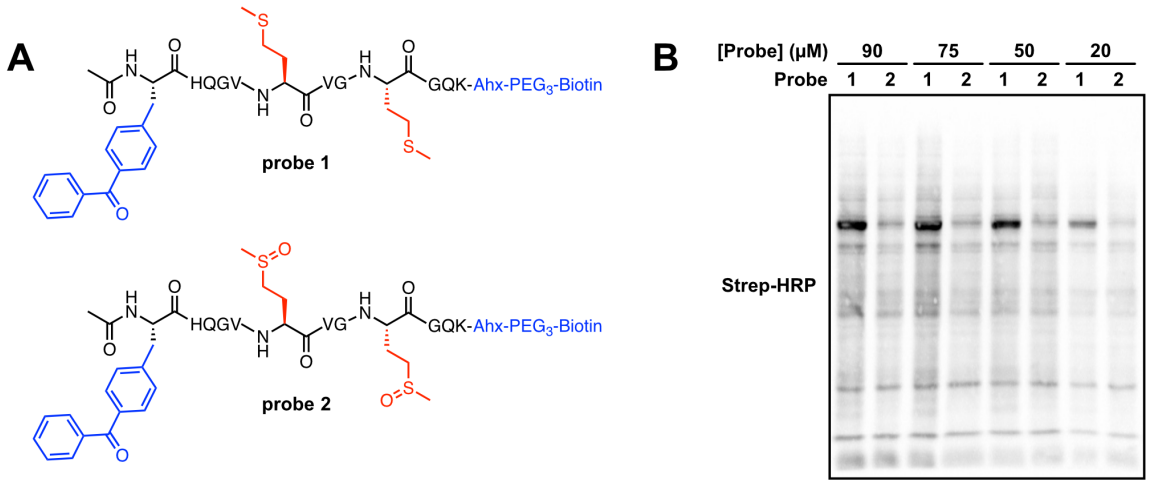


Figure 4.1: An approach to identify readers of methionine oxidation. **(A)** Structures of actin probes **1** and **2**. **(B)** Probes **1** and **2** label MCF7 lysates in a concentration dependent manner and show distinct labeling patterns.

4.2.2 SILAC experiments to identify readers of reduced and oxidized methionine

To achieve a robust analysis of Met(O) and Met specific readers, we performed two independent photo-crosslinking experiments in MCF7 lysates which had been SILAC labeled with isotopically heavy or light arginine and lysine. In a "forward" experiment, probe **1** was added to "light" lysate and probe **2** was added to "heavy" lysate. In a "reverse" experiment, the probes were switched. A plot was generated where the light to heavy ratio ($R_{L/H}$) for the forward experiment is plotted on the x-axis, and the $R_{L/H}$ for the reversed experiment is plotted on the y-axis. Proteins which did not show any selective affinity for either probe are clustered around 1, and proteins which are selective for probe **1** or probe **2** will be clustered at the top right or top left corners respectively (Figure 4.2B, Table 4.1). Of the 5 proteins enriched by either probe, only PFKL has been shown previously to directly bind to actin.[14] PFKL is stabilized by associating with F-actin and other cytoskeletal filaments.[14] HSPA8 has been shown to be involved in regulating myofibril integrity by interacting with actin binding protein CapZ, but a direct interaction with actin has not been

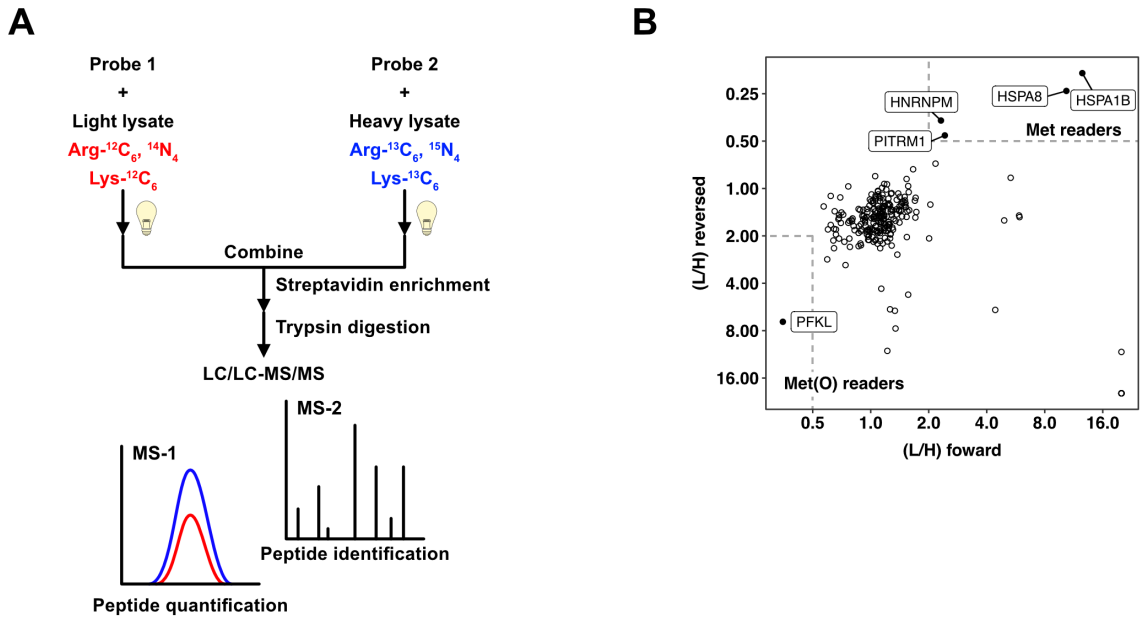


Figure 4.2: Identification of readers of reduced and oxidized methionine. **(A)** Workflow used to identify proteins which interact selectively with oxidized or reduced methionine. 50 µM probes **1** or **2** are crosslinked in isotopically heavy or light MCF7 lysates. After UV crosslinking, equal amounts of heavy and light lysate are combined. Proteins which are bound to the probe are enriched on streptavidin agarose beads, trypsinized, and subjected to LC-MS/MS analysis. **(B)** 2D plot showing SILAC $R_{L/H}$ for each protein in the "forward" (x-axis) and "reversed" (y-axis) experiments. Proteins which selectively interact preferentially with probe **1** are in the upper right portion of the plot and preferential interactors of probe **2** are in the bottom left portion.

ID	Protein	Description	$R_{L/H}$ forward	$R_{L/H}$ reverse
P08107	HSPA1B	Heat shock 70 kDa protein 1A/1B	12.54	0.18
P11142	HSPA8	Heat shock cognate 71 kDa protein	10.36	0.24
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	2.31	0.37
Q5JRX3	PITRM1	Presequence protease, mitochondrial	2.42	0.46
P17858	PFKL	6-phosphofructokinase, liver type	0.35	7.03

Table 4.1: Proteins which are selective interactors of probe **1** or **2**. The average SILAC $R_{L/H}$ from 3 technical replicates are shown.

reported.[15]

4.2.3 Verification of interacting proteins identified by LC-MS/MS

To confirm the MS results, we performed a crosslinking experiment, biotin enrichment, and eluted bound proteins from the beads by boiling in SDS. A western blot was then performed to detect HSPA8 and PFKL in the enriched and unenriched fractions. HSPA8 was shown to selectively bind to probe **1** and PFKL was shown to selectively bind to probe **2** (Figure 4.3), which provided a secondary confirmation of the results of the MS experiment.

4.2.3.1 Crosslinking With recombinant HSPA8 and PFKL

Next we measured the ability of probes **1** and **2** to bind to recombinant HSPA8 and PFKL. HSPA8 with an N-terminal HA tag was overexpressed in BL21 cells. PFKL was expressed in sf9 cells via baculovirus transduction.[16] To verify that PFKL was active we performed an activity assay at varying concentrations of fructose 6-phosphate (F6P) using an established protocol.[16,17] Briefly, PFKL activity is determined by a coupled enzyme assay in which the F6P and ATP are converted to fructose-1,6-diphosphate (F1,6BP). F1,6BP is converted into glyceraldehyde 3-phosphate (GAPD) by the activity of aldolase (ALDO) and triosephosphate isomerase (TPI). Then, the initial rate of the reaction is measured by monitoring the consumption of NADH by glycerol-3-phosphate dehydrogenase (GPDH) (Figure 4.4A). The V_{max} of PFKL was calculated as $2.45 \times 10^{-3}\,\mathrm{s}^{-1}$. We measured the $K_{m,F6P}$ as 2.3 µM which is consistent

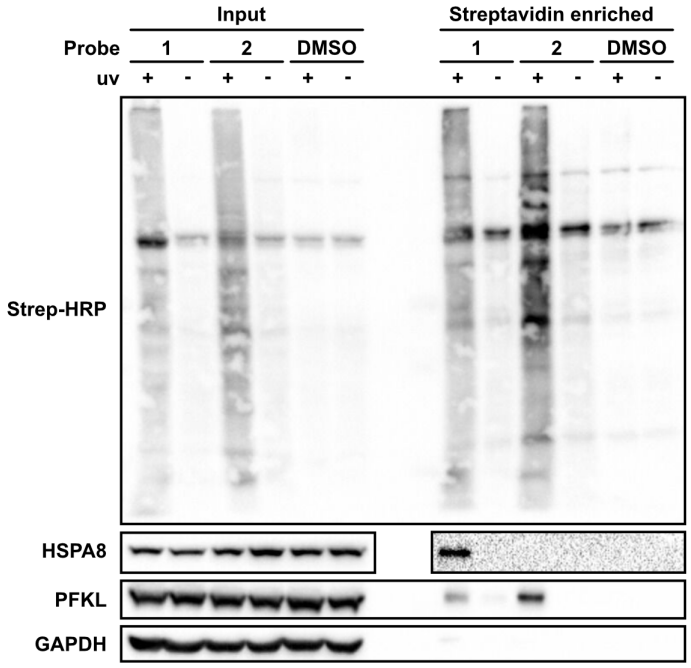


Figure 4.3: Detection of endogenous PFKL and HSPA8 in MCF7 lysates enriched with probe **1** or **2**. MCF7 lysates were crosslinked with 75 µM of the indicated probe. An aliquot of the labeled lysate was taken and run as the "input" fraction. The labeled proteins were pulled down on streptavidin-agarose beads. The bound proteins were eluted by heating in gel loading dye and 50 µM biotin to obtain the "enriched" fraction. Each fraction was analyzed by western blot antibodies for HSPA8 and PFKL and streptavidin-HRP to detect probe labeled proteins.

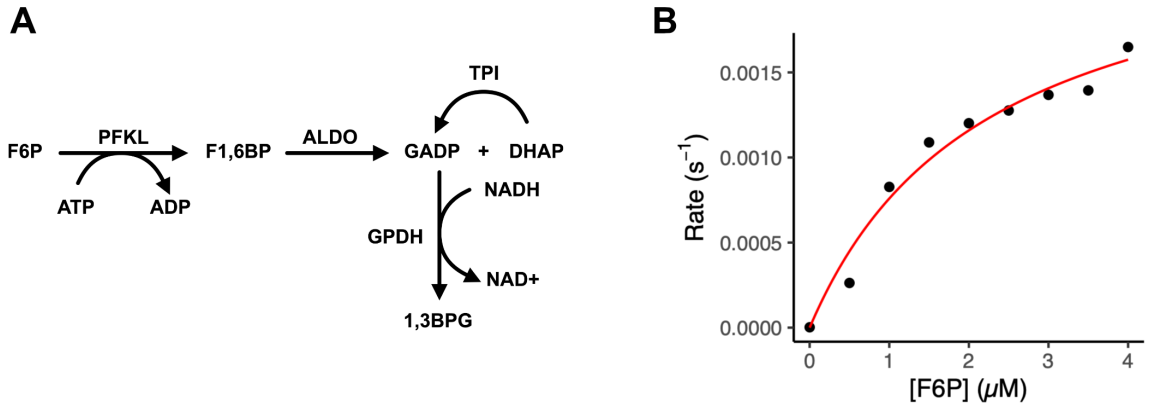


Figure 4.4: PFKL activity assay. **(A)** Coupled enzyme assay used to measure PFKL activity. **(B)** Michaelis–Menten plot of recombinant PFKL activity at varying concentrations of fructose 6-phosphate (F6P).

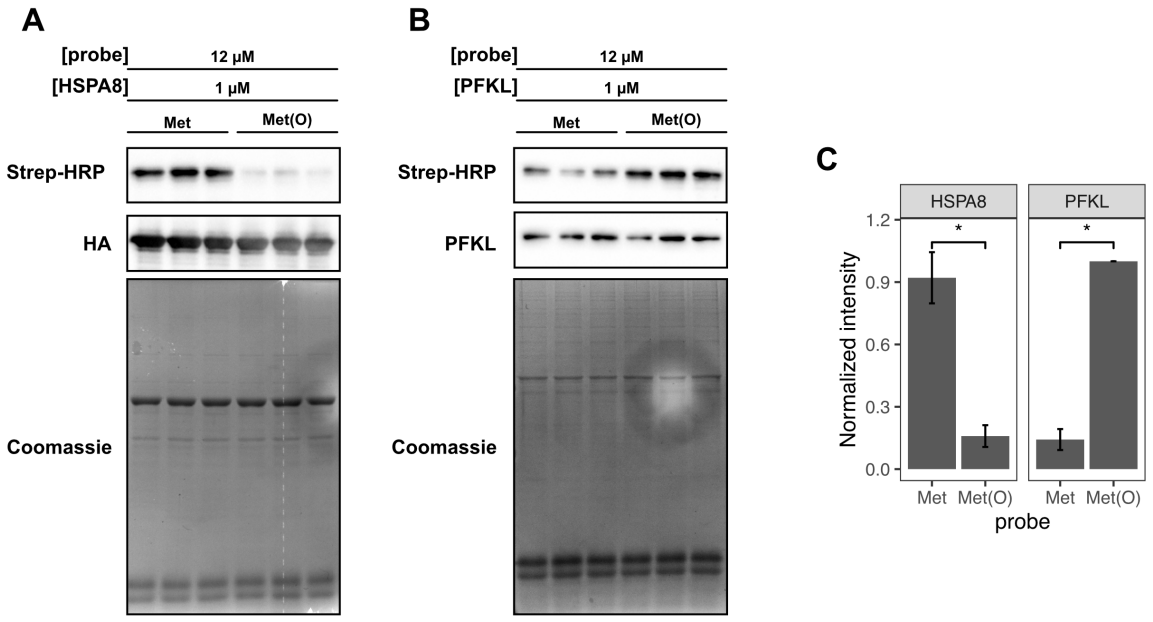


Figure 4.5: Crosslinking with recombinant PFKL and HSPA8. 12 µM of the indicated probe was combined with 1 µM purified **(A)** HSPA8 and **(B)** PFKL *in vitro*. Each lane is a biological replicate performed with a new preparation of protein. **(C)** Mean densitometry measurements of 3 biological replicates shown in **A** and **B**. Band intensities from the streptavidin blot were normalized to the intensity of the HA or PFKL band and then normalized intensity of the most intense replicate, such that the intensity of the most intense replicate is set to 1. Replicates were statistically analyzed by Student's t test ($^*p < 0.005$). Error bars represent the standard deviation of the 3 measurements.

with the value of $2.0 \pm 0.2\,\mu M$ reported by Webb *et al*[16] (Figure 4.4B).

Crosslinking reactions with recombinant protein were performed with 1 µM HSPA8 or PFKL and a 12 molar excess of probes **1** or **2**, and 0.2 mg mL^{-1} casein. The experiment was performed in 3 biological replicates. Both probes showed a statistically significant preference for the expected protein. Probe **1** showed a ~5 fold selectivity for HSPA8 and probe **2** showed a ~2 fold selectivity for PFKL (Figure 4.3).

4.2.4 Mical1 shows oxidase activity on probe 1

A limitation of of our bait peptide approach is that the peptide can only duplicate the amino acid sequence in actin between H42 and K52. Secondary and tertiary structural elements will not necessarily be preserved. Therefore, it can not be assumed that protein interactions with the actin peptide probes will also occur in the full length

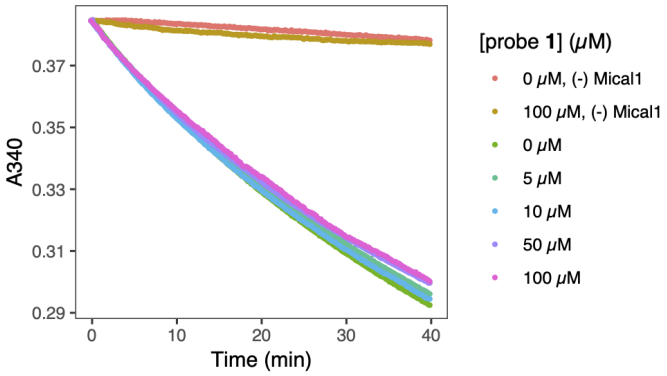


Figure 4.6: Mical1 NADPH consumption in the presence of probe **1**. The consumption of NADPH by the catalytic domain of human Mical1 was monitored by the A340 in the presence of 0 to 100 µM probe **1**, as well as in the presence of 0 and 100 µM probe **1** absent Mical1.

protein. As a first step at investigating how well the peptide probes approximate the full length protein, we tested the oxidase activity of Mical1 against probe **1**. Mical redox enzymes have been shown to induce actin disassembly through targeted oxidation of actin M46 and M49.[8–10,18] Although MICALs have some basal level of NADPH consumption even in the absence of substrate,[18,19] they are known to show an increased rate of NDAPH consumption in the presence of F-actin.[8,9,18,19] The purified catalytic domain of human Mical1 did not show any change in the rate of NADPH consumption at concentrations of probe **1** ranging from 0 to 100 µM probe **1** (Figure 4.6). Even though Mical1 activity was not affected by the presence of probe **1** as a substrate, probe **1** was slightly oxidized by Mical1 (Figure 4.7A,B). Taken together these data suggest that although our peptide probes are not perfect representations of full length actin, they do interact with some of the natural actin binding proteins.

4.2.5 Confirming oxidation state dependent interactions at the protein level

Next we sought to verify that the interactions with the actin peptides also occur at the protein level. PFKL but not HSPA8 were shown to co-immunoprecipitate

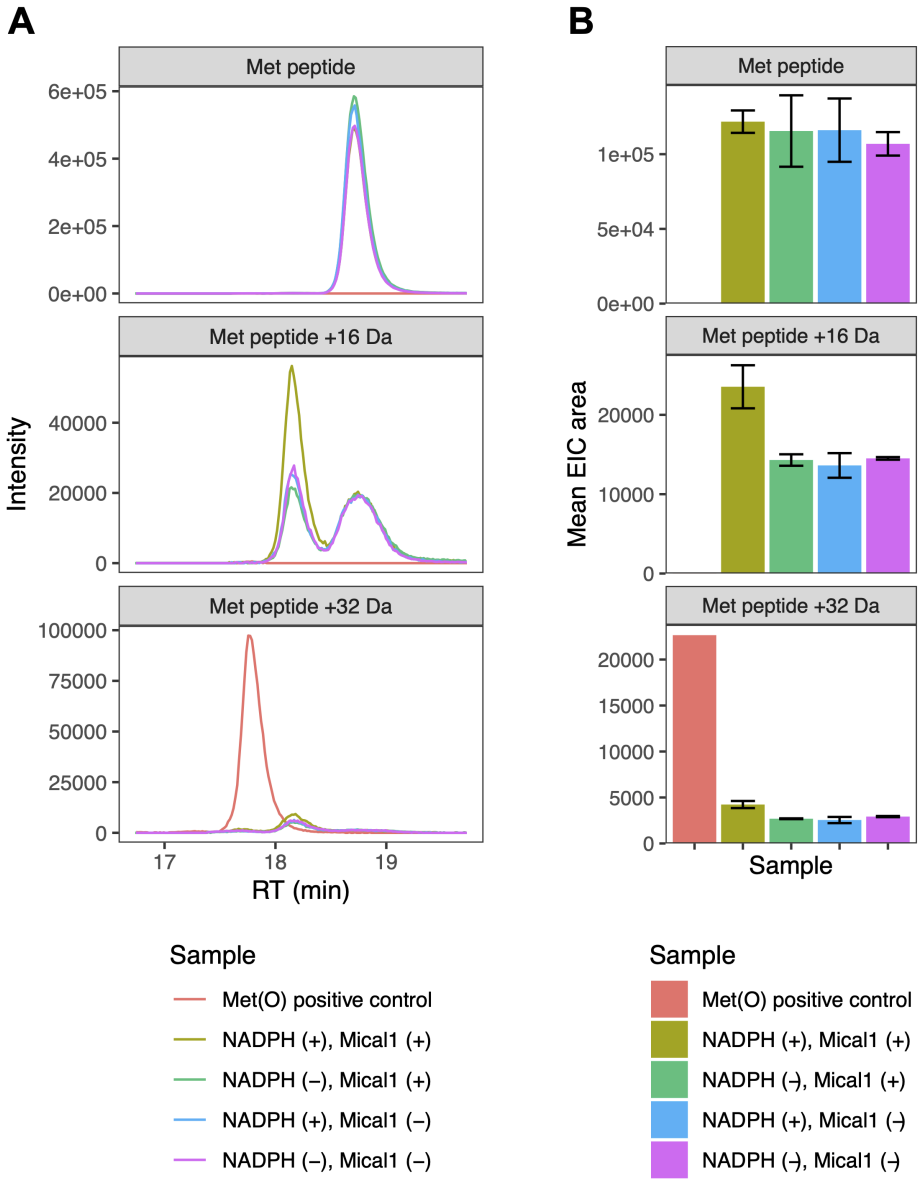


Figure 4.7: MicaI1 oxidizes probe **1**. 100 µM probe **1** was treated with 1 µM MicaI1 in the or absence of 200 µM NADPH. The doubly oxidized probe **2** was run as a positive control. **(A)** Representative extracted ion chromatograms (EICs) of the mass of fully reduced, singly oxidized, and double oxidized probe **1**. **(B)** Mean EIC area of 3 technical replicates. Error bars represent the standard deviation of the 3 measurements.

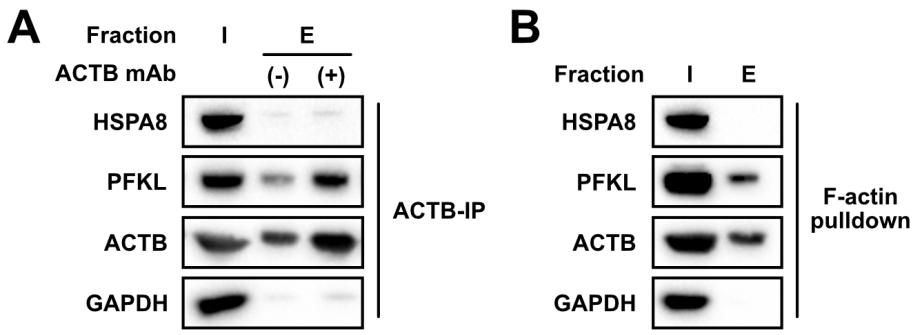


Figure 4.8: PFKL interacts with actin *in situ*. **(A)** Co-immunoprecipitation of ACTB in MCF7 lysates. **(B)** 25 µg biotin-phalloidin was added to MCF7 lysates with then incubated with streptavidin-agarose beads. The input (I) and bead bound (E) fractions were analyzed by western blot.

with ACTB in MCF7 lysates (Figure 4.8A). Biotin-phalloidin can be used to identify proteins which selectively associate with F-actin.[20] Only PFKL was shown to associate with F-actin (Figure 4.8B). PFKL binds to both probes **1** and **2**, although it preferentially binds to the oxidized form of the actin bait peptide. The interaction of all phosphofructokinase (PFK) isoforms with cytoskeletal filaments is well established.[14,21,22] Therefore, it is not surprising that PFKL binds to F-actin to some extent. To better understand how the oxidation state of actin affects PFKL binding, we directly compared the co-precipitation of reduced or Mical oxidized actin with PFKL. PFKL interacts with both forms of actin but interacts to a greater extent with the oxidized form (Figure 4.9A,B).

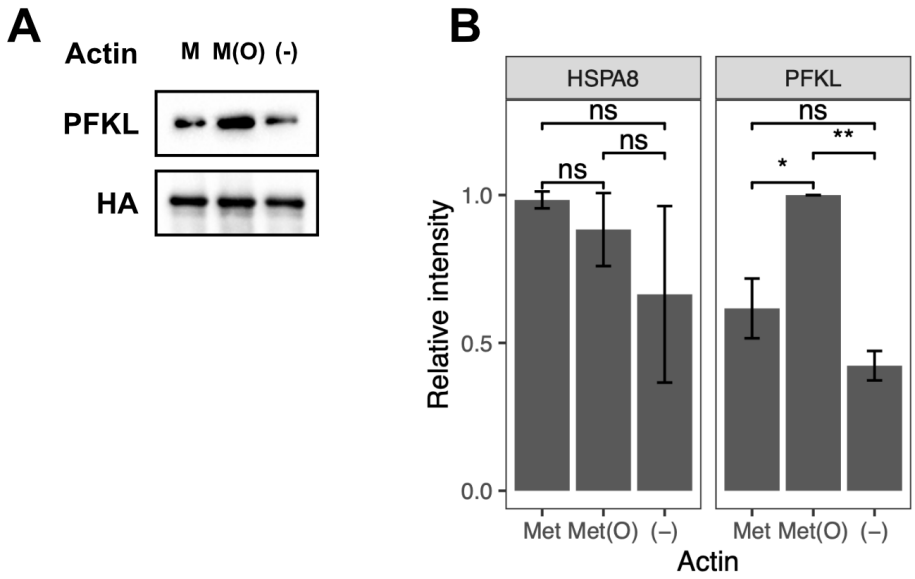


Figure 4.9: Actin pulldown. **(A)** 50 µg reduced or Mical oxidized G-actin was labeled with biotin-NHS then bound to streptavidin-agarose beads. 6 µM recombinant HSPA8 and PFKL were combined and incubated with the immobilized actin and the eluting fractions were analyzed by an anti HA (HSPA8) and PFKL western blot. A representative of 3 biological replicates is shown. The input fraction (I) was the combined HSPA8 prior to addition to pulldown. The elution fraction (E) was the soluble fraction after boiling the streptavidin-agarose beads in gel loading buffer. **(B)** Mean densitometry measurements of 3 biological replicates of the experiment shown in **A**. Replicates were statistically analyzed by Student's t test (*$p < 0.05$, **$p < 0.005$). Error bars represent the standard deviation of the 3 measurements.

4.3 Conclusion

In summary we show that HSPA8 and PFKL interact with actin peptides containing reduced and oxidized methionine respectively. It is well established that PFK interacts with cytoskeletal fillaments.[14,21,22] Several recent studies have also started to unveil a possible interplay between cell metabolism and cytoskeletal organization.[23–25] Indeed, a recent study showed that all 3 PFK isoforms are downregulated by changes in cellular microenvironment via polyubiquitination by TRIM21.[26] The preferential interaction between PFKL and oxidized actin suggests a possible mechanism through TRIM21 is recruited to ubiquitinate PFKL.

Although HSPA8 showed robust and selective labeling by probe **1**, the interaction could not be confirmed at the protein level. There are two possible reasons for this; (1) the interaction between HSPA8 and actin is transient and not strong enough to withstand immunoprecipitation or (2) HSPA8 only interacts with the actin peptide probe, but not the full length protein. Additional experiments are warranted to further characterize the biological consequences of methionine oxidation on the dynamics between actin HSPA8 and PFKL.

4.4 Materials and methods

Synthesis of Met/Met(O) peptide probes

Ac-F(p-Bz)-H-Q-G-V-**M**-V-G-**Met**-G-Q-Ahx-PEG_3-Biotin

Met probe

Ac-F(p-Bz)-H-Q-G-V-**M(O)**-V-G-**Met(O)**-G-Q-Ahx-PEG_3-Biotin

Met(O) probe

Peptides were synthesized by manual solid-phase synthesis on Fmoc-PEG Biotin NovaTag^{TM} resin (Millipore Sigma) using Fmoc as the protecting group for α-amino

groups. The success of each deprotection and coupling was confirmed by Kaiser test following the standard protocol. 6-(Fmoc-amino)hexanoic acid, Fmoc-Lys(Boc)-OH, Fmoc-Gln(trt)-OH, Fmoc-Gly-OH, Fmoc-Met-OH, Fmoc-Gly-OH, Fmoc-Val-OH, Fmoc-Met-OH, Fmoc-Val-OH, Fmoc-Gly-OH, Fmoc-Gln(trt)-OH, Fmoc-His(trt)-OH, Fmoc-p-Bz-Phe-OH residues were added under standard coupling conditions. The same same order of residues was used for the Met(O) probe with the substitution of Fmoc-Met(O)-OH for Fmoc-Met-OH. After the final Fmoc deprotection, dry DCM was added to the resin. Acetic anhydride (10 eq) and pyridine (10 eq) were added dropwise and the vessel was shaken for 3 hours. After acetyl capping the peptide was cleaved from the resin in 90 % TFA, 5 % DCM, 2.5 % TIS, 2.5 % water for 2 hours. The peptide was purified by preparative HPLC with a gradient of increasing ACN 0.1 % TFA (solvent) in water- 0.1 % TFA (solvent A) and analyzed on a Agilent triple TOF mass spectrometer coupled to a Agilent 1200 series LC to give the pure peptide. HRMS for Met probe ($C_{92}H_{144}N_{22}O_{22}S_3 + 3\,H^+$): 669.6667 *m/z* calcd.; obsd. $[M + 3H^+]$ 669.6755, HRMS for Met(O) probe ($C_{92}H_{144}N_{22}O_{24}S_3 + 3\,H^+$): 680.3333 *m/z* calcd.; obsd. $[M + 3H^+]$ 680.3427

Cell culture

MCF7 cells were grown in RPMI media minus L-lysine and L-arginine (Thermo Fisher) supplemented with 10 % dialyzed FBS (Gibco), 1 % Penicillin-Streptomycin-Amphotericin B (Gibco), and either 84 µg mL^{-1} [^{13}C/^{15}N]-L-arginine (R10) and 146 µg mL^{-1} [^{13}C]-L-lysine (K6) or 84 µg mL^{-1} L-arginine (R0) and 146 µg mL^{-1} L-lysine (K0) at 37 °C and 5 % CO_2 for a minimum of 6 passages.

Verification of Heavy Amino Acid Incorporation for SILAC

Soluble cell lysates were diluted to 1 mg mL^{-1} and precipitated with the addition of 10 % TCA followed by incubation at −80 °C for 1 hour. Proteins were pelleted by

centrifugation ($17,000 \times g$ for 10 min) at 4 °C and pellets washed twice with ice-cold acetone. Pellets were air dried at room temperature then resuspended in 8 M urea in PBS (30 µL) and solubilized by heating (65 °C for 5 min) and sonication. Once the pellet had dissolved, 70 µM NH_3HCO_3 (70 µL) was added, and the sample was treated with 15 mM DTT (15 min 65 °C) and 20 mM iodoacetamide (room temperature, 30 min). The urea concentration was diluted to 2 M with PBS and 1 mM $CaCl_2$ and 2 µg trypsin (Promega) were added and the sample was incubated overnight at 37 °C. The trypsin was quenched with the addition of formic acid (15 µL) and the samples were centrifuged at 15000 rpm for 20 min. Our observation of light/heavy intensity ratios ($R_{L/H}$) for most peptides conformed complete metabolic labeling with heavy amino acids (Figure 4A.1).

MS sample prep

Probes **1** or **2** (50 µM) were added to soluble, isotopically heavy or light, MCF7 lysates at 2 mg mL^{-1} (500 µL). The mixture was subjected to UV irradiation for 1 hour on ice. The heavy and light samples were combined and precipitated with the addition of 10 % TCA and incubated at −80 °C for 1 hour. Samples were centrifuged at $17,000 \times g$ for 10 min and the pellets were washed twice with ice cold acetone. After allowing the protein pellets to air dry, pellets were resuspended in 1.2 % SDS by repeated rounds of heating at 65 °C and sonication. The SDS was diluted to 0.2 % with the addition of 5 mL PBS and the solution was incubated with 170 µL streptavidin agarose beads (Thermo Fisher) overnight at 4 °C then 3 hours at room temperature. The beads were washed with PBS (3×5 mL), and water (3×5 mL). The beads were pelleted by centrifugation ($1400 \times g$, 3 min) between washes. The washed beads were resuspended in 500 µL 6 M urea / PBS and 10 mM DTT and incubated at 65 °C for 15 min. 20 mM iodoacetamide was then added and the samples were incubated at 37 °C for 30 min. After reduction and alkylation the beads were pelleted

by centrifugation ($1400 \times g$, 3 min), and resuspended in 200 µL 2 M urea / PBS, 1 mM $CaCl_2$, and 2 µg trypsin (Promega) and incubated overnight at 37 °C. The tryptic digest was collected, and the beads were washed 2×75 µL with water and the washes combined with the tryptic digest for a final sample volume of ∼350 µL. The trypsin was quenched with the addition of 15 µL formic acid and stored at −20 °C until MS analysis.

Liquid chromatography mass spectrometry

Mass spectrometry was performed on an LTQ Orbitrap Discovery (Thermo Fisher) coupled to an Agilent 1200 series HPLC. Protein digests were pressure loaded onto a 250 µm fused silica desalting column, packed with 4 cm Aqua C18 reverse phase resin (Phenomenex). Peptides were eluted onto a 100 µm fused silica biphasic column, packed with 4 cm strong cation exchange resin (Watman) and 10 cm C18 resin using a using a five-step multidimensional LC/LC-MS/MS protocol (MudPIT).[27] Each of the five runs begins with a salt push (0 %, 25 %, 50 %, 80 %, and 100 % buffer C), followed by a gradient 0–100 % buffer B in buffer A (buffer A: 95 % water, 5 % ACN, 0.1 % formic acid; buffer B: 20 % water, 80 % ACN, 0.1 % formic acid; buffer C 95 % water, 5 % ACN, 500 mM ammonium acetate). The flowrate was ∼0.25 µL min^{-1} and the spray voltage was set to 2.75 kV. One survey MS1 scan (400-1800 m/z) was followed by 8 data dependent scans of the n$^{\text{th}}$ most intense ion.

Peptide identification

Peptides were assigned to MS2 spectra using the SEQUEST[28] algorithm, searching against a human reverse-concatenated non-redundant protein database. Data sets were searched independently for peptides containing isotopically light and heavy amino acids. For the light search the default masses of lysine and arginine were used, for the heavy search a static modification was specified for lysine (+6.020 13 Da) and

arginine (+10.008 26 Da). A static modification was specified for cysteine residues (+57.0215 Da, iodoacetamide alkylation) in both heavy and light samples. MS2 matches were assembled by protein and filtered to achieve a false discovery rate of 5 % using DTASelect 2.0.[29]

Ratio quantification

Light/heavy intensity ratios ($R_{L/H}$) were calculated using the CIMAGE quantification package as described previously.[27,30] CIMAGE reports peptide $R_{L/H}$ for co-eluting light and heavy peptides. The peptides are grouped by protein and a representative $R_{L/H}$ for each protein is calculated from the median $R_{L/H}$ for each peptide identified for that protein.

PFKL expression and activity assay

Recombinant PFKL was prepared by baculovirus expression as described previously.[16,17] Briefly, complementary DNA (cDNA) encoding isoform b of *Homo sapiens* PFKL (NP_002617) was inserted into pFastBac HTa vectors. Baculovirus was used to infect sf9 cells at a density of $2 * 10^6$ cells/mL at a multiplicity of infection (MOI) of 2 for 48 hours in a shaking flask at 26 °C.

The activity of recombinant PFKL was confirmed with an auxiliary enzyme assay.[17] Reactions were performed at a final volume of 200 µL containing 50 mM HEPES at pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 0.15 mM NADH, 0.675 U mL^{-1} aldolase, 5 U mL^{-1} triosephosphate isomerase, 2 U mL^{-1} glycerol phosphate dehydrogenase, 0.25 mM ATP, and 0.25 mM ADP. Auxiliary enzymes were de-salted using an Amicon Ultracel-10K Centrifugal Filter Unit prior to use. The consumption of NADH was monitored by the measuring the absorbance at 340 nm using a SpectraMax M5 plate reader (Molecular Devices). The initial reaction rates were measured at various concentrations of F6P. The rate and the concentration of F6P were fit the Michaelis-Menten

equation having the form:

$$r = \frac{V_{max}C}{K_m + C} \tag{4.1}$$

where r is the reaction rate, C is the substrate concentration V_{max} is the maximum rate of the system, and K_m is the Michaelis–Menten constant.

Constructs and antibodies

Mouse Mical1 (NP_612188.1) (residues 1-611),[10] and transcript variant 1 of human HSPA8 (NP_006588.1), were cloned into pET28b and expressed in BL21 *E. Coli*. Streptavidin-HRP Conjugate (EMD Millipore); PFKL, and HSPA8 (1F2-H5) (Novus Biologicals); β-actin mouse mAb (15G5A11/E2) (Invitrogen); β-actin Rabbit Ab, HA-Tag (C29F4), GAPDH (Cell Signaling) antibodies were used for western blotting and/or immunoprecipitation assays. Purified reduced and Mical oxidized rabbit skeletal muscle actin were purchased from Cytoskeleton Inc.

Protein purification

Cell pellets were thawed on ice and resuspended in PBS (5 mL). Cells were lysed by sonication and centrifuged at $10,000 \times g$ for 20 minutes at 4 °C. The supernatant was loaded onto a Ni-NTA column equilibrated with PBS, 10 mM imidazole (pH 7.4). The column was washed with 10 column volumes 25 mM imidazole in PBS, and the purified protein was eluted with 4×1 column volume fractions with 250 mM imidazole in PBS. Elution fractions were analyzed for purity by SDS-PAGE and the protein containing fractions were combined. Protein fractions were desalted to remove imidazole using a NAP-5 column (Thermo Fisher).

Immunoprecipitation

200 µg MCF7 lysate at 1 mg mL^{-1} was incubated with 10 µg of a mouse α-ACTB mAb overnight at 4 °C. The mixture was then incubated with 50 µL protein-G agarose beads (Thermo Fisher) for 2 hours at room temperature. The bound proteins were eluted from the beads by heating to 85 °C in SDS-PAGE loading buffer for 30 minutes. The input and eluting fractions were then analyzed by western blot.

F-actin pulldown

200 µg MCF7 lysate at 1 mg mL^{-1} was incubated with 25 µg biotin-phalloidin (Thermo Fisher) incubated at room temperature for 1 hour. The mixture was incubated with 50 µL streptavidin-agarose beads overnight at 4 °C with constant rotation. The bound proteins were eluted from the beads by heating to 85 °C in SDS-PAGE loading buffer for 30 minutes. The input and eluting fractions were then analyzed by western blot.

Actin pulldown

10 µg reduced or Mical oxidized rabbit skeletal muscle actin was solubilized in G-actin buffer as described previously[8] and diluted to 6 µM, then labeled with 20 molar excess biotin-LC-LC-NHS ester (Thermo Fisher) according to the manufacturer's protocol. The biotinylated actin was desalted on a NAP-5 column to remove excess linker then immobilized on 50 µL streptavidin-agarose beads by incubation at room temperature for 2 hours with constant rotation (200 µL 6 µM actin per sample). Recombinant PFKL and HSPA8 were combined and diluted to a final concentration of 6 µM with actin G-buffer and added to the immobilized actin (200 µL per sample), and incubated overnight at 4 °C overnight with constant rotation. The beads were washed with PBS (3 × 1 mL) and pelleted by centrifugation (1400 × g, 3 min) between washes. The bound proteins were eluted from the beads by heating to 85 °C in

SDS-PAGE loading buffer for 10 minutes. The input and eluting fractions were then analyzed by western blot.

Western blotting

Samples were separated on a 10 % SDS-PAGE gel then transferred to nitrocellulose membranes at 75 V for 100 minutes. The membranes were blocked with 5 % non-fat powdered milk in TBS-T for 1 hour at room temperature. The membrane was probed with primary antibody (1:1000) in 5 % powdered milk / TBS-T overnight at 4 °C. The blot was washed (3×5 min) with TBS-T then incubated with a α-rabbit-HRP conjugated secondary antibody (1:2000) for 2 hours at room temperature. For biotin-blots, after electrotransfer, membranes were blocked overnight with 5 % non-fat dry milk in TBS-T at room temperature. The membrane was then washed (3×5 min) and probed with streptavidin-HRP (1:1000) in TBS-T for 1 hour at room temperature. After probing with secondary antibody or streptavidin-HRP, membranes were washed 3×5 min with TBS-T and treated with chemiluminescence reagents (Thermo Fisher) and developed on a Bio-Rad ChemDoc MP imaging system.

References

[1] Lee, B. C.; Dikiy, A.; Kim, H. Y.; Gladyshev, V. N. Functions and evolution of selenoprotein methionine sulfoxide reductases. *Biochim Biophys Acta* **2009**, *1790*, 1471–7.

[2] Kaya, A.; Lee, B. C.; Gladyshev, V. N. Regulation of protein function by reversible methionine oxidation and the role of selenoprotein MsrB1. *Antioxid Redox Signal* **2015**, *23*, 814–22.

[3] Drazic, A.; Miura, H.; Peschek, J.; Le, Y.; Bach, N. C.; Kriehuber, T.; Winter, J. Methionine oxidation activates a transcription factor in response to oxidative stress. *Proc Natl Acad Sci U S A* **2013**, *110*, 9493–8.

[4] Erickson, J. R. et al. A dynamic pathway for calcium-independent activation of CaMKII by methionine oxidation. *Cell* **2008**, *133*, 462–74.

[5] Lim, J. C.; Kim, G.; Levine, R. L. Stereospecific oxidation of calmodulin by methionine sulfoxide reductase A. *Free Radic Biol Med* **2013**, *61*, 257–64.

[6] Marimoutou, M.; Springer, D. A.; Liu, C.; Kim, G.; Levine, R. L. Oxidation of Methionine 77 in Calmodulin Alters Mouse Growth and Behavior. *Antioxidants (Basel)* **2018**, *7*, DOI: 10.3390/antiox7100140.

[7] Wilson, C.; Terman, J. R.; Gonzalez-Billault, C.; Ahmed, G. Actin filaments-A target for redox regulation. *Cytoskeleton (Hoboken)* **2016**, *73*, 577–595.

[8] Hung, R. J.; Yazdani, U.; Yoon, J.; Wu, H.; Yang, T.; Gupta, N.; Huang, Z.; van Berkel, W. J.; Terman, J. R. Mical links semaphorins to F-actin disassembly. *Nature* **2010**, *463*, 823–7.

[9] Hung, R. J.; Pak, C. W.; Terman, J. R. Direct redox regulation of F-actin assembly and disassembly by Mical. *Science* **2011**, *334*, 1710–3.

[10] Lee, B. C.; Peterfi, Z.; Hoffmann, F. W.; Moore, R. E.; Kaya, A.; Avanesov, A.; Tarrago, L.; Zhou, Y.; Weerapana, E.; Fomenko, D. E.; Hoffmann, P. R.; Gladyshev, V. N. MsrB1 and MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. *Mol Cell* **2013**, *51*, 397–404.

[11] Xiao, H.; Chatterjee, A.; Choi, S. H.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew Chem Int Ed Engl* **2013**, *52*, 14080–3.

[12] Li, X.; Kapoor, T. M. Approach to profile proteins that recognize post-translationally modified histone tails. *J Am Chem Soc* **2010**, *132*, 2504–5.

[13] Kalesh, K. A.; Tate, E. W. A succinyl lysine-based photo-cross-linking peptide probe for Sirtuin 5. *Org Biomol Chem* **2014**, *12*, 4310–3.

[14] Sola-Penna, M.; Da Silva, D.; Coelho, W. S.; Marinho-Carvalho, M. M.; Zancan, P. Regulation of mammalian muscle type 6-phosphofructo-1-kinase and its implication for the control of the metabolism. *IUBMB Life* **2010**, *62*, 791–6.

[15] Hishiya, A.; Kitazawa, T.; Takayama, S. BAG3 and Hsc70 interact with actin capping protein CapZ to maintain myofibrillar integrity under mechanical stress. *Circ Res* **2010**, *107*, 1220–31.

[16] Webb, B. A.; Dosey, A. M.; Wittmann, T.; Kollman, J. M.; Barber, D. L. The glycolytic enzyme phosphofructokinase-1 assembles into filaments. *J Cell Biol* **2017**, *216*, 2305–2313.

[17] Webb, B. A.; Forouhar, F.; Szu, F. E.; Seetharaman, J.; Tong, L.; Barber, D. L. Structures of human phosphofructokinase-1 and atomic basis of cancer-associated mutations. *Nature* **2015**, *523*, 111–4.

[18] Grintsevich, E. E.; Yesilyurt, H. G.; Rich, S. K.; Hung, R. J.; Terman, J. R.; Reisler, E. F-actin dismantling through a redox-driven synergy between Mical and cofilin. *Nat Cell Biol* **2016**, *18*, 876–85.

[19] Wu, H.; Yesilyurt, H. G.; Yoon, J.; Terman, J. R. The MICALs are a Family of F-actin Dismantling Oxidoreductases Conserved from Drosophila to Humans. *Sci Rep* **2018**, *8*, 937.

[20] Clarke, J. P.; Mearow, K. M. Cell stress promotes the association of phosphorylated HspB1 with F-actin. *PLoS One* **2013**, *8*, e68978.

[21] El-Bacha, T.; de Freitas, M. S.; Sola-Penna, M. Cellular distribution of phosphofructokinase activity and implications to metabolic regulation in human breast cancer. *Mol Genet Metab* **2003**, *79*, 294–9.

[22] Lee, J. H.; Liu, R.; Li, J.; Zhang, C.; Wang, Y.; Cai, Q.; Qian, X.; Xia, Y.; Zheng, Y.; Piao, Y.; Chen, Q.; de Groot, J. F.; Jiang, T.; Lu, Z. Stabilization of phosphofructokinase 1 platelet isoform by AKT promotes tumorigenesis. *Nat Commun* **2017**, *8*, 949.

[23] Hu, H. et al. Phosphoinositide 3-Kinase Regulates Glycolysis through Mobilization of Aldolase from the Actin Cytoskeleton. *Cell* **2016**, *164*, 433–46.

[24] Bays, J. L.; Campbell, H. K.; Heidema, C.; Sebbagh, M.; DeMali, K. A. Linking E-cadherin mechanotransduction to cell metabolism through force-mediated activation of AMPK. *Nat Cell Biol* **2017**, *19*, 724–731.

[25] Sullivan, W. J. et al. Extracellular Matrix Remodeling Regulates Glucose Metabolism through TXNIP Destabilization. *Cell* **2018**, *175*, 117–132 e21.

[26] Park, J. S.; Burckhardt, C. J.; Lazcano, R.; Solis, L. M.; Isogai, T.; Li, L.; Chen, C. S.; Gao, B.; Minna, J. D.; Bachoo, R.; DeBerardinis, R. J.; Danuser, G. Mechanical regulation of glycolysis via cytoskeleton architecture. *Nature* **2020**, *578*, 621–626.

[27] Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790–5.

[28] Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994**, *5*, 976–89.

[29] Tabb, D. L.; McDonald, W. H.; Yates J. R., 3. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **2002**, *1*, 21–6.

[30] Gao, J.; Liu, Y.; Yang, F.; Chen, X.; Cravatt, B. F.; Wang, C. CIMAGE2.0: An Expanded Tool for Quantitative Analysis of Activity-Based Protein Profiling (ABPP) Data. *J Proteome Res* **2021**, DOI: `10.1021/acs.jproteome.1c00455`.

Chapter 5

Development of Proteomic Tools to Study Protein Citrullination

A significant portion of the work described in this chapter has been published in:

Tilvawala, R. et al. The Rheumatoid Arthritis-Associated Citrullinome. *Cell Chem Biol* **2018**, *25*, 691–704.

Maurais, A. J. et al. A Streamlined Data Analysis Pipeline for the Identification of Sites of Citrullination. *Biochemistry* **2021**, *60*, 2902–2914.

The first part of the human RA and healthy biotin-PG sample preparation was performed by Son Nguyen. Target validation and SERPIN activity assays were performed by Ronak Tilvawala. Protein purifications, PAD autocitrullination MS sample preparation, and HEK-PAD2 MS sample preparation were performed by Ari Salinger.

5.1 Introduction

Protein citrullination is an enzyme-catalyzed post-translational modification (PTM) that converts an arginine residue into citrulline.[3,4] This irreversible hydrolysis reaction is catalyzed by the protein arginine deiminase (PAD) family of enzymes. There are five PAD isozymes in humans and other mammals, of which, only PADs 1-4 are catalytically active. Of note, PAD activity requires high, micromolar levels of calcium.

Aberrant PAD expression and citrullination is a hallmark of rheumatoid arthritis (RA).[5] In fact, patients with RA generate autoantibodies that target citrullinated proteins, and the occurrence of these anti-citrullinated protein antibodies (ACPAs) is the most specific diagnostic for RA. Antibodies targeting the citrullinated forms of vimentin, α-enolase, and fibrin correlate with different RA subtypes[6] and higher ACPA titers correlate with disease onset and severity.[7] In addition to RA, dysregulated PAD activity is associated with lupus, ulcerative colitis, sepsis, and idiopathic lung fibrosis.[8–13] The exact role of citrullination in driving the pathogenicity of such disparate diseases is not firmly established due to the challenges associated with identifying citrullinated proteins from complex biological samples.

Given the important role of protein citrullination in diverse disease pathologies, numerous chemical probes and mass-spectrometry (MS) strategies have been developed to enrich and identify citrullinated proteins.[4,14–16] Selective chemical derivatization of the urea group in citrulline (Cit) can be achieved with reagents such as 2,3-butanedione and phenyl glyoxal (PG).[17] PG derivatives functionalized with rhodamine and biotin have enabled the visualization and enrichment of citrullinated proteins and peptides.[1,3,18,19] However, assignment of the exact sites of citrullination within these proteins has been confounded by unpredictable fragmentation patterns and poor fragmentation efficiency of peptides containing the PG adduct. Alterna-

tively, the presence of Cit can be detected directly by the +0.984 Da mass shift that occurs upon deimination. However, the small mass shift, coupled with the low abundance of citrullinated peptides within a proteome, necessitate the use of stringent filtering criteria to limit false-positive assignments. Specifically, false-positive identifications can result from: (1) the incorrect assignment of the monoisotopic peak within the isotopic distribution for a given peptide, and (2) the presence of a deamidated glutamine (Gln) and/or asparagine (Asn) within a peptide, which results in a mass shift that is identical to that expected for Cit.

Several strategies have been developed to minimize the number of false-positive hits resulting from database searches for citrullinated peptides. The simplest method is to use stringent precursor mass tolerances of 5 ppm to minimize the incorrect assignment of the monoisotopic species.[20] However, this approach does not account for misannotations driven by the presence of deamidated Gln and Asn. A second approach utilizes a dual database searching strategy, where searches are performed with and without a +0.984 differential modification on Arg, Asn, and Gln. Peptides that score higher with the differential modification on Arg, relative to Asn and Gln, are considered true matches.[20,21] A third approach is to specifically search for diagnostic fragment ions that are unique to citrullination. These diagnostic ions include the immonium ion of Cit, and the −43.0058 Da neutral loss of isocyanic acid (HNCO) from Cit that occurs upon peptide fragmentation. In particular, the isocyanic acid neutral losses can be used as diagnostic fragment ions for the unambiguous detection of a Cit-containing peptide[16,22,23] because the fragmentation of peptides containing deamidated Gln and/or Asn do not generate similar neutral losses. Several recent studies have exploited this Cit-specific neutral loss to identify citrullinated peptides. For example, one study developed a logistic regression model to determine sites of citrullination based on the presence of predictive diagnostic ions, including the Cit-specific neutral loss species.[24] In a second study by Lee *et al*, a library of ~2200

citrullinated peptides were chemically synthesized as reference spectra to unambiguously differentiate citrullinated peptides from their corresponding deamidated counterparts. Manual comparisons of the experimental spectra to the synthetic standards enabled the high-confidence assignment and validation of citrullinated peptides from 30 human tissues. Additionally, these studies confirmed that the neutral loss of isocyanic acid from fragment ions was a key characteristic required for assignment of sites of citrullination with high confidence.[16] The requirement to detect neutral loss fragment ions was subsequently applied in a study by Salinger *et al* to identify sites of citrullination from neutrophil extracellular traps (NETs).[25] Both the Lee *et al* and Salinger et al studies confirmed the importance of applying stringent criteria to minimize false-positives in citrulline assignments, but relied on labor-intensive manual assessment of fragmentation spectra to identify determining neutral loss peaks. A recent study by Chaerkady *et al*[26] identified 833 citrullination sites in ionomycin-treated neutrophil and mast cells using neutral-loss ion-based searches in MaxQuant. Importantly, the generated data were then used to benchmark a machine-learning model to predict sites of citrullination to further facilitate sequence-based prediction of novel citrullination events.

Herein, we describe two studies which take distinct approaches to identifying targets of protein citrullination. In the first study, we describe the use of PG-based probes to identify novel citrullinated proteins from RA patient serum, synovial fluid (SF), and synovial tissues. We identified an increase in many ACPA-target proteins in RA samples consistent with the hypothesis that ACPAs are disease causing. Moreover, we identified numerous metabolic enzymes as well as serine protease inhibitors (SERPINs). Importantly, for a subset of SERPINs, we show that citrullination can abolish their inhibitory activity and thereby modulate SERPIN-regulated pathways including the complement pathway, blood clotting, fibrinolysis, and cell motility. Increased protease activity would allow for enhanced complement activation and degra-

dation of the extracellular matrix that would enable the influx of other immune cells and consequently potentiate the inflammatory response.

In the second study, we report a streamlined computational workflow for assigning sites of citrullination with high confidence, based on two newly developed algorithms, termed ionFinder and envoMatch. The first algorithm, ionFinder, can rapidly identify Cit-containing peptides from tandem MS data by identifying the presence of diagnostic neutral loss ions. The second algorithm, envoMatch, automates the matching of isotopic envelopes to confirm that the monoisotopic species displays the required +0.984 Da mass shift. Both algorithms are now available as open-source programs. The ionFinder program is designed to be implemented downstream of the database-searching step of standard proteomic workflows to differentiate between high-confidence and low-confidence identifications of Cit-containing peptides. As such, ionFinder is compatible with the outputs of common database-searching algorithms, including SEQUEST, Mascot and MaxQuant, and functions on both collision-induced dissociation (CID) and high-energy collisional dissociation (HCD) fragmentation datasets. The accuracy of ionFinder was evaluated using fragmentation spectra from verified Cit-containing, and non Cit containing, synthetic peptides generated in the study by Lee *et al*[16] as well as proteomic data from human cells and tissues.[16,26] To demonstrate the versatility of the ionFinder and envoMatch workflow, we comprehensively mapped the sites of autocitrullination in all four active PAD isozymes. Moreover, we used this workflow to map the citrullinome of cells expressing PAD2, resulting in the identification of over 350 unique Cit-containing peptides from 220 proteins.

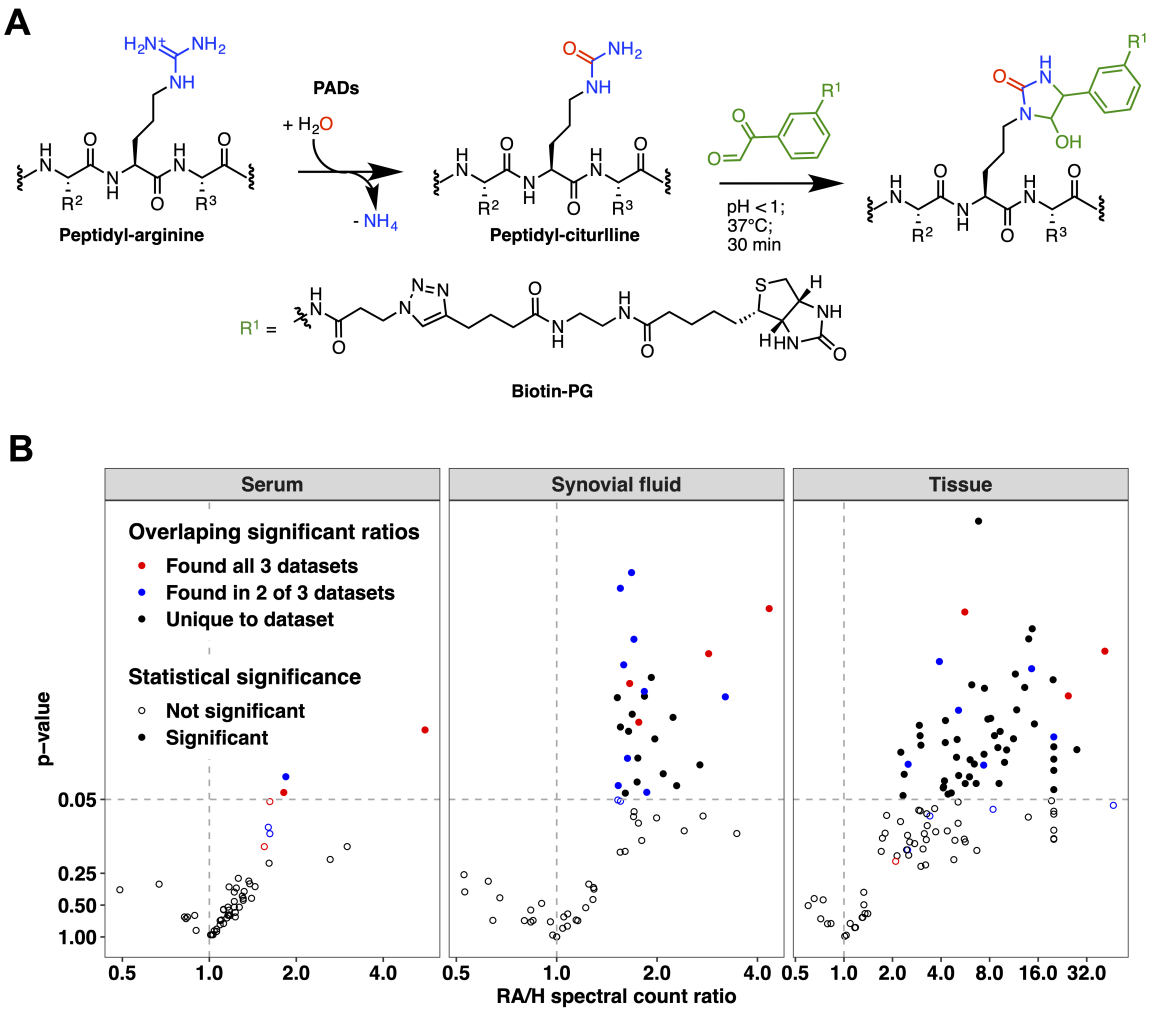


Figure 5.1: **(A)** PADs convert arginine to citrulline. Biotin-PG can selectively react with the urea on citrulline at low pH. Citrullinated proteins can then be isolated and identified by MS. **(B)** Volcano plot showing citrullinated proteins identified in serum, SF, and tissue from RA patients. The spectral count ratio is plotted on the x-axis and the p-value from three replicates of each pool of samples is on the y-axis. Proteins which show a statistically significant increase in citrullination in RA patients are on the top right corner of each plot.

5.2 Results

5.2.1 Identification of citrullinated proteins with biotin-PG

PG probes can be used to chemoselectively label the urea on citrulline under acidic conditions and facilitate the visualization and identification of citrullinated proteins in complex proteomes (Figure 5.1A). We sought to utilize biotin-PG to identify proteins that are selectively citrullinated in human serum, SF (the fluid surrounding joints), and synovial tissue samples (granulation tissue which surrounds joints) in patients with RA. Serum samples were obtained from healthy donors ($n = 16$; 7 females and 9 males) and RA patients ($n = 16$; 9 females and 7 males). The SF samples were collected from healthy donors ($n = 16$; 6 females and 10 males) and RA patients ($n = 16$; 6 females and 10 males). The synovial tissue samples were obtained from one healthy donor and two RA patients. The RA serum, SF and synovial tissue samples were obtained from confirmed RA patients but different individuals. The serum and SF samples were pooled to generate representative samples. RA and healthy controls were labeled with biotin-PG and citrullinated proteins were enriched on streptavidin-agarose. After extensive washing, we performed an on-bead trypsin digest and subsequent multidimensional chromatography coupled with tandem mass spectrometry analysis.

Relative levels of protein citrullination were determined by comparing the spectral counts for a given protein in RA vs. healthy samples. Only proteins with an average of greater than 10 spectral counts in RA samples were used to calculate spectral count ratios. Proteins with a $RA/healthy$ spectral count ratio > 1 and a p value < 0.05 were considered to be selectively citrullinated in the RA sample (Figure 5.1). One readily apparent trend is that the number of citrullinated proteins increases as the sample source progresses from serum to, SF, to synovial tissue. These observations are consistent with increased citrullination occurring within the affected area.

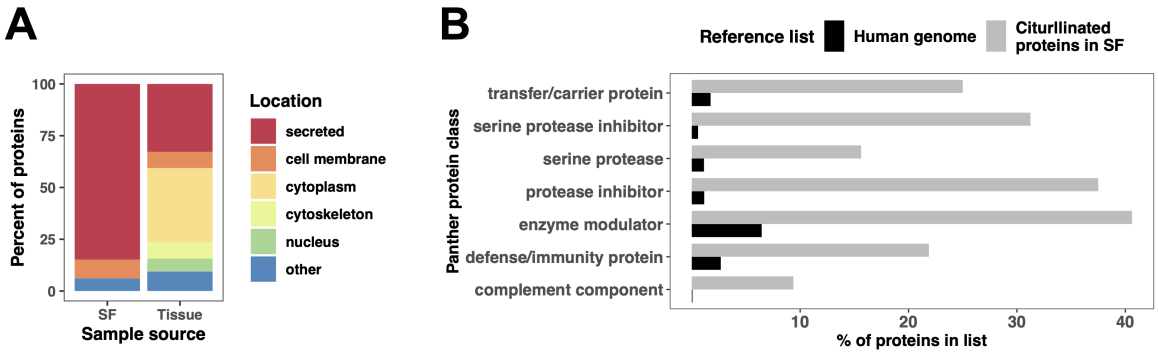
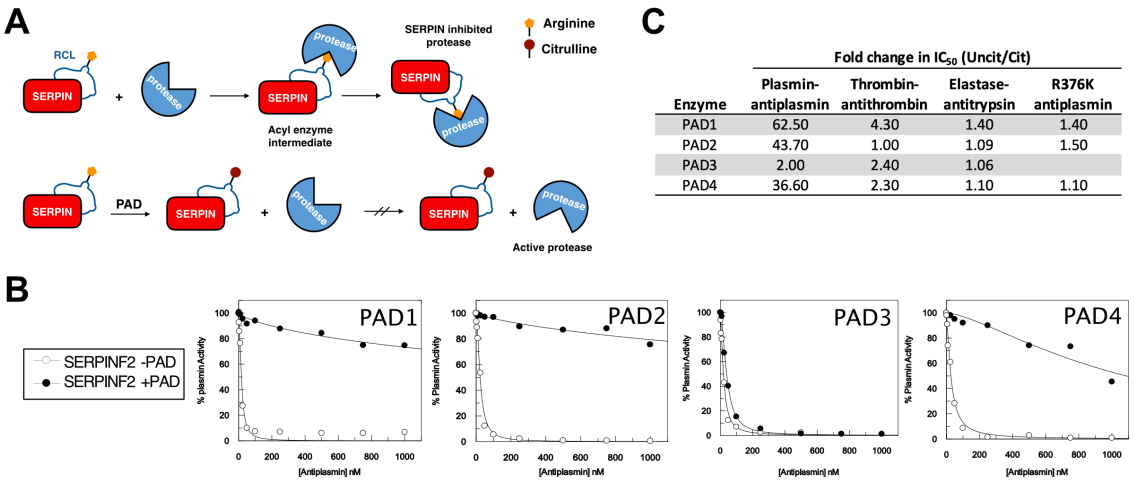


Figure 5.2: **(A)** Subcellular localizations of proteins with elevated citrullination in synovial fluid (SF) and synovial tissue. **(B)** Statistical over representation test for Panther protein classes enriched in SF. The percentage of proteins in which belong to a particular protein class are compared to the percentage of proteins in the human proteome belonging to the same class. The classes which showed the highest degree of enrichment in RA SF are shown.

Notably, several proteins that show high ratios with significant p values in SF, also appear in the serum data at lower levels and at lower confidence. This likely reflects the leaching of these proteins from SF into serum. Among the various other citrullinated proteins discovered, the majority were secreted (85 % in SF and 33 % in synovial tissue) and cytoplasmic (36 % in synovial tissue) proteins (Figure 5.2A), suggesting that both intra- and extracellular citrullination contributes to pathogenesis. We detected elevated levels of known citrullinated proteins including vimentin, fibrin, keratin, and enolase from synovial tissue and SF. In addition to these known substrates, we identified a wide variety of novel citrullinated proteins covering various functional classes, among which were SERPINs, serine proteases, carriers/transporters, and complement system components. A PANTHER overrepresentation test[27] was used to identify protein classes and biological processes that were enriched in SF. The complement pathway, SERPINs, and serine proteases were the most elevated in SF (Figure 5.2B).

SERPINs are ubiquitous secreted proteins which regulate protease activity.[28] SERPINs react with serine proteases through a loop called the reactive center loop (RCL). The RCL is an extended, exposed sequence above the body of the SERPIN scaffold. Upon RCL cleavage, the SERPIN undergoes a conformational change that traps the



Enzyme	Plasmin-antiplasmin	Thrombin-antithrombin	Elastase-antitrypsin	R376K antiplasmin
PAD1	62.50	4.30	1.40	1.40
PAD2	43.70	1.00	1.09	1.50
PAD3	2.00	2.40	1.06	
PAD4	36.60	2.30	1.10	1.10

Figure 5.3: (A) Schematic showing the effect of citrullination on serine protease inhibitor (SERPIN) activity. Citrullination of the P1-arginine in SERPINs abolishes their inhibitory activity against their cognate proteases. **(B)** IC_{50} plots showing inhibition of plasmin by citrullinated and uncitrullinated antiplasmin. Plasmin (50 nM) was mixed with different concentrations of citrullinated or uncitrullinated antiplasmin and incubated for 30 min at room temperature. Hydrolysis of the plasmin substrate D-Val-Leu-Lys-pNA (1 mM) was monitored spectrophotometrically at 405 nm. **(C)** Fold Change in SERPIN IC_{50} for (+/−) pretreatment with various PAD isoforms.

protease in an inactive acyl-enzyme complex. In some SERPINs, the P1 position in the RCL is occupied by an arginine residue, which is responsible for docking the RCL into the protease active site.[29] Given the important role of the P1-arginine in the SERPIN-protease interaction, a possible consequence of citrullination on SERPIN activity is that the binding affinity to their cognate protease would be reduced (Figure 5.3A). To determine how citrullination effects SERPIN activity, antiplasmin, anithrombin, and antitrypsin were recombinant expressed and purified. The IC_{50} of each SERPIN was determined by measuring the activity of the target protease while titrating in increasing concentrations of SERPIN, either pretreated or not pretreated with PAD. Antiplasmin and anithrombin were found to have a reduced IC_{50} when pretreated with PAD (Figure 5.3B,C). However PAD-treated antitrypsin did not show any change in IC_{50} of neutrophil elastase, which is consistent with the fact that antitrypsin has a Met residue in the P1 position instead of an Arg as in antiplasmin and anithrombin. To confirm that SERPIN inactivation is due to citrullination of the

P1-arginine, we generated an antiplasmin variant in which the P1-arginine was mutated to a lysine and evaluated its ability to inhibit plasmin. In contrast to wild-type antiplasmin, citrullination of the R376K mutant does not alter the inhibition profile of the mutant SERPIN; the IC_{50} values for citrullinated and uncitrullinated mutant enzymes are almost identical (Figure 5.3C). In total, these results demonstrate that citrullination inhibits a large subset of SERPINs via citrullination of the P1-arginine.

5.2.2 Development ionFinder

A limitation of the biotin-PG probe used in the previous section is that the peptides actually bearing the citrulline residue remain bound to the streptavidin beads, so the site of citrullination cannot be determined. To facilitate the unambiguous detection of citrulline and automate the process of validating peptide-spectrum matches (PSMs) of citrullinated peptides assigned by database-searching algorithms, we developed a software package named ionFinder. As a starting point for ionFinder, we adapted the empirically determined rules developed by Lee *et al*,[16] where the first step in the workflow was the elimination of peptides annotated as having a C-terminal Cit, because citrullinated Arg residues no longer serve as a recognition element for trypsin. In subsequent filtering steps, a requirement for diagnostic neutral loss (NL) peaks in the fragmentation spectra was instituted to unambiguously assign sites of citrullination. As a first step to automating the identification of NL species, we developed a decision tree to categorize diagnostic fragment ions for peptides annotated to contain Cit. For a given MS2 spectra, each fragment ion is classified into 1 of 4 categories (Figure 5.4): (1) a Cit-determining fragment ion (Det), where a +0.984 Da mass shift from the unmodified Arg peptide can be assigned to a Cit-containing peptide fragment that does not contain Asn or Gln; (2) a Cit-determining neutral loss fragment ion (DetNL), where a Cit-containing peptide fragment shows a −43.0058 Da neutral loss; (3) an artifact neutral loss fragment ion (ArtNL), where a −43.0058

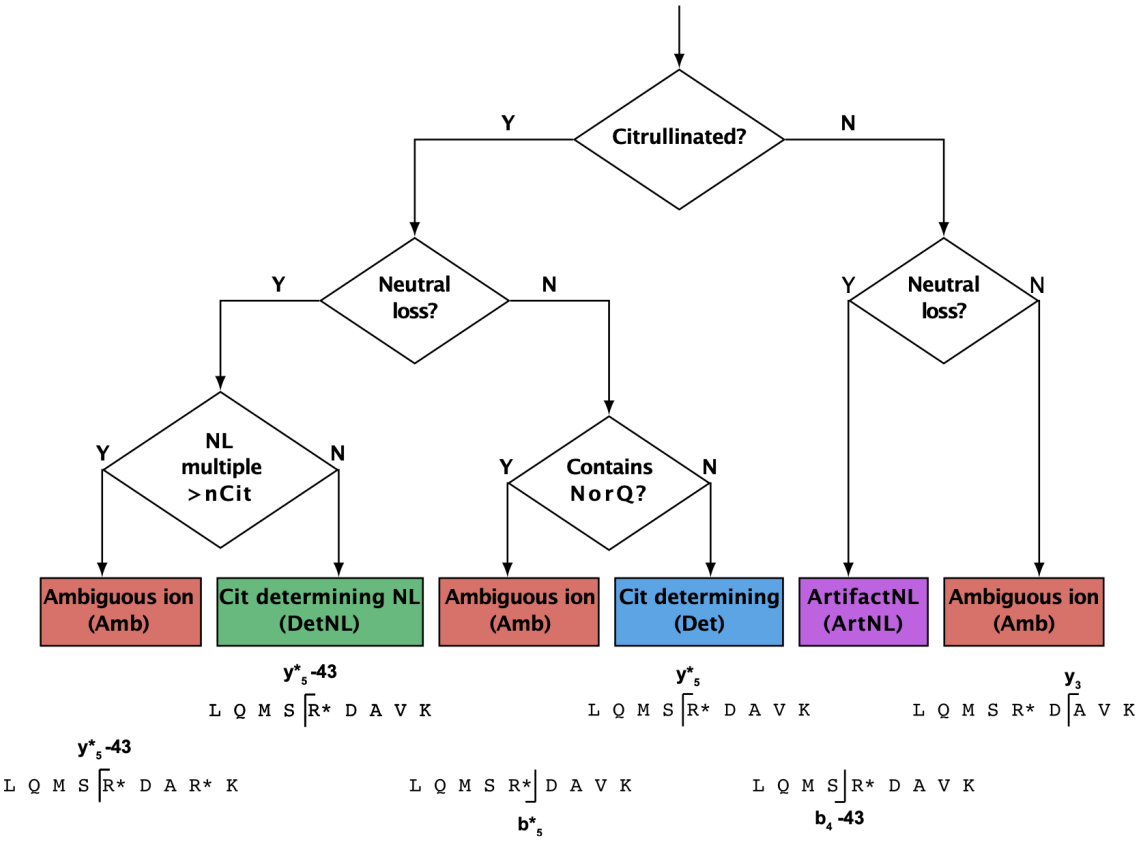


Figure 5.4: A decision tree depicting how each fragment ion within an MS/MS spectrum is binned into one of five categories: a citrulline-determining fragment (Det); (2) a citrulline-determining neutral loss fragment (DetNL); (3) an artifact neutral loss fragment (ArtNL); and, (4) an ambiguous fragment (Amb).

Da neutral loss is observed for a peptide fragment that does not contain Cit; and, (4) an ambiguous fragment ion (Amb), which lacks diagnostic peptide fragments to confidently assign the presence or absence of Cit.

After classifying the fragment ions from each MS2 scan into the different categories, the number of each type of fragment ion (Det, DetNL, ArtNL, and Amb) is then used to assign a value of "true", "likely", "ambiguous", or "false" to each peptide identification according to a second decision tree (Figure 5.5). Spectra containing two or more DetNL fragment ions are classified as "true", whereas spectra with no Det or DetNL fragment ions are classified as "false". Spectra containing a single DetNL fragment ion, or any number of Det fragment ions, are classified as "likely".

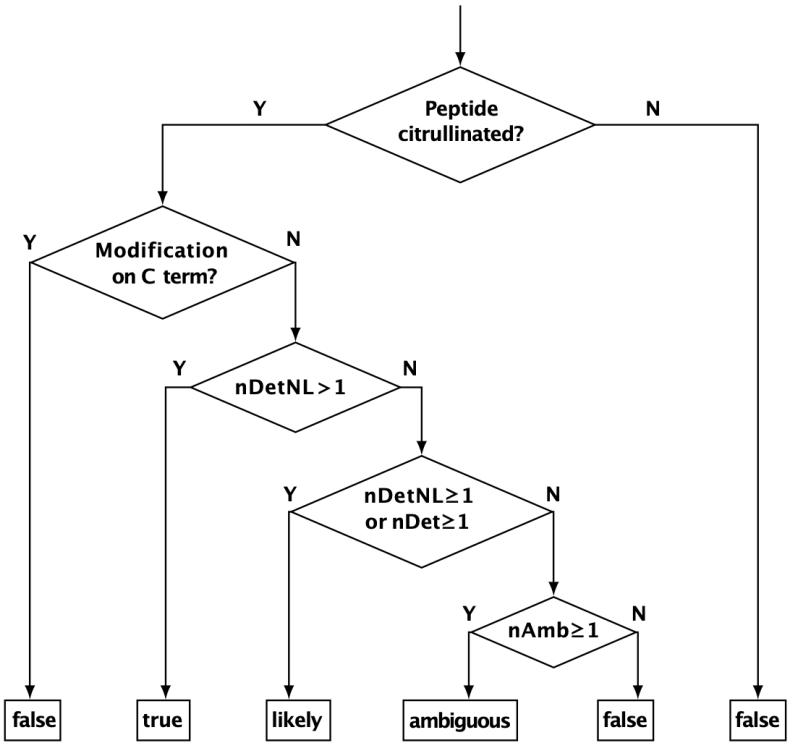


Figure 5.5: A decision tree depicting how each PSM is assigned categories of "true", "likely", "ambiguous" and "false", based on the number of DetNL, Det, and Amb fragments.

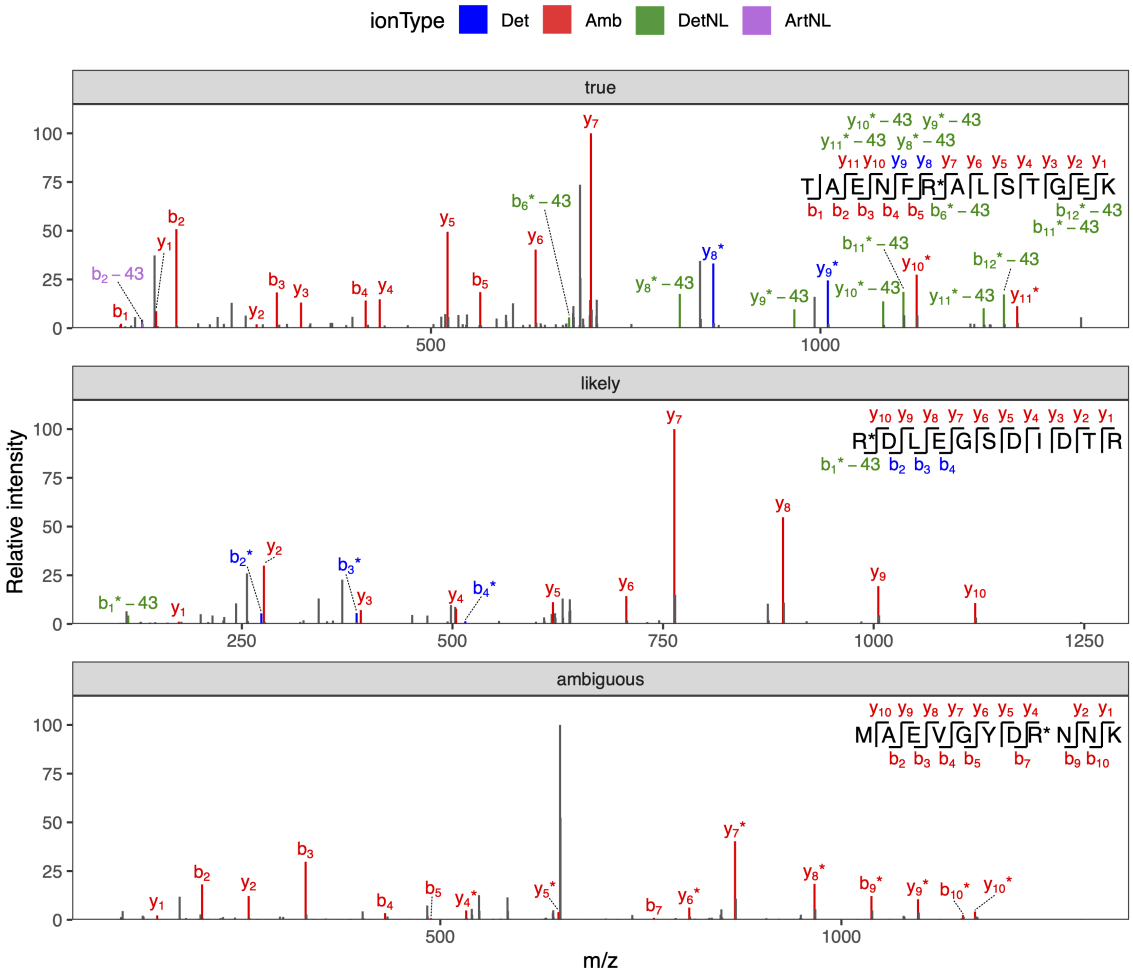


Figure 5.6: Representative fragmentation spectra from peptides classified as "true", "likely", and "ambiguous" by ionFinder.

Lastly, spectra containing no Det or DetNL, and only Amb fragment ions, are listed as "ambiguous". For example, if a peptide is annotated as containing a single Cit, ionFinder rapidly searches the fragmentation data and automates the categorization of each PSM into the "true", "likely", "ambiguous", or "false" categories. For those peptides that contain multiple annotated Cit residues, the categorization of "true", "likely", "ambiguous", or "false" is given to each individual site, and then the value assigned to the peptide as a whole is the lowest classification of all sites. For example, a doubly citrullinated peptide will only be classified as "true" if both individual Cit sites are categorized as "true". Representative PSMs with ion type classifications and the overall value assigned to the spectra are shown in Figure 5.6.

5.2.3 Validation of ionFinder Against Ground Truth Datasets

To validate the utility of ionFinder, we first applied this algorithm to a previously described dataset[16] that includes Cit-containing synthetic peptides, as well as non-Cit containing peptide controls. Specifically, we randomly selected fragmentation spectra from 100 Cit-containing peptides (validated "true" hits), and 100 non-Cit containing peptides that contain one or more deamidated Asn or Gln residues (validated "false" hits). Upon initial application of ionFinder to this dataset, we observed numerous ArtNL fragments, where a -43.0058 Da neutral loss was observed for fragments that did not contain a Cit residue. Upon manual inspection, many of these ArtNL fragment peaks displayed relatively low intensity, compared to the majority of DetNL peaks. Moreover, these ArtNL fragments were a very small fraction of the total NL species, where DetNL fragments clearly dominated (Figure 5.7A-B). Nevertheless, we sought to eliminate these ArtNL fragments because the presence of a NL species is a critical component of our Cit assignment. Initially, we evaluated the use of m/z ppm cutoffs, as well as signal-to-noise cutoffs, to minimize the number of ArtNL fragments and provide greater confidence in the assignment of DetNL fragments.

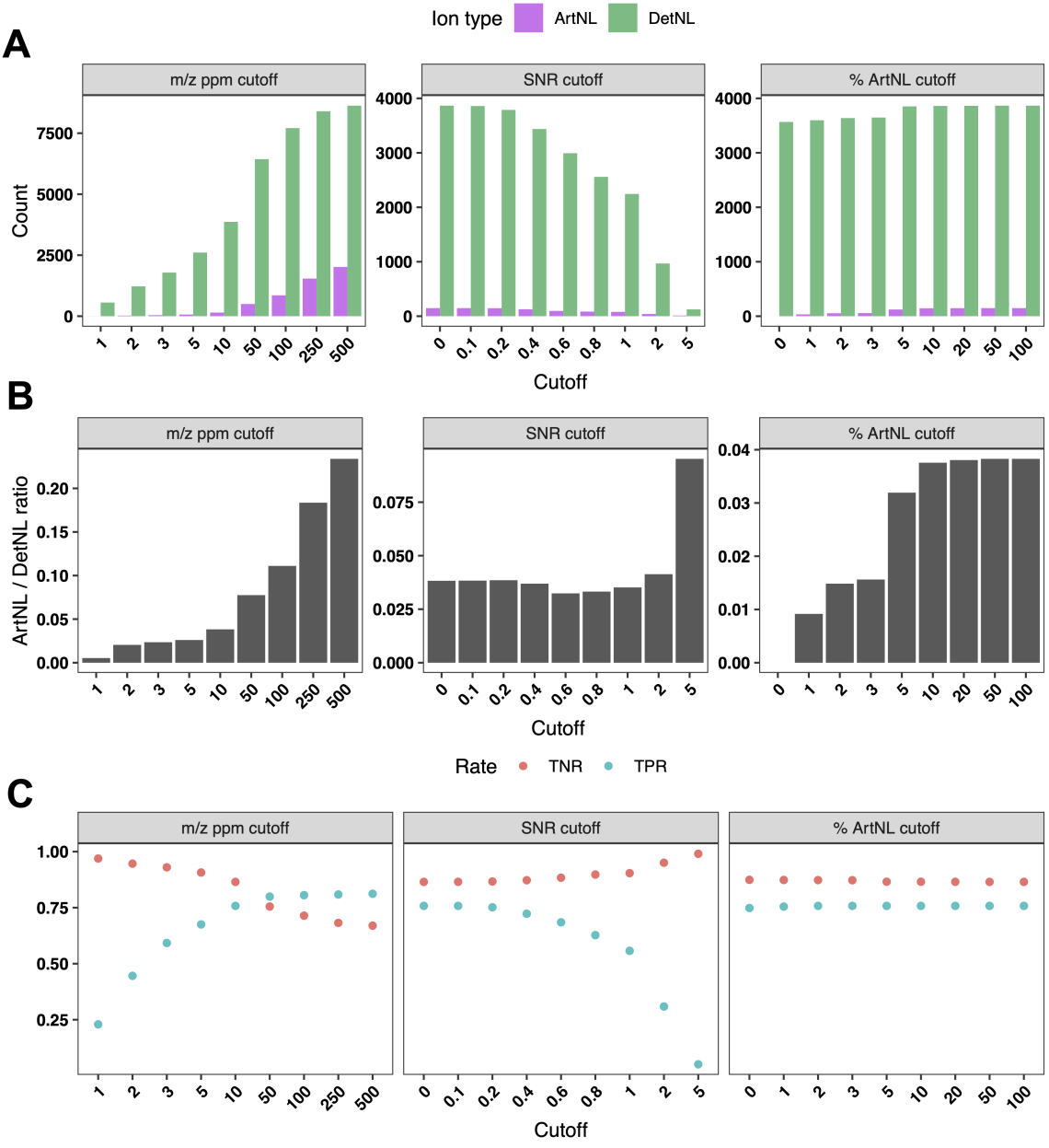


Figure 5.7: **(A)** Bar graphs depicting DetNL and ArtNL fragments upon varying the *m/z* ppm cutoff value (left panels), signal-to-noise cutoff (middle panels), and a dynamic % ArtNL cutoff (right panel). The data are depicted as total counts of fragments. **(B)** Bar graphs depicting DetNL and ArtNL fragments upon varying the *m/z* ppm cutoff value (left panels), signal-to-noise cutoff (middle panels), and a dynamic % ArtNL cutoff (right panel). The data are depicted as ArtNL/DetNL ratios (bottom panels). **(C)** Plot depicting how varying the *m/z* ppm cutoff value (left panels), signal-to-noise cutoff (middle panels), and a dynamic % ArtNL cutoff (right panel) affects the true positive rate (TPR) and true negative rate (TNR) rates.

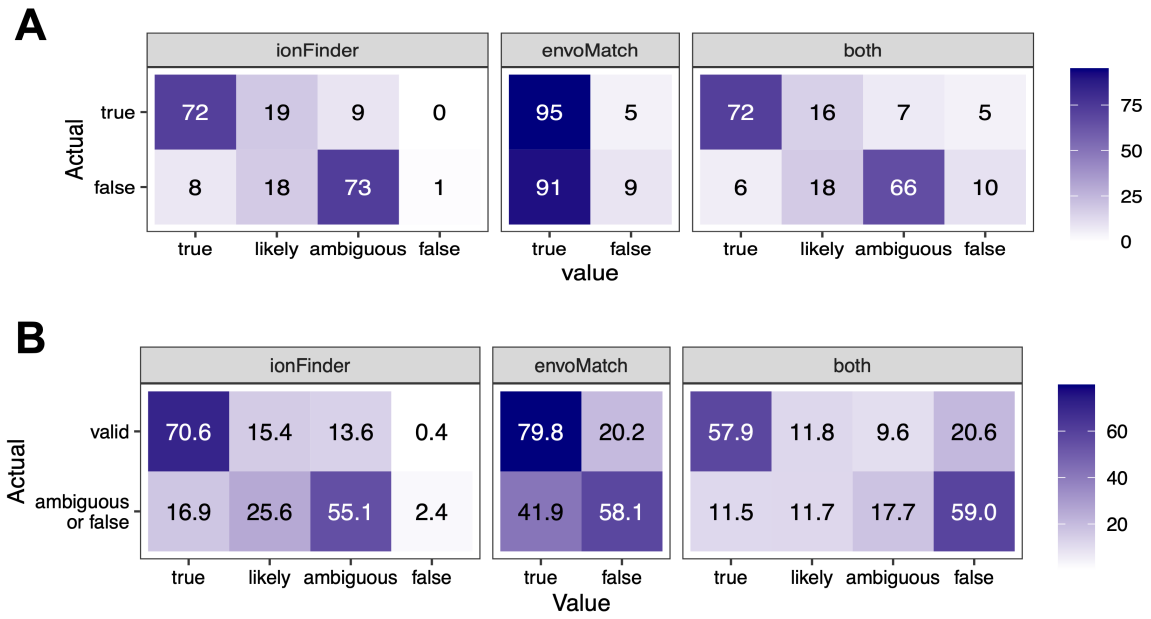


Figure 5.8: Verifying the accuracy of ionFinder and envoMatch using an established "true" and "false" dataset. **(A)** Fragmentation spectra from 100 Cit-containing and 100 non-Cit containing synthetic peptides that contain confirmed deamidation events from Lee *et al*[16] were analyzed by ionFinder, envoMatch, or both. The number of verified "true" and "false" hits that fell into each category by ionFinder and envoMatch are shown. **(B)** 2644 PSMs from a complex proteomic analysis were analyzed by ionFinder, envoMatch, or both. The percentage of verified "true" and "false" hits that fell into each category by ionFinder and envoMatch are shown.

However, the use of more stringent ppm and signal-to-noise cutoffs not only reduced the ArtNL fragment counts, but also led to a corresponding decrease in total fragment counts (Figure 5.7A-C). Consequently, we evaluated the use of a dynamic ion-intensity cutoff, wherein the cutoffs are set to self-adjust to provide an abundance of ArtNL fragments below a specified percentage of the total abundance of all other fragment types. We tested several potential cutoffs before establishing that a 1 % threshold reduced the number of ArtNL fragments significantly, without altering the number of Det and DetNL fragments (Figure 5.7A-C). In summary, the use of these criteria to minimize ArtNL fragments provides higher confidence of NL peak assignments that are used as diagnostic species for the presence of Cit residues.

When this optimized version of ionFinder was applied to the verified "true" and "false" spectra from the study by Lee *et al*,[16] 77 of the 100 "true" spectra were

automatically assigned as "true" by ionFinder (Figure 5.8A; Table 5A.1). By contrast, only 11 of the 100 "false" spectra were assigned as "true" by ionFinder. Examination of the annotated fragmentation spectra confirmed that the "true" spectra contained high-intensity NL peaks, whereas the annotated NL peaks in the "likely" assignments were of low intensity and precluded high-confident assignment of that PSM as a definitive Cit-containing peptide (Figure 5.6). We then expanded our test dataset to include deep proteomic data from human tissues used in the study by Lee *et al*,[16] where Cit-containing peptides were assigned by comparing to synthetic standards. We analyzed 2644 PSMs from these proteomic datasets, of which 456 and 2188 were classified as true and false, respectively. Analysis of these 2644 PSMs with ionFinder resulted in ∼71 % of the verified hits being classified as "true", while only ∼17 % of the false hits were annotated to be "true" (Figure 5.8B).

To further improve the reliability of our assignments, we developed an isotopic envelope matching algorithm, termed envoMatch, that would verify the proper monoisotopic peak identification of Cit-containing peptides. Note that isotopic envelope matching was excluded from the analysis by Lee et al.,[16] but successfully employed by Salinger *et al.*[25] to identify Cit-containing peptides from NETs. In this previous study, isotopic envelope matching was performed by using commercial software to predict MS1 chromatograms for both Arg- and Cit-containing peptides. Then, subsequent manual comparisons of the experimental and predicted envelopes were performed to determine an appropriate fit. The envoMatch program automates this process by comparing experimental isotopic envelopes to predicted Arg- and Cit-containing isotopic envelopes. The program then assigns a value of "true" or "false" quantitatively based on a Pearson correlation between the experimental and predicted isotopic envelopes. When the synthetic sample datasets were analyzed, envoMatch correctly classified the Cit-containing isotopic envelopes with ≤ 95 % accuracy (Figure 5.8A). As expected, 95 % of the "false" dataset was also classified as "true", because these

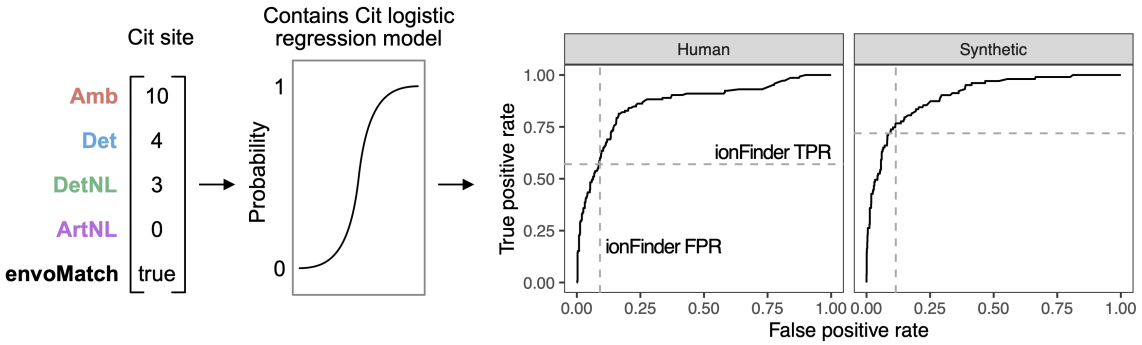


Figure 5.9: Receiver operating curve (ROC) for a logistic regression classifier trained against a complex proteomic dataset (human) and a dataset comprised of synthetic peptides. The true positive rate (TPR) and false positive rate (FPR) observed for ionFinder is also shown on each panel.

"false" peptides were all deaminated, and therefore have the same isotopic distribution as the citrullinated peptide. Isotopic envelope verification through envoMatch becomes more important when analyzing complex proteomic samples where there is a higher likelihood of observing co-eluting isobaric species. When the 2644 PSMs from the complex proteomic sample were analyzed by both ionFinder and envoMatch, the number of verified true positives decreased from 70 % to 58 % (Figure 5.8B). We also observed an ∼5 % decrease in the false-positive rate, where the detNL-containing species did not display the required isotopic distribution. We therefore recommend a workflow by which spectra are analyzed by both ionFinder and envoMatch to generate high-confidence assignments of sites of citrullination, i.e., those that receive annotations of "true" based on an analysis of both MS1 and MS2 spectra.

5.2.4 Comparison of ionFinder and envoMatch to a logistic regression classifier

Huh *et al* used spectra from synthetic peptides containing citrulline or arginine to train a logistic regression model which calculates the probability that a peptide in a given PSM is citrullinated. However, a significant disadvantage of their model is that it was not trained against PSMs from deamidated peptides.[24] We implemented

a modified ionFinder workflow to compare how ionFinder would compare to a binary classifier on data containing deamidated peptides. The ionFinder ion classification decision tree (Figure 5.4) would still be for feature engineering and reducing MS2 spectra into a list of ion type classifications, but the empirically designed decision tree in Figure 5.5 would be replaced by a logistic regression classifier. For each PSM, the fragment ion classifications given by the decision tree in Figure 5.4 and the output from envoMatch were used to engineer features for each site of citrullination comprised of the number of ions found in the MS2 in each category, and a boolean depicting whether the isotopic envelope in the MS1 was classified as valid. Both the synthetic peptide and complex proteomic ground truth datasets were split into two groups by random sampling. 70 % of each dataset was used to train the model and 30 % was used to test its performance. The receiver operating curve (ROC) plots had an area under the curve (AUC) of 0.861 and 0.901 for the human proteome and synthetic peptide datasets respectively (Figure 5.9). The output of the binary classifier is an odds ratio between 1 and 0 reflecting the odds that the given feature falls into one or the other category. Because ionFinder gives 1 of 4 ordinal, discrete values, a ROC curve cannot be generated. However, the static TPR and FPR can be used to access its performance. Interestingly, the ionFinder TPR and FPR are almost exactly the same as the binary classifier. Although the logistic regression classifier performed just as well as ionFinder, we chose to continue with the original ionFinder algorithm because ionFinder is designed to handle peptides with multiple sites of citrullination. The available training datasets have fewer multiply citrullinated peptides than we observe in our own data which hinders the training and assessment of a binary classifer in context of multiple sites of citrullination.

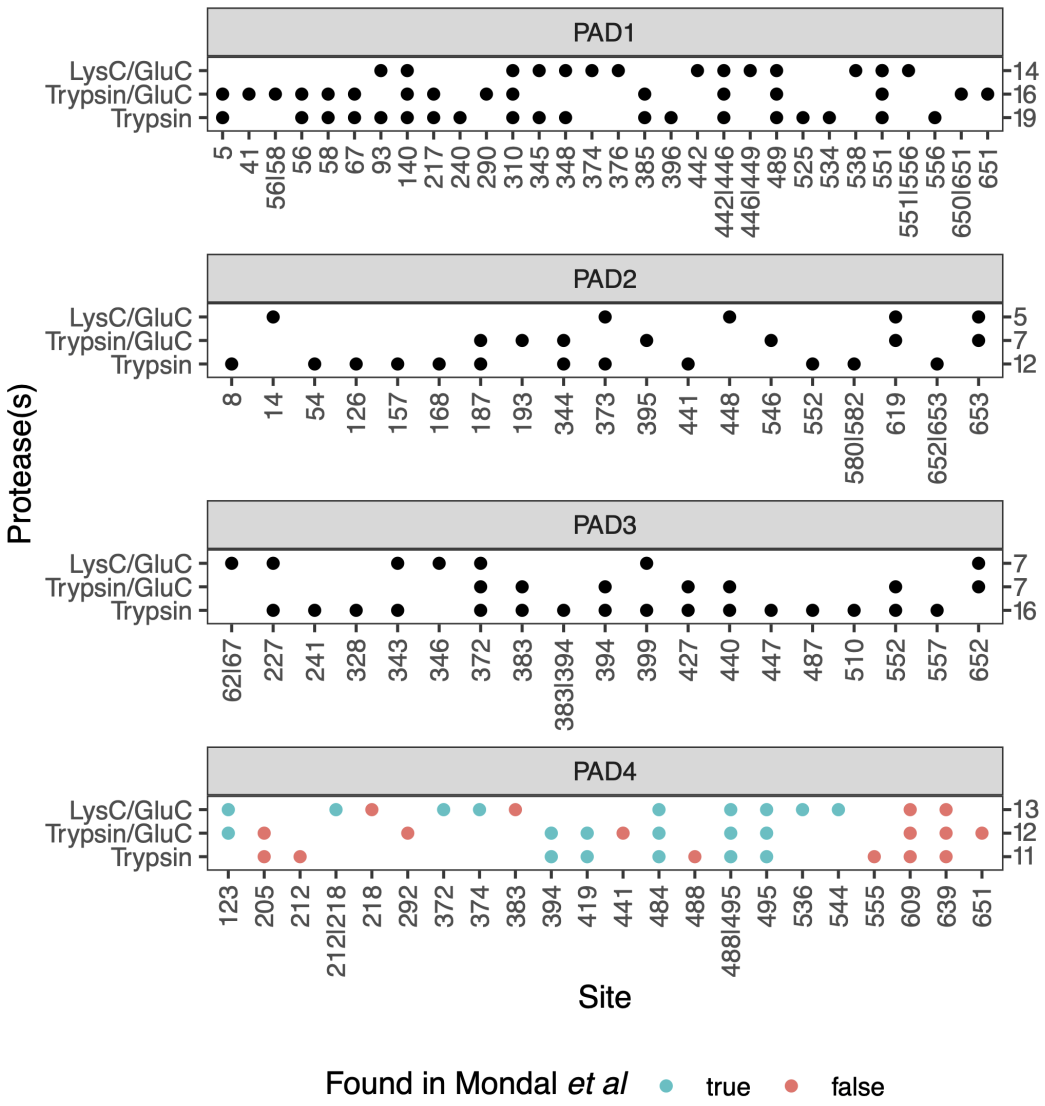


Figure 5.10: The residues that are identified as citrullinated for each individual PAD under three different digestion conditions; LysC/GluC, Trypsin/GluC, and Trypsin are shown. The PAD4 data were compared to a previous study. Citrullination sites identified in that Mondal *et al*[30] are shown in blue. Unique sites are shown in red.

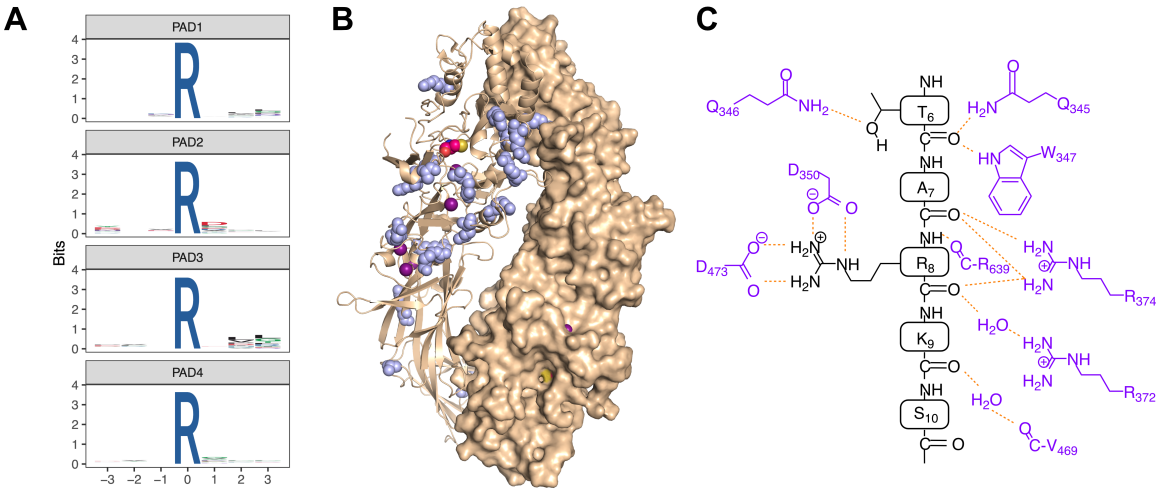


Figure 5.11: Specificity of autocitrullination. **(A)** Sequence logos depicting the substrate specificity determinants of autocitrullination by PAD1, PAD2, PAD3, and PAD4 (top to bottom). **(B)** Structure of the holo PAD2 dimer (PDBID: 4N2C) with citrullinated Arg residues shown in blue. **(C)** Cartoon showing how histone H3 binds in the active site of PAD4 (PDB ID: 2DEW). The interactions are primarily between the enzyme and the substrate backbone, indicating that specific side chains do not make a major contribution to substrate specificity.

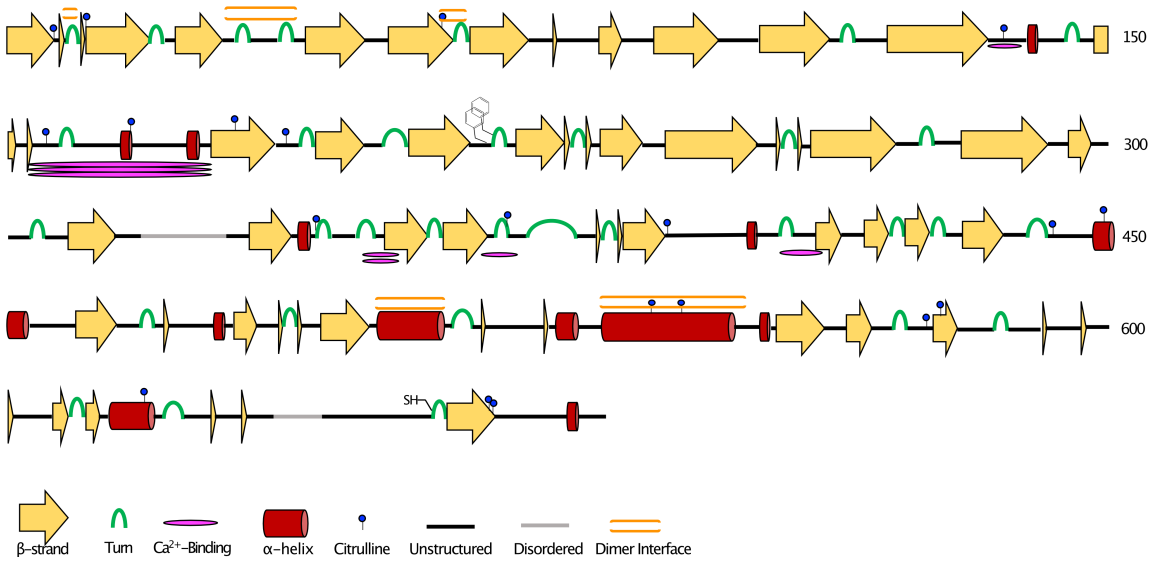


Figure 5.12: Secondary structural analysis of PAD2 autocitrullination shows that many of the sites of citrullination do not fall on canonical secondary structural regions. The cartoon highlights the sites of citrullination, different secondary structures, the FF-motif which plays a role in nuclear localization, and the location of the active site cysteine.

5.2.5 Sites of autocitrullination in PADs 1, 2, 3, and 4

To evaluate the utility of ionFinder in a "real world" situation, we examined the autocitrullination of PADs 1-4. Note that extensive research has established that the PADs autocitrullinate in biological systems, and once modified, protein-protein interactions are impacted.[30–32] Notably, recombinant, purified PADs readily undergo autocitrullination under reducing conditions in the presence of Ca^{2+}. PADs 1-4 were incubated with Ca^{2+} (5 mM) and TCEP (1 mM) at 37 °C for 1 h to induce autocitrullination. The proteins were processed for tandem mass spectrometry using different combinations of proteases (LysC/GluC, Trypsin/GluC, and Trypsin) to increase the depth of coverage and mitigate the lower efficiency of trypsin towards Cit-containing proteins. After processing the resulting data using envoMatch and ionFinder, we found that a large percentage of arginine residues from each PAD were autcitrullinated. Specifically, ionFinder unambiguously mapped 28 citrullinated residues on PAD1 (76 % of arginine residues), 20 on PAD2 (61 % of arginine residues), 19 on PAD3 (49 % of arginine residues), and 21 on PAD4 (78 % of arginine residues) (Figure 5.10). We compared the PAD4 autocitrullination sites obtained from ionFinder to a previously published dataset where sites of autocitrullination of PAD4 were evaluated in a time-dependent manner using tandem mass tags (TMT) labeling.[30] The sites identified by ionFinder displayed significant overlap with the previously determined sites of PAD4 autocitrullination.

Since these autocitrullination sites were obtained using purified, recombinant PADs, we cannot conclude that they all occur endogenously. However, we can use these data to search for patterns in PAD substrate preferences. First, we examined the citrullination data for each PAD to establish the existence of a consensus amino acid sequence flanking the site of citrullination (Figure 5.11A). Upon inspection, the results indicate that there are no obvious consensus sequences for PAD autocitrullination. For example, PADs 1, 2, and 4 show no discernable preference for a particular

residue, or type of residue, in positions that span a 3-residue window N-terminal and C-terminal to the site of citrullination. PAD3, however, does show a small preference for hydrophobic residues at the $R+2$ and $R+3$ positions.

Next, we mapped the sites of PAD2 autocitrullination onto high-resolution structures of PAD2 available in the protein databank.[33] These structural analyses revealed that 9 of the 20 citrullinated residues are not on strands or helices, and are instead located on looped or intermediate regions of the enzyme (Figure 5.12). The Cit residues that lie in regions of canonical secondary structure seem to be on the cap region, such as the C-termini of α-helices, and both termini of β-strands. Finally, we found that a quarter of the citrullinated residues were present at, or adjacent to, the dimer interface in PAD2 (Figure 5.12). These data indicate that PADs preferentially citrullinate Arg residues that are accessible and lie on flexible regions of proteins. Consistent with this notion is the fact that the PAD active site consists of a narrow ∼21 Å long channel that interacts with the substrate guanidinium and main chain carbonyls of a substrate. While this organization confers steric selectivity for peptidyl-arginine over free arginine, active site residues only contact water and the backbone of the substrate and not key side chains in the $R-3$ to $R+3$ positions, as revealed from structures of PAD4 bound to histone tail analogues (Figure 5.11).[34] As such, PADs preferentially citrullinate protein domains that are flexible enough to provide access to the enzyme, rendering looped regions, or caps of helices and strands, ideal locations for citrullination. Based on the MS data presented here, and the strong homology amongst PAD isozymes, we hypothesize that this explanation likely holds true for all PADs.

5.2.5.1 The HEK-PAD2 citrullinome

Having established the utility of envoMatch and ionFinder for identifying Cit-containing peptides from a simple purified protein system, we next sought to eval-

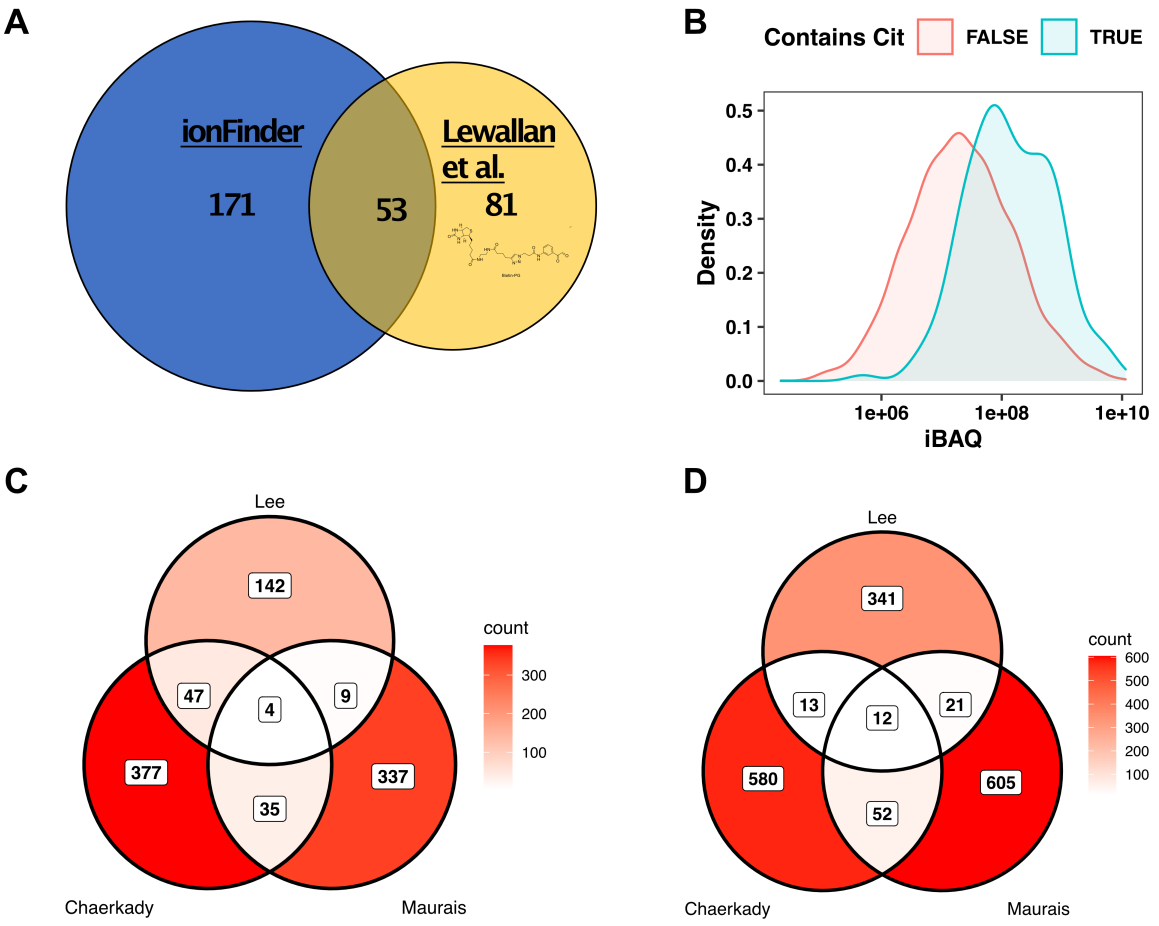


Figure 5.13: Global analysis of the HEK293T-PAD2 citrullinome. **(A)** Comparison between the citrullinome identified by ionFinder and Lewallen *et al.*[15] **(B)** Kernel density estimate plot depicting the probability distribution of protein families, TRUE and FALSE, abundance based on cumulative intensity-based absolute quantification (iBAQ) values from fractionated samples. The iBAQ at the maximum density in the plot was 7.45×10^7 for Cit-containing proteins and 1.94×10^7 for non-Cit-containing proteins. Thus, the maximum probability density of protein iBAQ values was 3.8 times greater for Cit-containing proteins than non-Cit containing proteins. **(C)** Number of overlapping citrulline containing proteins and **(D)** peptides in this work, Lee *et al.*,[16] and Chaerkady *et al.*[26]

uate the assignment of Cit residues within a complex proteome. For these studies, we used a previously described HEK293T-PAD2 cell line[15] that stably overexpresses PAD2. Addition of Ca^{2+} and ionomycin to the cell growth medium leads to robust citrullination of hundreds of proteins. After quenching with EDTA, cells were harvested, and proteomes were extracted and proteolytically digested for MS. Based on our prior work identifying citrullinated peptides in NETs,[25] we expected that a two-dimensional peptide fractionation step prior to MS analysis, would provide improved coverage of low-abundant Cit peptides. Therefore, samples were segmented offline into 12 fractions on a C18 column at high pH, followed by on-line reverse-phase separation at neutral pH. The resulting MS data were analyzed with envoMatch and ionFinder, resulting in the identification of ~350 unique Cit-containing PSMs on over 220 proteins (Table 5A.2).

Employing biotin-PG, we previously identified proteins that are citrullinated in the HEK-PAD2 cell line upon treatment with Ca^{2+} and ionomycin.[15] Comparing the data from the biotin-PG-enriched lysates (500 µg), and the unenriched lysate analyzed with envoMatch and ionFinder (50 µg), we found 53 common proteins amongst the two groups. Notably, ionFinder identified an additional 171 citrullinated proteins that were distinct from the biotin-PG dataset (Figure 5.13A). Differences between the two datasets are expected given the numerous variations in the analysis workflows. For example, the protein hits generated using biotin-PG are determined by the stoichiometry of the chemical reaction of the probe with citrullinated proteins. In addition, biotin-PG does not directly identify the sites of citrullination, instead protein identification is based on non-citrullinated tryptic peptides from citrullinated proteins that were modified by biotin-PG. Therefore, the abundance of MS-amenable tryptic peptides can further skew the protein identifications. By contrast, envoMatch and ionFinder directly identify the exact site of citrullination, without the need for biotin-PG labeling and enrichment. Here, peptide identifications are dependent on

the ionization and fragmentation characterizations of the Cit-containing peptide. Additionally, due to the lack of an enrichment step, Cit identifications are also biased by the relative abundance of the citrullinated protein. In fact, the proteins identified as being citrullinated by ionFinder were, on average, over the 70$^{\text{th}}$ percentile in protein abundance (as assessed with cumulative intensity-based absolute quantification (iBAQ) values) (Figure 5.13B). We also compared the Cit identifications from the Lee et al[16] and the Chaerkady *et al*[26] studies to our HEK-PAD2 data. These three studies were similar in that there was no specific enrichment of Cit peptides, but differed in the data analysis workflows used for Cit identification. There was little overlap across the three studies (Figure 5.13C,D), likely due to the significant differences in the proteomes that were analyzed in these three studies (various human tissues, neutrophils, and HEK cells, respectively).

As in the PAD autocitrullination experiments, there was no clear consensus sequence for citrullination in HEK cells (Figure 5.14A). However, many of the proteins we identified as being citrullinated with envoMatch and ionFinder were nuclear proteins, with an overrepresentation of DNA and RNA-binding proteins (Figure 5.14B). In fact, 58 % of the citrullinated proteins are annotated as localized to the nucleus based on information provided by Uniprot. In addition, 21 % of citrullinated proteins are classified as nucleotide binding proteins (PANTHER: PC00171). This is unsurprising when one considers that Arg-rich proteins are often responsible for binding to the net negatively charged backbone of DNA and RNA.[37] One interesting example is the splicing factor U1 70K snRNP, which is citrullinated at Arg222. Arg222 is in a S/R/E rich region adjacent to the N-terminal RNA-binding motif, and many studies have shown that the surrounding Ser residues can be phosphorylated to regulate alternative splicing, cell death, and autoantigenicity.[37–40] Another intriguing example is eEF1, an EF-Tu homologue, which acts as an elongation factor that facilitates aminoacyl-tRNA binding into the A-site of ribosomes during translation. The region

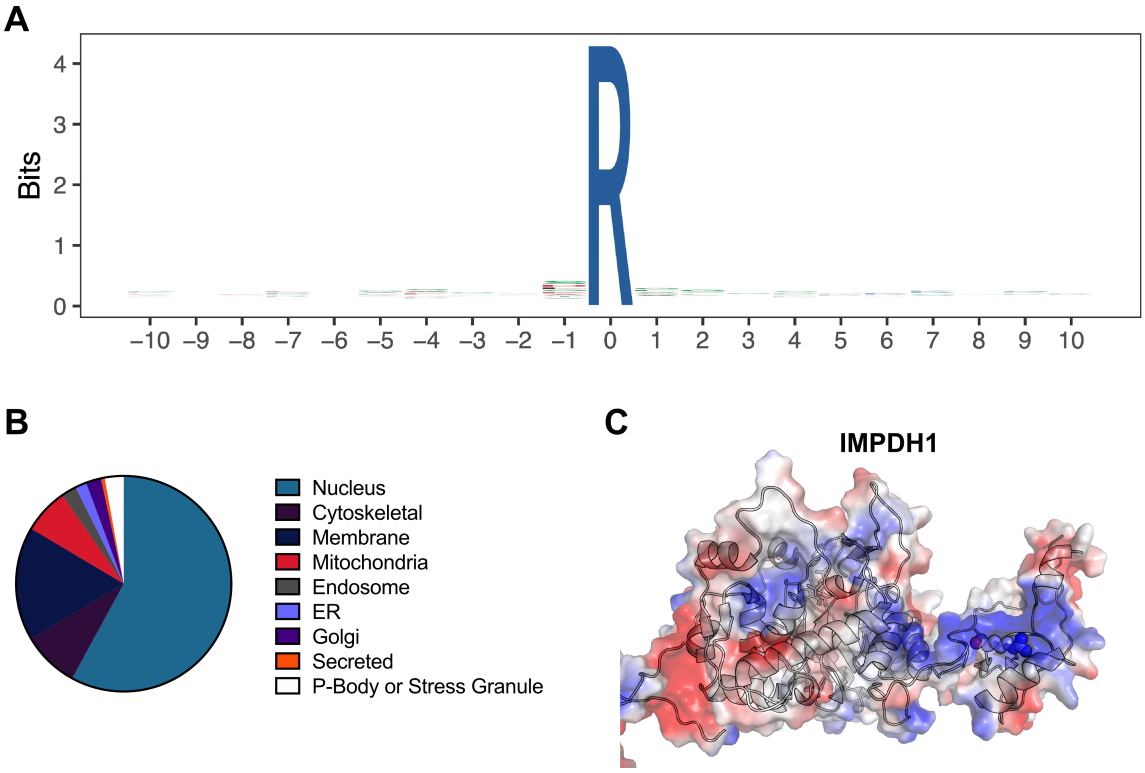


Figure 5.14: **(A)** Sequence logo depicting the substrate specificity determinants of citrullination of PAD2 in HEK293T cells. **(B)** Cellular localization of the citrullinated proteins identified by ionFinder, demonstrating that many of the proteins are, as expected, nuclear. **(C)** Crystallographic depiction of IMPDH2 electrostatic distribution (PDB 1NF7) showing the location of Arg224 in the apo state of the enzyme.[35,36] Arg224 is in an exposed, and positively charged region, making it a target for citrullination by PAD2. When GTP is bound, Arg224 is no longer as accessible, as it lies in the nucleotide binding pocket near the dimer interface.

that is citrullinated (Arg96, 427, and 430) is part of a positively charged cluster of amino acids involved in binding tRNA. Lastly, the enzyme inosine-5′-monophosphate dehydratase 2 (IMPDH2) is citrullinated at Arg224. This region of IMPDH2 is involved in nucleotide binding, where Arg224 contributes to GTP binding in the holo conformation.[41] In the apo form, Arg224 is on a positively charged, readily accessible loop (Figure 5.14C) where this residue is far more exposed and available to act as a substrate for a PAD. Fascinatingly, mutations of this arginine residue in IMPDH1 are associated with retinopathy.[42]

5.2.6 Conclusions

Aberrant citrullination is linked to many human diseases including RA.[5] To gain a better understanding of the role of citrullination in disease pathogenesis, we used citrulline selective probes to identify proteins with elevated citrullination in RA serum, synovial fluid, and synovial tissue. In addition to identifying known PAD substrates such as vimentin, fibrin, and enolase, we also identified more than 150 novel citrullinated proteins. Among the most notable classes of novel proteins were SERPINs which we demonstrated to show lower activity when citrullinated.

Citrullination remains an understudied area of research. While many advances have been made over the last several decades, site-specific decoding of the citrullinome has remained challenging. Identifying sites of citrullination by mass spectrometry is plagued by false-positive annotations due to the small mass shift that results from the modification, as well as confounding isobaric Asn and Gln deamidations. Accurate assignment of Cit residues within peptides requires extensive mining of fragmentation spectra for the presence of neutral loss species that are unique to Cit. Previous studies used manual spectral matching to confirm sites of citrullination. Here, we automated this process through the development of a suite of computational tools, envoMatch and ionFinder, which can mine fragmentation data to unambiguously map the sites of

citrullination. These tools are intended to be incorporated downstream of database-search algorithms, such as SEQUEST, Mascot, or MaxQuant, to further parse for high-confident assignments of Cit-containing peptides. In the recent Chaerkady *et al* study[26] a similar approach was taken, where MaxQuant data was parsed for hits that specifically displayed Cit-specific NL species. Our described workflow incorporates more stringent filtering than that available through MaxQuant by setting a dynamic NL cutoff to differentiate determining from artifactual NL species. Additionally, MaxQuant does not implement the isotopic envelope matching performed by envoMatch, which serves to further decrease false positives from the output and increase confidence in the identified hits. EnvoMatch and ionFinder were used to comprehensively map of the sites of PAD autocitrullination. Our data indicate that PADs prefer substrate Arg residues on flexible, surface-exposed regions of the protein, supporting the hypothesis that PADs are opportunistic in their substrate selection. Furthermore, many of these resulting Cit residues were in the dimer interface of the protein, suggesting that autocitrullination may therefore affect PAD dimerization. With these citrullination maps at our disposal, our future work is focused on discovering the molecular consequences of autocitrullination on enzyme structure, activity, and subcellular localization.

In addition to sites of autocitrullination on purified PAD proteins, envoMatch and ionFinder also identified a long list of cellular targets for PAD2. Upon activating cells with calcium and ionomycin, PAD2 re-localizes to the nucleus.[43] Our data confirms the nuclear localization of PAD2 because the majority of the identified sites of citrullination were found on nuclear proteins. Furthermore, many of the targets, nuclear or otherwise, are involved in nucleotide binding. Proteins which bind to nucleotides are often arginine-rich and have higher pI values to promote a favorable interface with the negatively charged sugar-phosphate backbone. These qualities make this family of proteins prime targets for the PADs. Future work entails mechanistic investigation of

these targets to understand the functional consequences of citrullination in vitro and in cells using our recently described approach to site-specifically incorporate citrulline into proteins.[30]

In summary, ionFinder and envoMatch, automate the process of verifying the assignment of Cit annotations by standard database-search algorithms, primarily through the automated assignment of diagnostic neutral loss species. The primary strength of ionFinder and envoMatch is the speed of analysis, which far exceeds the throughput amenable by manual spectral matching. Additionally, the programs are flexible in the types of instruments, fragmentations modes, and database-search algorithms that are used to generate the input data. The primary limitation in the described ionFinder workflow is the limited coverage of citrullination events on low abundance proteins. Since the input data for ionFinder is generated from cell lysates that have not specifically been enriched for Cit-containing proteins or peptides, there is an expected bias toward identification of sites of citrullination from abundant proteins within the proteome. Further advances in peptide fractionation methodology, and the advent of instruments with faster scan times, will likely help to overcome this abundance bias, and allow for a deeper interrogation of citrullination events from complex biological samples. Lastly, although specifically designed for analysis of sites of citrullination, ionFinder and envoMatch are easily adapted to other PTMs that show diagnostic NL species. We believe that the customizable nature of ionFinder and envoMatch will provide a useful tool for proteomic analyses in general, by setting more stringent criteria for assignment of modified peptides.

5.3 Materials and methods

5.3.1 Methods for biotin PG labeled samples

5.3.1.1 Biotin-PG labeling and proteomic analysis

Samples were labelled with biotin-PG similarly to previously described methods.[14,15] Briefly, samples (300 µg) were diluted in buffer (100 mM HEPES pH 7.6 to a final concentration of 1 µg µL^{-1} in a reaction volume of 300 µL) and incubated with 20 % trichloroacetic acid (60 µL of 100 % TCA) and 0.5 mM biotin-PG (5 µL of 5 mM stock) for 30 min at 37 °C. After a 30 min incubation, the reaction was quenched with 60 µL of 0.1 M citrulline dissolved in 50 mM HEPES pH 7.6. Proteins were precipitated by placing the reaction mixtures on ice for 30 min followed by centrifugation (13,500 rpm, 15 min) at 4 °C. The supernatants were discarded and the protein pellet was washed twice with cold acetone and dried.

Protein pellets were then resuspended in 1.2 % SDS in PBS, boiled for 10 min and sonicated for 12-15 min. Once dissolved, samples were diluted in PBS buffer (0.2 % final SDS concentration) and incubated with streptavidin agarose slurry (170 µL) overnight at 4 °C, and a further 3 h at 25 °C. Streptavidin beads were then washed with PBS containing 0.2 % SDS, with PBS alone (3 × 5 mL), and water (3 × 5 mL) in order to remove any unbound proteins. Streptavidin beads were heated in a buffer containing 500 µL of 6 M urea in PBS and 10 mM DTT (65 °C, 20 min) followed by iodoacetamide (20 mM) addition and further incubation for 30 min at 37 °C. The beads were pelleted by centrifugation ($1,400 \times g$ for 3 min), resuspended in a premixed solution of 2 M urea 1 mM $CaCl_2$ and 2 µg trypsin in PBS and incubated overnight at 37 °C. The beads were then pelleted by centrifugation and the supernatant, containing the peptide digests was collected. The pellet was washed twice with water (50 µL) and the collected supernatant fractions were combined. Formic acid (15 µL) was added to the peptide mixture and samples were stored at −20 °C until MS analysis

5.3.1.2 LC-MS/MS and data processing

Mass Spectrometry was performed on a LTQ Orbitrap Discovery (Thermo Fisher) coupled to an Agilent 1200 series HPLC. Protein digests were prepared as described above and pressure loaded onto a 250 µm fused silica desalting column, packed with 4 cm Aqua C18 reverse phase resin (Phenomenex). Peptides were eluted onto a 100 µm fused silica biphasic column, packed with 4 cm strong cation exchange (Watman) and 10 cm C18 resins using a five-step multidimensional LC/LC-MS/MS protocol (MudPIT).[44] Each of the five runs begins with a salt push (0 %, 25 %, 50 %, 80 %, and 100 %), followed by a gradient 0–100 % buffer B in buffer A (buffer A: 95 % water, 5 % acetonitrile, 0.1 % formic acid; buffer B: 20 % water, 80 % acetonitrile, 0.1 % formic acid) with 0.25 µL min^{-1} flow rate and 2.75 kV spray voltage. One full MS1 scan (400-1800 m/z) was followed by 8 data dependent scans of the nth most intense ion. Peptides were assigned to MS2s using the SEQUEST[45] algorithm, searching against a human reverse-concatenated non-redundant protein database. A static modification was specified for cysteine residues (+57.0215 m/z, iodoacetamide alkylation). MS2 matches were assembled by protein and filtered using DTASelect 2.0.[46] Spectral counts for proteins identified in each sample were combined into a single file using custom in house software.

5.3.1.3 Statistical overrepresentation test of PANTHER protein class and GO biological process

PANTHER gene list analysis was used to determine which protein classes/biological processes were significantly overrepresented in the lists of proteins identified in SF and tissue samples obtained from Healthy and RA patients.[27] The PANTHER software determines the number of proteins which would be expected to be found for a specific category in the input protein list, given a reference list (complete human genome). The expected number of proteins in each category is then compared to the actual

number of protein in that category in the input dataset. Proteins with an average of greater than 10 spectral counts, and a P-value for the RA/healthy spectral count ratio less than 0.05 were included. A binomial test is used to determine the statistical significance the overrepresentation value calculated for each category, using a P-value cutoff of 0.05. The test does not make any assumptions about how the datasets were generated, and only assumes that the input and reference datasets were taken from the same protein population.

5.3.2 Methods for ionFinder experiments

5.3.2.1 Automated assignment of "true" citrullinated residues using ionFinder

ionFinder was written in C++11. Results were exported from Scaffold in the form of a spectrum report and converted into a format suitable for ionFinder using a custom python script. After reading the required input files, ionFinder analyzes peptide-spectrum matches (PSM) in 3 major steps. In the first step, PSM assigned by PD are re-searched by ionFinder for a customized list of theoretical fragment ions. For each peptide, a list of masses for theoretical b and y ions are calculated. In addition, b and y -43.0058 Da NL ion masses are calculated for citrullination up to the multiplicity of citrulline residues on the peptide. In this work, the MS2 spectra were charge deconvoluted prior to searching with PD, so that only singly charged fragment ions were considered. Each MS2 spectrum was searched for theoretical fragment ions within a 10 ppm m/z tolerance. If multiple ions are found in the specified range, ties are broken by intensity. In cases where multiple fragments have the same predicted mass, all possible fragments are considered found if the ion is found in the MS2 spectrum. In the second analysis step, fragment ions which were identified in the spectrum are classified into 1 of 5 categories with respect to how they provide evidence supporting or contradicting the citrullination of a given peptide. These

assignments were based on the decision tree shown in Figure 5.4. Unmodified b or y ions which do not contain any citrullines are classified as ambiguous (Amb); fragment ions which are citrullinated, but also contain N or Q will also be classified as Amb; b or y ions which contain Cit are classified as Cit determining (Det). NL ions which unambiguously belong to a Cit containing fragment are classified as Cit determining NL (DetNL); NL ions on multiply modified peptides which cannot unambiguously be assigned to 1 or both modifications are classified as ambiguous (Amb); if a NL is observed for an ion which does not contain Cit it is classified as artifact NL (ArtNL). ArtNL ions in effect function as decoy fragment ions because they would not be expected to be observed in the fragmentation spectrum in significant abundance. Once each fragment ion has been classified, a dynamic ion intensity cutoff is set such that the percentage of ArtNL ion intensities in the spectrum is less than 1 %.
The sum of intensities for ion type t, above cutoff k is defined as:

$$A_t = \sum_{i=1}^{n} \begin{cases} \text{if } a_{i,t} \geq k, & a_{i,t} \\ \text{else,} & 0 \end{cases} \tag{5.1}$$

where $a_{t,i}$ is the intensity for an ion i of type t (out of n ions of type t).
The fraction of intensity from ArtNL ions out of all ion types is defined as:

$$P_{ArtNL} = \frac{A_{ArtNL}}{A_{ArtNL} + A_{AmbNL} + A_{DetNL} + A_{Det} + A_{Amb}} \tag{5.2}$$

k is set such that $P_{ArtNL} \leq x$ where x is the user specified allowable percentage of intensity from ArtNL ions.

In the third analysis step, the number of ions observed in each of the 5 classes are used to assign each site of citrullination as “true”, “likely”, “ambiguous”, or “false” according to the decision tree in (Figure 5.5). A classification of true, likely, ambigu-

ous, or false is assigned to each site on multiply-citrullinated peptides individually, then the value assigned to the peptide is the lowest classification of all sites; i.e. a doubly citrullinated peptide will only be classified as "true", if both sites are classified as "true".

5.3.2.2 Automated verification of precursor isotopic envelopes with envoMatch

envoMatch is written in Python3 and utilizes routines from the Pyteomics[47] package for calculation of isotopic envelopes and mzXML and mzML file parsing. envoMatch compares the observed isotopic envelope for the precursor peptide with the theoretical envelope of the PSM peptide. For a peptide with n citrulline residues, theoretical envelopes of a peptide with $\{n, n-1, \ldots, 0\}$ citrulline residues are calculated. An envelope similarity score comparing the observed envelope to each theoretical isotopic envelope is generated using a Pearson correlation. A peptide is considered to have a "true" site of citrullination if the best match was for the envelope with n cit residues and if the similarity score is greater than 0.8.

5.3.2.3 Software availability

ionFinder and envoMatch are open-source software and are distributed under MIT licenses. Source code and installation instructions can be found at their respective GitHub repositories. ionFinder (`https://github.com/weerapana-lab/ionFinder`), and envoMatch (`https://github.com/weerapana-lab/envoMatch`).

5.3.2.4 Expression and purification of PAD1,2, & 3

PAD1, PAD2, and PAD3 were expressed and purified similarly to previously described methods.[33,48] Briefly, E. coli BL21 (DE3) were transformed with a plasmid encoding a PAD protein with an N-terminal 10X-His tag. Single colonies were used

to inoculate a 5 mL starter culture that was grown overnight in LB (0.1 mg mL^{-1} ampicillin, 0.025 mg mL^{-1} chloramphenicol) agitating at 37 °C. The starter culture (1 mL) was used to inoculate 1 L of pre-warmed LB media (0.1 mg mL^{-1} ampicillin, 0.025 mg mL^{-1}). Expression cultures were grown to an OD600 of 0.6-0.8. Flasks were cooled on ice for 20 min before expression was induced by the addition of 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were agitated overnight at 16 °C and cell pellets were harvested via centrifugation and frozen in liquid nitrogen. Frozen cells pellets were stored at -80 °C.

To lyse the cells, pellets were thawed in a water bath at room temperature. Per each g of frozen cell pellet, 1 mL of lysis buffer (20 mM Tris pH 7.6, 400 mM NaCl, 5 mM $MgCl_2$, 5 mM imidazole, 0.5 mM TCEP, 1 % Triton X-100) was added in addition to PierceTM EDTA-free protease inhibitor tablets. The slurry was agitated for 30 min after which the mixture was spiked with universal nuclease (Promega). The mixture was briefly sonicated before clarifying the lysate of insoluble debris. Soluble lysate was applied to Ni-NTA resin via gravity. The resin was washed thoroughly with a buffer consisting of 20 mM Tris pH 7.6, 400 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, and 10 % (w/v) glycerol. PAD protein was eluted with wash buffer containing 250 mM imidazole. Purity was assessed via SDS-PAGE and Coomassie blue staining. The eluted protein was dialyzed overnight into a buffer of 20 mM Tris pH 7.6, 500 mM NaCl, 0.5 mM TCEP, and 10% (w/v) glycerol. Protein concentration was measured using the Bradford Assay.

5.3.2.5 Recombinant PAD4 expression and purification

PAD4 was expressed and purified as described previously[49] with the following alterations: after the GST tag was removed with Precision Protease, PAD4 protein was dialyzed overnight against a buffer consisting of 20 mM Tris pH 7.6, 100 mM NaCl, 1 mM EDTA, 10% (w/v) glycerol, and 2 mM DTT. This low-salt buffer caused full

length PAD4 to precipitate, whereas the GST tag and GST-tagged PAD4 remained in solution. The precipitate was collected via centrifugation in a 50 mL Falcon tube and washed 2 X with 20 mL of the low-salt buffer. Finally, pure PAD4 was resolubilized gently in 3 mL of buffer consisting of 20 mM Tris pH 8.1, 500 mM NaCl, 1 mM EDTA, 10% (w/v) glycerol, and 2 mM DTT. Purity was assessed via SDS-PAGE and Coomassie blue staining. Protein concentration was measured using the Bradford Assay.

5.3.2.6 Preparation of purified proteins for tandem MS analysis

Purified PADs (30 µg) were autocitrullinated for 1 h at 37 °C in 100 µL of buffer consisting of 100 mM HEPES pH 7.6, 100 mM NaCl, 1 mM TCEP, and 5 mM $CaCl_2$. Autocitrullination was stopped by the addition of trichloroacetic acid to a final concentration of 20% (w/v). Reactions were placed on ice for 1 h to promote protein precipitation. Sample tubes were centrifuged in a 4 °C tabletop centrifuge at top speed for 15 min. The precipitant was washed with 300 µL cold acetone, which was then removed via pipette after another 10 min centrifugation at top speed. The pellets were air dried and then resuspended in a solution of 8 M urea in PBS (30 µL). Once solubilized, 100 mM ammonium bicarbonate (70 µL) was added to dilute the urea concentration to 2.4 M. The samples were then reduced by the addition of 1 M DTT (1.5 µL), and further denatured by placing them at 65 °C for 15 min. Reduced cysteines were then alkylated with iodoacetamide (12.5 mM final concentration) at 21 °C in the dark for 30 min. The urea concentration was then further diluted by the addition of PBS (120 µL). Both GluC and LysC (Promega) were used in combination (1:30 ratio, enzyme:substrate). Trypsin was used both in combination with GluC and alone (1:50, enzyme:substrate) in the presence of 1 mM $CaCl_2$. Digests were agitated by rotating at 37 °C overnight. Protein digestion was stopped by the addition of formic acid to 5 % (v/v). Peptides were then desalted using Pierce^{TM}

C18 spin columns and dried to a powder. After resuspension in water, the peptide concentration was assessed using a PierceTM quantitative fluorometric peptide assay.

5.3.2.7 HEK-PAD2 cell culture and lysate prep

HEK293T cells that stably express full length PAD2 (HEK-PAD2 cells)[15] were propagated according to previously described methods.[15] Briefly, cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin / streptomycin. The cultures were maintained in a humidified atmosphere with 5 % CO2 at 37 °C. Cellular citrullination was induced according to previously described methods.13 Briefly, cells were grown to passage 5 on T-175 plates until ∼90 % confluency. They were then treated with Ca^{2+} (2 mM final concentration) and ionomycin (5 µM final concentration) for 1 h. After incubation, cells were scraped, washed with cold PBS, and then snap-frozen. Pellets were stored at -80 °C until lysis. To lyse the cells, pellets were thawed on ice in 5 mL of buffer consisting of PBS, 5 mM EDTA, 1 % Triton X-100, and Pierce protease inhibitor. Lysis was achieved in an ice bath via iterative rounds of sonication using a Sonic Dismembrator (Fisher Scientific) fitted with a microtip (amplitude 10, 10x1 sec bursts with 30 sec of rest between cycles, 4 cycles total). Lysates were cleared by centrifugation, and the protein concentration assessed via the DC assay.

5.3.2.8 Preparation of HEK-PAD2 lysate for tandem MS analysis

Cell lysate (50 µg) was precipitated on ice for 30 min by the addition of trichloroacetic acid to 20 % (w/v). Sample tubes were centrifuged in a 4 °C tabletop centrifuge at top speed for 15 min. The precipitate was washed with 300 µL cold acetone, which was then removed via pipette after another 10 min centrifugation at top speed. The pellets were air dried and then resuspended in a solution of 8 M urea in PBS (30 µL). Once solubilized, 100 mM ammonium bicarbonate (70 µL) was added to dilute

the urea concentration to 2.4 M. The samples were then reduced by the addition of 1 M DTT (1.5 µL), and further denatured by placing them at 65 °C for 15 min. Reduced cysteines were then alkylated with iodoacetamide (12.5 mM final concentration) at 21 °C in the dark for 30 min. The urea concentration was then further reduced by the addition of PBS (120 µL). Trypsin was used at a concentration of 1:50 (enzyme:substrate) in the presence of 1 mM $CaCl_2$, and samples were proteolyzed overnight agitating by rotation at 37 °C. After digestion, tryptic peptides were separated on a ZORBAX extended C18 column (Agilent) over a 1 h, biphasic gradient from 0 % Buffer A (10 mM ammonium bicarbonate) to 100 % Buffer B (10 mM ammonium bicarbonate, 90 % acetonitrile). Fractions (0.5 mL) were collected in a 96 well plate and pooled by column (as opposed to by row) to yield 12 samples. The 12 samples of pooled fractions were dried in a vacuum concentrator and resuspended in 5 % acetonitrile/0.1 % trifluoroacetic acid.

5.3.2.9 Tandem MS

Peptide mixtures were separated on a NanoAcquity UPLC (Waters Corporation, Milford, MA) using an in-house packed pre-column (C18, 200 A, 5 µm, 2 cm) and an in-house packed analytical column (C18, 100 A, 5 µm, 2 cm) using the aqueous mobile phase of water + 0.1 % formic acid (A) and an organic mobile phase of acetonitrile + 0.1 % formic acid (B). Peptide trapping was operated at 4 min at 4 µL min^{-1} µL/min and 5 % B. Then the peptides were transferred to the analytical column at the flow rate of 300 nL/min using the gradient of 5 % to 35 % B over 60 minutes, then 35 % to 60 % B for 30 min, followed by 15 min of high organic wash of 90 % B and 18 min of 5 % B for re-equilibration. Ions were then introduced to a Q Exactive hybrid quadrupole-Orbitrap (Thermo Fisher Scientific Inc., Waltham, MA) mass spectrometer, performing at positive electrospray ionization (ESI^+) with the ionization voltage set at 1.4 kV. The full MS (MS1) data scan was acquired in a *m/z*

scan range of 300-1750 Da, using an AGC target of 1×10^6, the maximum injection time of 30 ms and a resolution of 70000 at m/z 200. The MS/MS (MS2) data were acquired in data-dependent acquisition mode, performing MS/MS on the top 10 most abundant precursor ions using an AGC target of 1×10^5, a maximum injection time of 110 ms, an isolation width of 1.6 Da, a resolution of 17500 at m/z 200, and a collision energy of 27 volts using HCD.

5.3.2.10 MS data analysis

The LC-MS/MS raw data were processed using Thermo Proteome Discoverer (PD) 2.1.1.21 (Thermo Fisher Scientific Inc.). The data obtained for the autocitrullinated PADs was searched against an E. coli SwissProt database FASTA file that included the sequences corresponding to the PAD proteins, and the HEK-PAD2 samples were searched against the Human SwissProt database. Both searches were performed using Mascot Server 2.6.2 (Matrix Science Ltd). The search parameters included searching specific protease cleavage sites with 2 maximum missed cleavages. Carbamidomethyl cysteine modifications were set as a fixed modification, while variable modifications included: peptide N-terminal acetylation, methionine oxidation, N-terminal glutamine to pyroglutamate, and citrullination of arginine. In addition to these modifications, the deamidation of Asn and Gln was set as variable modifications. A 10 ppm m/z cutoff was employed for the precursor mass and 0.05 Da for the fragment ion mass tolerance. Protein identification and validation was done using Scaffold 4.10.0 (Proteome Software Inc.), employing 1 % FDR threshold for peptides, and a 99 % probability threshold for protein identification, using Peptide Prophet and Protein Prophet algorithms.29, 30 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE31 partner repository with the dataset identifier PXD027358.

References

[1] Tilvawala, R.; Nguyen, S. H.; Maurais, A. J.; Nemmara, V. V.; Nagar, M.; Salinger, A. J.; Nagpal, S.; Weerapana, E.; Thompson, P. R. The Rheumatoid Arthritis-Associated Citrullinome. *Cell Chem Biol* **2018**, *25*, 691–704.

[2] Maurais, A. J.; Salinger, A. J.; Tobin, M.; Shaffer, S. A.; Weerapana, E.; Thompson, P. R. A Streamlined Data Analysis Pipeline for the Identification of Sites of Citrullination. *Biochemistry* **2021**, *60*, 2902–2914.

[3] Fuhrmann, J.; Thompson, P. R. Protein Arginine Methylation and Citrullination in Epigenetic Regulation. *ACS Chem Biol* **2016**, *11*, 654–68.

[4] Tilvawala, R.; Thompson, P. R. Peptidyl arginine deiminases: detection and functional analysis of protein citrullination. *Curr Opin Struct Biol* **2019**, *59*, 205–215.

[5] Carmona-Rivera, C. et al. Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis. *Sci Immunol* **2017**, *2*, eaaz9319.

[6] Schellekens, G. A.; de Jong, B. A.; van den Hoogen, F. H.; van de Putte, L. B.; van Venrooij, W. J. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *J Clin Invest* **1998**, *101*, 273–281.

[7] Van Boekel, M. A.; Vossenaar, E. R.; van den Hoogen, F. H.; van Venrooij, W. J. Autoantibody systems in rheumatoid arthritis: specificity, sensitivity and diagnostic value. *Arthritis Res* **2002**, *4*, 87–93.

[8] Jones, J. E.; Causey, C. P.; Knuckley, B.; Slack-Noyes, J. L.; Thompson, P. R. Protein arginine deiminase 4 (PAD4): Current understanding and future therapeutic potential. *Curr Opin Drug Discov Devel* **2009**, *12*, 616–627.

[9] Chumanevich, A. A.; Causey, C. P.; Knuckley, B. A.; Jones, J. E.; Poudyal, D.; Chumanevich, A. P.; Davis, T.; Matesic, L. E.; Thompson, P. R.; Hofseth, L. J. Suppression of Colitis in Mice by Cl-Amidine: A Novel Peptidylarginine Deiminase (Pad) Inhibitor. *Am J Physiol Gastrointest Liver Physiol* **2011**, *300*, G929–G938.

[10] Knight, J. S.; Subramanian, V.; O'Dell, A. A.; Yalavarthi, S.; Zhao, W.; Smith, C. K.; Hodgin, J. B.; Thompson, P. R.; Kaplan, M. J. Peptidylarginine deiminase inhibition disrupts NET formation and protects against kidney, skin and vascular disease in lupus-prone MRL/lpr mice. *Ann Rheum Dis* **2015**, *74*, 2199–206.

[11] Liu, Y. et al. Peptidylarginine deiminases 2 and 4 modulate innate and adaptive immune responses in TLR-7-dependent lupus. *JCI Insight* **2018**, *3*, e124729.

[12] Tian, Y. et al. Peptidylarginine deiminase 2 has potential as both a biomarker and therapeutic target of sepsis. *JCI Insight* **2020**, *5*, e138873.

[13] Li, F. J. et al. Citrullinated vimentin mediates development and progression of lung fibrosis. *Sci Transl Med* **2021**, *13*, eaba2927.

[14] Clancy, K. W.; Weerapana, E.; Thompson, P. R. Detection and identification of protein citrullination in complex biological systems. *Curr Opin Chem Biol* **2015**, *30*, 1–6.

[15] Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. Chemical Proteomic Platform To Identify Citrullinated Proteins. *ACS Chem Biol* **2015**, *10*, 2520–8.

[16] Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. Mining the Human Tissue Proteome for Protein Citrullination. *Mol Cell Proteomics* **2018**, *17*, 1378–1391.

[17] Holm, A.; Rise, F.; Sessler, N.; Sollid, L. M.; Undheim, K.; Fleckenstein, B. Specific modification of peptide-bound citrulline residues. *Anal Biochem* **2006**, *352*, 68–76.

[18] Tutturen, A. E.; Holm, A.; Jorgensen, M.; Stadtmuller, P.; Rise, F.; Fleckenstein, B. A technique for the specific enrichment of citrulline-containing peptides. *Anal Biochem* **2010**, *403*, 43–51.

[19] Bicker, K. L.; Subramanian, V.; Chumanevich, A. A.; Hofseth, L. J.; Thompson, P. R. Seeing citrulline: development of a phenylglyoxal-based probe to visualize protein citrullination. *J Am Chem Soc* **2012**, *134*, 17015–8.

[20] Nepomuceno, A. I.; Gibson, R. J.; Randall, S. M.; Muddiman, D. C. Accurate identification of deamidated peptides in global proteomics using a quadrupole orbitrap mass spectrometer. *J Proteome Res* **2014**, *13*, 777–85.

[21] Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. Accurate Identification of Deamidation and Citrullination from Global Shotgun Proteomics Data Using a Dual-Search Delta Score Strategy. *J Proteome Res* **2020**, *19*, 1863–1872.

[22] Hao, G.; Wang, D.; Gu, J.; Shen, Q.; Gross, S. S.; Wang, Y. Neutral loss of isocyanic acid in peptide CID spectra: a novel diagnostic marker for mass spectrometric identification of protein citrullination. *J Am Soc Mass Spectrom* **2009**, *20*, 723–7.

[23] Jin, Z.; Fu, Z.; Yang, J.; Troncosco, J.; Everett, A. D.; Van Eyk, J. E. Identification and characterization of citrulline-modified brain proteins by combining HCD and CID fragmentation. *Proteomics* **2013**, *13*, 2682–91.

[24] Huh, S.; Hwang, D.; Kim, M. S. Statistical Modeling for Enhancing the Discovery Power of Citrullination from Tandem Mass Spectrometry Data. *Anal Chem* **2020**, *92*, 12975–12986.

[25] Salinger, A. J.; Dubuke, M. L.; Carmona-Rivera, C.; Maurais, A. J.; Shaffer, S. A.; Weerapana, E.; Thompson, P. R.; Kaplan, M. J. Technical comment on "Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis". *Sci Immunol* **2020**, *5*, eaax5672.

[26] Chaerkady, R.; Zhou, Y.; Delmar, J. A.; Weng, S. H. S.; Wang, J.; Awasthi, S.; Sims, D.; Bowen, M. A.; Yu, W.; Cazares, L. H.; Sims, G. P.; Hess, S. Characterization of Citrullination Sites in Neutrophils and Mast Cells Activated by Ionomycin via Integration of Mass Spectrometry and Machine Learning. *J Proteome Res* **2021**, *20*, 3150–3164.

[27] Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P. D. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* **2017**, *45*, D183–D189.

[28] Huntington, J. A.; Read, R. J.; Carrell, R. W. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **2000**, *407*, 923–6.

[29] Khan, M. S.; Singh, P.; Azhar, A.; Naseem, A.; Rashid, Q.; Kabir, M. A.; Jairajpuri, M. A. Serpin Inhibition Mechanism: A Delicate Balance between Native Metastable State and Polymerization. *J Amino Acids* **2011**, *2011*, 606797.

[30] Mondal, S.; Wang, S.; Zheng, Y.; Sen, S.; Chatterjee, A.; Thompson, P. R. Site-specific incorporation of citrulline into proteins in mammalian cells. *Nat Commun* **2021**, *12*, 45.

[31] Andrade, F.; Darrah, E.; Gucek, M.; Cole, R. N.; Rosen, A.; Zhu, X. Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. *Arthritis Rheum* **2010**, *62*, 1630–40.

[32] Slack, J. L.; Jones L. E., J.; Bhatia, M. M.; Thompson, P. R. Autodeimination of protein arginine deiminase 4 alters protein-protein interactions but not activity. *Biochemistry* **2011**, *50*, 3997–4010.

[33] Slade, D. J.; Fang, P.; Dreyton, C. J.; Zhang, Y.; Fuhrmann, J.; Rempel, D.; Bax, B. D.; Coonrod, S. A.; Lewis, H. D.; Guo, M.; Gross, M. L.; Thompson, P. R. Protein arginine deiminase 2 binds calcium in an ordered fashion: implications for inhibitor design. *ACS Chem Biol* **2015**, *10*, 1043–53.

[34] Arita, K.; Shimizu, T.; Hashimoto, H.; Hidaka, Y.; Yamada, M.; Sato, M. Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4. *Proc Natl Acad Sci U S A* **2006**, *103*, 5291–6.

[35] Risal, D.; Strickler, M.; Goldstein, B. Ternary complex of the human type II Inosine Monophosphate Dedhydrogenase with Ribavirin Monophosphate and C2-Mycophenolic Adenine Dinucleotide. *Unpublished* **2004**.

[36] Fernandez-Justel, D.; Nunez, R.; Martin-Benito, J.; Jimeno, D.; Gonzalez-Lopez, A.; Soriano, E. M.; Revuelta, J. L.; Buey, R. M. A Nucleotide-Dependent Conformational Switch Controls the Polymerization of Human IMP Dehydrogenases to Modulate their Catalytic Activity. *J Mol Biol* **2019**, *431*, 956–969.

[37] Lafarga, V.; Sirozh, O.; Diaz-Lopez, I.; Galarreta, A.; Hisaoka, M.; Zarzuela, E.; Boskovic, J.; Jovanovic, B.; Fernandez-Leiro, R.; Munoz, J.; Stoecklin, G.; Ventoso, I.; Fernandez-Capetillo, O. Widespread displacement of DNA- and RNA-binding factors underlies toxicity of arginine-rich cell-penetrating peptides. *EMBO J* **2021**, e103311.

[38] Kramer, A. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* **1996**, *65*, 367–409.

[39] Utz, P. J.; Hottelet, M.; van Venrooij, W. J.; Anderson, P. Association of phosphorylated serine/arginine (SR) splicing factors with the U1-small ribonucleoprotein (snRNP) autoantigen complex accompanies apoptotic cell death. *J Exp Med* **1998**, *187*, 547–60.

[40] Smith, C. W.; Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **2000**, *25*, 381–8.

[41] Hof, D.; Raats, J. M.; Pruijn, G. J. Apoptotic modifications affect the autoreactivity of the U1 snRNP autoantigen. *Autoimmun Rev* **2005**, *4*, 380–8.

[42] Aherne, A.; Kennan, A.; Kenna, P. F.; McNally, N.; Lloyd, D. G.; Alberts, I. L.; Kiang, A. S.; Humphries, M. M.; Ayuso, C.; Engel, P. C.; Gu, J. J.; Mitchell, B. S.; Farrar, G. J.; Humphries, P. On the molecular pathology of neurodegeneration in IMPDH1-based retinitis pigmentosa. *Hum Mol Genet* **2004**, *13*, 641–50.

[43] Zheng, L.; Nagar, M.; Maurais, A. J.; Slade, D. J.; Parelkar, S. S.; Coonrod, S. A.; Weerapana, E.; Thompson, P. R. Calcium Regulates the Nuclear Localization of Protein Arginine Deiminase 2. *Biochemistry* **2019**, *58*, 3042–3056.

[44] Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790–5.

[45] Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994**, *5*, 976–89.

[46] Weisser, H.; Nahnsen, S.; Grossmann, J.; Nilse, L.; Quandt, A.; Brauer, H.; Sturm, M.; Kenar, E.; Kohlbacher, O.; Aebersold, R.; Malmstrom, L. An automated pipeline for high-throughput label-free quantitative proteomics. *J Proteome Res* **2013**, *12*, 1628–44.

[47] Levitsky, L. I.; Klein, J. A.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. *J Proteome Res* **2019**, *18*, 709–714.

[48] Knuckley, B.; Causey, C. P.; Jones, J. E.; Bhatia, M.; Dreyton, C. J.; Osborne, T. C.; Takahara, H.; Thompson, P. R. Substrate specificity and kinetic studies of PADs 1, 3, and 4 identify potent and selective inhibitors of protein arginine deiminase 3. *Biochemistry* **2010**, *49*, 4852–4863.

[49] Kearney, P. L.; Bhatia, M.; Jones, N. G.; Luo, Y.; Glascock, M. C.; Catchings, K. L.; Yamada, M.; Thompson, P. R. Kinetic characterization of protein arginine deiminase 4: a transcriptional corepressor implicated in the onset and progression of rheumatoid arthritis. *Biochemistry* **2005**, *44*, 10570–10582.

Appendix I

Supplemental figures

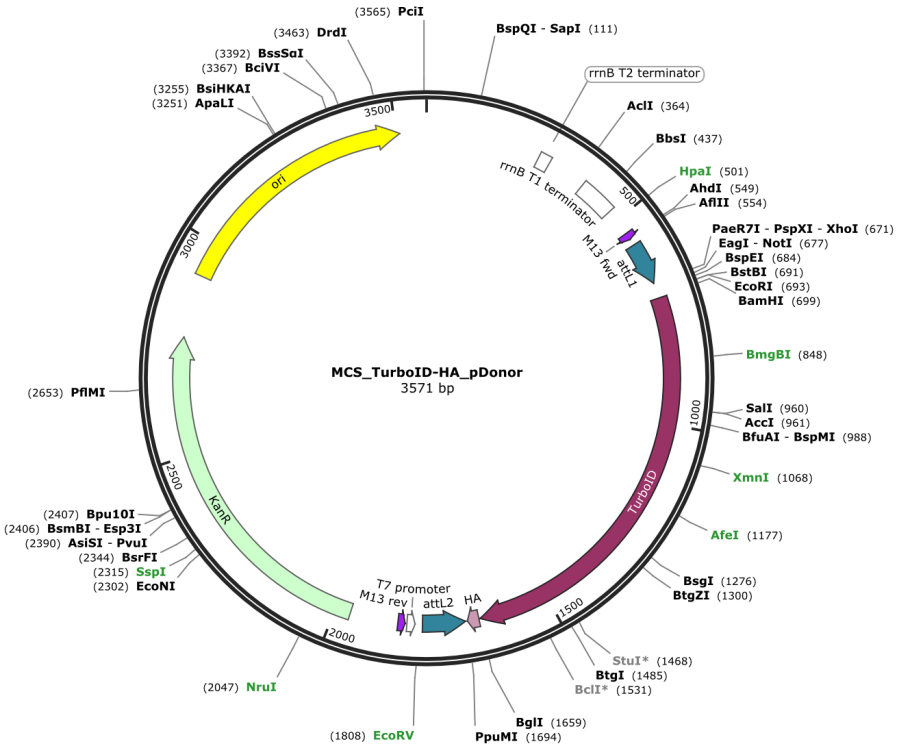


Figure 2A.1: MCS-TurboID-HA pDonor221 map.

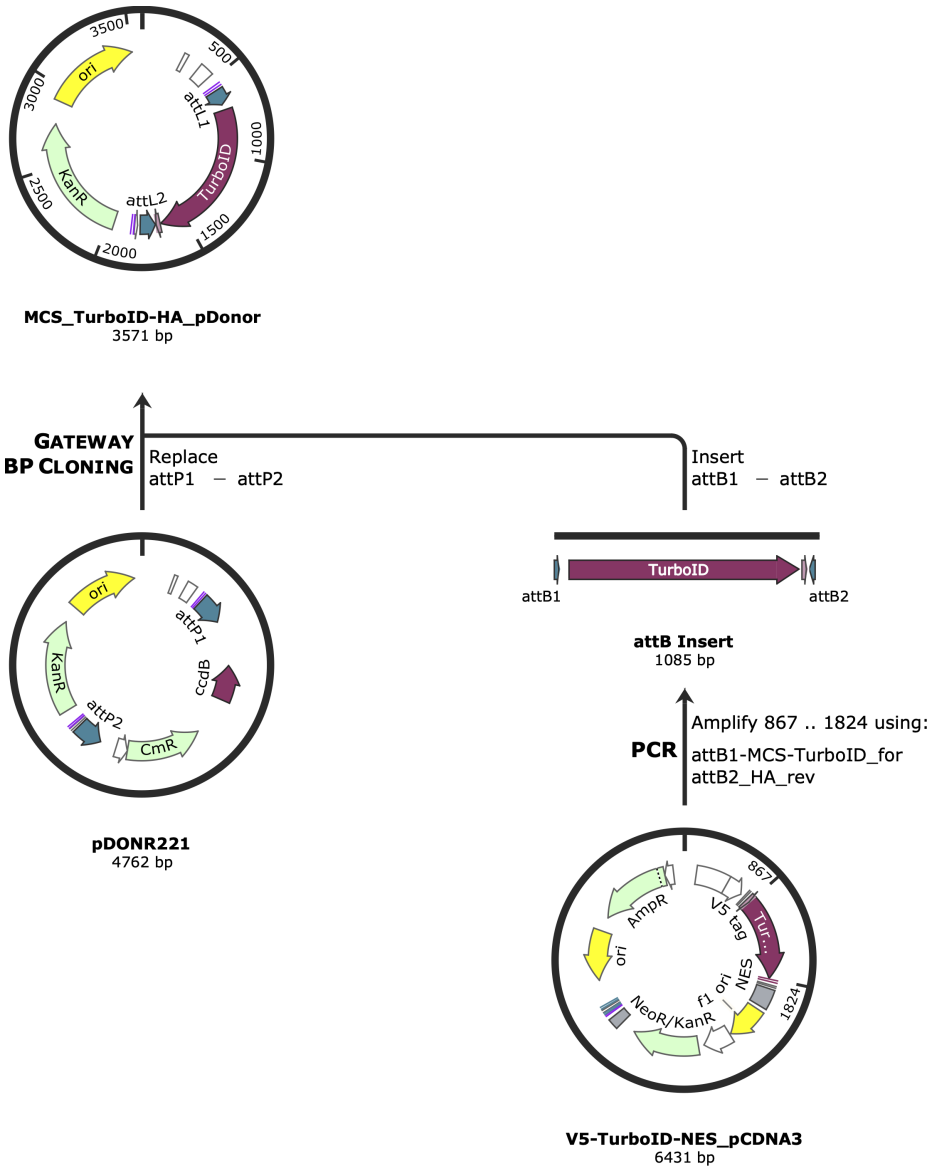


Figure 2A.2: Workflow to generate MCS-TurboID-HA pDonor221.

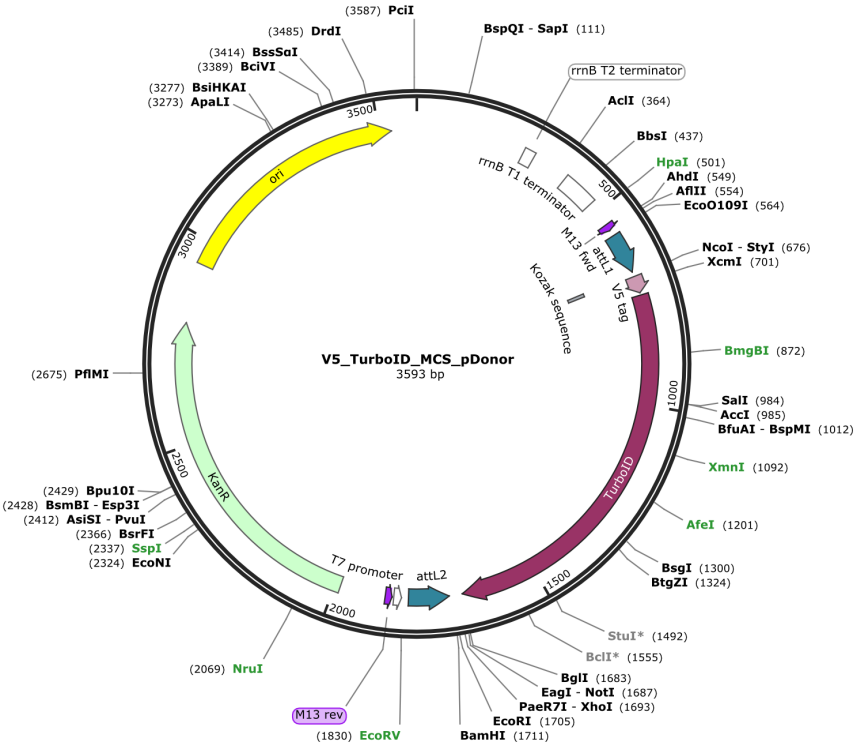


Figure 2A.3: V5-TurboID-MCS pDonor221 map.

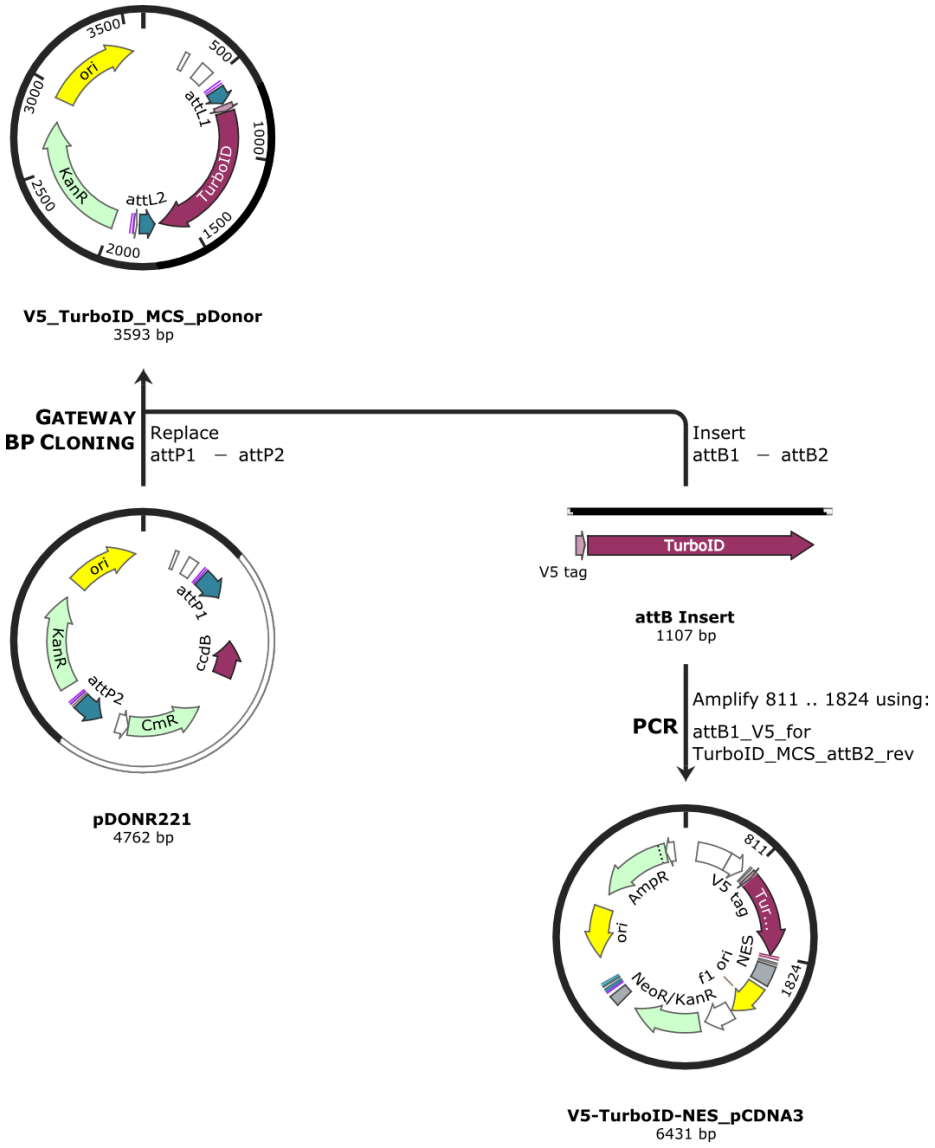


Figure 2A.4: Workflow to generate V5-TurboID-MCS pDonor221.

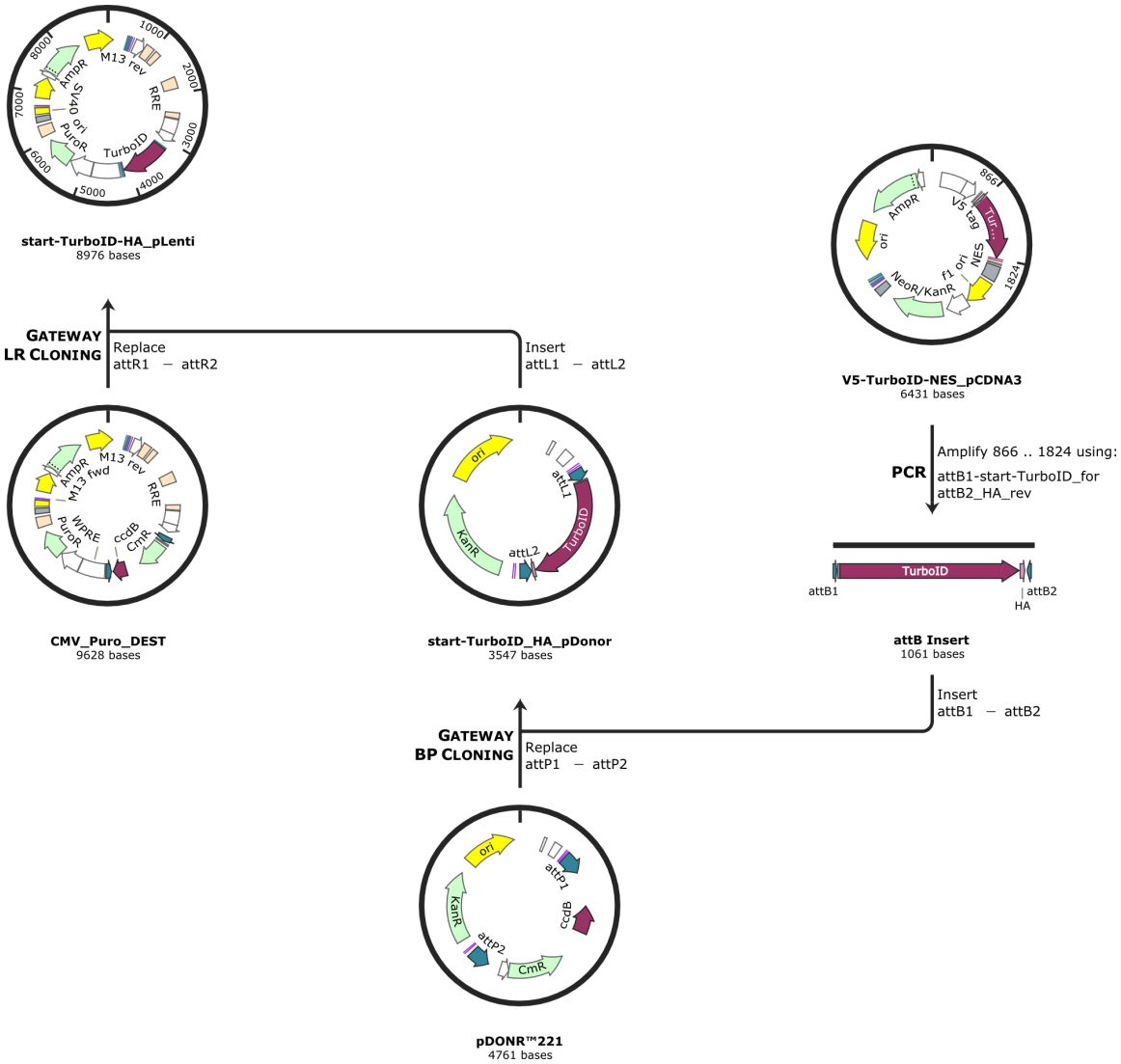


Figure 2A.5: Workflow to generate start-TurboID-HA pLenti control.

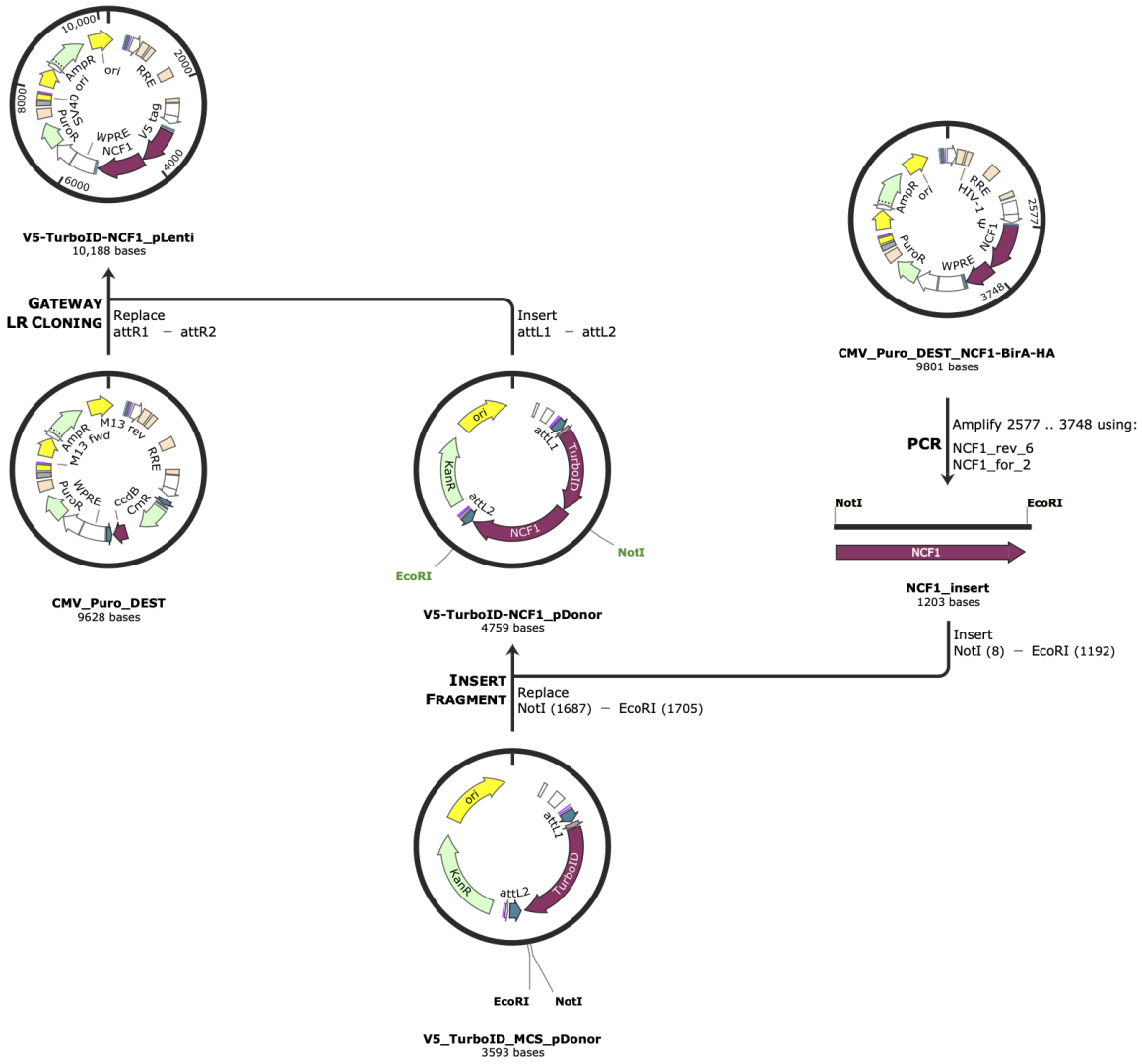


Figure 2A.6: Workflow to generate V5-TurboID-NCF1 pLenti.

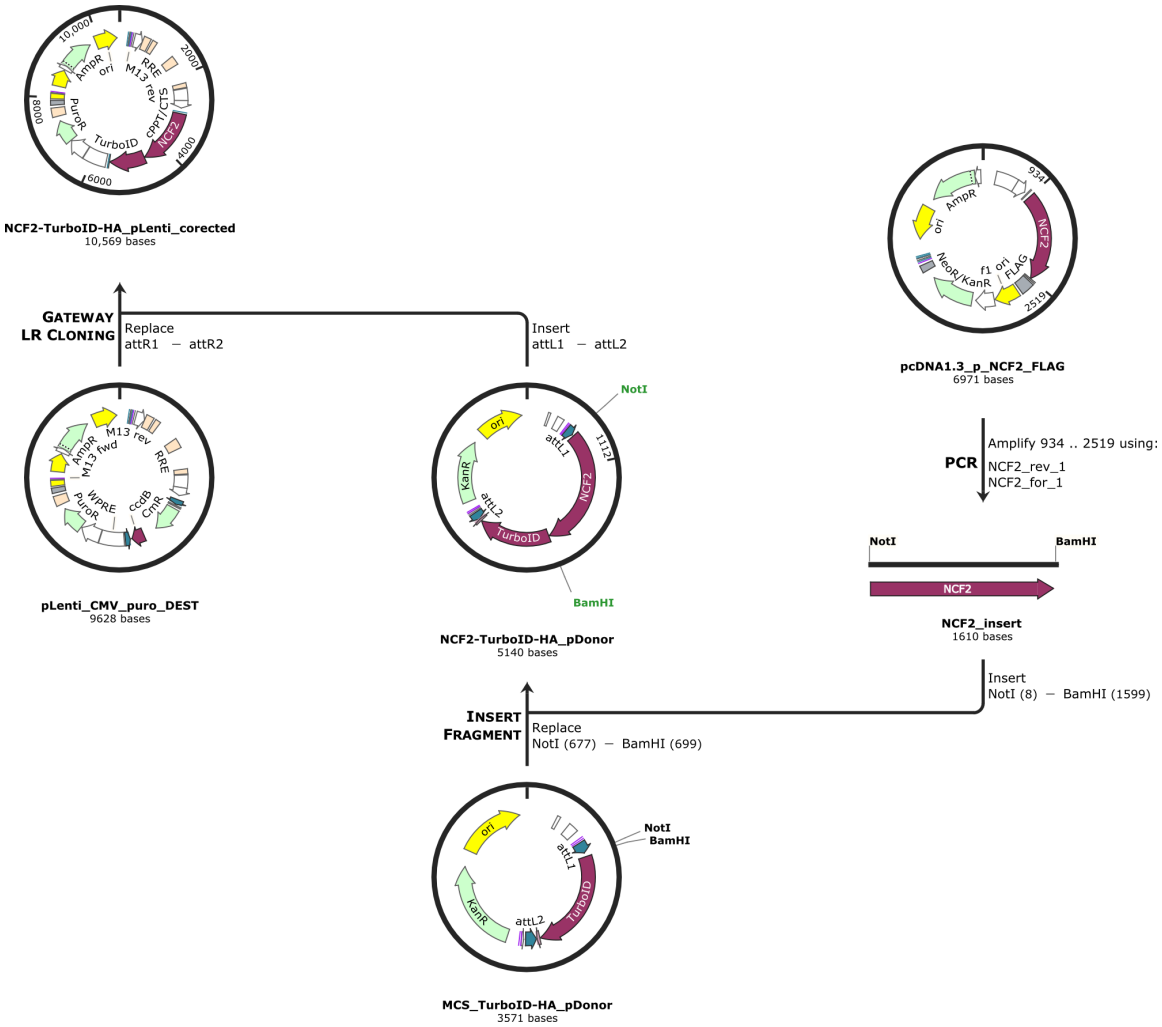


Figure 2A.7: Workflow to generate NCF2-TurboID-HA pLenti.

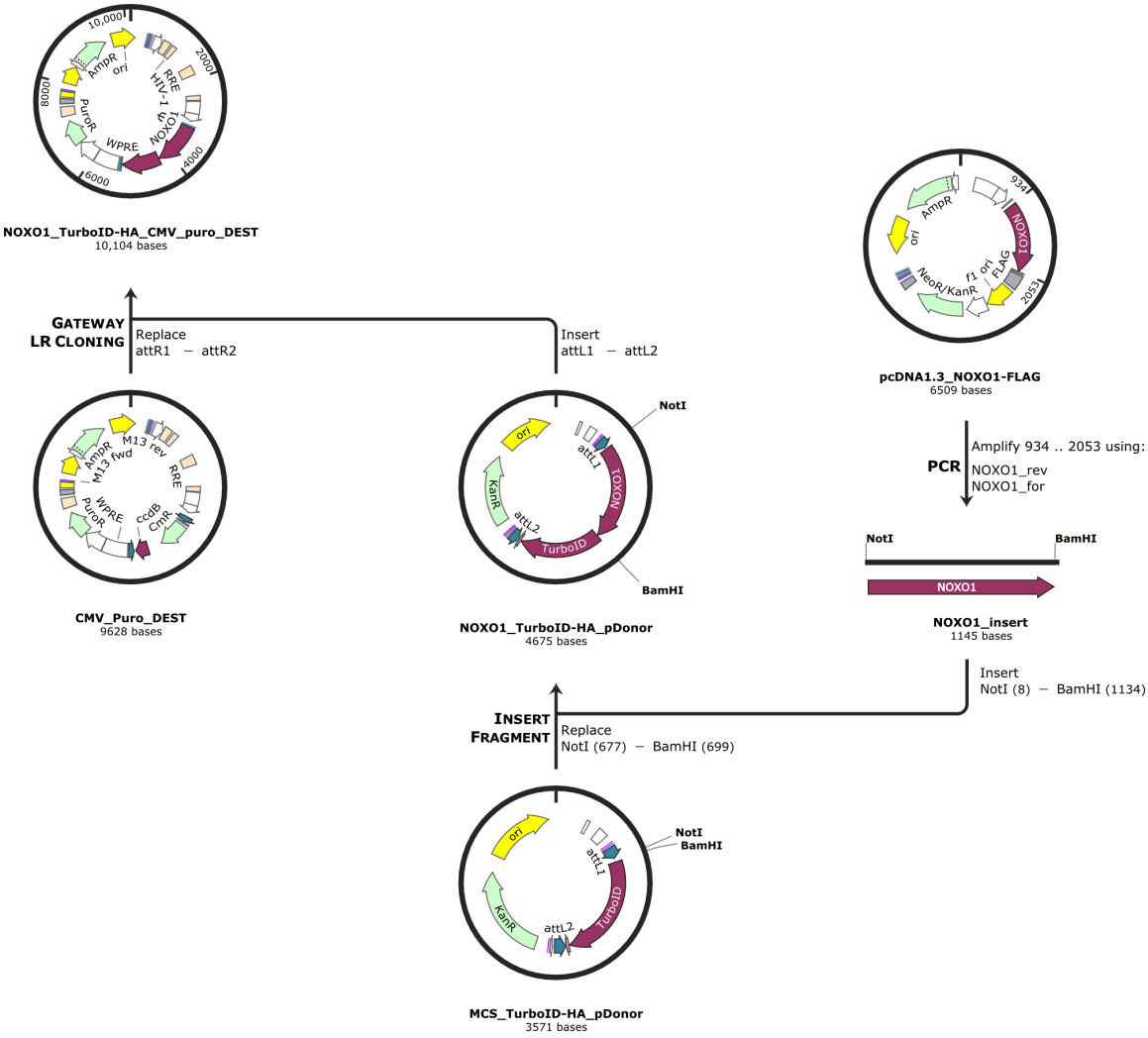


Figure 2A.8: Workflow to generate NOXO1-TurboID-HA pLenti.

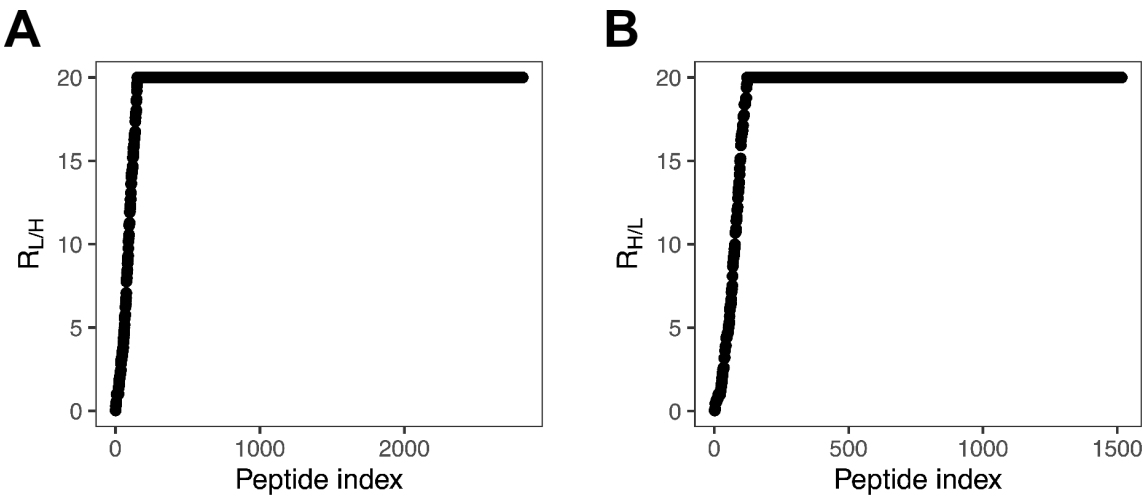


Figure 4A.1: Verification of SILAC labeling in MCF7 cells. **(A)** Light to heavy peptide ratios in unlabeled cell line. **(B)** Heavy to light peptide ratios in heavy labeled cell line.

Appendix II

Mass spectrometry tables

Table 2A.1: Proteins which were selectively enriched in either the NCF1, NCF2, or NOXO1 bait datasets. The $R_{L/H}$ from 2 technical replicates is are shown in the "Ratios" column. The median ratio is shown in the "Median ratios" column. The ratios (if identified) of the protein in the TurboID control dataset are shown in the "TurboID control" and "TID median" columns.

Accession	Protein	Description	Bait	Ratios	Median ratio	TurboID control	TID median
O15381	NVL	NVL Nuclear valosin-containing protein-like	NCF1	20.00|20.00	20.00		
O43719	HTATSF1	HTATSF1 HIV Tat-specific factor 1	NCF1	20.00|20.00	20.00	13.65	13.65
P19878	NCF2	NCF2 Neutrophil cytosol factor 2	NCF1	20.00|20.00	20.00	0.98|1.18	1.08
P43487	RANBP1	RANBP1 Ran-specific GTPase-activating protein	NCF1	20.00|20.00	20.00		
Q13148	TARDBP	TARDBP TAR DNA-binding protein 43	NCF1	20.00|20.00	20.00	1.28	1.28
P47756	CAPZB	CAPZB F-actin-capping protein subunit beta	NCF1	20.00|16.51	18.25	10.55	10.55
O43175	PHGDH	PHGDH D-3-phosphoglycerate dehydrogenase	NCF1	13.21|20.00	16.61		
P46777	RPL5	RPL5 60S ribosomal protein L5	NCF1	14.63|10.41	12.52	20.00|1.05	10.52
P11388	TOP2A	TOP2A DNA topoisomerase 2-alpha	NCF1	10.51|10.73	10.62	1.11|0.94	1.02
Q14315	FLNC	FLNC Filamin-C	NCF1	10.59|10.46	10.52	20	20.00
Q9NSE4	IARS2	IARS2 Isoleucine–tRNA ligase, mitochondrial	NCF1	8.50|10.47	9.49	0.75	0.75
Q9ULK4	MED23	MED23 Mediator of RNA polymerase II transcription subunit	NCF1	1.77|2.99	2.38	0.86	0.86
P14635	CCNB1	CCNB1 G2/mitotic-specific cyclin-B1	NCF1	2.97|1.53	2.25		
P10599	TXN	TXN Thioredoxin	NCF1	2.74|1.55	2.14	0.72|0.91	0.82
P01857	IGHG1	IGHG1 Immunoglobulin heavy constant gamma 1	NCF2	20.00|20.00	20.00		
P01861	IGHG4	IGHG4 Immunoglobulin heavy constant gamma 4	NCF2	20.00|20.00	20.00		
P02768	ALB	ALB Serum albumin	NCF2	20.00|20.00	20.00		
P26232	CTNNA2	CTNNA2 Catenin alpha-2	NCF2	20.00|20.00	20.00		
P47929	LGALS7	LGALS7 Galectin-7	NCF2	20.00|20.00	20.00		
P46736	BRCC3	BRCC3 Lys-63-specific deubiquitinase BRCC36	NCF2	20.00|10.31	15.16	0.95|1.08	1.01
O43303	CCP110	CCP110 Centriolar coiled-coil protein of 110 kDa	NCF2	10.49|10.48	10.49		
P31947	SFN	SFN 14-3-3 protein sigma	NCF2	4.64|5.70	5.17		
P04792	HSPB1	HSPB1 Heat shock protein beta-1	NCF2	4.13|2.25	3.19		
P57721	PCBP3	PCBP3 Poly(rC)-binding protein 3	NOXO1	20.00|20.00	20.00		
Q92614	MYO18A	MYO18A Unconventional myosin-XVIIIa	NOXO1	10.38|20.00	15.19		
P07204	THBD	THBD Thrombomodulin	NOXO1	10.44|10.75	10.60	2.02|20.00	11.01
P46777	RPL5	RPL5 60S ribosomal protein L5	NOXO1	10.35|10.46	10.41	20.00|1.05	10.52
Q9UM54	MYO6	MYO6 Unconventional myosin-VI	NOXO1	2.87|2.97	2.92	1.68|1.50	1.59
P62805	HIST1H4A	HIST1H4A Histone H4	NOXO1	2.80|1.82	2.31	1.34|0.45	0.90
P0CG47	UBB	UBB Polyubiquitin-B	NOXO1	2.42|2.16	2.29	1.3	1.30
P0CG48	UBC	UBC Polyubiquitin-C	NOXO1	2.42|2.16	2.29	1.3	1.30
P62979	RPS27A	RPS27A Ubiquitin-40S ribosomal protein S27a	NOXO1	2.42|2.16	2.29	1.3	1.30
P62987	UBA52	UBA52 Ubiquitin-60S ribosomal protein L40	NOXO1	2.42|2.16	2.29	1.3	1.30

Table 2A.2: Percent oxidation of proteins identified in each bait dataset. There is a row for each bait protein and peptide sequence. The mean percent oxidation for the peptide in the (-) CAT and (+) CAT datasets are shown on the same row. The "Different" column indicates whether the protein was selectively oxidized in either the (-) CAT, (+) CAT datasets.

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
O95372	LYPLA2	DLAILQC*HGELDPMVPVR	C171	NOXO1	25.38	0.00	(-) CAT	
O95373	IPO7	GIDQC*IPLFVEAALER	C757	NCF1	13.06	0.00	(-) CAT	
O95551-1	TDP2	C*GGLPNNIVDVWEFLGKPK	C273	NCF1	14.36	0.00	(-) CAT	
P00367	GLUD1	C*AVVDVPFGGAK	C172	NOXO1	16.43	0.00	(-) CAT	MOD_RES
P00533-1	EGFR	AC*GADSYEMEEDGVRK	C311	NCF1	81.52	60.56	(-) CAT	DISULFID
P05556-1	ITGB1	SNGLIC*GGNGVC*K	C568;C574	NCF2	100.00	78.90	(-) CAT	DISULFID; DISULFID
P14923	JUP	EAMC*PGVSGEDSSLLLATQVEGQAT NLQR	C90	NCF1	18.71	5.71	(-) CAT	
P21333	FLNA	YWPQEAGEYAVHVLC*NSEDIR	C649	NCF1	29.24	13.73	(-) CAT	
P35221	CTNNA1	C*VIALQEK	C526	NCF2	21.80	0.00	(-) CAT	
P49792	RANBP2	VIPDFVC*QGGDITK	C3122	NCF1	13.69	0.00	(-) CAT	
Q03701	CEBPZ	ASHLLETLLC*K	C411	NCF2	84.75	64.48	(-) CAT	
Q13835	PKP1	QQVYQLGGIC*K	C290	NOXO1	14.09	0.00	(-) CAT	
Q14247-1	CTTN	HC*SQVDSVR	C112	NOXO1	62.92	23.47	(-) CAT	
Q1KMD3	HNRNPUL2	DLLVQQASQC*LSK	C518	NCF1	27.61	0.00	(-) CAT	
Q69YH5-1	CDCA2	SPATPAC*R	C315	NCF2	30.48	9.76	(-) CAT	
Q8IWI9-1	MGA	SAFC*SDKLDEYLENEGK	C823	NCF1	15.42	0.00	(-) CAT	
Q9HD45	TM9SF3	YFSLPFC*VGSK	C66	NCF2	40.63	17.09	(-) CAT	
Q9NQS7	INCENP	MGTTAPGPIHLLELC*DQK	C15	NCF1	20.09	0.00	(-) CAT	
Q9NRD9	DUOX1	AC*QLINR	C1533	NOXO1	43.79	0.00	(-) CAT	
Q9ULV4	CORO1C	NDQC*YDDIR	C23	NCF2	13.82	0.00	(-) CAT	
Q9Y446-2	PKP3	TTSRPEAGVC*SLALPSDLQLDR	C33	NCF2	18.21	0.00	(-) CAT	
O60716-1	CTNND1	NC*DGVPALVR	C450	NCF1	3.01	21.57	(+) CAT	
P04075	ALDOA	ALANSLAC*QGK	C339	NCF2	0.00	17.08	(+) CAT	
P04406-1	GAPDH	VPTANVSVVDLTC*R	C247	NOXO1	9.91	25.63	(+) CAT	MOD_RES; MOD_RES
P05023	ATP1A1	C*SSILLHGK	C518	NOXO1	0.00	19.30	(+) CAT	
P05023	ATP1A1	LIIVEGC*QR	C705	NCF2	6.92	24.23	(+) CAT	
P14618	PKM	NTGIIC*TIGPASR	C49	NCF2	0.00	16.53	(+) CAT	
P14923	JUP	VAAGVLC*ELAQDKEAADAIDAEGAS APLMELLHSR	C609	NCF2	5.48	21.12	(+) CAT	
P30050-1	RPL12	C*TGGEVGATSALAPK	C17	NOXO1	0.00	91.01	(+) CAT	
P37802	TAGLN2	DGTVLC*ELINALYPEGQAPVKK	C63	NCF1	1.03	13.73	(+) CAT	
P39023	RPL3	VAC*IGAWHPAR	C253	NCF1	8.51	25.13	(+) CAT	
P62244	RPS15A	QVLIRPC*SK	C30	NCF2	0.00	100.00	(+) CAT	
P62280	RPS11	NMSVHLSPC*FR	C116	NCF2	0.00	14.79	(+) CAT	
P62879	GNB2	TFVSGAC*DASIK	C204	NCF2	0.00	21.31	(+) CAT	
Q13470-2	TNK1	LPRPPLC*SR	C348	NCF1	0.00	29.91	(+) CAT	
Q7L2H7	EIF3M	YTVYC*SLIK	C125	NOXO1	0.00	100.00	(+) CAT	
Q8IVF2-3	AHNAK2	VDPEC*SVEDSK	C4879	NCF2	0.00	20.51	(+) CAT	
Q9P206-2	KIAA1522	C*SLHSASPASVR	C380	NOXO1	73.22	85.74	(+) CAT	
Q9Y624	F11R	LVC*YNNK	C74	NOXO1	0.00	16.82	(+) CAT	DISULFID
A0MZ66	KIAA1598 SHTN1	VTFQPPSSIGC*R	C565	NOXO1	100.00		same	
F2Z333	C1orf233 FNDC10	SPARGFRC*QAPGC*VLHAPAGR	C64;C69	NCF2	100.00	100.00	same	
O00148	DDX39A	C*MALAQLLVEQNFPAIAIHR	C299	NCF1		100.00	same	
O00148	DDX39A	C*MALAQLLVEQNFPAIAIHR	C299	NCF2		100.00	same	
O00151	PDLIM1	C*GTGIVGVFVK	C263	NCF1		13.80	same	METAL
O00151	PDLIM1	C*GTGIVGVFVK	C263	NCF2		9.19	same	METAL
O00213	APBB1	SSVAVNNC*IR	C397	NCF1	0.00		same	
O00213	APBB1	NIATSLHEIC*SK	C499	NCF1	0.00	0.00	same	
O00443	PIK3C2A	STANC*HLER	C317	NCF1	29.19		same	
O00483	NDUFA4	LALFNPDVC*WDR	C44	NCF1	0.00		same	
O00505	KPNA3	DQVEYLVQQNVIPPFC*NLLSVK	C417	NCF1		0.00	same	
O00515	LAD1	GLPC*TELFVAPVGVASK	C428	NCF1	0.00		same	
O00515	LAD1	GLPC*TELFVAPVGVASK	C428	NCF2	0.00		same	
O00541-1	PES1	DLDDALSMC*FLFSTFPR	C141	NCF1	0.00		same	
O00541-1	PES1	C*HVQTIQLC*R	C153;C161	NCF1	0.00	0.00	same	
O00567	NOP56	IINDNATYC*R	C211	NCF1	0.00	0.00	same	
O00567	NOP56	IINDNATYC*R	C211	NCF2	0.00	0.00	same	
O00567	NOP56	IINDNATYC*R	C211	NOXO1	0.00	0.00	same	
O00571	DDX3X	DLMAC*AQTGSGK	C223	NCF1	0.00	0.00	same	
O00571	DDX3X	VRPC*VVYGGADIGQQIR	C298	NCF1	14.68	25.67	same	
O00571	DDX3X	VRPC*VVYGGADIGQQIR	C298	NCF2		0.00	same	
O00571	DDX3X	GC*HLLVATPGR	C317	NCF1	0.00	0.00	same	
O00571	DDX3X	IGLDFC*K	C341	NCF1	0.00	0.00	same	
O00571	DDX3X	IGLDFC*K	C341	NCF2	0.00	0.00	same	
O00571	DDX3X	IGLDFC*K	C341	NOXO1	0.00	0.00	same	
O14617-1	AP3D1	LEDPDPGVQSAAVNVIC*ELAR	C208	NCF1	0.00	0.00	same	
O14617-1	AP3D1	LEDPDPGVQSAAVNVIC*ELAR	C208	NCF2	0.00		same	
O14617-1	AP3D1	LEDPDPGVQSAAVNVIC*ELAR	C208	NOXO1	0.00	0.00	same	
O14617-1	AP3D1	ASC*ILQLVK	C574	NCF1		37.80	same	
O14672	ADAM10	TITLQPGSPC*NDFR	C632	NCF2	100.00	100.00	same	DISULFID
O14672	ADAM10	TITLQPGSPC*NDFR	C632	NOXO1	100.00	100.00	same	DISULFID
O14684	PTGES	HGGPQYC*R	C59	NCF1	0.00		same	
O14684	PTGES	HGGPQYC*R	C59	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
O14776-1	TCERG1	YLVLDC*VPEER	C1062	NCF1	0.00	0.00	same	
O15084-1	ANKRD28	LLVSHGAEVTC*K	C234	NCF1		0.00	same	
O15226	NKRF	LLTDGYAC*EVR	C242	NCF1	26.76		same	
O15226	NKRF	LLTDGYAC*EVR	C242	NOXO1	100.00		same	
O15260-1	SURF4	LC*LISTFLEDGIR	C32	NCF1	0.00	0.00	same	
O15260-1	SURF4	LC*LISTFLEDGIR	C32	NCF2		0.00	same	
O15440	ABCC5	LVELSGGC*IK	C1250	NCF1		0.00	same	
O15440	ABCC5	HESSDVNC*R	C146	NOXO1	0.00		same	
O43143	DHX15	FEDC*SSAK	C226	NCF1	0.00	0.00	same	
O43143	DHX15	FEDC*SSAK	C226	NCF2		0.00	same	
O43143	DHX15	FEDC*SSAK	C226	NOXO1		0.00	same	
O43143	DHX15	TC*TDIKPEWLVK	C750	NCF1	0.00	0.00	same	
O43172-1	PRPF4	C*IMFLEGHLK	C399	NCF1	0.00		same	
O43172-1	PRPF4	C*IMFLEGHLK	C399	NOXO1		0.00	same	
O43290	SART1	GAIVFNATSEFC*R	C560	NCF1	0.00	0.00	same	
O43290	SART1	KGAIVFNATSEFC*R	C560	NCF1	100.00		same	
O43290	SART1	KGAIVFNATSEFC*R	C560	NCF2	100.00	100.00	same	
O43290	SART1	KGAIVFNATSEFC*R	C560	NOXO1		100.00	same	
O43290	SART1	GLAAALLLC*QNK	C645	NCF1	0.00	0.00	same	
O43290	SART1	SLPSAVYC*IEDK	C674	NCF1	0.00	0.00	same	
O43290	SART1	SLPSAVYC*IEDK	C674	NCF2	0.00		same	
O43395	PRPF3	VLGFSEPTVVTAALNC*VGK	C37	NCF1	7.37	6.16	same	
O43451	MGAM	ILDGYLFC*K	C565	NCF1	100.00	100.00	same	
O43491-1	EPB41L2	VC*VEHHTFYR	C488	NCF1	0.00	0.00	same	
O43657	TSPAN6	LEDC*TPQR	C183	NCF1	100.00	100.00	same	
O43657	TSPAN6	LEDC*TPQR	C183	NCF2	100.00	100.00	same	
O43657	TSPAN6	LEDC*TPQR	C183	NOXO1	100.00	100.00	same	
O43795-2	MYO1B	LVMSYVAAVC*GK	C129	NCF2	0.00		same	
O43795-2	MYO1B	LVMSYVAAVC*GK	C129	NOXO1		11.50	same	
O60231	DHX16	TGMESLTVTPC*SK	C714	NCF1	0.00	0.00	same	
O60231	DHX16	TGMESLTVTPC*SK	C714	NCF2	0.00	0.00	same	
O60231	DHX16	TGMESLTVTPC*SK	C714	NOXO1	0.00		same	
O60264	SMARCA5	C*NTLITLIER	C1001	NCF1	0.00	0.00	same	
O60264	SMARCA5	SVC*LIGDKEQR	C259	NCF1	19.51		same	
O60264	SMARCA5	SVC*LIGDK	C259	NCF2	100.00	100.00	same	
O60264	SMARCA5	SVC*LIGDKEQR	C259	NCF2	0.00	0.00	same	
O60264	SMARCA5	SVC*LIGDKEQR	C259	NOXO1	0.00	0.00	same	
O60264	SMARCA5	VLDILEDYC*MWR	C519	NCF1	0.00	0.00	same	
O60437	PPL	C*DLEIYQLK	C1265	NCF1	0.00		same	
O60437	PPL	C*DLEIYQLK	C1265	NCF2	0.00	0.00	same	
O60437	PPL	ETPLKPIPVEALC*DFEGEQGLISR	C408	NCF1	0.00	0.00	same	
O60437	PPL	ETPLKPIPVEALC*DFEGEQGLISR	C408	NCF2	0.00	0.00	same	
O60437	PPL	LIAPAVC*FVIPPTDPEALALADSLG SQYR	C449	NCF1	0.00		same	
O60437	PPL	LIAPAVC*FVIPPTDPEALALADSLG SQYR	C449	NCF2	0.00		same	
O60437	PPL	FQEHC*PDLER	C694	NCF1	0.00	0.00	same	
O60437	PPL	FQEHC*PDLER	C694	NCF2	0.00		same	
O60716-1	CTNND1	SNAAAYLQHLC*YR	C394	NCF1	12.92	13.39	same	
O60716-1	CTNND1	SNAAAYLQHLC*YR	C394	NCF2	3.11	6.92	same	
O60716-1	CTNND1	SNAAAYLQHLC*YR	C394	NOXO1	0.00	6.35	same	
O60716-1	CTNND1	EVHLGAC*GALK	C429	NCF1	0.00	3.47	same	
O60716-1	CTNND1	EVHLGAC*GALK	C429	NCF2	0.00	0.00	same	
O60716-1	CTNND1	EVHLGAC*GALK	C429	NOXO1	0.00	0.00	same	
O60716-1	CTNND1	NC*DGVPALVR	C450	NCF2		7.36	same	
O60716-1	CTNND1	NC*DGVPALVR	C450	NOXO1	8.15	9.00	same	
O60716-1	CTNND1	EPNEDC*KPR	C516	NCF1	9.38	9.78	same	
O60716-1	CTNND1	EPNEDC*KPR	C516	NCF2	5.66	15.71	same	
O60716-1	CTNND1	EPNEDC*KPR	C516	NOXO1	6.09	8.25	same	
O60716-1	CTNND1	HIEWESVLTNTAGC*LR	C533	NCF1	0.00	0.00	same	
O60716-1	CTNND1	HIEWESVLTNTAGC*LR	C533	NCF2	0.00	0.00	same	
O60716-1	CTNND1	HIEWESVLTNTAGC*LR	C533	NOXO1		0.00	same	
O60716-1	CTNND1	LVENC*VC*LLR	C579;C581	NCF1	0.00	0.00	same	
O60716-1	CTNND1	LVENC*VC*LLR	C579;C581	NCF2	0.00	0.00	same	
O60716-1	CTNND1	LVENC*VC*LLR	C579;C581	NOXO1	1.62	0.00	same	
O60716-1	CTNND1	YQEAAPNVANNTGPHAASC*FGAK	C618	NCF1		100.00	same	
O60716-1	CTNND1	YQEAAPNVANNTGPHAASC*FGAK	C618	NOXO1	0.00	0.00	same	
O60716-1	CTNND1	TPAILEASAGAIQNLC*AGR	C692	NCF1	5.92	3.75	same	
O60716-1	CTNND1	TPAILEASAGAIQNLC*AGR	C692	NCF2	4.71	1.82	same	
O60716-1	CTNND1	TPAILEASAGAIQNLC*AGR	C692	NOXO1	3.75	0.00	same	
O60832-1	DKC1	TTHYTPLAC*GSNPLKR	C74	NCF1	0.00		same	
O75131	CPNE3	EALAQC*VLAEIPQQVVGYFNTYK	C506	NCF1	38.43		same	
O75165	DNAJC13	VIHALSENELC*VR	C2082	NCF1		0.00	same	
O75165	DNAJC13	VIHALSENELC*VR	C2082	NCF2		100.00	same	
O75165	DNAJC13	VIHALSENELC*VR	C2082	NOXO1		100.00	same	
O75179-1	ANKRD17	FC*ELLIGR	C1336	NCF1	0.00	0.00	same	
O75179-1	ANKRD17	FC*ELLIGR	C1336	NCF2	0.00	0.00	same	
O75323-1	GBAS NIPSNAP2	IC*QEVLPK	C93	NCF2	0.00	5.67	same	
O75330-1	HMMR	LLEYIEEISC*ASDQVEK	C242	NCF1	0.00	0.00	same	
O75330-1	HMMR	LLEYIEEISC*ASDQVEK	C242	NCF2		0.00	same	
O75367-1	H2AFY	NC*LALADDKK	C297	NCF1	0.00	0.00	same	
O75367-1	H2AFY	NC*LALADDKK	C297	NOXO1	0.00		same	
O75533-1	SF3B1	VQENC*IDLVGR	C1035	NCF1	12.49	9.99	same	
O75533-1	SF3B1	IC*FELLELLK	C1059	NCF1	0.00	0.00	same	
O75533-1	SF3B1	IC*FELLELLK	C1059	NCF2	0.00	0.00	same	
O75533-1	SF3B1	VAIGPC*R	C1244	NCF1	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
O75533-1	SF3B1	MLQYC*LQGLFHPAR	C1250	NCF1	0.00		same	
O75533-1	SF3B1	MLQYC*LQGLFHPAR	C1250	NCF1	0.00	0.00	same	
O75533-1	SF3B1	VKPYLPQIC*GTVLWR	C933	NCF1	0.00	0.00	same	
O75534	CSDE1	LLTSYGFIQC*SER	C42	NCF1	0.00	0.00	same	
O75534	CSDE1	QRPGQQVATC*VR	C506	NCF1	0.00	0.00	same	
O75643-1	SNRNP200	DILC*GAADEVLAVLK	C133	NCF1	0.00	0.00	same	
O75643-1	SNRNP200	DILC*GAADEVLAVLK	C133	NCF2	0.00	0.00	same	
O75643-1	SNRNP200	DILC*GAADEVLAVLK	C133	NOXO1	0.00	0.00	same	
O75643-1	SNRNP200	AALETDENLLLC*APTGAGK	C502	NCF1	0.00	0.00	same	
O75643-1	SNRNP200	TNVALMC*MLR	C516	NCF2	100.00		same	
O75643-1	SNRNP200	LATYGITVAELTGDHQLC*K	C576	NCF1	0.00	0.00	same	
O75683	SURF6	KAEEATEAQEVVEATPEGAC*TEPR	C189	NCF1	0.00	0.00	same	
O76021	RSL1D1	EINDC*IGGTVLNISK	C197	NOXO1	100.00	100.00	same	
O76021	RSL1D1	SGSC*SAIR	C211	NCF1	14.54	18.27	same	
O76021	RSL1D1	SGSC*SAIR	C211	NOXO1	0.00	3.21	same	
O76021	RSL1D1	AVDALLTHC*K	C47	NCF1	0.00	0.00	same	
O76021	RSL1D1	KAVDALLTHC*K	C47	NCF1	0.00	0.00	same	
O76021	RSL1D1	AVDALLTHC*K	C47	NCF2	0.00	0.00	same	
O76021	RSL1D1	AVDALLTHC*K	C47	NOXO1	0.00	0.00	same	
O95218-1	ZRANB2	VSDGDWIC*PDKK	C15	NCF1	0.00		same	
O95218-1	ZRANB2	C*GNVNFAR	C20	NCF1	23.92	28.65	same	
O95218-1	ZRANB2	TC*SNVNWAR	C74	NCF1	0.00	0.00	same	
O95361-1	TRIM16	EAELQC*TQLDLER	C178	NCF2		100.00	same	
O95372	LYPLA2	DLAILQC*HGELDPMVPVR	C171	NCF1	14.93	10.17	same	
O95372	LYPLA2	DLAILQC*HGELDPMVPVR	C171	NCF2	5.03	6.15	same	
O95372	LYPLA2	TYPGVMHSSC*PQEMAAVK	C213	NCF1	0.00	4.30	same	
O95372	LYPLA2	YIC*PHAPR	C56	NCF1	12.00	10.48	same	
O95372	LYPLA2	YIC*PHAPR	C56	NCF2	0.00	0.00	same	
O95372	LYPLA2	YIC*PHAPR	C56	NOXO1	0.00	0.00	same	
O95373	IPO7	VLTGVAGEDAEC*HAAK	C736	NCF1	0.00	0.00	same	
O95373	IPO7	VLTGVAGEDAEC*HAAK	C736	NCF2	0.00	0.00	same	
O95373	IPO7	VLTGVAGEDAEC*HAAK	C736	NOXO1	0.00		same	
O95373	IPO7	GIDQC*IPLFVEAALER	C757	NCF2		100.00	same	
O95391	SLU7	YSYC*TGEAGK	C455	NCF1	0.00		same	
O95551-1	TDP2	C*GGLPNNIVDVWEFLGKPK	C273	NCF2	0.00		same	
O95573	ACSL3	LLLC*GGAPLSATTQR	C450	NCF1		0.00	same	
O95573	ACSL3	LLLC*GGAPLSATTQR	C450	NCF2	0.00	0.00	same	
O95573	ACSL3	LLLC*GGAPLSATTQR	C450	NOXO1	0.00	0.00	same	
O95573	ACSL3	VGAPLVC*C*EIK	C503;C504	NCF1	0.00	0.00	same	
O95573	ACSL3	VGAPLVC*C*EIK	C503;C504	NCF2	0.00		same	
O95573	ACSL3	VGAPLVC*C*EIK	C503;C504	NOXO1	0.00	0.00	same	
O95782	AP2A1	QSAALC*LLR	C171	NCF2	0.00	0.00	same	
O95782	AP2A1	QSAALC*LLR	C171	NOXO1	0.00	0.00	same	
P00338-1	LDHA	DDVFLSVPC*ILGQNGISDLVK	C293	NCF1	0.00		same	
P00390	GSR	LPVMTMIPDVDC*LLWAIGR	C328	NCF1	0.00		same	
P00390	GSR	LPVMTMIPDVDC*LLWAIGR	C328	NCF2	0.00	0.00	same	
P00390	GSR	GIYAVGDVC*GK	C377	NCF1	100.00		same	
P00390	GSR	GIYAVGDVC*GK	C377	NCF2	100.00		same	
P00390	GSR	GIYAVGDVC*GK	C377	NOXO1	100.00		same	
P00505	GOT2	NLDKEYLPIGGLAEFC*K	C106	NCF1	0.00		same	
P00505	GOT2	NLDKEYLPIGGLAEFC*K	C106	NCF2	0.00	0.00	same	
P00505	GOT2	VGAFTMVC*K	C295	NCF1	0.00	0.00	same	
P00505	GOT2	VGAFTMVC*K	C295	NCF2	0.00	0.00	same	
P00505	GOT2	VGAFTMVC*K	C295	NOXO1	0.00	0.00	same	
P00533-1	EGFR	TPLLSSLSATSNNSTVAC*IDR	C1049	NCF1	17.74		same	LIPID
P00533-1	EGFR	TPLLSSLSATSNNSTVAC*IDR	C1049	NCF2	36.38	37.45	same	LIPID
P00533-1	EGFR	TPLLSSLSATSNNSTVAC*IDR	C1049	NOXO1	40.92	39.16	same	LIPID
P00533-1	EGFR	NGLQSC*PIKEDSFLQR	C1058	NOXO1	59.13		same	
P00533-1	EGFR	FSNNPALC*NVESIQWR	C157	NCF1	62.38	68.50	same	DISULFID
P00533-1	EGFR	FSNNPALC*NVESIQWR	C157	NCF2	68.59	70.87	same	DISULFID
P00533-1	EGFR	FSNNPALC*NVESIQWR	C157	NOXO1	92.99	96.11	same	DISULFID
P00533-1	EGFR	IIC*AQQC*SGR	C215;C219	NCF1	95.46	95.79	same	DISULFID; DISULFID
P00533-1	EGFR	IIC*AQQC*SGR	C215;C219	NCF2	100.00	100.00	same	DISULFID; DISULFID
P00533-1	EGFR	IIC*AQQC*SGR	C215;C219	NOXO1	100.00	100.00	same	DISULFID; DISULFID
P00533-1	EGFR	ESDC*LVC*R	C248;C251	NCF1	100.00	100.00	same	DISULFID; DISULFID
P00533-1	EGFR	ESDC*LVC*R	C248;C251	NCF2	100.00	100.00	same	DISULFID; DISULFID
P00533-1	EGFR	ESDC*LVC*R	C248;C251	NOXO1	100.00	100.00	same	DISULFID; DISULFID
P00533-1	EGFR	FRDEATC*K	C260	NCF1	71.13	67.21	same	DISULFID
P00533-1	EGFR	FRDEATC*K	C260	NCF2	77.61	80.09	same	DISULFID
P00533-1	EGFR	FRDEATC*K	C260	NOXO1	93.19	95.05	same	DISULFID
P00533-1	EGFR	DTC*PPLMLYNPTTYQMDVNPEGK	C264	NOXO1	100.00	100.00	same	DISULFID
P00533-1	EGFR	YSFGATC*VK	C291	NCF2	62.65	63.82	same	DISULFID
P00533-1	EGFR	YSFGATC*VK	C291	NOXO1	85.47	87.63	same	DISULFID
P00533-1	EGFR	NYVVTDHGSC*VR	C307	NCF1	62.64	60.65	same	DISULFID
P00533-1	EGFR	NYVVTDHGSC*VR	C307	NCF2	71.66	73.77	same	DISULFID
P00533-1	EGFR	NYVVTDHGSC*VR	C307	NOXO1	91.53	92.76	same	DISULFID
P00533-1	EGFR	AC*GADSYEMEEDGVR	C311	NCF1	88.95		same	DISULFID
P00533-1	EGFR	AC*GADSYEMEEDGVR	C311	NCF2	100.00	90.49	same	DISULFID

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P00533-1	EGFR	AC*GADSYEMEEDGVRK	C311	NCF2	79.82	84.63	same	DISULFID
P00533-1	EGFR	AC*GADSYEMEEDGVR	C311	NOXO1	96.31	100.00	same	DISULFID
P00533-1	EGFR	AC*GADSYEMEEDGVRK	C311	NOXO1	100.00	100.00	same	DISULFID
P00533-1	EGFR	AC*GADSYEMEEDGVRK	C311	NOXO1	90.57	94.61	same	DISULFID
P00533-1	EGFR	VC*NGIGIGEFK	C337	NCF1	62.71	62.20	same	DISULFID
P00533-1	EGFR	VC*NGIGIGEFK	C337	NCF2	76.33	73.40	same	DISULFID
P00533-1	EGFR	VC*NGIGIGEFK	C337	NOXO1	83.63	83.45	same	DISULFID
P00533-1	EGFR	NLC*YANTINWK	C470	NCF2		100.00	same	DISULFID
P00533-1	EGFR	NLC*YANTINWK	C470	NOXO1		100.00	same	DISULFID
P00533-1	EGFR	ATGQVC*HALC*SPEGC*WGPEPR	C506;C510; C515	NCF1		100.00	same	DISULFID; DISULFID; DISULFID
P00533-1	EGFR	ATGQVC*HALC*SPEGC*WGPEPR	C506;C510; C515	NCF2		100.00	same	DISULFID; DISULFID; DISULFID
P00533-1	EGFR	ATGQVC*HALC*SPEGC*WGPEPR	C506;C510; C515	NOXO1	100.00	100.00	same	DISULFID; DISULFID; DISULFID
P00533-1	EGFR	C*NLLEGEPR	C539	NCF1	82.93	88.35	same	DISULFID
P00533-1	EGFR	C*NLLEGEPR	C539	NCF2	100.00	100.00	same	DISULFID
P00533-1	EGFR	C*NLLEGEPR	C539	NOXO1	100.00	100.00	same	DISULFID
P00533-1	EGFR	EILDEAYVMASVDNPHVC*R	C775	NCF1	0.00	0.00	same	
P00533-1	EGFR	EILDEAYVMASVDNPHVC*R	C775	NCF2	0.00	0.00	same	
P00533-1	EGFR	EILDEAYVMASVDNPHVC*R	C775	NOXO1	0.00	0.00	same	
P00533-1	EGFR	C*WMIDADSRPK	C950	NCF1	0.00	0.00	same	
P00533-1	EGFR	C*WMIDADSRPK	C950	NCF2	0.00	0.00	same	
P00533-1	EGFR	C*WMIDADSRPK	C950	NOXO1	0.00	0.00	same	
P02533	KRT14	YC*MQLAQIQEMIGSVEEQLAQLR	C367	NCF1	0.00	0.00	same	DISULFID
P02533	KRT14	YC*MQLAQIQEMIGSVEEQLAQLR	C367	NCF1	25.17	21.02	same	DISULFID
P02533	KRT14	YC*MQLAQIQEMIGSVEEQLAQLR	C367	NCF2		90.16	same	DISULFID
P02533	KRT14	YC*MQLAQIQEMIGSVEEQLAQLR	C367	NOXO1	9.32		same	DISULFID
P04075	ALDOA	ALANSLAC*QGK	C339	NCF1	7.99	15.36	same	
P04075	ALDOA	ALANSLAC*QGK	C339	NOXO1	9.26		same	
P04083	ANXA1	C*LTAIVK	C263	NCF1	0.00	1.87	same	
P04083	ANXA1	C*LTAIVK	C263	NCF2	0.00	0.00	same	
P04083	ANXA1	C*LTAIVK	C263	NOXO1	3.51	0.00	same	
P04083	ANXA1	C*ATSKPAFFAEK	C270	NCF1	8.34	13.99	same	
P04083	ANXA1	C*ATSKPAFFAEK	C270	NCF2	0.00	0.00	same	
P04083	ANXA1	C*ATSKPAFFAEK	C270	NOXO1	0.00	0.00	same	
P04083	ANXA1	ILVALC*GGN	C343	NCF1	19.46	11.98	same	DISULFID
P04083	ANXA1	ILVALC*GGN	C343	NCF2	9.65	13.79	same	DISULFID
P04083	ANXA1	ILVALC*GGN	C343	NOXO1	0.00	0.00	same	DISULFID
P04406-1	GAPDH	IISNASC*TTNC*LAPLAK	C152;C156	NCF1	0.00	0.00	same	ACT_SITE; MOD_RES
P04406-1	GAPDH	IISNASC*TTNC*LAPLAK	C152;C156	NCF2	0.00	0.00	same	ACT_SITE; MOD_RES
P04406-1	GAPDH	IISNASC*TTNC*LAPLAK	C152;C156	NOXO1		9.69	same	ACT_SITE; MOD_RES
P04406-1	GAPDH	VPTANVSVVDLTC*R	C247	NCF1	27.86	19.60	same	MOD_RES; MOD_RES
P04406-1	GAPDH	VPTANVSVVDLTC*R	C247	NCF2	21.09	21.07	same	MOD_RES; MOD_RES
P04439	HLA-A	AYLDGTC*VEWLR	C188	NCF1	100.00	100.00	same	DISULFID
P04439	HLA-A	AYLDGTC*VEWLR	C188	NCF2	100.00	100.00	same	DISULFID
P04439	HLA-A	AYLDGTC*VEWLR	C188	NOXO1	100.00	100.00	same	DISULFID
P04843	RPN1	VAC*ITEQVLTLVNKR	C477	NCF2	0.00		same	
P04843	RPN1	TEGSDLC*DR	C545	NCF1	0.00	0.00	same	
P04843	RPN1	LKTEGSDLC*DR	C545	NCF1	28.76		same	
P04843	RPN1	TEGSDLC*DR	C545	NCF2	0.00	0.00	same	
P04843	RPN1	TEGSDLC*DR	C545	NOXO1	0.00	0.00	same	
P05023	ATP1A1	IISANGC*K	C211	NCF1	0.00	0.00	same	
P05023	ATP1A1	IISANGC*K	C211	NCF2	0.00	0.00	same	
P05023	ATP1A1	IISANGC*K	C211	NOXO1	0.00	0.00	same	
P05023	ATP1A1	NIAFFSTNC*VEGTAR	C249	NCF1	0.00	0.00	same	
P05023	ATP1A1	NIAFFSTNC*VEGTAR	C249	NCF2	4.97	0.00	same	
P05023	ATP1A1	NIAFFSTNC*VEGTAR	C249	NOXO1	0.00	0.00	same	
P05023	ATP1A1	IAGLC*NR	C428	NCF1	0.00	0.00	same	
P05023	ATP1A1	IAGLC*NR	C428	NCF2	0.00	0.00	same	
P05023	ATP1A1	IAGLC*NR	C428	NOXO1	0.00	0.00	same	
P05023	ATP1A1	C*IELC*C*GSVK	C459;C463; C464	NCF1	0.00	0.00	same	
P05023	ATP1A1	C*IELC*C*GSVK	C459;C463; C464	NOXO1	0.00		same	
P05023	ATP1A1	C*SSILLHGK	C518	NCF1	11.27		same	
P05023	ATP1A1	C*SSILLHGK	C518	NCF2	0.00	0.00	same	
P05023	ATP1A1	AC*VVHGSDLK	C663	NCF1	2.45	1.78	same	
P05023	ATP1A1	AC*VVHGSDLK	C663	NCF2	0.00	0.00	same	
P05023	ATP1A1	AC*VVHGSDLK	C663	NOXO1	0.00	0.00	same	
P05023	ATP1A1	LIIVEGC*QR	C705	NCF1	2.21	8.92	same	
P05023	ATP1A1	LIIVEGC*QR	C705	NOXO1	21.09		same	
P05109	S100A8	LLETEC*PQYIR	C42	NCF1	0.00	0.00	same	MOD_RES
P05109	S100A8	LLETEC*PQYIR	C42	NCF2	0.00	0.00	same	MOD_RES

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P05120	SERPINB2	YYSSEPQAVDFLEC*AEEAR	C161	NCF1	0.00	0.00	same	DISULFID
P05120	SERPINB2	YYSSEPQAVDFLEC*AEEAR	C161	NCF2	0.00	0.00	same	DISULFID
P05120	SERPINB2	YYSSEPQAVDFLEC*AEEAR	C161	NOXO1	0.00		same	DISULFID
P05120	SERPINB2	ITNC*ILFFGR	C405	NCF1	0.00	0.00	same	DISULFID
P05120	SERPINB2	ITNC*ILFFGR	C405	NCF2	2.77	2.07	same	DISULFID
P05120	SERPINB2	ITNC*ILFFGR	C405	NOXO1	0.00	0.00	same	DISULFID
P05141	SLC25A5	GTDIMYTGTLDC*WR	C257	NCF1	0.00	0.00	same	
P05141	SLC25A5	GTDIMYTGTLDC*WR	C257	NCF2	0.00	0.00	same	
P05141	SLC25A5	GTDIMYTGTLDC*WR	C257	NOXO1	0.00	0.00	same	
P05165-1	PCCA	MADEAVC*VGPAPTSK	C111	NCF1	6.45	5.34	same	
P05165-1	PCCA	MADEAVC*VGPAPTSK	C111	NCF2	100.00		same	
P05165-1	PCCA	MADEAVC*VGPAPTSK	C111	NCF2	12.16	11.90	same	
P05165-1	PCCA	MADEAVC*VGPAPTSK	C111	NOXO1		0.00	same	
P05165-1	PCCA	MADEAVC*VGPAPTSK	C111	NOXO1	6.97	4.82	same	
P05165-1	PCCA	C*LAAEDVVFIGPDTHAIQAMGDK	C155	NCF1	0.00		same	
P05165-1	PCCA	C*LAAEDVVFIGPDTHAIQAMGDK	C155	NCF2	0.00	6.40	same	
P05165-1	PCCA	C*LAAEDVVFIGPDTHAIQAMGDKIE SK	C155	NCF2	6.10	14.48	same	
P05165-1	PCCA	C*LAAEDVVFIGPDTHAIQAMGDK	C155	NOXO1	0.00	0.00	same	
P05165-1	PCCA	C*LAAEDVVFIGPDTHAIQAMGDKIE SK	C155	NOXO1	5.60	0.00	same	
P05165-1	PCCA	LQVEHPVTEC*ITGLDLVQEMIR	C363	NCF1	3.71	5.26	same	
P05165-1	PCCA	LQVEHPVTEC*ITGLDLVQEMIR	C363	NCF2	100.00		same	
P05165-1	PCCA	LQVEHPVTEC*ITGLDLVQEMIR	C363	NCF2	0.00	2.71	same	
P05165-1	PCCA	LQVEHPVTEC*ITGLDLVQEMIR	C363	NOXO1	1.81	2.23	same	
P05165-1	PCCA	INGWAVEC*R	C398	NCF1	14.65	4.48	same	
P05165-1	PCCA	INGWAVEC*R	C398	NCF2	4.33	8.73	same	
P05165-1	PCCA	INGWAVEC*R	C398	NOXO1	1.71	3.75	same	
P05165-1	PCCA	TVQC*LSR	C616	NCF1	15.88	16.86	same	
P05165-1	PCCA	TVQC*LSR	C616	NCF2	22.49	22.02	same	
P05165-1	PCCA	TVQC*LSR	C616	NOXO1	15.34	12.14	same	
P05165-1	PCCA	SVHC*QAGDTVGEGDLLVELE	C712	NCF2	0.00	0.00	same	
P05165-1	PCCA	SVHC*QAGDTVGEGDLLVELE	C712	NOXO1	0.00	0.00	same	
P05388	RPLP0	AGAIAPC*EVTVPAQNTGLGPEK	C119	NCF1	0.00		same	
P05556-1	ITGB1	FC*EC*DNFNC*DR	C553;C555; C560	NOXO1	100.00		same	DISULFID; DISULFID; DISULFID
P05556-1	ITGB1	SNGLIC*GGNGVC*K	C568;C574	NCF1		100.00	same	DISULFID; DISULFID
P05556-1	ITGB1	SNGLIC*GGNGVC*K	C568;C574	NOXO1	86.45		same	DISULFID; DISULFID
P05556-1	ITGB1	C*DDLEALK	C64	NCF1	100.00	100.00	same	DISULFID
P05556-1	ITGB1	C*DDLEALK	C64	NOXO1	100.00	100.00	same	DISULFID
P05556-1	ITGB1	C*DDLEALKK	C64	NOXO1	100.00	100.00	same	DISULFID
P06733-1	ENO1	SC*NC*LLLK	C337;C339	NCF1	0.00	0.00	same	
P06733-1	ENO1	VNQIGSVTESLQAC*K	C357	NCF1		0.00	same	
P07204	THBD	HIGTDC*DSGK	C480	NOXO1	100.00	100.00	same	DISULFID
P07339	CTSD	AIGAVPLIQGEYMIPC*EK	C329	NCF2	100.00	100.00	same	DISULFID
P07355	ANXA2	GLGTDEDSLIEIIC*SR	C133	NCF1	11.66	5.86	same	
P07355	ANXA2	GLGTDEDSLIEIIC*SR	C133	NCF2	6.26	11.42	same	
P07355	ANXA2	GLGTDEDSLIEIIC*SR	C133	NOXO1	4.57	1.12	same	
P07355	ANXA2	MSTVHEILC*K	C9	NCF1	16.00	16.74	same	
P07355	ANXA2	MSTVHEILC*K	C9	NCF2	20.52	17.64	same	
P07355	ANXA2	MSTVHEILC*K	C9	NOXO1	6.17	6.06	same	
P07737	PFN1	C*YEMASHLR	C128	NCF1	0.00	0.00	same	
P07737	PFN1	C*YEMASHLR	C128	NCF2	0.00	0.00	same	
P07737	PFN1	C*YEMASHLR	C128	NOXO1	0.00		same	
P07900	HSP90AA1	VFIMDNC*EELIPEYLNFIR	C374	NCF1	0.00	0.00	same	
P08238	HSP90AB1	VFIMDSC*DELIPEYLNFIR	C366	NCF1	0.00	0.00	same	
P08238	HSP90AB1	VFIMDSC*DELIPEYLNFIR	C366	NCF2	0.00	0.00	same	
P08238	HSP90AB1	C*LELFSELAEDKENYKK	C412	NCF1	0.00	0.00	same	
P08238	HSP90AB1	C*LELFSELAEDKENYKK	C412	NCF2	0.00	0.00	same	
P08240-1	SRPR SRPRA	NVAADIAVQLC*ESVANK	C363	NCF1	0.00	0.00	same	
P08240-1	SRPR SRPRA	NVAADIAVQLC*ESVANKLEGK	C363	NCF1	0.00	3.16	same	
P08574	CYC1	DVC*TFLR	C271	NCF1	0.00	0.00	same	
P08574	CYC1	DVC*TFLR	C271	NCF2	0.00	0.00	same	
P08574	CYC1	DVC*TFLR	C271	NOXO1	0.00	0.00	same	
P08670	VIM	QVQSLTC*EVDALK	C328	NCF2	100.00	100.00	same	
P08670	VIM	QVQSLTC*EVDALK	C328	NOXO1	100.00	100.00	same	
P08708	RPS17 RPS17L	VC*EEIAIIPSKK	C35	NCF1	0.00	0.00	same	
P08708	RPS17 RPS17L	VC*EEIAIIPSKK	C35	NCF2	0.00		same	
P08708	RPS17 RPS17L	VC*EEIAIIPSKK	C35	NOXO1	0.00		same	
P08865	RPSA	YVDIAIPC*NNK	C163	NCF1	0.00		same	
P08865	RPSA	YVDIAIPC*NNK	C163	NCF2	100.00	100.00	same	
P09651-1	HNRNPA1	YHTVNGHNC*EVR	C175	NCF1	10.93	17.21	same	
P09651-1	HNRNPA1	YHTVNGHNC*EVR	C175	NCF2	2.38	11.91	same	
P09651-1	HNRNPA1	YHTVNGHNC*EVR	C175	NOXO1	6.51	1.07	same	
P09651-1	HNRNPA1	SHFEQWGTLTDC*VVMR	C43	NCF1	0.00		same	
P09651-1	HNRNPA1	SHFEQWGTLTDC*VVMR	C43	NCF1	6.20	3.64	same	
P09651-1	HNRNPA1	SHFEQWGTLTDC*VVMR	C43	NCF2	0.00	0.00	same	
P09651-1	HNRNPA1	SHFEQWGTLTDC*VVMR	C43	NOXO1	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P09758	TACSTD2	TLVRPSEHALVDNDGLYDPDC*DPEGR	C108	NCF1		100.00	same	DISULFID
P09758	TACSTD2	TLVRPSEHALVDNDGLYDPDC*DPEGR	C108	NOXO1	100.00		same	DISULFID
P09758	TACSTD2	MTVC*SPDGPGGR	C44	NCF1	85.49	86.63	same	
P09758	TACSTD2	MTVC*SPDGPGGR	C44	NCF2	100.00	100.00	same	
P09758	TACSTD2	MTVC*SPDGPGGR	C44	NOXO1	100.00	100.00	same	
P09758	TACSTD2	ALGSGMAVDC*STLTSK	C66	NCF1	90.85	93.61	same	
P09758	TACSTD2	ALGSGMAVDC*STLTSK	C66	NCF2	65.88		same	
P09758	TACSTD2	ALGSGMAVDC*STLTSK	C66	NOXO1	78.49		same	
P09874	PARP1	WYHPGC*FVK	C162	NCF1	0.00	0.00	same	
P09874	PARP1	VC*STNDLK	C256	NCF1	0.00	0.00	same	
P09874	PARP1	VC*STNDLK	C256	NCF2	0.00	0.00	same	
P09874	PARP1	VADGMVFGALLPC*EEC*SGQLVFK	C295;C298	NCF1	0.00	0.00	same	
P09874	PARP1	SDAYYC*TGDVTAWTK	C311	NCF1	0.00		same	
P09874	PARP1	ASLC*ISTK	C429	NCF1	0.00	0.00	same	
P0DMV8	HSPA1B HSPA1A	C*QEVISWLDANTLAEKDEFEHK	C574	NCF1	12.27		same	
P0DMV8	HSPA1B HSPA1A	ELEQVC*NPIISGLYQGAGGPGPGGFGAQGPK	C603	NCF1	0.00	0.00	same	
P10599-1	TXN	C*MPTFQFFK	C73	NCF1	15.33		same	MOD_RES; DISULFID
P10599-1	TXN	C*MPTFQFFK	C73	NCF1	10.82	8.43	same	MOD_RES; DISULFID
P10599-1	TXN	C*MPTFQFFK	C73	NCF2	8.48	12.35	same	MOD_RES; DISULFID
P10809	HSPD1	C*EFQDAYVLLSEK	C237	NCF1	0.00	0.00	same	
P10809	HSPD1	C*EFQDAYVLLSEK	C237	NCF2	6.80	0.00	same	
P10809	HSPD1	C*EFQDAYVLLSEK	C237	NOXO1	0.00	0.00	same	
P10809	HSPD1	C*IPALDSLTPANEDQK	C447	NCF2	0.00		same	
P10809	HSPD1	C*IPALDSLTPANEDQK	C447	NOXO1		24.70	same	
P11142-1	HSPA8	VC*NPIITK	C603	NCF1	0.00	0.00	same	
P11142-1	HSPA8	VC*NPIITK	C603	NCF2	0.00		same	
P11171-1	EPB41	VSLLDDTVYEC*VVEK	C224	NCF1		100.00	same	
P11279	LAMP1	C*NAEEHVR	C338	NCF2	100.00	100.00	same	DISULFID
P11279	LAMP1	C*NAEEHVR	C338	NOXO1	100.00	100.00	same	DISULFID
P11388-1	TOP2A	SFGSTC*QLSEK	C392	NCF1	0.00		same	
P11498	PC	ADFAQAC*QDAGVR	C131	NCF1	18.00	18.61	same	
P11498	PC	ADFAQAC*QDAGVR	C131	NCF2	18.57	19.24	same	
P11498	PC	ADFAQAC*QDAGVR	C131	NOXO1	9.47	8.51	same	
P11498	PC	INGC*AIQC*R	C372;C376	NCF1	2.84	4.06	same	
P11498	PC	INGC*AIQC*R	C372;C376	NCF2	6.09	6.41	same	
P11498	PC	INGC*AIQC*R	C372;C376	NOXO1	1.98	1.83	same	
P11498	PC	AC*TELGIR	C56	NCF1	10.49	12.13	same	
P11498	PC	AC*TELGIR	C56	NCF2	12.84	14.46	same	
P11498	PC	AC*TELGIR	C56	NOXO1	8.08	5.31	same	
P11498	PC	FLYEC*PWR	C622	NCF1	13.49	12.85	same	
P11498	PC	FLYEC*PWR	C622	NCF2	10.16	11.21	same	
P11498	PC	FLYEC*PWR	C622	NOXO1	5.35	3.61	same	
P11498	PC	AGTHILC*IK	C739	NCF1	9.98	10.40	same	
P11498	PC	AGTHILC*IK	C739	NCF2	10.49	12.40	same	
P11498	PC	AGTHILC*IK	C739	NOXO1	7.68	6.75	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NCF1	7.12		same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NCF1	10.86	10.16	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NCF1	14.16	5.70	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NCF2	16.47	9.85	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NCF2	13.54	9.14	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NOXO1	7.87	7.97	same	
P11498	PC	DMAGLLKPTAC*TMLVSSLR	C752	NOXO1	3.24	2.81	same	
P11498	PC	GLYAAFDC*TATMK	C850	NCF1	17.83	14.04	same	
P11498	PC	GLYAAFDC*TATMK	C850	NCF2	18.88	20.26	same	
P11498	PC	GLYAAFDC*TATMK	C850	NOXO1	100.00	100.00	same	
P11498	PC	GLYAAFDC*TATMK	C850	NOXO1	7.06	6.79	same	
P12277	CKB	FC*TGLTQIETLFK	C254	NCF1		0.00	same	
P13010	XRCC5	C*FSVLGFC*K	C339;C346	NCF1	0.00	0.00	same	
P13010	XRCC5	C*FSVLGFC*K	C339;C346	NCF2	0.00		same	
P13010	XRCC5	LFQC*LLHR	C493	NCF1	0.00	4.63	same	
P13639	EEF2	TFC*QLILDPIFK	C290	NCF1	0.00	0.00	same	
P13639	EEF2	STLTDSLVC*K	C41	NCF1	0.00	0.00	same	
P13639	EEF2	STLTDSLVC*K	C41	NCF2	0.00	0.00	same	
P13639	EEF2	STLTDSLVC*K	C41	NOXO1	0.00	0.00	same	
P13639	EEF2	YVEPIEDVPC*GNIVGLVGVDQFLVK	C466	NCF1	6.30	7.62	same	
P13639	EEF2	DLEEDHAC*IPIK	C567	NCF2	100.00		same	
P13639	EEF2	DLEEDHAC*IPIK	C567	NOXO1		100.00	same	
P13639	EEF2	ETVSEESNVLC*LSK	C591	NCF1		0.00	same	
P13639	EEF2	IWC*FGPDGTGPNILTDITK	C651	NCF1	0.00		same	
P13639	EEF2	EGALC*EENMR	C693	NCF1	0.00	0.00	same	
P13639	EEF2	EGALC*EENMR	C693	NCF2	0.00	0.00	same	
P13639	EEF2	EGALC*EENMR	C693	NOXO1	0.00	0.00	same	
P13639	EEF2	C*LYASVLTAQPR	C728	NCF1	0.00	0.00	same	
P13639	EEF2	C*LYASVLTAQPR	C728	NOXO1		0.00	same	
P13639	EEF2	LMEPIYLVEIQC*PEQVVGGIYGVLNR	C751	NCF1		0.00	same	
P13645	KRT10	AETEC*QNTEYQQLLDIK	C427	NCF1	100.00	100.00	same	
P13645	KRT10	AETEC*QNTEYQQLLDIK	C427	NCF2		100.00	same	
P13646	KRT13	AGLENTVAETEC*R	C354	NCF1	19.46	19.95	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P13646	KRT13	AGLENTVAETEC*R	C354	NCF2	19.04		same	
P13646	KRT13	AGLENTVAETEC*R	C354	NOXO1	25.07	32.94	same	
P13646	KRT13	SEMEC*QNQEYK	C383	NCF1	0.00	0.00	same	
P13646	KRT13	SEMEC*QNQEYK	C383	NCF2	0.00	0.00	same	
P13646	KRT13	SEMEC*QNQEYK	C383	NOXO1	0.00	0.00	same	
P13647	KRT5	KQC*ANLQNAIADAEQR	C407	NCF1	0.00	0.00	same	
P13647	KRT5	QC*ANLQNAIADAEQR	C407	NCF1	3.13	1.32	same	
P13647	KRT5	KQC*ANLQNAIADAEQR	C407	NCF2	0.00	0.00	same	
P13647	KRT5	QC*ANLQNAIADAEQR	C407	NCF2	8.77	10.39	same	
P13647	KRT5	KQC*ANLQNAIADAEQR	C407	NOXO1	0.00		same	
P13647	KRT5	QC*ANLQNAIADAEQR	C407	NOXO1	5.54	8.32	same	
P13647	KRT5	VSLAGAC*GVGGYGSR	C55	NCF1	15.77	13.32	same	
P13647	KRT5	VSLAGAC*GVGGYGSR	C55	NCF2	21.21	27.97	same	
P13647	KRT5	VSLAGAC*GVGGYGSR	C55	NOXO1	21.51	20.34	same	
P13804-1	ETFA	TIYAGNALC*TVK	C155	NCF1	0.00	0.00	same	
P13804-1	ETFA	TIYAGNALC*TVK	C155	NCF2	0.00	0.00	same	
P13804-1	ETFA	TIYAGNALC*TVK	C155	NOXO1	0.00	0.00	same	
P13987	CD59	FEHC*NFNDVTTR	C70	NCF1	100.00	100.00	same	DISULFID
P13987	CD59	FEHC*NFNDVTTR	C70	NCF2	100.00	100.00	same	DISULFID
P13987	CD59	FEHC*NFNDVTTR	C70	NOXO1	100.00	100.00	same	DISULFID
P14174	MIF	LLC*GLLAER	C81	NCF1	0.00	0.00	same	
P14598	NCF1	C*PHLLDFFK	C111	NCF2	100.00		same	
P14598	NCF1	C*PHLLDFFK	C111	NOXO1	100.00	100.00	same	
P14598	NCF1	SESGWWFC*QMK	C196	NCF1	4.20	4.25	same	
P14598	NCF1	SESGWWFC*QMK	C196	NCF1	9.60	8.98	same	
P14618	PKM	C*DENILWLDYK	C152	NCF1	100.00	100.00	same	
P14618	PKM	C*DENILWLDYK	C152	NCF2	100.00	100.00	same	
P14618	PKM	AGKPVIC*ATQMLESMIK	C326	NCF1		0.00	same	
P14618	PKM	AGKPVIC*ATQMLESMIK	C326	NCF2		63.06	same	
P14618	PKM	AGKPVIC*ATQMLESMIK	C326	NOXO1	0.00		same	
P14618	PKM	C*C*SGAIIVLTK	C423;C424	NCF1	0.00	0.00	same	
P14618	PKM	C*C*SGAIIVLTK	C423;C424	NCF2	0.00	0.00	same	
P14618	PKM	C*C*SGAIIVLTK	C423;C424	NOXO1	0.00	0.00	same	
P14618	PKM	GIFPVLC*K	C474	NCF1	0.00	0.00	same	
P14618	PKM	GIFPVLC*K	C474	NCF2	0.00	0.00	same	
P14618	PKM	GIFPVLC*K	C474	NOXO1		0.00	same	
P14618	PKM	NTGIIC*TIGPASR	C49	NCF1	12.28	8.97	same	
P14618	PKM	NTGIIC*TIGPASR	C49	NOXO1	10.08	10.97	same	
P14678-1	SNRPB	C*ILQDGR	C19	NCF1	0.00	0.00	same	
P14678-1	SNRPB	C*ILQDGR	C19	NCF2	0.00	0.00	same	
P14678-1	SNRPB	C*ILQDGR	C19	NOXO1	0.00	0.00	same	
P14866	HNRNPL	VFNVFC*LYGNVEK	C404	NCF1	0.00	0.00	same	
P14866	HNRNPL	VFNVFC*LYGNVEK	C404	NCF2	0.00		same	
P14866	HNRNPL	VFNVFC*LYGNVEK	C404	NOXO1	0.00		same	
P14866	HNRNPL	LC*FSTAQHAS	C581	NCF2	69.08	65.97	same	
P14923	JUP	C*TTSILHNLSHHR	C204	NCF1	8.06		same	
P14923	JUP	C*TTSILHNLSHHR	C204	NCF2	16.83	21.68	same	
P14923	JUP	C*TTSILHNLSHHR	C204	NOXO1	0.00	2.61	same	
P14923	JUP	VLSVC*PSNKPAIVEAGGMQALGK	C341	NCF1	0.00	0.00	same	
P14923	JUP	VLSVC*PSNKPAIVEAGGMQALGK	C341	NCF2	0.00	0.00	same	
P14923	JUP	VLSVC*PSNKPAIVEAGGMQALGK	C341	NOXO1	7.42	3.23	same	
P14923	JUP	ILVNQLSVDDVNVLTC*ATGTLSNLT C*NNSK	C410;C420	NCF1		0.00	same	
P14923	JUP	ILVNQLSVDDVNVLTC*ATGTLSNLT C*NNSK	C410;C420	NOXO1	0.00		same	
P14923	JUP	AGDKDDITEPAVC*ALR	C457	NCF1	10.00	9.70	same	
P14923	JUP	AGDKDDITEPAVC*ALR	C457	NCF2	11.79	11.63	same	
P14923	JUP	AGDKDDITEPAVC*ALR	C457	NOXO1	9.56	7.45	same	
P14923	JUP	GIMEEDEAC*GR	C49	NCF1	13.73	11.62	same	
P14923	JUP	GIMEEDEAC*GR	C49	NCF2	22.24	22.74	same	
P14923	JUP	GIMEEDEAC*GR	C49	NOXO1		36.28	same	
P14923	JUP	NLALC*PANHAPLQEAAVIPR	C511	NCF1	5.76	8.58	same	
P14923	JUP	NLALC*PANHAPLQEAAVIPR	C511	NCF2	11.02	9.57	same	
P14923	JUP	NLALC*PANHAPLQEAAVIPR	C511	NOXO1	7.56	8.83	same	
P14923	JUP	VAAGVLC*ELAQDKEAADAIDAEGAS APLMELLHSR	C609	NCF1	29.03		same	
P14923	JUP	VAAGVLC*ELAQDKEAADAIDAEGAS APLMELLHSR	C609	NOXO1		10.19	same	
P14923	JUP	EAMC*PGVSGEDSSLLLATQVEGQAT NLQR	C90	NCF2	20.47	13.25	same	
P14923	JUP	EAMC*PGVSGEDSSLLLATQVEGQAT NLQR	C90	NOXO1	7.30	0.00	same	
P15170-3	GSPT1	LESGSIC*K	C465	NCF1	12.28	10.66	same	
P15170-3	GSPT1	LESGSIC*K	C465	NCF2	13.93	11.48	same	
P15170-3	GSPT1	GIEEEEILPGFILC*DPNNLC*HSGR	C519;C525	NCF1	0.00	0.00	same	
P15170-3	GSPT1	LKGIEEEEILPGFILC*DPNNLC*HS GR	C519;C525	NCF1	0.00	0.00	same	
P15311	EZR	EGILSDEIYC*PPETAVLLGSYAVQA K	C117	NCF1	0.00	0.00	same	
P15311	EZR	EGILSDEIYC*PPETAVLLGSYAVQA K	C117	NCF2	0.00		same	
P15311	EZR	ILQLC*MGNHELYMR	C284	NCF1	0.00	0.00	same	
P15311	EZR	ILQLC*MGNHELYMR	C284	NCF2	0.00	0.00	same	
P15880	RPS2	IGKPHTVPC*K	C182	NCF1	2.34	4.34	same	
P15880	RPS2	IGKPHTVPC*K	C182	NOXO1		0.00	same	
P15880	RPS2	C*GSVLVR	C188	NCF1	0.00	0.00	same	
P15880	RPS2	C*GSVLVR	C188	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P15880	RPS2	C*GSVLVR	C188	NOXO1	0.00	0.00	same	
P15880	RPS2	LLMMAGIDDC*YTSAR	C222	NCF1	0.00	0.00	same	
P15880	RPS2	LLMMAGIDDC*YTSAR	C222	NCF2		0.00	same	
P15880	RPS2	LLMMAGIDDC*YTSAR	C222	NOXO1	0.00		same	
P15880	RPS2	GC*TATLGNFAK	C229	NCF1	3.75	1.95	same	
P15880	RPS2	GC*TATLGNFAK	C229	NCF2	0.00	0.00	same	
P15880	RPS2	GC*TATLGNFAK	C229	NOXO1	0.60	0.36	same	
P15924-1	DSP	ELDEC*FAQANDQMEILDSLIR	C110	NOXO1	0.00	0.00	same	
P15924-1	DSP	QMGQPC*DAYQK	C135	NOXO1	0.00		same	
P15924-1	DSP	LLEAQAC*TGGIIHPTTGQK	C2656	NCF1		0.00	same	
P15924-1	DSP	LLEAQAC*TGGIIHPTTGQK	C2656	NCF2	0.00		same	
P15924-1	DSP	MSC*NGGSHPR	C3	NOXO1	0.00	0.00	same	
P16144-2	ITGB4	VPSVELTNLYPYC*DYEMK	C1190	NOXO1	0.00	0.00	same	
P16144-2	ITGB4	C*ERPLQGYSVEYQLLNGGELHR	C1483	NCF1	0.00	0.00	same	
P16144-2	ITGB4	C*ERPLQGYSVEYQLLNGGELHR	C1483	NCF2	0.00	0.00	same	
P16144-2	ITGB4	C*ERPLQGYSVEYQLLNGGELHR	C1483	NOXO1	0.00	0.00	same	
P16144-2	ITGB4	RPNGDIVGYLVTC*EMAQGGGPATAFR	C1608	NCF2	0.00	0.00	same	
P16144-2	ITGB4	RPNGDIVGYLVTC*EMAQGGGPATAFR	C1608	NOXO1	0.00	0.00	same	
P16144-2	ITGB4	ISGNLDAPEGGFDAILQTAVC*TR	C245	NOXO1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	C*HLDTTGTYTQYR	C288	NCF1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	C*HLDTTGTYTQYR	C288	NCF2	100.00	100.00	same	DISULFID
P16144-2	ITGB4	C*HLDTTGTYTQYR	C288	NOXO1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	ALEHVDGTHVC*QLPEDQK	C424	NCF2		100.00	same	DISULFID
P16144-2	ITGB4	ALEHVDGTHVC*QLPEDQK	C424	NOXO1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	EGEDKPC*SGR	C512	NCF1		60.63	same	DISULFID
P16144-2	ITGB4	EGEDKPC*SGR	C512	NCF2	92.71	83.71	same	DISULFID
P16144-2	ITGB4	EGEDKPC*SGR	C512	NOXO1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	TSGFLC*NDR	C551	NCF1		77.23	same	DISULFID
P16144-2	ITGB4	TSGFLC*NDR	C551	NCF2	100.00	100.00	same	DISULFID
P16144-2	ITGB4	TSGFLC*NDR	C551	NOXO1	100.00	100.00	same	DISULFID
P16144-2	ITGB4	C*NTQAELLAAGC*QR	C61;C72	NCF1	100.00	100.00	same	DISULFID; DISULFID
P16144-2	ITGB4	C*NTQAELLAAGC*QR	C61;C72	NCF2	100.00	100.00	same	DISULFID; DISULFID
P16144-2	ITGB4	C*NTQAELLAAGC*QR	C61;C72	NOXO1	100.00	100.00	same	DISULFID; DISULFID
P16144-2	ITGB4	TC*EEC*NFK	C648;C651	NCF2	100.00	100.00	same	DISULFID; DISULFID
P16144-2	ITGB4	TC*EEC*NFK	C648;C651	NOXO1	100.00	100.00	same	DISULFID; DISULFID
P16144-2	ITGB4	LC*TENLLKPDTR	C828	NCF1	0.00	0.00	same	
P16144-2	ITGB4	LC*TENLLKPDTR	C828	NCF2	0.00	0.00	same	
P16144-2	ITGB4	LC*TENLLKPDTR	C828	NOXO1	0.00	0.00	same	
P16591-1	FER	LIGVC*TQR	C627	NCF2		100.00	same	
P16591-1	FER	NC*LVGENNVLK	C690	NCF1	0.00	0.00	same	
P16591-1	FER	NC*LVGENNVLK	C690	NCF2	0.00	0.00	same	
P16591-1	FER	NC*LVGENNVLK	C690	NOXO1	0.00	0.00	same	
P16615	ATP2A2	TGTLTTNQMSVC*R	C364	NCF1	0.00	0.00	same	
P16615	ATP2A2	TGTLTTNQMSVC*R	C364	NCF2	0.00	0.00	same	
P16615	ATP2A2	TGTLTTNQMSVC*R	C364	NOXO1	0.00	0.00	same	
P16615	ATP2A2	VGEATETALTC*LVEK	C447	NCF1	0.00	0.00	same	
P16615	ATP2A2	VGEATETALTC*LVEK	C447	NCF2	0.00	0.00	same	
P16615	ATP2A2	VGEATETALTC*LVEK	C447	NOXO1	0.00	0.00	same	
P16615	ATP2A2	ANAC*NSVIK	C471	NCF1	0.00	0.00	same	
P16615	ATP2A2	ANAC*NSVIK	C471	NCF2	0.00	0.00	same	
P16615	ATP2A2	ANAC*NSVIK	C471	NOXO1	0.00	0.00	same	
P16615	ATP2A2	C*LALATHDNPLRR	C560	NOXO1	19.32		same	
P16615	ATP2A2	DAC*LNAR	C669	NCF1	0.00	0.00	same	
P16615	ATP2A2	DAC*LNAR	C669	NCF2	0.00	0.00	same	
P16615	ATP2A2	DAC*LNAR	C669	NOXO1	0.00	0.00	same	
P17844	DDX5	ELAQQVQQVAAEYC*R	C191	NCF1	0.00	0.00	same	
P17844	DDX5	ELAQQVQQVAAEYC*R	C191	NCF2	0.00	0.00	same	
P17844	DDX5	ELAQQVQQVAAEYC*R	C191	NOXO1	0.00	0.00	same	
P17844	DDX5	LIDFLEC*GK	C234	NCF1	2.31	0.00	same	
P17844	DDX5	LIDFLEC*GK	C234	NCF2	0.00	0.00	same	
P17844	DDX5	LIDFLEC*GK	C234	NOXO1	0.00	0.00	same	
P17987	TCP1	IC*DDELILIK	C357	NCF1	0.00	0.00	same	
P18124	RPL7	YGIIC*MEDLIHEIYTVGK	C186	NCF1	0.00		same	
P18583-5	SON	ANAAAMC*AK	C2070	NCF1	0.00	0.00	same	
P18583-5	SON	ANAAAMC*AK	C2070	NCF2	0.00	0.00	same	
P18583-5	SON	ANAAAMC*AK	C2070	NOXO1	0.00	0.00	same	
P18583-5	SON	C*VSVQTDPTDEIPTKK	C92	NCF1	0.00	0.00	same	
P18583-5	SON	C*VSVQTDPTDEIPTKK	C92	NCF2		0.00	same	
P18583-5	SON	C*VSVQTDPTDEIPTKK	C92	NOXO1		0.00	same	
P19012	KRT15	AGLENSLAETEC*R	C355	NCF1		25.00	same	
P19012	KRT15	AGLENSLAETEC*R	C355	NCF2	29.03		same	
P19012	KRT15	AGLENSLAETEC*R	C355	NOXO1	31.62	28.18	same	
P19012	KRT15	C*EMEAQNQEYK	C380	NCF1	0.00	0.00	same	
P19012	KRT15	C*EMEAQNQEYK	C380	NCF2	0.00	0.00	same	
P19012	KRT15	C*EMEAQNQEYK	C380	NOXO1	0.00	0.00	same	
P19338	NCL	EALNSC*NK	C543	NCF1	21.76		same	
P19338	NCL	EALNSC*NKR	C543	NCF1	100.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P19338	NCL	EALNSC*NK	C543	NOXO1	86.65		same	
P19878	NCF2	AMEC*VWK	C165	NCF2	12.21	14.17	same	
P19878	NCF2	GLVPC*NYLEPVELR	C291	NCF2	19.69	21.09	same	
P19878	NCF2	IC*FNIGC*MYTILK	C40;C45	NCF2	0.00	0.00	same	
P19878	NCF2	VFVEDC*ATTDLESTR	C514	NCF2	2.13	6.24	same	
P19878	NCF2	VFVEDC*ATTDLESTRR	C514	NCF2	19.25	21.08	same	
P19878	NCF2	VFVEDC*ATTDLESTR	C514	NOXO1	0.00	0.00	same	
P20700	LMNB1	C*QSLTEDLEFR	C198	NCF1	0.00		same	
P20700	LMNB1	C*QSLTEDLEFR	C198	NCF2	0.00	0.00	same	
P20700	LMNB1	C*QSLTEDLEFR	C198	NOXO1	0.00		same	
P21333	FLNA	IVGPSGAAVPC*K	C1018	NOXO1	100.00	100.00	same	
P21333	FLNA	AHVVPC*FDASK	C1157	NOXO1	100.00	100.00	same	
P21333	FLNA	C*SGPGLER	C1165	NCF1	26.40		same	
P21333	FLNA	C*SGPGLER	C1165	NCF2	11.40	17.44	same	
P21333	FLNA	LQVEPAVDTSGVQC*YGPGIEGQGVFR	C1260	NCF1	8.14	7.45	same	
P21333	FLNA	LQVEPAVDTSGVQC*YGPGIEGQGVFR	C1260	NCF2	0.00	0.00	same	
P21333	FLNA	ALGALVDSC*APGLC*PDWDSWDASKPVTNAR	C205;C210	NCF1	0.00	0.00	same	
P21333	FLNA	ATC*APQHGAPGPGPADASK	C2543	NCF1	0.00	0.00	same	
P21333	FLNA	ATC*APQHGAPGPGPADASK	C2543	NCF2	0.00	0.00	same	
P21333	FLNA	C*SYQPTMEGVHTVHVTFAGVPIPR	C444	NCF1	16.22	7.08	same	
P21333	FLNA	C*SYQPTMEGVHTVHVTFAGVPIPR	C444	NCF2	0.00	0.00	same	
P21333	FLNA	SPYTVTVGQAC*NPSAC*R	C478;C483	NCF1	0.00	0.00	same	
P21333	FLNA	SPYTVTVGQAC*NPSAC*R	C478;C483	NCF2	0.00		same	
P21333	FLNA	SPYTVTVGQAC*NPSAC*R	C478;C483	NOXO1	0.00	0.00	same	
P21333	FLNA	VGTEC*GNQK	C574	NCF1	21.02	16.97	same	
P21333	FLNA	VGTEC*GNQK	C574	NCF2	15.37	15.46	same	
P21333	FLNA	YWPQEAGEYAVHVLC*NSEDIR	C649	NCF2		0.00	same	
P21333	FLNA	VQVQDNEGC*PVEALVK	C717	NCF1	8.98	5.29	same	
P21333	FLNA	VQVQDNEGC*PVEALVK	C717	NCF2	0.00	0.00	same	
P21333	FLNA	C*APGVVGPAEADIDFDIIR	C810	NCF1	5.26	5.63	same	
P21333	FLNA	C*APGVVGPAEADIDFDIIR	C810	NCF2	0.00	0.00	same	
P21796	VDAC1	YQIDPDAC*FSAK	C232	NCF1	0.00		same	
P21796	VDAC1	YQIDPDAC*FSAK	C232	NCF2	0.00	0.00	same	
P21796	VDAC1	YQIDPDAC*FSAK	C232	NOXO1	0.00		same	
P21964-1	COMT	YLPDTLLLEEC*GLLR	C207	NCF1	0.00		same	
P22314	UBA1	VGEFC*HNR	C179	NCF1	0.00	0.00	same	
P22528	SPRR1B	TKEPC*HPK	C41	NCF2	59.90		same	
P22528	SPRR1B	TKEPC*HPK	C41	NOXO1	56.40		same	
P22528	SPRR1B	VPEPC*HPK	C49	NCF1	0.00	0.00	same	
P22528	SPRR1B	VPEPC*HPK	C49	NCF2	0.00		same	
P22528	SPRR1B	VPEPC*HPK	C49	NOXO1	0.00	0.00	same	
P22528	SPRR1B	VPEPC*QPK	C57	NOXO1	25.77	24.55	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NCF1	0.00	0.00	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NCF1	7.90	1.77	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NCF2	0.00	0.00	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NCF2	2.21	7.20	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NOXO1	0.00	0.00	same	
P22626	HNRNPA-2B1	LTDC*VVMR	C50	NOXO1	0.00	0.00	same	
P22695	UQCRC2	NALANPLYC*PDYR	C192	NCF2	0.00	0.00	same	
P22695	UQCRC2	NALANPLYC*PDYR	C192	NOXO1	0.00	0.00	same	
P23284	PPIB	DVIIADC*GK	C202	NCF2	100.00		same	MOD_RES
P23284	PPIB	DVIIADC*GK	C202	NOXO1	15.27		same	MOD_RES
P23396-1	RPS3	AC*YGVLR	C119	NCF1	4.44	4.87	same	
P23396-1	RPS3	AC*YGVLR	C119	NCF2		23.00	same	
P23396-1	RPS3	GLC*AIAQAESLR	C97	NCF1	0.00	0.00	same	
P23396-1	RPS3	GLC*AIAQAESLR	C97	NCF2	0.00	0.00	same	
P23396-1	RPS3	GLC*AIAQAESLR	C97	NOXO1	0.00	0.00	same	
P23526-1	AHCY	SKFDNLYGC*R	C195	NCF1	0.00	0.00	same	
P23526-1	AHCY	SKFDNLYGC*R	C195	NCF2	0.00	0.00	same	
P23526-1	AHCY	SKFDNLYGC*R	C195	NOXO1	0.00	0.00	same	
P23634-1	ATP2B4	EGDFGC*TVMELR	C24	NOXO1	0.00	0.00	same	
P25398	RPS12	LVEALC*AEHQINLIK	C69	NCF1	0.00	0.00	same	
P25398	RPS12	LGEWVGLC*K	C92	NCF1	0.00	0.00	same	
P26006	ITGA3	SETVLTC*ATGR	C904	NCF1	100.00	100.00	same	DISULFID
P26006	ITGA3	SETVLTC*ATGR	C904	NCF2		100.00	same	DISULFID
P26006	ITGA3	SETVLTC*ATGR	C904	NOXO1	100.00	100.00	same	DISULFID
P26368	U2AF2	YC*DPDSYHR	C464	NCF1	0.00	0.00	same	
P26368	U2AF2	YC*DPDSYHR	C464	NCF2		0.00	same	
P26368	U2AF2	YC*DPDSYHR	C464	NOXO1	0.00		same	
P27635	RPL10	MLSC*AGADR	C105	NCF1	0.00		same	
P27635	RPL10	MLSC*AGADR	C105	NCF1	3.75	4.47	same	
P27635	RPL10	MLSC*AGADR	C105	NCF2	25.48	31.80	same	
P27635	RPL10	MLSC*AGADR	C105	NOXO1	7.55	0.00	same	
P27635	RPL10	VDEFPLC*GHMVSDEYEQLSSEALEAAR	C49	NCF1		0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P27694	RPA1	DSLVDIIGIC*K	C323	NCF1	0.00		same	
P27695	APEX1	IC*SWNVDGLR	C65	NCF1	0.00	0.00	same	MOD_RES; DISULFID
P27695	APEX1	IC*SWNVDGLR	C65	NCF2	0.00	0.00	same	MOD_RES; DISULFID
P27695	APEX1	IC*SWNVDGLR	C65	NOXO1	0.00	0.00	same	MOD_RES; DISULFID
P27797	CALR	C*KDDEFTHLYTLIVRPDNTYEVK	C163	NCF1		0.00	same	
P27816-1	MAP4	C*SLPAEEDSVLEK	C635	NCF2	100.00	100.00	same	
P27816-1	MAP4	C*SLPAEEDSVLEK	C635	NOXO1	100.00	100.00	same	
P29034	S100A2	YSC*QEGDKFK	C22	NOXO1	0.00	0.00	same	
P29317	EPHA2	NILVNSNLVC*K	C753	NCF1	0.00	0.00	same	
P29317	EPHA2	NILVNSNLVC*K	C753	NCF2	0.00	0.00	same	
P29317	EPHA2	NILVNSNLVC*K	C753	NOXO1	9.21	0.00	same	
P29590	PML	C*FEAHQWFLK	C151	NCF1	0.00	0.00	same	METAL
P29590	PML	TPTLTSIYC*R	C204	NOXO1	0.00	0.00	same	
P29590	PML	C*DISAEIQQR	C227	NCF1	0.00	0.00	same	
P30050-1	RPL12	EILGTAQSVGC*NVDGR	C141	NCF1	0.00	0.00	same	
P30050-1	RPL12	EILGTAQSVGC*NVDGR	C141	NCF2	0.00	0.00	same	
P30050-1	RPL12	EILGTAQSVGC*NVDGR	C141	NOXO1	0.00	0.00	same	
P30050-1	RPL12	C*TGGEVGATSALAPK	C17	NCF2	0.00	0.00	same	
P30101	PDIA3	VDC*TANTNTC*NK	C85;C92	NOXO1	100.00		same	DISULFID; DISULFID
P30536	TSPO	MAPPWVPAMGFTLAPSLGC*FVGSR	C19	NCF1	2.47	0.00	same	
P30536	TSPO	MAPPWVPAMGFTLAPSLGC*FVGSR	C19	NCF2	0.00	0.00	same	
P30536	TSPO	MAPPWVPAMGFTLAPSLGC*FVGSR	C19	NOXO1	0.00		same	
P31040	SDHA	VGSVLQEGC*GK	C536	NCF1	0.00		same	
P31949	S100A11	C*IESLIAVFQK	C13	NCF1	4.66	2.25	same	DISULFID
P31949	S100A11	C*IESLIAVFQK	C13	NCF2	0.60	2.97	same	DISULFID
P31949	S100A11	C*IESLIAVFQK	C13	NOXO1	0.00	0.00	same	DISULFID
P32969	RPL9	MRPGVAC*SVSQAQK	C134	NCF1		0.00	same	
P32969	RPL9	MRPGVAC*SVSQAQK	C134	NCF2	0.00		same	
P32969	RPL9	MRPGVAC*SVSQAQK	C134	NOXO1	0.00		same	
P32969	RPL9	TIC*SHVQNMIK	C74	NCF1	0.00	0.00	same	
P32969	RPL9	TIC*SHVQNMIK	C74	NCF2	0.00	0.00	same	
P32969	RPL9	TIC*SHVQNMIK	C74	NOXO1	0.00	0.00	same	
P33240	CSTF2	LC*VQNSPQEAR	C150	NCF1	0.00	0.00	same	
P33240	CSTF2	LC*VQNSPQEAR	C150	NCF2		0.00	same	
P33240	CSTF2	LC*VQNSPQEAR	C150	NOXO1	0.00		same	
P33947-1	KDELR2	SC*AGISGK	C29	NCF1	11.58	11.91	same	
P33947-1	KDELR2	SC*AGISGK	C29	NCF2	7.94	3.83	same	
P33947-1	KDELR2	SC*AGISGK	C29	NOXO1	8.56	5.92	same	
P34897-1	SHMT2	MREVC*DEVK	C241	NCF1	0.00		same	
P34897-1	SHMT2	MREVC*DEVK	C241	NCF2	0.00	0.00	same	
P34897-1	SHMT2	MREVC*DEVK	C241	NOXO1	0.00	0.00	same	
P34897-1	SHMT2	AALEALGSC*LNNK	C91	NOXO1	0.00		same	
P35221	CTNNA1	AAAGEFADDPC*SSVKR	C116	NCF1	100.00	100.00	same	
P35221	CTNNA1	AAAGEFADDPC*SSVKR	C116	NCF2	100.00	100.00	same	
P35221	CTNNA1	NVPILYTASQAC*LQHPDVAAYK	C228	NCF1	0.00	0.00	same	
P35221	CTNNA1	NVPILYTASQAC*LQHPDVAAYK	C228	NCF2	0.00	0.00	same	
P35221	CTNNA1	NVPILYTASQAC*LQHPDVAAYK	C228	NOXO1	0.00	0.00	same	
P35221	CTNNA1	LIEVANLAC*SISNNEEGVK	C438	NCF1	4.47	0.00	same	
P35221	CTNNA1	LIEVANLAC*SISNNEEGVK	C438	NCF2	38.42		same	
P35221	CTNNA1	LIEVANLAC*SISNNEEGVK	C438	NOXO1	15.22	9.74	same	
P35221	CTNNA1	MSASQLEALC*PQVINAALALAAKPQSK	C461	NCF1	0.00	0.00	same	
P35221	CTNNA1	MSASQLEALC*PQVINAALALAAKPQSK	C461	NCF2	0.00	0.00	same	
P35221	CTNNA1	C*VIALQEK	C526	NCF1	0.00	0.00	same	
P35221	CTNNA1	C*VIALQEK	C526	NOXO1	0.00		same	
P35222	CTNNB1	C*TAGTLHNLSHHR	C213	NCF1	0.00	0.00	same	
P35222	CTNNB1	C*TAGTLHNLSHHR	C213	NCF2	0.00	0.00	same	
P35222	CTNNB1	C*TAGTLHNLSHHR	C213	NOXO1	0.00	0.00	same	
P35222	CTNNB1	AGDREDITEPAIC*ALR	C466	NCF1	0.00	0.00	same	
P35222	CTNNB1	AGDREDITEPAIC*ALR	C466	NCF2	0.00	0.00	same	
P35222	CTNNB1	AGDREDITEPAIC*ALR	C466	NOXO1	0.00	0.00	same	
P35222	CTNNB1	NLALC*PANHAPLR	C520	NCF1	0.00	0.00	same	
P35222	CTNNB1	NLALC*PANHAPLR	C520	NCF2	0.00	0.00	same	
P35222	CTNNB1	NLALC*PANHAPLR	C520	NOXO1	0.00	0.00	same	
P35268	RPL22	FTLDC*THPVEDGIMDAANFEQFLQER	C25	NCF1	0.00		same	
P35579-1	MYH9	ADFC*IIHYAGK	C569	NCF1	0.00	0.00	same	
P35579-1	MYH9	C*IIPNHEK	C671	NCF1	0.00	8.78	same	
P35579-1	MYH9	C*IIPNHEK	C671	NCF2	24.35	16.36	same	
P35579-1	MYH9	VEDMAELTC*LNEASVLHNLK	C91	NCF1	0.00	0.00	same	
P35579-1	MYH9	VEDMAELTC*LNEASVLHNLK	C91	NCF2	0.00	0.00	same	
P35579-1	MYH9	KQELEEIC*HDLEAR	C917	NCF1		100.00	same	
P35579-1	MYH9	C*QHLQAEK	C931	NOXO1		0.00	same	
P35606	COPB2	TFEVC*DLPVR	C56	NCF1	0.00	0.00	same	
P35606	COPB2	TFEVC*DLPVR	C56	NCF2	0.00		same	
P35658-1	NUP214	TSC*KDDEAVVQAPR	C1003	NCF1	0.00	0.00	same	
P35908	KRT2	QC*KNVQDAIADAEQR	C417	NCF1	0.00	0.00	same	
P35908	KRT2	QC*KNVQDAIADAEQR	C417	NOXO1	0.00	0.00	same	
P36578	RPL4	YAIC*SALAASALPALVMSK	C125	NCF1	0.00		same	
P36578	RPL4	YAIC*SALAASALPALVMSK	C125	NCF1	1.17	2.05	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P36578	RPL4	YAIC*SALAASALPALVMSK	C125	NCF2	0.00	0.00	same	
P36578	RPL4	GPC*IIYNEDNGIIK	C208	NCF1	0.00	0.00	same	
P36578	RPL4	RGPC*IIYNEDNGIIK	C208	NCF1		6.05	same	
P36578	RPL4	GPC*IIYNEDNGIIK	C208	NCF2	0.00	0.00	same	
P36578	RPL4	GPC*IIYNEDNGIIK	C208	NOXO1		0.00	same	
P36578	RPL4	FC*IWTESAFR	C250	NCF1	0.00	0.00	same	
P36578	RPL4	FC*IWTESAFR	C250	NCF2	0.00	0.00	same	
P36578	RPL4	SGQGAFGNMC*R	C96	NCF1	0.00		same	
P36578	RPL4	SGQGAFGNMC*R	C96	NCF1	7.62	8.90	same	
P36578	RPL4	SGQGAFGNMC*R	C96	NCF2	16.24	19.36	same	
P36578	RPL4	SGQGAFGNMC*R	C96	NOXO1	0.00	0.00	same	
P37802	TAGLN2	NMAC*VQR	C124	NCF1	9.47	8.90	same	
P37802	TAGLN2	NMAC*VQR	C124	NCF1	15.20	15.86	same	
P37802	TAGLN2	NMAC*VQR	C124	NCF2	13.06	13.58	same	
P37802	TAGLN2	NMAC*VQR	C124	NCF2	18.91	21.22	same	
P37802	TAGLN2	NMAC*VQR	C124	NOXO1	12.79	3.39	same	
P37802	TAGLN2	NMAC*VQR	C124	NOXO1	19.55	11.23	same	
P37802	TAGLN2	DGTVLC*ELINALYPEGQAPVK	C63	NCF1	1.27	3.85	same	
P37802	TAGLN2	DGTVLC*ELINALYPEGQAPVKK	C63	NCF2	0.00	0.00	same	
P38398	BRCA1	IDLLASDPHEALIC*K	C442	NCF1	0.00	0.00	same	
P38398	BRCA1	IDLLASDPHEALIC*K	C442	NCF2	0.00	0.00	same	
P39023	RPL3	TVFAEHISDEC*K	C114	NCF1	0.00	0.00	same	
P39023	RPL3	TVFAEHISDEC*KR	C114	NCF1	0.00	0.00	same	
P39023	RPL3	TVFAEHISDEC*K	C114	NCF2		0.00	same	
P39023	RPL3	TVFAEHISDEC*KR	C114	NCF2		0.00	same	
P39023	RPL3	VAC*IGAWHPAR	C253	NCF2	0.00	10.08	same	
P40429	RPL13A	C*EGINISGNFYR	C38	NCF1	0.00	0.00	same	
P40429	RPL13A	C*EGINISGNFYR	C38	NCF2	0.00	0.00	same	
P40429	RPL13A	C*EGINISGNFYR	C38	NOXO1	0.00	0.00	same	
P40926	MDH2	TIIPLISQC*TPK	C212	NCF1	0.00	0.00	same	
P40926	MDH2	TIIPLISQC*TPK	C212	NCF2	0.00	0.00	same	
P40926	MDH2	TIIPLISQC*TPK	C212	NOXO1	0.00		same	
P40926	MDH2	EGVVEC*SFVK	C275	NCF1	0.00	0.00	same	
P40926	MDH2	EGVVEC*SFVK	C275	NCF2	0.00	0.00	same	
P40926	MDH2	EGVVEC*SFVK	C275	NOXO1	0.00	0.00	same	
P40926	MDH2	SQETEC*TYFSTPLLLGK	C285	NCF1	0.00	0.00	same	
P40926	MDH2	SQETEC*TYFSTPLLLGK	C285	NCF2	0.00	0.00	same	
P40926	MDH2	SQETEC*TYFSTPLLLGK	C285	NOXO1	0.00	0.00	same	
P40926	MDH2	GYLGPEQLPDC*LK	C89	NCF1	0.00	9.17	same	
P40926	MDH2	GYLGPEQLPDC*LK	C89	NCF2	0.00	0.00	same	
P40926	MDH2	GYLGPEQLPDC*LK	C89	NOXO1	0.00	0.00	same	
P40926	MDH2	GC*DVVVIPAGVPR	C93	NCF1	0.00	0.00	same	
P40926	MDH2	GC*DVVVIPAGVPR	C93	NCF2	0.00	0.00	same	
P40926	MDH2	GC*DVVVIPAGVPR	C93	NOXO1	0.00	0.00	same	
P42166	TMPO	GGTLFGGEVC*K	C684	NCF1	0.00	0.00	same	
P42167	TMPO	EMFPYEASTPTGISASC*R	C363	NCF1	0.00		same	
P42167	TMPO	EMFPYEASTPTGISASC*R	C363	NCF1	10.81	0.00	same	
P42167	TMPO	EMFPYEASTPTGISASC*R	C363	NCF2	0.00		same	
P42677	RPS27	LTEGC*SFR	C77	NCF1	13.04	10.71	same	
P42677	RPS27	LTEGC*SFR	C77	NCF2	10.31	13.15	same	
P42677	RPS27	LTEGC*SFR	C77	NOXO1	6.96	11.89	same	
P45880	VDAC2	WNTDNTLGTEIAIEDQIC*QGLK	C103	NOXO1	0.00		same	
P45880	VDAC2	SC*SGVEFSTSGSSNTDTGK	C47	NOXO1	0.00	0.00	same	
P45974-1	USP5	VC*ASEKPPK	C838	NCF1	0.00		same	
P45974-1	USP5	VC*ASEKPPK	C838	NCF2	0.00		same	
P45974-1	USP5	VC*ASEKPPK	C838	NOXO1	0.00	0.00	same	
P46060	RANGAP1	SSAC*FTLQELK	C141	NCF1	0.00	0.00	same	
P46060	RANGAP1	LNNC*GMGIGGGK	C152	NCF1	0.00	0.00	same	
P46060	RANGAP1	ILAAALTEC*HR	C169	NCF1	0.00	0.00	same	
P46060	RANGAP1	ALAPLLLAFVTKPNSALESC*SFAR	C573	NCF1	0.00	0.00	same	
P46087	NOP2	VLLDAPC*SGTGVISK	C463	NCF1	0.00	0.00	same	
P46087	NOP2	VLLDAPC*SGTGVISK	C463	NOXO1	0.00		same	
P46736	BRCC3	VLYTC*FQSIQAQK	C172	NCF1	0.00	0.00	same	
P46736	BRCC3	VLYTC*FQSIQAQK	C172	NCF2	0.00		same	
P46736	BRCC3	VC*LESAVELPK	C228	NCF1	0.00		same	
P46736	BRCC3	ILC*QEEQDAYRR	C240	NCF1	0.00	0.00	same	
P46776	RPL27A	NQSFC*PTVNLDK	C70	NCF1	0.00	0.00	same	
P46776	RPL27A	NQSFC*PTVNLDK	C70	NCF2		0.00	same	
P46776	RPL27A	NQSFC*PTVNLDK	C70	NOXO1	0.00	0.00	same	
P46777	RPL5	VGLTNYAAAYC*TGLLLAR	C100	NCF2	100.00	100.00	same	
P46777	RPL5	VGLTNYAAAYC*TGLLLAR	C100	NOXO1	100.00	100.00	same	
P46777	RPL5	DIIC*QIAYAR	C62	NCF1	21.78	24.06	same	
P46777	RPL5	DIIC*QIAYAR	C62	NCF2	0.00		same	
P46777	RPL5	IEGDMIVC*AAYAHELPK	C76	NCF1	0.00		same	
P46777	RPL5	IEGDMIVC*AAYAHELPK	C76	NCF2		0.00	same	
P46777	RPL5	IEGDMIVC*AAYAHELPK	C76	NOXO1		0.00	same	
P46779	RPL28	NC*SSFLIK	C13	NCF1	9.01	4.08	same	
P46779	RPL28	NC*SSFLIK	C13	NCF2	9.05	0.00	same	
P46779	RPL28	NC*SSFLIK	C13	NOXO1	11.29	14.40	same	
P46782	RPS5	VNQAIWLLC*TGAR	C155	NCF1	0.00	0.00	same	
P46782	RPS5	VNQAIWLLC*TGAR	C155	NCF2	0.00	0.00	same	
P46782	RPS5	VNQAIWLLC*TGAR	C155	NOXO1	0.00	0.00	same	
P46782	RPS5	TIAEC*LADELINAAK	C172	NCF1	1.16	0.00	same	
P46782	RPS5	TIAEC*LADELINAAK	C172	NCF2	0.00	0.00	same	
P46782	RPS5	TIAEC*LADELINAAK	C172	NOXO1	0.00	0.00	same	
P46782	RPS5	AQC*PIVER	C66	NCF1	0.00		same	
P46782	RPS5	AQC*PIVER	C66	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P46782	RPS5	AQC*PIVER	C66	NOXO1	0.00	0.00	same	
P46977	STT3A	EGSPVLLNC*LMYK	C637	NCF1	0.00		same	
P46977	STT3A	EGSPVLLNC*LMYK	C637	NCF2	0.00		same	
P46977	STT3A	EGSPVLLNC*LMYK	C637	NOXO1	0.00	0.00	same	
P48509	CD151	VVPDSC*C*K	C184;C185	NCF1	100.00	100.00	same	
P48509	CD151	VVPDSC*C*K	C184;C185	NCF2	100.00	100.00	same	
P48509	CD151	VVPDSC*C*K	C184;C185	NOXO1	100.00	100.00	same	
P48509	CD151	TVVALC*GQR	C192	NCF1	0.00	0.00	same	
P48509	CD151	TVVALC*GQR	C192	NCF2	0.00	0.00	same	
P48509	CD151	TVVALC*GQR	C192	NOXO1	0.00		same	
P48509	CD151	VEGGC*LTK	C208	NCF1	100.00	100.00	same	
P48509	CD151	VEGGC*LTK	C208	NCF2	100.00	100.00	same	
P48509	CD151	VEGGC*LTK	C208	NOXO1	100.00	100.00	same	
P48668	KRT6C KRT6A	QC*ASLQAAIADAEQR	C402	NCF1	0.00		same	
P48668	KRT6C KRT6A	QC*ASLQAAIADAEQR	C402	NCF2	0.00	0.00	same	
P48668	KRT6C KRT6A	QC*ASLQAAIADAEQR	C402	NOXO1	0.00	0.00	same	
P49207	RPL34	SAC*GVC*PGR	C46;C49	NCF1	17.45		same	
P49207	RPL34	SAC*GVC*PGR	C46;C49	NCF2	18.09	17.86	same	
P49207	RPL34	AYGGSMC*AK	C83	NCF1	0.00		same	
P49207	RPL34	AYGGSMC*AK	C83	NCF1	2.48	6.61	same	
P49207	RPL34	AYGGSMC*AK	C83	NCF2	8.13	12.96	same	
P49207	RPL34	AYGGSMC*AK	C83	NOXO1	0.00	0.00	same	
P49327	FASN	AALQEELQLC*K	C1141	NCF1	0.00	6.78	same	
P49327	FASN	SFYGSTLFLC*R	C1403	NCF1	36.08		same	
P49327	FASN	WVC*SSLR	C1548	NCF1	0.00		same	
P49327	FASN	LGMLSPEGTC*K	C212	NCF1	0.00	0.00	same	
P49327	FASN	LGMLSPEGTC*K	C212	NCF2	0.00		same	
P49327	FASN	LSIPTYGLQC*TR	C2273	NCF1	0.00	0.00	same	
P49327	FASN	LSIPTYGLQC*TR	C2273	NCF2	0.00		same	
P49368-1	CCT3	AC*TILLR	C372	NCF1	0.00	0.00	same	DISULFID
P49756-1	RBM25	C*GLVLSWK	C110	NCF2		100.00	same	
P49756-1	RBM25	LQAFGFC*EYKEPESTLR	C132	NCF1	0.00	0.00	same	
P49790	NUP153	C*HQGPIKPYQQGR	C21	NCF1		8.94	same	
P49792	RANBP2	TGEEDEEEFFC*NR	C1196	NCF1	0.00	0.00	same	
P49792	RANBP2	IC*ANHYISPDMK	C1243	NCF1	0.00	0.00	same	
P49792	RANBP2	IC*ANHYISPDMK	C1243	NCF2	0.00		same	
P49792	RANBP2	C*KFEEAQSILK	C1296	NCF2	0.00		same	
P49792	RANBP2	EGQWDC*SVC*LVR	C1612;C1615	NCF1	0.00	0.00	same	
P49792	RANBP2	VWLWTAC*DFADGER	C2407	NCF1		0.00	same	
P49792	RANBP2	VWLWTAC*DFADGER	C2407	NCF2	0.00	0.00	same	
P49792	RANBP2	LKLPPTFFC*YK	C2659	NCF1	0.00	0.00	same	
P49792	RANBP2	LPPTFFC*YK	C2659	NCF1	0.00	0.00	same	
P49792	RANBP2	C*RPLEENTADNEK	C2696	NCF1	0.00	0.00	same	
P49792	RANBP2	EC*IIVWEK	C2710	NCF1	0.00	0.00	same	
P49792	RANBP2	VC*ANHVITK	C2982	NCF1	19.90	15.66	same	
P49792	RANBP2	VC*ANHVITK	C2982	NOXO1	0.00	0.00	same	
P49792	RANBP2	EVADC*FKK	C3032	NCF1	100.00	100.00	same	
P49792	RANBP2	EVADC*FKK	C3032	NCF2	100.00	100.00	same	
P49792	RANBP2	EVADC*FKK	C3032	NOXO1		100.00	same	
P49792	RANBP2	TFEEC*QQNLMK	C3040	NCF1	0.00	0.00	same	
P49792	RANBP2	VIPDFVC*QGGDITK	C3122	NCF2	0.00		same	
P49959	MRE11A MRE11	VTQAIQSFC*LEK	C336	NCF1	0.00	0.00	same	
P49959	MRE11A MRE11	VTQAIQSFC*LEK	C336	NOXO1		0.00	same	
P49959	MRE11A MRE11	TLHTC*LELLR	C75	NCF1	0.00	0.00	same	
P50479-1	PDLIM4	C*GHGIVGTIVK	C258	NCF1	76.70		same	
P50479-1	PDLIM4	C*GHGIVGTIVK	C258	NCF2		0.00	same	
P50479-1	PDLIM4	C*GHGIVGTIVK	C258	NOXO1	100.00	100.00	same	
P50914	RPL14	ALVDGPC*TQVR	C42	NCF1	100.00	100.00	same	
P50914	RPL14	ALVDGPC*TQVR	C42	NCF2	100.00	100.00	same	
P50914	RPL14	ALVDGPC*TQVR	C42	NOXO1	100.00	100.00	same	
P50914	RPL14	C*MQLTDFILK	C54	NCF1	0.00		same	
P50914	RPL14	C*MQLTDFILK	C54	NCF1	12.25	0.00	same	
P50914	RPL14	C*MQLTDFILK	C54	NCF2	9.42	12.04	same	
P50914	RPL14	C*MQLTDFILK	C54	NOXO1	2.77		same	
P50995	ANXA11	GVGTDEAC*LIEILASR	C294	NCF1	0.00	0.00	same	
P50995	ANXA11	GVGTDEAC*LIEILASR	C294	NCF2	0.00	0.00	same	
P50995	ANXA11	FNAVLC*SR	C384	NCF2	23.44		same	
P51149	RAB7A	AQAWC*YSK	C143	NCF2	0.00		same	
P51149	RAB7A	AQAWC*YSK	C143	NOXO1	0.00	0.00	same	
P51610-4	HCFC1	LVIYGGMSGC*R	C227	NCF1	0.00	0.00	same	
P51610-4	HCFC1	LVIYGGMSGC*R	C227	NCF2	0.00		same	
P51610-4	HCFC1	AGHC*AVAINTR	C326	NCF1	0.00	6.55	same	
P51610-4	HCFC1	AGHC*AVAINTR	C326	NOXO1	0.00		same	
P51858	HDGF	C*GDLVFAK	C12	NCF1	0.00	0.00	same	DISULFID
P51991-1	HNRNPA3	YHTINGHNC*EVK	C196	NCF1	0.00	0.00	same	
P51991-1	HNRNPA3	YHTINGHNC*EVK	C196	NOXO1	0.00	0.00	same	
P51991-1	HNRNPA3	WGTLTDC*VVMR	C64	NCF1	0.00		same	
P51991-1	HNRNPA3	WGTLTDC*VVMR	C64	NCF1	0.00	0.00	same	
P51991-1	HNRNPA3	WGTLTDC*VVMR	C64	NCF2	0.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P51991-1	HNRNPA3	WGTLTDC*VVMR	C64	NOXO1	0.00	0.00	same	
P52272	HNRNPM	GC*AVVEFK	C114	NCF1	0.00	0.00	same	
P52272	HNRNPM	GC*AVVEFK	C114	NCF2	0.00	0.00	same	
P52272	HNRNPM	GC*AVVEFK	C114	NOXO1	0.00		same	
P52272	HNRNPM	AC*QIFVR	C653	NCF1	0.00	0.00	same	
P52272	HNRNPM	AC*QIFVR	C653	NCF2		0.00	same	
P52272	HNRNPM	AC*QIFVR	C653	NOXO1	0.00		same	
P52272	HNRNPM	DKFNEC*GHVLYADIK	C676	NCF1	0.00	0.00	same	
P52294	KPNA1	NAVWALSNLC*R	C240	NCF1	0.00	0.00	same	
P52294	KPNA1	NAVWALSNLC*R	C240	NCF2	0.00		same	
P52597	HNRNPF	GLPFGC*TK	C122	NCF1	0.00	0.00	same	
P52597	HNRNPF	GLPFGC*TK	C122	NCF2	0.00	0.00	same	
P52597	HNRNPF	GLPFGC*TK	C122	NOXO1		0.00	same	
P52597	HNRNPF	DLSYC*LSGMYDHR	C267	NCF1	0.00		same	
P52597	HNRNPF	DLSYC*LSGMYDHR	C267	NCF1	0.00	0.00	same	
P52907	CAPZA1	ESC*DSALR	C124	NCF1	0.00	0.00	same	
P52907	CAPZA1	ESC*DSALR	C124	NOXO1	0.00	0.00	same	
P52948-1	NUP98	ISEAC*SLAQQSGDHR	C1347	NCF1	0.00	0.00	same	
P52948-1	NUP98	ISEAC*SLAQQSGDHR	C1347	NCF2	0.00		same	
P52948-1	NUP98	DVC*FHLLK	C1492	NCF1	0.00		same	
P54277	PMS1	C*DADHSFTSLSTPER	C243	NCF1	100.00	100.00	same	
P54277	PMS1	C*DADHSFTSLSTPER	C243	NCF2	100.00	100.00	same	
P54277	PMS1	C*DADHSFTSLSTPER	C243	NOXO1	100.00	100.00	same	
P54709	ATP1B3	EVTVEC*K	C250	NCF1		100.00	same	DISULFID
P54709	ATP1B3	EVTVEC*K	C250	NCF2		91.41	same	DISULFID
P54709	ATP1B3	EVTVEC*K	C250	NOXO1	87.55	87.95	same	DISULFID
P55010	EIF5	LC*TFILK	C138	NCF1	0.00	0.00	same	
P55010	EIF5	FVLC*PEC*ENPETDLHVNPK	C99;C102	NCF1	100.00		same	
P55010	EIF5	FVLC*PEC*ENPETDLHVNPK	C99;C102	NOXO1	100.00	100.00	same	
P55265-1	ADAR	LTEC*QLK	C499	NCF1	0.00	0.00	same	
P55265-1	ADAR	LTEC*QLK	C499	NCF2		0.00	same	
P55265-1	ADAR	LTEC*QLK	C499	NOXO1	0.00	0.00	same	
P55265-1	ADAR	SPVTTLLEC*MHK	C622	NCF1	0.00	0.00	same	
P55265-1	ADAR	SPVTTLLEC*MHK	C622	NCF2	0.00	0.00	same	
P55265-1	ADAR	LGNSC*EFR	C630	NOXO1		100.00	same	
P55265-1	ADAR	GETVNDC*HAEIISR	C909	NCF1	0.00		same	
P55769	NHP2L1 SNU13	LLDLVQQSC*NYK	C30	NCF1	0.00	0.00	same	
P55769	NHP2L1 SNU13	LLDLVQQSC*NYK	C30	NCF2	0.00	0.00	same	
P55769	NHP2L1 SNU13	LLDLVQQSC*NYK	C30	NOXO1	0.00		same	
P57088	TMEM33	LC*LQSIAFISR	C232	NCF1	0.00	0.00	same	
P57088	TMEM33	LC*LQSIAFISR	C232	NCF2	0.00	0.00	same	
P57088	TMEM33	LC*LQSIAFISR	C232	NOXO1		0.00	same	
P57737-3	CORO7-PAM16 CORO7	IVWVC*DGR	C691	NCF1	0.00	0.00	same	
P57737-3	CORO7-PAM16 CORO7	C*LLVSGFDSQSER	C695	NCF1	0.00	0.00	same	
P60709	ACTB	MDDDIAALVVDNGSGMC*K	C17	NCF1	10.76	0.00	same	
P60866	RPS20	VC*ADLIR	C36	NCF1	10.57	6.08	same	
P60866	RPS20	VC*ADLIR	C36	NCF2	12.05	0.00	same	
P60866	RPS20	VC*ADLIR	C36	NOXO1	11.02	0.00	same	
P60903	S100A10	DLDQC*RDGK	C62	NCF1	7.60	6.12	same	
P60903	S100A10	DLDQC*RDGK	C62	NCF2	12.02	0.00	same	
P60903	S100A10	DLDQC*RDGK	C62	NOXO1	0.00	0.00	same	
P61011-1	SRP54	TC*LIC*ADTFR	C133;C136	NCF1	0.00	0.00	same	
P61247	RPS3A	LFC*VGFTK	C139	NCF1	0.00	0.00	same	
P61247	RPS3A	LFC*VGFTK	C139	NCF2	0.00	0.00	same	
P61247	RPS3A	LFC*VGFTK	C139	NOXO1	0.00	0.00	same	
P61247	RPS3A	NC*LTNFHGMDLTR	C96	NCF1	0.00	0.00	same	
P61247	RPS3A	NC*LTNFHGMDLTR	C96	NCF2		0.00	same	
P61247	RPS3A	NC*LTNFHGMDLTR	C96	NOXO1	0.00	0.00	same	
P61513	RPL37A	YTC*SFC*GK	C39;C42	NCF1	0.00	0.00	same	
P61513	RPL37A	YTC*SFC*GK	C39;C42	NCF2	0.00	0.00	same	
P61513	RPL37A	YTC*SFC*GK	C39;C42	NOXO1	0.00	0.00	same	
P61513	RPL37A	AVGIWHC*GSC*MK	C57;C60	NCF1	0.00	0.00	same	
P61513	RPL37A	AVGIWHC*GSC*MK	C57;C60	NCF2		0.00	same	
P61586	RHOA	IGAFGYMEC*SAK	C159	NCF1	0.00	0.00	same	
P61586	RHOA	IGAFGYMEC*SAK	C159	NCF2	0.00	0.00	same	
P61586	RHOA	IGAFGYMEC*SAK	C159	NOXO1	0.00	0.00	same	
P61586	RHOA	LVIVGDGAC*GK	C16	NCF1	14.03	9.86	same	NP_BIND
P61586	RHOA	LVIVGDGAC*GK	C16	NCF2	9.35	0.00	same	NP_BIND
P61586	RHOA	LVIVGDGAC*GK	C16	NOXO1	0.00	0.00	same	NP_BIND
P61978	HNRNPK	IIPTLEEGLQLPSPTATSQLPLESDA VEC*LNYQHYK	C132	NCF2	0.00	0.00	same	
P61978	HNRNPK	IIPTLEEGLQLPSPTATSQLPLESDA VEC*LNYQHYK	C132	NOXO1	0.00		same	
P62195-1	PSMC5	AVAHHTDC*TFIR	C209	NCF1	0.00	0.00	same	
P62241	RPS8	NC*IVLIDSTPYR	C100	NCF2	64.41		same	
P62241	RPS8	NC*IVLIDSTPYR	C100	NOXO1		19.54	same	
P62241	RPS8	LDVGNFSWGSEC*C*TR	C71;C72	NCF1	0.00	0.00	same	
P62244	RPS15A	QVLIRPC*SK	C30	NCF1	4.31	8.26	same	
P62244	RPS15A	QVLIRPC*SK	C30	NOXO1		41.72	same	
P62263	RPS14	C*KELGITALHIK	C85	NCF1	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P62263	RPS14	C*KELGITALHIK	C85	NCF2	0.00	0.00	same	
P62263	RPS14	C*KELGITALHIK	C85	NOXO1	0.00	0.00	same	
P62266	RPS23	ITAFVPNDGC*LNFIEENDEVLVAGF GR	C90	NCF1	7.73	14.37	same	
P62280	RPS11	NMSVHLSPC*FR	C116	NCF1	0.00	7.39	same	
P62280	RPS11	NMSVHLSPC*FR	C116	NOXO1	0.00	0.00	same	
P62280	RPS11	DVQIGDIVTVGEC*RPLSK	C131	NCF1	14.95	11.46	same	
P62280	RPS11	DVQIGDIVTVGEC*RPLSK	C131	NCF2	0.00	0.00	same	
P62280	RPS11	DVQIGDIVTVGEC*RPLSK	C131	NOXO1	0.00	0.00	same	
P62280	RPS11	C*PFTGNVSIR	C60	NCF1	0.00	6.87	same	LIPID
P62280	RPS11	C*PFTGNVSIR	C60	NCF2	0.00	0.00	same	LIPID
P62306	SNRPF	C*NNVLYIR	C66	NCF1	0.00	0.00	same	
P62306	SNRPF	C*NNVLYIR	C66	NCF2	0.00		same	
P62306	SNRPF	C*NNVLYIR	C66	NOXO1	0.00	0.00	same	
P62316	SNRPD2	NNTQVLINC*R	C46	NCF1	8.63	8.36	same	
P62316	SNRPD2	NNTQVLINC*R	C46	NCF2		0.00	same	
P62316	SNRPD2	NNTQVLINC*R	C46	NOXO1	0.00		same	
P62316	SNRPD2	HC*NMVLENVK	C63	NCF1	4.06	1.12	same	
P62316	SNRPD2	HC*NMVLENVK	C63	NCF1	14.46	10.91	same	
P62316	SNRPD2	HC*NMVLENVK	C63	NCF2	0.00	0.00	same	
P62316	SNRPD2	HC*NMVLENVK	C63	NCF2	24.08	14.18	same	
P62316	SNRPD2	HC*NMVLENVK	C63	NOXO1	11.28		same	
P62318	SNRPD3	VLHEAEGHIVTC*ETNTGEVYR	C20	NCF1	0.00	0.00	same	
P62424	RPL7A	TC*TTVAFTQVNSEDK	C199	NCF2	0.00	0.00	same	
P62424	RPL7A	TC*TTVAFTQVNSEDKGALAK	C199	NCF2	0.00		same	
P62495	ETF1	YC*FGVEDTLK	C302	NCF1	0.00	0.00	same	
P62495	ETF1	YC*FGVEDTLK	C302	NCF2		0.00	same	
P62495	ETF1	YC*FGVEDTLK	C302	NOXO1	0.00		same	
P62701	RPS4X	FDTGNLC*MVTGGANLGR	C181	NCF2	100.00	100.00	same	
P62701	RPS4X	FDTGNLC*MVTGGANLGR	C181	NOXO1	100.00	100.00	same	
P62701	RPS4X	EC*LPLIIFLR	C41	NCF1	0.00	0.00	same	
P62701	RPS4X	EC*LPLIIFLR	C41	NCF2		0.00	same	
P62753	RPS6	LNISFPATGC*QK	C12	NCF1	17.20	28.69	same	
P62753	RPS6	LNISFPATGC*QK	C12	NCF2	38.64		same	
P62753	RPS6	LNISFPATGC*QK	C12	NOXO1		0.00	same	
P62826	RAN	VC*ENIPIVLC*GNK	C112;C120	NCF1	0.00	0.00	same	
P62826	RAN	VC*ENIPIVLC*GNK	C112;C120	NCF2	0.00	0.00	same	
P62826	RAN	VC*ENIPIVLC*GNK	C112;C120	NOXO1	0.00	0.00	same	
P62829	RPL23	EC*ADLWPR	C125	NOXO1	100.00		same	
P62829	RPL23	ISLGLPVGAVINC*ADNTGAK	C28	NCF1	0.00	1.11	same	
P62829	RPL23	ISLGLPVGAVINC*ADNTGAK	C28	NCF2	2.37	0.00	same	
P62829	RPL23	ISLGLPVGAVINC*ADNTGAK	C28	NOXO1	0.00	0.00	same	
P62873	GNB1	LFVSGAC*DASAK	C204	NCF2	0.00		same	
P62873	GNB1	LFVSGAC*DASAK	C204	NOXO1	0.00	0.00	same	
P62879	GNB2	TFVSGAC*DASIK	C204	NOXO1	0.00		same	
P62888	RPL30	LVILANNC*PALR	C52	NCF1	1.46	5.46	same	
P62888	RPL30	LVILANNC*PALR	C52	NCF2	12.60	4.72	same	
P62888	RPL30	LVILANNC*PALR	C52	NOXO1	0.00	6.68	same	
P62888	RPL30	TGVHHYSGNNIELGTAC*GK	C85	NCF1	0.00	0.00	same	
P62888	RPL30	TGVHHYSGNNIELGTAC*GK	C85	NCF2		0.00	same	
P62888	RPL30	TGVHHYSGNNIELGTAC*GK	C85	NOXO1	0.00	0.00	same	
P62888	RPL30	VC*TLAIIDPGDSDIIR	C92	NCF1	4.41	2.72	same	
P62888	RPL30	VC*TLAIIDPGDSDIIR	C92	NCF2	2.08	4.53	same	
P62888	RPL30	VC*TLAIIDPGDSDIIR	C92	NOXO1	0.00	0.00	same	
P62910	RPL32	ELEVLLMC*NK	C91	NCF1	0.00		same	
P62910	RPL32	ELEVLLMC*NK	C91	NCF2	36.98		same	
P62910	RPL32	ELEVLLMC*NK	C91	NOXO1		100.00	same	
P62910	RPL32	SYC*AEIAHNVSSK	C96	NCF1	0.00	0.00	same	
P62910	RPL32	SYC*AEIAHNVSSK	C96	NCF2	0.00	0.00	same	
P62910	RPL32	SYC*AEIAHNVSSK	C96	NOXO1	0.00	0.00	same	
P62987	UBA52	C*GHTNNLRPK	C115	NCF1	16.22		same	
P62987	UBA52	C*GHTNNLRPK	C115	NCF2	26.12	33.52	same	
P62987	UBA52	C*GHTNNLRPK	C115	NOXO1	6.48	5.82	same	
P63000-1	RAC1	YLEC*SALTQR	C157	NCF1	10.04	17.69	same	
P63000-1	RAC1	YLEC*SALTQR	C157	NOXO1	0.00	0.00	same	
P63000-1	RAC1	AVLC*PPPVK	C178	NCF1	0.00	0.00	same	
P63000-1	RAC1	AVLC*PPPVK	C178	NCF2	0.00		same	
P63000-1	RAC1	AVLC*PPPVK	C178	NOXO1	0.00	0.00	same	
P63010-1	AP2B1	C*VSTLLDLIQTK	C391	NCF1	0.00		same	
P63104-1	YWHAZ	DIC*NDVLSLLEK	C94	NCF1	0.00		same	
P63220	RPS21	TYAIC*GAIR	C56	NCF1	0.00	0.00	same	
P63244	GNB2L1 RACK1	LWNTLGVC*K	C138	NCF1		0.00	same	
P63244	GNB2L1 RACK1	LWNTLGVC*K	C138	NCF2	0.00		same	
P63244	GNB2L1 RACK1	FSPNSSNPIIVSC*GWDK	C168	NCF1	0.00	0.00	same	
P63244	GNB2L1 RACK1	FSPNSSNPIIVSC*GWDK	C168	NOXO1	0.00		same	
P63244	GNB2L1 RACK1	VWNLANC*K	C182	NCF1	29.63		same	
P63244	GNB2L1 RACK1	VWNLANC*K	C182	NCF2		43.28	same	
P63244	GNB2L1 RACK1	YWLC*AATGPSIK	C249	NCF1		45.85	same	
P63244	GNB2L1 RACK1	YWLC*AATGPSIK	C249	NCF2	100.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
P63244	GNB2L1 RACK1	YWLC*AATGPSIK	C249	NOXO1		100.00	same	
P68104	EEF1A1	NMITGTSQADC*AVLIVAAGVGEFEA GISK	C111	NCF1		0.00	same	
P68104	EEF1A1	DGNASGTTLLEALDC*ILPPTRPTDK PLR	C234	NCF1	4.70	14.00	same	
P68104	EEF1A1	DGNASGTTLLEALDC*ILPPTRPTDK PLR	C234	NCF2		11.37	same	
P68104	EEF1A1	KDGNASGTTLLEALDC*ILPPTRPTD KPLR	C234	NCF2	100.00	100.00	same	
P68104	EEF1A1	KDGNASGTTLLEALDC*ILPPTRPTD KPLR	C234	NOXO1		100.00	same	
P68104	EEF1A1	SGDAAIVDMVPGKPMC*VESFSDYPP LGR	C411	NCF1	0.00	0.00	same	
P68104	EEF1A1	SGDAAIVDMVPGKPMC*VESFSDYPP LGR	C411	NCF2	0.00		same	
P68366	TUBA4A	SIQFVDWC*PTGFK	C347	NCF1	14.22	11.14	same	
P68366	TUBA4A	SIQFVDWC*PTGFK	C347	NCF2	24.56		same	
P68431	H3C1	FQSSAVMALQEAC*EAYLVGLFEDTN LC*AIHAK	C97;C111	NCF1		0.00	same	
P78344-1	EIF4G2	MC*SLMLSK	C282	NCF1	0.00		same	
P78344-1	EIF4G2	MC*SLMLSK	C282	NCF1	0.00	0.00	same	
P78344-1	EIF4G2	MC*SLMLSK	C282	NCF2	0.00	0.00	same	
P78344-1	EIF4G2	LC*LELLNVGVESK	C97	NCF1	0.00	0.00	same	
P78344-1	EIF4G2	LC*LELLNVGVESK	C97	NCF2		0.00	same	
P78347	GTF2I	YC*VEEEEK	C80	NCF1	0.00	0.00	same	
P78347	GTF2I	YC*VEEEEK	C80	NOXO1	0.00	0.00	same	
P78527	PRKDC	LAAVVSAC*K	C1455	NCF1	0.00		same	
P78527	PRKDC	LAAVVSAC*K	C1455	NCF2	0.00	0.00	same	
P83731	RPL24	C*ESAFLSK	C36	NCF1	0.00	0.00	same	
P83731	RPL24	C*ESAFLSK	C36	NCF2	0.00	0.00	same	
P83731	RPL24	C*ESAFLSK	C36	NOXO1	0.00	0.00	same	
P83731	RPL24	VELC*SFSGYK	C6	NCF1	0.00	0.00	same	
P83731	RPL24	VELC*SFSGYK	C6	NCF2	0.00	0.00	same	
P83731	RPL24	VELC*SFSGYK	C6	NOXO1	0.00	0.00	same	
P98175-1	RBM10	INEDWLC*NK	C219	NCF1	0.00		same	
Q00610-1	CLTC	VIQC*FAETGQVQK	C491	NCF1	0.00	0.00	same	
Q00610-1	CLTC	AHIAQLC*EK	C617	NCF1	0.00	0.00	same	
Q00610-1	CLTC	AHIAQLC*EK	C617	NCF2	0.00		same	
Q00610-1	CLTC	IHEGC*EEPATHNALAK	C870	NCF1	0.00	0.00	same	
Q00610-1	CLTC	IHEGC*EEPATHNALAK	C870	NCF2	0.00	0.00	same	
Q00610-1	CLTC	IHEGC*EEPATHNALAK	C870	NOXO1	0.00	0.00	same	
Q00839	HNRNPU	APQC*LGK	C562	NCF1	3.30	3.95	same	
Q00839	HNRNPU	APQC*LGK	C562	NCF2	5.40	3.32	same	
Q00839	HNRNPU	APQC*LGK	C562	NOXO1	2.75	2.83	same	
Q00839	HNRNPU	MC*LFAGFQR	C594	NCF1	0.00	4.92	same	
Q00839	HNRNPU	MC*LFAGFQR	C594	NCF1	0.00	0.00	same	
Q00839	HNRNPU	MC*LFAGFQR	C594	NCF2	0.00	0.00	same	
Q00839	HNRNPU	MC*LFAGFQR	C594	NOXO1		0.00	same	
Q00839	HNRNPU	MC*LFAGFQR	C594	NOXO1	0.00	0.00	same	
Q00839	HNRNPU	AVVVC*PK	C607	NCF1	6.27	12.75	same	
Q00839	HNRNPU	KAVVVC*PK	C607	NCF1	7.26	10.14	same	
Q00839	HNRNPU	AVVVC*PK	C607	NCF2	0.00	6.09	same	
Q00839	HNRNPU	KAVVVC*PK	C607	NCF2	12.45	10.80	same	
Q00839	HNRNPU	AVVVC*PK	C607	NOXO1	0.00	0.00	same	
Q00839	HNRNPU	KAVVVC*PK	C607	NOXO1	0.00	0.00	same	
Q01780	EXOSC10	LLNC*QEFAVDLEHHSYR	C307	NCF1	27.91	29.85	same	
Q01780	EXOSC10	LYC*NVDSNK	C411	NCF1	0.00	0.00	same	
Q01780	EXOSC10	SFPGFQAFC*ETQGDR	C73	NCF1	0.00	0.00	same	
Q01780	EXOSC10	LLQC*MSR	C83	NCF1	3.31	0.00	same	
Q01780	EXOSC10	LLQC*MSR	C83	NCF2		0.00	same	
Q01780	EXOSC10	VMQYHGC*R	C93	NCF1	0.00	0.00	same	
Q01780	EXOSC10	VMQYHGC*R	C93	NCF1	8.44	9.05	same	
Q01780	EXOSC10	VMQYHGC*R	C93	NOXO1	0.00		same	
Q01813	PFKP	AAC*NLLQR	C112	NCF2	15.35	22.50	same	
Q01968-1	OCRL	GPLREPC*ALTLAQR	C28	NCF1	86.92		same	
Q01968-1	OCRL	GPLREPC*ALTLAQR	C28	NCF2	100.00	100.00	same	
Q01968-1	OCRL	FQISNNGQVPC*HFSFIPK	C597	NCF1	0.00	0.00	same	
Q01968-1	OCRL	FQISNNGQVPC*HFSFIPK	C597	NCF2	0.00	0.00	same	
Q01968-1	OCRL	C*VQEAEETLLIDIASNSGC*K	C64;C82	NCF1	0.00		same	
Q02543	RPL18A	DLTTAGAVTQC*YR	C109	NCF1	0.00	0.00	same	
Q02543	RPL18A	DLTTAGAVTQC*YR	C109	NCF2	0.00	0.00	same	
Q02543	RPL18A	DLTTAGAVTQC*YR	C109	NOXO1	0.00	0.00	same	
Q02543	RPL18A	C*HTPPLYR	C22	NCF1	0.00	0.00	same	
Q02543	RPL18A	C*HTPPLYR	C22	NCF2	0.00	0.00	same	
Q02543	RPL18A	C*HTPPLYR	C22	NOXO1	0.00	0.00	same	
Q03112-3	MECOM	C*FGQQTNLDR	C986	NCF1	0.00	0.00	same	
Q03112-3	MECOM	C*FGQQTNLDR	C986	NCF2		0.00	same	
Q03701	CEBPZ	ASHLLETLLC*K	C411	NCF1		84.01	same	
Q04695	KRT17	LSGGLGAGSC*R	C40	NCF1	29.00	29.14	same	
Q04695	KRT17	LSGGLGAGSC*R	C40	NCF2	27.77	37.23	same	
Q04695	KRT17	LSGGLGAGSC*R	C40	NOXO1	38.82	29.82	same	
Q07020	RPL18	GC*GTVLLSGPR	C134	NCF1	22.24		same	
Q07020	RPL18	GC*GTVLLSGPR	C134	NCF2	30.74		same	
Q08211	DHX9	LQISHEAAAC*ITGLR	C1099	NCF1	0.00	0.00	same	
Q08211	DHX9	NFLYAWC*GK	C12	NCF1	8.41	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q08211	DHX9	LAAQSC*ALSLVR	C242	NCF1	0.00	0.00	same	
Q08211	DHX9	LAAQSC*ALSLVR	C242	NCF2	0.00	0.00	same	
Q08211	DHX9	LAAQSC*ALSLVR	C242	NOXO1	0.00	0.00	same	
Q08211	DHX9	AAEC*NIVVTQPR	C438	NCF1	0.00	0.00	same	
Q08211	DHX9	AAEC*NIVVTQPR	C438	NCF2	0.00	0.00	same	
Q08211	DHX9	AAEC*NIVVTQPR	C438	NOXO1	0.00	0.00	same	
Q08211	DHX9	SC*GYSVR	C469	NCF1	0.00	0.00	same	
Q08211	DHX9	SC*GYSVR	C469	NCF2	0.00	0.00	same	
Q08211	DHX9	SC*GYSVR	C469	NOXO1	0.00	0.00	same	
Q08945	SSRP1	ADVIQATGDAIC*IFR	C200	NCF1	0.00	0.00	same	
Q08945	SSRP1	ADVIQATGDAIC*IFR	C200	NCF2	0.00	0.00	same	
Q08945	SSRP1	ELQC*LTPR	C207	NCF1	0.00	0.00	same	
Q09666-1	AHNAK	LEGDLTGPSVGVEVPDVELEC*PDAK	C1900	NCF1	23.77	19.42	same	
Q09666-1	AHNAK	LEGDLTGPSVGVEVPDVELEC*PDAK	C1900	NCF2	21.47	17.12	same	
Q09666-1	AHNAK	LEGDLTGPSVGVEVPDVELEC*PDAK	C1900	NOXO1	0.93	0.00	same	
Q09666-1	AHNAK	LEGDLTGPSVDVEVPDVELEC*PDAK	C2162	NCF1	18.98	18.35	same	
Q09666-1	AHNAK	LEGDLTGPSVDVEVPDVELEC*PDAK	C2162	NCF2	14.92	20.28	same	
Q09666-1	AHNAK	LEGDLTGPSVDVEVPDVELEC*PDAK	C2162	NOXO1	0.00	0.00	same	
Q09666-1	AHNAK	VDVEC*PDVNIEGPEGK	C2806	NCF1	24.37	21.50	same	
Q09666-1	AHNAK	VDVEC*PDVNIEGPEGK	C2806	NCF2	29.21	32.42	same	
Q09666-1	AHNAK	VDVEC*PDVNIEGPEGK	C2806	NOXO1	58.99	56.35	same	
Q09666-1	AHNAK	SSGC*DVNLPGVNVK	C5502	NCF1	13.50	9.76	same	
Q09666-1	AHNAK	SSGC*DVNLPGVNVK	C5502	NCF2	23.34	23.19	same	
Q09666-1	AHNAK	SSGC*DVNLPGVNVK	C5502	NOXO1	1.81	1.99	same	
Q12789-2	GTF3C1	LRPAAAPLEELTMGTSC*LPDTFTK	C1704	NCF1	0.00	0.00	same	
Q12789-2	GTF3C1	LRPAAAPLEELTMGTSC*LPDTFTK	C1704	NCF2	0.00		same	
Q12789-2	GTF3C1	LREELGLC*ER	C286	NCF1	0.00		same	
Q12789-2	GTF3C1	VPPFPLPLEPC*TQEFLWR	C42	NCF1	0.00		same	
Q12802-1	AKAP13	ESLASC*AK	C1838	NCF1	0.00		same	
Q12802-1	AKAP13	ESLASC*AK	C1838	NCF2	0.00	0.00	same	
Q12802-1	AKAP13	LGIPEC*ILLVTQR	C2142	NCF1	0.00	0.00	same	
Q12906-1	ILF3	C*LAALASLR	C203	NCF1	40.60		same	
Q12906-1	ILF3	C*LAALASLR	C203	NCF2		89.84	same	
Q12906-1	ILF3	C*LAALASLR	C203	NOXO1	0.00		same	
Q12906-1	ILF3	VLEC*LASGIVMPDGSGIYDPC*EK	C278;C295	NCF2		100.00	same	
Q12931	TRAP1	NIYYLC*APNR	C501	NCF1	0.00	0.00	same	
Q12931	TRAP1	NIYYLC*APNR	C501	NCF2	0.00	0.00	same	
Q12931	TRAP1	NIYYLC*APNR	C501	NOXO1		0.00	same	
Q12959-2	DLG1	VNDC*ILR	C275	NCF1	0.00		same	
Q12959-2	DLG1	VNDC*ILR	C275	NOXO1	0.00	0.00	same	
Q13111-1	CHAF1A	VLQQFAAC*FLETLPAQEEQTPK	C710	NCF1		0.00	same	
Q13111-1	CHAF1A	VLQQFAAC*FLETLPAQEEQTPK	C710	NCF2	100.00		same	
Q13111-1	CHAF1A	VLQQFAAC*FLETLPAQEEQTPK	C710	NOXO1		100.00	same	
Q13148-1	TARDBP	VTEDENDEPIEIPSEDDGTVLLSTVT AQFPGAC*GLR	C39	NCF1	4.89		same	
Q13177	PAK2	EC*LQALEFLHANQVIHR	C352	NCF1	0.00	0.00	same	
Q13177	PAK2	C*LEMDVEK	C480	NCF1		0.00	same	
Q13177	PAK2	C*LEMDVEKR	C480	NCF1	0.00	0.00	same	
Q13177	PAK2	C*LEMDVEK	C480	NCF2	0.00	0.00	same	
Q13177	PAK2	C*LEMDVEKR	C480	NCF2	0.00	0.00	same	
Q13206	DDX10	SIVFFSSC*K	C323	NCF1		0.00	same	
Q13206	DDX10	SIVFFSSC*K	C323	NCF2	0.00		same	
Q13206	DDX10	SIVFFSSC*K	C323	NOXO1	0.00		same	
Q13206	DDX10	C*FVSYVR	C471	NCF1	0.00	0.00	same	
Q13263	TRIM28	TVYC*NVHK	C209	NCF1		15.98	same	METAL
Q13263	TRIM28	TVYC*NVHK	C209	NCF2	100.00	100.00	same	METAL
Q13263	TRIM28	TVYC*NVHK	C209	NOXO1		100.00	same	METAL
Q13283	G3BP1	VMSQNFTNC*HTK	C73	NCF1		2.24	same	
Q13283	G3BP1	VMSQNFTNC*HTK	C73	NCF2		5.41	same	
Q13283	G3BP1	VMSQNFTNC*HTK	C73	NOXO1		13.68	same	
Q13427	PPIG	ILSC*GELIPK	C174	NCF1	0.00		same	
Q13427	PPIG	VVFELFSDVC*PK	C33	NCF1	0.00	0.00	same	
Q13470-2	TNK1	C*LIPEGAVC*R	C109;C117	NCF1	0.00	0.00	same	
Q13470-2	TNK1	C*LIPEGAVC*R	C109;C117	NCF2	0.00	0.00	same	
Q13470-2	TNK1	C*LIPEGAVC*R	C109;C117	NOXO1	0.00		same	
Q13470-2	TNK1	GELLGSGC*FGVVHR	C126	NCF1	0.00	0.00	same	NP_BIND
Q13470-2	TNK1	GELLGSGC*FGVVHR	C126	NCF2	0.00	0.00	same	NP_BIND
Q13470-2	TNK1	LPRPPLC*SR	C348	NCF2	23.30	22.49	same	
Q13470-2	TNK1	LPRPPLC*SR	C348	NOXO1	0.00	0.00	same	
Q13501-1	SQSTM1	DHRPPC*AQEAPR	C113	NCF1	0.00	0.00	same	
Q13501-1	SQSTM1	DHRPPC*AQEAPR	C113	NCF2		0.00	same	
Q13501-1	SQSTM1	DHRPPC*AQEAPR	C113	NOXO1	0.00	0.00	same	
Q13523	PRPF4B	DLLADLIGC*QR	C962	NCF1	0.00	0.00	same	
Q13586	STIM1	IDKPLC*HSEDEK	C56	NCF1	100.00	100.00	same	
Q13642-1	FHL1	C*SVNLANKR	C255	NCF1	0.00	0.00	same	
Q13642-1	FHL1	FWHDTC*FR	C65	NCF1	0.00		same	
Q13642-1	FHL1	FDC*HYC*R	C7;C10	NCF1	0.00	0.00	same	
Q13642-1	FHL1	C*LHPLANETFVAK	C71	NCF1	0.00	0.00	same	
Q13823	GNL2	VC*NVAPIAGETK	C336	NCF1	0.00	0.00	same	
Q13823	GNL2	VC*NVAPIAGETK	C336	NCF2		0.00	same	
Q13823	GNL2	VC*NVAPIAGETK	C336	NOXO1		0.00	same	
Q13823	GNL2	IFLIDC*PGVVYPSEDSETDIVLK	C362	NCF1	0.00		same	NP_BIND
Q13835	PKP1	TALAYEC*FQDQDNSTLALPSDQK	C14	NOXO1	9.84		same	
Q13835	PKP1	SEPDLYC*DPR	C161	NOXO1	8.28	0.00	same	
Q13835	PKP1	YSFYSTC*SGQK	C190	NCF2	0.00		same	
Q13835	PKP1	YSFYSTC*SGQK	C190	NOXO1	13.09	14.75	same	
Q13835	PKP1	IC*SEDIEC*SGLTIPK	C236;C242	NCF1	0.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q13835	PKP1	IC*SEDIEC*SGLTIPK	C236;C242	NCF2	0.00		same	
Q13835	PKP1	IC*SEDIEC*SGLTIPK	C236;C242	NOXO1	0.00	0.00	same	
Q13835	PKP1	QQVYQLGGIC*K	C290	NCF1	0.00	0.00	same	
Q13835	PKP1	QQVYQLGGIC*K	C290	NCF2	0.00	0.00	same	
Q13835	PKP1	VIIPFSGWC*DGNSNMSR	C387	NCF1	0.00	0.00	same	
Q13835	PKP1	VIIPFSGWC*DGNSNMSR	C387	NCF2	0.00		same	
Q13835	PKP1	VIIPFSGWC*DGNSNMSR	C387	NOXO1	0.00	0.00	same	
Q13835	PKP1	EVVDPEVFFNATGC*LR	C409	NCF2	0.00	0.00	same	
Q13835	PKP1	EVVDPEVFFNATGC*LR	C409	NOXO1	4.29	1.46	same	
Q13835	PKP1	SSTGC*FSNK	C513	NOXO1	0.00	0.00	same	
Q13835	PKP1	MMNNNYDC*PLPEEETNPK	C528	NOXO1	0.00	0.00	same	
Q13835	PKP1	DATLEAC*AGALQNLTASK	C568	NCF1	0.00	0.00	same	
Q13835	PKP1	DATLEAC*AGALQNLTASK	C568	NCF2	0.00	0.00	same	
Q13835	PKP1	DATLEAC*AGALQNLTASK	C568	NOXO1	0.00	0.00	same	
Q13835	PKP1	LLTSHTGNTSNSEDILSSAC*YTVR	C662	NCF1	0.00	0.00	same	
Q13835	PKP1	LLTSHTGNTSNSEDILSSAC*YTVR	C662	NCF2	0.00	0.00	same	
Q13835	PKP1	LLTSHTGNTSNSEDILSSAC*YTVR	C662	NOXO1	0.00	0.00	same	
Q13838-1	DDX39B	NC*PHIVVGTPGR	C165	NCF1	0.00	0.00	same	
Q13838-1	DDX39B	NC*PHIVVGTPGR	C165	NCF2	0.00	0.00	same	
Q13838-1	DDX39B	C*IALAQLLVEQNFPAIAIHR	C300	NCF1	0.00	0.00	same	
Q14103	HNRNPD	GFC*FITFK	C226	NCF1	0.00		same	
Q14126	DSG2	TLAEVC*LGQK	C840	NCF1	5.40	8.17	same	
Q14126	DSG2	TLAEVC*LGQK	C840	NCF2	0.00	5.87	same	
Q14126	DSG2	TLAEVC*LGQK	C840	NOXO1	0.00	6.44	same	
Q14134-1	TRIM29	SGSEEVLC*DSC*IGNK	C173;C176	NCF1	0.00	0.00	same	
Q14134-1	TRIM29	SGSEEVLC*DSC*IGNK	C173;C176	NCF2	0.00	0.00	same	
Q14134-1	TRIM29	DDLLNVC*MR	C420	NCF2	0.00	0.00	same	
Q14151	SAFB2	C*YGFVTMSTSDEATK	C449	NCF1	0.00	0.00	same	
Q14151	SAFB2	LQLEC*QR	C672	NCF1	0.00	0.00	same	
Q14210	LY6D	C*HVC*TSSSNC*K	C23;C26; C32	NCF1	100.00	100.00	same	DISULFID; DISULFID; DISULFID
Q14210	LY6D	C*HVC*TSSSNC*K	C23;C26; C32	NCF2	100.00	100.00	same	DISULFID; DISULFID; DISULFID
Q14210	LY6D	C*HVC*TSSSNC*K	C23;C26; C32	NOXO1	100.00	100.00	same	DISULFID; DISULFID; DISULFID
Q14210	LY6D	HSVVC*PASSR	C38	NCF1		64.44	same	DISULFID
Q14210	LY6D	HSVVC*PASSR	C38	NCF2	73.56	68.60	same	DISULFID
Q14210	LY6D	HSVVC*PASSR	C38	NOXO1	86.86	89.72	same	DISULFID
Q14241	TCEB3 ELOA	MMTLHQQC*IR	C581	NCF1	0.00	0.00	same	
Q14247-1	CTTN	HC*SQVDSVR	C112	NCF1	24.44	21.08	same	
Q14247-1	CTTN	HC*SQVDSVR	C112	NCF2	19.49		same	
Q14247-1	CTTN	C*ALGWDHQEK	C246	NCF1	14.41	13.93	same	
Q14247-1	CTTN	C*ALGWDHQEK	C246	NCF2	12.96	13.84	same	
Q14247-1	CTTN	GVC*KGRYGLFPANYVELR	C534	NCF1	0.00	0.00	same	
Q14258	TRIM25	NTVLC*NVVEQFLQADLAR	C70	NCF1	0.00	0.00	same	
Q14574-1	DSC3	DLDEGPEC*TPAAQYVR	C471	NOXO1	100.00	100.00	same	
Q14574-1	DSC3	VNLC*EC*THPTQC*R	C669;C671; C677	NCF2	100.00		same	
Q14574-1	DSC3	GGNQTLESC*R	C792	NCF1	0.00	0.00	same	
Q14574-1	DSC3	GGNQTLESC*R	C792	NCF2	0.00	0.00	same	
Q14574-1	DSC3	GGNQTLESC*R	C792	NOXO1	0.00	0.00	same	
Q14574-1	DSC3	C*NQNEDR	C836	NOXO1	0.00	0.00	same	
Q14574-1	DSC3	FITLAEAC*TK	C893	NCF1	0.00	0.00	same	
Q14574-1	DSC3	FITLAEAC*TK	C893	NCF2	0.00	0.00	same	
Q14574-1	DSC3	FITLAEAC*TK	C893	NOXO1	0.00	0.00	same	
Q14676	MDC1	C*NVEPVGR	C26	NCF2	100.00	100.00	same	
Q14676	MDC1	C*NVEPVGR	C26	NOXO1	100.00	100.00	same	
Q14678-1	KANK1	MLSAC*NLLK	C1093	NCF1	0.00	0.00	same	
Q14678-1	KANK1	MLSAC*NLLK	C1093	NCF2	0.00		same	
Q14678-1	KANK1	FC*LNTLQHEWFR	C1114	NCF1	0.00		same	
Q14678-1	KANK1	LLLDADVC*NVDHQNK	C1187	NCF1	0.00	0.00	same	
Q14678-1	KANK1	AASQINVC*GVR	C317	NCF1		0.00	same	
Q14692	BMS1	STLIQC*LIR	C101	NCF1	0.00	0.00	same	
Q14692	BMS1	STLIQC*LIR	C101	NCF2	0.00	0.00	same	
Q14692	BMS1	STLIQC*LIR	C101	NOXO1	0.00		same	
Q14692	BMS1	ADSLDC*SR	C734	NCF1	0.00	0.00	same	
Q14839	CHD4	FAEVEC*LAESHQHLSK	C1827	NCF1	9.34	5.65	same	
Q14CN4-2	KRT72	KC*ADLETAIADAEQR	C363	NCF2		0.00	same	
Q15020	SART3	ALVENC*LVPDLWIR	C341	NCF1	0.00	0.00	same	
Q15020	SART3	ALVENC*LVPDLWIR	C341	NCF2	0.00	0.00	same	
Q15020	SART3	ALVENC*LVPDLWIR	C341	NOXO1		0.00	same	
Q15020	SART3	EELEEIC*K	C821	NCF1	0.00	0.00	same	
Q15041	ARL6IP1	FHEIC*SNLVK	C109	NCF1	0.00	0.00	same	
Q15041	ARL6IP1	FHEIC*SNLVK	C109	NCF2	0.00	0.00	same	
Q15054	POLD3	FSAIQC*AAAVPR	C137	NCF1	0.00	0.00	same	
Q15070-1	OXA1L	C*LIFPLIVTGQR	C153	NCF1	0.00		same	
Q15070-1	OXA1L	C*LIFPLIVTGQR	C153	NCF2	0.00		same	
Q15070-1	OXA1L	C*LIFPLIVTGQR	C153	NOXO1	0.00	0.00	same	
Q15185	PTGES3	LTFSC*LGGSDNFK	C40	NCF1	0.00	0.00	same	
Q15233	NONO	FAC*HSASLTVR	C145	NCF1	4.92		same	
Q15233	NONO	FAC*HSASLTVR	C145	NOXO1		100.00	same	
Q15233	NONO	C*SEGSFLLTTFPRPVTVEPMDQLDD EEGLPEK	C208	NCF1	0.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q15365	PCBP1	LVVPATQC*GSLIGK	C109	NCF1	5.07	10.44	same	
Q15365	PCBP1	LVVPATQC*GSLIGK	C109	NCF2	0.00		same	
Q15365	PCBP1	AITIAGVPQSVTEC*VK	C158	NCF1	0.00	3.38	same	
Q15365	PCBP1	AITIAGVPQSVTEC*VK	C158	NCF2	0.00		same	
Q15365	PCBP1	QIC*LVMLETLSQSPQGR	C163	NOXO1	0.00	0.00	same	
Q15366-1	PCBP2	LVVPASQC*GSLIGK	C109	NCF1	0.00	3.42	same	
Q15366-1	PCBP2	LVVPASQC*GSLIGK	C109	NCF2	0.00	0.00	same	
Q15366-1	PCBP2	LVVPASQC*GSLIGK	C109	NOXO1	0.00	0.00	same	
Q15366-1	PCBP2	AITIAGIPQSIIEC*VK	C158	NCF1	0.00	0.00	same	
Q15366-1	PCBP2	AITIAGIPQSIIEC*VK	C158	NCF2	0.00	0.00	same	
Q15366-1	PCBP2	AITIAGIPQSIIEC*VK	C158	NOXO1	0.00	0.00	same	
Q15392	DHCR24	GLEAIC*AK	C252	NCF1	0.00	0.00	same	
Q15392	DHCR24	GLEAIC*AK	C252	NCF2	0.00	0.00	same	
Q15392	DHCR24	GLEAIC*AK	C252	NOXO1	0.00	0.00	same	
Q15393-1	SF3B3	FGNIC*VVR	C1054	NCF1	0.00	0.00	same	
Q15393-1	SF3B3	FGNIC*VVR	C1054	NCF2	0.00		same	
Q15393-1	SF3B3	SEHPPLC*GR	C1156	NCF1	8.48	12.70	same	
Q15393-1	SF3B3	SEHPPLC*GR	C1156	NCF2	11.15	14.89	same	
Q15393-1	SF3B3	SEHPPLC*GR	C1156	NOXO1	12.30		same	
Q15393-1	SF3B3	GMIFVC*SATHK	C289	NCF1	0.00		same	
Q15417	CNN3	DGIILC*ELINK	C59	NCF1	0.00	0.00	same	
Q15418-1	RPS6KA1	IC*DFGFAK	C556	NCF1	0.00	0.00	same	
Q15418-1	RPS6KA1	IC*DFGFAK	C556	NOXO1		0.00	same	
Q15424-1	SAFB	C*YGFVTMSTAEEATK	C448	NCF1	0.00		same	
Q15424-1	SAFB	C*YGFVTMSTAEEATK	C448	NCF1	0.00	0.00	same	
Q15424-1	SAFB	C*INHLHK	C463	NCF1	13.83	8.73	same	
Q15459	SF3A1	EVLDQVC*YR	C244	NCF1	0.00	0.00	same	
Q15637-2	SF1	SITNTTVC*TK	C279	NCF1	0.00	0.00	same	
Q15637-2	SF1	SITNTTVC*TK	C279	NCF2	0.00	0.00	same	
Q15637-2	SF1	SITNTTVC*TK	C279	NOXO1	0.00	0.00	same	
Q15942	ZYX	QNVAVNELC*GR	C384	NCF1	0.00	0.00	same	
Q15942	ZYX	QNVAVNELC*GR	C384	NCF2	0.00		same	
Q16630	CPSF6	VLISSLQDC*LHGIESK	C476	NCF1		0.00	same	
Q16630	CPSF6	VLISSLQDC*LHGIESK	C476	NCF2		100.00	same	
Q16666-1	IFI16	C*QVTPR	C191	NCF1	13.67	14.76	same	
Q16666-1	IFI16	C*QVTPR	C191	NCF2	13.89	14.02	same	
Q16666-1	IFI16	C*QVTPR	C191	NOXO1	7.23	4.90	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDK	C351;C356	NCF1	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDK	C351;C356	NCF1	0.00		same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDK	C351;C356	NCF2	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDK	C351;C356	NCF2		0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDK	C351;C356	NOXO1	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDKLQLFC FR	C351;C356; C366	NCF1	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDKLQLFC FR	C351;C356; C366	NCF1	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDKLQLFC FR	C351;C356; C366	NCF2	0.00	0.00	same	
Q16666-1	IFI16	MDVVGTGQC*HNIPC*EEGDKLQLFC FR	C351;C356; C366	NOXO1	0.00	4.74	same	
Q16666-1	IFI16	LQLFC*FR	C366	NCF1	4.92	4.62	same	
Q16666-1	IFI16	LQLFC*FR	C366	NCF2	0.00	7.84	same	
Q16666-1	IFI16	LQLFC*FR	C366	NOXO1	0.00	0.00	same	
Q16666-1	IFI16	IIAIANYVC*R	C637	NCF1	0.30	0.79	same	
Q16666-1	IFI16	KIIAIANYVC*R	C637	NCF1	4.24	4.15	same	
Q16666-1	IFI16	IIAIANYVC*R	C637	NCF2	1.62	0.30	same	
Q16666-1	IFI16	KIIAIANYVC*R	C637	NCF2	6.16	6.80	same	
Q16666-1	IFI16	IIAIANYVC*R	C637	NOXO1	2.10	2.03	same	
Q16666-1	IFI16	KIIAIANYVC*R	C637	NOXO1	2.73	0.62	same	
Q16666-1	IFI16	INQLC*SQTK	C679	NCF1	12.67	12.09	same	
Q16666-1	IFI16	INQLC*SQTK	C679	NCF2	12.98	14.41	same	
Q16666-1	IFI16	INQLC*SQTK	C679	NOXO1	8.21	7.00	same	
Q16666-1	IFI16	LTTINC*EEGDK	C727	NCF1	5.53	8.11	same	
Q16666-1	IFI16	LTTINC*EEGDKLK	C727	NCF1	10.05	7.87	same	
Q16666-1	IFI16	LTTINC*EEGDK	C727	NCF2	7.82	9.22	same	
Q16666-1	IFI16	LTTINC*EEGDKLK	C727	NCF2	9.29	9.89	same	
Q16666-1	IFI16	LTTINC*EEGDK	C727	NOXO1	2.22	1.20	same	
Q16666-1	IFI16	LTTINC*EEGDKLK	C727	NOXO1	3.85	2.24	same	
Q16666-1	IFI16	LTC*FELAPK	C737	NCF1	10.99	9.74	same	
Q16666-1	IFI16	LTC*FELAPK	C737	NCF2	10.19	11.05	same	
Q16666-1	IFI16	LTC*FELAPK	C737	NOXO1	5.93	4.29	same	
Q1ED39	C16orf88 KNOP1	DVGDTC*SVGK	C184	NCF1	0.00	0.00	same	
Q1ED39	C16orf88 KNOP1	GVSTLC*EEHVEPETTLPAR	C78	NCF1		12.77	same	
Q1KMD3	HNRNPUL2	ALLPHVLC*K	C405	NCF1		0.00	same	
Q1KMD3	HNRNPUL2	NC*VVELNFGQK	C408	NCF1	0.00		same	
Q1KMD3	HNRNPUL2	NC*VVELNFGQK	C408	NCF2	0.00	0.00	same	
Q1KMD3	HNRNPUL2	NC*VVELNFGQK	C408	NOXO1	0.00	0.00	same	
Q1KMD3	HNRNPUL2	DLLVQQASQC*LSK	C518	NOXO1		100.00	same	
Q1KMD3	HNRNPUL2	NFILDQC*NVYNSGQR	C538	NCF2	0.00	0.00	same	
Q1KMD3	HNRNPUL2	NFILDQC*NVYNSGQR	C538	NOXO1	0.00	0.00	same	
Q49AN0-1	CRY2	C*VYILDPWFAASSSVGINR	C52	NOXO1		53.97	same	
Q4VCS5	AMOT	LSIPSLTC*NPDKTDGPVFHSNTLER	C1047	NCF1	0.00	0.00	same	
Q4VCS5	AMOT	LSIPSLTC*NPDKTDGPVFHSNTLER	C1047	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q4VCS5	AMOT	LSIPSLTC*NPDKTDGPVFHSNTLER	C1047	NOXO1	0.00	0.00	same	
Q4VCS5	AMOT	NLRQELEGC*YEK	C455	NCF1		0.00	same	
Q4VCS5	AMOT	QELEGC*YEK	C455	NCF1	7.68	11.26	same	
Q4VCS5	AMOT	NLRQELEGC*YEK	C455	NOXO1		0.00	same	
Q4VCS5	AMOT	MQQALVQLQAAC*EK	C617	NCF1	0.00	0.00	same	
Q4VCS5	AMOT	MQQALVQLQAAC*EK	C617	NCF1	0.00	0.00	same	
Q4VCS5	AMOT	MQQALVQLQAAC*EK	C617	NCF2	0.00	0.00	same	
Q4VCS5	AMOT	MQQALVQLQAAC*EK	C617	NOXO1	0.00		same	
Q4VCS5	AMOT	QGNC*QPTNVSEYNAAALMELLR	C647	NCF1	0.00		same	
Q4VCS5	AMOT	C*LDMEGR	C741	NCF1	0.00		same	
Q4VCS5	AMOT	C*LDMEGR	C741	NCF1	14.98	15.80	same	
Q4VCS5	AMOT	TEQLSC*MRPAK	C781	NCF1		18.67	same	
Q4VCS5	AMOT	DC*STQTER	C862	NCF1	20.53		same	
Q562R1	ACTBL2	HQGVMVGMGQKDC*YVGDEAQSK	C53	NCF1		76.24	same	
Q562R1	ACTBL2	HQGVMVGMGQKDC*YVGDEAQSK	C53	NCF2	79.68		same	
Q562R1	ACTBL2	HQGVMVGMGQKDC*YVGDEAQSK	C53	NOXO1	64.02	65.55	same	
Q56VL3	OCIAD2	QSLLFC*PK	C27	NCF1	85.24	86.35	same	
Q56VL3	OCIAD2	QSLLFC*PK	C27	NCF2	57.06	65.95	same	
Q56VL3	OCIAD2	QSLLFC*PK	C27	NOXO1	91.83	86.72	same	
Q5JTH9-1	RRP12	C*LLHIVR	C839	NCF1	0.00	0.00	same	
Q5QJE6	DNTTIP2	QILIAC*SPVSSVR	C116	NCF1	0.00	0.00	same	
Q5QJE6	DNTTIP2	QILIAC*SPVSSVR	C116	NOXO1		0.00	same	
Q5QJE6	DNTTIP2	NHC*VPPYSESK	C618	NCF1	0.00	0.00	same	
Q5SSJ5-1	HP1BP3	TC*STTALK	C359	NCF1	0.00	0.00	same	
Q5SSJ5-1	HP1BP3	TC*STTALK	C359	NCF2	0.00		same	
Q5SSJ5-1	HP1BP3	TC*STTALK	C359	NOXO1	0.00		same	
Q5T0W9	FAM83B	FLLVDC*QK	C236	NCF1	12.98		same	
Q5T0W9	FAM83B	FLLVDC*QK	C236	NCF2	13.11	18.20	same	
Q5T0W9	FAM83B	SC*VPSSFAQEESAR	C281	NCF1	0.00	0.00	same	
Q5T0W9	FAM83B	SC*VPSSFAQEESAR	C281	NCF2	0.00		same	
Q5T0W9	FAM83B	SC*VPSSFAQEESAR	C281	NOXO1	0.00	0.00	same	
Q5T0W9	FAM83B	LC*SSSDTLVSEGEENQKPK	C801	NOXO1	0.00		same	
Q5T3I0	GPATCH4	ILTDEMLLQAC*EGR	C152	NCF1	5.58	8.64	same	
Q5T3I0	GPATCH4	ILTDEMLLQAC*EGR	C152	NCF1	18.38	14.29	same	
Q5T3I0	GPATCH4	AYADPC*SR	C333	NCF1	10.22	15.27	same	
Q5T7W7	TSTD2	FQGC*LAPDIR	C324	NCF1	100.00		same	
Q5UIP0	RIF1	SLIDNFALNPDILC*SAK	C312	NCF1	100.00		same	
Q5VT06	CEP350	LC*TPLLDLLTR	C2716	NCF1	0.00	0.00	same	
Q5VT06	CEP350	LC*TPLLDLLTR	C2716	NCF2	0.00	0.00	same	
Q5VT06	CEP350	LC*TPLLDLLTR	C2716	NOXO1	0.00	0.00	same	
Q5VTU8	ATP5EP2	YSQIC*AK	C19	NCF1	0.00	0.00	same	
Q5VTU8	ATP5EP2	YSQIC*AK	C19	NCF2	0.00	0.00	same	
Q5VTU8	ATP5EP2	YSQIC*AK	C19	NOXO1	0.00	0.00	same	
Q5VWN6-1	FAM208B	YLC*ASSVGGETLDK	C1125	NCF1	0.00		same	
Q5VWN6-1	FAM208B	AVC*SLQK	C1139	NCF1	8.98	7.55	same	
Q5VWN6-1	FAM208B	AVC*SLQK	C1139	NCF2		0.00	same	
Q5VWN6-1	FAM208B	IVLSC*DDSVK	C739	NCF1	0.00	0.00	same	
Q5VWN6-1	FAM208B	YVSINSTLESC*ELR	C833	NCF1	0.00		same	
Q5VWN6-1	FAM208B	YVSINSTLESC*ELR	C833	NOXO1	0.00	0.00	same	
Q5XXA6-1	ANO1	IHAPWNVLC*R	C166	NCF1	0.00		same	
Q5XXA6-1	ANO1	IHAPWNVLC*R	C166	NCF2		0.00	same	
Q5XXA6-1	ANO1	IHAPWNVLC*R	C166	NOXO1	0.00	0.00	same	
Q5XXA6-1	ANO1	MSSAC*ATAR	C395	NCF1	90.05	100.00	same	DISULFID
Q5XXA6-1	ANO1	MSSAC*ATAR	C395	NCF2	100.00	100.00	same	DISULFID
Q5XXA6-1	ANO1	MSSAC*ATAR	C395	NOXO1	98.40	100.00	same	DISULFID
Q5XXA6-1	ANO1	QQSPPDHEEC*VK	C706	NOXO1	0.00	0.00	same	DISULFID
Q69YH5-1	CDCA2	VADC*VVGK	C280	NCF1	0.00	0.00	same	
Q69YH5-1	CDCA2	SPATPAC*R	C315	NCF1		29.06	same	
Q69YH5-1	CDCA2	SPATPAC*R	C315	NOXO1		0.00	same	
Q6FI81-1	CIAPIN1	ILRPGGC*LFLK	C92	NCF1	0.00	0.00	same	
Q6IBS0	TWF2	AVLPLLDAQQPC*YLLYR	C67	NCF2	19.58		same	
Q6IBS0	TWF2	AVLPLLDAQQPC*YLLYR	C67	NOXO1		100.00	same	
Q6N043	ZNF280D	LQFLTC*K	C487	NCF1		0.00	same	
Q6N043	ZNF280D	LQFLTC*K	C487	NCF2		0.00	same	
Q6P158-1	DHX57	VPLEQLC*LR	C1003	NCF1	0.00	0.00	same	
Q6P1J9	CDC73	IEDEEC*VR	C145	NCF1	0.00		same	
Q6P2E9-1	EDC4	LC*TQLEGLQSTVTGHVER	C976	NCF1	0.00	0.00	same	
Q6PD62	CTR9	TNPGC*PAEVR	C196	NOXO1	100.00		same	
Q6PJT7-1	ZC3H14	LC*EPEVLNSLEETYSPFFR	C261	NCF1	11.25	5.38	same	
Q6PJT7-1	ZC3H14	C*LFVHPNC*K	C634;C641	NCF1	0.00	0.00	same	
Q6ZMH5	SLC39A5	QFALLC*PALLYQIDSR	C185	NCF1		87.68	same	
Q6ZMH5	SLC39A5	QFALLC*PALLYQIDSR	C185	NCF2	63.42	56.17	same	
Q6ZRV2	FAM83H	FEVFC*K	C1140	NCF1	0.00	0.00	same	
Q6ZRV2	FAM83H	FEVFC*K	C1140	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q6ZRV2	FAM83H	FLATEGAPDFLC*PEELEHVSR	C57	NCF1	0.00	0.00	same	
Q6ZRV2	FAM83H	FLATEGAPDFLC*PEELEHVSR	C57	NCF2		0.00	same	
Q6ZRV2	FAM83H	SLESC*LLDLR	C789	NCF1	0.00	0.00	same	
Q6ZRV2	FAM83H	SLESC*LLDLR	C789	NCF2	0.00	0.00	same	
Q6ZRV2	FAM83H	SLESC*LLDLR	C789	NOXO1	0.00	0.00	same	
Q6ZTQ4-2	CDHR3	AVC*HHFGLHIASGSPR	C215	NCF2	100.00	100.00	same	
Q6ZTQ4-2	CDHR3	AVC*HHFGLHIASGSPR	C215	NOXO1		100.00	same	
Q71DI3	H3C15	FQSSAVMALQEASEAYLVGLFEDTNL C*AIHAK	C111	NCF1		14.63	same	LIPID
Q7L014	DDX46	SWVQC*GISMK	C377	NCF1	0.00	0.00	same	
Q7L014	DDX46	SWVQC*GISMK	C377	NOXO1	0.00	0.00	same	
Q7L014	DDX46	VVC*VYGGTGISEQIAELKR	C478	NCF1	0.00	0.00	same	
Q7L014	DDX46	GAEIIVC*TPGR	C501	NCF1	0.00	0.00	same	
Q7L014	DDX46	GAEIIVC*TPGR	C501	NCF2	0.00	0.00	same	
Q7L014	DDX46	SVVC*SDVEQQVIVIEEEKK	C590	NCF2	62.82		same	
Q7L014	DDX46	SVVC*SDVEQQVIVIEEEKK	C590	NOXO1		100.00	same	
Q7L014	DDX46	ASYPC*MSLHGGIDQYDR	C646	NCF1	0.00		same	
Q7L2H7	EIF3M	YTVYC*SLIK	C125	NCF2	100.00		same	
Q7Z2W4	ZC3HAV1	LHIC*DHFTR	C174	NCF1	0.00	0.00	same	
Q7Z2W4	ZC3HAV1	EHGLNPDVVQNIQDIC*NSK	C219	NCF1	0.00	0.00	same	
Q7Z5K2-1	WAPAL WAPL	IVEDDASISSC*NK	C160	NCF1	0.00	0.00	same	
Q7Z5K2-1	WAPAL WAPL	IVEDDASISSC*NK	C160	NCF2		0.00	same	
Q7Z5K2-1	WAPAL WAPL	C*LSVISLATK	C675	NCF1	0.00	0.00	same	
Q7Z6E9	RBBP6	KPLGPPPPSYTC*FR	C161	NCF1	0.00	0.00	same	
Q7Z6E9	RBBP6	KPLGPPPPSYTC*FR	C161	NOXO1	0.00	0.00	same	
Q7Z6E9	RBBP6	C*GKPGHYIK	C164	NCF2	100.00	100.00	same	
Q7Z6E9	RBBP6	C*GKPGHYIK	C164	NOXO1	13.29	13.50	same	
Q7Z6E9	RBBP6	NC*PTNGDKNFESGPR	C174	NOXO1	0.00		same	
Q7Z6E9	RBBP6	AADC*DLQITNAQTK	C44	NOXO1	6.52	7.40	same	
Q86TP1-4	PRUNE PRUNE1	VAISAIYMDLEIC*EVLER	C100	NCF2	100.00		same	
Q86U42-1	PABPN1	VTILC*DK	C205	NCF1	0.00	0.00	same	
Q86V48-1	LUZP1	SC*LFEQGTR	C969	NCF1	0.00		same	
Q86V48-1	LUZP1	SC*LFEQGTR	C969	NCF2	0.00	0.00	same	
Q86V48-1	LUZP1	SC*LFEQGTR	C969	NOXO1	0.00	0.00	same	
Q86Y46	KRT73	QC*ANLETAIADAEQR	C371	NCF1	0.00		same	
Q86Y46	KRT73	QC*ANLETAIADAEQR	C371	NCF2	0.00	0.00	same	
Q8IVF2-3	AHNAK2	EKEDTDVADGC*R	C125	NCF1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	EKEDTDVADGC*R	C125	NCF2	0.00	0.00	same	
Q8IVF2-3	AHNAK2	EKEDTDVADGC*R	C125	NOXO1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	LC*EGTPQEGGLR	C305	NCF1	100.00	100.00	same	
Q8IVF2-3	AHNAK2	LC*EGTPQEGGLR	C305	NCF2	100.00	100.00	same	
Q8IVF2-3	AHNAK2	LC*EGTPQEGGLR	C305	NOXO1	100.00		same	
Q8IVF2-3	AHNAK2	VQMPC*LK	C4123	NCF1	3.98	0.00	same	
Q8IVF2-3	AHNAK2	VQMPC*LK	C4123	NCF2		10.56	same	
Q8IVF2-3	AHNAK2	VQMPC*LK	C4123	NOXO1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	LDLTGPHFESSILSPC*EDVTLTK	C4687	NCF1	12.80	17.42	same	
Q8IVF2-3	AHNAK2	LDLTGPHFESSILSPC*EDVTLTK	C4687	NCF2	31.98	27.22	same	
Q8IVF2-3	AHNAK2	LDLTGPHFESSILSPC*EDVTLTK	C4687	NOXO1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	VDPEC*SVEDSK	C4879	NCF1	9.02	9.82	same	
Q8IVF2-3	AHNAK2	VDPEC*SVEDSK	C4879	NOXO1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	C*DLDSTGLK	C5171	NCF1	28.21	25.84	same	
Q8IVF2-3	AHNAK2	C*DLDSTGLK	C5171	NCF2	17.14	15.77	same	
Q8IVF2-3	AHNAK2	EVQC*PEANIDTALC*K	C5315; C5325	NCF1	0.00	0.00	same	
Q8IVF2-3	AHNAK2	EVQC*PEANIDTALC*K	C5315; C5325	NCF2	0.00	0.00	same	
Q8IWI9-1	MGA	MPVVYLEPC*AVTR	C474	NOXO1		100.00	same	
Q8IWX8	CHERP	LLEETQLDMNEFDNLLQPIIDTC*TK	C168	NCF1	0.00		same	
Q8IWX8	CHERP	SPPHC*ELMAGHLR	C190	NCF1	0.00		same	
Q8IWX8	CHERP	SPPHC*ELMAGHLR	C190	NCF2		100.00	same	
Q8IWX8	CHERP	SPPHC*ELMAGHLR	C190	NOXO1	100.00	100.00	same	
Q8IWX8	CHERP	VVVPIYC*TSFLAVEEDKQQK	C246	NCF1	0.00		same	
Q8IWX8	CHERP	VVVPIYC*TSFLAVEEDKQQK	C246	NOXO1		0.00	same	
Q8IWX8	CHERP	LALEQQQLIC*K	C69	NCF1	0.00	0.00	same	
Q8IWX8	CHERP	LALEQQQLIC*K	C69	NCF2	0.00	0.00	same	
Q8IWX8	CHERP	LALEQQQLIC*K	C69	NOXO1	0.00		same	
Q8IX01-1	SUGP2	FSQLFQTLFELETETC*AK	C362	NCF1	1.05	2.66	same	
Q8IX01-1	SUGP2	C*SLKPEHR	C371	NCF2		100.00	same	
Q8IX01-1	SUGP2	DFC*FFTIK	C381	NCF1	0.00	0.00	same	
Q8IX01-1	SUGP2	NC*FFEIIKPFDK	C417	NCF1	2.94	2.74	same	
Q8IX01-1	SUGP2	SVTPLLMAC*NAYELSVK	C447	NCF1	0.00	0.00	same	
Q8IX01-1	SUGP2	TLSNPLDLALALETTNSLC*R	C476	NCF1	3.16	4.38	same	
Q8IX01-1	SUGP2	TLSNPLDLALALETTNSLC*R	C476	NCF2		0.00	same	
Q8IX01-1	SUGP2	AWLVSSGC*PLQVK	C540	NCF1	9.98	4.02	same	
Q8IX01-1	SUGP2	GADQKPTSADC*AVR	C656	NCF1	4.15	5.51	same	
Q8IX01-1	SUGP2	SEGSTPADGLPGEAAEDDLAGAPALS QASSGTC*FPR	C947	NCF1		17.10	same	
Q8IX01-1	SUGP2	RVC*LIQEPK	C970	NCF1	5.76	11.04	same	
Q8IX01-1	SUGP2	VC*LIQEPK	C970	NCF1		14.97	same	
Q8IX01-1	SUGP2	RVC*LIQEPK	C970	NCF2	0.00	6.19	same	
Q8IX01-1	SUGP2	VC*LIQEPK	C970	NCF2		100.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q8IX01-1	SUGP2	RVC*LIQEPK	C970	NOXO1	0.00		same	
Q8IX01-1	SUGP2	VC*LIQEPK	C970	NOXO1	100.00	100.00	same	
Q8IXZ2-1	ZC3H3	EASLVVTC*R	C301	NCF1	0.00		same	
Q8IY81	FTSJ3	LAC*DFLAR	C145	NCF1	0.00	0.00	same	
Q8IY81	FTSJ3	LAC*DFLAR	C145	NCF2	0.00	0.00	same	
Q8IY81	FTSJ3	LAC*DFLAR	C145	NOXO1	0.00	0.00	same	
Q8N0X7	SPG20 SPART	C*IVNNVSAETVQTVR	C562	NCF2		0.00	same	
Q8N0X7	SPG20 SPART	C*IVNNVSAETVQTVR	C562	NOXO1	0.00		same	
Q8N163-1	CCAR2 KIAA1967	VHLTPYTVDSPIC*DFLELQR	C238	NCF1	0.00		same	
Q8N1G0	ZNF687	TSC*GNITR	C339	NCF1	0.00	0.00	same	
Q8N1G0	ZNF687	C*SLLLHAR	C576	NCF1	0.00	0.00	same	
Q8N1G0	ZNF687	C*SFSAHQR	C718	NCF1	0.00	0.00	same	
Q8N1G0	ZNF687	C*SFSAHQR	C718	NCF2	0.00	0.00	same	
Q8N1G0	ZNF687	C*SFSAHQR	C718	NOXO1	0.00	0.00	same	
Q8N1G2	CMTR1 FTSJD2	SNESHC*SLQIK	C504	NCF1	0.00	0.00	same	
Q8N1G2	CMTR1 FTSJD2	DSTFDLPADSIAPFHIC*YYGR	C794	NCF1	0.00	0.00	same	
Q8N1G2	CMTR1 FTSJD2	RTDPEC*TAPIKK	C9	NCF1	0.00	0.00	same	
Q8N1G2	CMTR1 FTSJD2	RTDPEC*TAPIKK	C9	NOXO1	0.00		same	
Q8N3C0	ASCC3	TVEELFVNC*K	C1638	NCF2	100.00	100.00	same	
Q8N3C0	ASCC3	TVEELFVNC*K	C1638	NOXO1		100.00	same	
Q8N3C0	ASCC3	LKPEC*STEELLSILSDAEEYTDLPV R	C1840	NCF1	0.00		same	
Q8N3C0	ASCC3	FLNEHLQEAC*TPELKPVEK	C208	NCF1	0.00	0.00	same	
Q8N3C0	ASCC3	FLNEHLQEAC*TPELKPVEK	C208	NCF2	0.00	0.00	same	
Q8N3C0	ASCC3	FQALQDNC*K	C310	NCF1	0.00	0.00	same	
Q8N3C0	ASCC3	FQALQDNC*K	C310	NCF2	0.00	0.00	same	
Q8N3C0	ASCC3	FQALQDNC*KK	C310	NCF2	35.88		same	
Q8N5K1	CISD2	VVNEINIEDLC*LTK	C92	NCF1	0.00	0.00	same	
Q8N5K1	CISD2	VVNEINIEDLC*LTK	C92	NCF2	0.00	0.00	same	
Q8N5K1	CISD2	VVNEINIEDLC*LTK	C92	NOXO1	0.00	0.00	same	
Q8ND82	ZNF280C	SERPC*DEDKTDSETGK	C278	NCF1	0.00	0.00	same	
Q8NE71-1	ABCF1	IC*IVGPNGVGK	C655	NCF1		0.00	same	
Q8NE71-1	ABCF1	FGLESHAHTIQIC*K	C741	NCF1	0.00	0.00	same	
Q8NE71-1	ABCF1	VVFAELAC*R	C758	NCF1	0.00	0.00	same	
Q8NFA2-3	NOXO1	EGGPSLGSSGPQFC*ASR	C238	NOXO1	0.00	0.00	same	
Q8TDX7	NEK7	AAC*LLDGVPVALK	C53	NCF1	100.00	100.00	same	
Q8TDX7	NEK7	AAC*LLDGVPVALK	C53	NCF2	100.00	100.00	same	
Q8TDX7	NEK7	AAC*LLDGVPVALK	C53	NOXO1		100.00	same	
Q8TF68-1	ZNF384	C*AHPGC*EK	C375;C380	NCF1	0.00	0.00	same	
Q8TF68-1	ZNF384	VYTC*TIC*SR	C435;C438	NCF1	0.00		same	
Q8WU76	SCFD2	LLEAVGGPDC*HLR	C54	NCF1	0.00	0.00	same	
Q8WU76	SCFD2	LLEAVGGPDC*HLR	C54	NOXO1	0.00	0.00	same	
Q8WUM0	NUP133	LLLC*EHAEK	C641	NCF1	0.00	0.00	same	
Q8WUM0	NUP133	LLLC*EHAEK	C641	NCF2	0.00	0.00	same	
Q8WUM0	NUP133	LLLC*EHAEK	C641	NOXO1	0.00	0.00	same	
Q8WUY1	THEM6	AHTVLAASC*AR	C104	NCF1	11.65	0.00	same	
Q8WUY1	THEM6	AHTVLAASC*AR	C104	NCF2	0.00		same	
Q8WUY1	THEM6	AHTVLAASC*AR	C104	NOXO1	11.52	0.00	same	
Q8WUY1	THEM6	DGFVC*ALLR	C146	NCF1	0.00		same	
Q8WUY1	THEM6	DGFVC*ALLR	C146	NCF2	0.00	0.00	same	
Q8WUY1	THEM6	DGFVC*ALLR	C146	NOXO1	0.00	0.00	same	
Q8WUY1	THEM6	VVQHLC*QR	C168	NCF1	21.54	26.26	same	
Q8WUY1	THEM6	VVQHLC*QR	C168	NCF2	8.27	11.74	same	
Q8WUY1	THEM6	VVQHLC*QR	C168	NOXO1	23.78	28.30	same	
Q8WV24	PHLDA1	TVDC*VER	C242	NCF1	10.68	10.01	same	
Q8WV24	PHLDA1	TVDC*VER	C242	NCF2	27.52	39.08	same	
Q8WV24	PHLDA1	TVDC*VER	C242	NOXO1	18.68		same	
Q8WWM7	ATXN2L	MLHFLTAVVGSTC*DVK	C135	NCF1	0.00	0.00	same	
Q8WXI9	GATAD2B	SC*ASLLR	C409	NCF1	100.00	100.00	same	
Q8WXI9	GATAD2B	VEPFVC*AQC*R	C420;C423	NCF1	0.00	0.00	same	
Q8WXI9	GATAD2B	VEPFVC*AQC*R	C420;C423	NCF2		0.00	same	
Q8WXI9	GATAD2B	VEPFVC*AQC*R	C420;C423	NOXO1	0.00	0.00	same	
Q8WYP5	AHCTF1	C*LVAGLLSPR	C521	NCF1	0.00	0.00	same	
Q8WYP5	AHCTF1	C*LVAGLLSPR	C521	NCF2		0.00	same	
Q8WYP5	AHCTF1	C*LVAGLLSPR	C521	NOXO1	0.00	0.00	same	
Q92482-1	AQP3	C*GEMLHIR	C11	NCF1	13.84	12.29	same	
Q92482-1	AQP3	C*GEMLHIR	C11	NCF2	13.97	18.57	same	
Q92482-1	AQP3	C*GEMLHIR	C11	NOXO1	0.00	0.00	same	
Q92547	TOPBP1	LKPDDSGVNIAEAAAQNVYC*LR	C1463	NCF1	100.00	100.00	same	
Q92547	TOPBP1	LKPDDSGVNIAEAAAQNVYC*LR	C1463	NCF2		100.00	same	
Q92547	TOPBP1	LKPDDSGVNIAEAAAQNVYC*LR	C1463	NOXO1	100.00		same	
Q92551	IP6K1	YPC*VLDLK	C221	NCF1	86.94		same	
Q92551	IP6K1	YPC*VLDLK	C221	NCF2		100.00	same	
Q92551	IP6K1	YPC*VLDLK	C221	NOXO1	0.00		same	
Q92621	NUP205	C*QDVSAGSLQELALLTGIISK	C1662	NCF1	0.00		same	
Q92665	MRPS31	HFMELVTC*GLSK	C356	NCF1	0.00		same	
Q92665	MRPS31	HFMELVTC*GLSK	C356	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	FVIGFIEEAC*KR	C100	NCF1	0.00		same	
Q92797-1	SYMPK	AVAC*SGAAQVR	C578	NCF1	14.47	21.14	same	
Q92797-1	SYMPK	AVAC*SGAAQVR	C578	NCF2	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q92797-1	SYMPK	AVAC*SGAAQVR	C578	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	YC*EDESR	C686	NCF1	3.58	0.00	same	
Q92797-1	SYMPK	KYC*EDESR	C686	NCF2	100.00	100.00	same	
Q92797-1	SYMPK	YC*EDESR	C686	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	GMGMNSPELLLLVENC*PK	C848	NCF1	0.00		same	
Q92797-1	SYMPK	GMGMNSPELLLLVENC*PK	C848	NCF1	3.56	0.00	same	
Q92797-1	SYMPK	GMGMNSPELLLLVENC*PK	C848	NCF2	0.00		same	
Q92797-1	SYMPK	GMGMNSPELLLLVENC*PK	C848	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDK	C859	NCF1	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDKVPPSPELVK	C859	NCF1	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDKVPPSPELVKR	C859	NCF1	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDK	C859	NCF2	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDKVPPSPELVK	C859	NCF2	0.00		same	
Q92797-1	SYMPK	C*LHSLTDK	C859	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	C*LHSLTDKVPPSPELVK	C859	NOXO1	0.00	0.00	same	
Q92797-1	SYMPK	ATNLC*FAER	C969	NCF1	4.04	7.05	same	
Q92797-1	SYMPK	ATNLC*FAER	C969	NOXO1	0.00	0.00	same	
Q92879	CELF1	VMFSSFGQIEEC*R	C137	NCF1	0.00	0.00	same	
Q92900	UPF1	ADSVVVLLC*R	C209	NCF1	0.00	0.00	same	
Q92900	UPF1	ADSVVVLLC*R	C209	NCF2	0.00	0.00	same	
Q92900	UPF1	ADSVVVLLC*R	C209	NOXO1	0.00		same	
Q92900	UPF1	C*FLSWLVK	C237	NCF1	0.00	0.00	same	
Q92900	UPF1	LMQGDEIC*LR	C374	NCF1	0.00	0.00	same	
Q92900	UPF1	LMQGDEIC*LR	C374	NCF2		0.00	same	
Q92900	UPF1	LMQGDEIC*LR	C374	NOXO1		0.00	same	
Q92900	UPF1	QGNGPVLVC*APSNIAVDQLTEK	C531	NCF1	0.00	0.00	same	
Q92900	UPF1	QGNGPVLVC*APSNIAVDQLTEK	C531	NCF2	0.00		same	
Q92900	UPF1	QGNGPVLVC*APSNIAVDQLTEK	C531	NOXO1	0.00		same	
Q92900	UPF1	EKDFIILSC*VR	C852	NCF1	0.00	0.00	same	
Q92900	UPF1	EKDFIILSC*VR	C852	NCF2	0.00		same	
Q92900	UPF1	EKDFIILSC*VR	C852	NOXO1	0.00		same	
Q92922	SMARCC1	TQDEC*ILHFLR	C657	NCF1	0.00	0.00	same	
Q92973-1	TNPO1	SITC*WTLSR	C467	NCF1	0.00	0.00	same	
Q92973-1	TNPO1	SITC*WTLSR	C467	NCF2	0.00	0.00	same	
Q92973-1	TNPO1	SITC*WTLSR	C467	NOXO1	0.00		same	
Q93009	USP7	VDVIFC*DK	C799	NCF1	0.00	0.00	same	
Q93009	USP7	LSESVLSPPC*FVR	C90	NOXO1		0.00	same	
Q96AG4	LRRC59	VAGDC*LDEK	C131	NCF1	0.00	0.00	same	
Q96AG4	LRRC59	VAGDC*LDEK	C131	NCF2	0.00	0.00	same	
Q96AG4	LRRC59	VAGDC*LDEK	C131	NOXO1	0.00		same	
Q96AG4	LRRC59	ATILDLSC*NK	C48	NCF2	0.00	0.00	same	
Q96AG4	LRRC59	LTTLPSDFC*GLTHLVK	C59	NCF1	0.00		same	
Q96AG4	LRRC59	LTTLPSDFC*GLTHLVK	C59	NCF2	0.00	0.00	same	
Q96C19	EFHD2	AAAGELQEDSGLC*VLAR	C172	NCF1	0.00	0.00	same	
Q96C19	EFHD2	AAAGELQEDSGLC*VLAR	C172	NCF2	0.00	0.00	same	
Q96C19	EFHD2	AAAGELQEDSGLC*VLAR	C172	NOXO1		0.00	same	
Q96C24	SYTL4	C*QESLGR	C66	NCF1	0.00	0.00	same	
Q96C24	SYTL4	C*QESLGR	C66	NCF2	0.00	0.00	same	
Q96C24	SYTL4	C*QESLGR	C66	NOXO1	0.00	0.00	same	
Q96C24	SYTL4	GC*NHLVC*R	C83;C88	NCF1	0.00	0.00	same	
Q96D03	DDIT4L	LSSTEPC*GLR	C101	NCF1	100.00	100.00	same	
Q96D03	DDIT4L	LSSTEPC*GLR	C101	NCF2	100.00	100.00	same	
Q96D03	DDIT4L	LSSTEPC*GLR	C101	NOXO1	100.00	100.00	same	
Q96DT7	ZBTB10	SFSC*DIC*GK	C752;C755	NCF1	0.00	0.00	same	
Q96DT7	ZBTB10	SFSC*DIC*GK	C752;C755	NCF2	0.00	0.00	same	
Q96DT7	ZBTB10	SFSC*DIC*GK	C752;C755	NOXO1	0.00	0.00	same	
Q96F86	EDC3	SQDVAVSPQQQQC*SK	C137	NCF1	0.00	0.00	same	
Q96F86	EDC3	SQDVAVSPQQQQC*SK	C137	NCF2	0.00	0.00	same	
Q96F86	EDC3	SQDVAVSPQQQQC*SK	C137	NOXO1	0.00		same	
Q96F86	EDC3	EFC*TDSGLVVPSISYELHK	C272	NCF1	98.47	97.84	same	
Q96F86	EDC3	EFC*TDSGLVVPSISYELHK	C272	NCF2	100.00	100.00	same	
Q96F86	EDC3	EFC*TDSGLVVPSISYELHK	C272	NOXO1	100.00	100.00	same	
Q96F86	EDC3	NVHQRPTVALLC*GPHVK	C341	NCF1	0.00	0.00	same	
Q96F86	EDC3	NVHQRPTVALLC*GPHVK	C341	NCF2	0.00	0.00	same	
Q96F86	EDC3	NVHQRPTVALLC*GPHVK	C341	NOXO1	0.00	0.00	same	
Q96F86	EDC3	GAQGISC*GR	C353	NCF1	9.85	12.02	same	
Q96F86	EDC3	GAQGISC*GR	C353	NOXO1	0.00		same	
Q96F86	EDC3	DLPTSPVDLVINC*LDC*PENVFLR	C410;C413	NCF1	0.00	0.00	same	
Q96F86	EDC3	C*LVPEVTFR	C47	NCF1	5.01	5.27	same	
Q96F86	EDC3	C*LVPEVTFR	C47	NCF2	0.00		same	
Q96FQ6	S100A16	MSDC*YTELEK	C4	NCF1	0.00	0.00	same	
Q96FQ6	S100A16	MSDC*YTELEK	C4	NCF2	0.00	0.00	same	
Q96FQ6	S100A16	MSDC*YTELEK	C4	NOXO1	0.00	0.00	same	
Q96I24-1	FUBP3	IQAESGC*K	C109	NCF1	0.00		same	
Q96I24-1	FUBP3	C*QHAAHIISELILTAQER	C310	NCF1	0.00		same	
Q96I25	RBM17	C*VIFEIPGAPDDEAVR	C339	NCF1	0.00	0.00	same	
Q96I25	RBM17	AC*FYNLDK	C385	NCF1	0.00	0.00	same	
Q96I82-1	KAZALD1	GEVPEPLC*AC*R	C126;C128	NCF1	100.00	100.00	same	
Q96I82-1	KAZALD1	GEVPEPLC*AC*R	C126;C128	NCF2	100.00	100.00	same	
Q96I82-1	KAZALD1	GEVPEPLC*AC*R	C126;C128	NOXO1	100.00	100.00	same	
Q96JM3	CHAMP1	LEC*DHC*SFR	C16;C19	NCF1	0.00	0.00	same	
Q96JM3	CHAMP1	C*NFESNFPR	C770	NCF1	0.00	0.00	same	
Q96JP5-1	ZFP91	C*EMEGC*GTVLAHPR	C313;C318	NCF1	0.00	0.00	same	
Q96JP5-1	ZFP91	YVC*PHPSC*GR	C344;C349	NCF1	0.00	0.00	same	
Q96JP5-1	ZFP91	DYIC*EYC*AR	C374;C377	NCF1	0.00	0.00	same	
Q96JP5-1	ZFP91	MIHTGEKPLQC*EIC*GFTC*R	C402;C405; C409	NCF1	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q96KG9	SCYL1	LGC*LIWEVFNGPLPR	C210	NCF2		100.00	same	
Q96KR1	ZFR	NVNLVLLC*SEKPSK	C812	NCF1		0.00	same	
Q96PK6-1	RBM14	VIEC*DVVK	C108	NCF1	0.00	0.00	same	
Q96PK6-1	RBM14	VIEC*DVVK	C108	NCF2	0.00	0.00	same	
Q96PK6-1	RBM14	VIEC*DVVK	C108	NOXO1	0.00	0.00	same	
Q96PK6-1	RBM14	IFVGNVDGADTTPEELAALFAPYGTV MSC*AVMK	C31	NCF1		0.00	same	
Q96PK6-1	RBM14	IFVGNVDGADTTPEELAALFAPYGTV MSC*AVMK	C31	NOXO1	5.26	0.00	same	
Q96PK6-1	RBM14	IFVGNVSAAC*TSQELR	C90	NCF1	0.00	0.00	same	
Q96PK6-1	RBM14	IFVGNVSAAC*TSQELR	C90	NCF2		0.00	same	
Q96PK6-1	RBM14	IFVGNVSAAC*TSQELR	C90	NOXO1	0.00	0.00	same	
Q96Q89-1	KIF20B	VSELSLC*DLAGSER	C379	NCF1		0.00	same	
Q96Q89-1	KIF20B	VSELSLC*DLAGSER	C379	NCF2	0.00		same	
Q96Q89-1	KIF20B	VSELSLC*DLAGSER	C379	NOXO1		0.00	same	
Q96QC0	PPP1R10	C*TYLNILLQTR	C48	NCF1	0.00		same	
Q96QD5-1	DEPDC7	C*HPDIFIEHFGD	C500	NCF1	0.00	0.00	same	
Q96QD5-1	DEPDC7	C*HPDIFIEHFGD	C500	NCF2	0.00	0.00	same	
Q96RN5	MED15	TFVPAMTAIHGPPITAPVVC*TR	C660	NCF1	0.00	0.00	same	
Q96RN5	MED15	TFVPAMTAIHGPPITAPVVC*TR	C660	NCF1	0.00	0.00	same	
Q96RN5	MED15	TFVPAMTAIHGPPITAPVVC*TR	C660	NOXO1	0.00	0.00	same	
Q96RQ3	MCCC1	TSAAQAIHPGC*GFLSENMEFAELC*K	C129;C142	NCF1	0.00	2.10	same	
Q96RQ3	MCCC1	TSAAQAIHPGC*GFLSENMEFAELC*K	C129;C142	NCF2	6.98	17.30	same	
Q96RQ3	MCCC1	TSAAQAIHPGC*GFLSENMEFAELC*K	C129;C142	NOXO1	3.49	4.88	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NCF1	0.00	0.00	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NCF1	15.09	10.31	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NCF2		0.00	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NCF2	17.21	15.23	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NOXO1	0.00	0.00	same	
Q96RQ3	MCCC1	SIMAAAGVPVVEGYHGEDQSDQC*LK	C190	NOXO1	5.20	1.08	same	
Q96RQ3	MCCC1	HNFC*FMEMNTR	C332	NCF1		0.00	same	
Q96RQ3	MCCC1	HNFC*FMEMNTR	C332	NCF1	4.42	2.48	same	
Q96RQ3	MCCC1	HNFC*FMEMNTR	C332	NCF2	8.06	7.05	same	
Q96RQ3	MCCC1	HNFC*FMEMNTR	C332	NOXO1		0.00	same	
Q96RQ3	MCCC1	HNFC*FMEMNTR	C332	NOXO1	0.54	0.00	same	
Q96RQ3	MCCC1	ESLC*QAALGLILK	C509	NCF1	6.87	5.50	same	
Q96RQ3	MCCC1	ESLC*QAALGLILK	C509	NCF2	7.37	10.66	same	
Q96RQ3	MCCC1	ESLC*QAALGLILK	C509	NOXO1	7.49	5.55	same	
Q96RQ3	MCCC1	TFQVLGNLYSEGDC*TYLK	C595	NCF1	2.39	0.78	same	
Q96RQ3	MCCC1	TFQVLGNLYSEGDC*TYLK	C595	NCF2	10.76	12.78	same	
Q96RQ3	MCCC1	TFQVLGNLYSEGDC*TYLK	C595	NOXO1	0.00	0.00	same	
Q96RQ3	MCCC1	C*SVNGVASK	C600	NCF1	5.48	5.42	same	
Q96RQ3	MCCC1	C*SVNGVASK	C600	NCF2	8.43	7.40	same	
Q96RQ3	MCCC1	C*SVNGVASK	C600	NOXO1	4.06	3.77	same	
Q96RT1	ERBB2IP ERBIN	AQVAFEC*DEDKDER	C464	NCF1	0.00	0.00	same	
Q96RT1	ERBB2IP ERBIN	AQVAFEC*DEDKDER	C464	NOXO1	0.00	0.00	same	
Q96SI9	STRBP	SIGTC*NRPLGAGEALR	C254	NCF1	100.00	100.00	same	
Q96SI9	STRBP	SIGTC*NRPLGAGEALR	C254	NCF2		100.00	same	
Q96ST2	IWS1	ALRPGDPGFC*AR	C749	NCF1	78.74		same	
Q96ST2	IWS1	ALRPGDPGFC*AR	C749	NOXO1	86.20		same	
Q99442	SEC62	AVDC*LLDSK	C55	NCF1	100.00		same	
Q99459	CDC5L	TAAQC*LEHYEFLLDK	C96	NCF1	0.00	0.00	same	
Q99613	EIF3C	GC*ILTLVER	C443	NCF1	86.73		same	
Q99613	EIF3C	GC*ILTLVER	C443	NCF2	100.00		same	
Q99613	EIF3C	GC*ILTLVER	C443	NOXO1	100.00	100.00	same	
Q99613	EIF3C	GTTEEVC*R	C500	NCF1	0.00	0.00	same	
Q99613	EIF3C	TMVQLGIC*AFR	C619	NCF1	0.00	0.00	same	
Q99613	EIF3C	TMVQLGIC*AFR	C619	NOXO1		0.00	same	
Q99613	EIF3C	TC*HSFIINEK	C752	NCF1	0.00	0.00	same	
Q99613	EIF3C	C*LEEFELLGK	C79	NCF1	0.00	0.00	same	
Q99613	EIF3C	C*LEEFELLGK	C79	NCF2	0.00		same	
Q99714-1	HSD17B10	VDVAVNC*AGIAVASK	C91	NCF1		0.00	same	
Q99832	CCT7	TC*TFILR	C370	NCF1	0.00	0.00	same	
Q99832	CCT7	TC*TFILR	C370	NCF2	0.00	0.00	same	
Q99959	PKP2	AVC*GALR	C416	NOXO1	0.00	0.00	same	
Q99959	PKP2	C*DGLIDSLVHYVR	C579	NOXO1	0.00	0.00	same	
Q99961	SH3GL1	QNFIDPLQNLC*EK	C147	NCF1	21.97		same	
Q99961	SH3GL1	SMPPLDQPSC*K	C311	NCF1	0.00	10.11	same	
Q99961	SH3GL1	NPGYPQSEGLLGEC*MIR	C96	NCF1	0.00	0.00	same	
Q9BQ39	DDX50	AIIFC*ETK	C396	NCF1	0.00	0.00	same	
Q9BQ39	DDX50	AIIFC*ETK	C396	NCF2		0.00	same	
Q9BQG0	MYBBP1A	IFTHHLC*R	C884	NCF1	0.00		same	
Q9BQI9	NRIP2	ISAGC*LSR	C188	NCF2	100.00	100.00	same	
Q9BQI9	NRIP2	ISAGC*LSR	C188	NOXO1	100.00	100.00	same	
Q9BRR8	GPATCH1	LPPVFC*PNAR	C839	NCF1	0.00	0.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q9BSJ8	ESYT1	LGTQTFC*SR	C370	NCF1	0.00	0.00	same	
Q9BSJ8	ESYT1	LGTQTFC*SR	C370	NCF2		0.00	same	
Q9BSJ8	ESYT1	LGTQTFC*SR	C370	NOXO1	0.00		same	
Q9BSJ8	ESYT1	KLVSIVHGC*R	C995	NCF1	0.00	0.00	same	
Q9BTC0	DIDO1	AYPVSGC*FDYLSEDLPDTIHIGGR	C1079	NCF1	0.00		same	
Q9BTC0	DIDO1	AYPVSGC*FDYLSEDLPDTIHIGGR	C1079	NOXO1		0.00	same	
Q9BTT6	LRRC1	NEIPEIPESISFC*K	C104	NCF1	0.00	0.00	same	
Q9BTT6	LRRC1	NEIPEIPESISFC*K	C104	NOXO1		0.00	same	
Q9BTT6	LRRC1	C*GALENLVNDVSDEAWNER	C432	NCF1		0.00	same	
Q9BTT6	LRRC1	C*GALENLVNDVSDEAWNER	C432	NOXO1	0.00	0.00	same	
Q9BU76-1	C1orf35	AEDQTESSC*ESHR	C179	NCF1	0.00	2.18	same	
Q9BU76-1	C1orf35	AEDQTESSC*ESHR	C179	NOXO1		0.00	same	
Q9BUJ2-1	HNRNPUL1	AEPYC*SVLPGFTFIQHLPLSER	C391	NCF1	0.00	0.00	same	
Q9BUJ2-1	HNRNPUL1	WDVLIQQATQC*LNR	C487	NCF1		82.61	same	
Q9BUJ2-1	HNRNPUL1	WDVLIQQATQC*LNR	C487	NCF2		100.00	same	
Q9BUJ2-1	HNRNPUL1	WDVLIQQATQC*LNR	C487	NOXO1	100.00	100.00	same	
Q9BUJ2-1	HNRNPUL1	AIVIC*PTDEDLKDR	C532	NCF2	100.00		same	
Q9BV81	EMC6	GNAAVLDYC*R	C29	NCF2	100.00	100.00	same	
Q9BY42	RTFDC1	RPIVAC*ELGR	C51	NCF1	0.00	0.00	same	
Q9BY89-1	KIAA1671	VSSNQDPDSC*R	C585	NCF2	0.00	0.00	same	
Q9BY89-1	KIAA1671	ATVAVSEEHC*APGATSVR	C842	NCF1	0.00	0.00	same	
Q9BZE4	GTPBP4	TLLLC*GYPNVGK	C174	NCF1	0.00		same	
Q9C0C2	TNKS1BP1	EHGVGGVSQC*PEPGLR	C1373	NCF1		0.00	same	
Q9H089	LSG1	QFLC*IPR	C112	NCF1	0.00	0.00	same	
Q9H089	LSG1	QFLC*IPR	C112	NCF2		0.00	same	
Q9H0A0	NAT10	FILSLASC*K	C194	NCF1	0.00	0.00	same	
Q9H0A0	NAT10	FILSLASC*K	C194	NCF2		0.00	same	
Q9H0A0	NAT10	FILSLASC*K	C194	NOXO1	0.00	0.00	same	
Q9H0A0	NAT10	WLNDLLC*LDC*LNITR	C486;C489	NCF1	0.00		same	
Q9H0A0	NAT10	WLNDLLC*LDC*LNITR	C486;C489	NOXO1	0.00	0.00	same	
Q9H0C8	ILKAP	C*LLDTFK	C190	NCF1	0.00	0.00	same	
Q9H0C8	ILKAP	C*GVTSVPDIR	C301	NCF1	0.00	0.00	same	
Q9H0C8	ILKAP	FILLAC*DGLFK	C325	NCF1	0.00	0.00	same	
Q9H2P0	ADNP	EESSIHC*K	C223	NCF1	0.00	0.00	same	
Q9H2P0	ADNP	IC*TIC*NELFPENVYSVHFEK	C449;C452	NCF1	0.00		same	
Q9H2P0	ADNP	TLC*PLC*FSILK	C624;C627	NCF1	0.00	0.00	same	
Q9H3N1	TMX1	FIITALPTIYHC*K	C106	NCF1	0.00		same	
Q9H5V9-1	CXorf56	SVVC*SDTR	C11	NCF2	100.00	100.00	same	
Q9H5V9-1	CXorf56	SVVC*SDTR	C11	NOXO1	100.00	100.00	same	
Q9H6R4-1	NOL6	LAELLTQQHGLQC*R	C777	NOXO1	81.38		same	
Q9H6S0	YTHDC2	IFC*TQPR	C245	NCF1	0.00	0.00	same	
Q9H8V3	ECT2	VTHLVANC*TQGEK	C221	NCF1	0.00	0.00	same	
Q9H8V3	ECT2	INQAKPEC*GR	C584	NCF1	0.00	0.00	same	
Q9H8V3	ECT2	VETISLGEHPC*DR	C686	NCF2	100.00	100.00	same	
Q9H8Y5	ANKZF1	APRTSC*SGSGERESPER	C48	NCF2	0.00		same	
Q9H8Y5	ANKZF1	APRTSC*SGSGERESPER	C48	NOXO1	0.00	0.00	same	
Q9H910-1	HN1L JPT2	DHVFLC*EGEEPK	C118	NCF1	0.00	0.00	same	
Q9H920	RNF121	LSC*NHVFHEFC*IR	C251;C259	NCF1	0.00	0.00	same	
Q9H9B1	EHMT1	APLLVLC*EDHR	C575	NCF1	0.00	9.43	same	
Q9H9B1	EHMT1	APLLVLC*EDHR	C575	NCF2	0.00		same	
Q9H9B1	EHMT1	APLLVLC*EDHR	C575	NOXO1	0.00		same	
Q9HA47	UCK1	STVC*EKIMELLGQNEVEQR	C40	NCF1	100.00	100.00	same	
Q9HA47	UCK1	STVC*EKIMELLGQNEVEQR	C40	NCF2	100.00		same	
Q9HB71	CACYBP	TDTVLILC*R	C154	NCF1		0.00	same	
Q9HCS7	XAB2	LWLDYC*QFLMDQGR	C117	NCF1	0.00		same	
Q9HCS7	XAB2	LWLDYC*QFLMDQGR	C117	NCF1	4.92	6.76	same	
Q9HCS7	XAB2	LWLDYC*QFLMDQGR	C117	NCF2	0.00	0.00	same	
Q9HCS7	XAB2	SNYQLWHELC*DLISQNPDK	C225	NCF1	0.00		same	
Q9HCS7	XAB2	LWC*SLADYYIR	C260	NCF1	0.00	0.00	same	
Q9HCS7	XAB2	LWC*SLADYYIR	C260	NCF2	0.00	0.00	same	
Q9HCS7	XAB2	QVDDLASVWC*QC*GELELR	C437;C439	NCF1	0.00	0.00	same	
Q9HCS7	XAB2	DLFEQALDGC*PPK	C587	NCF1	0.00	1.69	same	
Q9HCS7	XAB2	ARDLFEQALDGC*PPK	C587	NCF2		100.00	same	
Q9HCS7	XAB2	DLFEQALDGC*PPK	C587	NCF2	0.00	0.00	same	
Q9HCS7	XAB2	ARDLFEQALDGC*PPK	C587	NOXO1	100.00	100.00	same	
Q9HCS7	XAB2	DLFEQALDGC*PPK	C587	NOXO1	0.00	0.00	same	
Q9HCS7	XAB2	LLPC*SYK	C66	NCF1	12.96	12.29	same	
Q9HCS7	XAB2	LLPC*SYK	C66	NCF2	18.99	18.26	same	
Q9HCS7	XAB2	LLPC*SYK	C66	NOXO1	32.38		same	
Q9HCS7	XAB2	FADMEC*K	C676	NCF1		2.82	same	
Q9HCS7	XAB2	FADMEC*K	C676	NCF1	5.62	5.10	same	
Q9HCS7	XAB2	FADMEC*K	C676	NCF2		100.00	same	
Q9HCS7	XAB2	FADMEC*K	C676	NCF2	1.83	1.54	same	
Q9HCS7	XAB2	FADMEC*K	C676	NOXO1	0.00	0.00	same	
Q9HCS7	XAB2	AIYSFC*SQIC*DPR	C691;C695	NCF1	0.00	0.00	same	
Q9HCS7	XAB2	AIYSFC*SQIC*DPR	C691;C695	NCF2	0.00	0.00	same	
Q9HCS7	XAB2	AIYSFC*SQIC*DPR	C691;C695	NOXO1	0.00	0.00	same	
Q9HCS7	XAB2	C*VTDPAYEDVNNC*HER	C86;C98	NCF1	0.00	0.00	same	
Q9HCS7	XAB2	C*VTDPAYEDVNNC*HER	C86;C98	NCF2	0.00	0.00	same	
Q9HCS7	XAB2	C*VTDPAYEDVNNC*HER	C86;C98	NOXO1	0.00	0.00	same	
Q9HCY8	S100A14	DLVTQQLPHLMPSNC*GLEEK	C62	NCF2	0.00	0.00	same	
Q9HCY8	S100A14	IANLGSC*NDSK	C74	NCF1	0.00	0.00	same	
Q9HCY8	S100A14	IANLGSC*NDSK	C74	NCF2	0.00	0.00	same	
Q9HCY8	S100A14	IANLGSC*NDSK	C74	NOXO1	0.00	0.00	same	
Q9NQS7	INCENP	MGTTAPGPIHLLELC*DQK	C15	NCF2	0.00	0.00	same	
Q9NQS7	INCENP	MGTTAPGPIHLLELC*DQK	C15	NOXO1	0.00		same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q9NQS7	INCENP	LMEFLC*NMDNK	C24	NCF1	0.00	0.00	same	
Q9NQS7	INCENP	IIC*HSYLER	C355	NCF1	11.82	3.36	same	
Q9NQS7	INCENP	IIC*HSYLER	C355	NOXO1		0.00	same	
Q9NQS7	INCENP	TPSSPC*PASK	C483	NCF1	0.00	0.00	same	
Q9NR30-1	DDX21	TIIFC*ETK	C445	NCF1	6.93	0.00	same	
Q9NR30-1	DDX21	TIIFC*ETK	C445	NCF2	0.00	0.00	same	
Q9NR30-1	DDX21	TIIFC*ETK	C445	NOXO1	0.00	0.00	same	
Q9NR30-1	DDX21	LGVC*FDVPTASVTEIQEK	C682	NCF1	100.00	100.00	same	
Q9NR30-1	DDX21	LGVC*FDVPTASVTEIQEK	C682	NCF2	100.00	100.00	same	
Q9NR30-1	DDX21	LGVC*FDVPTASVTEIQEK	C682	NOXO1	100.00	100.00	same	
Q9NRD9	DUOX1	AC*QLINR	C1533	NCF1	0.00		same	
Q9NRD9	DUOX1	TDHLASC*LQR	C423	NCF1		0.00	same	
Q9NRD9	DUOX1	TDHLASC*LQR	C423	NOXO1	0.00	0.00	same	
Q9NRD9	DUOX1	EALTC*ELSR	C800	NCF1		0.00	same	
Q9NRD9	DUOX1	EALTC*ELSR	C800	NOXO1	7.83	0.00	same	
Q9NRD9	DUOX1	SFIEISNNC*LSK	C890	NOXO1	0.00	0.00	same	
Q9NRP0-1	OSTC	VPFLVLEC*PNLK	C14	NCF1	0.00	0.00	same	
Q9NRP0-1	OSTC	VPFLVLEC*PNLK	C14	NCF2	0.00	0.00	same	
Q9NRP0-1	OSTC	VPFLVLEC*PNLK	C14	NOXO1	0.00	0.00	same	
Q9NRY5	FAM114A2	SLAELTAC*SIELFHK	C388	NCF1	0.00		same	
Q9NVP1	DDX18	NLQC*LVIDEADR	C329	NCF1	9.51	13.79	same	
Q9NVP1	DDX18	NLQC*LVIDEADR	C329	NCF2	0.00	0.00	same	
Q9NVP1	DDX18	NLQC*LVIDEADR	C329	NOXO1	0.00	0.00	same	
Q9NX02	NLRP2	LGPC*SFAELVFR	C256	NCF1	0.00		same	
Q9NX02	NLRP2	LGPC*SFAELVFR	C256	NCF2	0.00		same	
Q9NYF8-1	BCLAF1	C*DSADLR	C688	NCF1	0.00	4.57	same	
Q9NYF8-1	BCLAF1	C*DSADLR	C688	NOXO1	0.00		same	
Q9NYL9	TMOD3	C*FSLAATR	C231	NCF1	0.00	0.00	same	
Q9NYL9	TMOD3	C*FSLAATR	C231	NCF2	0.00		same	
Q9NYZ3	GTSE1	ESC*TAHAASQAATQR	C198	NCF1	0.00	0.00	same	
Q9NYZ3	GTSE1	ESC*TAHAASQAATQR	C198	NCF2		0.00	same	
Q9NYZ3	GTSE1	AVGSPLC*VPAR	C539	NCF1	0.00		same	
Q9NZ01-1	TECR	LC*FLDK	C18	NCF1	0.00	4.54	same	
Q9NZ01-1	TECR	LC*FLDK	C18	NCF2	3.31	5.36	same	
Q9NZ01-1	TECR	LC*FLDK	C18	NOXO1	0.00	0.00	same	
Q9NZB2-1	FAM120A	LLVDADNC*LHR	C53	NCF1	0.00	0.00	same	
Q9NZB2-1	FAM120A	LLVDADNC*LHR	C53	NCF2		0.00	same	
Q9NZB2-1	FAM120A	LLVDADNC*LHR	C53	NOXO1	0.00	0.00	same	
Q9NZM1	MYOF	KPVVGQC*TIER	C1392	NCF1	0.00	0.00	same	
Q9NZM1	MYOF	KPVVGQC*TIER	C1392	NCF2	0.00	0.00	same	
Q9NZM1	MYOF	KPVVGQC*TIER	C1392	NOXO1		0.00	same	
Q9NZM1	MYOF	GLELQPQDNNGLC*DPYIK	C1574	NCF1	100.00	100.00	same	
Q9P0L0	VAPA	LMEEC*KR	C179	NCF1	0.00	0.00	same	
Q9P0L0	VAPA	KLMEEC*KR	C179	NCF1	11.96		same	
Q9P0L0	VAPA	LMEEC*KR	C179	NCF2		0.00	same	
Q9P0L0	VAPA	KLMEEC*KR	C179	NCF2	6.84	0.00	same	
Q9P0L0	VAPA	LMEEC*KR	C179	NOXO1	0.00		same	
Q9P258	RCC2	LIEGLSHEVIVSAAC*GR	C209	NCF1	0.00		same	
Q9P2X0-1	DPM3	VATFHDC*EDAAR	C67	NCF1	0.00	0.00	same	
Q9P2X0-1	DPM3	VATFHDC*EDAAR	C67	NOXO1	0.00	0.00	same	
Q9UBC9	SPRR3	VPEPGC*TK	C66	NCF1	0.00	0.00	same	
Q9UBC9	SPRR3	VPEPGC*TK	C66	NCF2	0.00	0.00	same	
Q9UBC9	SPRR3	VPEPGC*TK	C66	NOXO1	0.00	0.00	same	
Q9UBD5-1	ORC3	SYC*ENHLGSTAK	C483	NCF1	0.00	0.00	same	
Q9UBD5-1	ORC3	ENVVNFIDC*LVR	C561	NCF1	0.00	0.00	same	
Q9UBM7	DHCR7	VIEC*SYTSADGQR	C380	NCF1	0.00	0.00	same	
Q9UBM7	DHCR7	VIEC*SYTSADGQR	C380	NCF2	0.00		same	
Q9UBM7	DHCR7	VIEC*SYTSADGQR	C380	NOXO1	0.00	0.00	same	
Q9UG63	ABCF2	LAHEDAEC*EK	C186	NCF2	0.00	0.00	same	
Q9UGU0	TCF20	SVIC*DISPLR	C868	NCF1	15.48		same	
Q9UGU0	TCF20	SVIC*DISPLR	C868	NCF2	0.00		same	
Q9UGU0	TCF20	SVIC*DISPLR	C868	NOXO1		0.00	same	
Q9UHB6-4	LIMA1	LLANQQVFHISC*FR	C415	NCF2		22.21	same	
Q9UHB6-4	LIMA1	IYC*KPHFNQLFK	C439	NCF1	0.00	0.00	same	
Q9UHB7-1	AFF4	C*ESLLYLR	C1005	NCF1	0.00	0.00	same	
Q9UHB9	SRP68	FETFC*LDPSLVTK	C562	NCF2	100.00	100.00	same	
Q9UHD1	CHORDC1	MALLC*YNR	C5	NCF1	0.00	0.00	same	METAL
Q9UHD1	CHORDC1	MALLC*YNR	C5	NCF2	0.00		same	METAL
Q9UHQ9	CYB5R1	VPEDPTQC*FLLFANQTEK	C208	NCF1	100.00		same	
Q9UHQ9	CYB5R1	VPEDPTQC*FLLFANQTEK	C208	NCF2		100.00	same	
Q9UHQ9	CYB5R1	VPEDPTQC*FLLFANQTEK	C208	NOXO1		100.00	same	
Q9UHW9	SLC12A6	HIDVC*SK	C390	NCF2	100.00	100.00	same	
Q9UHW9	SLC12A6	HIDVC*SK	C390	NOXO1	100.00	100.00	same	
Q9UHX1-1	PUF60	ALAIMC*R	C129	NCF2		0.00	same	
Q9UIG0	BAZ1B	ELDELLNC*LHPQGIR	C1006	NCF1		100.00	same	
Q9UIG0	BAZ1B	ELDELLNC*LHPQGIR	C1006	NOXO1		100.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q9UIG0	BAZ1B	SALSC*VISK	C497	NCF1	0.00	0.00	same	
Q9UIG0	BAZ1B	LQILTALC*HR	C753	NCF1	0.00	0.00	same	
Q9UK61	FAM208A	LLC*YISLR	C371	NCF1	0.00	0.00	same	
Q9UK61	FAM208A	ELINLIQC*R	C690	NCF1	0.00	0.00	same	
Q9UKV3-1	ACIN1	SHC*FVTYSTVEEAVATR	C1052	NCF1	0.00	0.00	same	
Q9UKV3-1	ACIN1	SHC*FVTYSTVEEAVATR	C1052	NCF2		0.00	same	
Q9UKV3-1	ACIN1	AAPC*IYWLPLTDSQIVQK	C1223	NCF1	0.00	0.00	same	
Q9UKV3-1	ACIN1	AAPC*IYWLPLTDSQIVQK	C1223	NCF2	0.00	0.00	same	
Q9UKX7	NUP50	AC*VGNAYHK	C151	NCF1	4.35	16.49	same	
Q9UKX7	NUP50	AC*VGNAYHK	C151	NOXO1	0.00		same	
Q9UKX7	NUP50	HVNTNPLC*DLTPIFK	C181	NCF1	0.00	0.00	same	
Q9UKX7	NUP50	ADTNLGNILLNVLIPPNMPC*TR	C416	NCF1	6.97	9.17	same	
Q9ULA0	DNPEP	NDTPC*GTTIGPILASR	C413	NCF1	100.00	100.00	same	
Q9ULA0	DNPEP	NDTPC*GTTIGPILASR	C413	NCF2	100.00	100.00	same	
Q9ULA0	DNPEP	NDTPC*GTTIGPILASR	C413	NOXO1	100.00	100.00	same	
Q9ULJ7-1	ANKRD50	TALSVAALC*VPASK	C720	NCF1	0.00		same	
Q9ULJ7-1	ANKRD50	TALSVAALC*VPASK	C720	NCF2	0.00	0.00	same	
Q9ULJ7-1	ANKRD50	TALSVAALC*VPASK	C720	NOXO1	0.00	0.00	same	
Q9ULK4	MED23	DNRPQGWC*LSDTYLK	C1043	NCF1	0.00		same	
Q9ULK4	MED23	IIC*NLKPALK	C1319	NCF1	0.00	0.00	same	
Q9ULK4	MED23	DMVC*NMLGLNK	C284	NCF1	0.00		same	
Q9ULK4	MED23	LISC*LGAFR	C40	NCF1	5.76	14.40	same	
Q9ULV4	CORO1C	NDQC*YDDIR	C23	NCF1	6.27	8.45	same	
Q9ULW0	TPX2	HTVPC*MPPAK	C198	NCF1	0.00	0.00	same	
Q9ULW0	TPX2	IC*RDPQTPVLQTK	C364	NCF1	0.00	0.00	same	
Q9ULW0	TPX2	TVEIC*PFSFDSR	C536	NCF1	0.00	0.00	same	
Q9UNY4-1	TTF2	C*SPFDEFNLWR	C785	NCF1	0.00	0.00	same	
Q9UPN9-2	TRIM33	TC*IEAHQR	C241	NCF2		0.00	same	METAL
Q9UQR1	ZNF148	SHVC*EHC*NAAFR	C173;C176	NCF1	0.00		same	
Q9Y2U8	LEMD3	LAQLAGDHEC*GSSSQR	C547	NCF1	100.00	100.00	same	
Q9Y2X3	NOP58	LNLSC*IHSPVVNELMR	C106	NCF1	0.00	0.00	same	
Q9Y2X3	NOP58	LNLSC*IHSPVVNELMR	C106	NCF2	0.00	0.00	same	
Q9Y2X3	NOP58	LNLSC*IHSPVVNELMR	C106	NOXO1	0.00	0.00	same	
Q9Y2X9	ZNF281	TC*GEVIVK	C367	NCF1	15.33	15.27	same	
Q9Y2X9	ZNF281	TC*GEVIVK	C367	NCF2		0.00	same	
Q9Y314	NOSIP	TVTC*PMSGKPLR	C185	NCF1	0.00	0.00	same	
Q9Y314	NOSIP	YVC*AVTR	C223	NCF2	81.36		same	
Q9Y314	NOSIP	DSLSNATPC*AVLRPSGAVVTLEC*VEK	C236;C250	NCF1	0.00	0.00	same	
Q9Y314	NOSIP	DSLSNATPC*AVLRPSGAVVTLEC*VEK	C236;C250	NCF2	0.00	0.00	same	
Q9Y314	NOSIP	DSLSNATPC*AVLRPSGAVVTLEC*VEK	C236;C250	NOXO1	0.00	0.00	same	
Q9Y314	NOSIP	NC*TAGAVYTYHEK	C8	NCF1	0.00	0.00	same	
Q9Y3U8	RPL36	EVC*GFAPYER	C48	NCF2	0.00	4.42	same	
Q9Y3U8	RPL36	EVC*GFAPYER	C48	NOXO1	10.80	12.26	same	
Q9Y446-2	PKP3	SAVDLSC*SR	C144	NCF1	0.00	0.00	same	
Q9Y446-2	PKP3	SAVDLSC*SR	C144	NCF2	6.88	19.25	same	
Q9Y446-2	PKP3	SAVDLSC*SR	C144	NOXO1	12.34	11.05	same	
Q9Y446-2	PKP3	TTSRPEAGVC*SLALPSDLQLDR	C33	NOXO1	21.62		same	
Q9Y446-2	PKP3	C*YSDAAAK	C362	NOXO1	0.00	0.00	same	
Q9Y446-2	PKP3	SVENAVC*VLR	C544	NCF2	60.02		same	
Q9Y446-2	PKP3	SVENAVC*VLR	C544	NOXO1	37.50	38.91	same	
Q9Y446-2	PKP3	DLAGAPPGEVVGC*FTPQSR	C584	NCF1	0.00	0.00	same	
Q9Y446-2	PKP3	DLAGAPPGEVVGC*FTPQSR	C584	NCF2		0.00	same	
Q9Y446-2	PKP3	DLAGAPPGEVVGC*FTPQSR	C584	NOXO1	0.00	0.00	same	
Q9Y4E8	USP15	IGNILDLC*TALSALSGIPADK	C506	NCF1	0.00	0.00	same	
Q9Y4E8	USP15	IGNILDLC*TALSALSGIPADK	C506	NCF2	0.00	0.00	same	
Q9Y4K1	AIM1 CRYBG1	ISSVQPIC*LDSFTGPR	C1404	NCF1	5.59	15.65	same	
Q9Y4K1	AIM1 CRYBG1	ISSVQPIC*LDSFTGPR	C1404	NCF2	14.72	9.92	same	
Q9Y4K1	AIM1 CRYBG1	SPEAVGSEC*PSR	C541	NCF1	12.12	16.01	same	
Q9Y4K1	AIM1 CRYBG1	SPEAVGSEC*PSR	C541	NCF2	26.48	21.31	same	
Q9Y4K1	AIM1 CRYBG1	SPEAVGSEC*PSR	C541	NOXO1	0.00		same	
Q9Y4K1	AIM1 CRYBG1	SNLPNC*ANSDTDFMGLFK	C976	NCF2	0.00	0.00	same	
Q9Y4P8-1	WIPI2	NIC*SLATIQK	C334	NCF1	0.00		same	
Q9Y4W2	LAS1L	VEC*VLAELK	C306	NCF1	0.00	0.00	same	
Q9Y4W2	LAS1L	VEC*VLAELK	C306	NCF2	0.00	0.00	same	
Q9Y4W2	LAS1L	VEC*VLAELK	C306	NOXO1	0.00	0.00	same	
Q9Y5B9	SUPT16H	SYC*SNLVR	C283	NCF1		0.00	same	
Q9Y5B9	SUPT16H	INFYC*PGSALGR	C574	NCF1	0.00	0.00	same	
Q9Y5K6	CD2AP	AQIIELLC*IVEALKK	C595	NCF1	0.00	0.00	same	
Q9Y5X1	SNX9	GFLGC*FPDIIGTHK	C502	NCF1	0.00		same	
Q9Y624	F11R	LVC*YNNK	C74	NCF1	0.00	0.00	same	DISULFID
Q9Y624	F11R	LVC*YNNK	C74	NCF2	0.00	0.00	same	DISULFID
Q9Y6C9	MTCH2	LC*SGVLGTVVHGK	C79	NCF2	100.00		same	
Q9Y6I9	TEX264	LFTESC*SISPK	C68	NCF1	100.00	100.00	same	

Continued on next page

Table 2A.2 – continued from previous page

Accession	Protein	Sequence	Positions	Bait	(-) CAT	(+) CAT	Different	Function
Q9Y6I9	TEX264	LFTESC*SISPK	C68	NOXO1		100.00	same	
Q9Y6Y8	SEC23IP	SIIEC*VDDFR	C467	NCF1	0.00	0.00	same	

Table 5A.1: ionFinder and envoMatch output from 200 PSMs from synthetic deaminated and citrullinated peptides.

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	Det	Amb	DetNL	ArtNL
GFGrGGAESHTFK	GFGrGGAESHTFK	True	17824	02207a_BA7-TUM_mod_citrullination_l_7_01_01-DDA-1h-R1	Likely	1.00	True	1	11	0	0
EDHGrGYFEYIEENKYSR	EDHGrGYFEYIEENKYSR	True	32954	02207a_BD2-TUM_mod_citrullination_s_2_01_01-DDA-1h-R1	True	0.94	True	4	10	3	0
AEGGGGGGrPGAPAAGDGK	AEGGGGGGrPGAPAAGDGK	True	7190	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	True	0.99	True	1	6	2	0
GDFSSANNrDNTYNR	GDFSSANNrDNTYNR	True	15436	02207a_BA10-TUM_mod_citrullination_l_10_01_01-DDA-1h-R1	True	0.99	True	0	21	6	0
SGQSEDrQPVPGQQMTLK	SGQSEDrQPVPGQQMTLK	True	24935	02207a_BB4-TUM_mod_citrullination_l_16_01_01-DDA-1h-R1	True	0.94	True	0	20	4	2
KGPGPGGPGGAGVArGGAGGGPSGD	KGPGPGGPGGAGVArGGAGGGPSGD	True	13576	02207a_BB10-TUM_mod_citrullination_l_22_01_01-DDA-1h-R1	True	0.99	True	11	19	10	0
sSAHFNrGPAYGLSAEVK	sSAHFNrGPAYGLSAEVK	True	33818	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	True	0.97	True	0	17	6	0
YYGGGSEGGrAPK	YYGGGSEGGrAPK	True	11256	02207a_BA8-TUM_mod_citrullination_l_8_01_01-DDA-1h-R1	True	1.00	True	11	10	10	0
EGrGESENAGTNQETR	EGrGESENAGTNQETR	True	5754	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	True	0.99	True	3	15	4	0
rSYVSSGEMmVGGLAPGR	rSYVSSGEMmVGGLAPGR	True	33252	02207a_BD2-TUM_mod_citrullination_s_2_01_01-DDA-1h-R1	True	0.99	True	3	14	2	0
EQTEGEYSSLEHESArGVIEcLK	EQTEGEYSSLEHESArGVIEcLK	True	38470	02207a_BD2-TUM_mod_citrullination_s_2_01_01-DDA-1h-R1	True	0.85	True	2	8	2	0
IGGGIDVPVPrHSVGVVIGR	IGGGIDVPVPrHSVGVVIGR	True	40271	02207a_BD2-TUM_mod_citrullination_s_2_01_01-DDA-1h-R1	True	0.99	True	3	16	2	0
McrTLEDQLSELK	McrTLEDQLSELK	True	40126	02207a_BA7-TUM_mod_citrullination_l_7_01_01-DDA-1h-R1	Likely	0.99	True	0	9	1	0
rGMDDDRGPR	rGMDDDRGPR	True	6275	02207a_BA2-TUM_mod_citrullination_l_2_01_01-DDA-1h-R1	Likely	0.99	True	0	3	1	0
MAAPIDrVGQTIER	MAAPIDrVGQTIER	True	35805	02207a_BD1-TUM_mod_citrullination_s_1_01_01-DDA-1h-R1	True	0.98	True	1	17	5	0
GLDrVQDEYSR	GLDrVQDEYSR	True	22441	02207a_BA4-TUM_mod_citrullination_l_4_01_01-DDA-1h-R1	True	0.99	True	2	11	5	1
DSYSSrDYPSSR	DSYSSrDYPSSR	True	14464	02207a_BD1-TUM_mod_citrullination_s_1_01_01-DDA-1h-R1	True	0.99	True	3	9	5	0
rKGTDVNVFNTILTTR	rKGTDVNVFNTILTTR	True	43883	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	Likely	0.99	True	1	9	0	0
ITPGArGAFSEEYK	ITPGArGAFSEEYK	True	26922	02207a_BD1-TUM_mod_citrullination_s_1_01_01-DDA-1h-R1	True	0.99	True	4	8	4	0
DYYSSrSQSGGYSDR	DYYSSrSQSGGYSDR	True	17335	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	True	1.00	True	2	17	7	0
sIHFSSPVFTSrSAAFSGR	sIHFSSPVFTSrSAAFSGR	True	45213	02207a_BB6-TUM_mod_citrullination_l_18_01_01-DDA-1h-R1	True	0.98	True	8	13	7	0
IAYQrNDDDEEEAAR	IAYQrNDDDEEEAAR	True	12176	02207a_BD3-TUM_mod_citrullination_s_3_01_01-DDA-1h-R1	Likely	0.99	True	0	10	1	0
LrNEDLGLPPLFPPR	LrNEDLGLPPLFPPR	True	52932	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	True	0.97	True	1	19	5	0
YFDrINENDPEYIR	YFDrINENDPEYIR	True	34873	02207a_BA9-TUM_mod_citrullination_l_9_01_01-DDA-1h-R1	True	1.00	True	2	12	3	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNL	nArtNL
rVFImDScDELIP EYLNFIR	rVFImDScDELIP EYLNFIR	True	59872	02207a_BB8-TUM_ mod_citrullination_ l_20_01_01-DDA-1h-R1	True	0.94	True	6	15	2	0
DYYDrGYDR	DYYDrGYDR	True	21011	02207a_BA2-TUM_ mod_citrullination_ l_2_01_01-DDA-1h-R1	True	0.99	True	3	7	3	0
DYTYrDYGHSSSR	DYTYrDYGHSSSR	True	18102	02207a_BA8-TUM_ mod_citrullination_ l_8_01_01-DDA-1h-R1	True	0.98	True	4	11	6	0
TAWrLDcAQLK	TAWrLDcAQLK	True	35879	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	3	12	7	1
rDYDDmSPR	rDYDDmSPR	True	8427	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	Likely	0.99	True	1	7	1	0
VSAQrLQEAGILS AEELQR	VSAQrLQEAGILS AEELQR	True	43341	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	True	0.98	True	0	26	9	0
rGETESEEFEK	rGETESEEFEK	True	18337	02207a_BA4-TUM_ mod_citrullination_ l_4_01_01-DDA-1h-R1	Likely	1.00	True	2	10	1	0
TVLSNVQEELDrM TR	TVLSNVQEELDrM TR	True	53722	02207a_BA12-TUM_ mod_citrullination_ l_12_01_01-DDA-1h-R1	True	0.98	True	5	16	9	0
ErDFTSLENTVEE R	ErDFTSLENTVEE R	True	37968	02207a_BA9-TUM_ mod_citrullination_ l_9_01_01-DDA-1h-R1	Likely	0.96	True	2	12	1	1
ITrLQEKEDLQEL NDR	ITrLQEKEDLQEL NDR	True	28335	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.99	True	2	14	3	0
SLEKQrQDLANLQ K	SLEKQrQDLANLQ K	True	24354	02207a_BA9-TUM_ mod_citrullination_ l_9_01_01-DDA-1h-R1	True	1.00	True	0	20	9	0
EEKErLEQDLQQm QAK	EEKErLEQDLQQm QAK	True	23335	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	Amb	0.99	True	0	5	0	0
GGTSrEGGQTAPA STR	GGTSrEGGQTAPA STR	True	5823	02207a_BA12-TUM_ mod_citrullination_ l_12_01_01-DDA-1h-R1	True	1.00	True	1	16	5	0
rDLEGSDIDTR	rDLEGSDIDTR	True	22464	02207a_BA4-TUM_ mod_citrullination_ l_4_01_01-DDA-1h-R1	Likely	0.99	True	3	10	1	0
rGEEGHDPKEPEQ LR	rGEEGHDPKEPEQ LR	True	11178	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.98	True	1	7	2	0
SQSGGYSDrSSGG SYR	SQSGGYSDrSSGG SYR	True	10179	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.98	True	7	14	9	0
EVTINQSLLAPLr LDADPSLQR	EVTINQSLLAPLr LDADPSLQR	True	54701	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.98	True	0	23	6	0
WGDrDSEGTWR	WGDrDSEGTWR	True	28509	02207a_BA4-TUM_ mod_citrullination_ l_4_01_01-DDA-1h-R1	True	0.99	True	2	9	2	0
YEELQVTAGrHGD DLR	YEELQVTAGrHGD DLR	True	25997	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.87	True	4	11	4	0
LQTESGEFSrQLD EKEALVSQLSR	LQTESGEFSrQLD EKEALVSQLSR	True	42615	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	Amb	0.96	True	0	18	0	0
ELDrNYLNYGEEG APGK	ELDrNYLNYGEEG APGK	True	31930	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.99	True	1	18	7	0
DSYSSrDYPSSR	DSYSSrDYPSSR	True	14852	02207a_BA6-TUM_ mod_citrullination_ l_6_01_01-DDA-1h-R1	True	1.00	True	7	10	7	0
rDLEEATLQHEAT AATLR	rDLEEATLQHEAT AATLR	True	33016	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	Likely	0.42	False	0	8	1	0
ENAAAPSPVrAPA PSPAK	ENAAAPSPVrAPA PSPAK	True	21552	02207a_BB8-TUM_ mod_citrullination_ l_20_01_01-DDA-1h-R1	True	0.99	True	6	9	5	1
LTrDAQSPDESKT NEK	LTrDAQSPDESKT NEK	True	6579	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	True	0.99	True	2	14	2	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNL	nArtNL
SSGGGYSGDrSGG GYGGDR	SSGGGYSGDrSGG GYGGDR	True	11986	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	1.00	True	5	13	4	0
GGGYGGDrGGGYG GDR	GGGYGGDrGGGYG GDR	True	12990	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	True	1.00	True	5	10	5	0
rSPESLPAGPGAA ALEGGTR	rSPESLPAGPGAA ALEGGTR	True	32287	02207a_BB7-TUM_ mod_citrullination_ l_19_01_01-DDA-1h-R1	True	0.98	True	4	10	3	0
ENPrNFSDNQLQE GK	ENPrNFSDNQLQE GK	True	21467	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	Amb	0.98	True	0	6	0	0
MGAGLGHGMDrVG SEIER	MGAGLGHGMDrVG SEIER	True	30187	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	Likely	1.00	True	2	14	1	0
sNPrSLEEEKYDm SGAR	sNPrSLEEEKYDm SGAR	True	23417	02207a_BB3-TUM_ mod_citrullination_ l_15_01_01-DDA-1h-R1	True	0.97	True	0	20	5	0
TTEYQLSTLEErD IKK	TTEYQLSTLEErD IKK	True	30672	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	True	0.90	True	6	11	3	0
TAENFrALSTGEK	TAENFrALSTGEK	True	26464	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.98	True	2	14	7	1
SESrSASEHSSSA ESER	SESrSASEHSSSA ESER	True	3948	02207a_BB3-TUM_ mod_citrullination_ l_15_01_01-DDA-1h-R1	Likely	0.93	True	0	9	1	0
LTrGEADRDTYR	LTrGEADRDTYR	True	11326	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	Amb	1.00	True	0	3	0	0
DYSrGFGGR	DYSrGFGGR	True	19235	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	1.00	True	3	7	4	0
YLYLrNNQIDHID EK	YLYLrNNQIDHID EK	True	33054	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.98	True	0	15	6	0
DKDAYSSFGSrSD SR	DKDAYSSFGSrSD SR	True	18250	02207a_BA12-TUM_ mod_citrullination_ l_12_01_01-DDA-1h-R1	True	1.00	True	7	8	6	0
APQVLVLAPTrEL ANQVSK	APQVLVLAPTrEL ANQVSK	True	47052	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.99	True	0	23	6	0
DrDYSDHPSGGSY R	DrDYSDHPSGGSY R	True	10900	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	3	11	4	0
SHrSYScQVTHEG STVEK	SHrSYScQVTHEG STVEK	True	12181	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	Likely	0.99	True	1	7	1	0
SSGGGYSGDrSGG GYGGDR	SSGGGYSGDrSGG GYGGDR	True	12143	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.97	True	2	11	2	0
KATDNEPSQFSEP rK	KATDNEPSQFSEP rK	True	14689	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	True	1.00	True	4	12	5	0
aDFDTYDDrAYSS FGGGR	aDFDTYDDrAYSS FGGGR	True	42767	02207a_BB4-TUM_ mod_citrullination_ l_16_01_01-DDA-1h-R1	Likely	0.78	False	1	11	1	0
YDDYSSSrDGYGG SR	YDDYSSSrDGYGG SR	True	16409	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.99	True	7	12	9	0
LEVErDNLAQDLA TVR	LEVErDNLAQDLA TVR	True	42781	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.98	True	2	20	9	0
DKADrTQTGHVLG NPQR	DKADrTQTGHVLG NPQR	True	12503	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	Likely	0.81	True	0	10	1	0
SrGSGGLGGAcGG AGFGSR	SrGSGGLGGAcGG AGFGSR	True	22289	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	True	1.00	True	12	17	8	0
QEMQEVQSSrSGR	QEMQEVQSSrSGR	True	10670	02207a_BA7-TUM_ mod_citrullination_ l_7_01_01-DDA-1h-R1	True	1.00	True	3	10	8	0
TAQLErSLQEEHV AVAQLR	TAQLErSLQEEHV AVAQLR	True	36953	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	Amb	0.95	True	0	14	0	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNL	nArtNL
SGGGYGGDrGGGY GGDR	SGGGYGGDrGGGY GGDR	True	12023	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	5	13	5	0
SrAEAESWYQTK	SrAEAESWYQTK	True	25825	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	True	1.00	True	5	10	4	0
EDSLEAGLPLQVr GYPEEK	EDSLEAGLPLQVr GYPEEK	True	43378	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.99	True	2	15	3	0
GGTSrEGGQTAPA STR	GGTSrEGGQTAPA STR	True	5517	02207a_BA12-TUM_ mod_citrullination_ l_12_01_01-DDA-1h-R1	True	0.99	True	3	16	4	0
AQLrELNITAAK	AQLrELNITAAK	True	32606	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	0	14	5	0
SSGGSYrDSYDSY ATHNE	SSGGSYrDSYDSY ATHNE	True	22050	02207a_BD3-TUM_ mod_citrullination_ s_3_01_01-DDA-1h-R1	Amb	0.96	True	0	7	0	0
GGSrGEEVGELSR	GGSrGEEVGELSR	True	20171	02207a_BA7-TUM_ mod_citrullination_ l_7_01_01-DDA-1h-R1	Likely	0.98	True	1	9	1	0
VIVDKYrDGTK	VIVDKYrDGTK	True	13928	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	2	5	2	0
TVEMrDGEVIK	TVEMrDGEVIK	True	24613	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	4	10	8	0
DYTYrDYGHSSSR	DYTYrDYGHSSSR	True	17957	02207a_BA8-TUM_ mod_citrullination_ l_8_01_01-DDA-1h-R1	True	0.99	True	4	11	7	0
GIAGrQDILDDSG YVSAYK	GIAGrQDILDDSG YVSAYK	True	38875	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	True	0.99	True	1	19	4	0
EEAPSLrPAPPPI SGGGYR	EEAPSLrPAPPPI SGGGYR	True	33391	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	True	0.98	True	3	14	4	0
TGTYrQLFHPEQL ITGK	TGTYrQLFHPEQL ITGK	True	42389	02207a_BB3-TUM_ mod_citrullination_ l_15_01_01-DDA-1h-R1	True	0.99	True	0	16	3	0
LGrQGNLSSSGNQ EGSQK	LGrQGNLSSSGNQ EGSQK	True	10841	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	True	0.92	True	1	17	2	0
YQDDFLLDESEcr VVK	YQDDFLLDESEcr VVK	True	41854	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.87	True	7	7	5	0
EDHGrGYFEYIEE NK	EDHGrGYFEYIEE NK	True	34667	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	Likely	0.80	False	0	11	1	0
SGGGYGGDrSGGG YGGDR	SGGGYGGDrSGGG YGGDR	True	12449	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.98	True	5	10	4	0
TTEYQLSTLEErD IK	TTEYQLSTLEErD IK	True	35499	02207a_BA11-TUM_ mod_citrullination_ l_11_01_01-DDA-1h-R1	True	0.99	True	7	14	9	0
SSGGSYrDSYDSY ATHNE	SSGGSYrDSYDSY ATHNE	True	21466	02207a_BB10-TUM_ mod_citrullination_ l_22_01_01-DDA-1h-R1	True	0.99	True	0	15	2	0
ITPGArGAFSEEY KR	ITPGArGAFSEEY KR	True	22906	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	1.00	True	4	10	3	0
AAQDrDQIYR	AAQDrDQIYR	True	16675	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.99	True	0	11	5	0
GVVDSEDLPLNIS rEMLQQSK	GVVDSEDLPLNIS rEMLQQSK	True	51608	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	True	0.95	True	0	19	3	0
HSATGrGQASSAV SDR	HSATGrGQASSAV SDR	True	6812	02207a_BA12-TUM_ mod_citrullination_ l_12_01_01-DDA-1h-R1	True	1.00	True	2	15	4	0
SGGYGGSrDYYSS R	SGGYGGSrDYYSS R	True	17850	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	True	0.98	True	4	11	4	0
SGGGYGGDrSSGG GYSGDR	SGGGYGGDrSSGG GYSGDR	True	11709	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	True	0.99	True	4	15	5	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNL	nArtNL
SGGGYGGDrSGGGYGGDR	SGGGYGGDrSGGGYGGDR	True	11694	02207a_BB4-TUM_mod_citrullination_l_16_01_01-DDA-1h-R1	True	0.97	True	6	15	7	0
AKPALEDLrQGLLPVLESFK	AKPALEDLRqGLLPVLESFK	False	57372	02207a_BB7-TUM_mod_citrullination_l_19_01_01-DDA-1h-R1	Amb	0.99	True	0	4	0	0
EDHGrGYFEYIEENK	EDHGRGYFEYIEEnK	False	31936	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	Amb	0.99	True	0	1	0	0
ILNEMrDQYEK	ILNEMRDqYEK	False	22741	02207a_BA4-TUM_mod_citrullination_l_4_01_01-DDA-1h-R1	Amb	1.00	True	0	14	0	0
SKEEEQQrLINDLTAQR	SKEEEQQRLINDLTAqR	False	30437	02207a_BB3-TUM_mod_citrullination_l_15_01_01-DDA-1h-R1	Amb	1.00	True	0	3	0	0
IETIEVMEDrQSGK	IETIEVMEDRqSGK	False	31899	02207a_BA9-TUM_mod_citrullination_l_9_01_01-DDA-1h-R1	Amb	1.00	True	0	2	0	0
AHAAIrENPVYEK	AHAAIREnPVYEK	False	18332	02207a_BA7-TUM_mod_citrullination_l_7_01_01-DDA-1h-R1	True	0.99	True	2	18	10	0
ALSrQLSSGVSEIR	ALSRqLSSGVSEIR	False	30063	02207a_BA9-TUM_mod_citrullination_l_9_01_01-DDA-1h-R1	True	0.98	True	1	15	2	0
ESGSEQErVSK	ESGSEqERVSK	False	4973	02207a_BA4-TUM_mod_citrullination_l_4_01_01-DDA-1h-R1	True	0.99	True	2	9	5	0
DDEFrKNFDK	DDEFRKnFDK	False	15889	02207a_BA3-TUM_mod_citrullination_l_3_01_01-DDA-1h-R1	Amb	1.00	True	0	7	0	0
ErNGLSLAALK	ERnGLSLAALK	False	34972	02207a_BA4-TUM_mod_citrullination_l_4_01_01-DDA-1h-R1	Amb	1.00	True	0	10	0	0
LGrQGNLSSSGNQEGSQK	LGRqGNLSSSGNQEGSQK	False	11134	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	Amb	0.99	True	0	5	0	0
DVTILAATNrPDR	DVTILAATnRPDR	False	30196	02207a_BA8-TUM_mod_citrullination_l_8_01_01-DDA-1h-R1	Amb	1.00	True	0	3	0	0
YAFDFArQSK	YAFDFARqSK	False	35297	02207a_BA3-TUM_mod_citrullination_l_3_01_01-DDA-1h-R1	True	1.00	True	0	14	6	0
rSYESVEGNIKQGmSmR	RSYESVEGnIKQGmSmR	False	19361	02207a_BB2-TUM_mod_citrullination_l_14_01_01-DDA-1h-R1	Likely	0.99	True	1	6	0	0
DDEFrKNFDK	DDEFRKnFDK	False	20483	02207a_BA2-TUM_mod_citrullination_l_2_01_01-DDA-1h-R1	Amb	1.00	True	0	7	0	0
SGQAQErQESHEQEVNR	SGQAQERqESHEQEVNR	False	5517	02207a_BD2-TUM_mod_citrullination_s_2_01_01-DDA-1h-R1	Amb	0.97	True	0	11	0	0
SYLSrLMLANR	SYLSRLMLAnR	False	45765	02207a_BA4-TUM_mod_citrullination_l_4_01_01-DDA-1h-R1	Amb	0.97	True	0	2	0	0
VQDNHDPPNQSGLQrFVIR	VQDNHDPPNQSGLqRFVIR	False	33276	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	Amb	0.98	True	0	5	0	0
SGTASVVcLLNNFYPrEAK	SGTASVVcLLNnFYPREAK	False	51701	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	Amb	0.91	True	0	3	0	0
DYNGrNQGGYDR	DYnGRNQGGYDR	False	9397	02207a_BA7-TUM_mod_citrullination_l_7_01_01-DDA-1h-R1	Amb	1.00	True	0	8	0	0
rAPPPPGSGPPVQDK	RAPPPPGSGPPVqDK	False	11796	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	Amb	1.00	True	0	2	0	0
ErQELGSPEER	ERqELGSPEER	False	14414	02207a_BA1-TUM_mod_citrullination_l_1_01_01-DDA-1h-R1	Amb	0.99	True	0	6	0	0
EKAENrVVQIEK	EKAEnRVVQIEK	False	15204	02207a_BA6-TUM_mod_citrullination_l_6_01_01-DDA-1h-R1	Likely	0.99	True	0	8	1	0
SGrDPNHFRPAGLPEK	SGRDPnHFRPAGLPEK	False	18762	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	Amb	0.95	True	0	3	0	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNI	nArtNL
ATMQNLNDrLASY LDK	ATMQNLnDRLASY LDK	False	46872	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.99	True	1	1	2	0
ISELIErNQR	ISELIERnQR	False	23509	02207a_BA6-TUM_ mod_citrullination_ l_6_01_01-DDA-1h-R1	Likely	0.99	True	1	4	1	0
DITSDTSGDFrNA LLSLAK	DITSDTSGDFRnA LLSLAK	False	56941	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	Amb	1.00	True	0	9	0	0
DnTYNrVSEDLR	DNTYnRVSEDLR	False	30104	02207a_BD1-TUM_ mod_citrullination_ s_1_01_01-DDA-1h-R1	Likely	0.99	True	1	7	1	0
DDEFrKNFDK	DDEFRKnFDK	False	20170	02207a_BA2-TUM_ mod_citrullination_ l_2_01_01-DDA-1h-R1	Amb	1.00	True	0	7	0	0
NSLLLMQrK	NSLLLMqRK	False	30062	02207a_BA1-TUM_ mod_citrullination_ l_1_01_01-DDA-1h-R1	True	0.99	True	1	9	3	1
SYrYALQDmDK	SYRYALqDmDK	False	20036	02207a_BA3-TUM_ mod_citrullination_ l_3_01_01-DDA-1h-R1	Amb	1.00	True	0	8	0	0
rYYFLSEEAEEGD ELQQR	RYYFLSEEAEEGD ELqQR	False	32860	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	Likely	0.92	True	1	2	0	0
QEYDESGPSIVHr K	qEYDESGPSIVHR K	False	16205	02207a_BA7-TUM_ mod_citrullination_ l_7_01_01-DDA-1h-R1	Likely	0.99	True	1	2	0	0
AKPALEDLrQGLL PVLESFK	AKPALEDLRqGLL PVLESFK	False	56859	02207a_BB7-TUM_ mod_citrullination_ l_19_01_01-DDA-1h-R1	Amb	0.98	True	0	10	0	0
LrVLQFIR	LRVLqFIR	False	42374	02207a_BA1-TUM_ mod_citrullination_ l_1_01_01-DDA-1h-R1	Likely	1.00	True	1	6	0	0
DnTYNrVSEDLR	DnTYNRVSEDLR	False	26621	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	Amb	0.94	False	0	5	0	0
QrLEELQLFK	qRLEELQLFK	False	46146	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	Amb	0.99	True	0	8	0	0
LTGITYQALrPDP LR	LTGITYqALRPDP LR	False	38941	02207a_BA11-TUM_ mod_citrullination_ l_11_01_01-DDA-1h-R1	Amb	0.96	True	0	5	0	0
QTQEATNDQNrGT ETHGQGR	QTqEATNDQNRGT ETHGQGR	False	2916	02207a_BB9-TUM_ mod_citrullination_ l_21_01_01-DDA-1h-R1	Amb	0.97	True	0	1	0	0
TAQLErSLQEEHV AVAQLR	TAQLERSLQEEHV AVAqLR	False	32717	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	Amb	0.98	True	0	7	0	0
LMEEcDELVEIIQ QrK	LMEEcDELVEIIQ qRK	False	47188	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	True	0.83	True	1	17	2	0
YAFDFArQSK	YAFDFARqSK	False	35592	02207a_BA3-TUM_ mod_citrullination_ l_3_01_01-DDA-1h-R1	True	0.97	True	0	12	4	0
DGPPLrGSNMDFR EPTEEER	DGPPLRGSnMDFR EPTEEER	False	27900	02207a_BB7-TUM_ mod_citrullination_ l_19_01_01-DDA-1h-R1	Amb	0.98	True	0	7	0	0
LTMQNLNDrLASY LDK	LTMQnLNDRLASY LDK	False	48895	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	Amb	0.99	True	0	5	0	0
QrLEELQLFK	qRLEELQLFK	False	41481	02207a_BA2-TUM_ mod_citrullination_ l_2_01_01-DDA-1h-R1	Amb	0.99	True	0	11	0	0
NDGKLYDAYVVYP rNYK	NDGKLYDAYVVYP RnYK	False	34324	02207a_BB3-TUM_ mod_citrullination_ l_15_01_01-DDA-1h-R1	Likely	0.89	True	0	10	1	0
SKEEEQQrLINDL TAQR	SKEEEQqRLINDL TAQR	False	31402	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	Amb	0.98	True	0	4	0	0
DDEFrKNFDK	DDEFRKnFDK	False	21437	02207a_BA2-TUM_ mod_citrullination_ l_2_01_01-DDA-1h-R1	Amb	1.00	True	0	7	0	0
DITSDTSGDFrNA LLSLAK	DITSDTSGDFRnA LLSLAK	False	56909	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	Likely	0.97	True	0	16	1	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNI	nArtNL
WSGSASrNHLGSAWEQSR	WSGSASRnHLGSAWEQSR	False	22655	02207a_BB9-TUM_mod_citrullination_l_21_01_01-DDA-1h-R1	Likely	1.00	True	1	11	0	0
GELDcHQLADSFrE	GELDcHqLADSFRE	False	31280	02207a_BB10-TUM_mod_citrullination_l_22_01_01-DDA-1h-R1	Amb	0.97	True	0	10	0	0
rAPPPPGSGPPVQDK	RAPPPPGSGPPVqDK	False	12569	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	Likely	0.99	True	1	2	0	0
rAPPPPGSGPPVQDK	RAPPPPGSGPPVqDK	False	12104	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	False	1.00	True	0	0	0	0
mETLSNASGTFAIrLLK	mETLSnASGTFAIRLLK	False	50827	02207a_BB3-TUM_mod_citrullination_l_15_01_01-DDA-1h-R1	Likely	0.44	False	1	1	0	0
VLLTVQrK	VLLTVqRK	False	21089	02207a_BA1-TUM_mod_citrullination_l_1_01_01-DDA-1h-R1	Amb	1.00	True	0	10	0	0
rYYFLSEEAEEGDELQQR	RYYFLSEEAEEGDELQqR	False	36132	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	Likely	0.99	True	1	1	0	0
rLEALAAEFSSNWQK	RLEALAAEFSSNWqK	False	42591	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	Amb	0.98	True	0	1	0	0
NQPArFKNYAELTIMK	NQPARFKnYAELTIMK	False	34336	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	Amb	0.98	True	0	11	0	0
QAGDLrSVIR	qAGDLRSVIR	False	25856	02207a_BA3-TUM_mod_citrullination_l_3_01_01-DDA-1h-R1	Amb	0.99	True	0	7	0	0
ArYnAqcqETIR	ARYNAQcqETIR	False	14390	02207a_BA6-TUM_mod_citrullination_l_6_01_01-DDA-1h-R1	Amb	0.81	False	0	6	0	0
rFPEVSGYcHcADSQGR	RFPEVSGYcHcADSqGR	False	21960	02207a_BB2-TUM_mod_citrullination_l_14_01_01-DDA-1h-R1	Amb	0.89	True	0	2	0	0
ErQELGSPEER	ERqELGSPEER	False	11317	02207a_BA1-TUM_mod_citrullination_l_1_01_01-DDA-1h-R1	Likely	1.00	True	1	13	0	0
YGVQADrVDK	YGVqADRVDK	False	14209	02207a_BA3-TUM_mod_citrullination_l_3_01_01-DDA-1h-R1	Amb	0.96	True	0	8	0	0
DDEFrKNFDK	DDEFRKnFDK	False	32249	02207a_BA2-TUM_mod_citrullination_l_2_01_01-DDA-1h-R1	Amb	0.96	True	0	4	0	0
SGrDPNHFRPAGLPEK	SGRDPnHFRPAGLPEK	False	22310	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	Amb	0.93	True	0	3	0	0
ALArEVDLKDYEDQQK	ALAREVDLKDYEDqQK	False	25583	02207a_BB1-TUM_mod_citrullination_l_13_01_01-DDA-1h-R1	Likely	1.00	True	1	3	0	0
rcEScAPGYEGNPIQPGGK	RcEScAPGYEGNPIqPGGK	False	20813	02207a_BB5-TUM_mod_citrullination_l_17_01_01-DDA-1h-R1	False	0.99	True	0	0	0	0
rTIAQDYGVLK	RTIAqDYGVLK	False	29418	02207a_BA3-TUM_mod_citrullination_l_3_01_01-DDA-1h-R1	Amb	1.00	True	0	11	0	0
TEELNrEVAGHTEQLQMSR	TEELnREVAGHTEQLQMSR	False	26556	02207a_BB7-TUM_mod_citrullination_l_19_01_01-DDA-1h-R1	Amb		False	0	13	0	0
rEPWLLPSQHNDIIR	REPWLLPSQHnDIIR	False	36718	02207a_BA11-TUM_mod_citrullination_l_11_01_01-DDA-1h-R1	Amb	0.99	True	0	4	0	0
QrIAEETILK	qRIAEETILK	False	25999	02207a_BA2-TUM_mod_citrullination_l_2_01_01-DDA-1h-R1	Amb	0.99	True	0	16	0	0
SGTASVVcLLNNFYPrEAK	SGTASVVcLLNnFYPREAK	False	52133	02207a_BB6-TUM_mod_citrullination_l_18_01_01-DDA-1h-R1	True	0.96	True	3	12	6	0
WSGSASrNHLGSAWEQSR	WSGSASRnHLGSAWEQSR	False	22491	02207a_BB9-TUM_mod_citrullination_l_21_01_01-DDA-1h-R1	Amb	0.99	True	0	10	0	0
mETLSNASGTFAIrLLK	mETLSnASGTFAIRLLK	False	57647	02207a_BB3-TUM_mod_citrullination_l_15_01_01-DDA-1h-R1	Likely	-0.11	False	3	1	0	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNInArtNL	
DGPPLrGSNMDFR EPTEEER	DGPPLRGSnMDFR EPTEEER	False	28048	02207a_BB7-TUM_ mod_citrullination_ l_19_01_01-DDA-1h-R1	Amb	0.99	True	0	8	0	0
GLArLALQFTTNP K	GLARLALQFTTnP K	False	42216	02207a_BA8-TUM_ mod_citrullination_ l_8_01_01-DDA-1h-R1	Amb	0.98	True	0	2	0	0
ISELIErNQR	ISELIERNqR	False	23976	02207a_BA6-TUM_ mod_citrullination_ l_6_01_01-DDA-1h-R1	True	1.00	True	0	13	5	0
SKEEEQQrLINDL TAQR	SKEEEQqRLINDL TAQR	False	35520	02207a_BB3-TUM_ mod_citrullination_ l_15_01_01-DDA-1h-R1	Amb	0.99	True	0	5	0	0
VQDNHDPPNQSGL QrFVIR	VQDNHDPPnQSGL QRFVIR	False	32907	02207a_BB9-TUM_ mod_citrullination_ l_21_01_01-DDA-1h-R1	Amb	0.97	True	0	5	0	0
ELEDrFASEASGY QDNIAR	ELEDRFASEASGY QDnIAR	False	32539	02207a_BB6-TUM_ mod_citrullination_ l_18_01_01-DDA-1h-R1	Likely	0.96	True	2	4	0	0
DTASSRcQScSEr EEAGK	DTASSRcqScSER EEAGK	False	4973	02207a_BB4-TUM_ mod_citrullination_ l_16_01_01-DDA-1h-R1	Amb	0.97	True	0	6	0	0
LMEEcDELVEIIQ QrK	LMEEcDELVEIIQ qRK	False	45074	02207a_BB1-TUM_ mod_citrullination_ l_13_01_01-DDA-1h-R1	Amb	0.94	True	0	1	0	0
MAEVGYDrNNK	MAEVGYDRnNK	False	12406	02207a_BA3-TUM_ mod_citrullination_ l_3_01_01-DDA-1h-R1	Amb	1.00	True	0	17	0	0
TSQPNYrPQGmEG FLKSDER	TSQPnYRPQGmEG FLKSDER	False	22758	02207a_BB9-TUM_ mod_citrullination_ l_21_01_01-DDA-1h-R1	Amb	0.98	True	0	9	0	0
rYYFLSEEAEEGD ELQQR	RYYFLSEEAEEGD ELQqR	False	35252	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	Amb	0.92	True	0	1	0	0
GGGGNFGPGPGSN FrGGSDGYGSGR	GGGGNFGPGPGSn FRGGSDGYGSGR	False	27417	02207a_BD2-TUM_ mod_citrullination_ s_2_01_01-DDA-1h-R1	Likely	0.98	True	1	8	0	0
DSEINrLRSELK	DSEInRLRSELK	False	24501	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	Amb	0.98	True	0	4	0	0
rSYESVEGNIKQG mSmR	RSYESVEGnIKQG mSmR	False	19663	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	Amb	0.99	True	0	5	0	0
rLYQSTLLQDGLK	RLYQSTLLqDGLK	False	34699	02207a_BA7-TUM_ mod_citrullination_ l_7_01_01-DDA-1h-R1	Amb	1.00	True	0	1	0	0
VGVSSrINEWLTK	VGVSSRInEWLTK	False	44135	02207a_BA7-TUM_ mod_citrullination_ l_7_01_01-DDA-1h-R1	Likely	0.99	True	0	3	1	0
ITMQNLNDrLASY LDK	ITMqNLNDRLASY LDK	False	44155	02207a_BB9-TUM_ mod_citrullination_ l_21_01_01-DDA-1h-R1	Amb	0.99	True	0	10	0	0
QWTSLSLrVTAK	qWTSLSLRVTAK	False	36711	02207a_BA5-TUM_ mod_citrullination_ l_5_01_01-DDA-1h-R1	Amb	0.93	True	0	4	0	0
DYSGYQrDGYQQN FK	DYSGYqRDGYQQN FK	False	25947	02207a_BA11-TUM_ mod_citrullination_ l_11_01_01-DDA-1h-R1	Likely	0.98	True	0	3	1	0
rEALVKGNPPIYR	REALVKGnPPIYR	False	21146	02207a_BA8-TUM_ mod_citrullination_ l_8_01_01-DDA-1h-R1	Amb	0.99	True	0	6	0	0
LLErPPEFTLPLY NK	LLERPPEFTLPLY nK	False	46516	02207a_BA11-TUM_ mod_citrullination_ l_11_01_01-DDA-1h-R1	Amb	0.98	True	0	3	0	0
VNNLISDQDFFAL rSIK	VNNLISDqDFFAL RSIK	False	48718	02207a_BB2-TUM_ mod_citrullination_ l_14_01_01-DDA-1h-R1	Amb	0.99	True	0	10	0	0
AALEALLGrLQAE R	AALEALLGRLqAE R	False	51189	02207a_BA8-TUM_ mod_citrullination_ l_8_01_01-DDA-1h-R1	Likely	1.00	True	1	3	0	0
LLGrSQAIQVGTS WK	LLGRSqAIQVGTS WK	False	38724	02207a_BA9-TUM_ mod_citrullination_ l_9_01_01-DDA-1h-R1	True	0.97	True	1	17	4	0
IGArVYALPEDLV EVNPK	IGARVYALPEDLV EVnPK	False	43405	02207a_BB4-TUM_ mod_citrullination_ l_16_01_01-DDA-1h-R1	Likely	0.99	True	1	2	0	0

Continued on next page

Table 5A.1 – continued from previous page

Sequence	Reference sequence	Ground truth	Scan	Parent file	Contains Cit	Env. score	Good env.	nDet	nAmb	nDetNL	nArtNL
GWLrDPSASPGDA GEQAIR	GWLRDPSASPGDA GEqAIR	False	30924	02207a_BB5-TUM_ mod_citrullination_ l_17_01_01-DDA-1h-R1	False	0.98	True	0	0	0	0

Table 5A.2: Unique sites of citrullination identified in the HEK-PAD2 citrullinome.

Accession	Protein	Description	Sequence	Residue	Spectral counts
B5ME19	EIFCL	Eukaryotic translation initiation factor 3 subunit C-like protein	QTALASR*FLK	R255	2
O00193	SMAP	Small acidic protein	FLR*LMGAGK	R43	3
O00422	SAP18	Histone deacetylase complex subunit SAP18	AVESR*VTQEEIKK	R6	1
O00422	SAP18	Histone deacetylase complex subunit SAP18	MAVESR*VTQEEIK	R6	1
O00422	SAP18	Histone deacetylase complex subunit SAP18	AVESR*VTQEEIK	R6	2
O00567	NOP56	Nucleolar protein 56	EAMVQAEEAAAEITR*K	R437	1
O00571	DDX3X	ATP-dependent RNA helicase DDX3X	SDYDGIGSR*GDR*SGFGK	R110|R113	1
O14654	IRS4	Insulin receptor substrate 4	LYFCVDR*GATK	R661	2
O14787-2	TNPO2	Isoform 2 of Transportin-2	GDVEEDEAVPDSEQDIKPR*F HK	R336	1
O14979	HNRDL	Heterogeneous nuclear ribonucleoprotein D-like	LLESR*YHQIGSGK	R294	2
O14979-2	HNRDL	Isoform 2 of Heterogeneous nuclear ribonucleoprotein D-like	LLESR*YHQIGSGK	R175	2
O15042-2	SR140	Isoform 2 of U2 snRNP-associated SURP motif-containing protein	TFVR*GGVVNAAKEEHETDEK	R137	1
O15042-2	SR140	Isoform 2 of U2 snRNP-associated SURP motif-containing protein	TFVR*GGVVNAAK	R137	2
O43290	SNUT1	U4/U6.U5 tri-snRNP-associated protein 1	ADDLLPLGDQTQDGDFGSR*L R	R438	1
O43809	CPSF5	Cleavage and polyadenylation specificity factor subunit 5	SVVPPNR*SQTGWPR	R8	1
O60216	RAD21	Double-strand-break repair protein rad21 homolog	DVIDEPIIEEPSR*LQESVME ASR	R450	1
O60664-4	PLIN3	Isoform 4 of Perilipin-3	DTVATQLSEAVDATR*GAVQS GVDK	R143	1
O60749	SNX2	Sorting nexin-2	AVNTQALSGAGILR*MVNK	R283	2
O60763	USO1	General vesicular transport factor p115	MNFLR*GVMGGQSAGPQHTEA ETIQK	R5	2
O60841	IF2P	Eukaryotic translation initiation factor 5B	ADR*ETVAVKPTENNEEEFTS K	R76	1
O75521	ECI2	Enoyl-CoA delta isomerase 2, mitochondrial	WLSDECTNAVVNFLSR*K	R390	1
O75533	SF3B1	Splicing factor 3B subunit 1	MNAR*TYMDVMR	R150	2
O75533	SF3B1	Splicing factor 3B subunit 1	IWDPTPSHTPAGAATPGR*GD TPGHATPGHGGATSSAR	R270	2
O75534-3	CSDE1	Isoform 3 of Cold shock domain-containing protein E1	AVAAPRPDR*LVNR	R767	1
O75643	U520	U5 small nuclear ribonucleoprotein 200 kDa helicase	ADVTAR*SLQYEYK	R7	2
O75663	TIPRL	TIP41-like protein	VACAEEWQESR*TEGEHSK	R95	1
O75940	SPF30	Survival of motor neuron-related-splicing factor 30	WQQFNNR*AYSK	R184	1
O94776	MTA2	Metastasis-associated protein MTA2	GTTEPHSRGHLSR*PEAQSLS PYTTSANR	R428	1
O94826	TOM70	Mitochondrial import receptor subunit TOM70	TPEGR*ASPAPGSGHPEGPGA HLDMNSLDR	R89	1
O95218-2	ZRAB2	Isoform 2 of Zinc finger Ran-binding domain-containing protein 2	SR*GLFSANDWQCK	R61	1
O95218-2	ZRAB2	Isoform 2 of Zinc finger Ran-binding domain-containing protein 2	NFR*VSDGDWICPDKK	R7	1
O95232	LC7L3	Luc7-like protein 3	MISAAQLLDELMGR*DR	R14	1
O95757	HS74L	Heat shock 70 kDa protein 4L	SAVSDKQDR*LNQTLK	R567	1
P02769	ALBU	Serum albumin	AWSVAR*LSQK	R241	2
P04350	TBB4A	Tubulin beta-4A chain	MR*EIVHLQAGQCGNQIGAK	R2	2
P04350	TBB4A	Tubulin beta-4A chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
P06576	ATPB	ATP synthase subunit beta, mitochondrial	VAAAPASGALR*R	R18	1
P07437	TBB5	Tubulin beta chain	GHYTEGAELVDSVLDVVR*K	R121	4
P07437	TBB5	Tubulin beta chain	MR*EIVHIQAGQCGNQIGAK	R2	2
P07437	TBB5	Tubulin beta chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
P07900-2	HS90A	Isoform 2 of Heat shock protein HSP 90-alpha	GVVDSEDLPLNISR*EMLQQS K	R522	6
P07900-2	HS90A	Isoform 2 of Heat shock protein HSP 90-alpha	AQALR*DNSTMGYMAAK	R742	2
P07910-2	HNRPC	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	DYYDR*MYSYPAR	R122	3
P07910-2	HNRPC	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	VPPPPPIAR*AVVPSKR	R138	4
P07910-2	HNRPC	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	VPPPPPIAR*AVVPSK	R138	7
P07910-2	HNRPC	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	R*SAAEMYGSSFDLDYDFQR	R99	1
P08238	HS90B	Heat shock protein HSP 90-beta	GVVDSEDLPLNISR*EMLQQS K	R392	6
P08238	HS90B	Heat shock protein HSP 90-beta	LVSSPCCIVTSTYGWTANMERIMK	R604	1
P08238	HS90B	Heat shock protein HSP 90-beta	AQALR*DNSTMGYMMAK	R612	2

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
P08243-2	ASNS	Isoform 2 of Asparagine synthetase [glutamine-hydrolyzing]	WINATDPSAR*TLTHYK	R529	2
P08621	RU17	U1 small nuclear ribonucleoprotein 70 kDa	DDTSRYDER*PGPSPLPHR	R222	1
P08621	RU17	U1 small nuclear ribonucleoprotein 70 kDa	YDER*PGPSPLPHR	R222	1
P08670	VIME	Vimentin	NTR*TNEKVELQELNDR	R100	1
P08670	VIME	Vimentin	SR*LGDLYEEEMR	R145	1
P08670	VIME	Vimentin	DVR*QQYESVAAK	R273	1
P08670	VIME	Vimentin	FADLSEAANR*NNDALR	R304	5
P08670	VIME	Vimentin	FADLSEAANRNNDALR*QAK	R310	1
P08670	VIME	Vimentin	LLEGEESR*ISLPLPNFSSLN LR	R410	1
P08670	VIME	Vimentin	ISLPLPNFSSLNLR*ETNLDS LPLVDTHSK	R424	1
P08670	VIME	Vimentin	TVETR*DGQVINETSQHHDDL E	R450	5
P08670	VIME	Vimentin	SSAVR*LR*SSVPGVR	R69|R71	1
P09651-2	ROA1	Isoform A1-A of Heterogeneous nuclear ribonucleoprotein A1	IEVIEIMTDR*GSGK	R140	3
P09651-2	ROA1	Isoform A1-A of Heterogeneous nuclear ribonucleoprotein A1	GGNFGGR*SSGPYGGGGQYFA KPR	R284	2
P09651-2	ROA1	Isoform A1-A of Heterogeneous nuclear ribonucleoprotein A1	SSGPYGGGGQYFAKPRNQGGY GGSSSSSSYGSGR*R	R318	3
P10809	CH60	60 kDa heat shock protein, mitochondrial	LVQDVANNTNEEAGDGTTTAT VLAR*SIAK	R121	2
P10809	CH60	60 kDa heat shock protein, mitochondrial	FDR*GYISPYFINTSK	R221	1
P12268	IMDH2	Inosine-5'-monophosphate dehydrogenase 2	LPIVNEDDELVAIIAR*TDLK	R224	1
P13639	EF2	Elongation factor 2	FSVSPVVR*VAVEAK	R506	2
P14866	HNRPL	Heterogeneous nuclear ribonucleoprotein L	IEYAKPTR*LNVFK	R272	2
P14866	HNRPL	Heterogeneous nuclear ribonucleoprotein L	MAAAGGGGGGGR*YYGGGSEG GR*APK	R46|R56	1
P15311	EZRI	Ezrin	EKEELMLR*LQDYEEK	R350	1
P15311	EZRI	Ezrin	EELER*QAVDQIK	R405	1
P16104	H2AX	Histone H2AX	SSR*AGLQFPVGR	R21	1
P16615	AT2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	SMSVYCTPNKPSR*TSMSK	R505	3
P17844	DDX5	Probable ATP-dependent RNA helicase DDX5	ENYDR*GYSSLLK	R516	3
P20338	RAB4A	Ras-related protein Rab-4A	LQIWDTAGQER*FR	R74	2
P20340-2	RAB6A	Isoform 2 of Ras-related protein Rab-6A	LQLWDTAGQER*FR	R74	1
P20700	LMNB1	Lamin-B1	MGSR*AGGPTTPLSPTR	R14	3
P20700	LMNB1	Lamin-B1	SMYEEEINETR*R	R220	1
P20700	LMNB1	Lamin-B1	LENAR*LSSEMNTSTVNSAR	R276	1
P20700	LMNB1	Lamin-B1	LSPSPSSR*VTVSR	R397	5
P20700	LMNB1	Lamin-B1	LSPSPSSR*VTVSR*ASSSR	R397|R402	1
P22626	ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	IDTIEIITDR*QSGK	R147	3
P22626	ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	GGGGNFGPGPGSNFR*GGSDG YGSGR	R228	1
P22626	ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	SGNFGGSR*NMGGPYGGGNYG PGGSGGSGGYGGR	R325	2
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	ANLSLLR*RPGEK	R286	2
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	DKLESEMEDAYHEHQANLLR* QDLMR	R536	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	QREESYSR*MGYMDPR	R599	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	EESYSR*MGYMDPR	R599	2
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	ER*DMRMGGGGAMNMGDPYGS GGQK	R608	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	ER*DMR*MGGGGAMNMGDPYG SGGQK	R608|R611	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	ERDMR*MGGGGAMNMGDPYGS GGQK	R611	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	FPPLGGGGGIGYEANPGVPPA TMSGSMMGSDMR*TER	R663	1
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	FGQGGAGPVGGQGPR*GMGPG TPAGYGR	R681	2
P23588	IF4B	Eukaryotic translation initiation factor 4B	EDDSSASTSQSTR*AASIFGG AKPVDTAAR	R356	3
P26373	RL13	60S ribosomal protein L13	LATQLTGPVMPVR*NVYK	R158	1
P26373	RL13	60S ribosomal protein L13	ANAR*LFGIR	R190	1
P27635	RL10	60S ribosomal protein L10	LQTGMR*GAFGKPQGTVAR	R116	1
P27635	RL10	60S ribosomal protein L10	R*LIPDGCGVK	R189	1
P27635	RL10	60S ribosomal protein L10	IR*IFDLGR	R32	1
P27708	PYR1	CAD protein	IHR*ASDPGLPAEEPK	R1857	1
P27816	MAP4	Microtubule-associated protein 4	TTTLSGTAPAAGVVPSR*VK	R888	1
P27816	MAP4	Microtubule-associated protein 4	LSR*LATNTSAPDLK	R922	3
P29692-2	EF1D	Isoform 2 of Elongation factor 1-delta	ATAPQTQHVSPMR*QVEPPAK	R502	1
P29692-3	EF1D	Isoform 3 of Elongation factor 1-delta	ATAPQTQHVSPMR*QVEPPAK	R112	1

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
P31350	RIR2	Ribonucleoside-diphosphate reductase subunit M2	GLSLVDKENTPPALSGTR*VLASK	R41	1
P31943	HNRH1	Heterogeneous nuclear ribonucleoprotein H	FGR*DLNYCFSGMSDHR	R262	2
P32322-3	P5CR1	Isoform 3 of Pyrroline-5-carboxylate reductase 1, mitochondrial	LLPR*SLAPAGKD	R338	1
P33992	MCM5	DNA replication licensing factor MCM5	VGTDR*TGFTFK	R48	2
P33993	MCM7	DNA replication licensing factor MCM7	GQTAR*TQRPADVIFATVR	R653	1
P35579	MYH9	Myosin-9	ELEDATETADAMNR*EVSSLK	R1912	1
P35580-3	MYH10	Isoform 3 of Myosin-10	AQR*TGLEDPER	R4	1
P36551	HEM6	Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	TSGTR*ATSLGRPEEEEDELAHR	R109	1
P36578	RL4	60S ribosomal protein L4	GHR*IEEVPELPLVVEDK	R143	1
P36578	RL4	60S ribosomal protein L4	MINTDLSR*ILK	R291	2
P37802	TAGL2	Transgelin-2	GPAYGLSR*EVQQK	R12	2
P38159	RBMX	RNA-binding motif protein, X chromosome	GPPSR*GGHMDDGGYSMNFNMSSSR*GPLPVK	R125|R144	4
P40227	TCPZ	T-complex protein 1 subunit zeta	GLQDVLR*TNLGPK	R35	1
P42166	LAP2A	Lamina-associated polypeptide 2, isoform alpha	LREQGTESR*SSTPLPTISSSAENTR	R157	1
P42166	LAP2A	Lamina-associated polypeptide 2, isoform alpha	LASER*NLFISCK	R274	1
P42167	LAP2B	Lamina-associated polypeptide 2, isoforms beta/gamma	LREQGTESR*SSTPLPTISSSAENTR	R157	1
P42167	LAP2B	Lamina-associated polypeptide 2, isoforms beta/gamma	HASPILPITEFSDIPR*R	R319	1
P42167	LAP2B	Lamina-associated polypeptide 2, isoforms beta/gamma	EMFPYEASTPTGISASCR*RPIK	R364	1
P42167	LAP2B	Lamina-associated polypeptide 2, isoforms beta/gamma	GAAGR*PLELSDFR*MEESFSSK	R373|R381	1
P43243	MATR3	Matrin-3	GPLPLSSQHR*GDADQASNILASFGLSAR	R102	1
P43243	MATR3	Matrin-3	DLDELSR*YPEDK	R127	2
P43243	MATR3	Matrin-3	TEEGPTLSYGR*DGR*SATR	R160|R163	2
P43243	MATR3	Matrin-3	GPSLNPVLDYDHGSR*SQESGYYDR	R207	1
P43243	MATR3	Matrin-3	FDSEYER*MGR	R252	1
P43243	MATR3	Matrin-3	FDSEYER*MGR*GPGPLQER	R252|R255	1
P43243	MATR3	Matrin-3	GPGPLQER*SLFEK	R263	2
P43243	MATR3	Matrin-3	GNLGAGNGNLQGPR*HMQK	R387	4
P43243	MATR3	Matrin-3	QELGR*VIHLSNLPHSGYSDSAVLK	R496	1
P43243	MATR3	Matrin-3	LVLR*IPNR*GIDLLK	R578|R582	1
P43243	MATR3	Matrin-3	IPNR*GIDLLK	R582	1
P43243	MATR3	Matrin-3	IPNR*GIDLLKK	R582	1
P46087-2	NOP2	Isoform 2 of Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	GPQPPTVSPIR*SSRPPPAK	R785	1
P46776	RL27A	60S ribosomal protein L27a	TGAAPIIDVVR*SGYYK	R105	3
P46776	RL27A	60S ribosomal protein L27a	LWTLVSEQTR*VNAAK	R87	1
P46781	RS9	40S ribosomal protein S9	PVAR*SWVCR	R5	1
P46783	RS10	40S ribosomal protein S10	GLEGERPAR*LTR	R116	1
P49005	DPOD2	DNA polymerase delta subunit 2	VILAGNLLSHSTQSR*DSINK	R255	1
P49368	TCPG	T-complex protein 1 subunit gamma	KGESQTDIEITR*EEDFTR	R260	1
P49756	RBM25	RNA-binding protein 25	R*FPVAPLIPYPLITK	R247	1
P49756	RBM25	RNA-binding protein 25	EDINAIEMEEDKR*DLISR	R274	1
P49790	NU153	Nuclear pore complex protein Nup153	LR*NTPYQAPVR	R279	1
P49790	NU153	Nuclear pore complex protein Nup153	RIPSIVSSPLNSPLDR*SGIDITDFQAK	R342	1
P49792	RBP2	E3 SUMO-protein ligase RanBP2	FGQGDLPKPINSDFR*SVFSTK	R1473	2
P49915	GUAA	GMP synthase [glutamine-hydrolyzing]	VINAAHSFYNGTTTLPISDEDR*TPR	R317	3
P50402	EMD	Emerin	TYGEPESAGPSR*AVR	R115	1
P50454	SERPH	Serpin H1	LGSR*LYGPSSVSFADDFVR	R133	1
P50454	SERPH	Serpin H1	LYGPSSVSFADDFVR*SSK	R148	1
P51610-4	HCFC1	Isoform 4 of Host cell factor 1	FR*VAGINACGR	R1909	1
P51991	ROA3	Heterogeneous nuclear ribonucleoprotein A3	IETIEVMEDR*QSGK	R161	1
P51991	ROA3	Heterogeneous nuclear ribonucleoprotein A3	SSGSPYGGGYGSGGGSGGYGSR*R	R376	2
P51991	ROA3	Heterogeneous nuclear ribonucleoprotein A3	MEVKPPPGR*PQPDSGR	R9	1
P51991-2	ROA3	Isoform 2 of Heterogeneous nuclear ribonucleoprotein A3	IETIEVMEDR*QSGK	R139	1
P51991-2	ROA3	Isoform 2 of Heterogeneous nuclear ribonucleoprotein A3	SSGSPYGGGYGSGGGSGGYGSR*R	R354	2
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	FGSGMNMGR*INEILSNALK	R371	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGPGIDR*LGGAGMER	R410	2
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGAGLGHGMDR*VGSEIER	R429	4
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGLVMDR*MGSVER	R443	3
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGSGIER*MGPLGLDHMASSIER	R456	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGQTMER*IGSGVER	R478	1

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MAAPIDR*VGQTIER	R503	5
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGSGVER*MGPAIER	R517	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGPAIER*MGLSMER	R524	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MVPAGMGAGLER*MGPVMDR	R543	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGPVMDR*MATGLER	R550	3
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGANNLER*MGLER	R565	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	MGLER*MGANSLER	R570	4
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	GGNR*FEPYANPTK	R60	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	GGNR*FEPYANPTKR	R60	1
P52272	HNRPM	Heterogeneous nuclear ribonucleoprotein M	AIEMER*GNFGGSFAGSFGGA GGHAPGVAR	R627	1
P52597	HNRPF	Heterogeneous nuclear ribonucleoprotein F	FMSVQR*PGPYDRPGTAR	R206	1
P52597	HNRPF	Heterogeneous nuclear ribonucleoprotein F	FMSVQRPGPYDR*PGTAR	R212	2
P52597	HNRPF	Heterogeneous nuclear ribonucleoprotein F	FMSVQRPGPYDR*PGTAR*R* YIGIVK	R212|R217|R218	1
P52948-5	NUP98	Isoform 5 of Nuclear pore complex protein Nup98-Nup96	TR*SLVGGLLQSK	R1025	2
P54886	P5CS	Delta-1-pyrroline-5-carboxylate synthase	LR*HEILLSQSVR	R128	1
P55072	TERA	Transitional endoplasmic reticulum ATPase	KYEMFAQTLQQSR*GFGSFR	R766	2
P55196-5	AFAD	Isoform 5 of Afadin	FRPDMR*MLSSPK	R84	1
P55795	HNRH2	Heterogeneous nuclear ribonucleoprotein H2	FGR*DLNYCFSGMSDHR	R262	2
P61106	RAB14	Ras-related protein Rab-14	LQIWDTAGQER*FR	R72	2
P61978-2	HNRPK	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	GGDLMAYDR*R	R325	1
P61978-2	HNRPK	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	GGDLMAYDR*R*GRPGDR	R325|R326	1
P61978-2	HNRPK	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	SR*NTDEMVELR	R37	1
P61978-3	HNRPK	Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	GGDLMAYDR*R	R301	1
P61978-3	HNRPK	Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	GGDLMAYDR*R*GRPGDR	R301|R302	1
P61978-3	HNRPK	Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	SR*NTDEMVELR	R37	1
P62633-2	CNBP	Isoform 2 of Cellular nucleic acid-binding protein	DCDLQEDACYNCGR*GGHIAK	R72	1
P62701	RS4X	40S ribosomal protein S4, X isoform	LTGVFAPR*PSTGPHK	R30	1
P62753	RS6	40S ribosomal protein S6	EDDVR*QYVVR	R154	1
P62753	RS6	40S ribosomal protein S6	IQRLVTPR*VLQHK	R183	1
P62753	RS6	40S ribosomal protein S6	LVTPR*VLQHK	R183	1
P62753	RS6	40S ribosomal protein S6	R*MATEVAADALGEEWK	R31	1
P62820	RAB1A	Ras-related protein Rab-1A	LQIWDTAGQER*FR	R72	2
P62854	RS26	40S ribosomal protein S26	FVIR*NIVEAAAVR	R42	1
P62917	RL8	60S ribosomal protein L8	VISSANR*AVVGVVAGGGR	R163	1
P62917	RL8	60S ribosomal protein L8	VISSANRAVVGVVAGGGR*ID KPILK	R174	2
P67809	YBOX1	Nuclease-sensitive element-binding protein 1	WFNVR*NGYGFINR*NDTK	R69|R77	1
P68104	EF1A1	Elongation factor 1-alpha 1	DGNASGTTLLEALDCILPPTR PTDKPLR*LPLQDVYK	R247	1
P68104	EF1A1	Elongation factor 1-alpha 1	KDGNASGTTLLEALDCILPPT RPTDKPLR*LPLQDVYK	R247	1
P68104	EF1A1	Elongation factor 1-alpha 1	SGDAAIVDMVPGKPMCVESFS DYPPLGR*FAVR	R423	1
P68104	EF1A1	Elongation factor 1-alpha 1	FAVR*DMRQTVAVGVIK	R427	1
P68104	EF1A1	Elongation factor 1-alpha 1	DMR*QTVAVGVIK	R430	2
P68104	EF1A1	Elongation factor 1-alpha 1	YYVTIIDAPGHR*DFIK	R96	2
P68366-2	TBA4A	Isoform 2 of Tubulin alpha-4A chain	EDAANNYAR*GHYTIGK	R90	3
P68371	TBB4B	Tubulin beta-4B chain	GHYTEGAELVDSVLDVVR*K	R121	4
P68371	TBB4B	Tubulin beta-4B chain	MR*EIVHLQAGQCGNQIGAK	R2	2
P68371	TBB4B	Tubulin beta-4B chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
P78347-2	GTF2I	Isoform 2 of General transcription factor II-I	RPELLTHSTTEVTQPR*TNTP VKEDWNVR	R514	1
P78371	TCPB	T-complex protein 1 subunit beta	IFGSR*VR*VDSTAK	R255|R257	2
P83731	RL24	60S ribosomal protein L24	FQR*AITGASLADIMAK	R80	1
P84098	RL19	60S ribosomal protein L19	R*LASSVLR*CGK	R9|R16	1
P84103-2	SRSF3	Isoform 2 of Serine/arginine-rich splicing factor 3	MHR*DSCPLDCK	R3	2
P98175-5	RBM10	Isoform 5 of RNA-binding protein 10	LPLGTR*LDQQTLPLGGR	R317	2
Q00341	VIGLN	Vigilin	SSVAVLTQESFAEHR*SGLVP QQIK	R16	1
Q00839	HNRPU	Heterogeneous nuclear ribonucleoprotein U	SSGPTSLFAVTVAPPGAR*QG QQQAGGK	R204	2
Q00839	HNRPU	Heterogeneous nuclear ribonucleoprotein U	EDHGR*GYFEYIEENK	R255	3
Q00839	HNRPU	Heterogeneous nuclear ribonucleoprotein U	LNTLLQR*APQCLGK	R558	2
Q00839	HNRPU	Heterogeneous nuclear ribonucleoprotein U	FIEIAAR*K	R572	2
Q00839-2	HNRPU	Isoform 2 of Heterogeneous nuclear ribonucleoprotein U	SSGPTSLFAVTVAPPGAR*QG QQQAGGDGK	R204	5
Q00839-2	HNRPU	Isoform 2 of Heterogeneous nuclear ribonucleoprotein U	EDHGR*GYFEYIEENK	R236	3
Q00839-2	HNRPU	Isoform 2 of Heterogeneous nuclear ribonucleoprotein U	LNTLLQR*APQCLGK	R539	2

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
Q00839-2	HNRPU	Isoform 2 of Heterogeneous nuclear ribonucleoprotein U	FIEIAAR*K	R553	2
Q02878	RL6	60S ribosomal protein L6	VLATVTKPVGGDKNGGTR*VVK	R105	1
Q02878	RL6	60S ribosomal protein L6	NPVLVR*GIGR	R52	1
Q03252	LMNB2	Lamin-B2	LSPSPSSR*VTVSR	R411	5
Q04637-4	IF4G1	Isoform C of Eukaryotic translation initiation factor 4 gamma 1	ITKPGSIDSNNQLFAPGGR*LSWGK	R1003	5
Q04637-4	IF4G1	Isoform C of Eukaryotic translation initiation factor 4 gamma 1	TPLRPLDPTR*LQGINCGPDFTPSFANLGR	R569	2
Q04637-4	IF4G1	Isoform C of Eukaryotic translation initiation factor 4 gamma 1	LQGINCGPDFTPSFANLGR*TTLSTR	R588	1
Q04637-4	IF4G1	Isoform C of Eukaryotic translation initiation factor 4 gamma 1	GGPPGPPISR*GLPLVDDGGWNTVPISK	R955	1
Q05639	EF1A2	Elongation factor 1-alpha 2	FAVR*DMRQTVAVGVIK	R427	1
Q05639	EF1A2	Elongation factor 1-alpha 2	DMR*QTVAVGVIK	R430	2
Q07666-3	KHDR1	Isoform 3 of KH domain-containing, RNA-binding, signal transduction-associated protein 1	SGSMDPSGAHPSVR*QTPSR	R31	1
Q07666-3	KHDR1	Isoform 3 of KH domain-containing, RNA-binding, signal transduction-associated protein 1	GAYR*EHPYGR*Y	R397|R403	1
Q07812-2	BAX	Isoform Beta of Apoptosis regulator BAX	TGALLLQGFIQDR*AGR	R34	2
Q08211	DHX9	ATP-dependent RNA helicase A	ILTTEGR*NALIHK	R1018	2
Q08211	DHX9	ATP-dependent RNA helicase A	QISRPSAAGINLMIGSTR*YGDGPRPPK	R1154	1
Q08211	DHX9	ATP-dependent RNA helicase A	AENNSEVGASGYGVPGPTWDRGANLR	R141	5
Q12931	TRAP1	Heat shock protein 75 kDa, mitochondrial	FEDR*SPAAECLSEK	R567	2
Q12931	TRAP1	Heat shock protein 75 kDa, mitochondrial	LDTHPAMVTVLEMGAAR*HFLR	R619	1
Q12931	TRAP1	Heat shock protein 75 kDa, mitochondrial	HFLR*MQQLAK	R623	2
Q13162	PRDX4	Peroxiredoxin-4	TR*EEECHFYAGGQVYPGEASR	R47	2
Q13247-3	SRSF6	Isoform SRP55-3 of Serine/arginine-rich splicing factor 6	DGYSYGSR*SGGGGYSSR	R85	1
Q13263	TIF1B	Transcription intermediary factor 1-beta	IVAERPGTNSTGPAPMAPPR*APGPLSK	R427	1
Q13428-2	TCOF	Isoform 2 of Treacle protein	ILQVR*AASAPAK	R224	2
Q13442	HAP28	28 kDa heat- and acid-stable phosphoprotein	GVEGLIDIENPNR*VAQTTK	R88	1
Q13442	HAP28	28 kDa heat- and acid-stable phosphoprotein	KGVEGLIDIENPNR*VAQTTK	R88	1
Q13492-2	PICAL	Isoform 2 of Phosphatidylinositol-binding clathrin assembly protein	SGQSLTDR*ITAAQHSVTGSAVSK	R9	1
Q13541	4EBP1	Eukaryotic translation initiation factor 4E-binding protein 1	SGGSSCSQTPSR*AIPATR	R13	5
Q13541	4EBP1	Eukaryotic translation initiation factor 4E-binding protein 1	FLMECR*NSPVTK	R63	2
Q13573	SNW1	SNW domain-containing protein 1	GMDSGFAGGEDEIYNVYDQAWR*GGK	R438	1
Q13838-2	DX39B	Isoform 2 of Spliceosome RNA helicase DDX39B	ILVATNLFGR*GMDIER	R364	1
Q13885	TBB2A	Tubulin beta-2A chain	GHYTEGAELVDSVLDVVR*K	R121	4
Q13885	TBB2A	Tubulin beta-2A chain	MR*EIVHIQAGQCGNQIGAK	R2	2
Q13885	TBB2A	Tubulin beta-2A chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
Q13895	BYST	Bystin	HAPLADQILAGNAVR*AGVR	R30	1
Q14008-2	CKAP5	Isoform 2 of Cytoskeleton-associated protein 5	LNQAR*SMSGHPEAAQMVR	R1468	2
Q14126	DSG2	Desmoglein-2	VVPSFLPVDQGGSLVGR*NGVGGMAK	R684	1
Q14152	EIF3A	Eukaryotic translation initiation factor 3 subunit A	GLDEDR*GSWR	R1026	1
Q14152	EIF3A	Eukaryotic translation initiation factor 3 subunit A	LGDDEDREPSLRPDDDR*VPR	R956	1
Q14152	EIF3A	Eukaryotic translation initiation factor 3 subunit A	GADDDRPSWR*NTDDDRPPR	R990	2
Q14247-3	SRC8	Isoform 3 of Src substrate cortactin	HCSQVDSVR*GFGGK	R119	1
Q14692	BMS1	Ribosome biogenesis protein BMS1 homolog	LFR*IQGQK	R1260	1
Q14692	BMS1	Ribosome biogenesis protein BMS1 homolog	AGLSPANCQSDR*VNLEK	R560	1
Q14739	LBR	Delta(14)-sterol reductase	LTPLILKPFGNSISR*YNGEPEHIER	R131	1
Q14739	LBR	Delta(14)-sterol reductase	LTPLILKPFGNSISRYNGEPEHIER*NDAPHK	R141	2
Q14739	LBR	Delta(14)-sterol reductase	ELAVR*TFEVTPIR	R195	1
Q14839-2	CHD4	Isoform 2 of Chromodomain-helicase-DNA-binding protein 4	NILSR*LANR	R1923	2
Q14839-2	CHD4	Isoform 2 of Chromodomain-helicase-DNA-binding protein 4	ENEFSFEDNAIR*GGK	R829	2

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
Q14C86-3	GAPD1	Isoform 3 of GTPase-activating protein and VPS9 domain-containing protein 1	LSAQAQVAEDILDKYR*NAIK	R1010	1
Q15020	SART3	Squamous cell carcinoma antigen recognized by T-cells 3	VLSR*AVAAATYK	R41	1
Q15056-2	IF4H	Isoform Short of Eukaryotic translation initiation factor 4H	ADFDTYDDR*AYSSFGGGR	R10	3
Q15056-2	IF4H	Isoform Short of Eukaryotic translation initiation factor 4H	DGPPLR*GSNMDFREPTEEER	R171	1
Q15056-2	IF4H	Isoform Short of Eukaryotic translation initiation factor 4H	AQR*PR*LQLKPR	R188|R190	1
Q15058	KIF14	Kinesin-like protein KIF14	VLGTGNLYHR*SIGK	R257	1
Q15233	NONO	Non-POU domain-containing octamer-binding protein	FGQAATMEGIGAIGGTPPAFNR*AAPGAEFAPNKR	R456	5
Q15286	RAB35	Ras-related protein Rab-35	LQIWDTAGQER*FR	R69	2
Q15417	CNN3	Calponin-3	GMSVYGLGR*QVYDPK	R265	1
Q15428	SF3A2	Splicing factor 3A subunit 2	MDFQHR*PGGK	R6	1
Q15637-5	SF01	Isoform 5 of Splicing factor 1	EDDNR*ILRPWQSSETR	R385	1
Q15637-5	SF01	Isoform 5 of Splicing factor 1	ILRPWQSSETR*SITNTTVCTK	R396	1
Q15637-6	SF01	Isoform 6 of Splicing factor 1	EDDNR*ILRPWQSSETR	R260	1
Q15637-6	SF01	Isoform 6 of Splicing factor 1	ILRPWQSSETR*SITNTTVCTK	R271	1
Q15637-6	SF01	Isoform 6 of Splicing factor 1	YACGLWGLSPASR*K	R467	1
Q15717	ELAV1	ELAV-like protein 1	SNGYEDHMAEDCR*GDIGR	R14	3
Q15717	ELAV1	ELAV-like protein 1	NVALLSQLYHSPAR*R	R205	2
Q15717	ELAV1	ELAV-like protein 1	NVALLSQLYHSPARR*FGGPVHHQAQR	R206	1
Q15836	VAMP3	Vesicle-associated membrane protein 3	LQQTQNQVDEVVDIMR*VNVDK	R30	1
Q16777	H2A2C	Histone H2A type 2-C	SSR*AGLQFPVGR	R21	1
Q3ZCQ8	TIM50	Mitochondrial import inner membrane translocase subunit TIM50	QNLFLGSLTSR*LWPR	R345	4
Q4VCS5	AMOT	Angiomotin	SLSER*LMQMSLATSGVK	R177	1
Q52LJ0-2	FA98B	Isoform 2 of Protein FAM98B	VLMGR*VPDR	R304	1
Q5JSH3-2	WDR44	Isoform 2 of WD repeat-containing protein 44	HLAEEYGER*AINK	R449	1
Q5QNW6-2	H2B2F	Isoform 2 of Histone H2B type 2-F	IAGEASR*LAHYNK	R80	3
Q6DN03	H2B2C	Putative histone H2B type 2-C	IAGEASR*LAHYNK	R80	3
Q6P1J9	CDC73	Parafibromin	APEQRPAPNAAPVDPTLR*TK	R281	1
Q6P2E9	EDC4	Enhancer of mRNA-decapping protein 4	ASCASIDIEDATQHLR*DILK	R17	1
Q6PJT7-2	ZC3HE	Isoform 2 of Zinc finger CCCH domain-containing protein 14	LCEPEVLNSLEETYSPFFR*NNSEK	R278	1
Q6PKG0	LARP1	La-related protein 1	KSEESR*FSHLTSLPQQLPSQQLMSK	R578	1
Q71U36-2	TBA1A	Isoform 2 of Tubulin alpha-1A chain	HVPR*AVFVDLEPTVIDEVR	R29	1
Q71U36-2	TBA1A	Isoform 2 of Tubulin alpha-1A chain	AVFVDLEPTVIDEVR*TGTYR	R44	5
Q71U36-2	TBA1A	Isoform 2 of Tubulin alpha-1A chain	TGTYR*QLFHPEQLITGK	R49	1
Q71U36-2	TBA1A	Isoform 2 of Tubulin alpha-1A chain	EDAANNYAR*GHYTIGK	R70	3
Q7Z2W4	ZCCHV	Zinc finger CCCH-type antiviral protein 1	KFTYLGSQDR*ARPPSGSSK	R305	1
Q7Z2W4	ZCCHV	Zinc finger CCCH-type antiviral protein 1	FTYLGSQDR*ARPPSGSSK	R305	2
Q7Z434	MAVS	Mitochondrial antiviral-signaling protein	LPINSTR*AGMVPSK	R355	2
Q7Z7F7-2	RM55	Isoform 2 of 39S ribosomal protein L55, mitochondrial	KEYEQELSDDLHVER*YR	R154	1
Q86U42-2	PABP2	Isoform 2 of Polyadenylate-binding protein 2	AAAAAAAAAAGAAGGR*GSGPGR	R17	1
Q8IWS0	PHF6	PHD finger protein 6	GLSEDTR*STSSHGTDEMESSSYR	R180	1
Q8IX01-4	SUGP2	Isoform 4 of SURP and G-patch domain-containing protein 2	YSLSGSVAHSR*DAGR	R73	1
Q8IYB3-2	SRRM1	Isoform 2 of Serine/arginine repetitive matrix protein 1	MDAGFFR*GTSAEQDNR	R7	1
Q8N1F7	NUP93	Nuclear pore complex protein Nup93	ASVLLGSR*GLDISHISQR	R67	2
Q8N684-3	CPSF7	Isoform 3 of Cleavage and polyadenylation specificity factor subunit 7	TPAILYTYSGLR*NR	R121	1
Q8N7H5-3	PAF1	Isoform 3 of RNA polymerase II-associated factor 1 homolog	TEYISTEFNR*YGISNEKPEVK	R133	1
Q8NC51	PAIRB	Plasminogen activator inhibitor 1 RNA-binding protein	EFDR*HSGSDR*SSFSHYSGLK	R195|R201	2
Q8NC51	PAIRB	Plasminogen activator inhibitor 1 RNA-binding protein	HSGSDR*SSFSHYSGLK	R201	2
Q8NC51	PAIRB	Plasminogen activator inhibitor 1 RNA-binding protein	VEFNIR*KPNEGADGQWK	R309	1
Q8NC51-3	PAIRB	Isoform 3 of Plasminogen activator inhibitor 1 RNA-binding protein	EFDR*HSGSDR*SSFSHYSGLK	R195|R201	2
Q8NC51-3	PAIRB	Isoform 3 of Plasminogen activator inhibitor 1 RNA-binding protein	HSGSDR*SSFSHYSGLK	R201	2
Q8NC51-3	PAIRB	Isoform 3 of Plasminogen activator inhibitor 1 RNA-binding protein	VEFNIR*KPNEGADGQWK	R294	1
Q8NCF5	NF2IP	NFATC2-interacting protein	SSLR*LIPDDLSLLK	R135	2

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
Q8ND56-2	LS14A	Isoform 2 of Protein LSM14 homolog A	DFDFESANAQFNKEEIDR*EFHNK	R312	1
Q8WWM7-2	ATX2L	Isoform 2 of Ataxin-2-like protein	YIPLPQR*VR	R355	1
Q92597	NDRG1	Protein NDRG1	YFVQGMGYMPSASMTR*LMR	R322	1
Q92665	RT31	28S ribosomal protein S31, mitochondrial	SELLSQLQQHEEESR*AQR	R189	1
Q92841	DDX17	Probable ATP-dependent RNA helicase DDX17	LMQLVDHR*GGGGGGGGR	R555	1
Q92841-3	DDX17	Isoform 4 of Probable ATP-dependent RNA helicase DDX17	LMQLVDHR*GGGGGGGK	R476	1
Q92922	SMRC1	SWI/SNF complex subunit SMARCC1	EKPVDLQNFGLR*TDIYSK	R602	1
Q92945	FUBP2	Far upstream element-binding protein 2	VPDGMVGLIIGR*GGEQINK	R162	1
Q92945	FUBP2	Far upstream element-binding protein 2	DAFADAVQR*AR*QIAAK	R80|R82	1
Q92973-2	TNPO1	Isoform 2 of Transportin-1	GDVEEDETIPDSEQDIRPR*FHR	R338	2
Q96AE4	FUBP1	Far upstream element-binding protein 1	NEYGSR*IGGNEGIDVPIPR	R271	1
Q96AE4-2	FUBP1	Isoform 2 of Far upstream element-binding protein 1	NEYGSR*IGGNEGIDVPIPR	R270	1
Q96C36	P5CR2	Pyrroline-5-carboxylate reductase 2	LLTR*SLALGGK	R311	3
Q96E39	RMXL1	RNA binding motif protein, X-linked-like-1	GPPSR*GGHMDDGGYSMNFNMSSSR*GPLPVK	R125|R144	4
Q96HA1-2	P121A	Isoform 2 of Nuclear envelope pore membrane protein POM 121	FSR*SAIPEQIISSTLSSPSSNAPDPCAK	R19	1
Q96I24	FUBP3	Far upstream element-binding protein 3	AEGFVDALHR*VR	R25	1
Q96I24	FUBP3	Far upstream element-binding protein 3	AEGFVDALHR*VRQIAAK	R25	1
Q99459	CDC5L	Cell division cycle 5-like protein	EHLR*LGLLGLPAPK	R477	1
Q99733-2	NP1L4	Isoform 2 of Nucleosome assembly protein 1-like 4	LTDQVMQNPR*VLAALQER	R36	1
Q99961	SH3G1	Endophilin-A2	IAASSSFR*SSDKPIR	R290	3
Q9BQE3	TBA1C	Tubulin alpha-1C chain	EDAANNYAR*GHYTIGK	R105	3
Q9BQE3	TBA1C	Tubulin alpha-1C chain	HVPR*AVFVDLEPTVIDEVR	R64	1
Q9BQE3	TBA1C	Tubulin alpha-1C chain	AVFVDLEPTVIDEVR*TGTYR	R79	5
Q9BQE3	TBA1C	Tubulin alpha-1C chain	TGTYR*QLFHPEQLITGK	R84	1
Q9BU76-4	MMTA2	Isoform 4 of Multiple myeloma tumor-associated protein 2	GVDR*LLGLGSASGSVGR	R117	2
Q9BUF5	TBB6	Tubulin beta-6 chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
Q9BUJ2-2	HNRL1	Isoform 2 of Heterogeneous nuclear ribonucleoprotein U-like protein 1	RPYEENR*GR*GYFEHR	R179|R181	2
Q9BVA1	TBB2B	Tubulin beta-2B chain	GHYTEGAELVDSVLDVVR*K	R121	4
Q9BVA1	TBB2B	Tubulin beta-2B chain	MR*EIVHIQAGQCGNQIGAK	R2	2
Q9BVA1	TBB2B	Tubulin beta-2B chain	LHFFMPGFAPLTSR*GSQQYR	R276	4
Q9BWJ5	SF3B5	Splicing factor 3B subunit 5	TDR*YTIHSQLEHLQSK	R4	1
Q9BZE4	NOG1	Nucleolar GTP-binding protein 1	VNEVLNR*LHLAIPTR	R359	1
Q9C005	DPY30	Protein dpy-30 homolog	MEPEQMLEGQTQVAENPHSEYGLTDNVER*IVENEK	R29	1
Q9GZT3-2	SLIRP	Isoform 2 of SRA stem-loop-interacting RNA-binding protein, mitochondrial	AASAAR*GAAALR	R7	2
Q9H0U4	RAB1B	Ras-related protein Rab-1B	MGPGAASGGER*PNLK	R183	3
Q9H0U4	RAB1B	Ras-related protein Rab-1B	LQIWDTAGQER*FR	R69	2
Q9H1E3	NUCKS	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	VVDYSQFQESDDADEDYGR*DSGPPTK	R28	1
Q9H307	PININ	Pinin	AVAVR*TLQEQLEK	R6	2
Q9H307	PININ	Pinin	DLEGAVSR*LGGER	R85	1
Q9H910	JUPI2	Jupiter microtubule associated homolog 2	TSDIFGSPVTATSR*LAHPNKPK	R104	2
Q9H910	JUPI2	Jupiter microtubule associated homolog 2	GSGIFDESTPVQTR*QHLNPPGGK	R81	1
Q9NNW5	WDR6	WD repeat-containing protein 6	AR*AGAGAPVVGSGSSGGGNAFTGLGPVSTLPSLHGK	R535	1
Q9NPD8	UBE2T	Ubiquitin-conjugating enzyme E2 T	ADEEEMLDNLPEAGDSR*VHNSTQK	R173	1
Q9NR30	DDX21	Nucleolar RNA helicase 2	LR*SDAGLESDTAMK	R6	1
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	FTDEESR*VFLLDR	R344	2
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	VFLLDR*GNTR	R350	1
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	SFATASHR*NTEEEGLK	R427	1
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	ETGYVVER*PSTTK	R468	1
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	LKETGYVVER*PSTTK	R468	2
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	DR*LLASTLVHSVK	R567	2
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	LAGEER*VFKEENQK	R671	3
Q9NYF8-2	BCLF1	Isoform 2 of Bcl-2-associated transcription factor 1	LR*CDSADLR*HDIDR	R685|R692	1
Q9NYV4-2	CDK12	Isoform 2 of Cyclin-dependent kinase 12	FAR*VPLALHPVVGQPFLK	R1398	2

Continued on next page

Table 5A.2 – continued from previous page

Accession	Protein	Description	Sequence	Residue	Spectral counts
Q9P258	RCC2	Protein RCC2	AAAAAWEEPSSGNGTAR*AGPR	R22	2
Q9UID3	VPS51	Vacuolar protein sorting-associated protein 51 homolog	AAAAAAGPSPGSGPGDSPEGPEGEAPER*R	R29	1
Q9UK61-3	TASOR	Isoform 3 of Protein TASOR	FLYSAPR*NK	R593	1
Q9UKM9-2	RALY	Isoform 1 of RNA-binding protein Raly	LFAR*STAVTTSSAK	R153	1
Q9UKM9-2	RALY	Isoform 1 of RNA-binding protein Raly	VLAGQTLDINMAGEPKPDR*PK	R97	1
Q9UKV3	ACINU	Apoptotic chromatin condensation inducer in the nucleus	EFKEEGEEIPR*VKPEEMMDERPK	R313	1
Q9UKV3	ACINU	Apoptotic chromatin condensation inducer in the nucleus	TRSQEQEVLER*GGR	R336	1
Q9UKY7	CDV3	Protein CDV3 homolog	AETEER*SLDNFFAK	R7	1
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2	MSQVPAPVPLMSLR*TAPAANLASR	R2221	1
Q9Y262-2	EIF3L	Isoform 2 of Eukaryotic translation initiation factor 3 subunit L	QDLAYER*QYEQQTYQVIPEVIK	R38	1
Q9Y266	NUDC	Nuclear migration protein nudC	LSDLDSETR*SMVEK	R284	2
Q9Y2W1	TR150	Thyroid hormone receptor-associated protein 3	NFR*VTAYK	R522	1
Q9Y2W1	TR150	Thyroid hormone receptor-associated protein 3	R*NREEEWDPEYTPK	R863	1
Q9Y2W1	TR150	Thyroid hormone receptor-associated protein 3	R*NR*EEEWDPEYTPK	R863|R865	1
Q9Y2W1	TR150	Thyroid hormone receptor-associated protein 3	RNR*EEEWDPEYTPK	R865	1
Q9Y2W2	WBP11	WW domain-binding protein 11	TSAYGPPTR*AVSILPLLGHGVPR	R178	1
Q9Y2W2	WBP11	WW domain-binding protein 11	AEITR*FVPTALR	R577	1
Q9Y2W2	WBP11	WW domain-binding protein 11	FVPTALR*VR	R584	1
Q9Y3B9	RRP15	RRP15-like protein	AAAAPDSR*VSEEENLKK	R9	1
Q9Y450	HBS1L	HBS1-like protein	R*DKPSVEPVEEYDYEDLK	R45	1
Q9Y520-7	PRC2C	Isoform 7 of Protein PRRC2C	IDNR*LLEKPYVR	R1318	1
Q9Y520-7	PRC2C	Isoform 7 of Protein PRRC2C	SAR*DHAISLSEPR	R796	1
Q9Y570-4	PPME1	Isoform 4 of Protein phosphatase methylesterase 1	SMHLGR*LPSRPPLPGSGGSQSGAK	R12	1
Q9Y5M8	SRPRB	Signal recognition particle receptor subunit beta	GGR*GDVGSADIQDLEK	R252	2