THE UNIVERSITY OF CHICAGO

METABOLITE-INDUCED PROTEIN MODIFICATIONS IN CELL SIGNALING: A DIRECT LINK BETWEEN METABOLISM AND REGULATORY PATHWAYS OF THE CELL

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

GIHOON LEE

CHICAGO, ILLINOIS

JUNE 2019

"Ask, and it will be given you. Seek, and you will find. Knock, and it will be opened for you.

Matthew 7:7

Table of Contents

List of Figures

List of Tables

Abstract

Metabolite-induced Protein Modifications in Cell Signaling: A Direct Link Between Metabolism and Regulatory Pathways of the Cell

By Gihoon Lee Under Direction of Professor Raymond E. Moellering

Glycolysis is a primary and central metabolic pathway of the carbohydrates that supplies the energy source and building block precursors to the living system. A strong correlation between deregulated glucose metabolism and various human diseases has been studied, however mechanisms that integrate the metabolic state with regulatory pathways in cellular systems are largely unexplored.

Here we report the direct link between glycolysis and KEAP1-NRF2 transcriptional program via a novel reactive metabolite-induced posttranslational modification. We demonstrate that reactive dicarbonyl metabolite methylglyoxal, mainly generated from glycolysis in cells, is a signaling messenger of the activation of NRF2 transcriptional program. Methylglyoxal induces the formation of a methylimidazole crosslink between cysteine electrophile sensor and proximal arginine residues posttranslational modification between two monomeric KEAP1 which was firstly characterized by our study. *In-vivo* experiment of NRF2-dependent UV-damage mouse model with the PGK1 inhibitor showed therapeutic efficacy, demonstrating the physiological relevance of regulating glucose metabolism for activating NRF2 signaling cascade. In summary, our work presented herein highlights the role of reactive metabolite-induced posttranslational modifications in cell signaling.

Furthermore, we developed the novel chemical proteomic technologies to unveil the underlying protein communication networks and their dynamic and temporal interactions with other proteins or ligands in live cells.

Our P3 profiling technology was designed to profile high resolution protein proximity maps by mild and fast proximity biotin labeling with light activation. Application of our approach to KEAP1 identified potential protein-protein interaction partners including well-characterized KEAP1 binding protein PGAM5, demonstrating that our novel chemical proteomic tool is applicable to study challenging research targets such as redox signaling networks of the cell.

We also designed SILAC surface mapping assay to study the protein-ligand interactions in live cells, and we confirmed our strategy with photoaffinity analogue probe of GNF-2 and its binding site in c-ABL that is well-characterized. Indeed, the results of our quantitative proteomic workflow with PGK1 and its novel inhibitor CBR-470-1 suggest that CBR-470-1 may inhibit PGK1 activity by interfering its interaction with glycolytic intermediates 1,3-BPG and 3PG.

Overall, these new chemical proteomic technologies with photoreactive functionalities and quantitative mass spectrometry may provide an unbiased, global insight into the nature of protein communications and regulations in complex biological systems that lead to elucidate the network of reactive metabolite-induced posttranslational modifications in cell signaling.

Acknowledgements

I would like to express my sincerest gratitude to all who supports me in my graduate career. The work presented in this dissertation would not be possible without their support.

First and foremost, I would like to thank my doctorate advisor Dr. Raymond E. Moellering for his insightful mentorship, incredible supports, and encouragement throughout my graduate program. Starting my doctoral research in the field of chemical biology was a big challenge to me because I had no research experience in chemical biology and lack of my background in biology. I still remember the moment when I met Ray at the first time after starting my doctoral program. In the meetings with Ray in my first quarter, I was very impressed by his enthusiasm for science and his ambition to build up the forefront research field to address the fundamental biological questions. Thus, I was convinced of my successful graduate career in chemical biology under his guidance and decided to join the Moellering lab. Then, it became a one of my best decisions ever in my life. Ray gave me the wonderful opportunities to explore the diverse research projects that aim to discover the underlying nature of proteins in complex biological system via novel chemical proteomic technologies. Involving in these challenging projects sometimes brought me into the hard and stressful moments, but Ray always kept me on the right path of research direction and encouraged me to have critical thinking to the research, ranging from fine details of each experiment to the big picture of whole research projects. As with my dedication, his guidance led to the successful completion of the work presented in this thesis, and I appreciated his patience with me to be trained as a chemical biologist in the Moellering lab. I also appreciate his supports for my applications and extensions of my fellowships, research items and maintenance of Q-

Exactive HF, and my attendance to the conference. I am very grateful and honored to be the first recipient of doctoral degree awarded by Ray.

Secondly, I would like to thank all in the Moellering lab. I felt lucky to do my research with the amazing lab members who are smart, kind and supportive. Dr. Jae Won Chang is a first postdoc of the Moellering lab, and I've worked several research projects with him since I joined the lab. He taught me the basic skills of chemical proteomics and metabolomics, and he contributed to design, perform, and analyze the targeted metabolomics with LC-MS described in Chapter 2. I am thankful for his support for me to familiarize with chemical biology studies and his dedicated contribution to Chapter 2 work. Indeed, his advice and comments were helpful for me to continue my experiments in the lab and prepare for the next steps in academic career, and I am thankful for his kindness as well. Dr. David C. McCutcheon is collaborating with me for the research work described in Chapter 3, and he contributed to the project as a one of two co-first authors. He designed and synthesized the proximity profiling probes, generated the SNAP-POI genes and stably expressed cell lines, and performed the platform validation tests and *in vitro* assays. I am grateful for his help and discussion for synthesis and purification of the photoaffinity probe GNF-PAP that is described in Chapter 4, and also thankful for his strong dedication to the work in Chapter 3. I also want to thank John S. Coukos who collaborated with me for the work described in Chapter 2, and he made a significant contribution to the characterization of MICA modification via model peptide experiments followed by measurements of the modification with LC-MS and NMR spectroscopy. I also want to thank Dr. Gang Li and Jeffrey E. Montgomery for their help and discussion of the molecular biology techniques, and thanks to Jun X. Huang for our collaboration to develop novel thermal profiling platform that globally screens the effect of

phosphorylation in target protein stability in cells. I sincerely thank all the other group members for their help and discussions.

I would also like to thank our research collaborators at the Scripps Research Institute who initiate the work described in Chapter 2 with us. Dr. Michael J. Bollong contributed to the work as a one of two co-first authors, and he significantly contributed to the discovery of the direct link between glycolytic enzyme PGK1 and KEAP1-NRF2 signaling axis via novel PGK1 inhibitor CBR-470-1. We brought our expertise together and finally report a beautiful story of our discovery in *Nature*, and I greatly appreciate his strong dedication to the research project. I also want to thank Dr. Peter G. Schultz and Dr. Luke L. Lairson who guided the project with Ray as corresponding authors. Dr. Hwayoung Yun contributed to the project by synthesizing and characterizing the molecules, designing and performing biochemical assays, and cell-based biological experiments. Emily N. Chin contributed to the project by designing, perform, and analyzing *in vivo* experiments of mouse skin UV damage model with Michael and Luke. I also want to thank Claudio Zambaldo, Insha Ahmad, and Arnab K. Chatterjee who contributed to the project by designing and performing biochemical assays and cell-based biological experiments.

I would like to thank my committee members of the thesis defense, Dr. Marsha R. Rosner and Dr. Chuan He. I greatly appreciate the time and effort they devoted and all the suggestions and comments they made. I also want to thank Dr. Sergey A. Kozmin and Dr. Jared C. Lewis. I am grateful for their time and effort to provide the references of my postdoctoral position application.

Finally, I would like to thank my family. My graduate student life in United States wouldn't go well without the support and help from my only one brother, Giman. I am greatly thankful for his constant love and support over the years. I dedicate my thesis to my parents. Pursuing my longing dream in science would not be processed without their endless love, support, and

encouragement. To me and my family, Chicago is a meaningful place. I grew up in Chicago for five years, during which my father lived the most fierce and challenging life for his doctoral program in Physics. My father's challenging for a dream ultimately motivated me to pursue the life of a scientist, and I truly wish to move forward to the next path in science from the city where I lived during my childhood and complete my doctoral program in 2019.

CHAPTER 1

Introduction

1.1 Deregulated glucose metabolism in human diseases

Since the first evidence of sugar catabolism was discovered from live yeast cells by Louis Pasteur in 1860s[1] and cell-free extracts in 1897 by Eduard Buchner[2,3], glycolysis has been characterized and studied over the century as a primary, major metabolic process[4,5] of the carbohydrates. In glycolysis, one glucose molecule is finally broken down into two molecules of pyruvate with the generation of two ATP and two NADH molecules, therefore the host cells obtain the source of energy and building blocks, which are necessary for sustaining the life, from this sequential enzymatic digestion processes.

One interesting feature of glycolysis is the generation of two ATP molecules without the involvement of molecular oxygen, therefore complete oxidation of glucose is not achieved only by this pathway. Pyruvate, a final product of glycolysis, is transferred into the mitochondria and further catabolized into CO_2 with the generation of ~36 ATP per glucose via tricarboxylic acid (TCA) cycle. A sequential carbohydrate oxidation processes, from glycolysis to TCA cycle, is called oxidative phosphorylation (OXPHOS) which is the highly efficient metabolic process to produce ATP in the aerobic organisms, including animal- and plant tissues. On the other hand, in anaerobic organisms or in tissues which are exposed to the limited oxygen, pyruvate is converted to lactate (e.g. in human muscle) or ethanol (e.g. in yeast). This different catabolic process of glucose is called, anaerobic glycolysis or fermentation.

Numerous published studies have demonstrated a strong connection between the deregulated glucose metabolism and various human metabolic diseases: cancer[6-9], diabetes[10,11]and

central obesity[12,13], neurodegenerations[14,15], and aging[16]. However, the mechanisms that integrate the metabolic state with the regulatory signaling pathways in biological systems, thus leading to understand the roles of deregulated glucose metabolism in diseases and develop the novel therapeutic strategies by targeting dysfunction in metabolic pathways, are largely unexplored.

Initially discovered by Otto Warburg in 1924[8], tumors have been observed to produce lactate from glucose even in the presence of sufficient oxygen that is called 'Warburg effect' or aerobic glycolysis. Many attempts to explain a Warburg effect has been performed. A model of imparted mitochondrial metabolism in cancer cells was originally proposed by Otto Warburg[17,18], but the results from following studies demonstrated that the activity of mitochondria in cancer cells is independent of aerobic glycolysis, indicating that alternative mechanism leads to the unique glucose metabolism in cancer[19-21]. Development and application of various metabolomic technologies, such as positron-emission tomography (PET) imaging with ^{18}fluorodeoxyglucose (FdG)[22-25] to trace the tissue's glucose uptakes *in vivo* and stable isotope-resolved metabolomics[26-28] with 2D-NMR or high-resolution mass spectrometry to trace actual metabolic fluxes in cells and *in vivo*, have led to the acceleration of the research of cancer glucose metabolism. The results of several published studies with empirical data and mathematic models have proposed the alternative models to explain Warburg effect[29]. FdG-PET imaging and quantitation studies revealed a prevalence of upregulated glycolysis in tumor cells[30,31] and a correlation of glucose uptake rate with the aggressive phenotypes[20], suggesting a model that Warburg effect induces powerful growth advantages to the tumor tissues by the alteration to its local microenvironment to occupy scarce nutritional resources[32,33]. Another model proposed by many published studies is that cancer cells have significant metabolic requirements for producing building blocks (e.g. acetyl-CoA,

nonessential amino acids, ribose) beyond the production of energy[6,20,34]. Indeed, lactate was recently discovered as an energy source of human lung tumor as evidenced by an lactate transporter expression and a metabolic flux of tumor tissues from extracellular stable isotope-labeled lactate[9]. The direct- and indirect links between the expression of oncogenes and the regulation of several glycolytic enzymes in cancer cells have been identified[7,35-38], however the complex connection between the altered glucose metabolism and tumor proliferation still need further investigation.

Type 2 diabetes (T2D) is another representative disease strongly correlated with impaired glucose metabolism in human body. Insulin, generated from pancreatic β cells, is a peptide hormone serving as a primary regulator of plasma glucose levels by stimulating glucose uptake in muscle, fat, brain cells and decreasing hepatic glucose production with increasing uptake[39] that lead to keep the overall glucose homeostasis with steady blood glucose concentration between 4 and 7 mM in normal body condition[40,41]. T2D occurs when hepatic cells, adipocytes, and muscle tissues become resistant to insulin that leads to increase blood glucose concentration in the body. Mechanistically, insulin stimulates translocation of glucose transporter GLUT4, thought to be a rate-limiting step in glucose uptake, from an intracellular compartment to the membrane[42] in skeletal muscle tissues and adipocytes[43]. Indeed, the study of adipose-selective GLUT4 knockout mice reported the insulin resistance occurred secondarily in muscle and liver, demonstrating a critical role of adipose tissues and their insulin-sensitive glucose transporter in glucose homeostasis *in vivo*[44]. Insulin also participates in reprogramming the glucose metabolism of hepatocytes. Liver-specific insulin receptor knockout mice failed to suppress hepatic glucose metabolism with insulin resistance and following severe glucose intolerance[45] and insulin resistance in central adipocytes was suggested to indirectly affect the hepatic glucose metabolism by supplying plasma free fatty acids (FFA) to the liver[46,47]. Published studies of insulin-signaling

pathways with cDNA and protein sequencing measurements demonstrated that insulin-signaling cascades directly regulate the metabolic enzymes by phosphorylation and affect the expression of metabolic enzyme levels involving in gluconeogenesis and glycolysis in hepatocytes[48,49]. Besides insulin, glycolysis has been reported to be involved in the regulation of hepatic glucose production and insulin secretion in pancreatic β cells, as evidenced by stable isotopic tracing of liver glucose metabolism in dysregulation of hepatic glucose production (HGP)[11,50,51] and effect of hexokinase family activity on insulin secretion in β cells[52-54]. Obesity is strongly associated with the risk of insulin resistance and T2D[13,55,56], however, understanding the mechanistic roles of glucose metabolism and specific metabolic signals in obesity, insulin resistance and T2D requires further studies.

Aging, which is characterized as the persistent and chronic accumulation of damages and stresses, has been recognized to be strongly associated with glycolysis, as recent measurements of FdG-PET suggested that age-related alteration of glucose metabolism in brain resulted in the loss of aerobic glycolysis[57,58] and the formation of advanced glycation end products (AGEs) by non-enzymatic reactions between proteins and the glycolytic intermediates resulted in skin aging[59]. Indeed, high-risk of glucose intolerance and insulin resistance has observed from older adults with altered glucose metabolism[60-62], indicating the effects of deregulated glucose metabolism in aging. Likewise, several studies have reported a correlation between neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) and cerebral glucose metabolism. Upregulation of glycolysis in astrocytes, the major non-neuronal glial cells, may resulted in the accumulation of amyloid beta protein of AD[63], and deregulation of metabolic enzymes involving in pentose phosphate pathway was observed in early stage of sporadic PD[64,65].

In summary, glucose metabolism has been reported to play a major role in various human metabolic diseases, but actual regulatory roles of the glucose metabolism and its connection to specific regulatory pathways in these metabolic diseases are not fully understood.

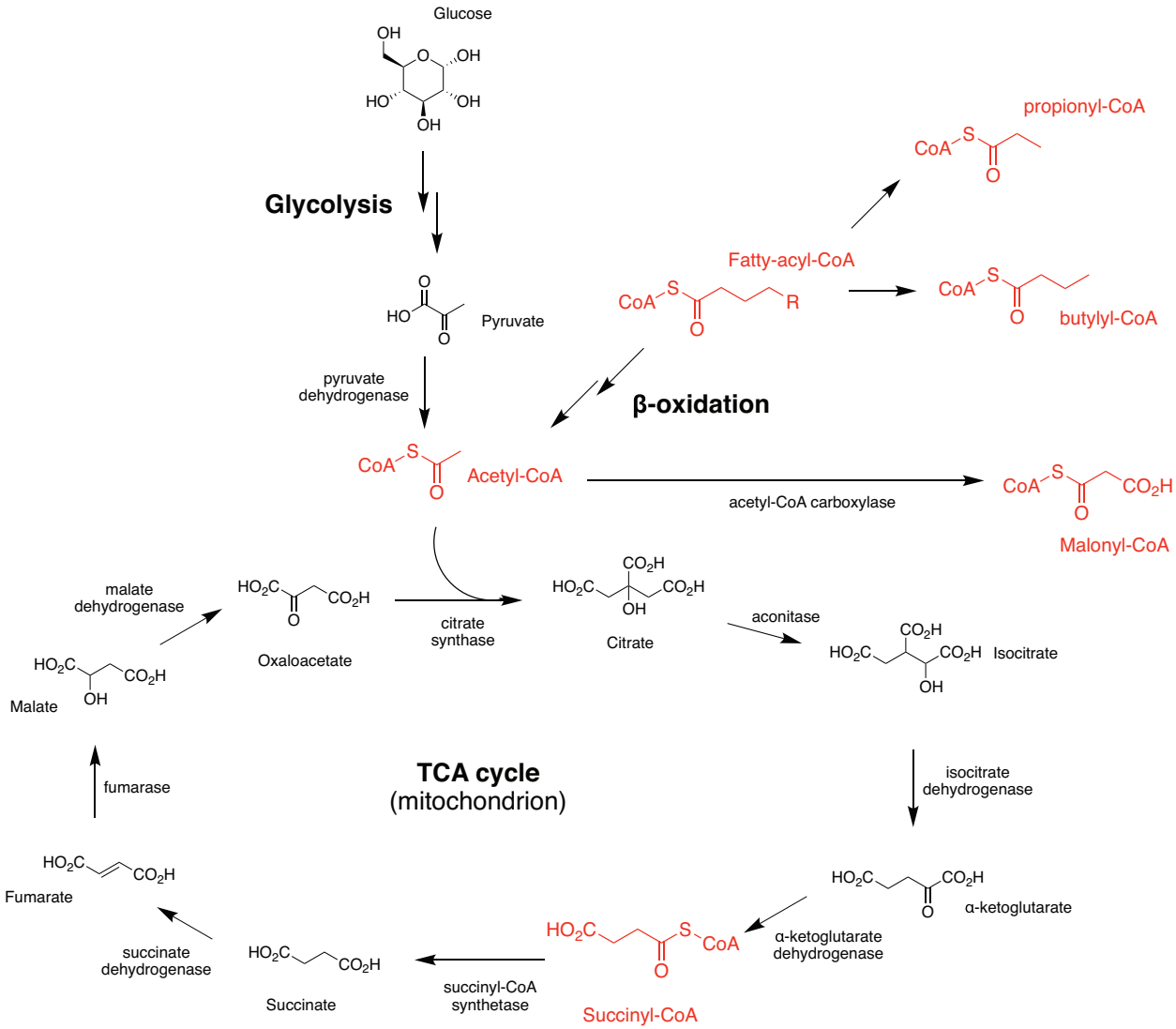


Figure 1.1 Generation of cellular acyl-CoA
Metabolic pathways that provide acyl-CoA metabolites which mediate acylation of lysine and cysteine residues.

1.2 Reactive metabolite-induced protein modifications and glucose metabolism

Along with the altered glucose metabolism in diseases, the most well-characterized biochemical means of conveying metabolic conditions in cells include the reversible allosteric modulation or inhibition of protein function by small molecule metabolites[66], as well as covalent posttranslational modifications of proteins[67]. However, the regulatory roles of primary and secondary metabolites in intra- or inter-pathway communication remains largely unexplored.

Recent studies of endogenous, intrinsically reactive metabolites have demonstrated that several of these metabolites are capable of forming functional, covalent modifications on proteins without the aid of enzymes, and regulate an array of cellular functions including metabolism and transcription.

Acyl coenzyme A (Acyl-CoA) group is a family of coenzymes generated from oxidative degeneration of fatty acids by Acyl-CoA synthetase with ATP and coenzyme A (CoA). Most of these metabolic intermediates were originally recognized to be involved in energy metabolism pathways, especially acetyl-CoA which is generated from both glycolytic metabolite pyruvate converted by pyruvate dehydrogenase and fatty acids broken down to acetyl-CoA through β-oxidation process is entering to TCA cycle in mitochondrion and succinyl-CoA is directly involved in TCA cycle to generate energy source (Figure 1.1)[68]. More recently, these coenzymes have been characterized as precursors of protein post-translational modifications (PTMs) that covalently modify target lysine and cysteine residues[69-72] (Figure 1.2) of the proteins including metabolic enzymes and regulate their activity and functions[70,71,73-76]. Especially, acylation of functional lysine residues of enzymes that are involved in glucose metabolism have been identified and observed to affect the overall metabolic flux, such as succinylation of isocitrate dehydrogenase (IDH)[77] and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in *E.Coli*[71] and acetylation of

all enzymes involving in TCA cycle and glycolysis except hexokinase (HK) and glucose-6-phosphate isomerase (GPI) of human leukemia cells[78] and liver tissues[76]. These results indicate the significant role of Acyl-CoA mediated protein modifications in glucose metabolism.

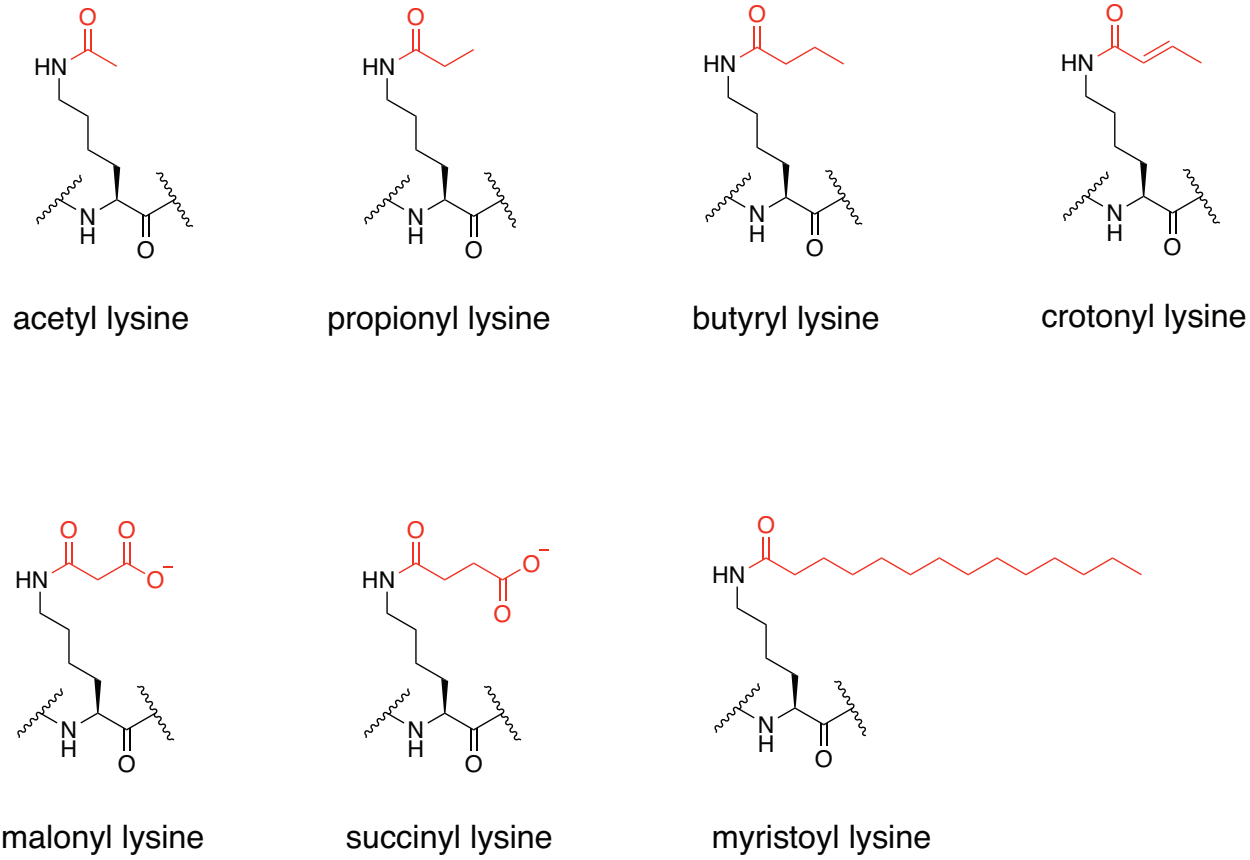


Figure 1.2 Acylation of lysine residues
Structures of protein modifications mediated by several acyl-CoA metabolites.

Acyl-CoA metabolites-mediated lysine modifications reversibly regulate the target protein functions by acylation and deacylation of functional lysine residues. Mechanisms of deacylation have been characterized by numerous studies, identifying that mammalian sirtuins (SIRTs) and

histone deacetylases (HDACs)[79-81] catalyze specific lysine modifications such as deacetylation by SIRT1-SIRT3[82,83], demalonylation, desuccinylation, and deglutarylation by SIRT5[70,74,77], and long-chain acyl groups by SIRT6[84]. Efforts to elucidate the lysine acylation mechanisms have led to identify histone acetyltransferases (HATs) that specifically catalyze the acetylation of lysine residues with acetyl-CoA[69] as well as non-enzymatic modifications of lysine residues by acyl-CoA intermediates[85].

Acyl-CoA intermediates possess thioester bond which is highly reactive electrophile that induces non-enzymatic acetylation of active lysine residues in target proteins, observed in histones[86] and mitochondrial matrix[87,88] and mediated by sirtuins[89]. Among the study of non-enzymatic modifications by acyl-CoA group, non-enzymatic acetylation has been characterized that the efficiency of acetylation with acetyl-CoA is highly affected by physiological condition of specific environments. The level of acetylation was increased by higher pH[90] in mitochondrial matrix[91] which has higher pH (pH 7.9-8)[92] and concentration of acetyl-CoA (0.1-1.5 mM in *E.Coli*)[93-95] than cytosol and nucleus (pH 7.2, 3-30 μM acetyl-CoA), indicating that non-enzymatic modifications of the proteins rely on chemical reactivity of the both active lysine residues and acetyl-CoA thioester bond in the specific condition, yet stoichiometry of acetylation in mammalian cells has not performed. Intriguingly, recent study of acylation identified a certain species of acyl-CoAs that form 5- or 6-membered ring cyclic anhydride intermediates by base catalysis-mediated intramolecular reaction between terminal carboxylic acid and thioester bond of acyl-CoA in pH over pKa of carboxylic acid (pH 4-5)[85]. Succinyl-CoA, glutaryl-CoA, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), and 3-methylglytaryl-CoA were observed to form highly reactive cyclic anhydride intermediates, and *in vitro* comparison of non-enzymatic modification rate between succinyl-CoA and acetyl-CoA indicated that cyclic anhydride is more reactive than

thioester to acetylate the protein lysine residues. Indeed, this study demonstrated that functional lysine residues of metabolic enzymes in TCA cycle are main targets of these acylation, demonstrating that acyl-CoAs which are generated from metabolic pathways directly regulate the glucose metabolism in cells. Recent application of activity-based thioester profiling[96] and chemo-proteomic experiments[97] were performed to identify the non-enzymatic processes in complex biological system directly and understand the features of non-enzymatic acylation, such as kinetics, sites of modification, and local environment, implying that elucidation of the nature of non-enzymatic PTMs necessitates novel chemical proteomic strategies.

Representative example of reactive metabolite-induced posttranslational modifications (rmPTMs) by primary glycolytic metabolite is non-enzymatic lysine acylation by 1,3-bisphosphoglycerate (1,3-BPG) discovered by Moellering and Cravatt in 2013[98]. Their studies demonstrated that 1,3-BPG, one of primary glycolytic metabolites possessing a strong electrophilic acyl phosphate substrate that is inherently more reactive than acyl-CoA thioester[89], induced the acylation of target lysine residues to form 3-phosphoglyceryl-lysine (pgK) (Figure 1.3). The level of pgK modifications of target proteins and whole proteome in cells have shown to be directly proportional to the level of glucose uptake and the level of 1,3- and 2,3-BPG isomers, a source of pgK modification. Indeed, pgK modifications on target proteins altered their physical properties (pH and charges) and regulate the functions, as evidenced by the inhibitory effects of pgK modification on target glycolytic enzymes observed from the results of *in vitro* enzymatic assay and targeted metabolomics with the $^{13}C_6$-glucose. To sum up, the studies of 1,3-BPG induced protein modifications indicated that reactive-metabolites may be a signaling messenger which dynamically and directly reflect the information of nutrient uptakes and convert the information to appropriate responses keep the homeostasis in the cellular system.

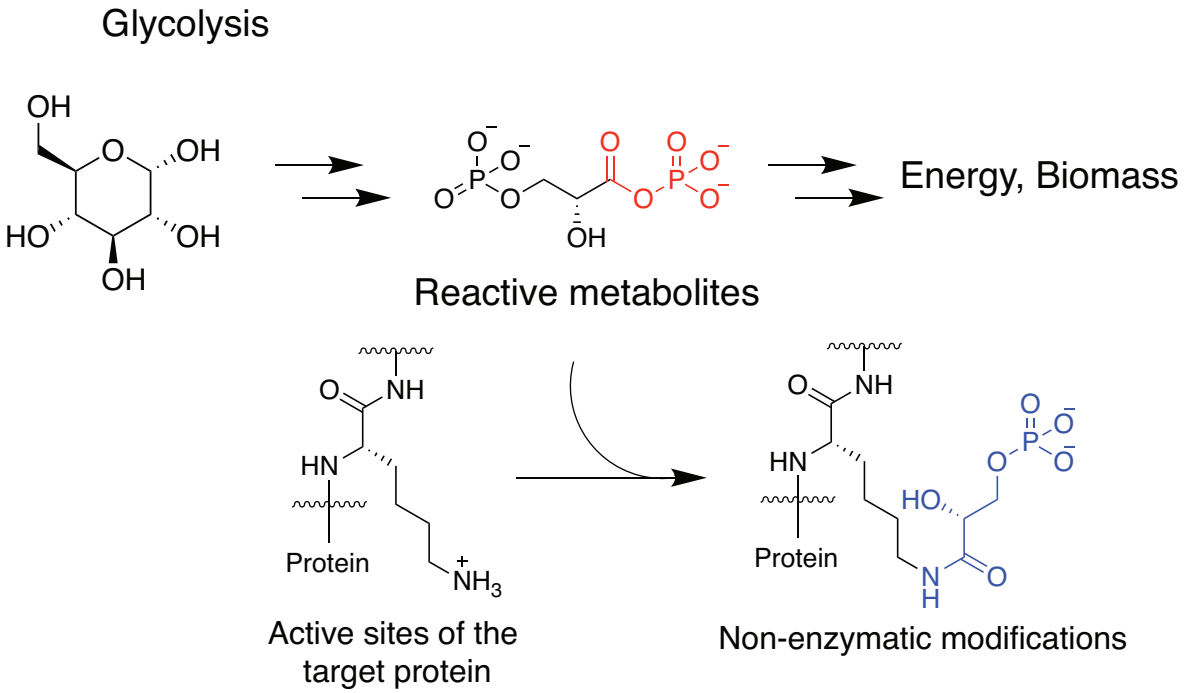


Figure 1.3 Phosphoglyceryl-lysine (pgK) modifications
A primary glycolytic metabolite which possesses intrinsic reactivity induces non-enzymatic reaction of lysine residues on target proteins, representatively indicating a direct connection between glycolytic flux and alterations in the structure and function of key proteins, eventually lead to the phenotypic influence on the biological system.

Glycation is a non-enzymatic modification of the protein residues by sugar molecules with inherent reactivity. Carbonyl groups of sugar molecules induce covalent modifications of nucleophilic amino acid residues of the proteins by complex steps of chemical reactions such as Schiff base reactions[99,100], Amadori rearrangements[101], and Maillard reactions[102] that lead to advanced glycation end-products (AGEs)[103]. Methylglyoxal (MGx) is a glycation inducer mostly produced from decomposition of glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone

phosphate (DHAP) in glycolysis. MGx possesses α,β-dicarbonyl moiety which is a highly reactive electrophile, and several studies characterized the various AGEs to nucleophilic amino acid residues including lysine, arginine, and cysteine by MGx (Figure 1.4)[104,105]. The correlation between endogenous MGx levels and tumor growths has been reported that increasing doses of cellular MGx, but lower levels, enhance cancer-promoting effects[106-109]. These observations may indicate the role of MGx and MGx-induced AGEs in cancer proliferation.

Likewise, many non-enzymatic PTMs in proteome, generated by the intermediates of metabolism, co-factors, and ROS, have been recognized to closely correlate with glucose metabolism. For instance, the loss of fumarate hydratase (FH) activity by mutations resulted in the accumulation of intracellular fumarate which is capable of inducing succination of hyperreactive cysteine residues of target proteins by non-enzymatic reactions[110-112]. As the loss-of-function mutations of FH has been known to associate with the development of leiomyomata, renal cysts, and tumors, TCA cycle intermediate fumarate is suggested as a oncometabolite with its regulatory roles in target proteins[113] including inhibitory effect of glycolytic enzyme GAPDH[112]. Another example is the regulation of glycolytic enzyme GAPDH by non-enzymatic S-glutathionylation to cysteine 152 residue which is inhibited by the treatment of antioxidant Vitamin C in *KRAS* and *BRAF* mutant cells[114]. Indeed, S-sulfhydration of GAPDH cysteine 150 residue by hydrogen sulfide (H_2S) in liver tissue was reported under physiological condition with the role of this modification on increasing catalytic activity of GAPDH[115].

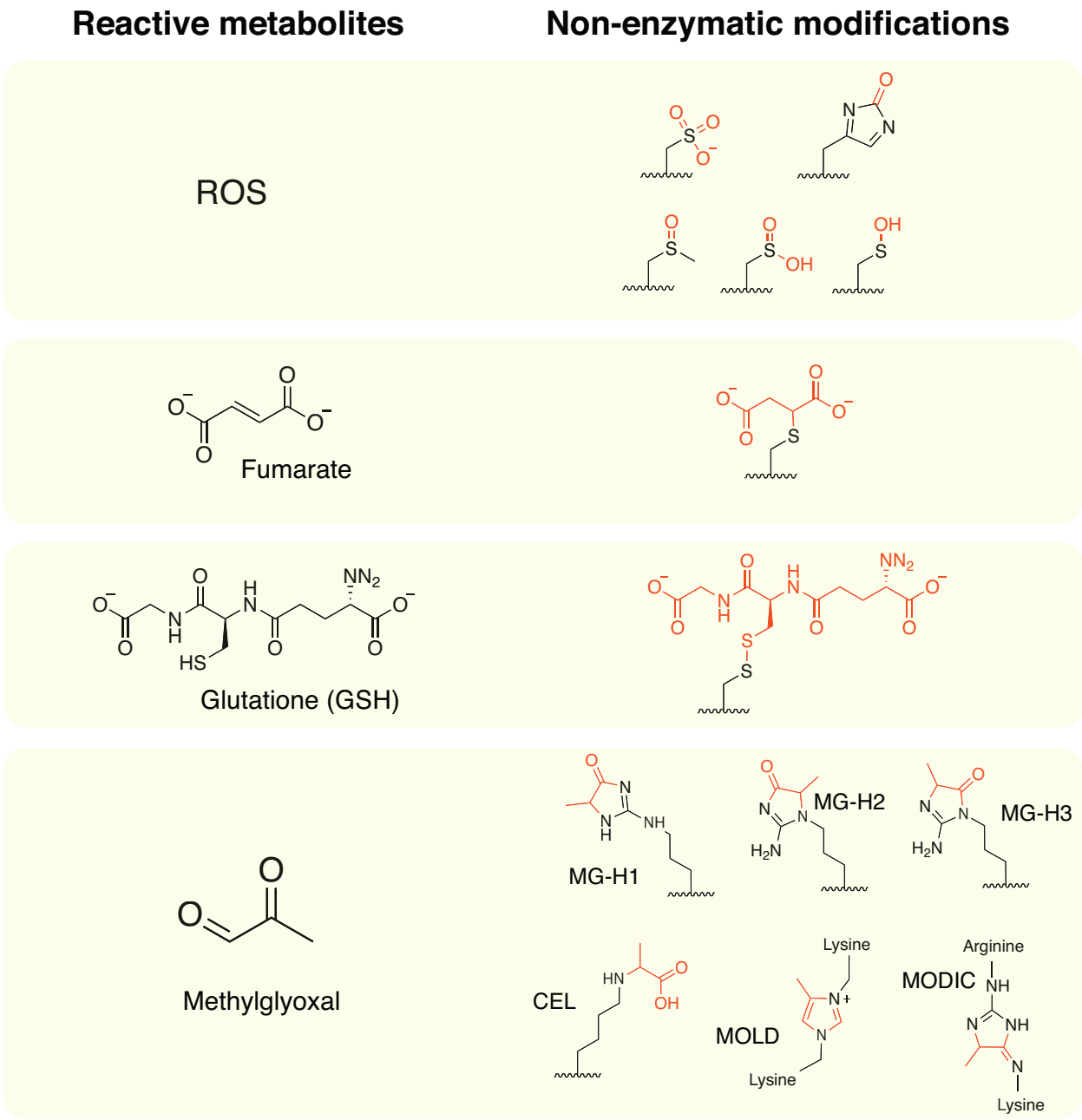


Figure 1.4 Reactive metabolites and corresponding rmPTMs in cells
A list of reactive metabolites and the structures of their corresponding non-enzymatic protein modifications which have been reported.

Intriguingly, study of rmPTMs including pgK modification, acylation of lysine residues by acyl-CoA, succination, and glycation by methylglyoxal indicates the unique features of these non-enzymatic protein modifications (Figure1.5).

1) Specificity and efficiency of non-enzymatic PTMs might be governed by intrinsic reactivity of the metabolites and sites of the modification in proteins[85,96-98,110], local environment such as pH[90-92], and amount of the precursor[93-95]. These specific modifications regulate the function and activity of the proteins in cells[78,98,112,114].
2) Spatial compartmentalization of non-enzymatic PTMs has observed[87,88,98], demonstrating that diffusion of reactive metabolites to proximal proteins may occur ahead of the modifications of these target proteins.
3) Specific rmPTMs such as labile glycations by MGx[116], pgK modifications[117], and acylation of lysine residues[89] are reversibly regulated by erasing mechanisms.
4) Regardless of the presence of discernible motifs, rmPTM sites and target proteins are highly conserved throughout evolution[91,98,110].

Therefore, it is conceivable that reactive metabolites may serve as intracellular messengers, coupling metabolic status to other pathways via the formation and functional outcomes of rmPTMs on specific sites within one or more proteins.

Taken together, reactive metabolites may play a role as messengers in cell signaling and directly link between the metabolism and regulatory pathways of the cell. Future study of rmPTMs will focused on elucidation of the underlying networks of the proteome communicated by reactive metabolite signals and identify their roles in biological system and in diseases.

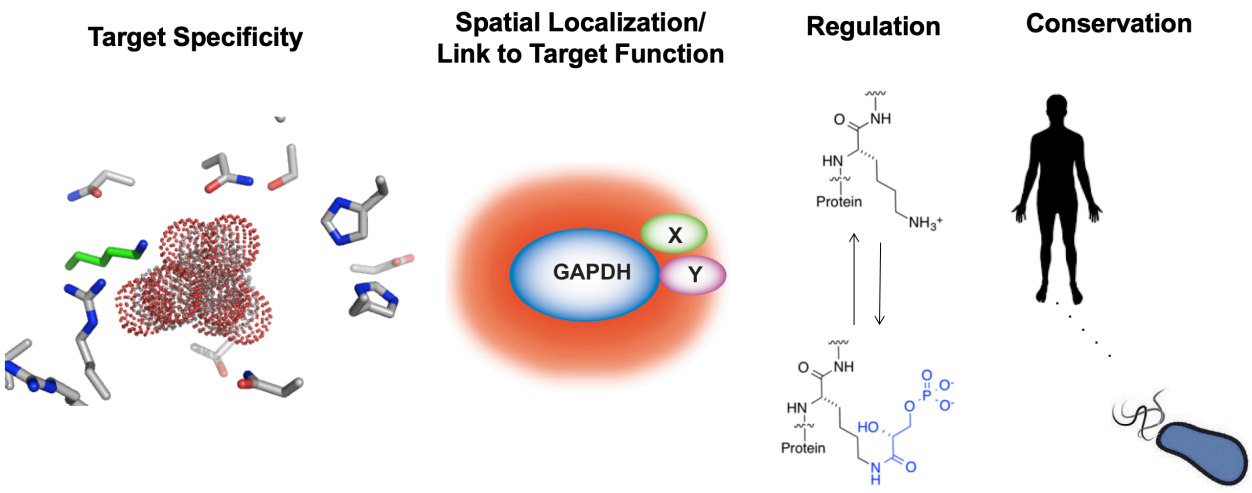


Figure 1.5 Empirical regulatory mechanisms for pgK modifications
Non-enzymatic posttranslational modifications via intrinsically reactive metabolites have been shown to occur in the specific residues of target proteins and regulate the functions of the targets in living system.

1.3 Scope of the dissertation

My dissertation will address the fundamental questions of reactive metabolite-induced protein modifications in cell signaling: Do reactive metabolites signal to glucose-responsive pathways? What are the roles of rmPTMs in disease? Will rmPTMs be therapeutic targets? To find the answers to these questions, I was mainly involved in developing novel chemical proteomic technologies and utilizing them to elucidate the direct link between glycolysis and central stress modulator program in cells.

Chapter 2 discusses our recent discovery of the direct communication between glycolysis and KEAP1-NRF2 signaling axis mediated by a reactive metabolite methylglyoxal.

Chapter 3 discusses the development of novel protein proximity labeling tool with quantitative proteomics to discover the dynamic protein-protein interaction networks.

Chapter 4 discusses the development of new quantitative proteomic workflow to profile the protein-ligand interactions in live cells.

Chapter 5 summarizes the future aspects of each research project in respond to the study of reactive metabolite-induced posttranslational modifications and envisions the future research directions.

CHAPTER 2

Reactive metabolite-induced protein modification integrates glycolysis with NRF2 signaling*

The work of this chapter is published: Bollong# and Lee# et. al*., Nature, 562,* 600-604 (2018).

2.1. Introduction

Elucidation of the regulatory roles of the reactive metabolite-induced protein modifications in cellular system also necessitates the study of 'receiver' proteins which perceive those specific metabolites and convert the metabolic information to the appropriate responses.

2.1.1. KEAP1-NRF2 pathway: central regulator of cell stress

The representative sensor protein in the cellular system that regulates the stress via recognizing various reactive species is KEAP1 (Kelch-like erythroid cell-derived protein with CNC homology-associated protein). KEAP1 is a cysteine-rich protein containing 27 cysteine residues among 624 amino acids of human protein, and sensitive cysteine residues (e.g. C151, C273, C288) specifically sensing electrophilic molecules in cells have been known to regulate

* The author's contribution to the studies presented in this chapter: The author contributed equally to this work with Michael J. Bollong as a one of two co-first authors, discovering 'Methylglyoxal' as a signaling molecule of KEAP1-NRF2 signaling axis and identifying the sites and structure of the MICA modification in KEAP1. The author designed, performed, and analyzed biochemical assays, cell-based biological experiments, metabolomic, and all proteomic experiments described in this chapter. The author contributed to writing the manuscript of published paper.

KEAP1 conformations and behaviors (Figure 2.1a)[118,119]. KEAP1 forms non-covalent homodimer in cytosol and KELCH repeats domain of the KEAP1 binds to the transcription factor NRF2 (nuclear factor erythroid 2 like 2, also called NFE2L2).

As a BTB family protein forming a E3-ubiquitin ligase complex with Cullin3 and RBX1, the role of KEAP1 in the KEAP1-NRF2 signaling is to regulate cellular NRF2 levels via inducing ubiquitination of NRF2 transcription factor followed by proteolytic degradation of this ubiquitinated protein in normal condition (Figure 2.1b)[120,121]. Upon exposure to the increasing level of various chemical insults causing cellular stresses, covalent modifications of the cysteine thiol sensor residues of KEAP1 with those reactive species initiate the activation of KEAP1-NRF2 pathway by inducing accumulation of NRF2[122-124] followed by the trans-localization of NRF2 transcriptional activator into nuclei. Then, NRF2 forms heterodimer with small MAF proteins and activates target genes for the expression of antioxidant-responsive elements (AREs), resulting in cytoprotective effects through a variety of mechanisms such as generating glutathione (GSH), reducing cellular ROS levels, detoxifying xenobiotics, and expressing drug transporters[125]. Overall, the KEAP1-NRF2 pathway is a key regulator to keep cellular homeostasis by sensing and limiting cellular damages under normal and deregulated metabolic states.

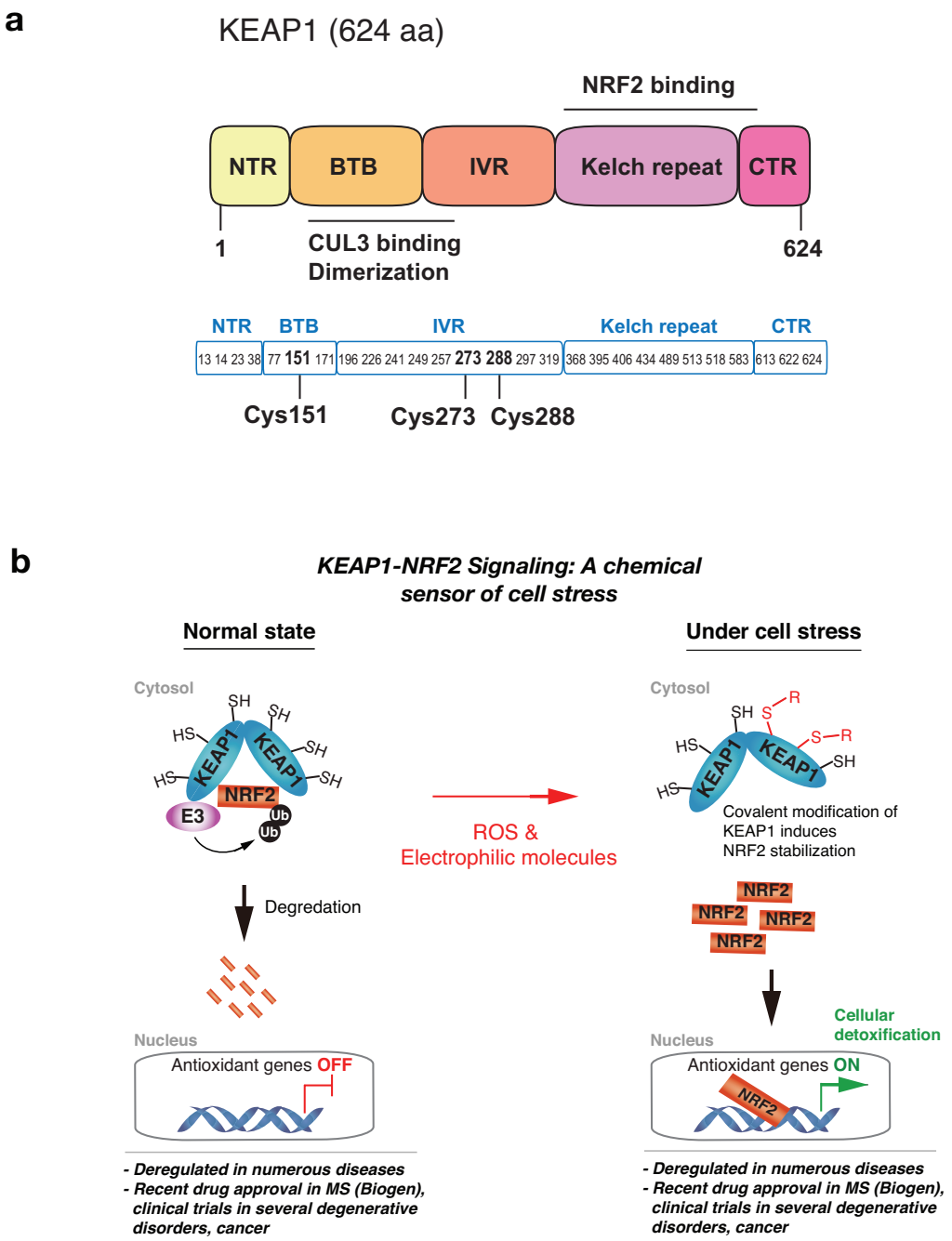


Figure 2.1 KEAP1-NRF2 signaling serves as a central regulator of cellular stress
a. KEAP1 consists of 624 amino acid residues and has five domains. BTB domain involves in mediating its homodimerization and binding with Cullin3, and KELCH domain has interaction with NRF2 transcription factor. Also, Keap1 is a cysteine-rich protein and reactive cysteine residues have been known as sensitive sensors of oxidants and electrophilic species. **b.** KEAP1 captures specific metabolic information by covalent modification of reactive cysteine sensor residues. Modifications induce conformational changes of KEAP1 that leads the accumulation of NRF2 which initiates the transcription of the antioxidant response element (ARE) to detoxify the system.

2.1.2. Regulation of the KEAP1-NRF2 pathway in diseases

Numerous studies have demonstrated a strong correlation of deregulated KEAP1-NRF2 pathway, as known to be caused by somatic mutations of KEAP1 or NRF2, level of cysteine modifications and KEAP1 expressions, and activity of binding partners affecting KEAP1-NRF2 interactions and downstream signaling cascades, to a variety of human diseases, including cancer[118,126], diabetes[127,128], chronic inflammatory diseases[129], Alzheimer's and Parkinson's diseases[130], and aging[131].

While the exact mechanism explaining how covalent modifications of cysteine sensors in KEAP1 induce the structural and behavioral alterations of KEAP1 that lead to the loss of its catalytic activity toward NRF2 proteolytic degradation is still enigmatic, even though several models have been suggested[119,132,133], the efforts to therapeutically target the KEAP1-NRF2 signaling activation for antioxidant therapy have largely focused on covalent small molecule agonist of KEAP1 cysteine sensor residues.

The covalent modifications of highly reactive cysteine thiols of KEAP1 depend on the nature of electrophiles, and currently identified xenobiotic inducers of NRF2 signaling via targeting KEAP1 modifications are chemically classified into several classes with their electrophilic substrates, including Michael acceptors, oxidizable diphenols and quinones, and isothiocyanates and sulfoxythiocarbamates (Figure 2.2)[134].

Michael acceptor is a prominent class of KEAP1 cysteine alkylators. As sorted to soft Lewis acids, Michael acceptors enable the alkylation of reactive cysteine thiols possessing low pKa values. Dimethyl fumarate (DMF, trade name Tecfidera) is an FDA-approved molecule and has been thought to modulate immune-response with the activation of NRF2 signaling axis. An

α,β-unsaturated electrophilic moiety of DMF and its metabolized product monomethyl fumarate (MMF) could be involved in alkylation of reactive cysteine thiols in a Michael addition reaction[135-137], yet the precise mechanisms of action and functions of DMF in live cells are still undefined. Another representative example of Michael acceptors is a Phase III clinical candidate Bardoxolone methyl (BARD)[138] which was evaluated for type 2 diabetes (T2D) in CDK patients[139,140]. The adoption of highly reactive Michael acceptor moiety (α,β-unsaturated α-cyano ketone) to the oleanolic acid scaffold of BARD results in potent NRF2 activation with specifically targeting the alkylation of C151 residue in KEAP1[141]. More recently, AI-1, a quinolinone derivative discovered in a cell-based high throughput screening (HTS) study, has been demonstrated to alkylate KEAP1 C151 residue through a Michael addition-elimination reaction[142].

Tert-butyl hydroquinone (TBHQ)[143], a compound of phenolic antioxidant family, undergoes oxidation to tert-butyl quinone (TBQ) in presence of oxygen and transition metals (e.g. Copper redox oxidation)[144] followed by quick Michael addition-mediated alkylation of KEAP1 cysteine residues to activate KEAP1-NRF2 signaling axis. Isothiocyanates sulforaphane (SFN) is well known natural compound of the chemical family sharing highly electrophilic isothiocyanate group (-N=C=S) that forms dithiocarbamate modifications[145] of reactive cysteine thiol residues in KEAP1. Indeed, iodoacetamide-alkyne (IAA-alkyne) probe[146], which is cysteine reactive, has been widely used to identify reactive cysteine residues in activity-based protein profiling (ABPP) studies[147,148].

The nature of intrinsic reactivity of these direct KEAP1 modulators has been associated with the issues of a range of toxicities, potentially associated with the accumulation of off-target covalent adducts resulting in potential adverse effects. Therefore, investigation of alternative

strategies in activation of KEAP1-NRF2 pathway, instead of direct leveraging of KEAP1, is required to discover a breakthrough in modulating cellular stresses for future clinical applications.

For the discovery of the novel regulatory nodes in the KEAP1-NRF2 pathway, we performed a cell-based, high-throughput phenotypic screen to identify small molecules that activate NRF2 signaling by mechanisms that are independent of direct KEAP1 covalent modifications, the default route for pathway activation. Herein, we report the discovery of a novel inhibitor of the glycolytic enzyme PGK1, which activates the KEAP1-NRF2 pathway activation by inducing a reactive metabolite-mediated posttranslational modification of KEAP1. This modification involves the unprecedented crosslinking of proximal reactive cysteine and arginine residues by the electrophilic metabolite methylglyoxal and provides evidence for a direct link between glycolysis and the NRF2 transcriptional cascade[149].

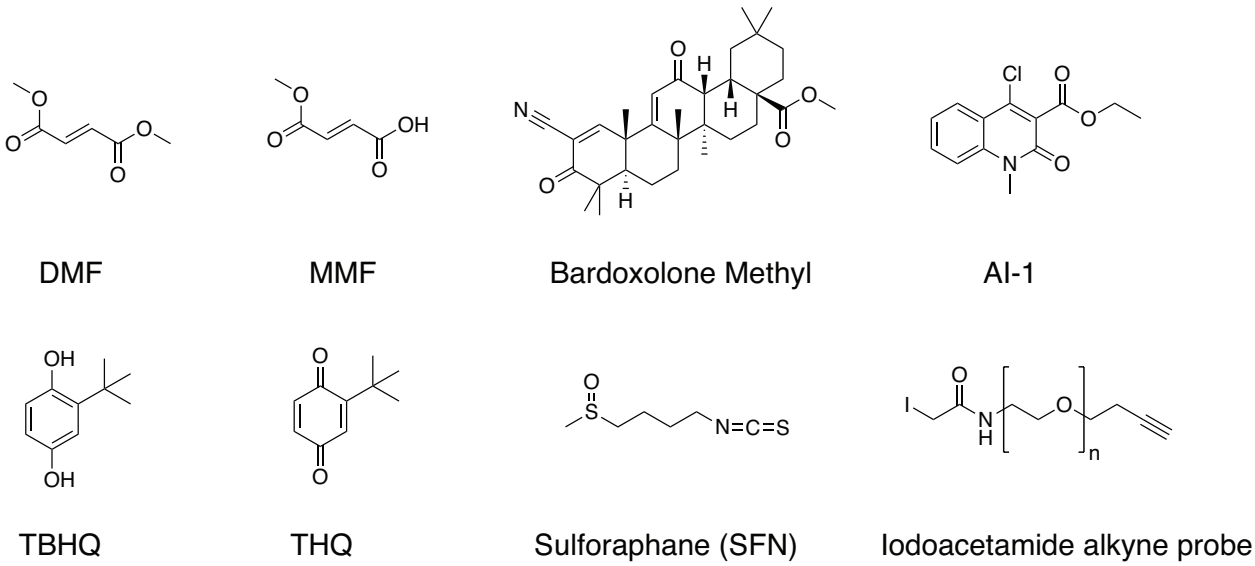


Figure 2.2 Targeting KEAP1-NRF2 activation by direct KEAP1 modification
Structure of KEAP1 cysteine alkylators.

2.2. Results

2.2.1. The novel KEAP1-NRF2 signaling activator CBR-470 series

Since the majority of KEAP1-NRF2 signaling activators possesses the electrophilic features inducing direct KEAP1 modifications, the discovery of non-electrophilic activator series implies the existence of alternative regulatory mechanisms in KEAP1-NRF2 signaling. We performed a cell-based, high-throughput phenotypic screen and identified a new class of NRF2 activators.

2.2.1.1 Identification of a non-electrophilic NRF2 activator

Cell based high-throughput screen assay was designed to read the ARE inducing activity by luciferase signals. To set up the assay, highly NRF2-responsive neuroblastoma IMR32 cell line was plated and then luciferase reporter construct (pTI-ARE-LUC) that bears the antioxidant responsive element (ARE) from *NQO1* gene[142] was transfected to the cell. Around 30,000 heterocyclic compounds from the library were initially screened with the assay, and 27 small molecules by which ARE-LUC signals are dose-dependently increased at least 5-fold above the plate mean with half maximal effective concentrations (EC_{50}) in the low micromolar range were identified (Figure 2.3a). Among these compounds, one small molecule which does not contain any obvious electrophilic substrates, termed CBR-470-0, induced a high magnitude of ARE-LUC signal with the similar levels of previously reported KEAP1 modifiers tert-butylhydroquinone (TBHQ) and AI-1 (Figure 2.3b and 2.3c). We suspected that CBR-470-0 induces NRF2 activation with the alternative mechanism, thus, the activity of CBR-470-0 was investigated further.

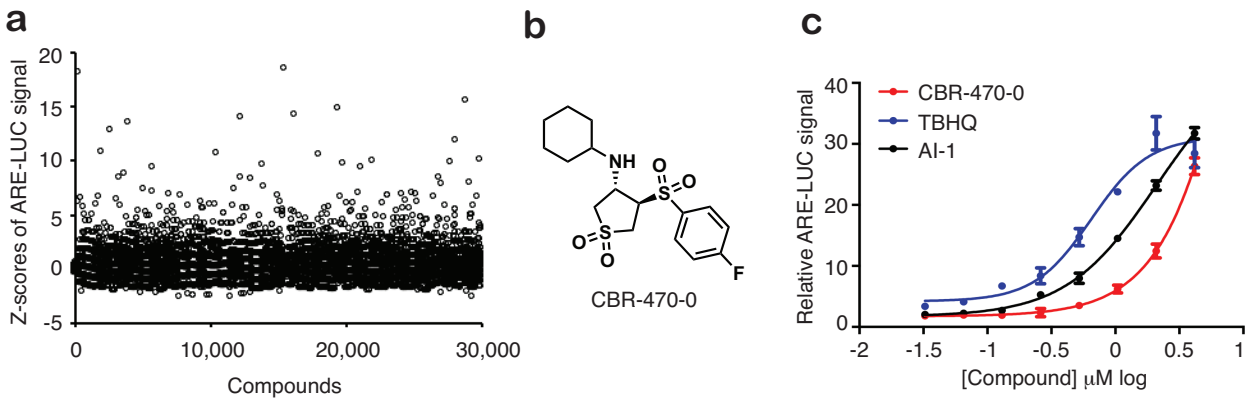


Figure 2.3 A high throughput screen identifies a novel NRF2 activator CBR-470-0 which activate a robust NRF2 transcriptional program
a. Plate-based Z-scores of ARE-LUC luminance measurements of all test compounds from a 30k compound screen in IMR32 cells. **b.** Structure of screening hit CBR-470-0. **c.** Relative ARE-LUC luminance measurements from IMR32 cells treated for 24 hours with a concentration response of CBR-470-0 and reported NRF2 activators TBHQ and AI-1 (n=3 biologically independent samples, mean and s.e.m.).

To develop the small molecule series of CBR-470 with higher potency, as well as identify the substrates of the compound enabling to be substituted for target engagement study, structure activity relationship (SAR) study was performed around the cyclic sulfone scaffold of CBR-470-0. Consequently, we further identified more potent CBR-470 derivatives including CBR-470-1, an isobutylamine substituted analogue that was displayed a five-fold increase in potency (EC50 = 962 nM) compared to CBR-470-0 in ARE-LUC assay (Figure 2.4a and 2.4b). Indeed, CBR-470-1 was stable in the mixture of high concentration of reduced glutathione (GSH), indicating that the compound is not reacting with nucleophilic thiol cysteine residues and is stable in live cells (Figure 2.4c).

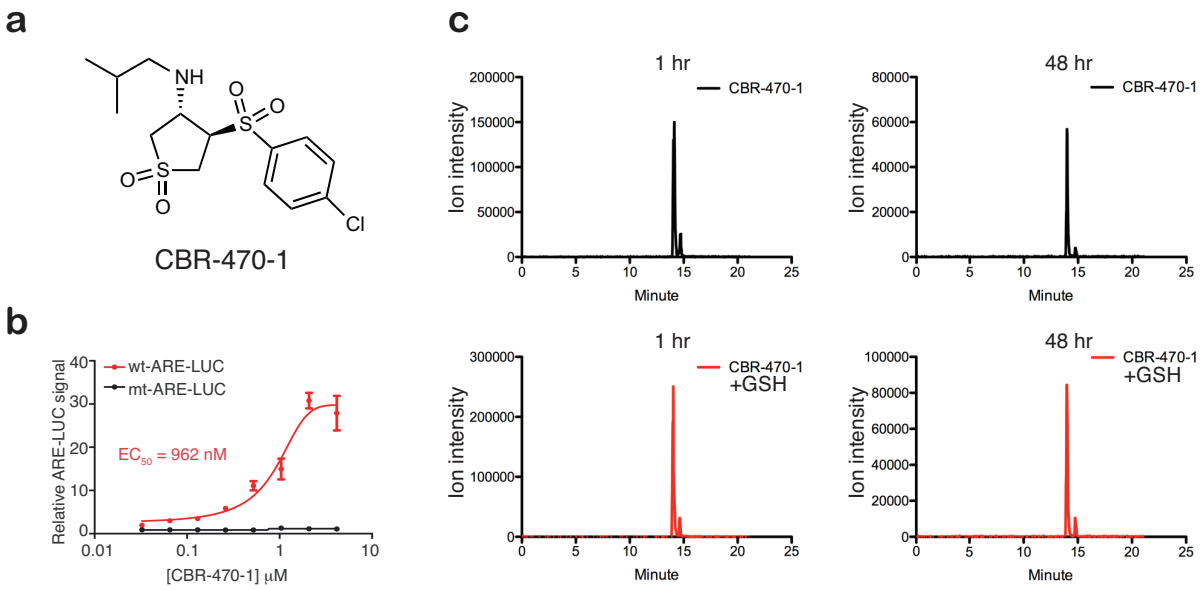


Figure 2.4 Development of CBR-470-1
a. Structure of CBR-470-1. **b.** Relative ARE-LUC luminance values from IMR32 cells transfected with wild type (wt) or mutant (mt, two core nucleotides necessary for NRF2 binding were changed from GC to AT) ARE-LUC reporter constructs and treated with the indicated doses of CBR-470-1 for 24 hours (n=3, mean and s.e.m.). **c.** LC-MS quantification of CBR-470-1 (50 μM) treated in the presence or absence of GSH (1 mM) in PBS for 1 hour (left) and 48 hours (right). Relative ion intensities within each time point were compared with representative chromatograms shown (n=2)

Western blot and qRT-PCR analyses revealed that CBR-470-1 induces the accumulation of NRF2 proteins in IMR32 cells with a dose- and time dependent manners and also increases the levels of both mRNA and protein of *NQO1* and *HMOX1*, representative NRF2-response genes (Figure 2.5a, 2.5b, and 2.5c). We also performed RNA-sequencing expression profiling of IMR32 cells and applied gene set enrichment analysis (GESA) to the data to read transcriptome-wide alterations of the cells induced by CBR-470-1. After 24 hours of CBR-470-1 treatment to cells and following data analysis, the results showed that NRF2 target genes, which were validated by

focused qRT-PCR analysis, are the most significantly enriched genes among cell transcriptome (ES = 0.83, P < 0.0001) (Figure 2.5d and 2.5e).

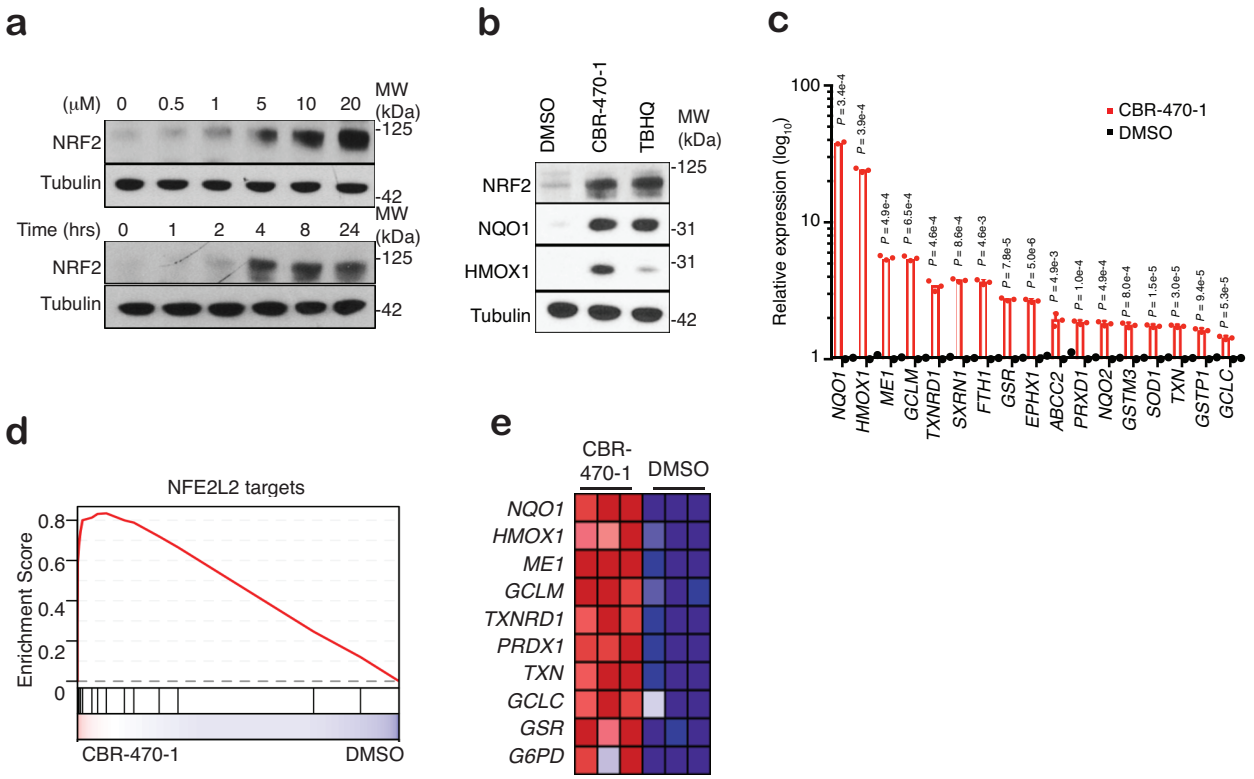


Figure 2.5 CBR-470-1 activate NRF2 signaling in IMR32 cells

a. NRF2 protein levels from IMR32 cells treated with the indicated concentrations of CBR-470-1 for 4 hours (top) or 5 μM CBR-470-1 for the indicated time periods (bottom). Blots are representative of 3 independent experiments. **b.** Western blot analyses of total NRF2 protein content or NRF2-controlled genes (*NQO1*, *HMOX1*) from IMR32 cells treated for 24 hours with 5 μM CBR-470-1 (n=3, mean and s.d.). **c.** Relative abundance of NRF2-dependent transcripts as determined by qRT-PCR from IMR32 cells treated for 24 hours with 5 μM CBR-470-1 (n=3). **d.** GSEA enrichment plot depicting the enrichment of a NRF2 target gene set ("Singh_NRF2L2_Targets" in MSigDB) from IMR32 cells treated for 24 hours with 5 μM CBR-470-1 (n=3, P < 0.0001, nom. p-value in GSEA). **e.** Heat map representation of the leading-edge subset of the most upregulated NRF2-regulated transcripts upon CBR-470-1 treatment. Data are biologically independent samples.

Besides IMR32 cells, enhanced NRF2 protein stabilization as well as *NQO1* and *HMOX* transcript levels were also observed in other mammalian cell lines HEK293T (~2.5-fold increase of both *NQO1* and *HMOX1*, $P < 0.001$), SH-SY5Y (~13-fold increase of *NQO1* and ~24-fold increases of *HMOX1*, $P < 0.001$), and primary human lung fibroblasts (HLF, ~ 2.2-hold increase of *NQO1*, $P < 0.001$) with CBR-470-1 treatment (Figure 2.6a and 2.6b). The data indicate that the activation of NRF2 pathway induced by CBR-470-1 is universal and not restricted to highly NRF2-responsive cell lines.

To see if CBR-470-1 activity relies on NRF2, we performed a genetic depletion of NRF2 protein levels in cells by transfections of corresponding shRNA and then we treated CBR-470-1 or TBHQ to HEK293T cells. ARE-LUC signals from the cell-based reporter assay were significantly dropped with the knockdown of NRF2, indicating that CBR-470-1 does not induce an alternative, compensatory transcriptional program but NRF2 pathway (Figure 2.6c).

To test the cytoprotective effect by CBR-470-1-induced activation of NRF2 pathway in cells, CBR-470-1 was pretreated to SH-SY5Y cells for 48 hours to induce NRF2 activation and then cell permeable peroxide TBHP was treated to cells with increasing doses. Data from cell viability measurements of these cells indicate that the non-electrophilic NRF2 activator CBR-470-1 protects SH-SY5Y cells with similar levels of efficacy to the direct KEAP1 modifier TBHQ upon under peroxide induced cell stress (Figure 2.6d).

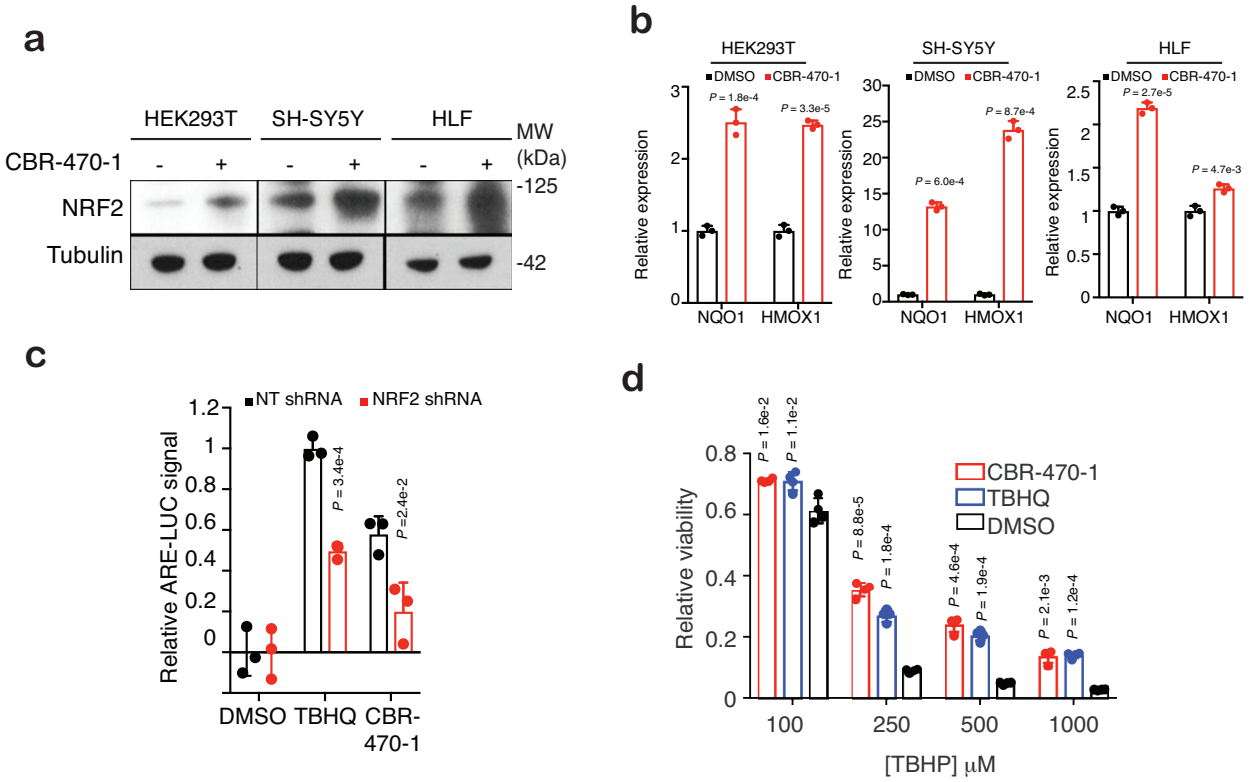


Figure 2.6 CBR-470-1 activate NRF2 signaling in other cancer cells
a. Western blot analyses of total NRF2 protein content from the indicated cell types treated for 4 hours with 5 μM CBR-470-1 (n=3). **b.** Relative expression levels of NQO1 and HMOX1 from the indicated cell types treated for 24 hours with 5 μM CBR-470-1 (n=3, mean and s.d.). **c.** Relative ARE-LUC luminescence values from HEK293T cells treated with the indicated shRNA constructs and pTI-ARE-LUC and then treated with TBHQ (10 μM) or CBR-470-1 (5 μM) for 24 hours (n=3). **d.** Relative viability measurements of SH-SY5Y cells treated with either CBR-470-1 (5 μM) or TBHQ (10 μM) for 48 hours and then challenged with the indicated doses of tert-Butyl hydroperoxide (TBHP) for 8 hours (n=4). Data are mean and s.d. of biologically independent samples ($P^{*} < 0.05$, $P^{**} < 0.005$, $P^{***} < 0.001$, univariable two-sided t-test).

2.2.1.2 CBR-470-2 activates a cytoprotective NRF2 program *in vivo*

From the previous cell-based high-throughput assays, SAR studies, Western blot and mRNA expression analyses identified a non-electrophilic activator of NRF2 signaling, CBR-470-1. With animal model *in vivo* experiments, we wondered if the molecule was physiologically relevant thus implying the therapeutic efficacy.

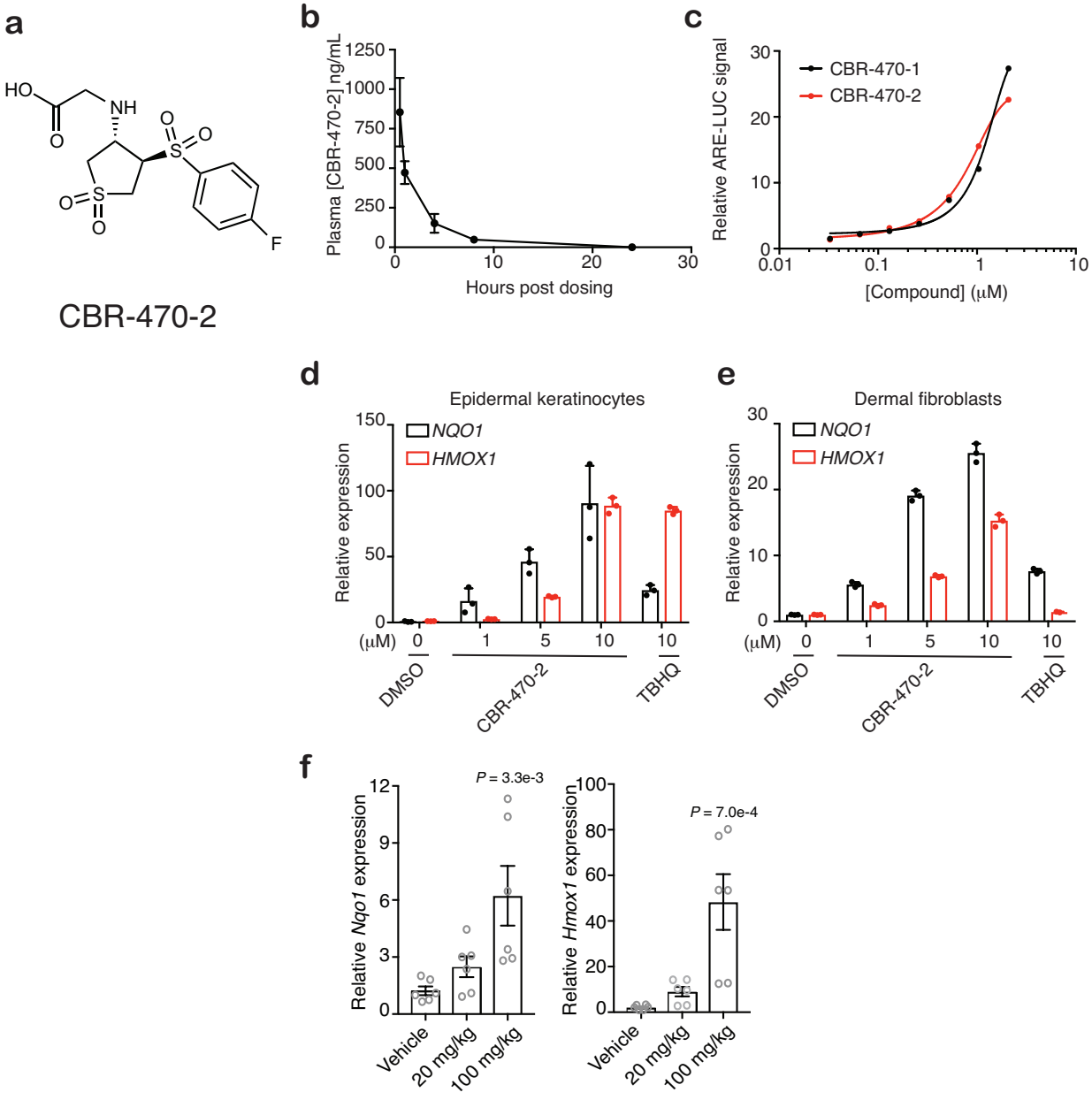


Figure 2.7 CBR-470-2 pharmacokinetics

a. Structure of CBR-470-2. **b.** Plasma concentrations of CBR-470-2 from mice treated with a single 20 mg/kg dose of compound (n=3 animals, mean and s.e.m.). **c.** Relative ARE-LUC luminance values from IMR32 cells transfected with pTI-ARE-LUC and treated with the indicated doses of CBR-470-1 and CBR-470-2 for 24 hours (n=3 biologically independent samples). **d. e.** Relative transcript levels of *NQO1* and *HMOX1* from mouse epidermal keratinocytes (**d**) and mouse dermal fibroblasts (**e**) treated for 24 hours with the indicated doses of compound (n=3 biologically independent samples, mean and s.d.). **f.** Relative *NQO1* and *HMOX1* transcript levels in mouse skin tissue 24 hours after the indicated oral doses of CBR-470-2 (n=6, biologically independent samples). Statistical analyses are one-way ANOVA with Dunnett's correction. Data are mean and s.e.m.

Among the series of CBR-470 derivatives, a glycine-substituted analogue CBR-470-2 showed more favorable bio-ability profile (20 mg/kg PO; $t_{1/2}$ = 2.7 hr; C_{max} = 956.9 ng/mL) than CBR-470-1 in mice allowing for the compound to be more tolerated with the sustained micromolar serum concentrations, but keeping almost equal potency in cells (~ 1 μM) (Figure 2.7a, 2.7b, and 2.7c). NRF2 activations by CBR-470-2 in mice were analyzed with the reading of NRF2 target gene expressions, and cells from several organs were harvested after systemic oral dosing of CBR-470-2. As a result, we found that relative transcript levels of *NQO1* and *HMOX1* were increased in the skin cells (epidermal keratinocytes and dermal fibroblasts) with dose-dependent manners (Figure 2.7d, 2.7e and 2.7f).

Several published studies with NRF2-knockout mice demonstrate that NRF2 is essential to protect the skin cells against photo-aging phenotypes (e.g. wrinkle formation and epidermal thickening) and skin carcinogenesis[150] resulting from UV irradiation[151]. As UV skin damage mouse model is well established to investigate the cytoprotective effects of NRF2 regulation *in vivo*, we evaluated the physiologic relevance of CBR-470-2 induced NRF2 activation in this acute animal model. In the study, Badoxolone methyl (BARD), a clinical Phase III NRF2 agonist by directly targeting KEAP1, was chosen as a positive control. Mice were prophylactically dosed with CBR-470-2 or BARD for five days before exposure to a single dose of UV irradiation (200 mJ/cm^2). After light-induced damages followed by five days of the compounds dosed in mice, UV damage of the sacrificed animal skin cells was measured and quantified. Compared to the vehicle treatment, both CBR-470-2 and BARD showed effects on decreasing total would areas and erythema (redness) histological scores of the UV damaged skin in mice pretreated with these compounds (Figure 2.8a, 2.8b, and 2.8c).

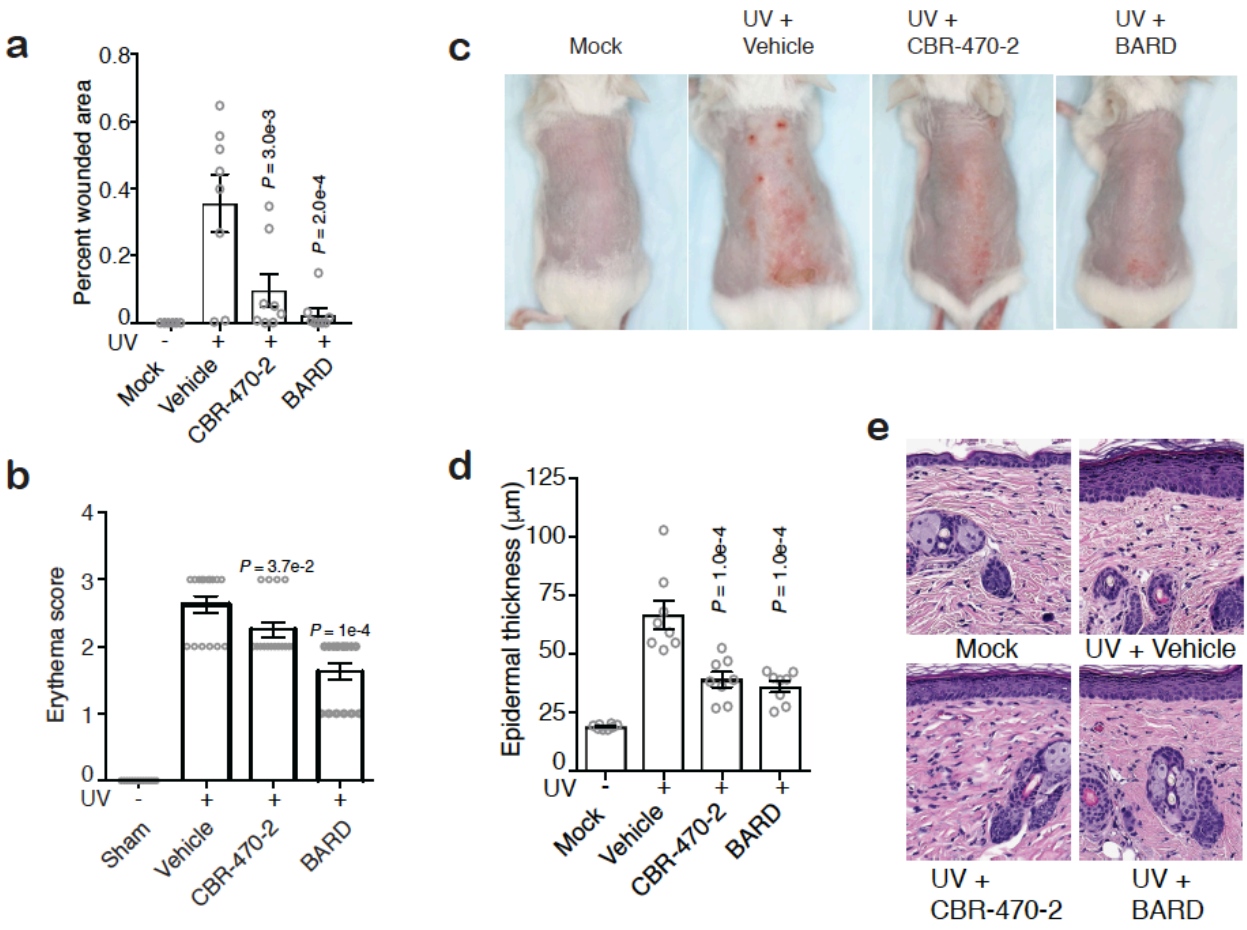


Figure 2.8 CBR-470-2 *in vivo* activity

a. Quantification of wounded area by automated image analysis from animals of the indicated treatment groups at study end (day 10). **b.** Blinded erythema scores from mice treated with vehicle, CBR-470-2 or Bardoxolone after UV exposure (n=8 animals, $P^{*} < 0.05$, $P^{**} < 0.005$, one-way ANOVA with Dunnett's correction, mean and s.e.m.). **c.** Representative images of UV-exposed dorsal regions of animals at day 10 of the study. **d.** Quantification of epidermal thickness from H&E stained skin sections from the indicated groups at study end. **e.** Representative images of H&E stained sections from animals sacrificed at day 10 of the study. CBR-470-2, 50 mg/kg BID PO; BARD, bardoxolone methyl, 3 mg/kg BID PO; UV, 200 mJ/cm^2; data are mean and s.e.m., n=8 animals. Statistical analyses are one-way ANOVA with Dunnett's correction. Data are mean and s.e.m.

After UV-induced skin damages, epidermal layer of mouse was thickened approximately 3-fold resulting from the proliferation of epidermal keratinocytes in an attempt to close cutaneous wounds[152], quantified by hematoxylin eosin (H&E) stain. Both CBR-470-2 and BARD were observed to reduce epidermal thickness of mice in response to UV exposure, consistent with activation of the NRF2 cytoprotective program (Figure 2.8d and 2.8e).

In summary, both cell-based and *in vivo* animal experiments were performed to identify a new class of NRF2 activators and evaluate its cytoprotective effects in line with NRF2 activation by measuring pharmacodynamic effects (*HMOX1* and *NQO1* induction) and NRF2-dependent phenotype modulations. All combined data suggest that CBR-470 series treatment is capable of activating NRF2 pathway in cells and *in vivo*, side-by-side checking with known NRF pathway activators.

2.2.2. CBR-470-1 targets PGK1 and reprograms the glycolytic flux

Non-electrophilic feature of CBR-470 series was queried the existent mechanism of NRF2 activation that is independent of direct KEAP1 binding. Thus, we performed a set of experiments to identify the actual binding target of CBR-470-1 and the effects of CBR-470-1 on cellular system to elucidate the NRF2 activation route.

2.2.2.1 Glycolytic enzyme PGK1 is a target of CBR-470-1

Our previous targeted SAR study of CBR-470 revealed a subtle effect of the substitution of CBR-470-0 cyclic sulfone substrate on the activation of NRF2 pathway. To apply a photoactivable affinity probe-based approach to identify the binding target of CBR-470-series, we designed CBR-470-PAP, a linker substituted analogue probe of CBR-470-1 possessing biotin and

diazirine substrates (Figure 2.9a). The probe was displayed to retain a potency (EC_{50} = 2.4 μM) in cell-based ARE-LUC assay (Figure 2.9b).

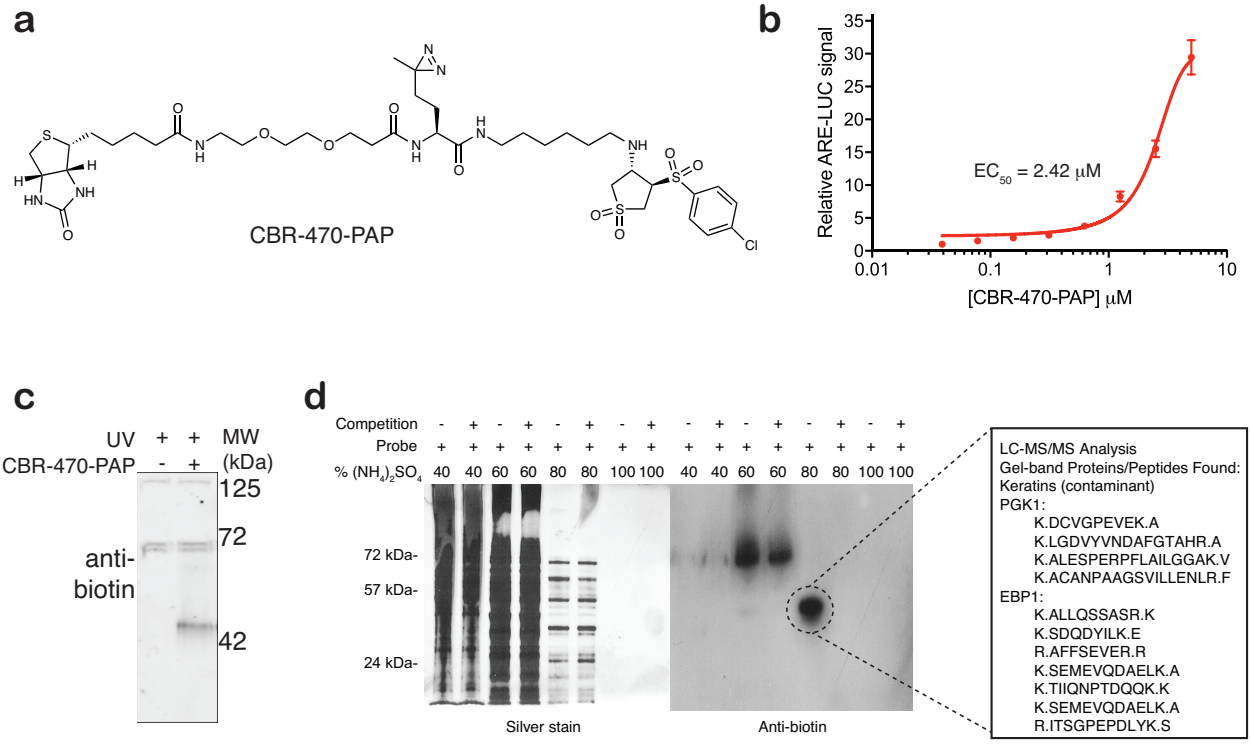


Figure 2.9 A photoactivable affinity probe-based approach identifies PGK1 as the relevant cellular target of CBR-470-PAP

a. Structure of CBR-470-PAP. **b.** Relative ARE-LUC luminance values from IMR32 cells transfected with pTI-ARE-LUC and then treated with the indicated doses of CBR-470-PAP for 24 hours (n=3). **c.** Anti-biotin western blot analysis of IMR32 cells treated with CBR-470-PAP (10 μM) for 1 hour and exposed to UV light to induce photocrosslinking (representative shown from n=4 biological replicates). **d.** Silver staining and anti-biotin Western blots of ammonium sulfate fractionated lysates from UV-irradiated IMR32 cells treated with 5 μM for 1 hour with or without CBR-470-1 competition (250 μM) (n=3). Shown on the right are initial proteomic target results from gel-band digestion and LC-MS/MS analysis.

To generate light-activated irreversible crosslinks between the probe and its reversible binding partner in live cells, 5 μM CBR-470-PAP was treated to IMR32 cells for 1 hour followed by UV irradiation to the cells in 10 minutes to activate the diazirine substrate of the probe. An apparent biotin band was observed near 45 kDa in α-biotin Western blots of whole cell and ammonium sulfate fractionated lysates, and LC-MS/MS analyses of the isolated gel fraction of this band (45 kDa) revealed that phosphoglycerate kinase 1 (PGK1) is a candidate of CBR-470-PAP target (Figure 2.9c).

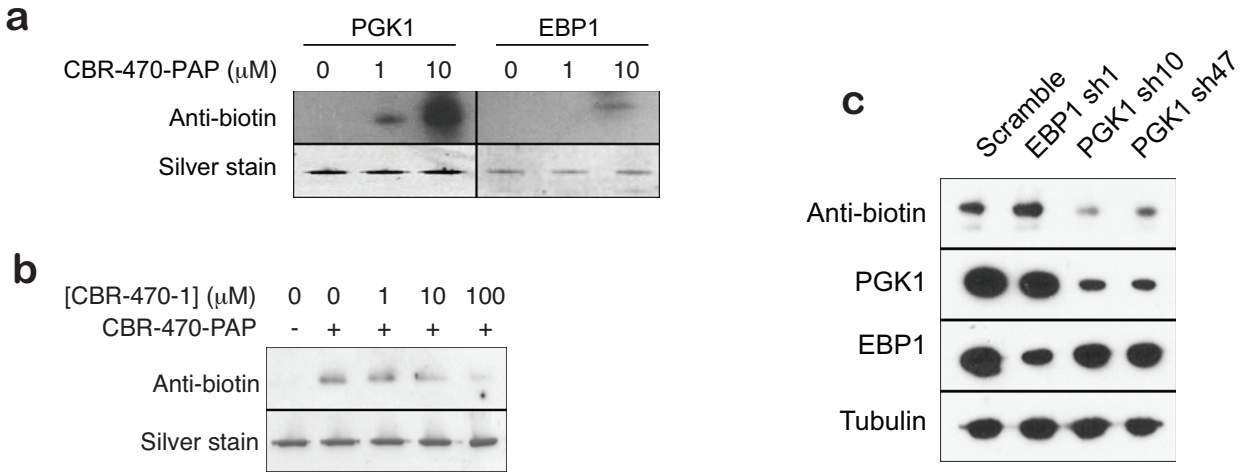


Figure 2.10 *In vitro* and *in situ* validation of the target of CBR-470-PAP and CBR-470-1 as PGK1
a. Anti-biotin Western blots from *in vitro* crosslinking assays with recombinant PGK1 and EBP1 in the presence of the indicated doses of CBR-470-PAP (n=2). **b.** Anti-biotin Western blot analyses from an *in vitro* crosslinking assays with recombinant PGK1 in the presence of CBR-470-PAP (1 μM) and indicated concentration of soluble CBR-470-1 competitor (n=2). **c.** Anti-biotin Western blot analyses of cells treated with 5 μM CBR-470-PAP after transduction with anti-PGK1 and anti-EBP1 shRNA for 48 hours. Depletion of PGK1 protein selectively reduces CBR-470-PAP-dependent labeling (n=2).

Isolation of the biotin band also identified a proliferation-associated protein 2G4 (EBP1) with PGK1, thus, *in vitro* binding experiments of CBR-470-PAP with recombinant PGK1 and EBP1 were firstly performed to verify the actual target of the compound (Figure 2.9d). Compared to EBP1, intensity of the biotin bands of recombinant PGK1 was increased with higher doses of CBR-470-PAP and decreased by the co-treatment of molar excess CBR-470-1, a binding competitor, in a dose-dependent manner, up to 100 μM *in vitro* (Figure 2.10a and 2.10b). Corresponding to the results of *in vitro* tests, shRNA-based knockdown of PGK1 in IMR32 cells dropped the intensity of both PGK1 bands and biotin bands, whereas CBR-470-PAP induced biotin labeling was not reduced by EBP1 knockdown (Figure 2.10c). Collectively, these results indicate that CBR-470-PAP selectively interacts with glycolytic enzyme PGK1 *in vitro* and in live cells.

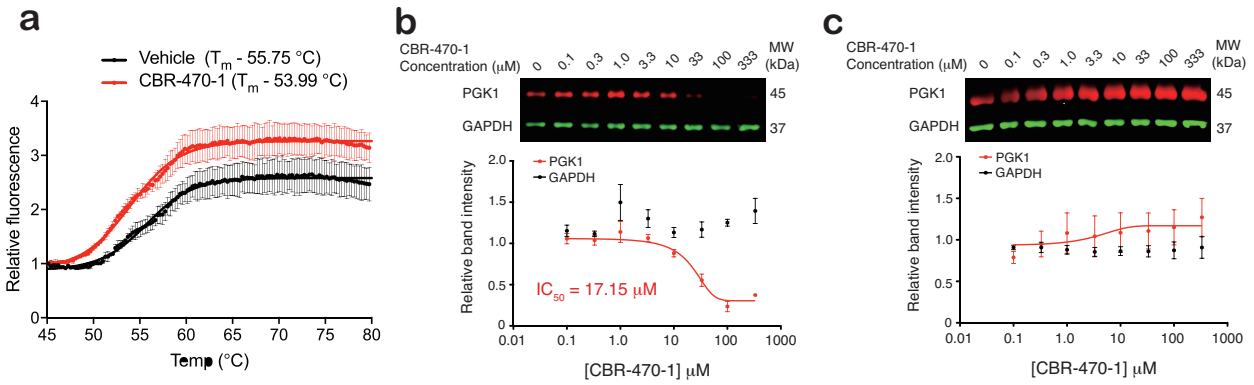


Figure 2.11 Label-free approaches identify PGK1 as the relevant target of CBR-470-1
a. Dye-based thermal denaturation assay with recombinant PGK1 in the presence of CBR-470-1 (20 μM) or vehicle alone (n=3). Calculated T_m values are listed **b. c.** Dose-dependent thermal stability assay of recombinant PGK1 and GAPDH in the presence of increasing doses of CBR-470-1 near the T_m of both proteins (57 °C) **(b)** (n=5) or room temperature **(c)** (n=3). Western blot of sample supernatants after centrifugation (13,000 rpm) detected total PGK1 and GAPDH protein, which were plotted in Prism (below). Data shown present mean ± SEM of biologically independent samples.

To evaluate the binding affinity of a novel NRF2 activator CBR-470-1 with the identified target PGK1, we further performed label-free *in vitro* assays. *In vitro* thermal stability assay (TSA) is a traditional ligand identification method to measure the intact protein-ligand interactions by quantification of a protein denaturation under high temperatures and a shift of protein stability by ligand bindings. We confirmed that CBR-470-1 resulted in a consistent shift in PGK1 thermal stability (Vehicle T_m = 55.75 °C, CBR-470-1 T_m = 53.99 °C) (Figure 2.11a). Similarly, isothermal dose response (ITDR)[153] profiling was performed against the mixture of recombinant PGK1 and GAPDH with increasing doses of CBR-470-1 and dose-dependent alteration of PGK1 stability in the presence of CBR-470-1 (IC_{50} = 17.15 μM), but not affected to GAPDH stability, was observed by Western blots (Figure 2.11b and 2.11c).

Recently, cellular thermal shift assay (CETSA)[154,155] was introduced as a new label-free target identification tool that allows thermal stability measurements of intact proteins in live cells. Indeed, combination of CETSA with quantitative proteomic analyses enables the screening of protein stability in a global manner. Thus, we also performed CETSA with HeLa cells treated with 10 μM CBR-470-1 for 1 hour and measured the melting curve (T_m) shifts of proteins with SILAC-based quantitative proteomic approach[156] (Figure 2.12a). Among the glycolytic enzymes, we found that PGK1 consistently showed highest T_m shifts across the biological replicates (ΔT_m < -1 °C) which is consistent with the results of *in vitro* TSA experiments with recombinant PGK1 (Figure 2.12b and 2.12c). Together, we confirmed PGK1 as a specific target of CBR-470-1.

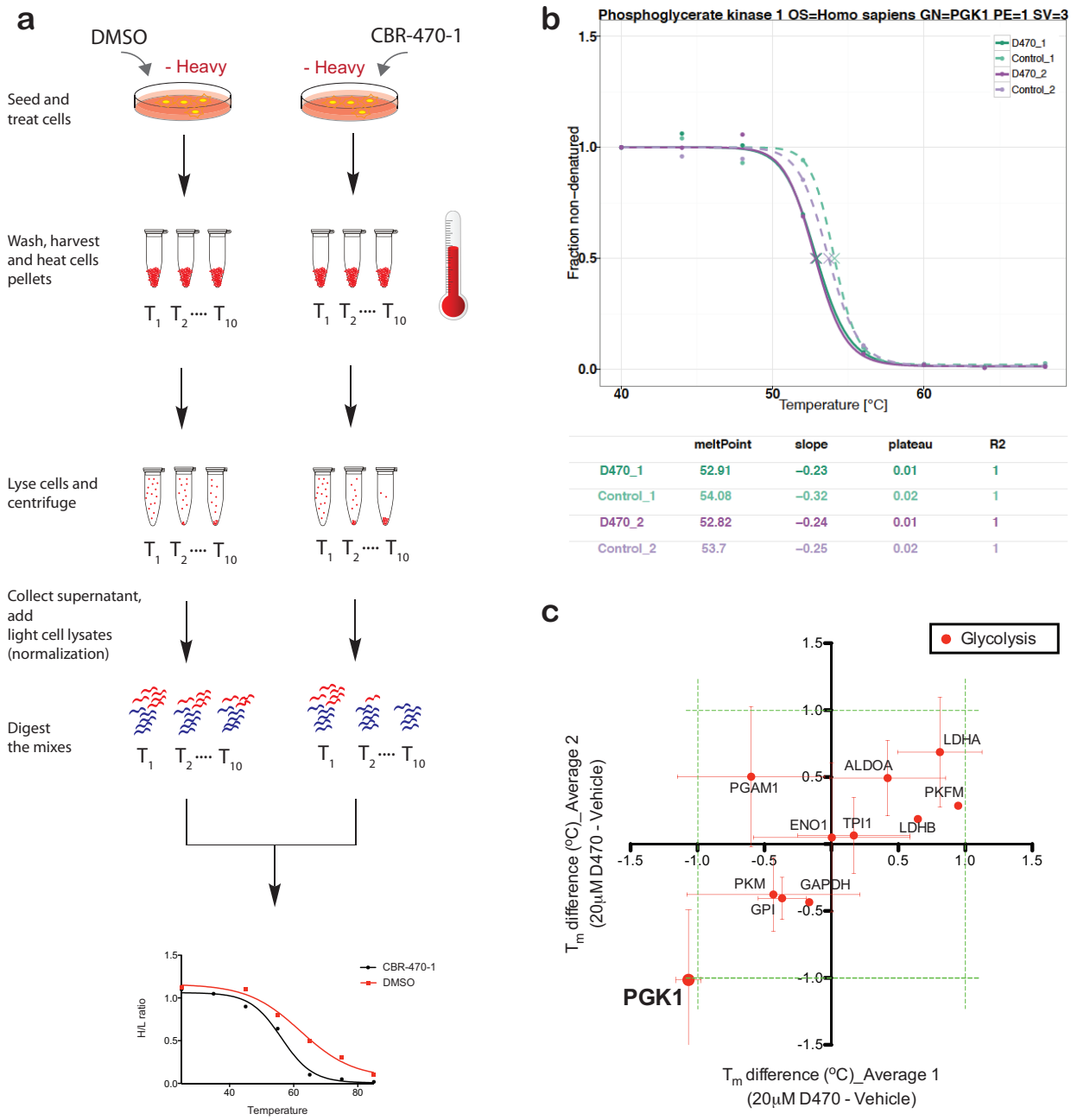


Figure 2.12 *In situ* CETSA-MS assay for proteome-wide target identification of CBR-470-1
a. Representative workflow of *in situ* CETSA-MS with SILAC quantification. Cells are treated with either CBR-470-1 (10 μM) or DMSO to each heavy-isotope labeled HeLa cell plate, washed and harvested. Equal amounts of cells are then heated to increasing temperatures and lysed, and the supernatant obtained after centrifugation. The amount of remained proteins in each fraction is normalized by the addition of the lysates of light-isotope labeled Hela cells (not heated), and each sample is digested and analyzed by SILAC quantitative mass spectrometry (MS) analysis. Detection of changes in protein stability in CBR-470-1 versus DMSO samples is enabled by bioinformatic processing and data plotting in the form of differential protein-abundance graphs. **b.**

(Figure 2.12, continued) Melt curve of PGK1 from cells treated with either CBR-470-1 or DMSO. Calculated T_m values are listed. **c.** Scatter plot of T_m shifts calculated from the four biological replicates of the CBR-470-1 versus DMSO treatment experiment. Each axis value (x-axis and y-axis values of each protein) reflected the mean ± s.e.m. of T_m shift values from two biological replicates each.

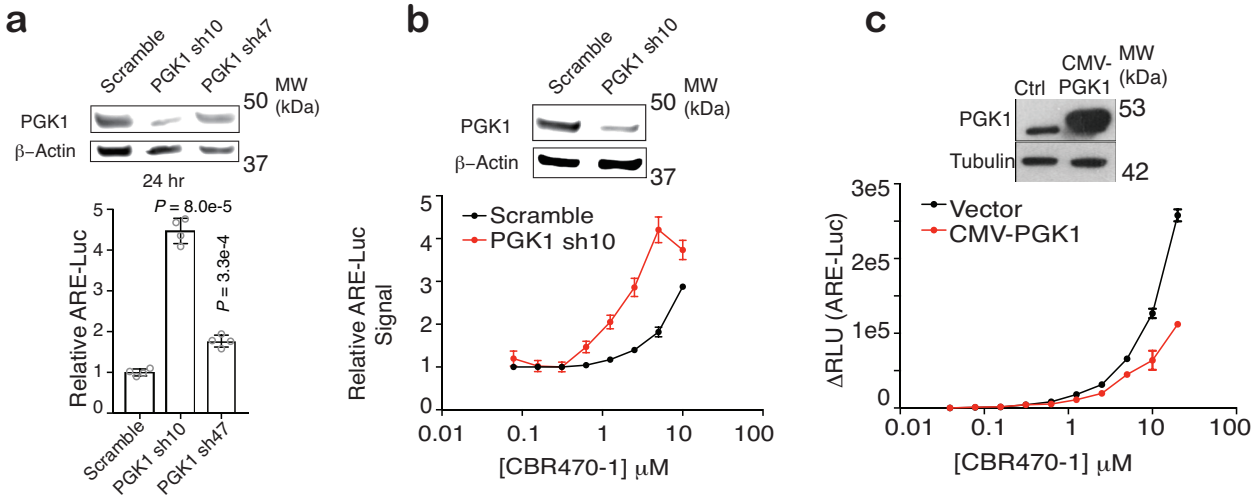


Figure 2.13 Modulation of cellular PGK1 levels by transient knockdown and expression affects activation of NRF2 pathway

a. Transient transfection of shRNA constructs targeting PGK1 in HEK293T cells activates the ARE-LUC reporter. PGK1 and β-actin protein levels shown from representative experiments (n=4 biological replicates). **b. c.** CBR-470-1 activation of ARE-LUC reporter in HEK293T cells with transient knockdown **(b)** or overexpression **(c)** of PGK1 demonstrates opposing effects on compound potency. PGK1, Actin, and Tubulin protein levels are shown from representative experiments (n=3).

2.2.2.2 Inhibition of PGK1 induces alteration of glycolytic flux and activation of NRF2 pathway.

Indirect activation of NRF2 signaling cascade by CBR-470-1, actually binding to PGK1, suggests the physiological relevance and the mechanism connecting PGK1 with KEAP1-NRF2

pathway. Instead of treating CBR-470-1 to cells, transient shRNA-based knockdown of PGK1 also increased ARE-LUC signals in the reporter assay (Figure 2.13a). Indeed, modulation of PGK1 levels in cells affected the sensitivity of CBR-470-1 toward activation of NRF2 pathway, with knockdown or overexpression of PGK1 resulting in decreased or increased EC_{50} values of CBR-470-1 (Figure 2.13b and 2.13c).

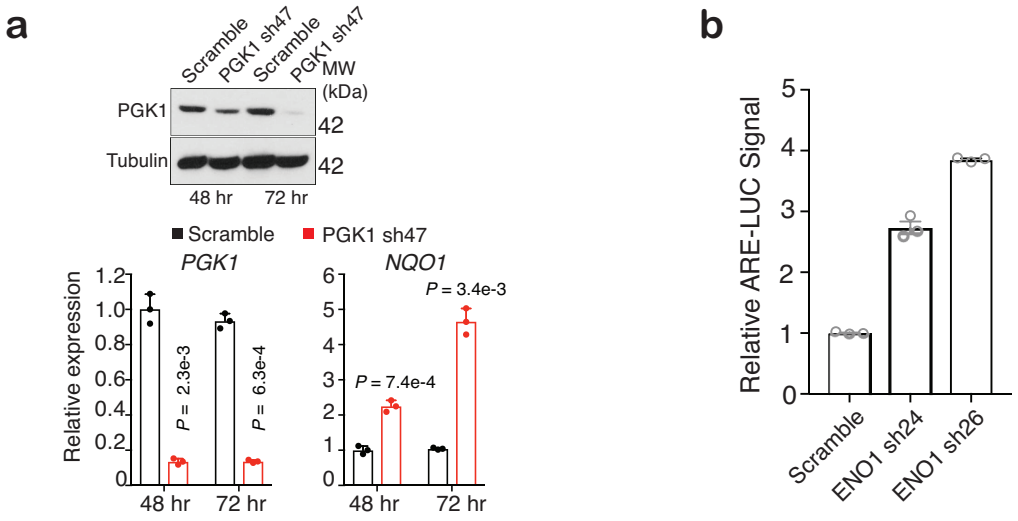


Figure 2.14 Viral knockdown of cellular PGK1 and ENO1 levels affects activation of NRF2 pathway
a. Viral shRNA knockdown of PGK1 induces NQO1 mRNA levels in IMR32 cells. PGK1 and Tubulin protein levels are shown from representative experiments (n=3). **b.** ARE-LUC reporter activity in HEK293T cells with transient shRNA knockdown of ENO1 (n=3). Data shown represent mean ± SEM of biologically independent samples.

Consistent with the data from the reporter assay, depletion of cellular PGK1 levels via viral shRNA knockdown in IMR32 cells resulted in significant increases of the expression of endogenous NRF2 target gene *NQO1*, thereby validating a connection between active population

of PGK1 and NRF2 activation in the absence of CBR-470-1 (Figure 2.14a). We also found the luciferase signal increased with the viral depletion of enolase (ENO1), a glycolytic enzyme downstream of PGK1 in glycolysis (Figure 2.14b). Overall, these results suggested that CBR-470-1 modulation of PGK1 activity, and therefore glycolysis, may result in NRF2 activation.

To reveal a role of CBR-470-1 in regulation of PGK1 function, thus effecting on cellular system and especially glycolytic flux, we first optimized and performed the *in vitro* enzyme assay[157,158] to see if CBR-470-1 selectively affects the activity of PGK1 (Figure 2.15a). In the assay, PGK1 shifts the equilibrium of the substrates from D-glyceraldehyde-3-phosphate (GAP) to 3-phosphoglycerate (3PG) kinetically monitored by the absorbance of NADH, a co-product of GAPDH reaction. Unlike to GAPDH, pretreatment of CBR-470-1 to recombinant PGK1 decreased the accumulation of NADH with increasing doses (IC_{50} = 2.4 μM), indicating that CBR-470-1 inhibits PGK1 activity and potentially affects the metabolism (Figure 2.15b). Thus, we next sought to evaluate the effect of CBR-470-1 on glucose metabolism in cells by targeted metabolomic profiling[98,159] (Figure 2.16a). By comparing the levels of metabolites in CBR-470-1 treated- and DMSO treated IMR32 cells, a profile was marked by accumulation of glycolytic metabolites upstream of PGK1 such as GAP and 1,3- and 2,3-bisphosphoglycerate (BPG), and depletion of downstream metabolites such as 3PG and lactate (Lac) (Figure 2.10d). This profile was highly consistent with the profile of glycolytic metabolites observed upon knockdown of PGK1 in IMR32 cells (Figure 2.16b and 2.16c). Indeed, kinetic analysis of metabolic levels after CBR-470-1 treatments revealed a rapid elevation of upstream metabolites within an hour and suggested that CBR-470-1 mediated alteration of glycolytic flux by PGK1 inhibition results in NRF2 activation (Figure 2.17).

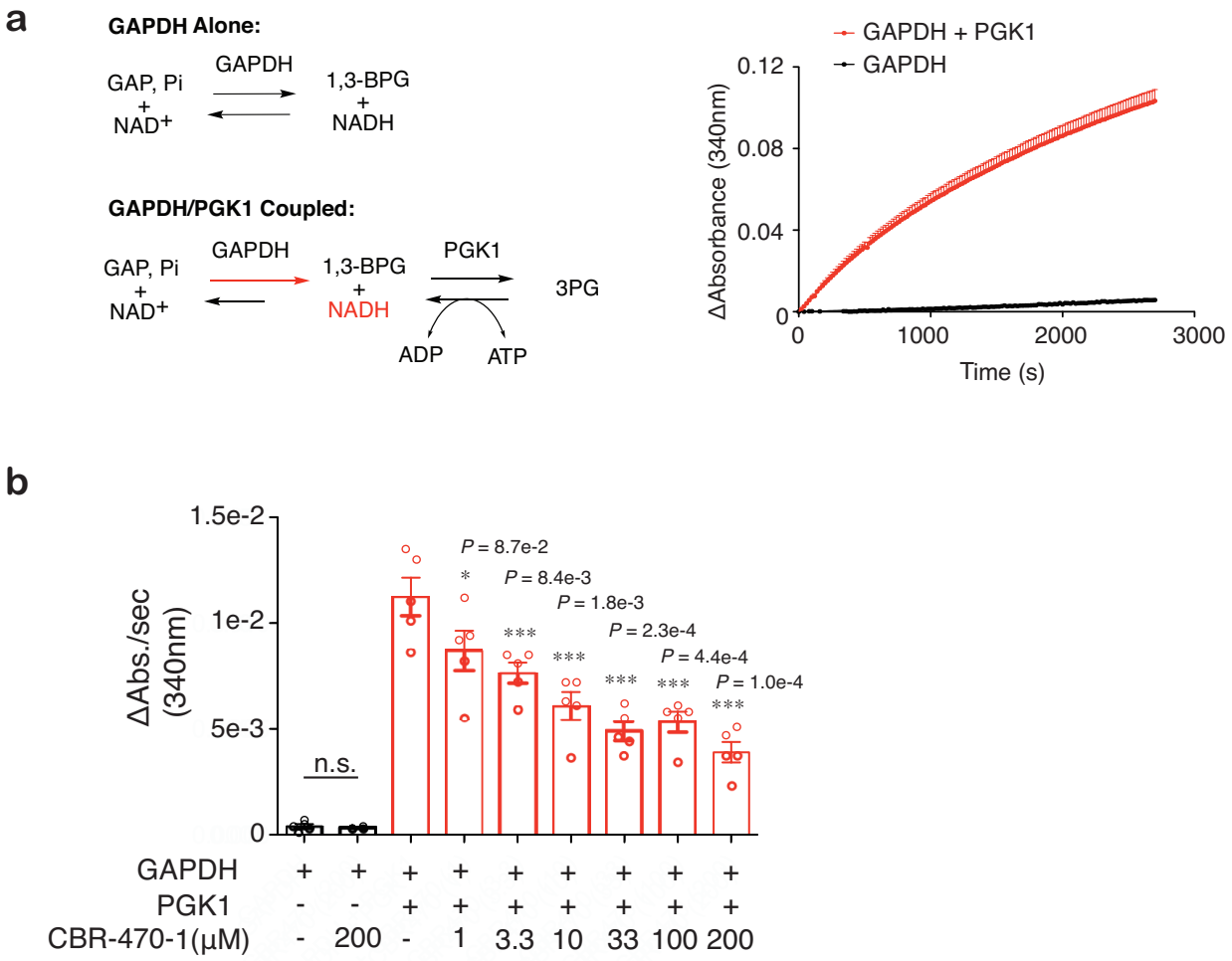


Figure 2.15 CBR-470-1 inhibits PGK1 *in vitro*

a. Schematic of the GAPDH/PGK1 coupled assay. Pre-equilibration of the GAPDH reaction resulted in an NAD^+/NADH equilibrium, which upon addition of PGK1 and ADP pulls the reaction to the right producing more NADH. Monitoring NADH absorbance after addition of PGK1 can be used to monitor PGK1 activity in the forward direction. Kinetic monitoring of NADH absorbance (340 nm) after established equilibrium with GAPDH shows little changes (black curve), but is significantly increased upon addition of PGK1, pulling the equilibrium to the right (red**)** curve). **b.** CBR-470-1 does not affect the GAPDH equilibrium alone, but significantly inhibits PGK1-dependent activity and accumulation of NADH (n=5). Each metabolite is normalized to the control condition at each time point. Univariate two-sided t-test (Fig 2.15 b); data shown represent mean ± SEM of biologically independent samples.

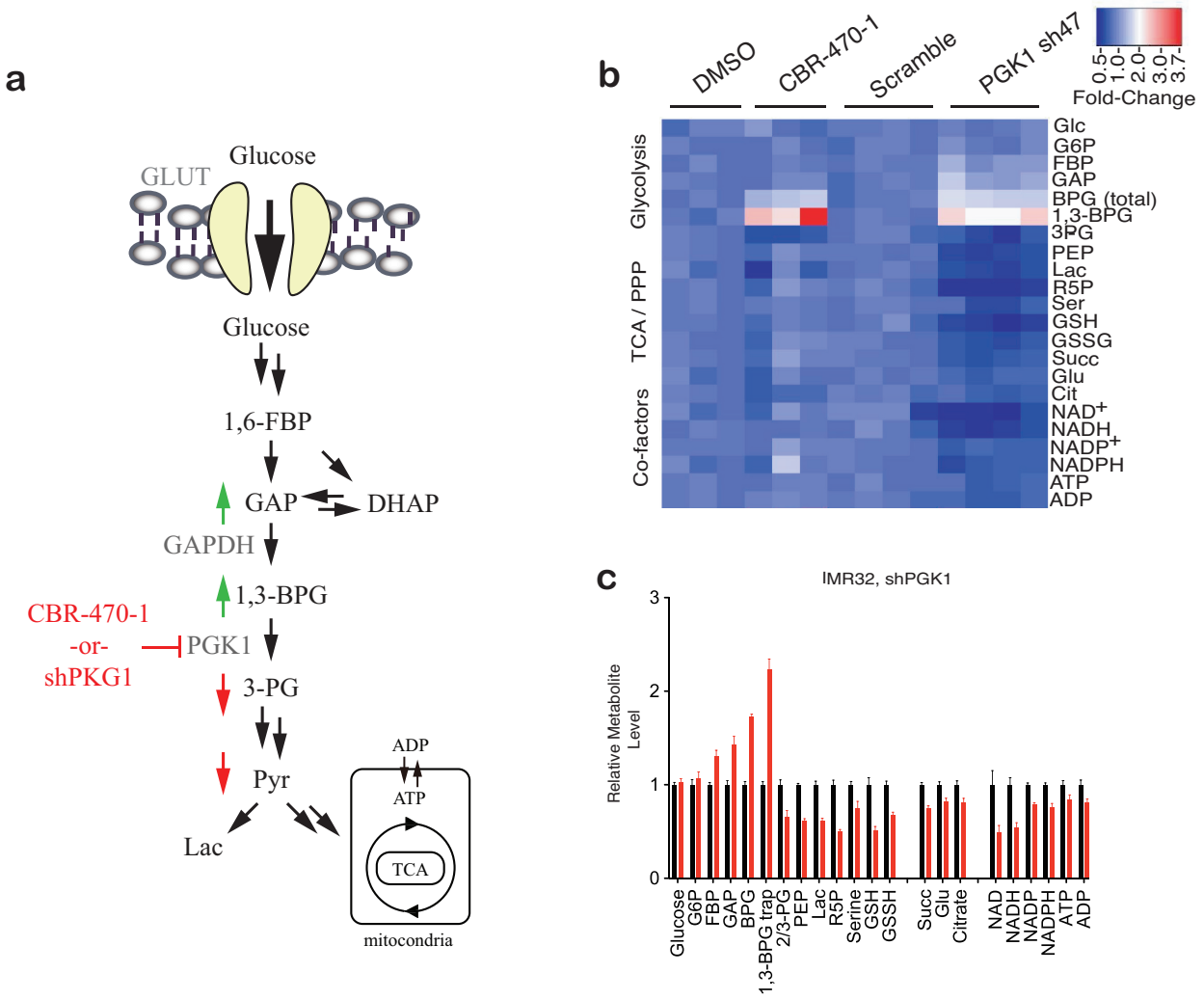


Figure 2.16 CBR-470-1 and PGK1 in cells

a. Schematic of the effect of CBR-470-1 on glycolytic flux. **b.** Heat map depiction of relative metabolite levels in IMR32 cells treated for 30 minutes with CBR-470-1 (left) or viral shRNA knockdown of PGK1 (right) relative to DMSO and scramble shRNA controls, respectively. BPG refers to both 2,3-BPG and 1,3-BPG, whereas 1,3-BPG specifically refers to the 1,3-isomer. **c.** Relative levels of central metabolites in IMR32 cells treated with viral knockdown of PGK1 for 72 hours (n=4). Each metabolite is normalized to the control condition at each time point. Data shown represent mean ± SEM of biologically independent samples.

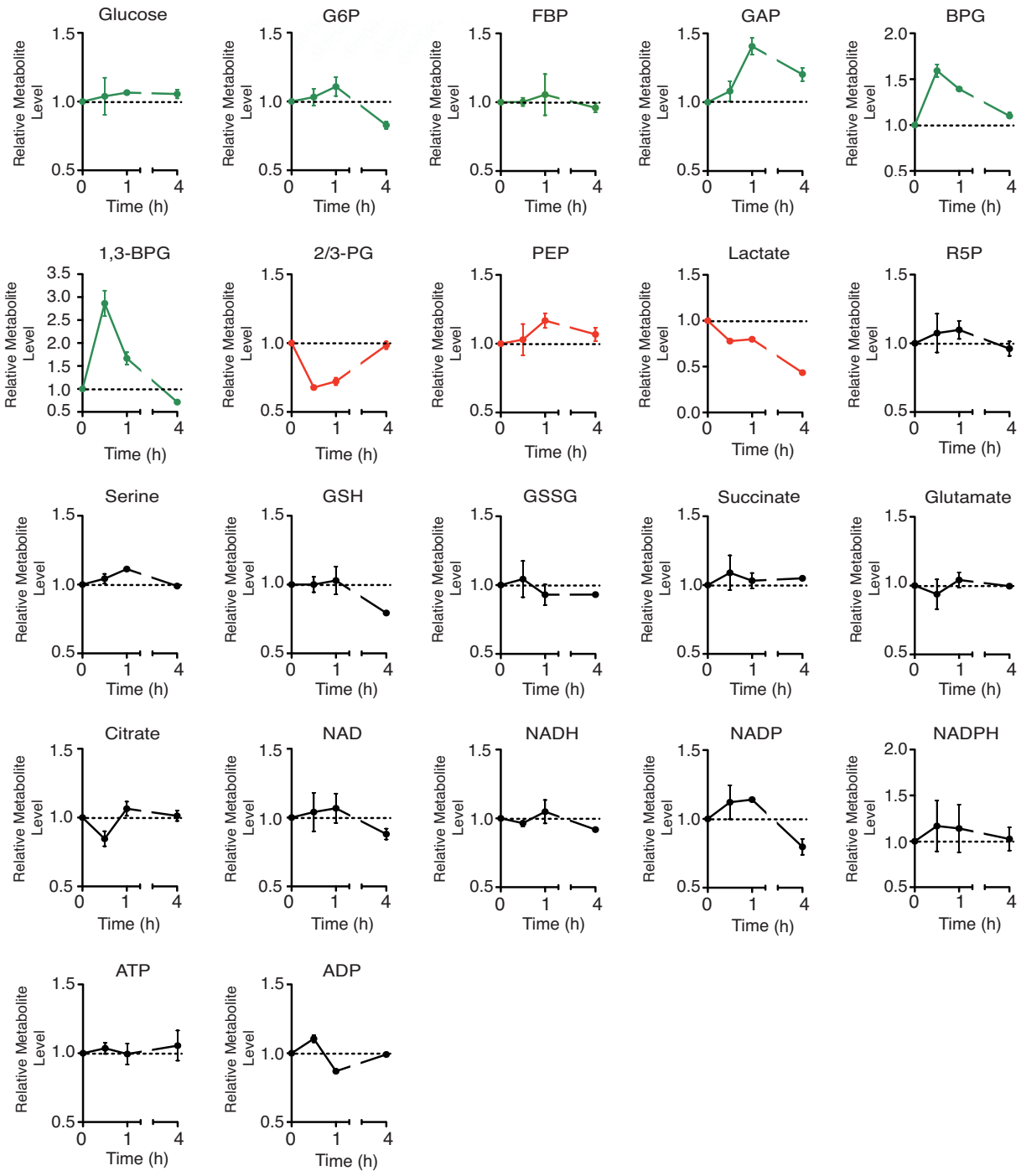


Figure 2.17 CBR-470-1 induces alteration of glycolytic flux
Relative levels of central metabolites in IMR32 cells treated with CBR-470-1 relative DMSO alone for the indicative times (n=3). Each metabolite is normalized to the control condition at each time point. Data shown represent mean ± SEM of biologically independent samples.

Investigation of a mechanism explaining a sequence of CBR-470-1-induced NRF2 activation requires a piece of information; if NRF2 activation is caused by a specific metabolic signal or by wholesale deregulation of cellular metabolism? To address the question, we utilized cell-based ARE-LUC assay and validated the activation of NRF2 pathway with CBR-470-1 and 2-deoxyglycose (2DG), an inhibitor of glucose entry into glycolysis. Co-treatment of CBR-470-1 and 2DG resulted in diminishing luciferase signals with increasing doses of 2DG (Figure 2.18), suggesting that central glycolytic metabolites that are accumulated in altered glucose metabolism may serve as a signaling messenger to the NRF2 signaling axis.

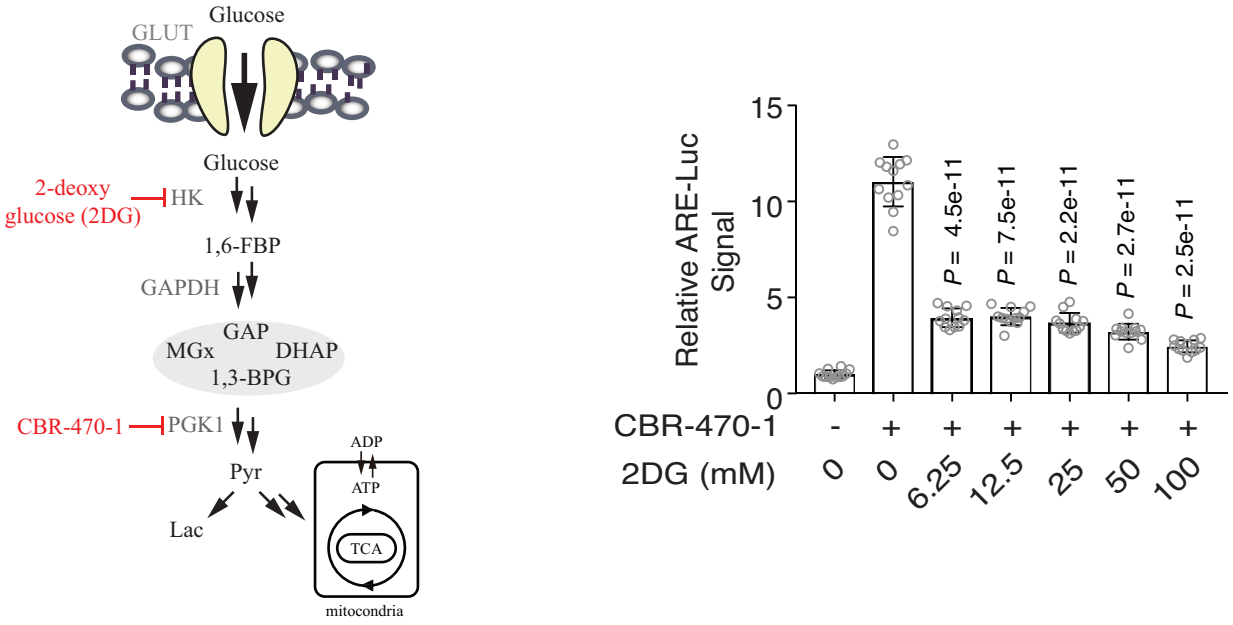


Figure 2.18 Co-inhibition of PGK1 and upstream glycolytic step reduces NRF2 activation
ARE-LUC reporter activity in IMR32 cells co-treated with CBR-470-1 (5 μM) and 2DG for 24 hours. Statistical analyses are univariate two-sided t-tests. Data are mean and s.d. of biologically independent samples.

2.2.3. Modulation of glycolysis accumulates cellular methylglyoxal directly inducing a covalent KEAP1 dimerization

Correlation between the elevated levels of central glycolytic metabolites in altered glucose metabolism and the activation of NRF2 pathway implies the involvement of specific metabolite in modulating KEAP1 behaviors[71,91,98,159]. Discovery of the trigger of NRF2 signaling cascade and the novel KEAP1 regulatory mechanism was aimed in the study.

2.2.3.1 KEAP1 forms covalent dimer under the inhibition of PGK1

Temporal accumulation of 1,3-BPG was previously observed with the inhibition of PGK1, approximately 3-fold elevated in IMR32 cells after 30 minutes of CBR-470-1 treatment. We initially assumed if intrinsically reactive metabolite 1,3-BPG could be involved in signaling to NRF2 pathway via non-enzymatic modifications of KEAP1, the sentinel electrophile sensor protein regulating the NRF2 stabilization (Figure 2.19a). Detection of phosphoglyceryl-lysine (pgK) modifications, as well as potential phosphoglyceryl-cysteine (pgC) modifications of reactive cysteine residues in KEAP1 is essential for proving our hypothesis, thus we generated polyclonal α-pgK antibody from rabbit and utilized it to read pgK or pgC epitopes by Western blot analysis. CBR-470-1 treatment of IMR32 cells for 30 minutes, a time at which 1,3-BPG levels are elevated, did result in α-pgK immunoreactive bands of proteins including GAPDH, previously characterized pgK target (Figure 2.19b). However, we couldn't find any α-pgK bands of affinity-enriched FLAG-KEAP1 expressed in cells and exposed to altered glycolytic flux by CBR-470-1, as determined by Western blot analysis of both reducing SDS-PAGEs and non-reducing SDS-PAGEs (Figure 2.19c). These data may suggest another regulatory mechanism of KEAP1 by CBR-470-1.

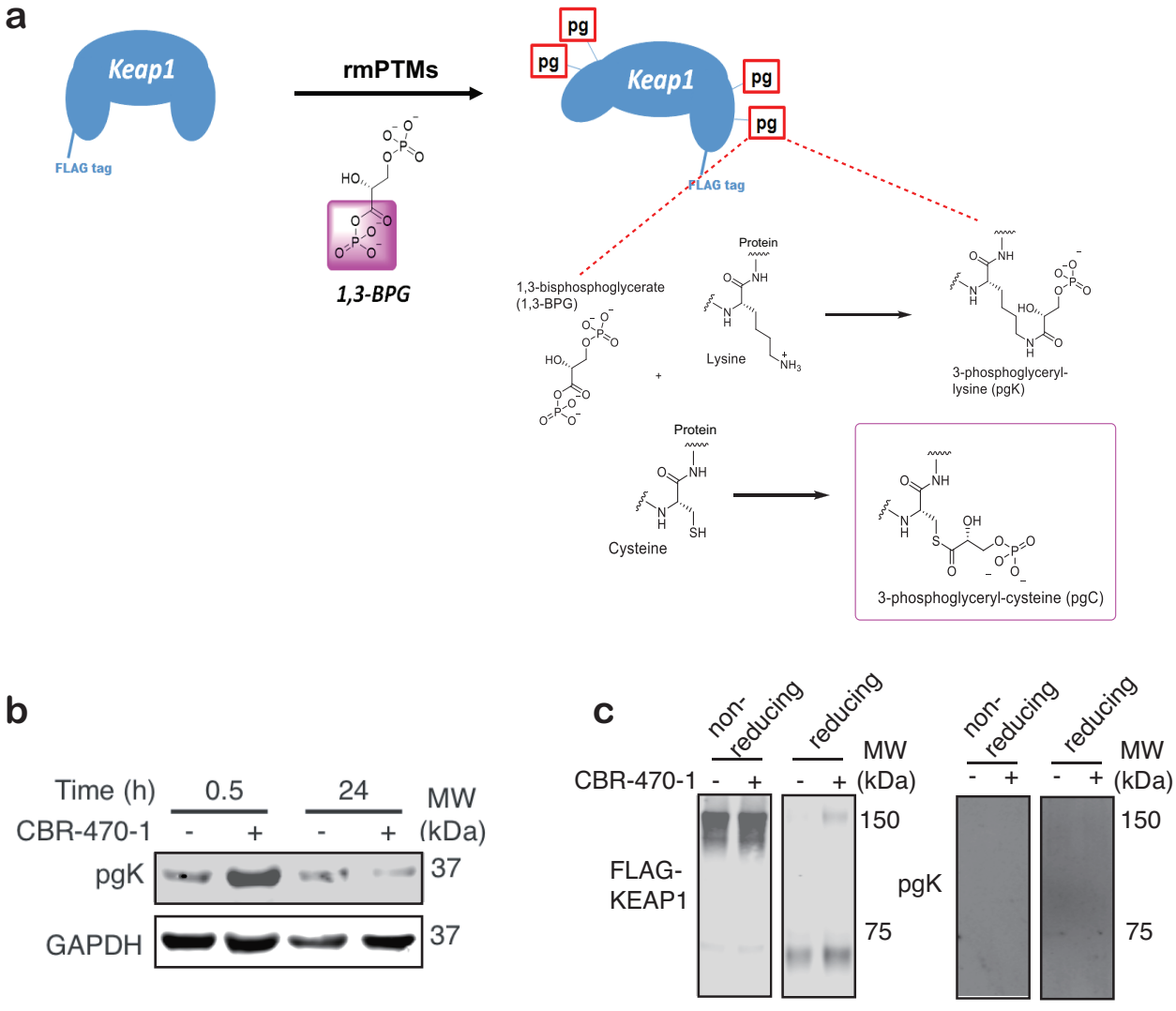


Figure 2.19 1,3-BPG induced KEAP1 modifications were not observed with CBR-470-1
a. The first model. 1,3-BPG possibly induces covalent modification in active lysine residues to form pgK (phosphoglyceryl-lysine) modifications or in active cysteine residues to form pgC (phosphoglycryl-cysteine) modifications. **b.** Anti-pgK and anti-GAPDH Western blots analyses of CBR-470-1 or DMSO-treated IMR32 cells at early (30 minutes) and late (24 hours) time pointes (n=6). **c.** Anti-FLAG (left) and anti-pgK (right) Western blot analyses of affinity purified FLAG-KEAP1 from HEK293T cells treated with DMSO or CBR-470-1 for 30 minutes. Duplicate samples were run under non-reducing (left) and reducing (DTT, right) conditions (n=6).

To test if CBR-470-1 treatment alters KEAP1 protein levels, we employed a kinetic Western blot analysis of KEAP1 in cells treated with 20 μM CBR-470-1, up to 24 hours. The effect of CBR-470-1 on altered KEAP1 levels in cells were not observed, determined by densitometric analysis of KEAP1 followed by the normalization with β-actin band (Figure 2.20a). Unexpectedly, we found the appearance of a strong, CBR-470-1 dependent height molecular weight KEAP1 (HMW-KEAP1) band at roughly twice (~140 kDa) the molecular weight of monomeric KEAP1 (~70 kDa) by these Western blots (Figure 2.20b). To test if HMW-KEAP1 is stable to reduction, thus indicating irreversible covalent crosslink formations, we reduced the proteome in samples with 4x SDS-PAGE loading buffer containing 50 mM 1,4-dithiothreitol (DTT) or 6 % 2-mercaptoethanol (βME) and found that this HMW-KEAP1 band is stable (Figure 2.20c).

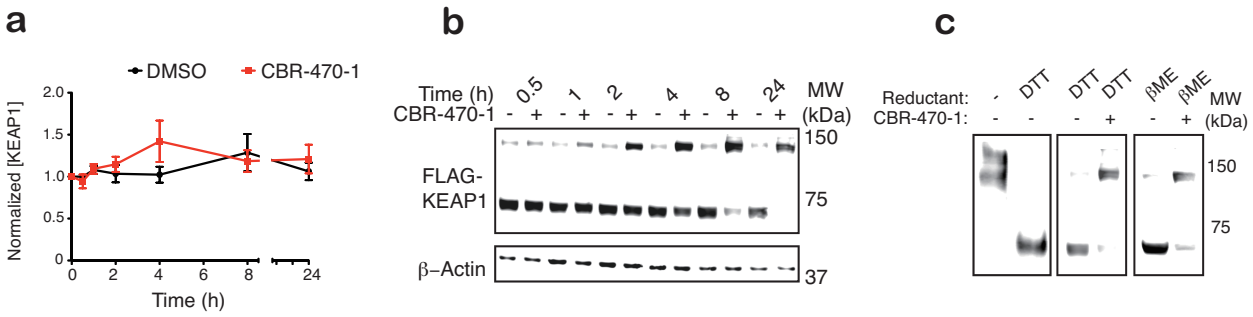


Figure 2.20 Treatment of CBR-470-1 induces HMW-KEAP1
a. Densitometry quantification of total endogenous KEAP1 levels (combined bands at ~70 and 140 kDa) in IMR32 cells treated with DMSO or CBR-470-1 for the indicated times (n=6). **b.** Time-course, anti-FLAG Western blot analyses of whole cell lysates from HEK293T cells expressing FLAG-KEAP1 treated with DMSO or CBR-470-1. **c.** Western blot detection of FLAG-KEAP1 in HEK293T cells comparing no-reducing reagent to DTT (left), and stability of CBR-470-1-dependent HMW-KEAP1 to the presence of DTT (12.5 mL final concentration, middle) and beta-mercaptoethanol (5% v/v final concentration, right) during sample preparation treated with DMSO or CBR-470-1 for 8 hours (n=8).

The level of KEAP1 modification was readily quantified by monitoring the population of HMW-KEAP1 and monomeric KEAP1, and CBR-470-1 mediated HMW-KEAP1 formation was observed in both exogenous FLAG-KEAP1 of HEK293T cells and endogenous KEAP1 of IMR32 cells (Figure 2.21a, 2.21b, and 2.21c).

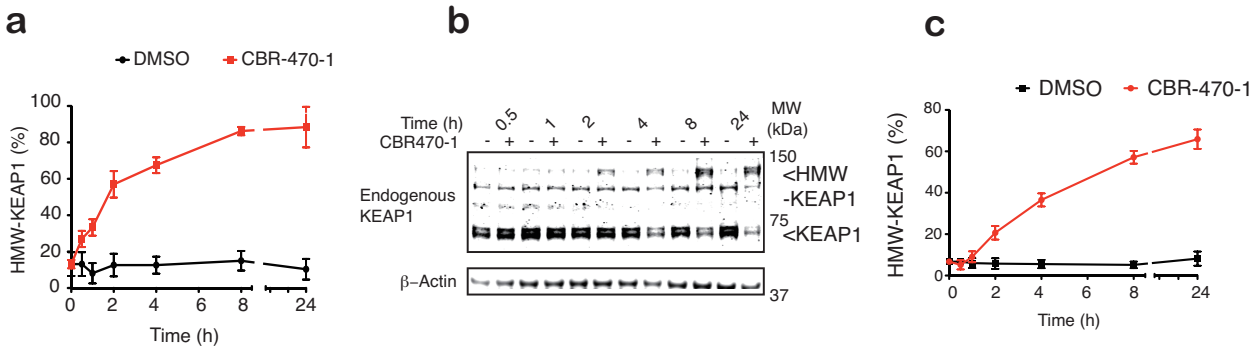


Figure 2.21 HMW-KEAP1 is time-dependently formed by CBR-470-1 treatment *in situ*
a. Time-dependent CB-470-1 treatment of HEK293T cells expressing FLAG-KEAP1. Time-dependent assays were run with 20 μM CBR-470-1 with Western blot analysis at the indicated time-points (n=8). **b. c.** Western blot detection **(b)** and quantification **(c)** of endogenous KEAP1 and β-actin in IMR32 cells treated with DMSO or CBR-470-1 for the indicated times (n-6). Arrows indicate monomeric (~70 kDa) and HMW-KEAP1 in HEK293T cells exposed to increasing doses of CBR-470-1 (n=3). Data shown present mean ± SEM of biologically independent samples.

Increasing doses of CBR-470-1 treated to HEK293T cells in FLAG-KEAP1 expression induced the elevation of HMW-KEAP1 formation (IC_{50} = ~2 μM) (Figure 2.22a and 2.22b), and kinetics of HMW-KEAP1 formation in cells mediated by CBR-470-1 also corresponded to the kinetics of NRF2 stabilization and induction of *NQO1* gene but distinct from the direct KEAP1 alkylator TBHQ (Figure 2.22c).

To evaluate the physiological relevance of HMW-KEAP1 formation in cells, we modulated redox system or glycolysis. Co-treatment of cells with CBR-470-1 and either reduced glutathione

(GSH) or N-acetylcysteine (NAC) resulted in dose-dependent inhibition of the HMW-KEAP1 band formation (Figure 2.22d). Indeed, knockdown of PGK1 by which NRF2 target genes were expressed also induced HMW-KEAP1 formation that was competed by co-treatment of GSH (Figure 2.22e).

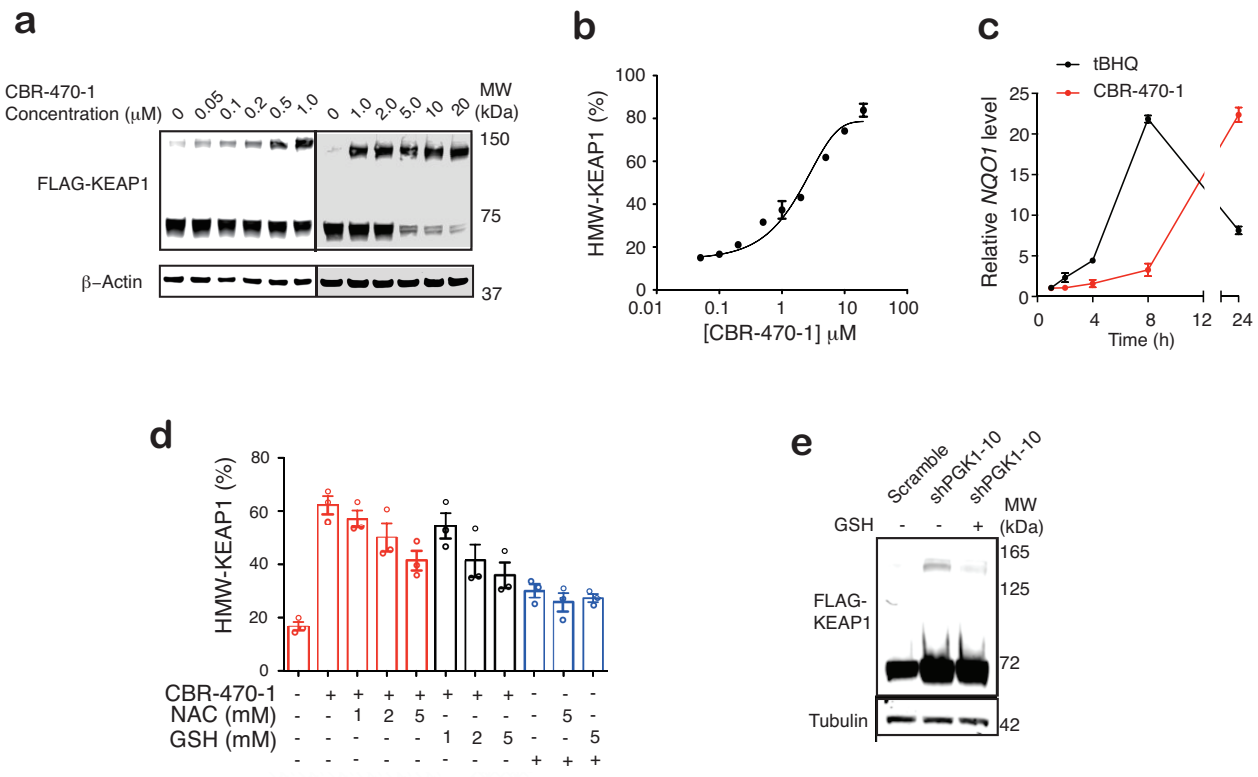


Figure 2.22 Modulation of PGK1 induces HMW-KEAP1
a. b. Western blot **(a)** detection and quantification **(b)** of FLAG-KEAP1 in HEK293T cells exposed to increasing doses of CBR-470-1 (n=3). **c.** Kinetic qRT-PCR measurement of NQO1 mRNA levels from IMR32 cells treated with TBHQ (10 μM) or CBR-470-1 (10 μM) for the indicated times (n=3). **d.** Quantification of HMW-KEAP1 formation upon treatment with CBR-470-1 or the direct KEAP1 alkylator TBHQ, in the presence or absence of reduced glutathione (GSH) or *N*-acetylcysteine (NAC) (n=3). All measurements taken after 8 hours of treatment in FLAG-KEAP1 expressing HEK293T cells. **e.** Transient shRNA knockdown of PGK1 induced HMW-KEAP1 formation, which was blocked by co-treatment of cells by GSH (n=3). Data shown present mean ± SEM of biologically independent samples.

In summary, data suggested that a modulation of glycolysis by PGK1 inhibition results in the formation of HMW-KEAP1, which is consistent with a molecular mass of covalent KEAP1 dimer that has been previously observed[160-164], however the mechanistic basis for the formation of this species was undefined.

2.2.3.2 Methylglyoxal is a signaling molecule inducing non-enzymatic KEAP1 dimerization and NRF2 activation.

Metabolomic, proteomic and NRF2 activation data suggested that the formation and accumulation of central glycolytic metabolites via PGK1 inhibition is necessary for activation of NRF2 signaling. Thus, we sought to find metabolite(s) could be responsible for HMW-KEAP1 formation and activation of NRF2 signaling.

The insight into chemistry leads to identification of the signaling messenger candidates potentially inducing non-enzymatic protein modifications. Several central glycolytic metabolites upstream of PGK1, besides 1,3-BPG, contain reactive functionalities such as triosephosphate isomers D-glyceraldehyde-3-phosphate (GAP, possessing aldehyde moiety) and dihydroxyacetone phosphate (DHAP, possessing ketone moiety), as well as their non-enzymatic elimination product methylglyoxal (MGx), an electrophilic α,β-dicarbonyl compound that has been discovered to form numerous glycated structures on nucleophilic residues in proteome[104,165-167] (Figure 2.23a). To determine if any of these metabolites could be generating HMW-KEAP1 by non-enzymatic modifications, we chose these candidates (GAP, DHAP, MGx, BPG), directly treated them to the aliquots of FLAG-KEAP1 expressed HEK293T cell lysates in DPBS, and screened HMW-KEAP1 by Western blot analysis after 2.5 hours of the treatment. Interestingly, in contrast to other reactive metabolites, MGx resulted in HMW-KEAP1 formation in high specificity, without exhibiting any

random oligomerizations of the abundant cytoskeletal protein β-actin in lysate (Figure 2.23b). The formation of HMW-KEAP1 was also observed in live cells directly treated with MGx, which is cell permeable (Figure 2.24a).

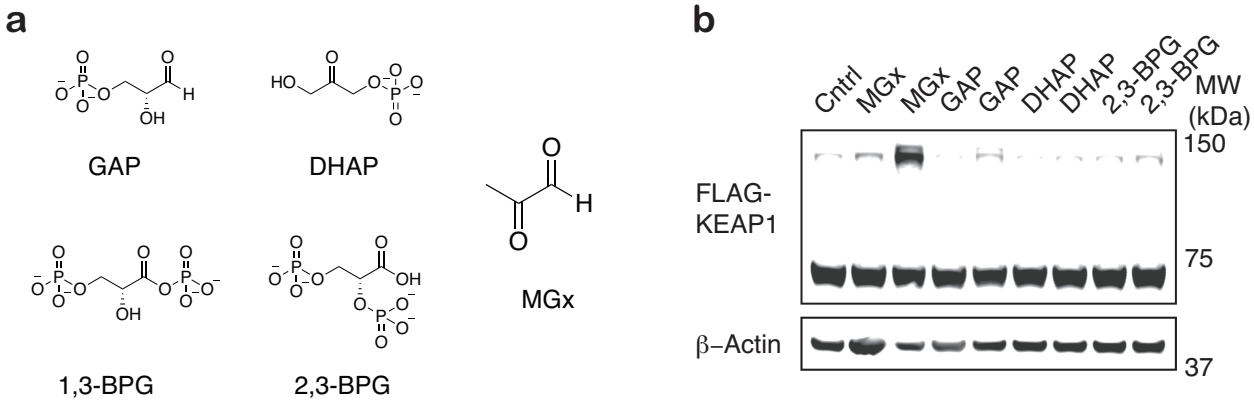


Figure 2.23 Screening reactive glycolytic metabolites identifies a glycolytic intermediate inducing HMW-KEAP1 formation in lysate
a. Structure of central glycolytic metabolites that are accumulated by CBR-470-1 induced PGK1 inhibition. **b.** Western blot monitoring of FLAG-KEAP1 migration in HEK293T lysates after treating with central glycolytic metabolites *in vitro* (1 and 5 mM, left and right for each metabolite).

Actual composition of commercial MGx solution (40% in H_2O) is a mixture of polymerized MGx and byproducts, thereby purified MGx is required for the study of quantitative MGx effects. To see if HMW-KEAP1 is directly formed by MGx in a dose-dependent manner, we treated freshly distilled MGx to the aliquots of FLAG-KEAP1 containing lysates with increasing MGx doses. In lysate HMW-KEAP1 formation was dose-dependently increased in mid-μM concentrations of MGx (IC_{50} = ~350 μM) which is consistent with the range of MGx concentrations previously reported in mammalian cell lines (Figure 2.24b)[168,169]. MGx-derived

modification studies have reported that most MGx exists as reversible protein adducts when exposed to complex proteome after lysis, and therefore the concentrations used *in vitro* likely mimic lower free MGx levels in the cell. Likewise, MGx treatment to purified FLAG-KEAP1 under reducing conditions (1 mM TCEP or 1 mM DTT) resulted in HMW-KEAP1 formation (Figure 2.24c). Overall, the data indicate that MGx is a promising signaling messenger of NRF2 activation by inducing covalent KEAP1 dimerization.

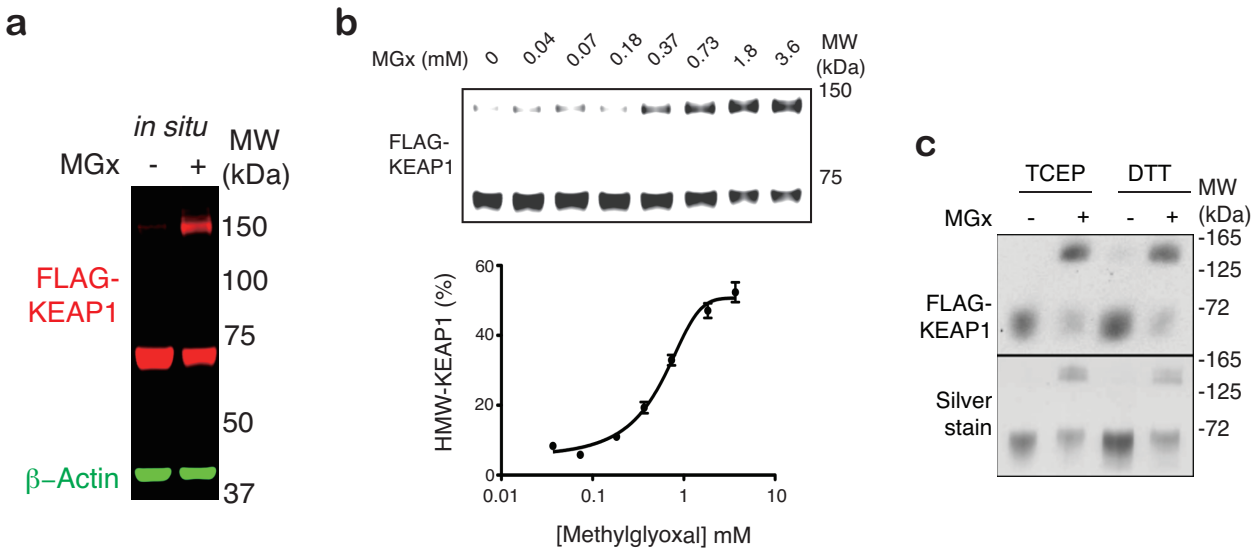
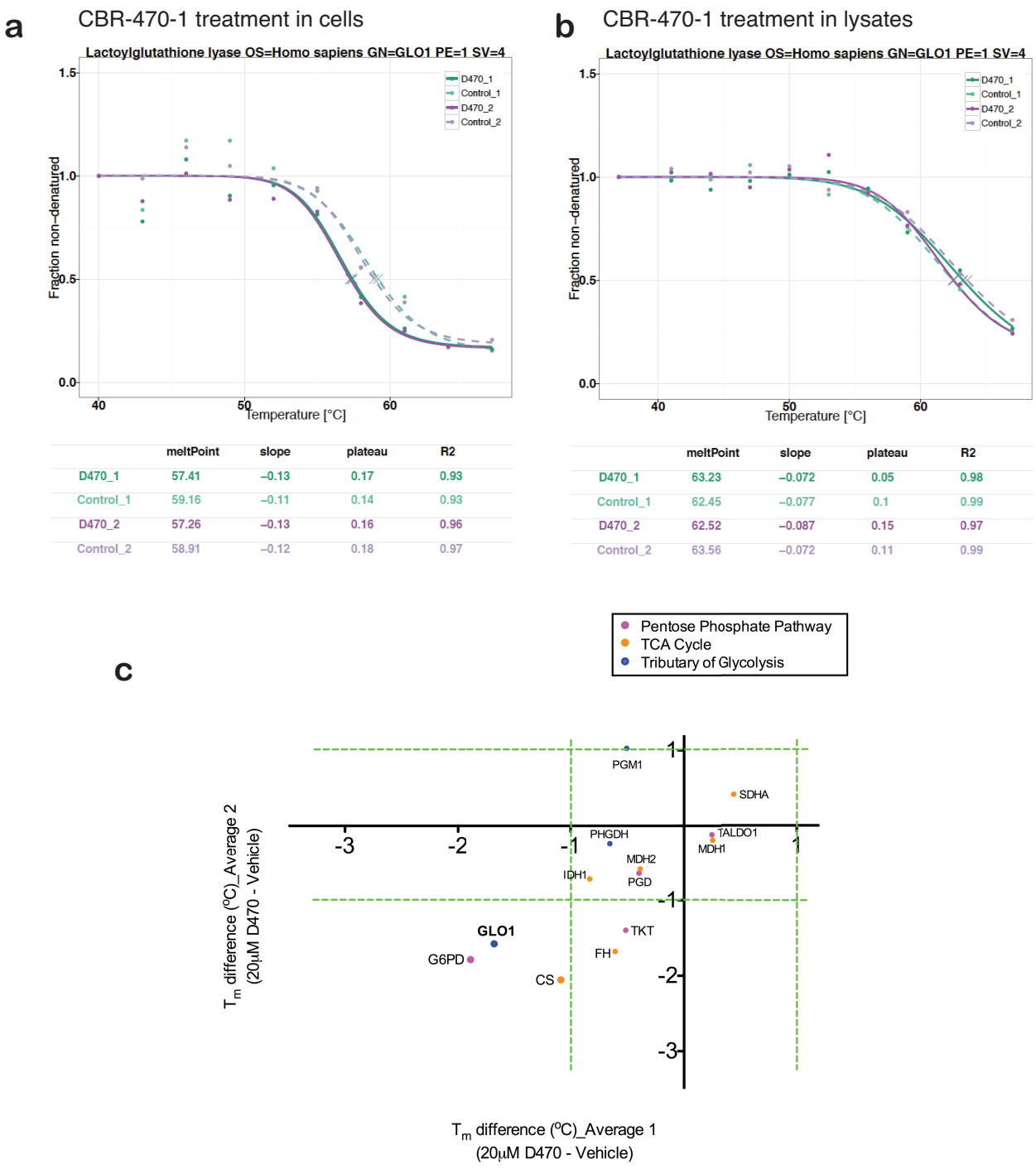


Figure 2.24 Methylglyoxal induces HMW-KEAP1 *in vitro* and in cells
a. FLAG-KEAP1 (red) and β-actin (green) from HEK293T cells treated with MGx (1 mM) or water control (n=3). **b.** Anti-FLAG Western blot analyses of FLAG-KEAP1 monomer and HMWKEAP1 fraction with dose-dependent incubation of distilled MGx in lysate from HEK293T cells expressing FLAG-KEAP1 (n=4). **c.** SDS-PAGE gel (silver stain) and anti-FLAG Western blot analyses of purified KEAP1 treated with the MGx under the indicated reducing conditions for 2 hours at 37 °C (n=3). Purified protein reactions were quenched in 4x SDS loading buffer containing βMe and processed for gel analysis as in (d). Data shown present mean ± SEM of biologically independent samples.



	meltPoint	slope	plateau	R2
D470_1	57.41	−0.13	0.17	0.93
Control_1	59.16	−0.11	0.14	0.93
D470_2	57.26	−0.13	0.16	0.96
Control_2	58.91	−0.12	0.18	0.97

	meltPoint	slope	plateau	R2
D470_1	63.23	−0.072	0.05	0.98
Control_1	62.45	−0.077	0.1	0.99
D470_2	62.52	−0.087	0.15	0.97
Control_2	63.56	−0.072	0.11	0.99

Figure 2.25 CBR-470-1 affects methylglyoxal detoxification system
a. b. Melt curve of GLO1 from cells treated with either CBR-470-1 or DMSO **(a)** or from lysates treated with either CBR-470-1 or DMSO **(b)**. Calculated T_m values are listed. **c.** Scatter plot of T_m shifts calculated from the four biological replicates of the CBR-470-1 versus DMSO treatment experiment. Each axis value (x-axis and y-axis values of each protein) reflected the mean ± s.e.m. of T_m shift values from two biological replicates each.

To globally screen the shift of protein stabilities as a consequence of CBR-470-1 treatment in live cells, we applied CETSA *in situ* and analyzed the data with quantitative proteomics workflow. Treatment of 10 μM CBR-470-1 to HeLa cells in 1 hour induced consistent high T_m shifts of proteins, including PGK1 and Glyoxalase 1 (GLO1, ΔT_m < -1.6 °C) (Figure 2.25a and 2.25b). Thermal stability of GLO1 was not shifted in cell lysate with CBR-470-1 treatment, suggesting that GLO1 is not a direct target of CBR-470-1 but a target of altered metabolic fluxes mediated by PGK1 inhibition (Figure 2.25c). Endogenous MGx generation in mammalian cells is primarily attributed to non-enzymatic degradation of the intermediate of glycolytic triosephosphates in triosephosphate isomerase 1 (TPI1). After generated and accumulated in cells, cellular MGx level is mainly regulated by glyoxalase system. The role of GLO1 in glyoxalase pathway is to catalyze the oxidation of the hemithioacetal product generated from the reaction of MGx and reduced glutathione (GSH), which is further metabolized by GLO2 to regenerate GSH and lactate. Data from CETSA indicates that modulation of glycolysis by PGK1 inhibition may induces the accumulation of cellular MGx that covalently modifies KEAP1 and activates NRF2 signaling cascade.

Cellular MGx was chemically trapped by molar excess 1,2-diaminobenzene after the lysis, and targeted LC-MS measurements of the product were performed to investigate the effect of CBR-470-1 cell treatment on endogenous MGx levels (Figure 2.26a, 2.26b, and 2.26c). Kinetic analysis of MGx levels after 20 μM CBR-470-1 treatments in IMR32 cells revealed a significant elevation of cellular MGx levels, reaching to the peak within the first few hours of treatment, approximate 4-fold elevation after 1 hour (Figure 2.26d). This result matched the kinetics of glycolytic flux perturbation by CBR-470-1, HMW-KEAP1 formation, and activation of NRF2 pathway. Consistent with the ability of CBR-470-1 to increase cellular MGx levels and the

formation of HMW-KEAP1 by MGx treatment in cells, we observed the expression of NRF2-dependent genes *NQO1* and *HMOX1* with MGx treatment in IMR32 cells (Figure 2.26e).

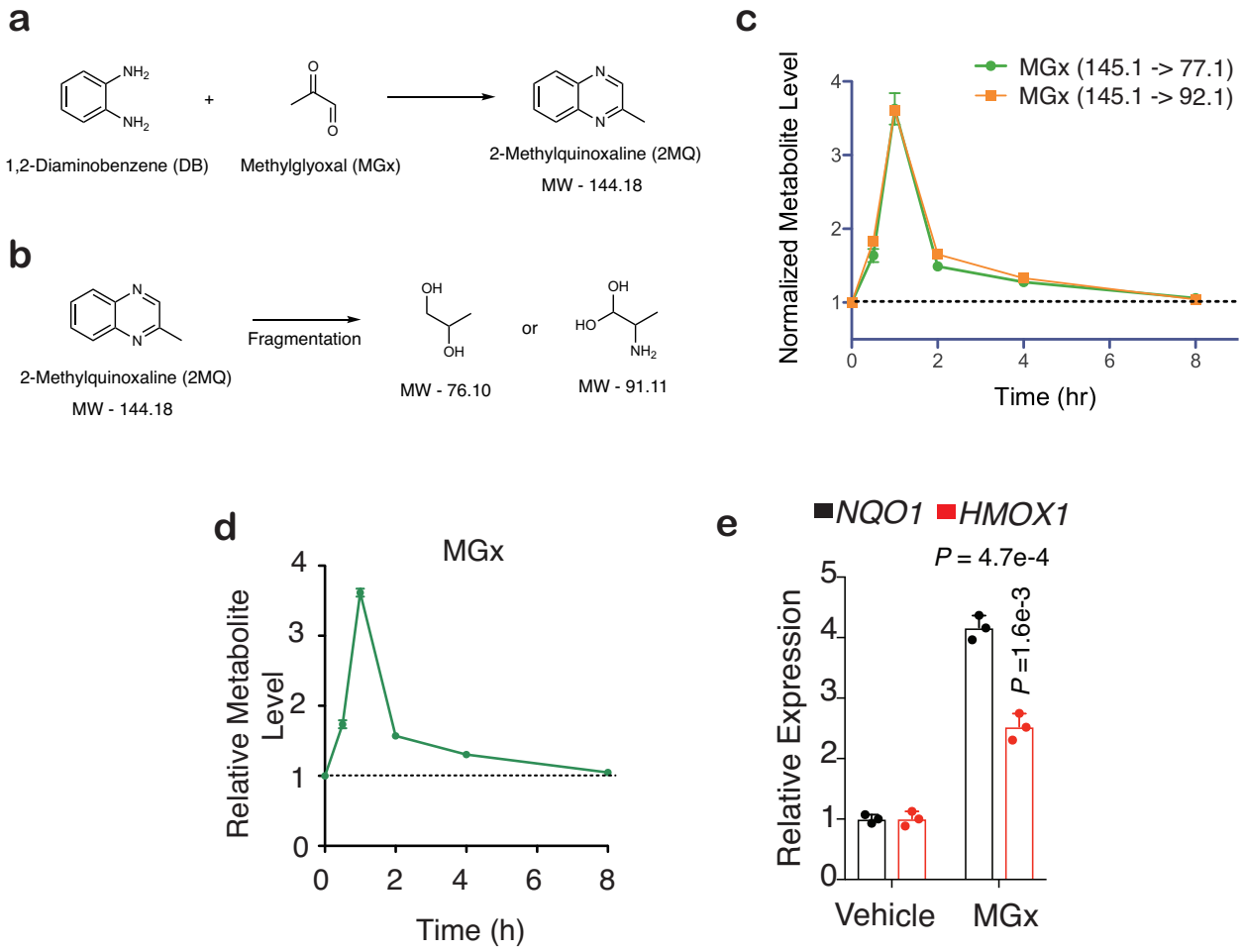


Figure 2.26 CBR-470-1 causes elevated methylglyoxal levels in cells
a. Schematic depicting chemical derivatization and trapping of cellular MGx for analysis by targeting metabolomics using two unique fragment ions. **b. c.** Daughter ion fragments **(b)** and resulting MS/MS quantification of MGx levels **(c)** in IMR32 cells treated with CBR-470-1, relative to DMSO (n=4). **d.** LC-MS/MS quantitation of cellular MGx levels in IMR32 cells treated with CBR-470-1 relative to DMSO (n=4). **e.** Relative *NQO1* and *HMOX1* mRNA levels in IMR32 cells treated with MGx (1 mM) or water control (n=3). Univariate two-sided t-test (d); Data shown present mean ± SEM of biologically independent samples.

Next, we sought to modulate cellular MGx levels and investigate the direct connection between MGx and KEAP1-NRF2 pathway. Data from CETSA *in situ* already pointed out the endogenous MGx mediator GLO1, thus we regulated endogenous MGx levels by either intervening GLO1 or supplying co-factor GSH in glyoxalase system. To read out NRF2 activation, ARE-LUC assay was used in the study. shRNA-based knockdown of GLO1 resulted in ARE-LUC signals and sensitized the cell to CBR-470-1 mediated activation of NRF2 signaling (Figure 2.27a, and 2.27b). Likewise, direct modulation of GLO1 activity with a cell-permeable inhibitor (GLOi) also amplified the reporter activation by CBR-470-1 (Figure 2.27c). Collectively, interruption of MGx detoxification pathway makes NRF2 axis more susceptible to perturbed glycolytic pathway.

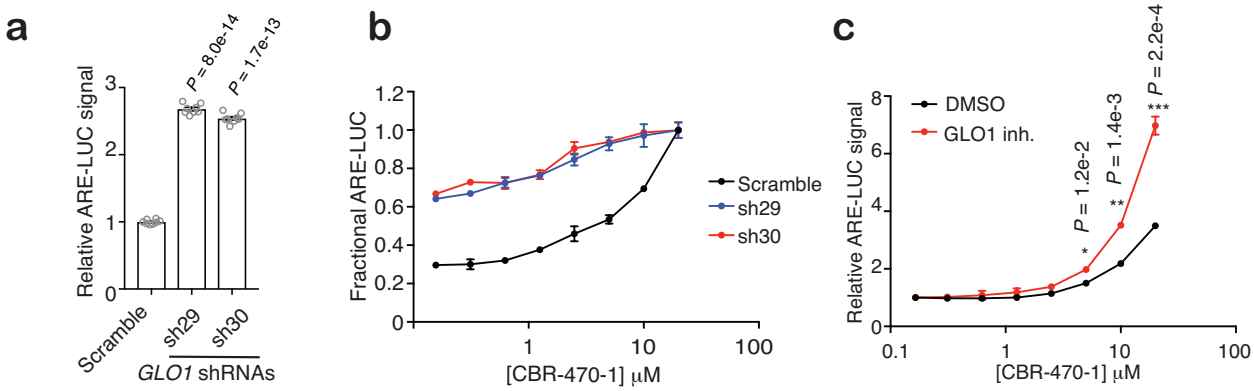


Figure 2.27 Glyoxalase activity regulates NRF2 activation in cells
a. ARE-LUC reporter activity in HEK293T cells with transient shRNA knockdown of GLO1 (n=8) **b.** Fractional ARE-LUC values from HEK293T cells transiently co-transfected with pTI-ARE-LUC and the indicated shRNAs and then treated for 24 hours with the indicated doses of CBR-470-1 (n=3) **c.** ARE-LUC reporter activity in HEK293T cells treated with CBR-470-1 alone (black) and with a cell-permeable small molecule GLO1 inhibitor (red) (n=3).

Cellular glutathione is an essential cofactor of MGx detoxification process, thus regulation of GSH levels should affect the activation of glyoxalase system and cellular MGx levels. Co-treatment of 2 mM NAC, a glutathione precursor, with 20 μM CBR-470-1 in IMR32 cells resulted in reduction of CBR-470-1 dependent MGx accumulation which is consistent with dose-dependent inhibition of HMW-KEAP1 formation by GSH or NAC (Figure 2.28a). Indeed, dose-dependent reduction of ARE-LUC signals and induction of HMOX and NQO1 mRNA induction were observed with GSH (Figure 2.28b and 2.28c). In summary, these metabolomic, proteomic, and transcriptomic data established a direct link between glycolysis and the KEAP1-NRF2 signaling pathway mediated by the non-enzymatic, direct modification of KEAP1 by a reactive glycolytic metabolite, MGx.

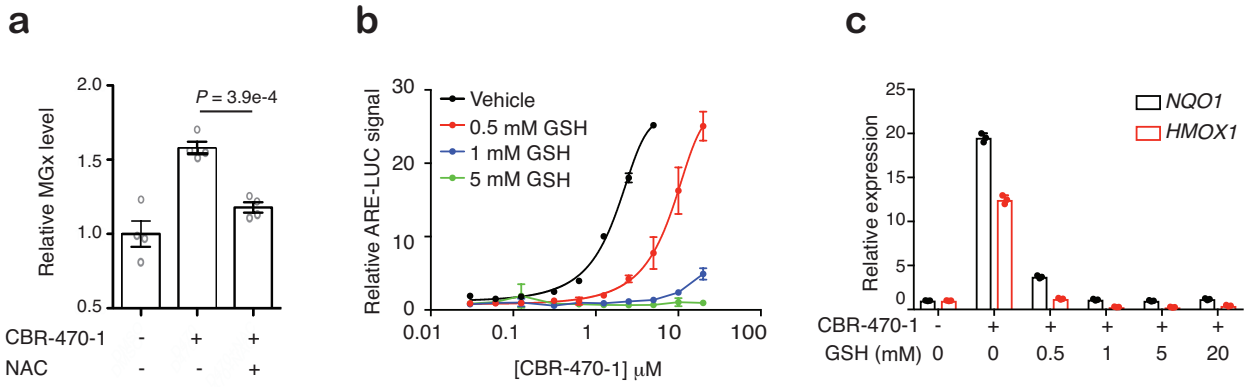


Figure 2.28 GSH activates glyoxalase system and reduces MGx in cells
a. Quantitative LC-MS/MS measurement of cellular MGx levels in IMR32 cells treated for 2 hours with CBR-470-1 or co-treated for 2 hours with CBR-470-1 and NAC (2 mM) relative to DMSO (n=4). **b.** Relative ARE-LUC luminance values from IMR32 cells co-treated with CBR-470-1 (10 μM) and the indicated concentrations of GSH for 24 hours (n=3). **c.** Relative levels of transcripts NQO1 and HMOX1 from IMR32 cells co-treated with CBR-470-1 (10 μM) and the indicated concentrations of GSH for 24 hours (n=3). Univariate two-sided t-test (c,d); Data shown present mean ± SEM of biologically independent samples.

2.2.4. Methylglyoxal forms a novel posttranslational modification between cysteine and arginine residues in KEAP1

Although an endogenous activator of NRF2 signaling that generates covalent HMW-KEAP1 was revealed with our experiments, the structural features of HMW-KEAP1 were not identified. Therefore, we addressed the following questions: 1) Does HMW-KEAP1 homodimer? 2) Which residues are covalently crosslinked by MGx? 3) What is the chemical structure of the modification?

2.2.4.1 Methylglyoxal forms irreversible crosslink between cysteine and arginine residues to generate covalent KEAP1 homodimer

HMW-KEAP1 has been observed to have molecular weight of approximately 140 kDa, which is closed to the double mass of monomeric KEAP1. However, it was still undefined if MGx induces covalent homodimerization between two KEAP1 monomers or heterodimerization between monomeric KEAP1 and another binding partner owning a similar mass.

In order to identify potential binding partners possibly consisting heterodimeric HMW-KEAP1 with monomeric KEAP1, we employed a SILAC-quantitative proteomic approach to identify the proteins co-enriched with HMW-KEAP1 from cells. From both normal SILAC treatment (Heavy: CBR-470-1 / Light: DMSO) and label-swapped treatment (Heavy: DMSO / Light: CBR-470-1), 71 kDa heat-shock protein HSPA1A was shown to be consistently enriched with HMW-KEAP1 (Figure 2.29a). To further validate if HSPA1A is covalently crosslinked with monomeric KEAP1 to form heterodimer, we performed Western blot analysis to find α-HSPA1A bands of HMW-KEAP1. However, α-HSPA1A immune-reactive bands were not overlapped with

HMW-KEAP1 bands but only observed near 70 kDa, indicating that monomeric KEAP1 does not form HMW-KEAP1 with HSPA1A (Figure 2.29b).

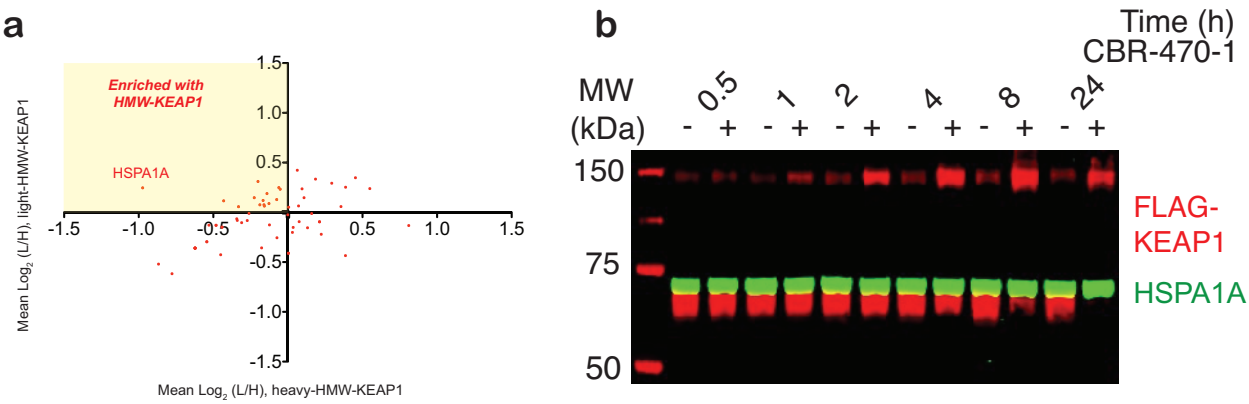


Figure 2.29 HMW-KEAP1 may not be a heterodimer
a. Scatter plot of the list of proteins of which SILAC ratio is calculated from Log_2(HMW-KEAP1/monomeric KEAP1) of normal- and label-swapped SILAC experiments of the CBR-470-1 versus DMSO treatment experiment. **b.** Time-course, anti-FLAG and anti-HSPA1A Western blot analysis of whole cell lysates from HEK293T cells expressing FLAG-KEAP1 treated with DMSO or CBR-470-1.

Understanding structural alterations and motions of KEAP1 induced by covalent modifications of electrophile sensor residues has been remained as a challenge, yet most of structural studies have focused on individual domains in isolation[138,160,170]. So far, crystal structures of isolated BTB and KELCH domains were revealed, however, new approach is demanded to screen the entire protein and to identify potential sites of the crosslink in KEAP1 with unbiased ways. Thus, we developed a SILAC-based quantitative proteomic method based on changes in tryptic peptide ratios across the entire protein (Figure 2.30a). This approach enables the

differentiation between unmodified- and modified regions of entire protein by quantifying the relative population of unmodified, each tryptic peptide from heavy HMW-KEAP1 and light monomeric KEAP digests. The assay was optimized, with covering above 95% regions of KEAP1, and specific tryptic peptides (16-39, 16-50, and 151-169) were distinguished by skewed SILAC ratios (Figure 2.30b and 2.30c). The results suggested that residues in NTR and BTB domains could be involved in HMW-KEAP1 formation in response to CBR-470-1 induced cellular MGx elevations in cells and MGx induced non-enzymatic crosslink in lysate (Figure 2.31).

Given hints from the quantitative proteomic assay, we generate focused set of FLAG-KEAP1 constructs containing mutations of nucleophilic amino acids, including cysteine, lysine and arginine that could be involved in known or unknown glycations by MGx[104]. More than a dozen C-to-S, K-to-M/R, and R-to-A mutations within the NTR and BTB domains identified by SILAC mapping, as well as other known functional cysteine residues in KEAP1, were queried for their effect on HMW-KEAP1 formation. Among 15 mutation sites of KEAP1, R-to-A mutations of two arginine residues (R15 of the NTR domain and R135 of the BTB domain) significantly, but incompletely, reduced the CBR-470-1 induced HMW-KEAP1 formation in cells. Surprisingly, near complete inhibition of HMW-KEAP1 formation was observed in C-to-S mutation of C151 residue in KEAP1 BTB domain, and R15/135A & C151S triple mutant completely blocked CBR-470-1 induced HMW-KEAP1 formation (Figure 2.32a, 2.32b, and 2.32c). Consistent with the results from SILAC mapping and mutant screening studies, pre-treatment of FLAG-KEAP1 transfected HEK293T cells with Bardoxolone methyl, which selectively alkylates C151, inhibited CBR-470-1 induced HMW-KEAP1 formation (Figure 2.32d).

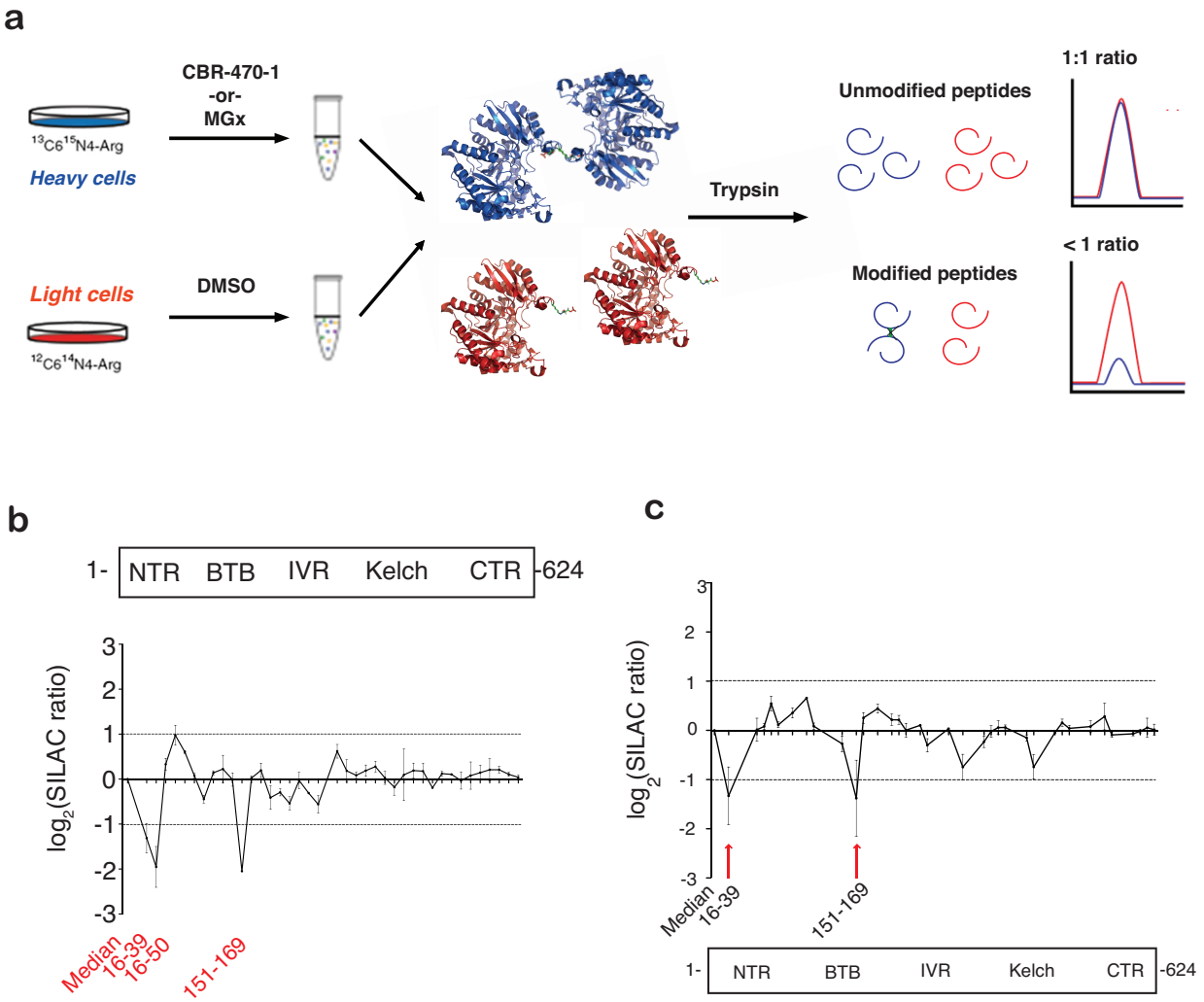


Figure 2.30 SILAC-based protein studies of HMW-KEAP1

a. Stable isotope-labeled cells (stable isotope labeling with amino acids in cell culture, SILAC) expressing FLAG-tagged KEAP1 were treated with vehicle ('light') and CBR-470-1 or MGx ('heavy'), respectively. Subsequent mixing of the cell lysates, anti-FLAG enrichment, tryptic digestion and LC-MS/MS analysis permitted detection of unmodified portions of KEAP1, which retained ~1:1 SILAC ratios relative to the median ratios for all detected KEAP1 peptides. In contrast, peptides that are modified under one condition will no longer match tryptic MS/MS searches, resulting skewed SILAC ratios that "drop out" (bottom). **b.** SILAC ratios for individual tryptic peptides from FLAG-KEAP1 enriched DMSO treated 'light' cells and CBR-470-1 treated 'heavy' cells, relative to the median ration of all KEAP1 peptides. Highlighted tryptic peptides were significantly reduced by 3- to 4- fold upon relative to the KEAP1 median, indicative of structural modification (n=8). **c.** SILAC ratios for individual tryptic peptides from FLAG-KEAP1 enriched MGx treated 'heavy' cell lysates and no treated 'light' cell lysates, relative to the median ratio of all KEAP1 peptides. Highlighted tryptic peptides were significantly reduced by 2- to 2.5-fold upon relative to the KEAP1 median, indicative of structural modification (n=12). Data shown present mean ± SEM of biologically independent samples.

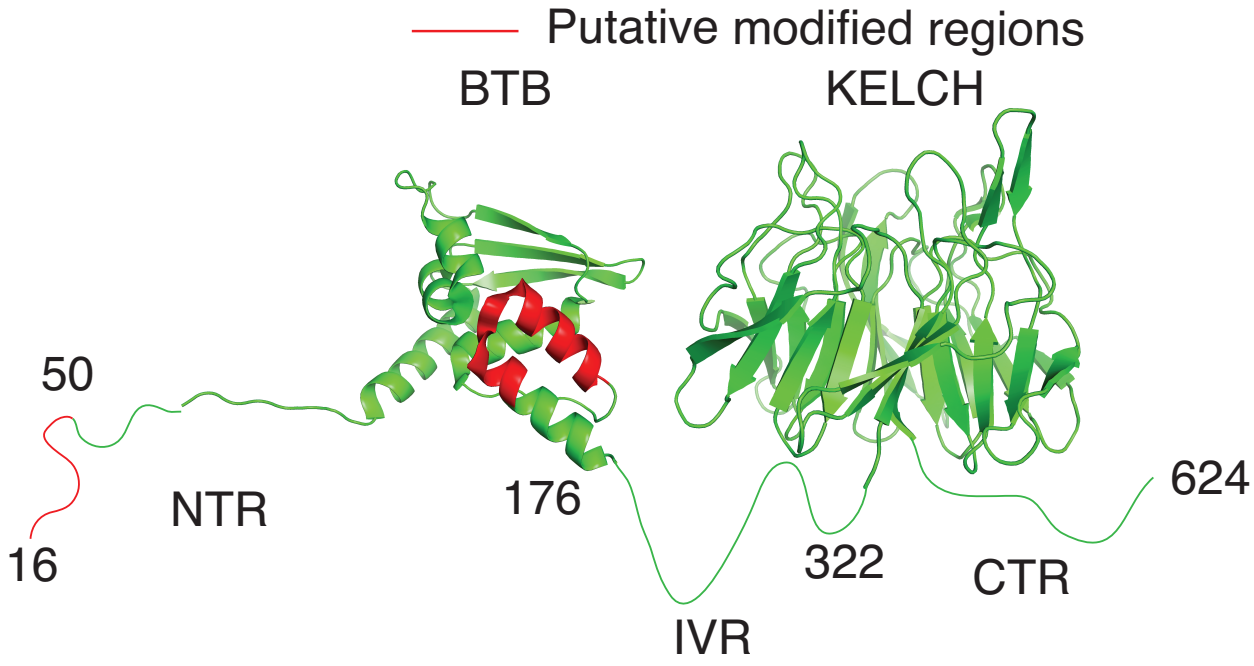


Figure 2.31 Structural depiction of potentially crosslinked sites of HMW-KEAP1
Structural depiction of potentially modified stretches of human KEAP1 (red) using published x-ray crystal structure of the BTB (PDB: 4CXI) and KELCH (PDB: 1U6D) domains. Intervening protei stretches are depicted as unstructured loops in green.

C151 is a well characterized reactive cysteine residue that has been targeted for NRF2 activation[171], such as Bardoxolone methyl[138]. This residue is situated in an exposed region of BTB domain that is predicted to mediate homo-dimeric interface between two KEAP1 monomers with NTR and BTB domains of another monomeric KEAP1, which is necessary for proper NRF2 binding and degradation of NRF2 by E3-dependent ubiquitination[123,160]. Thus, strong abrogation of HMW-KEAP1 formation through mutation of C151 and proximal arginine residues in the BTB and NTR regions suggested that MGx may be mediating an uncharacterized modification between these residues to form intermolecular crosslink in two KEAP1 monomers.

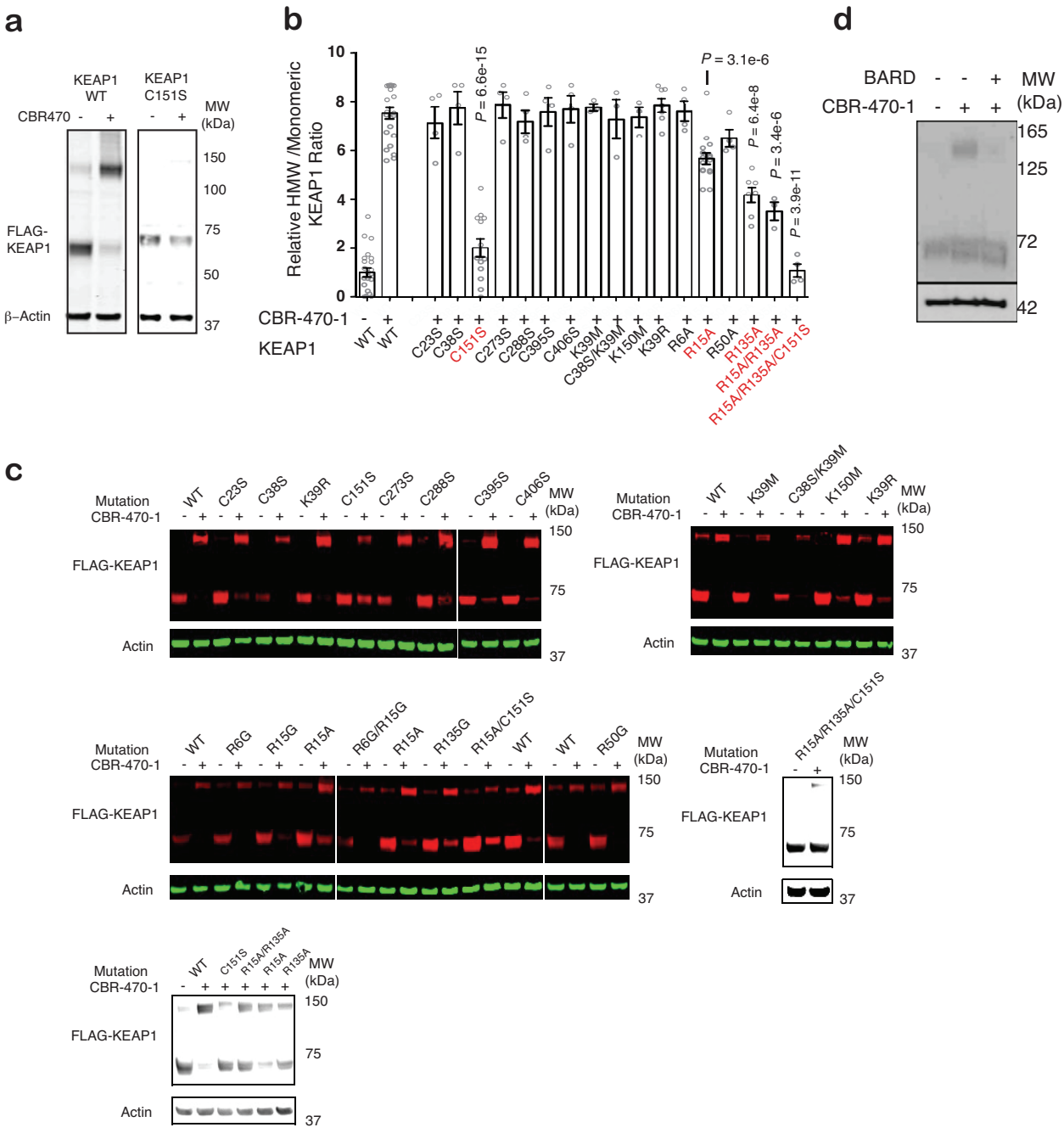


Figure 2.32 Identification of modified sites in HMW-KEAP1

a. Anti-FLAG western blot analyses of wild-type and C151S mutant of FLAG-KEAP1 from HEK293T cells treated with DMSO or CBR-470-1 for 8 hours. **b.** Quantified HMW-KEAP1 formation of wild-type or mutant FLAG-KEAP1 from HEK293T cells treated with DMSO or CBR-470-1 for 8 hours (n=23 for WT; n=16 for R15A; n=13 for C151S; n=7 for K39R, R135; n=4 for R6A, R50A, all other C-to-S mutations, and R15/135A & C151S triple-mutant; n=3 for R15/135A, and all K-to-M mutations). **c.** Representative Western blot images of mutant FLAG-KEAP1 screens with CBR-470-1. **d.** Representative Western blotting analysis of FLAG-KEAP1 dimerization from HK293T cells pre-treated with Bardoxolone methyl followed by CBR-470-1.

(Figure 2.32, continued) treatment for 4 hours (n=3). Univariate two-sided t-test (d); Data shown present mean ± SEM of biologically independent samples.

2.2.4.2 Methylimidazole crosslink between cysteine and arginine residues is formed between two KEAP1 monomers by Methylglyoxal

In order to identify a modification that would occur between nucleophilic amino acids, we synthesize model peptides containing two reactive amino acids from the combination of cysteine, arginine, and lysine, with a separation of these amino acids by a three-glycine linker. The design of these peptides was intended to mimic high inter- and intramolecular proximity of potential crosslinked amino acid residues by MGx (Figure 2.33a). Overnight treatment of the peptides with 12.5 mM MGx at physiologic temperature (37 °C) and pH (pH = 7.4) followed by LC-MS analysis identified a distinct peak from the mixture of MGx and cysteine-arginine (CR) peptide (Figure 2.33b). The m/z value of this new peak was 838 Da, which corresponds to a mass increase of 36 Da and is consistent with the insertion of 3 carbon atoms.

A strong candidate of this modification is mercapto methylimidazole crosslink between cysteine and arginine residues. Chemical structure of this potential modification is consistent with nucleophilic attack of the MGx dicarbonyl moiety by proximal and appropriately activated side chains of cysteine and arginine, coupled with subsequent dehydration-mediated cyclization reactions, which results in the formation of a novel methylimidazole crosslink between cysteine and arginine (MICA) posttranslational modification.

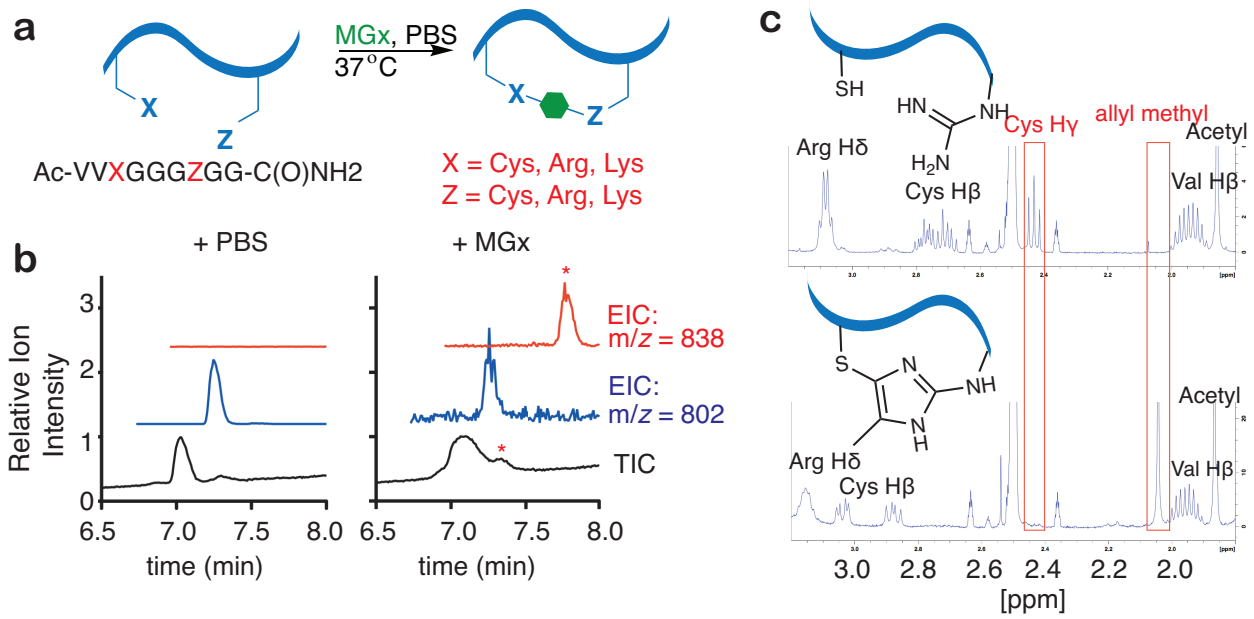


Figure 2.33 Structural characterization of CR-MGx peptide
a. Schematic of the model peptide screen for intramolecular modifications formed by MGx and nucleophilic residues. **b.** Total ion- (TIC) and extracted ion chromatograms (EIC) from MGx- and mock-treated peptide, with a new peak in the former condition marked with an asterisk. EICs are specific to the indicated m/z. (n=3 independent biological replicates). **c.** ^{1}H-NMR spectra of the unmodified (top) and MICA-modified (bottom) model peptide, with pertinent protons highlighted in each. Notable changes in the MICA-modified spectrum include the appearance of a single at 2.04 p.p.m. (allyl methyl in MICA), loss of the thiol proton at 2.43 p.p.m., and changes in chemical shift and splitting pattern of the cysteine beta protons and the arginine delta epsilon protons.

To investigate this crosslink in greater detail and elucidate its chemical structure, we first purified CR-MGx product by HPLC runs and then performed NMR experiments with ^{1}H-NMR, homonuclear correlation spectroscopy double quantum filter (COSY-DOF), and total correlation spectroscopy (TOCSY) measurements (Figure 2.34a, 2.34b, 2.34c, and 2.34d). Comparison of two-dimensional NMR spectrum between modified peptide (CR-MGx) and unmodified (CR) peptide resulted in similar coupling patterns, allowing for additional confirmation of protein assignments that provides spatial information of additional protons adopted in MICA peptide modification Indeed, direct ^{1}H-NMR spectra comparison highlighted few key differences,

disappearance of cysteine thiol proton peak at 2.43 p.p.m. and appearance of 3-proton singlet peak of allyl methyl in MICA at 2.04 p.p.m., confirming the proposed crosslink (Figure 2.33c). MICA protein modification was not previously reported thus we discovered and characterized a novel glycation structure.

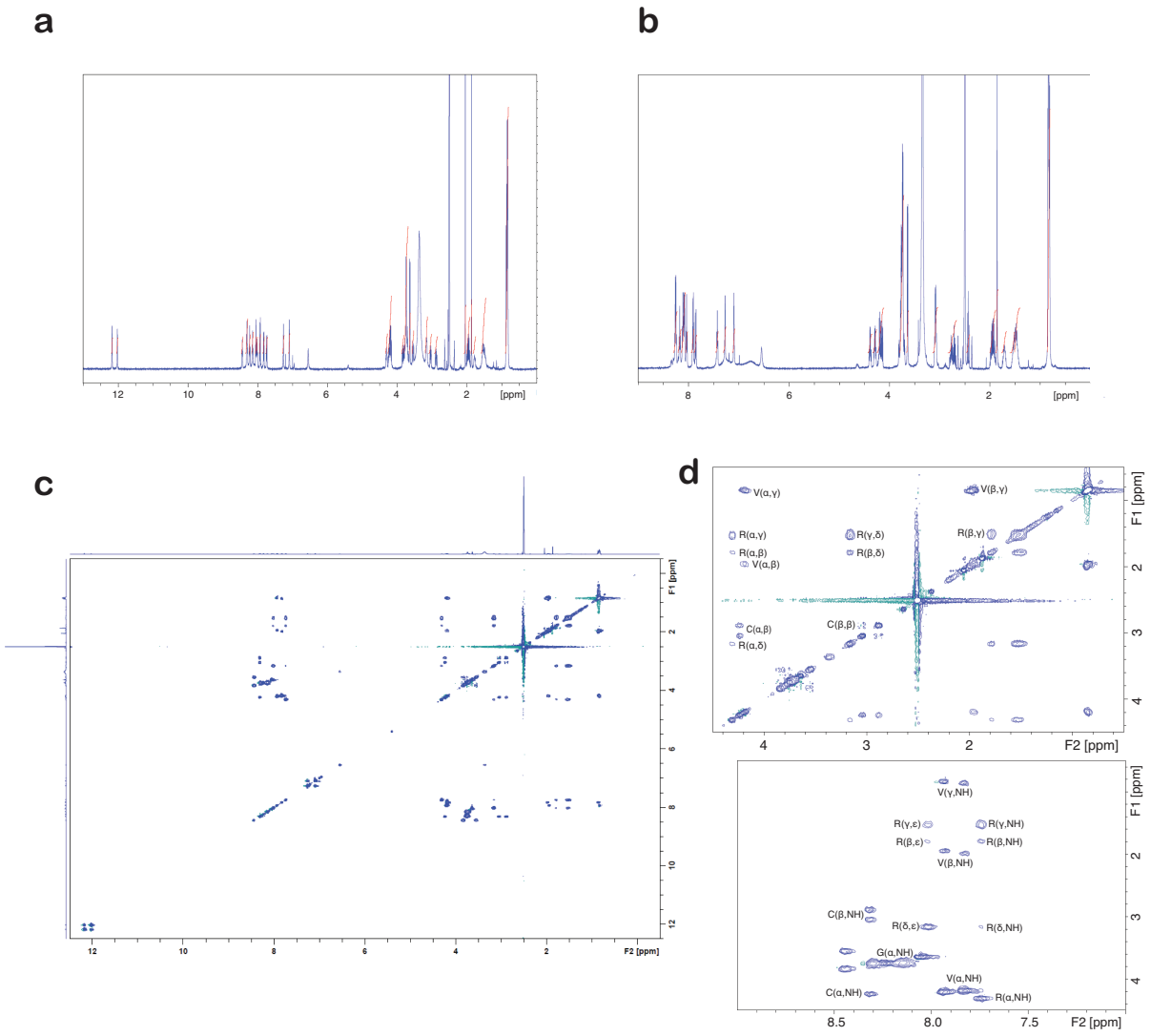


Figure 2.34 NMR spectrum of CR-MGx peptide

a. ^{1}H-NMR of CR-MGx peptide (isolated product of MGx incubated with Ac-NH-VVCGGGRGG-C(O)NH_2 peptide). ^{1}H NMR (500MHz, d6-DMSO) δ 12.17 (s, 1H), 12.02 (s, 1H), 8.44 (t, J = 5.6 Hz, 1H), 8.32-8.29 (m, 2H), 8.23 (t, J = 5.6 Hz, 1H), 8.14 (t, J = 5.9 Hz, 1H), 8.05 (t, J = 5.9 Hz, 1H), 8.01 (t, J = 5.9 Hz, 1H), 7.93 (d, J = 8.5 Hz, 1H), 7.74 (d, J = 8.0 Hz, 1H), 7.26 (s, 1H), 7.09 (s, 1H), 4.33-4.28 (m, 1H), 4.25-4.16 (m, 3H), 3.83 (dd, J = 6.9 Hz, J = 16.2 Hz, 1H),

(Figure 2.34, continued) 3.79-3.67 (m, 6H), 3.63 (d, J = 5.7 Hz, 2H), 3.54 (dd, J = 4.9 Hz, J = 16.2 Hz, 1H), 3.18-3.13 (m, 2H), 3.04 (dd, J = 4.9 Hz, J = 13.9 Hz, 1H), 2.88 (dd, J = 8.6 Hz, J = 13.6 Hz, 1H), 2.04 (s, 3H), 1.96 (sep, J = 6.8 Hz, 2H), 1.87 (s, 3H), 1.80-1.75 (m, 1H), 1.56-1.47 (m, 3H), .87-.82 (m, 12H). **b.** ^{1}H-NMR of CR peptide (Ac-NH-VVCGGGRGG-C(O)NH_2). ^{1}H NMR (500MHz, d6-DMSO) δ 8.27-8.24 (m, 2H), 8.18 (t, J = 5.7 Hz, 1H), 8.13-8.08 (m, 3H), 8.04 (t, J = 5.7 Hz, 1H), 7.91 (d, J = 8.8 Hz), 7.86 (d, J = 8.8 Hz, 1H), 7.43 (t, J = 5.4 Hz, 1H), 7.28 (s, 1H), 7.10 (s, 1H), 4.39 (dt, J = 5.6 Hz, J = 7.4 Hz, 1H), 4.28 (dt, J = 5.7 Hz, J = 7.2 Hz, 1H), 4.21-4.13 (m, 2H), 3.82-3.70 (m, 8H), 3.64 (d, J = 5.8, 2H), 3.08 (dt, J = 6.5 Hz, J = 6.5 Hz, 2H), 2.80-2.67 (m, 2H), 2.43 (t, J = 8.6 Hz, 1H), 1.94 (sep, J = 6.8 Hz, 2H), 1.85 (s, 3H), 1.75-1.68 (m, 1H), 1.54-1.42 (m, 3H), .85-.81 (m, 12H) **c.** ^{1}H-^{1}H TOCSY of CR-MGx peptide. **d.** Peak assignment for CR-MGx peptide TOCSY spectrum.

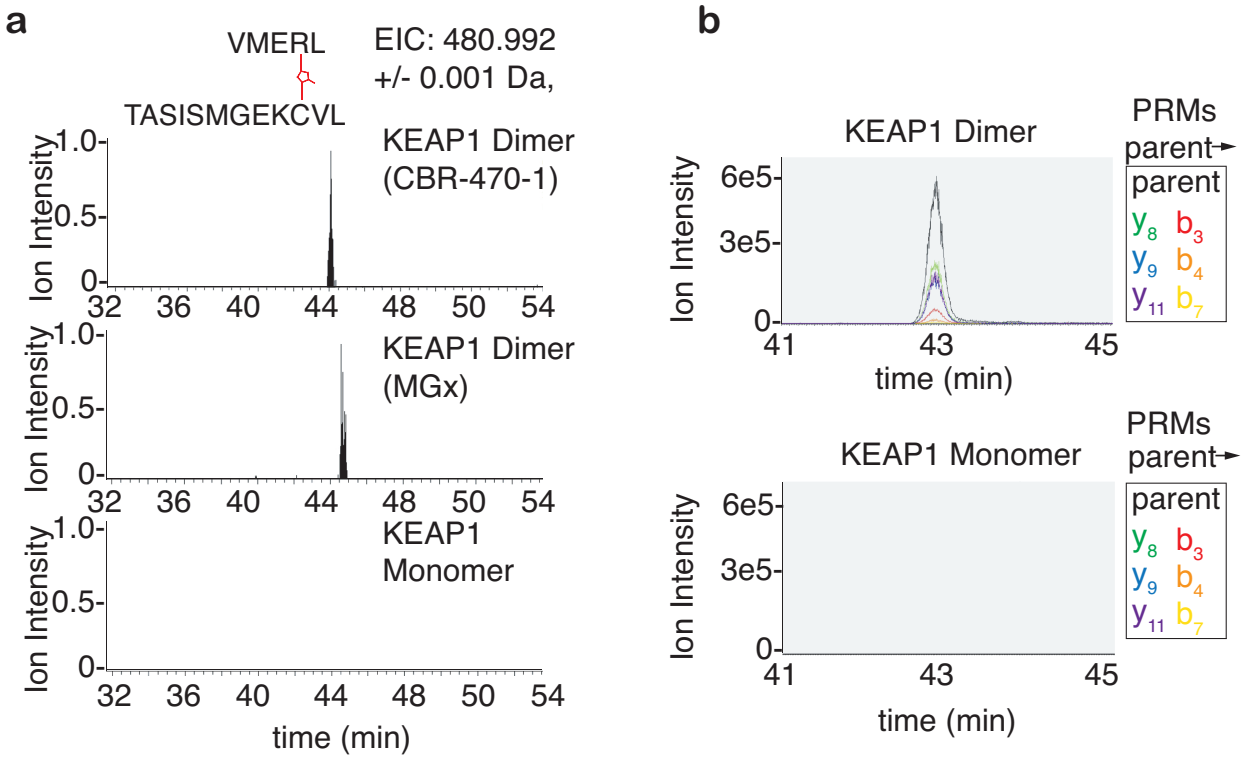


Figure 2.35 Methylglyoxal forms a novel posttranslational modification between proximal cysteine and arginine residues in KEAP1

a. EIC from LC-MS/MS analyses of gel-isolated and digested HMW-KEAP1 (CBR-470-1 and MGx-induced) and monomeric KEAP1 (DMSO and left over in MGx reaction mixtures) for the C151-R135 crosslinked peptide. Slight retention time variation was observed on commercial columns (n=3 independent biological replicates). **b.** PRM chromatograms for the parent and six-parent-to-daughter transitions in representative targeted proteomic runs from HMW-KEAP1 and monomeric digests (n=6).

In order to investigate whether MICA modification occurs between two KEAP1 monomers to form covalent homodimer, we first treated CBR-470-1 to FLAG-KEAP1 expressed cells or MGx to cell lysates followed by isolation of HMW-KEAP1 or monomeric KEAP1 by SDS-PAGE gels and then digested these discrete populations for LC-MS/MS analysis. MICA protein modification bears the crosslink between two distinct residues which brings complexity of the analysis of digests by standard search algorithms, thus m/z values of candidate MICA peptides calculated by considering predicted digestion sites, crosslinked sites (C151 and proximal arginine residues), and possible oxidation states of methionine residues were manually applied to identify extracted MS1 peaks only present in HMW-KEAP1 but not in monomeric KEAP1, with applying ± 10 p.p.m filter. A KEAP1 peptide bearing a MICA crosslink between C151 and R135 was identified in the digests of isolated HMW-KEAP formed by both CBR-470-1 in cell and MGx in lysate treatments, but not in the digests of all isolated monomeric KEAP1 from DMSO- cell treatments or even present in the MGx-lysate mixture (Figure 2.35a).

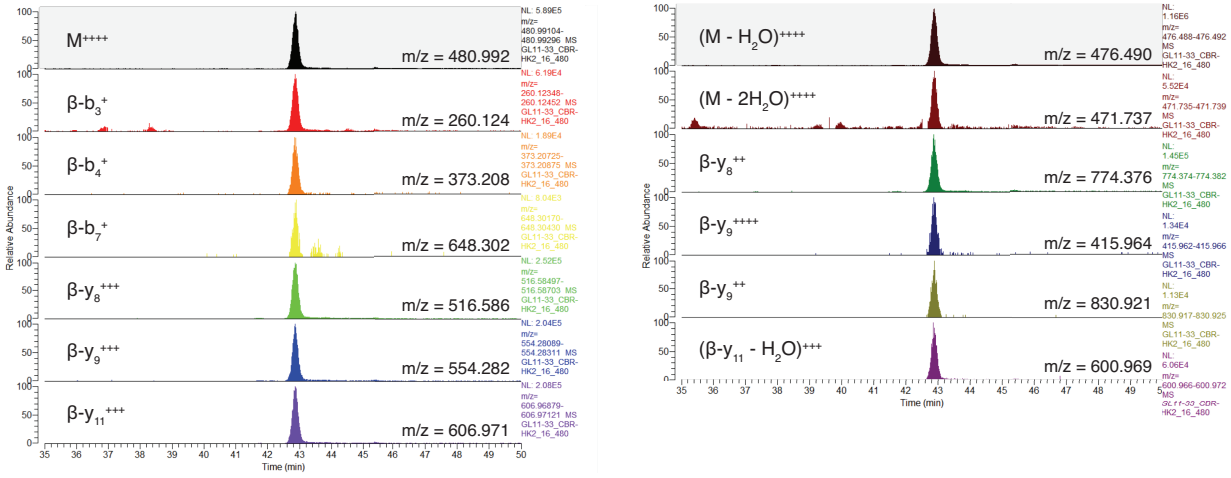


Figure 2.36 MS2 analysis of CR-MGx crosslinked KEAP1 peptide
Targeted Parallel Reaction Monitoring (PRM) transitions (n=6).

Final verification of the extracted species was performed in targeted parallel-reaction monitoring (PRM) mass spectrometry. b- and y- ions of the MICA KEAP1 peptide corresponding to the successfully extracted MS1 peak were estimated, and we confirmed the presence and co-elution of more than a dozen parent-to-daughter ion transitions of these b- and y- ions that were uniquely present in HMW-KEAP1 (Figure 2.35b, 2.36, and 2.37). These high-resolution spectroscopic and spectrometric analyses confirm the potential for proximal arginine and cysteine residues to form a novel imidazole crosslink that is mediated by MGx, and this modification directly forms intermolecular crosslink between proximal residues in KEAP1, hence, covalent homodimerization of two KEAP1 monomers.

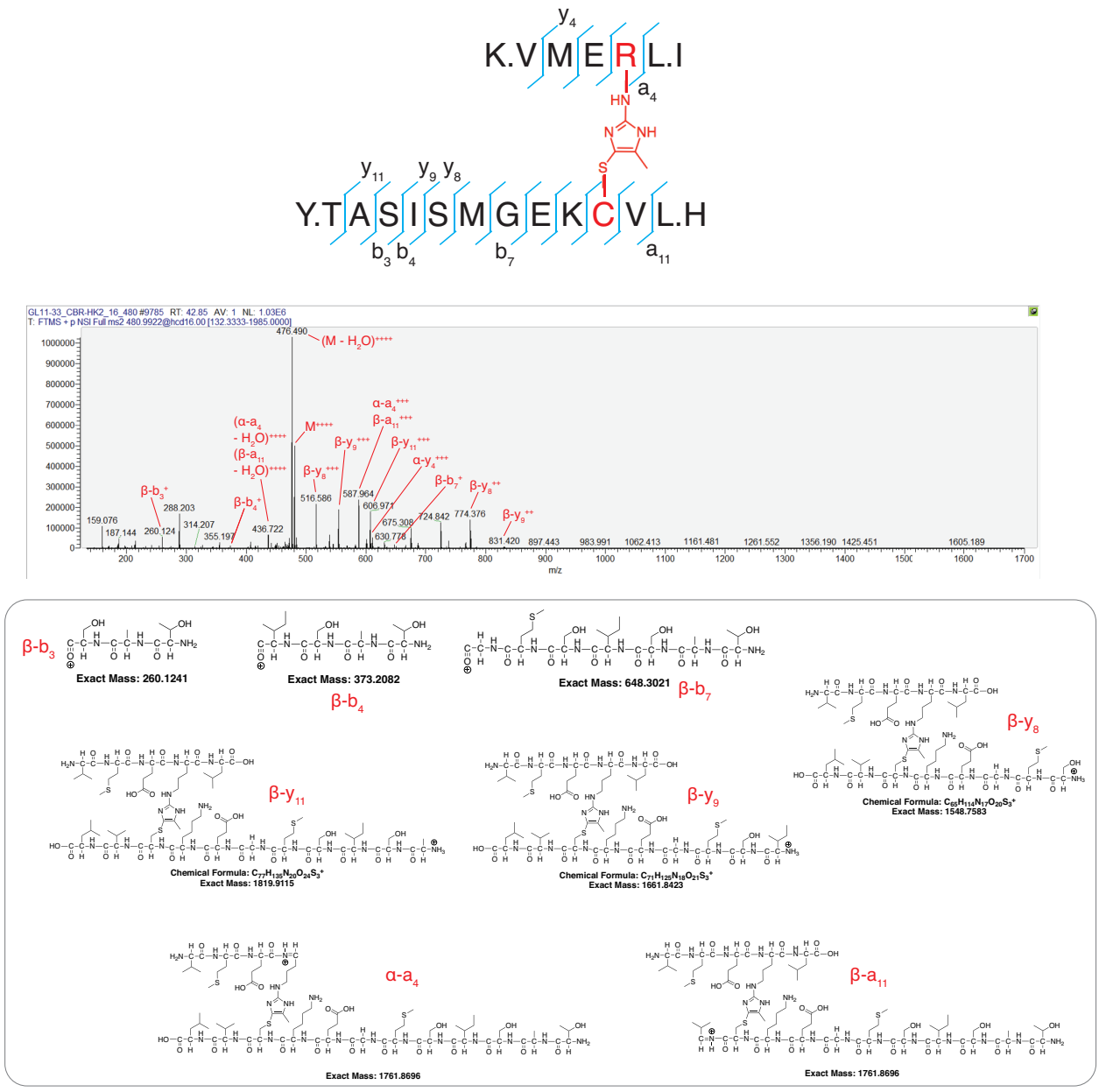


Figure 2.37 Annotated MS spectrum from the crosslinked C151-R135 KEAP1 peptide
Estimated b- and y- ions of the MICA KEAP1 peptide corresponding to the successfully extracted MS1 (Figure 2.35a).

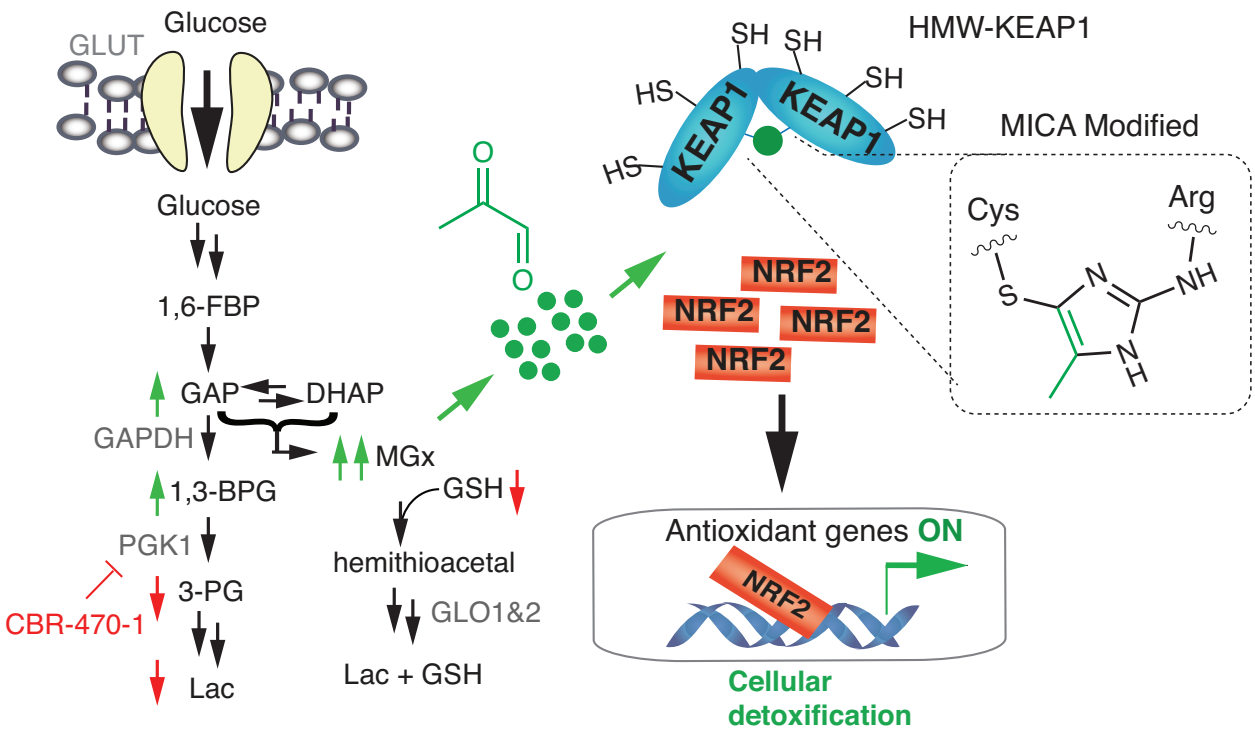


Figure 2.38 A model of direct communication between glucose metabolism and KEAP1-NRF2 signaling axis

The direct communication between glucose metabolism and NRF2 signaling mediated by MGx modification of KEAP1 and subsequent activation of the NRF2 transcriptional program.

2.3. Conclusion and Discussion

The KEAP1-NRF2 pathway, serving as a central sensor to detect and respond to a variety of cellular chemical insults, has been emphasized its clinical importance in connection to various human diseases. A longing for development of regulatory means of KEAP1-NRF2 pathway has led to the efforts mostly focused on direct KEAP1 modulations, therefore, elucidation of alternative regulatory routes would offer a variety of novel therapeutic strategies targeting NRF2 signaling cascade. Here, we proposed a model where glycolytic metabolic status is coupled to NRF2-dependent gene expression through the direct interaction of a reactive glycolytic metabolite MGx and the sentinel protein KEAP1 via formation of a stable and mechanistically novel protein posttranslational modification (Figure 2.38).

A powerful tool in adjusting cellular levels of specific reactive metabolites, thus allowing us to investigate the roles of non-enzymatic protein modifications in cell signaling, is a small molecular inhibitor of target metabolic enzyme. Beginning with cell-based high throughput screens and SAR studies, we discovered a novel NRF2 activator, CBR-470-1 which does not share an electrophilic feature implying an indirect regulation of NRF2 pathway. Application of various label- and label-free target identification technologies resulted in identification of central glycolytic enzyme PGK1 as a target of CBR-470-1. Following mechanism of action studies, including integrated transcriptomic, proteomic and metabolomic profiling, highlighted an unexpected direct link between the CBR-470-1 mediated elevation of central glycolytic metabolites, HMW-KEAP1 formation, and NRF2 activation. This suggests a future application of novel metabolic enzyme regulators in the study of the underlying communication networks between metabolism and cellular regulatory pathways via metabolite-derived non-enzymatic modifications in living system.

MGx, mainly generated from glycolytic pathway in cells, has been reported to form reversible and irreversible covalent modifications on nucleophilic amino acid residues in diverse proteins, however targets and regulatory functions of these modifications in cells and in diseases have still remained as a mystery. In our studies, we demonstrated that cellular MGx, which is accumulated in few hours after pharmacologic inhibition of glycolytic enzyme PGK1, is a signaling molecule inducing a non-enzymatic covalent homodimerization of KEAP1. The advanced glycation end products (AGEs) generated by MGx have been reported to be existed as various structures, modifications on single residue (e.g. MG-H1, MG-H2, MG-H3, CEL) or crosslinks between two residues (e.g. MOLD and MODIC). We discovered a MGx-derived novel posttranslational modification, MICA, that is formed a covalent linkage between proximal cysteine and arginine residues in HMW-KEAP1. A MICA crosslink between C151 and R135 residues in HMW-KEAP1, as evidenced by the results of SILAC mapping assay applied to HMW-KEAP1, the selective and dominant effect of single C151S mutation within the whole KEAP1 protein on CBR-470-1 induced covalent KEAP1 dimerization, and the identification of the KEAP1 MICA peptide only presenting in the digest of isolated HMW-KEAP1, indicates the selectivity of the non-enzymatic modification. While many of the cysteine residues in KEAP1 are known to be reactive electrophile sensors, C151 likely exhibits preferential modification by MGx and it could be explained by unique features of C151 residue: hyperreactivity of this residue (low pKa), inherent structural tolerance or affinity of MGx near the site, and the presence of properly oriented arginine residues inferred from the characterized structure of isolated BTB domain.

Recent studies of MGx-induced arginine modifications in histones also implied the role of MGx in regulations of target proteins[117,172]. In order to provide a global view of selective MGx modifications in the proteome and elucidate the roles of MGx-induced protein modifications in

cells and in diseases, as well as evaluating AGEs as promising biomarkers of disease pathology, our future studies will focus on full target profile of MGx-induced non-enzymatic protein modifications, especially MICA formation in inter- and intramolecular manners, and protein proximity profiles to find the mechanisms by which MGx delivers to the targets.

Finally, modulation of NRF2 signaling through alteration of glycolytic flux was shown to be cytoprotective in an established NRF2-responsive UV-induced damage mice model that is responsive to the clinical candidate Bardoxolone methyl. These data suggest that mechanisms beyond direct covalent targeting of KEAP1 using electrophilic small molecules may represent a viable therapeutic approach to the activate NRF2 in disease.

Reactive species in cells were considered as a flaw of the metabolism that must be removed from the system. Excess amounts of these intrinsically reactive molecules in the cell had been reported to react with the nucleophilic residues of proteins and other essential components of cells to form various types of modifications by inducing non-specific modifications that lead to the damages of the targets – DNAs, RNAs, and Proteins. However, recent studies with increasing number of publications have demonstrated a role of reactive metabolites as a signaling messenger involving in the regulation of target protein functions with high specificity and modulating the downstream signaling cascades. Likewise, many literatures have implicated MGx in the pathogenesis of diseases such as diabetes and aging, yet the discrete molecular targets of MGx in these contexts are unknown.

Our studies revealed that MGx, which was previously considered as a toxic metabolite inducing non-specific glycations of the cellular components, is a signaling molecule that directly deliver the information of altered glycolytic flux to the KEAP1-NRF2 signaling axis, adding on

additional layer of regulation to both pathways and highlights a conduit for regulation of global metabolic status. First, beyond providing a source of ATP and biosynthetic precursors, glycolysis a major regulator of cellular redox status through production of NADPH and reduced glutathione. These reducing equivalents are primarily responsible for the regulation of a wide range of reactive species in cells and when these stress levels exceed the capacity of the redox regulation by glycolysis, KEAP1-NRF2 pathway is poised to respond and limit cellular damage. Recently published studies have also implicated that targets of NRF2 transcriptional program include the glutathione synthesis[173] and GLO1 transcription which are directly involved in glyoxalase program to detoxify the cellular MGx levels, as well as metabolic enzymes that involved in the Pentose Phosphate Pathway (PPP)[174] to redirect the glucose carbon away from central metabolites (e.g. MGx). These observations indicate that direct coupling of glucose metabolism with KEAP1 function through MGx creates an intrinsic feedback loop to sense and respond to changing metabolic demands in the cell. This feedback loop may also explain the observed pulsatile nature of central metabolite elevation and cell tolerance in response to direct PGK1 inhibition, as elevation of methylglyoxal leads to increased NRF2 activation of protein targets that catalyze its depletion. Our study reinforces the integrated role of cellular reducing equivalents (e.g. GSH) as a simultaneous regulator of both ROS and reactive metabolites to respond to and limit biomolecular damage.

Besides the elucidation of MGx targets and corresponding protein communication networks, future studies will focus on determine whether MGx, through MICA or other modifications, as well as other reactive metabolites, participate in similar mechanisms by targeting other proteins and pathways. Additionally, the strong associations between advanced glycated end products (AEGs) and pathology in diseases such as neurodegeneration, diabetes, aging, cancer and other

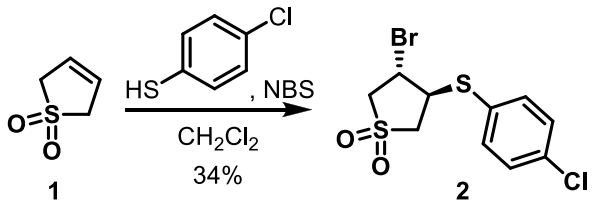
represent areas of study to interrogate whether aberrant modification of KEAP1 or other unknown MICA targets, may represent a significant portion of these poorly characterized modifications. Indeed, given the overwhelming evidence connecting altered glucose metabolism and NRF2 signaling to the pathogenesis of disease, our work should inspire future exploration of the role of this mechanism, as well as other reactive metabolite-mediated signals, as causes and therapeutic targets in human disease.

2.4. Materials and Methods

Chemicals

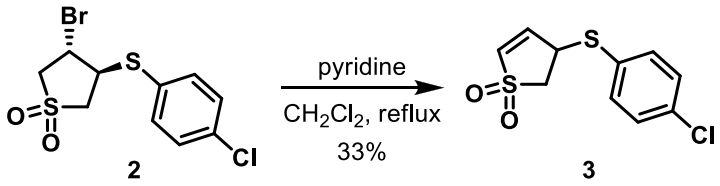
TBHQ, 2DG, MGx and GSH were obtained from Sigma Aldrich. The synthesis of AI-1 has been described previously[175]. GDC-0941 and Go 6976 were from Selleck. The GLO1 inhibitor (CAS No. 174568-92-4) was from MedChemExpress. CBR-470-0 and CBR-581-9 were from ChemDiv. CBR-470-1 (initially from ChemDiv as D470-2172) and related analogues were synthesized in house according to the below methods. All commercially obtained chemicals were dissolved in DMSO and used without further purification with the exception of 2DG, MGx and GSH, which were delivered as aqueous solutions.

Synthesis of CBR-470-1

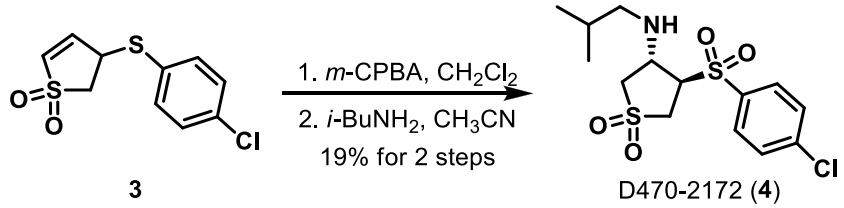


To a stirred solution of *N*-bromosuccinimide (3.15 g, 17.7 mmol) in CH_2Cl_2 (25 mL) was added dropwise a solution of 4-chlorothiophenol (2.56 g, 17.7 mmol). After stirred for 30 min, to the reaction mixture was added dropwise a solution of 3-sulfolene **1** (2.09 g, 17.7 mmol). After stirring for 2 h, the reaction mixture was quenched with H_2O. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (EtOAc : *n*-hexane = 1 : 4) to afford 2.04 g (34%) of **2** as white solid: ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.45 (d, 2H, J = 8.5 Hz), 7.37 (d, 2H, J = 8.5 Hz), 4.31 (q, 1H, J = 7.0 Hz), 4.21 (q, 1H, J = 7.3 Hz), 3.93 (q, 1H, J =

7.0 Hz), 3.70 (q, 1H, *J* = 7.1 Hz), 3.49 (q, 1H, *J* = 6.9 Hz), 3.12 (q, 1H, *J* = 7.1 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 135.9, 135.5, 130.0, 128.5, 59.3, 55.8, 51.7, 43.5; LR-MS (ESI+) *m*/*z* 340 ($M + H^+$); HR-MS (ESI+) calculated for $C_{10}H_{11}BrClO_2S_2$ ($M + H^+$) 340.9067; found 340.9068.

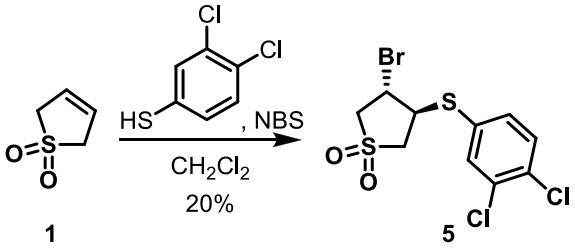


To a stirred solution of **2** (2.04 g, 5.94 mmol) in CH_2Cl_2 (15 mL) was added pyridine (1.20 mL, 14.9 mmol). After stirring for 3 h at 70 °C, the reaction mixture was cooled to room temperature and quenched with saturated *aq.* NH_4Cl. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (EtOAc : *n*-hexane = 1 : 3) to afford 513 mg (33%) of **3** as white solid: ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.39 (d, 2H, *J* = 8.6 Hz), 7.33 (d, 2H, *J* = 8.6 Hz), 6.67 (m, 2H), 4.43 (m, 1H), 3.63 (q, 1H, *J* = 7.4 Hz), 3.19 (q, 1H, *J* = 6.2 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 139.3, 135.9, 135.5, 133.2, 129.9, 128.9, 54.3, 45.0; LR-MS (ESI+) *m*/*z* 260 ($M + H^+$); HR-MS (ESI+) calculated for $C_{10}H_{10}ClO_2S_2$ ($M + H^+$) 260.9805; found 260.9807.



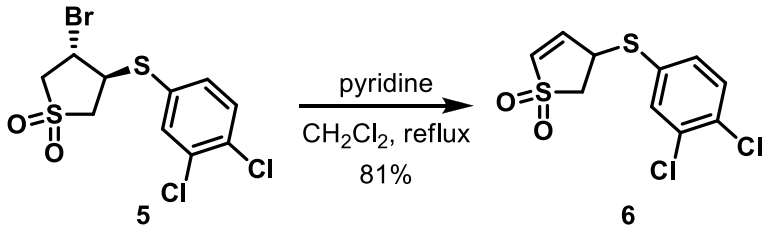
To a stirred solution of **3** (6.01 g, 23.0 mmol) in CH_2Cl_2 (100 mL) was added *m*-CPBA (14.2 g, 57.5 mmol, 70~75%). After stirring for 3 h, the reaction mixture was filtered and quenched with saturated *aq.* $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The crude mixture was used for next step without further purification. To a stirred solution of the sulfone in CH_3CN (50 mL) was added *i*-butylamine (2.29 mL, 23.0 mmol). After stirring for 2 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 1.60 g (19%) of CBR-470-1 (D470-2172, **4**) as white solid: ^{1}H NMR (DMSO-d_6, 400 MHz) δ 7.97 (d, 2H, J = 8.6 Hz), 7.76 (d, 2H, J = 8.6 Hz), 4.30 (m, 1H), 3.64 (m, 3H), 3.37 (m, 1H), 3.13 (q, 1H, J = 6.9 Hz), 2.16 (m, 1H), 2.05 (m, 1H), 1.87 (q, 1H, J = 7.5 Hz), 1.33 (m, 1H), 0.69 (d, 3H, J = 2.4 Hz), 0.67 (d, 3H, J = 2.5 Hz); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 139.6, 136.6, 130.7, 129.5, 63.5, 55.2, 55.1, 54.4, 49.5, 27.9, 20.3; LR-MS (ESI+) *m/z* 366 (M + H^+); HR-MS (ESI+) calculated for $C_{14}H_{21}ClNO_4S_2$ (M + H^+) 366.0595; found 366.0595.

Synthesis of CBR-470-2 and related analogues

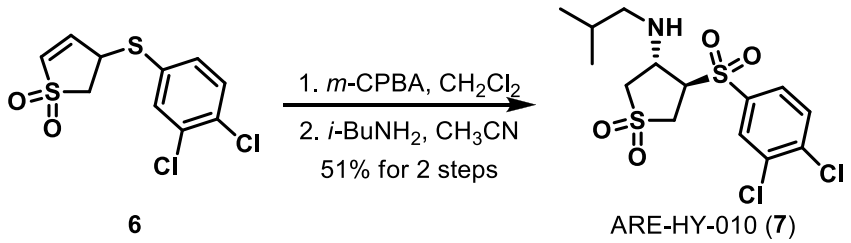


To a stirred solution of *N*-bromosuccinimide (2.26 g, 12.7 mmol) in CH_2Cl_2 (20 mL) was added dropwise a solution of 3,4-dichlorobenzenethiol (1.62 mL, 12.7 mmol). After stirred for 30 min, to the reaction mixture was added dropwise a solution of 3-sulfolene **1** (1.50 g, 12.7 mmol). After stirring for 2 h, the reaction mixture was quenched with H_2O. The aqueous layer was

extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (EtOAc : *n*-hexane = 1 : 5) to afford 974 mg (20%) of **5** as white solid: ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.59 (d, 1H, *J* = 2.1 Hz), 7.46 (d, 1H, *J* = 8.3 Hz), 7.33 (dd, 1H, *J* = 8.3, 2.1 Hz), 4.34 (q, 1H, *J* = 7.0 Hz), 4.03 (q, 1H, *J* = 7.3 Hz), 3.93 (q, 1H, *J* = 7.1 Hz), 3.73 (q, 1H, *J* = 7.1 Hz), 3.50 (q, 1H, *J* = 7.0 Hz), 3.14 (q, 1H, *J* = 7.1 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 135.3, 134.2, 133.8, 133.0, 131.6, 130.4, 59.4, 55.9, 51.9, 43.4; LR-MS (ESI+) *m/z* 374 (M + H^+); HR-MS (ESI+) calculated for $C_{10}H_{10}BrCl_2O_2S_2$ (M + H^+) 374.8677; found 374.8679.

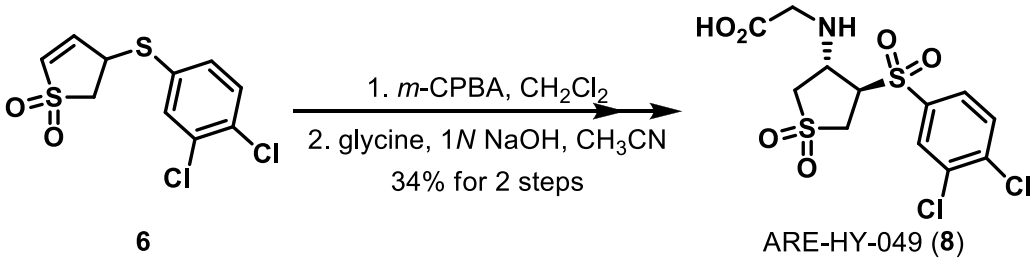


To a stirred solution of **5** (487 mg, 1.29 mmol) in CH_2Cl_2 (10 mL) was added pyridine (0.260 mL, 3.23 mmol). After stirring for 1 h at 70 °C, the reaction mixture was cooled to rt and quenched with saturated *aq.* NH_4Cl. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (EtOAc : *n*-hexane = 1 : 2) to afford 310 mg (81%) of **6** as white solid: ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.55 (d, 1H, *J* = 2.1 Hz), 7.44 (d, 1H, *J* = 8.3 Hz), 7.28 (dd, 1H, *J* = 8.4, 2.2 Hz), 6.70 (m, 2H), 4.47 (m, 1H), 3.66 (dd, 1H, *J* = 14.1, 8.2 Hz), 3.21 (dd, 1H, *J* = 14.1, 4.5 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 138.8, 135.2, 134.1, 133.7, 133.6, 132.9, 131.4, 130.7, 54.4, 45.0; LR-MS (ESI+) *m/z* 294 (M + H^+); HR-MS (ESI+) calculated for $C_{10}H_9Cl_2O_2S_2$ (M + H^+) 294.9416; found 294.9416.



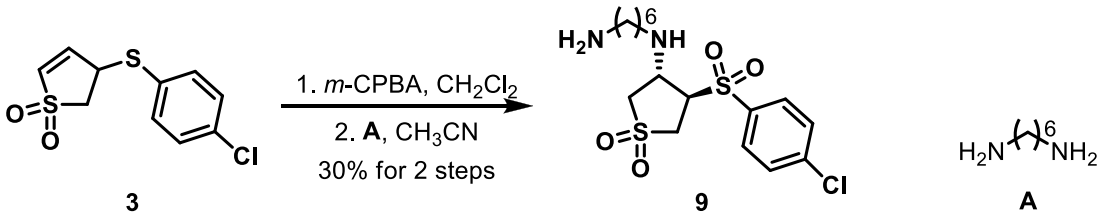
6 ARE-HY-010 (**7**)

To a stirred solution of **6** (310 mg, 1.05 mmol) in CH_2Cl_2 (10 mL) was added *m*-CPBA (648 mg, 2.63 mmol, 70~75%). After stirring for 1 h, the reaction mixture was filtered and quenched with saturated *aq.* $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The crude mixture was used for next step without further purification. To a stirred solution of the sulfone in CH_3CN (10 mL) was added *i*-butylamine (0.104 mL, 1.05 mmol). After stirring for 1 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 215 g (51%) of ARE-HY-010 (**7**) as white solid: ^{1}H NMR (DMSO-d_6, 400 MHz) δ 8.21 (d, 1H, J = 2.0 Hz), 7.96 (d, 2H, J = 8.4 Hz), 7.90 (dd, 1H, J = 8.4, 2.1 Hz), 4.38 (m, 1H), 3.66 (m, 3H), 3.42 (dd, 1H, J = 14.1, 9.0 Hz), 3.12 (q, 1H, J = 6.1 Hz), 2.17 (m, 1H), 2.03 (m, 1H), 1.87 (m, 1H), 1.31 (m, 1H), 0.67 (d, 3H, J = 3.9 Hz), 0.65 (d, 3H, J = 3.9 Hz); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 138.5, 137.6, 132.4, 131.6, 130.6, 128.7, 63.5, 55.4, 55.1, 54.4, 49.4, 28.0, 20.3; LR-MS (ESI+) m/z 400 (M + H^+); HR-MS (ESI+) calculated for $C_{14}H_{20}Cl_2NO_4S_2$ (M + H^+) 400.0205; found 400.0204.

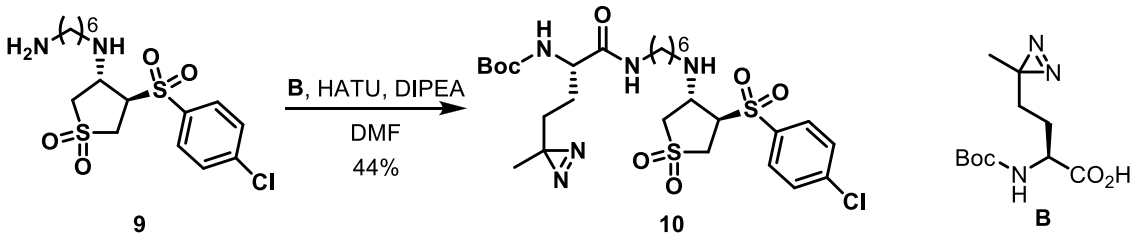


To a stirred solution of **6** (1.05 g, 3.54 mmol) in CH_2Cl_2 (20 mL) was added *m*-CPBA (2.18 mg, 8.85 mmol, 70~75%). After stirring for 2 h, the reaction mixture was filtered and quenched with saturated *aq.* $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The crude mixture was used for next step without further purification. To a stirred solution of the sulfone in CH_3CN (20 mL) were added glycine (0.104 mL, 1.05 mmol) and 1*N* NaOH (2.0 mL). After stirring for 1 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 483 mg (34%) of ARE-HY-049 (CBR-470-2) (**8**) as white solid: ^{1}H NMR (DMSO-d_6, 400 MHz) δ 8.24 (s, 1H), 7.93 (s, 1H), 7.89 (m, 2H), 7.71 (d, 1H, J = 8.9 Hz), 7.55 (t, 1H, J = 5.4 Hz), 4.48 (q, 1H, J = 8.2 Hz), 3.89 (q, 1H, J = 6.9 Hz), 3.62 (m, 2H), 3.41 (dd, 1H, J = 14.6, 8.2 Hz), 2.22 (m, 1H); LR-MS (ESI+) *m*/*z* 401 (M + H^+); HR-MS (ESI+) calculated for $C_{12}H_{14}Cl_2NO_6S_2$ (M + H^+) 401.9634; found 401.9637.

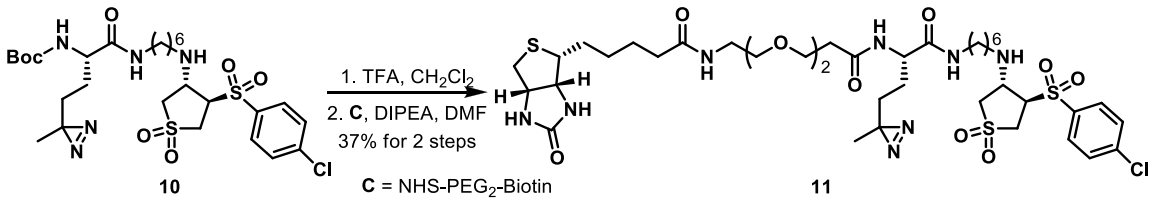
Synthesis of CBR-470-PAP



To a stirred solution of **3** (76.3 mg, 0.293 mmol) in CH_2Cl_2 (5 mL) was added *m*-CPBA (288 mg, 1.17 mmol, 70~75%). After stirring for 1 h, the reaction mixture was filtered and quenched with saturated *aq.* $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2 and the combined organic layers were dried over $MgSO_4$ and concentrated *in vacuo*. The crude mixture was used for next step without further purification. To a stirred solution of the sulfone in CH_3CN (5 mL) was added 1,6-hexanediamine **A** (0.115 mL, 0.879 mmol). After stirring for 1 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 35.9 mg (30%) of **9** as pale yellow solid: LR-MS (ESI+) *m*/*z* 409 ($M + H^+$).



To a stirred solution of **9** (17.5 mg, 42.8 μmol) and **B** (11.0 mg, 42.8 μmol) in DMF (2 mL) were added HATU (32.5 mg, 85.6 μmol) and DIPEA (29.8 μL, 0.171 mmol). After stirring for 24 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 12.2 mg (44%) of **10** as white solid: LR-MS (ESI+) *m*/*z* 648 ($M + H^+$).



To a solution of **10** (5.2 mg, 6.98 μmol) in CH_2Cl_2 (1 mL) was added TFA (0.5 mL). After stirring for 6 h, the residue was concentrated *in vacuo* and used for next step without further purification. To a stirred solution of the amine in DMF (1 mL) were added NHS-PEG_2-Biotin **C** (3.49 mg, 6.98 μmol) and DIPEA (4.86 μL, 27.9 μmol). After stirring for 6 h, the reaction mixture was concentrated *in vacuo*. The residue was purified by preparative HPLC and lyophilized to afford 2.41 mg (37%) of **11** as white solid: LR-MS (ESI+) *m*/*z* 933 ($M + H^+$).

Cell culture

IMR32, SH-SY5Y, HeLa, and HEK293T cell lines were purchased from ATCC. Human lung fibroblasts (HLF) and mouse dermal fibroblasts (MDFs, C57BL/6-derived) were obtained from Sciencell and used before passage 3. IMR32, HLF, SH-SY5Y, HeLa, and HEK293T cells were propagated in DMEM (Corning) supplemented with 10% fetal bovine serum (FBS, Corning) and Anti-anti (Gibco). MDFs were propagated in fibroblast medium 2 from Sciencell. Mouse epidermal keratinocytes (MPEK-BL6) were obtained from Zen Bio and propagated in epidermal keratinocyte medium (Zen Bio).

High throughput screening and ARE-LUC reporter assay

For high throughput screening, IMR32 cells were plated at 5 x 10^3 cells per well in white 384-well plates in 40 μL of growth medium. The next day 100 ng of pTI-ARE-LUC reporter

plasmid in 10 μL of Optimem medium (Gibco) was transfected into each well using Fugene HD at a dilution of 1 μg plasmid DNA: 4 μL of Fugene. 24 hours later, compounds were transferred using a 100 nL pintool head affixed to PerkinElmer FX instrument such that the final screening concentration was 2 μM. After 24 hours incubation, ARE-LUC luminance values were recorded on an Envision instrument after the addition of 30 μL of Bright Glo reagent solution (Promega, diluted 1:3 in water). Compounds which increased ARE-LUC signal greater than 4 Z-scores from plate mean were deemed hits. For overexpression and knockdown experiments in HEK293T with ARE-LUC reporter readouts, 5 x 10^5 cells were plated on poly-d-lysine coated plates and transfected with 1.5 μg of overexpression or shRNA plasmid and 500 ng of pTI-ARE-LUC using Optimem medium and Fugene in the same mode as above. 24 hours later, 10^3 transfected cells were plated in 50 μL of growth medium in white 96-well plates. After a 24 hours incubation, an additional 50 μL of growth medium with compound at the indicated concentrated was added to each well. ARE-LUC luminance values were recorded on an Envision plate reader 24 hours later by the addition of 75 μL of Bright Glo reagent solution (1:3 in water).

Peroxide stress model

10^4 SH-SY5Y cells were plated in 100 μL of growth medium in white 96-well plates. After 48 hours of compound treatment, 20 μL of tert-butyl peroxide diluted to the indicated concentrations was added to each well. After 8 hours incubation, cell viability measurements were recorded on an Envision plate reader after the addition of 50 μL of a Cell Titer Glo solution (Promega, diluted 1:6 in water). Relative viabilities are reported as a fraction relative to the same dose of compound treatment without TBHP.

shRNA knockdown studies

PGK1-targeting shRNA vectors sh10 and sh47 refer to Sigma Mission shRNA lentiviral clones NM_000291.2-338s1c1 and NM_000291.2-935s1c1 respectively. GLO1-targeting shRNA vectors sh29 and sh30 refer to Sigma Mission shRNA lentiviral clones NM_006708.1-195s1c1 and NM_006708.1-292s1c1 respectively. The non-targeting scrambled control vector refers to SHC002 (Sigma). Lentiviruses were generated in HEK293T cells by transient expression of the above vectors with pSPAX2 and pMD2.G packaging vectors (Addgene plasmids 11260 and 12259). Viral supernatants were collected after 48 hours of expression and passed through a 70 μm syringe filter before exposure to target cells.

Quantitative RT-PCR

Cells were collected by trypsinization and subsequent centrifugation at 500*g*. RNA was isolated using RNeasy kits from Qiagen and concentrations obtained using a NanoDrop instrument. 500ng-5μg of RNA was then reverse transcribed with oligo dT DNA primers using a SuperScript III First-Strand Synthesis kit from Invitrogen. Quantitative RT-PCR reactions were measured on a Viia 7 Real-Time PCR system (Thermo) using a Clontech SYBR green-based master mix. Gene specific primer sets are provided in **Table 2.1**. Reactions were normalized to TUBG1 levels for each biological replicate and relative transcript abundance calculated using the comparative C_t method.

Gene set enrichment analyses (GSEA)

Total RNA was extracted from IMR32 cells treated for 24 hours with either DMSO or 5 μM CBR-470-1 (3 biological replicates per condition) using a RNeasy kit (Qiagen). RNA-seq

experiments were performed by the Scripps Next Generation Sequencing Core according to established in house methods. Gene set enrichment analyses and leading edge heatmaps were generated with TPM values from the above experiment using the java GSEA package. "NFE2L2 targets" gene set refers to Molecular Signature Database (http://software.broadinstitute.org/gsea/msigdb) gene set ID M2662.

Quantitative metabolomic profiling

For polar metabolite profiling experiments, cells were grown in 15 cm plates and cultured in RPMI supplemented with 10% FBS, 2 mM L-glutamine and 1% P/S prior to media replacement containing either vehicle (DMSO) or the indicated dose of CBR-470-1. Following incubation for the appropriate time, cells were scraped into ice-cold PBS and isolated by centrifugation at 1,400*g* at 4°C. Cell pellets were resuspended in 300 µl of an 80:20 mixture of cold MeOH/H_2O, an internal standard was added (10 nmol d_3-serine; Sigma Aldrich), and the suspension was sonicated (Fisher Scientific FB-505) for 5 seconds followed by a 10 minute centrifugation at 16,000*g*. The supernatant was collected, dried under N_2 gas and resulting dried metabolites resuspended in 30 µl of 40% MeOH/H_2O for analysis on an Agilent triple quadrupole LC-MS/MS (Agilent Technologies 6460 QQQ). For negative mode operation, metabolites were separated by hydrophilic interaction chromatography with a Luna-NH_2 column (Phenomenex) running mobile phase A (CH_3CN supplemented with 0.2% NH_4OH) and B (95:5 v/v H_2O:CH_3CN supplemented with 50 mM NH_4OAc and 0.2% NH_4OH) and the following gradient: 0% B for 3 min; linear increase to 100% B for 27 min at a flow rate of 0.4 ml/min, followed by an isocratic flow of 100% B for 3 min. The spectrometer settings were: capillary voltage = 4.0 kV, drying gas temperature = 350°C at 10 L/min, and the nebulizer pressure was 45 psi. Metabolite peak transitions and retention

times are listed in **Table 2.2** and were confirmed by running standards for measured glycolytic, PPP, CAC and co-factor metabolites. 2-phosphoglycerate and 3-phosphoglycerate isomers were quantified in aggregate (2PG/3PG). Relative metabolite abundance was quantified by integrated peak area for the given MRM-transition, and all metabolite levels were normalized to internal standard extracted ion intensity values for d_3-serine. These parameters were used to quantify all metabolites, with the exception of 1,3-BPG and MGx, which required chemical derivitization to stable intermediates prior to LC-MS/MS quantification, as previously reported[98,159,165]. MGx deviated from all other metabolites, as it was separated on a Gemini reverse-phase C18 column (5 mm, 4.6 mm x 50 mm; Phenomenex) together with a pre-column (C18, 3.5 mm, 2 mm x 20 mm) and detected in positive mode analysis, with mobile phase A (H_2O) and B (50:50 v/v H_2O:CH_3CN) supplemented with 0.1% TFA. The gradient started with 0% B for 2 min and increased linearly to 100% B over 10 min with a flow rate of 0.4 ml/min, followed by an isocratic gradient of 100% B for 5 min at 0.4 ml/min. The QQQ settings were the same as above.

FLAG-tagged protein expression and Western blot

Full-length, human PGK1 (NM_000291, Origene) transiently expressed from a pCMV6 entry vector with a C-terminal Myc-DDK tag; full-length, human KEAP1 (28023, Addgene) was transiently expressed from a pcDNA/FRT/TO plasmid with a C-terminal 3xFLAG tag. All references to FLAG-PGK1 or FLAG-KEAP1 represent the proteins in the aforementioned vectors, respectively. Transient protein expression was performed in confluent 10 cm plates of HEK293T cells by transfection of 1 μg plasmid with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. For *in situ* compound or metabolite treatment experiments, compounds were added approximately 24 hours after transfection and incubated for the indicated duration. For

FLAG-KEAP1 western blotting and immunoprecipitation experiments, cells were harvested by scraping, pelleted by centrifugation, washed twice with PBS and lysed in 8 M urea, 50 mM NH_4HCO_3, phosphatase inhibitor cocktail (Sigma Aldrich), and EDTA-free complete protease inhibitor (Roche), pH 8.0, at 4 ˚C. Cells were sonicated (Fisher Scientific FB-505), insoluble debris cleared by centrifugation, and the supernatant was diluted into 4X Laemmli buffer containing 50 mM dithiothreitol (DTT) as a reducing agent. Samples were prepared for SDS-PAGE by heating to 95 ˚C for 5 minutes, cooled to room temperature, resolved on NuPAGE Novex 4-12% Bis-Tris Protein Gels (Invitrogen), and transferred onto nitrocellulose membranes by standard western blotting methods. Membranes were blocked in 2% BSA in TBS containing 0.1% tween-20 (TBST) and probed with primary and secondary antibodies. Primary antibodies used in this study include: anti-FLAG-M2 (1:1000, F1804, Sigma Aldrich), anti-KEAP1 (1:500, SC-15246, Santa Cruz), anti-HSPA1A (1:1000, 4872, Cell Signaling), anti-ACTB (1:1000, 4790, Cell Signaling), anti-GAPDH (1:1000, 2118S, Cell Signaling) and TUBG (1:1000, 5886, Cell Signaling). Rabbit polyclonal anti-pgK antibody was generated using pgK-modified KLH and affinity purification as described[98] at a 1:400 dilution of a 0.33 mg/mL stock in 10mM sodium HEPES (pH 7.5), 150mM NaCl, 30% glycerol and 0.02% sodium azide. Secondary donkey anti-rabbit, donkey anti-goat, and donkey anti-mouse (Licor), were used at 1:10,000 dilution in 2% BSA-containing TBST and incubated for 1 hour prior to washing and imaging on a Licor infrared scanner. Densitometry measurements were performed with ImageJ software.

Time- and dose-dependent CBR-470-1 treatment studies were performed in HEK293T cells 24 hours after transient transfection of FLAG-KEAP1, or in IMR32 cells for endogenous KEAP1. Fresh RPMI media with 10% FBS, 2 mM L-glutamine, 1% P/S and the indicated concentration of CBR470-1 (20 μM for time-dependent experiments) or equivalent DMSO was

added to cells in 10 cm dishes. Following the indicated incubation time cells were lysed in lysis buffer [50 mM Tris, 150 mM NaCl, 1% Triton-X 100, phosphatase inhibitor cocktail (Sigma Aldrich), and EDTA-free complete protease inhibitor (Roche), pH 7.4] and processed for western blot as indicated above.

Target identification studies with CBR-470-PAP

10 cm dishes of confluent IMR32 cells were exposed to 5 μM CBR-470-PAP with the addition of either DMSO or a 50-fold molar excess of CBR-470-1 (250 μM) for 1 hour at 37 °C. Samples were then UV crosslinked using a Stratalinker 2400 instrument for 10 minutes. RIPA extracted lysates were then fractionated with ammonium sulfate with percent increments of 20. These fractions were then separated via SDS-PAGE and relevant probe-labeling was determined by anti-biotin (1:500, ab1227, Abcam) western blotting as above. A parallel gel was silver stained using the Pierce silver stain kit. Relevant gel slices from the 80 percent fraction were excised and PGK1 identity was determined by LC-MS/MS by the Scripps Center for Metabolomics and Mass Spectrometry. Follow up shRNA knock down studies confirmed PGK1 as the target within this fraction.

Dye-based thermal denaturation assay

Thermal denaturation experiments were performed using a Protein Thermal Shift Dye Kit (ThermoFisher, 4461146). Reactions contained 2 μM recombinant PGK1 with the indicated dose of aqueously-delivered CBR-470-1 with 1x supplied thermal shift dye and reaction buffer in 20uL reaction volumes. Fluorescence values were recorded using a Viia7 Real-Time PCR instrument according to supplied instructions.

Recombinant PGK1 assay

PGK1 enzymatic activity in the forward direction was measured with a coupled enzymatic assay[157]. Three PGK1 conditions were prepared by dissolving recombinant PGK1 in potassium phosphate buffer (10 mM KH_2PO_4, 10 mM $MgSO_4$, pH 7.0), and transferring the aliquots of PGK1 solution to the microtubes being treated with same amount of DMSO and indicated concentrations of CBR-470-1. Final concentration of PGK1 is 20 ng/mL and DMSO is 1% for each sample. Two blank conditions, 0 μM and 100 μM of CBR-470-1 with no PGK1, were also prepared for the control measurements. All PGK1 samples and blank samples were pre-incubated for 20 minutes and then transferred to the UV-transparent 96 well plate (Corning). The assay solution (10 mM KH_2PO_4, 2 mM G3P, 0.6 mM NAD^+, 200 mM Glycine, 0.4 mM ADP, pH 7.0) was activated by adding GAPDH with 10 μg/mL final concentration, and then the assay solution was added to the wells containing PGK1 samples and blank samples. The change in absorbance at 340 nm at room temperature was measured every 20 seconds for 45 minutes, by Tecan Infinite M200 plate reader. Each condition was performed with three independent replications.

Isothermal dose response profiling of PGK1

In vitro thermal profiling assay for recombinant proteins was performed by dissolving pure recombinant PGK1 and GAPDH into PBS and dividing equal amount of mixture into 9 aliquots. Each aliquot was transferred to 0.2 mL PCR microtubes being treated with different amounts of CBR-470-1 added from DMSO stock, and equal amount of DMSO for the control. Each microtube contains 50 μL of mixture with final concentration of 45 μg/mL for each protein and DMSO

concentration 1% with following final concentrations of CBR-470-1; 0 μM, 0.1 μM, 0.3 μM, 1 μM, 3 μM, 10 μM, 33 μM, 100 μM, 333 μM. After 30 minutes incubation at 25°C, samples were heated at 57°C for 3 minutes followed by cooling at 25°C for 3 minutes using Thermal Cycler. The heated samples were centrifuged at 17,000g for 20 minutes at 4°C, and the supernatants were transferred to new Eppendorf tubes. Control experiments were performed with heating at 25°C for 3 minutes, instead of 57°C. Samples were analyzed by SDS-PAGE and Western blot.

Metabolite treatments and HMW-KEAP1 screening

For *in vitro* screening of glycolytic metabolites, HEK293T cells expressing FLAG-KEAP1 were lysed by snap-freeze-thaw cycles (3x) in PBS, pH 7.4, containing EDTA-free complete protease inhibitor (Roche). Lysates were cleared by centrifugation and the supernatants normalized for concentration by Bradford reagent (2 mg/mL). Concentrated stocks of each metabolite were made in PBS, which were added to the lysate samples for the final indicated concentrations and incubated at 37˚C for 2.5 hours with shaking. Following incubation, samples were denatured with 6 M urea and processed for SDS-PAGE and western blotting. Methylglyoxal (40% v/v with H_2O), glyceraldehyde 3-phosphate (GAP), dihydroxyacetone phosphate (DHAP), and 2,3-bisphosphoglycerate (2,3-BPG) were all obtained from Sigma Aldrich and used as PBS stocks. *In situ* metabolite treatments were performed in HEK293T cells 24 hours after transfection of FLAG-KEAP1, treated with MGx (1 or 5 mM) in H_2O or equivalent vehicle alone for 8 hours. Cells were collected by scraping, washed in PBS and centrifuged, and lysed in urea lysis buffer and analysis by SDS-PAGE and western blot.

Dose-response experiments were performed with high purity MGx prepared by acidic hydrolysis of MG-1,1-dimethylacetal (Sigma Aldrich) followed by fractional distillation under

reduced pressure and colorimetric calibration of the distillates, as previously reported. For *in vitro* MGx dose response dimerization of KEAP1, HEK293T cells expressing FLAG-KEAP1 were lysed in PBS as indicated above, then serial dilutions of high purity MGx in 50 mM Sodium Phosphate, pH 7.4, were added to the equal volume of lysate aliquots with final protein concentration of 1 mg/mL. Each mixture was incubated at 37°C for 8 hours with rotating, HMW-KEAP1 formation was analyzed by SDS-PAGE and western blot.

Site-directed mutagenesis of KEAP1

KEAP1 mutants were generated with PCR primers in **Table 2.3** according to the Phusion site-directed mutagenesis kit protocol (F-541, Thermo Scientific) and the QuikChange site-directed mutagenesis kit protocol (200523, Agilent). Mutant KEAP1 plasmids were verified by sequencing [CMV (forward), wild-type primers in the middle of KEAP1 sequence (forward) and BGH (reverse)] and were transiently expressed in HEK293T cells in the same manner as wild type KEAP1. Screening of CBR-470-1-induced HMW-KEAP1 formation with mutant constructs was performed just as with wild type KEAP1, after 8 hours CBR-470-1 treatment (20 μM). Following treatment, cells were harvested and prepared for SDS-PAGE and western blotting as indicated above.

SILAC cell culture methods and proteomic sample preparation

SILAC labeling was performed by growing cells for at least five passages in lysine- and arginine-free SILAC medium (RPMI, Invitrogen) supplemented with 10% dialyzed fetal calf serum, 2 mM L-glutamine and 1% P/S. “Light” and “heavy” media were supplemented with

natural lysine and arginine (0.1 mg/mL), and ^{13}C-, ^{15}N-labeled lysine and arginine (0.1 mg/mL), respectively.

General protein digestion for LC-MS/MS analysis was performed by diluting protein (e.g. whole lysate or enriched proteins) in digestion buffer (8 M urea, 50 mM NH_4HCO_3, pH 8.0), followed by disulfide reduction with DTT (10 mM, 40 minutes, 50 ˚C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Promega) at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted on ZipTip C18 tips (100 µL, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized. Lyophilized peptides were dissolved in LC-MS/MS Buffer A (H_2O with 0.1% formic acid, LC-MS grade, Sigma Aldrich) for proteomic analysis.

CETSA proteomic sample preparation

In situ Cellular thermal shift assay with SILAC quantitative proteomics was performed with “heavy” and “light” labeled Hela cells. SILAC-labeled HeLa cells, grown to 80-90% confluency in 2 of 10 cm cell-culture treated plates (Denville) each, were treated with DMSO alone (2 of 10 cm light cell plates and 1 of 10 cm heavy cell plate) or CBR-470-1 (20 µM, 1 of 10 cm heavy cell plate) for 1 hour. After treatment, heavy and light cells in each plate were gently washed with DPBS, detached with 0.05% trypsin-EDTA, collected by centrifugation (350 G, 2 minutes), and resuspended in 1.05 mL DPBS with EDTA-free protease inhibitor tablet (Roche) before being

divided into 10 equal aliquots in 0.2 mL PCR microtubes (100 μL each). After centrifugation (300 G, 2 minutes) followed by taking out 80 μM of supernatant and resuspending cell pellets in DPBS leftover (20 μM), each aliquot of heavy cells was exposed to a steady temperature between 37 °C to 67 °C for 3 minutes in parallel and then incubated at 25 °C for another 3 minutes using Thermal Cycler. As a comparison, each aliquot of light cells was exposed to 37 °C for 3 minutes in parallel and then incubated at 25 °C for another 3 minutes. Then, all cells were transferred to 1.6 mL Eppendorf tubes, lysed by 3 cycles of rapid freeze-thawing, and insoluble proteins and cell debris were removed by centrifugation at 17,000 G for 10 minutes at 4 °C. The amount of soluble proteome remained in each aliquot of heavy cell lysates were internally normalized by the addition of same volume light cell lysates. These samples were then denatured in 8 M urea and prepared for proteomic analysis by following the proteomic sample preparation protocol described above.

Proteomic LC-MS/MS and data analysis

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra high-pressure LC system (ThermoFisher) using a PepMap RSLC C18 column heated to 40˚C (column: 75 μm x 15 cm; 3 μm, 100 Å) coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). Digested peptides (500 ng) in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% Formic acid acetonitrile) at 300 nL/min: 0-2% buffer B over 10 minutes, 2-40% buffer B over 120 minutes, 40-70% buffer B over 10 minutes, and 70-100% buffer B over 5 minutes. MS/MS spectra were collected from 0 to 150 minutes using a data-dependent, top-20 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 400-1600 m/z, maximum IT of 50 ms, AGC target of 1e6, and data collection in profile mode. MS2 scans was performed by HCD

fragmentation with a resolution of 15,000, AGC target of 1e5, maximum IT of 30 ms, NCE of 26, and data type in centroid mode. Isolation window for precursor ions was set to 1.5 m/*z* with an underfill ratio of 0.5%. Peptides with charge state >5, 1 and undefined were excluded and dynamic exclusion was set to eight seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.60kV and ionization chamber temperature of 300 ˚C.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database (UniProt_Human_reviewed_04-10-2017.fasta). Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation); two total differential modification sites per peptide, including oxidized methionine (+15.9949); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting statistics were performed with a Δmass cutoff = 15 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%, peptide modification requirement (-m) was set to 1, and spectra display mode (-t) was set to 1. SILAC searchers were performed as above with “light” and “heavy” database searches of MS1 and MS2 files by including static modification of +8.014168 for lysine and +10.0083 for arginine in a parallel heavy search. SILAC quantification was performed using the QuantCompare algorithm, with a mass tolerance of 10 p.p.m. or less in

cases where co-eluting peptide interfere. In general all quantified peptides has mass error within 3 p.p.m..

CETSA melt curve analyses

SILAC data of CETSA samples were combined in response to DMSO- and CBR-470-1 treated cells. Relative amount of each proteins from heavy cells were normalized by the SILAC ratio of same proteins in the lowest temperature fraction. To generate protein melt curves, changes of heavy-to-light ratios as a function of temperature was fitted to the equation derived from the chemical denaturation theory using the R package developed by Savitski et al. T_m values were calculated at which the sigmoidal curve crosses the 0.5-fold change level. Only T_m values calculated from melting curves with curve $R^2 > 0.8$ are used in subsequent analyses. Shift in T_m values induced by CBR-470-1 *in situ* treatment were determined by subtracting the protein T_m value of DMSO treated cells from the protein T_m value of CBR-470-1 treated cells: $\Delta T_m = T_m$ (CBR-470-1) – T_m (DMSO)

Quantitative proteomic detection of potential KEAP1 modification sites

Quantitative surface mapping with SILAC quantitative proteomics was performed with "heavy" and "light" labeled HEK293T cells expressing FLAG-KEAP1. Cells were incubated with DMSO alone (light cells) or CBR-470-1 (20 μM, heavy cells) for 8 hours. After incubation cells were scraped, washed with PBS (3x) and combined prior to lysis in Urea lysis buffer [8 M Urea, 50 mM NH_4HCO_3, nicotinamide (1 mM), phosphatase inhibitor cocktail (Sigma Aldrich), and EDTA-free complete protease inhibitor (Roche), pH 8.0] by sonication at 4 ˚C. After sonication, insoluble debris was cleared by centrifugation (17,000 *g*, 10 min), diluted with Milli-Q water to

give 1 M urea, and lysate was incubated with Anti-FLAG M2 resin (100 µL slurry, A2220, Sigma Aldrich) at 4 ˚C overnight while rotating. For SILAC label-swap experiments, "light" HEK293T cells were incubated with CBR-470-1 and "heavy" cells were incubated with DMSO and processed as above. FLAG resin was washed with PBS (7x1 mL), FLAG-KEAP1 protein eluted with glycine-HCl buffer (0.1 M glycine, pH 3.5, 2x500 µL), followed by 8 M urea (2x100 µL). The combined eluent was brought up to 8 M urea total concentration and processed for trypsin digestion and LC-MS/MS analysis as indicated above.

The SILAC maps were generated by comparing SILAC ratios for each peptide, relative to the median value for all KEAP1 peptides. KEAP1 peptides are listed in **Table 2.4**. SILAC ratios were converted to Log_2 values and plotted to visualize peptides that are significantly perturbed, for example by modification, relative to the rest of the protein. A minimum of three SILAC ratios for each peptide was required for inclusion in KEAP1 surface maps, which allowed for ~85-90% coverage of the KEAP1 protein. Missing sequences were caused by the lack or close spacing of tryptic sites, resulting in inadequate peptides for MS/MS detection.

Co-enriched proteins with HMW-KEAP1 was analyzed from the same samples described above and identified by SILAC ratios. These proteins were considered as a possible heterodimerization partners.

In vitro MGx- peptide reactions

'CR' peptide was synthesized using standard solid phase peptide synthesis with FMOC-protected amino acids on MBHA rink amide resin. Peptides were cleaved in a solution of 94% trifluoroacetic acid, 2.5% triisopropyl silane, 2.5% H_2O, 1% β-mercaptoethanol (βME) and precipitated with ether. Peptide identity was confirmed using an Agilent 1100 series LC-MS.

Peptides were purified via reverse phase HPLC on an Agilent Zorbax SB-C18 250mm column and dried via lyophilization. For methylglyoxal reactions CR peptide (1 mM) was incubated with 12.5 mM methylglyoxal (diluted from 40% solution in water; Sigma Aldrich) or equivalent amount of water (mock) in 1x PBS pH 7.4 at 37°C overnight. Reactions were diluted 1:25 in 95/5 H_2O/Acetonitrile + 0.1% trifluoroacetic acid and analyzed by LC-MS.

For NMR experiments, approximately 1.5 mg of the CR or CR-MGx crosslinked peptide was purified by reverse phase HPLC, lyophilized and dissolved in 700 µL d_6-DMSO. The peptides were dried via lyophilization. All NMR experiments were performed on a Bruker Avance II+ 500 MHz 11.7 Tesla NMR. Data was processed and plotted in Bruker Topspin 3.5. CR peptide NMR experiments were run with a spectral width of 8.5 for 2D experiments (in both dimensions) and 15 for 1D proton NMR with a pulse width of 13.5 µs and an interscan delay of 3 s. For the proton NMR, 256 scans were taken. For the COSY-DQF experiment, 128 and 2048 complex points were acquired in the F1 and F2 dimensions respectively, with 8 scans per point. For the TOCSY experiment, a mixing time of 60 µs was used, and 256 and 1024 complex points were acquired with 8 scans per point. All CR-MGx peptide NMR experiments were run with a spectral width of 13 (in both dimensions) with a pulse width of 11.5 µs and an interscan delay of 2.2 s. For the proton NMR, 256 scans were taken. For the COSY-DQF experiment, 128 and 2048 complex points were acquired in the F1 and F2 dimensions respectively, with 8 scans per point. For the TOCSY experiment, a mixing time of 80 µs was used, and 256 and 1024 complex points were acquired with 8 scans per point.

In-gel digestion and double digestion of KEAP1

Targeted proteomic analyses of KEAP1 protein were performed by running anti-FLAG enriched HMW-KEAP1 and LMW-KEAP1 on SDS-PAGE gels, and isolated proteins in gel pieces were digested in-gel with sequencing grade trypsin (Promega), as previously reported[176].

Double digests of HMW-KEAP1 and LMW-KEAP1 were used for direct MS1 and MS2 analysis of CR-MGx crosslinked KEAP1 peptide. For preparation of double digests, peptides from in-gel tryptic digestions were dissolved in 100 mM Tris-HCl, pH 8.0, with 2 mM of $CaCl_2$, and further digested with mass spectrometry-grade chymotrypsin (Thermo Scientific) according to manufacturer's protocol. Chymotryptic digestion reactions were stopped by acidification, and double digests were desalted on Ziptip C18 tips as mentioned above.

MS1 and MS2 analysis of CR-MGx crosslinked KEAP1 peptide

CR-MGx crosslinked KEAP1 peptide was initially validated by manual inspection of chromatograms and MS1 spectra with theoretical m/z values of peptide candidates, followed by comparison of the extracted ion intensity in HMW-KEAP1 with the intensity of same ion in LMW-KEAP1; extracted MS1 ions of the candidates were present in HMW-KEAP1 digests but not in LMW-KEAP1 digests.

Validated MS1 ions were further analyzed by Parallel reaction monitoring (PRM) to identify CR-MGx crosslinked peptide with desired b- and y- ions. PRM experiments were performed with an Easy-nLC 1000 ultra high-pressure LC system coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource as indicated above. MS/MS spectra were collected from 0 to 160 minutes using a PRM setting with the following settings: Global and general settings included lock masses of off, chromatography peak width of 15 seconds, polarity of positive, in-source CID of 0.0 eV, inclusion of on, and a m/z value of validated MS1 ion with its charge state in the

inclusion list. MS2 scans were performed by HCD fragmentation with microscans of 1, resolution of 120,000, AGC target of 5e5, maximum IT of 200 ms, loop count of 1, MSX count of 1, isolation window of 2.0 m/z, isolation offset of 0.0 m/z, NCE of 16, and spectrum data type in profile mode. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.20kV and ionization chamber temperature of 275 ˚C.

UVB skin damage model

32 5-week old Balb/c male mice were randomized into 4 groups of 8 animals such that each group had similar body weight means. Mice were prepared for removal of hair from their entire back two days prior to UVB exposure (day 3) by using an electric shaver and depilatory cream. On day 5, mice received exposure to UVB (200 mJ/cm^2) produced by a broad band UVB lamp (Dermapal UVB Rev 2) powered by a Kernel UV Phototherapy system. UVB exposure was confined to a rectangular area of ~8 cm^2 by a lead shielding mask. UVB doses were confirmed by dosimeter measurements (Daavlin X96). Sham animals were shaved but received no UVB treatment. Mice were dosed from day 0 to study end at day 10 via oral gavage twice daily (CBR-470-2, 50mg/kg BID PO; BARD, 3 mg/kg PO; Vehicle, 0.5% methyl cellulose/0.5% Tween80). Mice were monitored daily for body weight changes and erythema scoring from days 5 to 10. Mice were sacrificed at day 10 and specimens collected for histological analysis from the wounded area. These studies were performed at Biomodels, LLC (Watertown, MA). Blinded erythema scores were recorded by a blinded, trained investigator according to established in house scale. In short, a scale of 0 to 4 was generated with a score of 0 referring to normal skin and a score of 4 indicating severe ulceration.

Percent wounded area measurements

Photographs of animals on day 10 of the study were taken such that the distance from camera, aperture, and exposure settings were identical. Images were then cropped such that only the shaved, wounded area encompassed the imaging field. These images were then processed with a custom ImageJ macro which first performed a three-color image deconvolution to separate the red content of the image[177]. The thresholding function within ImageJ software was then used to separate clear sites of wounding from red background present in normal skin. Red content corresponding to wounds was then quantified as a fraction of the whole imaging field and reported as percent wounded area.

Epidermal thickness measurements.

H&E stained skin sections corresponding to the wounded area were generated by Histotox Labs and accessed via pathxl software. 24 individual measurements of epidermal thickness from 8 sections spanning a 400 μm step distance was recorded per animal by a non-blinded, trained investigator. These measurements were then averaged to generate a mean epidermal thickness measurement per animal.

Cloning of lentiviral expression vector

To generate lentiviral vectors for constitutive expression, FLAG-KEAP1 were cloned into the pLenti6 backbone. pLenti-6-TP53-R273H (Addgene #22934) was digested with BamH1-HF (New England BioLabs #R3136S) and Apa1-HF (New England BioLabs #R0114S) and extracted with phenol-chloroform. Following electrophoresis (0.8% agarose) stained by SYBR Safe reagent (Invitrogen #S33102, 1:10,000 dilution), linearized backbone was excised and frozen. DNA was

eluted through a polyethylene filter and phenol-chloroform extracted followed by ethanol precipitation.

FLAG-KEAP1 construct (28023, Addgene) was digested using BamH1-HF (New England BioLabs #R3136S) and Apa1-HF (New England BioLabs #R0114S), followed by heat inactivation. Following electrophoresis (0.8% agarose) stained by SYBR Safe reagent (Invitrogen #S33102, 1:10,000 dilution), linearized insert was excised and frozen. DNA was eluted through a polyethylene filter and phenol-chloroform extracted followed by ethanol precipitation.

Backbone and insert were ligated using T4 DNA ligase (New England BioLabs #M0202). NEB 5-alpha Competent *E. coli* (High Efficiency) cells (New England BioLabs #C2987I) were transformed with the ligated plasmid. Transformed bacteria were plated on LB+Amp (100 μg/ml) agar plates and incubated at 37 ^{o}C overnight. Plasmid sequences were verified with Sanger sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing Facility using CMV-f and pBABE-r primers.

Forward sequencing primer: CMV-f 5' CGCAAATGGGCGGTAGGCGTG 3'
Reverse sequencing primer: pBABE-r 5' ACCCTAACTGACACACATTCC 3'

Lentiviral production and generation of stable cell line

HEK 293T cells were seeded in 6 cm dishes (BD Biosciences #353004) at 1.0 x 10^6 cells per dish and transfected after 24 hours with transfer plasmid (1 μg FLAG-KEAP1 pLenti6) and packaging vectors (0.1 μg pCMV-VSV-G, Addgene #8454; 0.9 μg pCMV-dR8.2, Addgene #12263) using Lipofectamine 2000 (Invitrogen #11668-027). Following overnight transfection, media was exchanged and allowed to incubate for an additional 24 hours. Viral collection was

performed at 24, 48 and 72 hours. Viral media was filtered with a Millex®-AA 0.8 µm filter (Fisher Scientific #SLAAV255F) and Polybrene (Sigma #H9268) was added to a concentration of 8 µg/mL before infection of target cell lines.

HEK293T cell lines were infected with 48-hour viral harvest. After 24 hours, cells were allowed to recover by exchanging the media. Cells were selected with Blasticidin (5 µg/mL; Fisher Scientific #20-335-025MG).

2.5. Tables

Gene	Forward Primer Sequence	Reverse Primer Sequence
NQO1	GCCTCCTTCATGGCATAGTT	GGACTGCACCAGAGCCAT
HMOX1	GAGTGTAAGGACCCATCGGA	GCCAGCAACAAAGTGCAAG
ME1	GGAGACGAAATGCATTCACA	ACGAATTCATGGAGGCAGTT
GCLM	GCTTCTTGGAAACTTGCTTCA	CTGTGTGATGCCACCAGATT
TXNRD1	TCAGGGCCGTTCATTTTTAG	GATCTGCCCGTTGTGTTTG
FTH1	GGCAAAGTTCTTCAAAGCCA	CATCAACCGCCAGATCAAC
GSR	TTGGAAAGCCATAATCAGCA	CAAGCTGGGTGGCACTTG
EPHX1	CTTCACGTGGATGAAGTGGA	CTGGCGGAATGAATTTGACT
ABCC2	GGGATCTCTTCCACACTGGAT	CATACAGGCCCTGAAGAGGA
PRDX1	GGGCACACAAAGGTGAAGTC	GCTGTTATGCCAGATGGTCAG
NQO2	TGCGTAGTCTCTCTTCAGCG	GCAACTCCTAGAGCGGTCCT
GSTM3	GGGTGATCTTGTTCTTCCCA	GGGGAAGCTCCTGACTATGA
SOD1	CCACACCTTCACTGGTCCAT	CTAGCGAGTTATGGCGACG
TXNRD1	TCAGGGCCGTTCATTTTTAG	GATCTGCCCGTTGTGTTTG
GSTP1	CTCAAAAGGCTTCAGTTGCC	ACCTCCGCTGCAAATACATC
GCLC	CTTTCTCCCCAGACAGGACC	CAAGGACGTTCTCAAGTGGG
GLO1	TGGATTAGCGTCATTCCAAG	GCGGACCCCAGTACCAAG
PGK1	CTTGGGACAGCAGCCTTAAT	CAAGCTGGACGTTAAAGGGA
TUBG1	ATCTGCCTCCCGGTCTATG	TACCTGTCGGAACATGGAGG

Table 2.1 Primer sequences for real-time qPCR analysis.

Metabolite	Precursor mass	MS1 Resolution	Product ion	MS2 Resolution	Dwell	Fragmentor	Collision Energy	Polarity	Retention time (min)
Glucose	179.05	Wide	89.2	Unit	5	68	12	Neg	12.2
G6P	258.9	Wide	138.9	Unit	100	100	5	Neg	22.3
FBP	339.1	Wide	96.9	Unit	100	100	20	Neg	26.8
GAP	169	Wide	96.9	Unit	100	100	5	Neg	22.1
BPG	264.9	Wide	96.9	Unit	5	86	21	Neg	30.9
2/3-PG	184.98	Wide	78.9	Unit	5	86	21	Neg	24.6
PEP	166.97	Wide	79	Unit	5	78	9	Neg	25.4
Pyruvate	87.1	Wide	43	Unit	100	100	10	Neg	14.8
Lac	89.1	Wide	43	Unit	100	100	20	Neg	13.5
D3-Serine	107.05	Wide	75.1	Unit	5	18	9	Neg	13.9
R5P	228.7	Wide	78.8	Unit	100	100	35	Neg	19.9
Serine	104.2	Wide	73.8	Unit	5	100	5	Neg	13.9
GSH	305.7	Wide	143.0	Unit	100	100	15	Neg	16.7
GSSG	610.7	Wide	305.9	Unit	100	100	15	Neg	20.5
Succ	117	Wide	73.1	Unit	100	100	5	Neg	18.8
Glu	146.1	Wide	102.1	Unit	100	100	5	Neg	15.9
Cit	191	Wide	111	Unit	5	100	5	Neg	24.4
NAD^+	662.1	Wide	540	Unit	100	100	15	Neg	16.1
NADH	663.4	Wide	407.9	Unit	100	100	35	Neg	16.1
$NADP^+$	742	Wide	619.9	Unit	100	100	25	Neg	24.1
NADPH	743.5	Wide	407.8	Unit	100	100	25	Neg	24.1
ATP	506	Wide	159	Unit	100	100	25	Neg	27.5
ADP	425.8	Wide	134	Unit	100	100	15	Neg	26.5
3PGha	199.98	Wide	199.98	Unit	5	116	0	Neg	22.4
3PGha	199.98	Wide	79	Unit	5	116	15	Neg	22.4
2MQ	145.1	Wide	77.1	Unit	5	100	24	Pos	8.5
2MQ	145.1	Wide	92.1	Unit	5	100	20	Pos	8.5
D3-Serine	109.07	Wide	63.1	Unit	5	40	12	Pos	4.3

Table 2.2 Acquisition parameters used for targeted metabolomic measurements on a triple quadrupole mass spectrometer. Glucose 6-phosphate, G6P; Fructose 1, 6-bisphosphate, FBP; Glyceraldehyde-3-phosphate, GAP; 1,3- or 2,3-Bisphosphoglycerate, BPG; 2- or 3-Phosphoglycerate, 2/3-PG; Phosphoenolpyruvate, PEP; Lactate, Lac; Ribose-5-phosphate, R5P; Glutathione, GSH; Glutathione disulfide, GSSG; Succinate, Succ; Glutamate, Glu; Citrate, Cit; Nicotinamide adenine dinucleotide, NAD^+; Nicotinamide adenine dinucleotide (reduced), NADH; Nicotinamide adenine dinucleotide phosphate, $NADP^+$; Nicotinamide adenine dinucleotide phosphate (reduced), NADPH; Adenosine triphosphate, ATP; Adenosine diphosphate, ADP; 3-Phosphoglyceroyl hydroxamic acid, 3PGha (derivatization product of 1,3-BPG); 2-Methylquinoxaline, 2MQ (derivatization product of MGx). D3-Serine is an isotopically labeled serine standard included in all runs as an internal normalization control.

Mutation	Primer (Forward)	Primer (Reverse)
C23S	5'-/5Phos/GCA GGG GAC GCG GTG ATG TAC -3'	5'-/5Phos/CCC CTC AGG AGA CTG TGA CTG CAG GGG C -3'
C38S	5'-/5Phos/GCC CTC CCA GCA TGG CAA -3'	5'-/5Phos/GTC ACC TCC GCC TTG GAC TCA GT -3'
C151S	5'-/5Phos/TGA ACG GTG CTG TCA TGT ACC AGA TC -3'	5'-/5Phos/TGA CGT GGA GGA CAG ACT TCT CGC -3'
C273S	5'-/5Phos/CCG AAC TTC CTG CAG ATG CAG CT -3'	5'-/5Phos/CGT CAA CGA GTG GGA GCG CAC G -3'
C288S	5'-/5Phos/GTC CGA CTC CCG CTG CAA GGA CT -3'	5'-/5Phos/TGC AGG ATC TCG GAC TTC TGC AGC T -3'
C396S	5'-/5Phos/GAC CAA TCA GTG GTC GCC CTG -3'	5'-/5Phos/ATG GGG TTG TAA GAG TCC AGG GC -3'
C405S	5'-/5Phos/CGT GCC CCG TAA CCG CAT CG -3'	5'-/5Phos/CTC ATG GGG GCG CTG GGC G -3'
K39R	5'-/5Phos/GCC CTC CCA GCA TGG CAA -3'	5'-/5Phos/GTC ACC TCC GCC CTG CAC TCA GT -3'
K39M	5'-/5Phos/GCC CTC CCA GCA TGG CAA -3'	5'- GTC ACC TCC GCC ATG CAC TCA GT -3'
C38S/K39M	5'-/5Phos/GCC CTC CCA GCA TGG CAA -3'	5'- GTC ACC TCC GCC ATG GAC TCA GT -3'
K150M	5'-/5Phos/TGA ACG GTG CTG TCA TGT ACC AGA TC -3'	5'- TGA CGT GGA GGA CAC ACA TCT CGC C -3'
R6A	5'- GCA GCC AGA TCC CGC GCC TAG CGG GGC TG -3'	5'- CAG CCC CGC TAG GCG CGG GAT CTG GCT GC -3'
R15A	5'- GGG CCT GCT GCG CAT TCC TGC CCC TGC A -3'	5'- TGC AGG GGC AGG AAT GCG CAG CAG GCC C -3'
R50A	5'- CTC CCA GCA TGG CAA CGC CAC CTT CAG CTA CAC -3'	5'- GTG TAG CTG AAG GTG GCG TTG CCA TGC TGG GAG -3'
R135A	5'- CCC AAG GTC ATG GAG GCC CTC ATT GAA TTC GCC T -3'	5'- AGG CGA ATT CAA TGA GGG CCT CCA TGA CCT TGG G -3'

Table 2.3 Primers used for site-directed mutagenesis of KEAP1 residues.

	Sequence	Amino acid
1	R.FLPLQSQCPEGAGDAVMYASTECK.A	16-39
2	R.FLPLQSQCPEGAGDAVM(15.9949)YASTECK.A	16-39
3	R.FLPLQSQCPEGAGDAVMYASTECKAEVTPSQHGNR.T	16-50
4	R.FLPLQSQCPEGAGDAVM(15.9949)YASTECKAEVTPSQHGNR.T	16-50
5	K.AEVTPSQHGNR.T	40-50
6	R.TFSYTLEDHTK.Q	51-61
7	K.QAFGIMNELR.L	62-71
8	K.QAFGIM(15.9949)NELR.L	62-71
9	R.LSQQLCDVTLQVK.Y	72-84
10	R.LSQQLCDVTLQVKYQDAPAAQFMAHK.V	72-97
11	K.YQDAPAAQFMAHK.V	85-97
12	K.YQDAPAAQFM(15.9949)AHK.V	85-97
13	K.YQDAPAAQFMAHKVVLASSSPVFK.A	85-108
14	K.YQDAPAAQFM(15.9949)AHKVVLASSSPVFK.A	85-108
15	K.VVLASSSPVFK.A	98-108
16	K.AMFTNGLR.E	109-116
17	K.AMFTNGLREQGMEVVSIEGIHPK.V	109-131
18	K.AM(15.9949)FTNGLREQGMEVVSIEGIHPK.V	109-131
19	R.EQGMEVVSIEGIHPK.V	117-131
20	R.EQGM(15.9949)EVVSIEGIHPK.V	117-131
21	R.LIEFAYTASISMGEK.C	135-150
22	R.LIEFAYTASISM(15.9949)GEK.C	135-150
23	K.CVLHVMNGAVMYQIDSVVR.A	151-169
24	K.CVLHVM(15.9949)NGAVMYQIDSVVR.A	151-169
25	R.ACSDFLVQQLDPSNAIGIANFAEQIGCVELHQR.A	170-202
26	R.AREYIYMHFGEVAK.Q	203-216
27	R.AREYIYM(15.9949)HFGEVAK.Q	203-216
28	R.EYIYMHFGEVAK.Q	205-216
29	R.EYIYM(15.9949)HFGEVAK.Q	205-216
30	R.EYIYMHFGEVAKQEEFFNLSHCQLVTLISR.D	205-234
31	K.QEEFFNLSHCQLVTLISR.D	217-234
32	K.QEEFFNLSHCQLVTLISRDDLNVR.C	217-240
33	R.CESEVFHACINWVK.Y	241-254
34	K.YDCEQR.R	255-260
35	R.RFYVQALLR.A	261-269
36	R.FYVQALLR.A	262-269
37	R.CHSLTPNFLQMQLQK.C	273-287
38	R.CHSLTPNFLQM(15.9949)QLQK.C	273-287

39	R.CHSLTPNFLQMQLQKCEILQSDSR.C	273-296
40	K.CEILQSDSR.C	288-296
41	R.CKDYLVK.I	297-303
42	K.IFEELTLHKPTQVMPCR.A	304-320
43	K.IFEELTLHKPTQVMPCRAPK.V	304-323
44	R.APKVGRLIYTAGGYFRQSLSYLEAYNPSDGTWLR.L	321-354
45	K.VGRLIYTAGGYFR.Q	324-336
46	R.LIYTAGGYFR.Q	327-336
47	R.QSLSYLEAYNPSDGTWLR.L	337-354
48	R.LADLQVPR.S	355-362
49	R.SGLAGCVVGGLLYAVGGR.N	363-380
50	R.NNSPDGNTDSSALDCYNPMTNQWSPCAPMSVPR.N	381-413
51	R.NNSPDGNTDSSALDCYNPM(15.9949)TNQWSPCAPMSVPR.N	381-413
52	R.NNSPDGNTDSSALDCYNPMTNQWSPCAPM(15.9949)SVPR.N	381-413
53	R.IGVGVIDGHIYAVGGSHGCIHHNSVER.Y	416-442
54	R.YEPERDEWHLVAPMLTR.R	443-459
55	R.YEPERDEWHLVAPM(15.9949)LTR.R	443-459
56	R.YEPERDEWHLVAPMLTRR.I	443-460
57	R.DEWHLVAPMLTR.R	448-459
58	R.DEWHLVAPM(15.9949)LTR.R	448-459
59	R.RIGVGVAVLNR.L	460-470
60	R.IGVGVAVLNR.L	461-470
61	R.LLYAVGGFDGTNR.L	471-483
62	R.LLYAVGGFDGTNRLNSAECYYPER.N	471-494
63	R.LLYAVGGFDGTNRLNSAECYYPERNEWR.M	471-498
64	R.LNSAECYYPER.N	484-494
65	R.LNSAECYYPERNEWR.M	484-498
66	R.MITAMNTIR.S	499-507
67	R.M(15.9949)ITAMNTIR.S	499-507
68	R.MITAM(15.9949)NTIR.S	499-507
69	R.M(15.9949)ITAM(15.9949)NTIR.S	499-507
70	R.SGAGVCVLHNCIYAAGGYDGQDQLNSVER.Y	508-536
71	R.SGAGVCVLHNCIYAAGGYDGQDQLNSVERYDVETETWTFVAPMK.H	508-551
72	R.YDVETETWTFVAPMK.H	537-551
73	R.YDVETETWTFVAPM(15.9949)K.H	537-551
74	R.RSALGITVHQGR.I	554-565
75	R.SALGITVHQGR.I	555-565
76	R.IYVLGGYDGHTFLDSVECYDPDTDTWSEVTR.M	566-596
77	R.SGVGVAVTMEPCR.K	602-614

78	R.SGVGVAVTM(15.9949)EPCR.K	602-614
79	R.SGVGVAVTMEPCRK.Q	602-615
80	R.SGVGVAVTM(15.9949)EPCRK.Q	602-615

Table 2.4 List of KEAP1 peptides collected for SILAC surface mapping assay

CHAPTER 3

Light-activated profiling of protein-protein interactions in live cells*

We are preparing for the publication of the work described in this chapter.

3.1 Introduction

Elucidation of the underlying protein communication networks and corresponding signaling pathways in live cells necessitates high-resolution protein proximity maps. The given technical challenges to achieve the exquisite protein network maps are to isolate, identify, and quantify the spatially compartmentalized interaction partners of the protein of interest (POI).

The traditional and predominant approach to investigate the protein-protein interactions (PPIs) and subcellular protein complexes is affinity-purification mass spectrometry (AP-MS)[178-181] in which interaction partners are co-enriched with immuno-affinity pulldown of POI[182] and identified by LC-MS/MS based proteomic analysis[183-186]. However, the challenges still remain with AP-MS to keep reliable information of temporal and weak PPIs and to get the maps of membrane protein communication networks due to the hydrophobic nature of these proteins specifically localized in the transmembrane.

* The author's contribution to the studies presented in this chapter: The author contributed equally to this work with David C. McCutcheon. The author designed, performed, and analyzed the all proteomic experiments described in this chapter and wrote the manuscript with Raymond E. Moellering for the publication of this work.

The state-of-the-art technologies to address these challenges are biotin-tagging systems by which proximal proteins (prey proteins) are covalently biotinylated with the aid of functional enzymes fused to the protein of interest (bait protein). Hence, proximity-labeled proteins could be selectively handled with streptavidin-pulldown and identified in the following proteomic analysis.

Enzymatic biotin ligation systems are the prominent biotin-tagging approaches. BioID[187], the most popular enzymatic tagging system, utilizes the engineered *E.Coli* biotin ligase (BirA) to generate biotinoyl-5'-AMP (bioAMP) intermediate from biotin and ATP, and then biotinylates the lysine residues in proximal proteins (Figure 3.1b)[188-192]. The rate of enzymatic biotinylation completely relies on the kinetics of the expressed biotin ligases. BioID requires long period of time (18-24 hours) to achieve sufficient amount of biotinylated proteins[190,193], thus, capturing the temporal and dynamic protein interactions between prey- and bait- proteins is still challenging. To accelerate the rate of enzymatic biotin tagging, hence solving a slow kinetics issue, advanced biotin ligase system (TurboID) engineered by yeast display-based directed evolution was recently developed[194], demonstrating the successful identification of spatially compartmentalized proteins in 10 minutes ligation. Indeed, other enzymatic tagging systems were recently introduced, such as NEDDylatior system[195] that labels prey proteins with NEDD8, and PUP-IT system[196] which labels small bacterial protein Pup to prey proteins by a fused bacterial Pup ligase PafA. Nonetheless, the efficiency of tagging by enzymatic labeling systems is time-dependent, and

The apparent merit of non-enzymatic biotin tagging systems is rapid kinetics of the labeling reaction enabling to interrogate the dynamic communication networks of the protein of interest in live cells. The advanced system of this category is APEX[197,198] which induces covalent modification of biotin-phenol to proximal proteins of the bait protein fused by an engineered ascorbic acid peroxidase (Figure 3.1a). Biotin-phenol radical formation and following covalent

labeling reaction in APEX system are completed within a minute with labeling radius of 10-20 nm[199], thus, APEX has been harnessed to get high resolution protein interaction maps of various organelles[200-204] in diverse types of cells[193,200] and species[205,206], as well as to resolve the spatiotemporal dynamics[207] of protein interaction networks. However, the major flaw of current non-enzymatic tagging systems, of both APEX and horseradish peroxidase (HRP) system[208,209], is a cytotoxicity issue caused by a supply of hydroperoxide (H_2O_2) for the generation of locally populated radicals. Infusion of H_2O_2 also brings a detrimental effect to study redox sensitive signaling pathways, which are highly sensitive toward and affected by reactive oxygen species, in live cells. Therefore, these defects limit the universal application of peroxidase-based proximity labeling systems.

Here we propose a new proximity tagging platform, termed Photochemoproteomic Proximity Profiling (P3 Profiling), which induces a non-enzymatic labeling of the proximal proteins of POI in live cells initiated by light irradiation. The results from *in vitro* validation tests and quantitative proteomics data from actual proximity profiling experiments in live cells demonstrate the potential versatility of this novel proximity labeling technology.

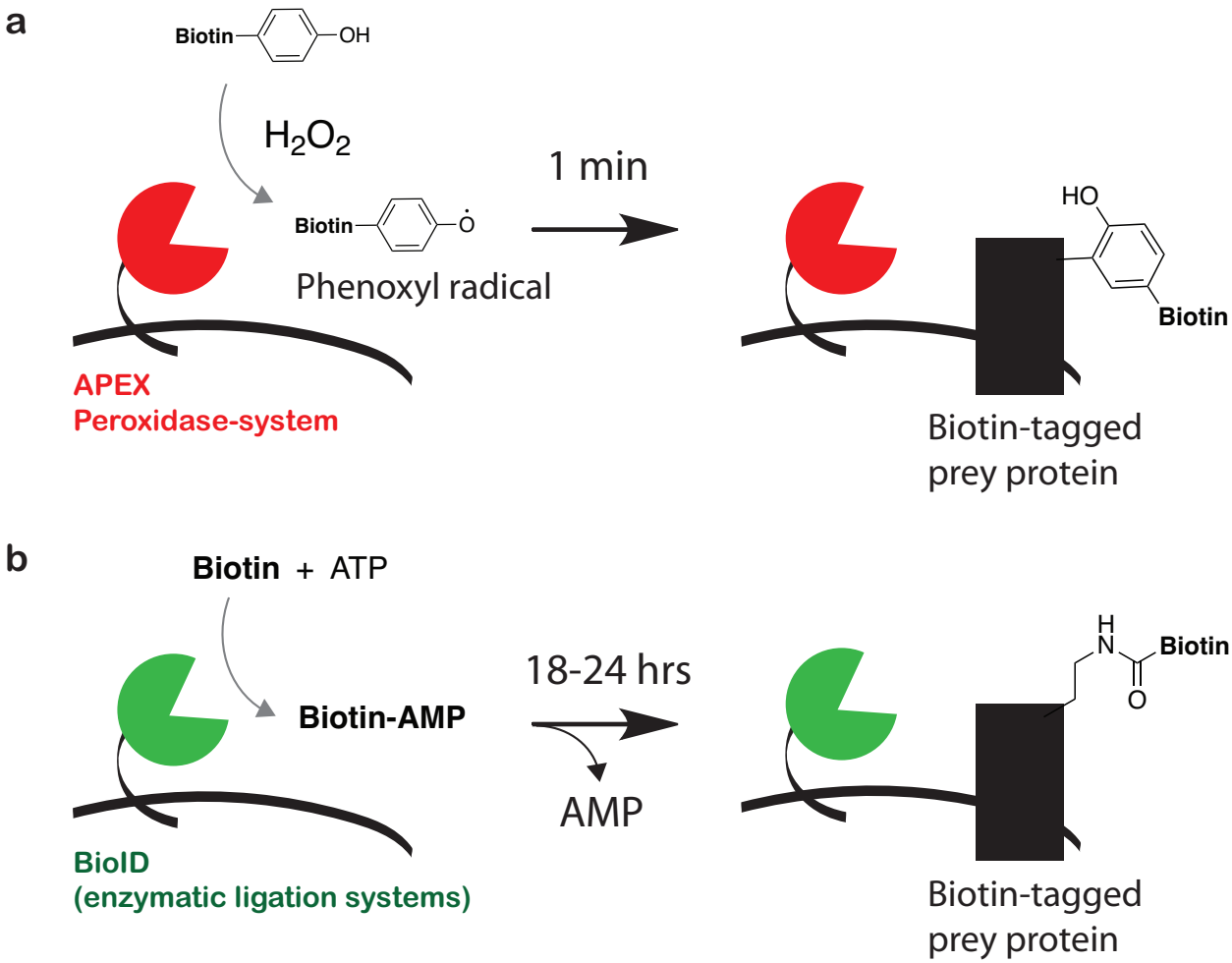


Figure 3.1 Major biotin-tagging proximity profiling systems to date
a. Peroxidase-based labeling systems require the infusion of hydroperoxide 1 minute to cells preloaded with biotin-phenol to initiate labeling. APEX or other peroxidase-based systems oxidizes biotin-phenol into a phenoxy radical which covalently tags proximal endogenous proteins preferentially at electron-rich side chains such as tyrosine. **b.** Enzymatic ligation system, such as BioID, labels the target with exogenously supplied free biotin and endogenous ATP, taking 18-24 hours to get sufficient labeling. BioID converts biotin to reactive bioAMP which is released from the enzyme's active site to react with lysine residues on proximal prey proteins.

3.2 Results

3.2.1 Design of the P3 probes

P3 profiling probe is composed of three main parts: tracer, linker, and biotin-photoreactive group (Figure 3.2a). We selected O^6-benzylguanine substrate (BnG) as a 'tracer' that selectively targets a bait protein fused by a SNAP protein. SNAP[210,211] is an engineered O^6-methylguanine-DNA methyltransferase (MGMT) specifically accepting BnG derivatives and covalently binding with these ligands, thus BnG-SNAP system has been applied for numerous protein labeling studies[212-216]. 'Linker' of the probe was designed to connect the tracer with biotin-photoreactive group and to be cleaved by UV irradiation (λ = 365 nm). Thereby, biotin-photoreactive group is diffused from the bait protein after irreversible covalent bond formation of the P3 probe to SNAP followed by light-induced cleavages of the linker. The 2-Nitrobenzyl substrate, which has been widely utilized as a photocleavable linker due to its high photocleavage efficiency, was employed in the P3 profiling probe. 'Biotin-photoreactive group' of the probe was designed to possess a biotin and a photoreactive substrate which generates a covalent crosslink to the prey proteins after the light exposure. We designed 1st generation P3 probes, termed PP1 and PP2, by introducing a diazirine substrate, which generates carbene reactive intermediate by light activation (λ = 365 nm), into the biotin-photoreactive group part. For the simple evaluation of the P3 profiling system in live cells, we also designed PP-FITC, a P3 profiling probe analogue with the substitution of biotin to a fluorescein isothiocyanate (FITC) dye.

To synthesize the P3 probes, we first synthesized three distinct parts and then connected them together (Tracer-Linker-Biotin Photoreactive group).

3.2.2 Validation of the P3 profiling platform in live cells

To evaluate the specificity of BnG substrate of the P3 probe to SNAP protein, as well as the photocleavage efficiency and kinetics of the linker, we treated the PP-FITC to FLAG-SNAP expressed HEK293T cells (15 μM, 2 hours) to induce BnG-SNAP irreversible covalent bond formation, and then we applied steady UV irradiation time between 0 minute to 10 minutes to the each cell plate followed by SDS-PAGEs. Strong FITC bands were observed at 25 kDa, which corresponds to the mass of FLAG-SNAP, and the intensity of FITC signals were diminished by longer light exposure, approximately 20 % left in 2.5 minutes and completely vanished in 5-10 minutes irradiation (Figure 3.2b). These results indicate the successful labeling of SNAP protein with BnG substrate of the P3 probe in live cells and photocleavage of the linker enabling the dissociation and diffusion of FITC dye from FLAG-SNAP in few minutes of light exposure.

Next, we sought to test if actual P3 probe, besides P3 analogue FITC probe, also specifically labels the SNAP protein in FLAG-SNAP expressed HEK293T cells with dose-dependent manners, thereby we treated PP1 to the live cells in 2 hours. The apparent biotin bands were observed near 25 kDa in streptavidin Western blots, which are consistent with the results from FITC tests, and the intensity of these biotin bands were increased with higher doses treatment and saturated at 15 μM of PP1 (Figure 3.2c). Interestingly, we also observed the slight mass shift of FLAG-SNAP with the covalent labeling of PP1, as evidenced by FLAG bands, due to the addition of probe mass to the protein.

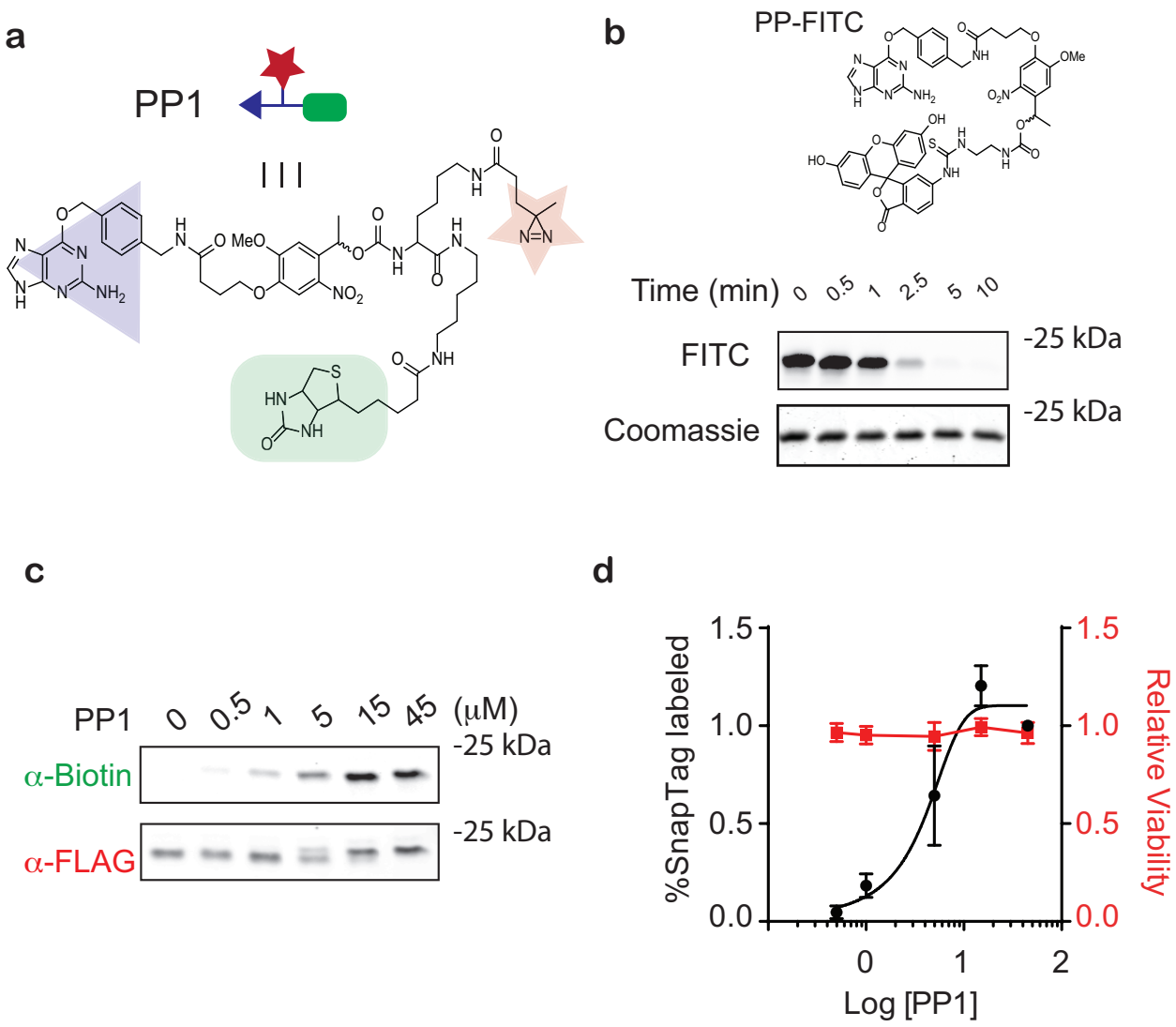


Figure 3.2 Design and validation of the P3 profiling platform
a. Design of the P3 probe. BnG substrate (purple) is selectively recognized and covalently linked by SNAP fused to a bait protein. Photoreactive substrate (diazirine, dark orange) and biotin (green) are employed in the "biotin-photoreactive group" component which is activated and diffused from the bait protein by few minutes light irradiation. Photocleavable linker (2-Nitrobenzyl linker) links between BnG substrate and biotin-diazirine component in our first P3 probe, PP1. **b.** Selectivity and efficiency of the compound to tag SNAP, and diffusion of the functional component by light-activated cleavages of linker were initially tested with PP-FITC and recombinant, purified FLAG-SNAP. **c.** Cell permeability and tagging efficiency to SNAP by PP1 were examined with FLAG-SNAP expressed HEK293T cells with indicated doses. **d.** Effect of PP1 dose treatments on cell viability and SNAP labeling efficiency within the time (2 hours) of P3 profiling test.

In order to check the potential cytotoxic effects of the P3 profiling system, we treated PP1 to the FLAG-SNAP expressed HEK293T cells with increasing doses, up to 45 μM, in 2 hours in which P3 probe enables the sufficient labeling of SNAP proteins. As confirmed by CTG readout, the results of cell viability test imply the no cytotoxic issue in our P3 profiling system (Figure 3.2d).

In summary, data suggested that our design of P3 profiling system successfully works in live cells, as confirmed by the selective binding of the P3 probe to SNAP and light activated cleavage of the probe linker part, without bringing the cytotoxicity.

3.2.3 *In vitro* proximity labeling with the P3 profiling platform

As 'tracer' and 'linker' of the P3 probe are working in FLAG-SNAP expressed mammalian cells confirmed by validation tests, we wondered if 'biotin-photoreactive group' successfully labels the proximal proteins via P3 profiling platform and light activation.

To address the question, we designed and performed an *in vitro* assay with FLAG-SNAP protein and α-FLAG agarose resin (mouse monoclonal) (Figure 3.3a). After the lysis of FLAG-SNAP expressed HEK293T cells, α-FLAG resin was added to the lysate to enrich FLAG-SNAP protein, washed, resuspended in PBS, and then PP1 or PP2 were added to the mixture. P3 profiling system in the workflow recognized α-FLAG as a proximal protein of FLAG-SNAP. After 1 hour of probe treatments, UV light was irradiated to the resin mixtures to induce diffusion and labeling of the biotins to FLAG-SNAP and α-FLAG protein complexes. These proteins were eluted by heating the resin with SDS-PAGE loading buffer containing reducing reagent (DTT or βME), and biotin labeling was analyzed by Western blots. Strong biotin bands were observed on FLAG-SNAP with PP1 or PP2 treatments, but not in the light and heavy chains of α-FLAG, in the absence

of light activation, indicating a localization of the P3 probe biotin substrate on the SNAP protein. By contrast, UV irradiation led to the formation of weak but distinguishable biotin bands on both light and heavy chains of the α-FLAG, as well as thinner biotin bands on FLAG-SNAP with the pre-treatment of PP1 or PP2, indicating a diffusion of biotin-photoreactive group from FLAG-SNAP followed by biotin labeling to the α-FLAG (Figure 3.3b). These biotin bands on both light and heavy chains of the α-FLAG were not shown when α-FLAG was pre-mixed with SNAP protein which has no FLAG-tag, suggesting that biotin-labeling only occurs on proximal proteins of SNAP-fused bait protein (Figure 3.3c). Overall, data suggest that P3 probes are capable of labeling biotins to the proximal proteins of bait protein via light irradiation, confirmed by *in vitro* tests.

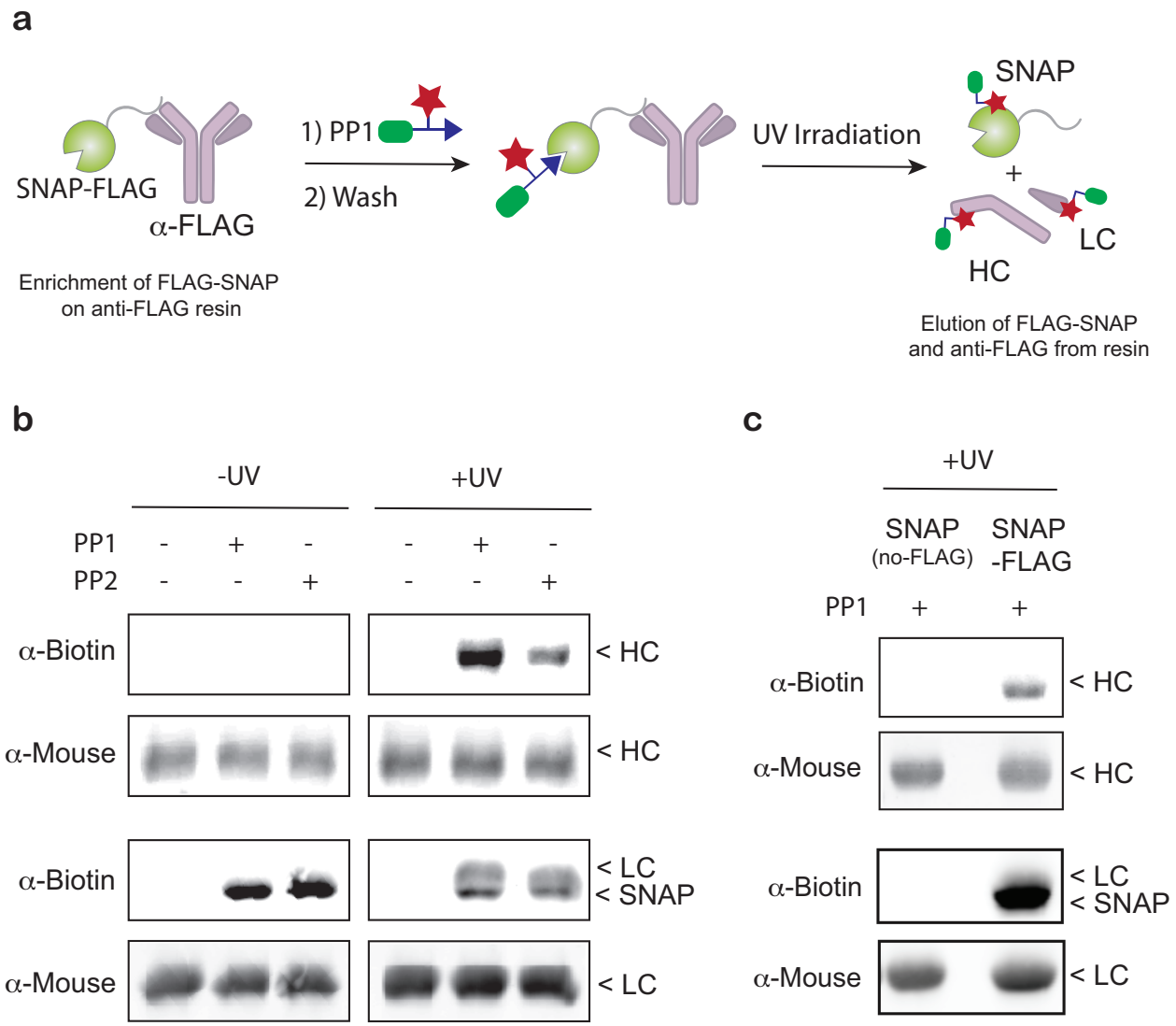


Figure 3.3 *In vitro* proximity profiling test with P3 profiling platform
a. Schematic of the *in vitro* validation test of proximity labeling efficiency with the P3 profiling platform. The protein complex of anti-FLAG with FLAG-tagged SNAP on resins mimics the proximity between SNAP fused bait protein and target prey proteins in the subcellular area. After inducing the FLAG-SNAP bound to anti-FLAG on resins, PP1 was treated to the mix followed by UV irradiation and Western blot analyses to see biotin labeling in anti-FLAG. **b.** Anti-biotin bands in the light- and heavy chains of FLAG antibody before and after the light irradiation to PP1 bound SNAP-anti FLAG complex *in vitro*. **c.** Anti-biotin bands in the light- and heavy chains of FLAG antibody pre-treated with either SNAP or FLAG-SNAP followed by the PP1 treatment and light irradiation *in vitro*: LC, light chain; HC, heavy chain.

3.2.4 Proximity profiling of KEAP1 with the P3 profiling platform in live cells

We next employed the P3 profiling platform to a specific POI in live cells and examined the system in full-scale proteomic experiments and quantitative proteomic analyses (Figure 3.4a). To investigate the dynamic communication partners of the bait protein via our proximity labeling system, we chose KEAP1 as a target.

KEAP1 is a representative sensor protein in the cellular system that regulates the stress via recognizing various reactive species. In KEAP1-NRF2 signaling axis[120-123,162], electrophilic molecules and reactive oxygen species directly modify sensitive cysteine sensor residues of KEAP1 that lead to the behavioral alteration of KEAP1 and stabilization of its binding partner NRF2 transcription factor that induces the activation of downstream signaling cascade to produce antioxidant-responsive elements (AREs) and detoxifies the cellular system to keep homeostasis.

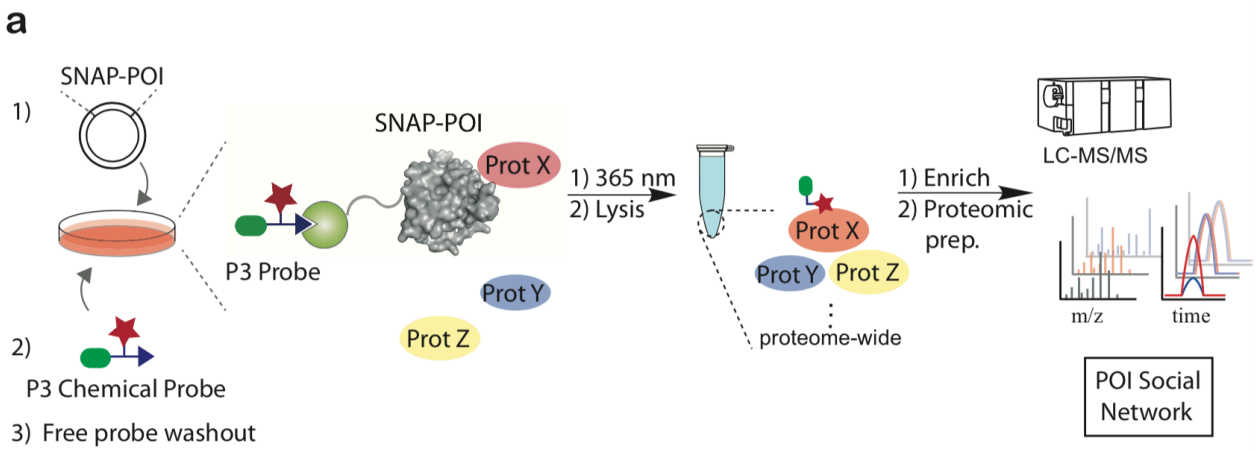
Because of its biological importance, numerous studies have been performed to elucidate the dynamic communication networks of KEAP1 in live cells, however, application of current biotin-tagging technologies to study KEAP1-NRF2 pathway has not reported to date. This is because 1) slow kinetics of the enzymatic labeling precludes the use of enzymatic biotin-tagging systems to investigate dynamic interactions of sensitive sensor proteins and 2) electrophile sensor proteins, such as KEAP1, are tuned to sensitively recognize the reactive species, thereby infusion of external hydroperoxide affects the behavior of these proteins in live cells and makes harder to study redox signaling pathways via peroxidase-based proximity labeling systems. Instead, the P3 profiling system, which is simply activated by a short light exposure and labels the prey proteins by photoreactive functionalities, enables its application to study KEAP1 protein interaction network.

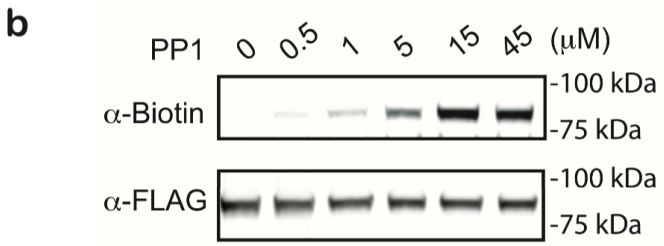
SNAP protein was fused to the N-terminus (SNAP-KEAP1) or C-terminus (KEAP1-SNAP) sites in KEAP1, and these constructs were stably expressed in HEK293T cells by viral transduction. As consistent with the results from FLAG-SNAP cells, 15 μM of PP1 was shown to saturate the stably expressed SNAP-KEAP1 protein in live cells (Figure 3.4b).

In order to identify the proteins specifically enriched by P3 profiling system, we employed a SILAC (stable isotope labeling with amino acids in cell culture)-quantitative proteomics[156] by generating a pair of heavy- and light- isotope labeled cells of SNAP-KEAP1 and KEAP1-SNAP proteins. PP1 was treated to each heavy cell plate for 2 hours and excess probes were washed off from these cells. Then, cells were exposed to UV light for inducing light activation of PP1 photoreactive substrates followed by photoaffinity labeling of the proximal proteins. After the process of the enrichment of the target proteins and following proteomic sample preparation, enriched samples were analyzed by LC-MS/MS based proteomics and the list of enriched proteins by P3 profiling method was identified with the SILAC quantitation analysis after the normalization of SILAC ratios of enriched proteins by a median SILAC ratio of the corresponding bulk proteome (Figure 3.4c).

Among the list of enriched proteins which passed the filter (Probe-to-DMSO ratio > 2, $P < 0.05$), KEAP1 and MGMT were shown to be highly and consistently enriched by P3 profiling system (Figure 3.5a and 3.5b). MGMT gene is corresponding to SNAP protein fused to KEAP1, indicating that P3 profiling system is capable of tagging the biotin substrates to the most proximal protein, bait protein itself and another monomeric KEAP1 non-covalently[120,160] or covalently[149] binding to form KEAP1 homodimer in cytosol, and SILAC ratio of these self-labeled proteins could be considered as a barometer of the profiling efficiency. Intriguingly, we also observed PGAM5, another protein consistently shown to be enriched with P3 profiling. Yet the precise

mechanisms in actions and roles of this protein in cells are still not well characterized, many published studies have demonstrated PGAM5 as a strong substrate for KEAP1[217] that forms the PGAM5-KEAP1-NRF2 complex[218,219] localized to the cytosolic surface of the outer mitochondrial membrane[220]. Enrichments of PGAM5 were detected from both SNAP-KEAP1 and KEAP1-SNAP expressed HEK293T cells, with high -$\log_{10}$ *P* values, and the actual chromatogram of PGAM5 peptide MS1 peak from both heavy and light digests suggests that PGAM5 is a promising target of P3 probe and SNAP fused KEAP1 construct in live cells (Figure 3.5d). Other proteins that passed the criteria (Probe-to-DMSO ratio > 2, $P < 0.05$) were also identified in SNAP-KEAP1 and KEAP1-SNAP cells and listed (Figure 3.5c).





(Figure 3.4, continued)

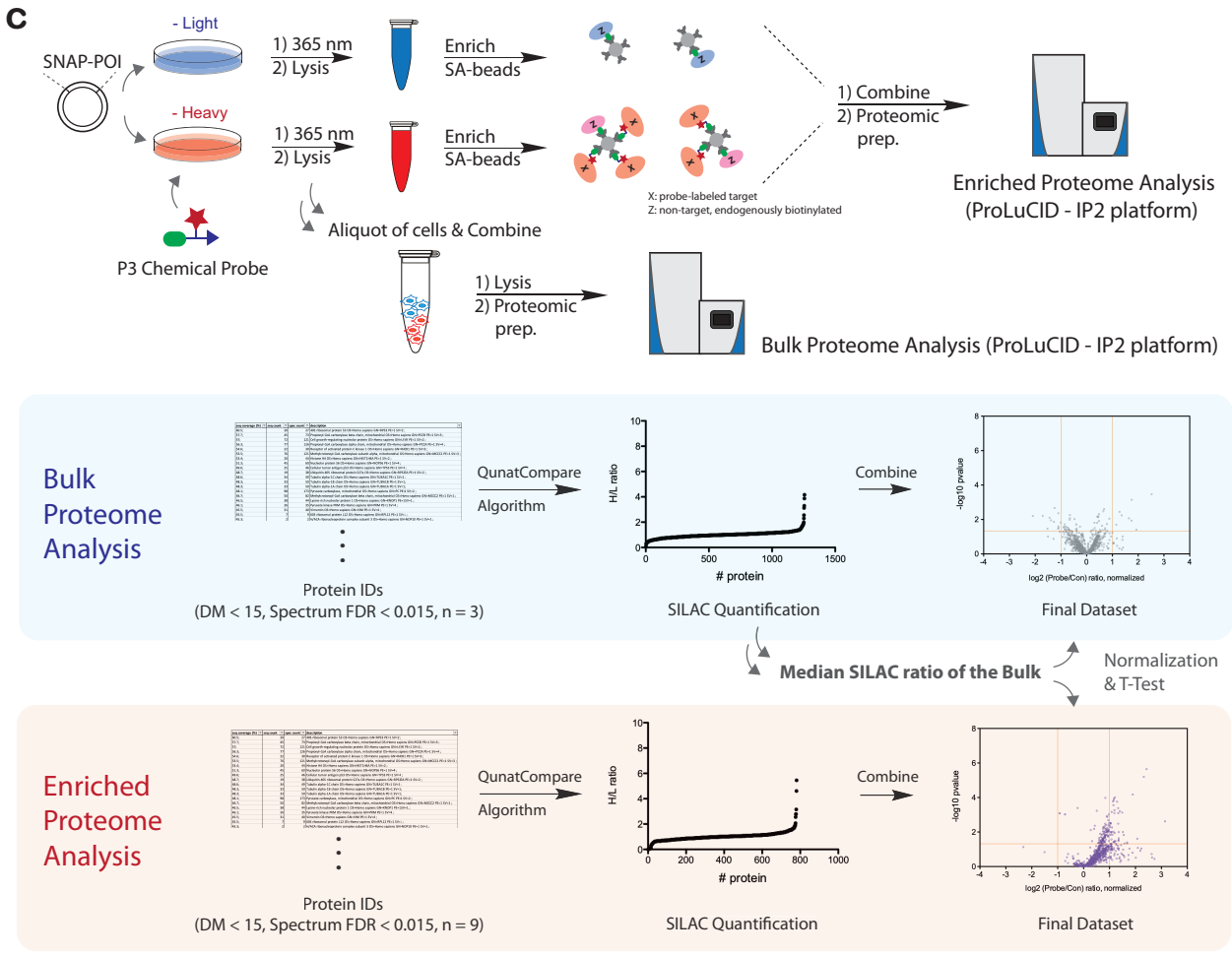
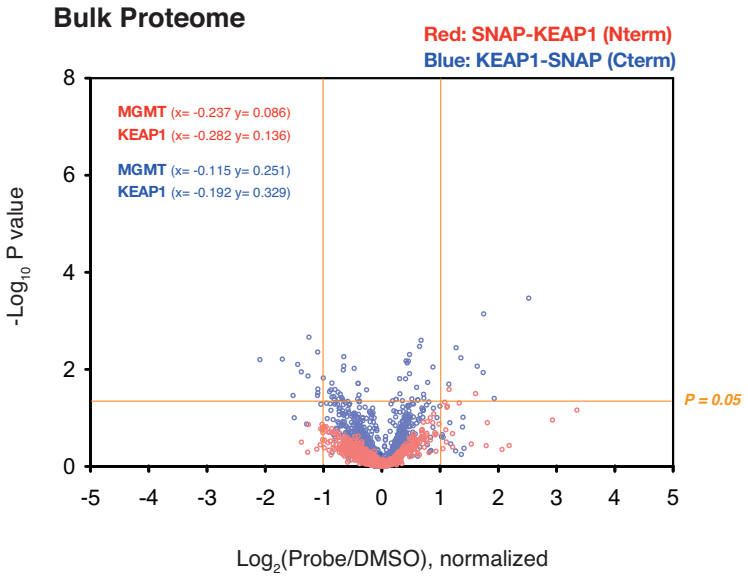


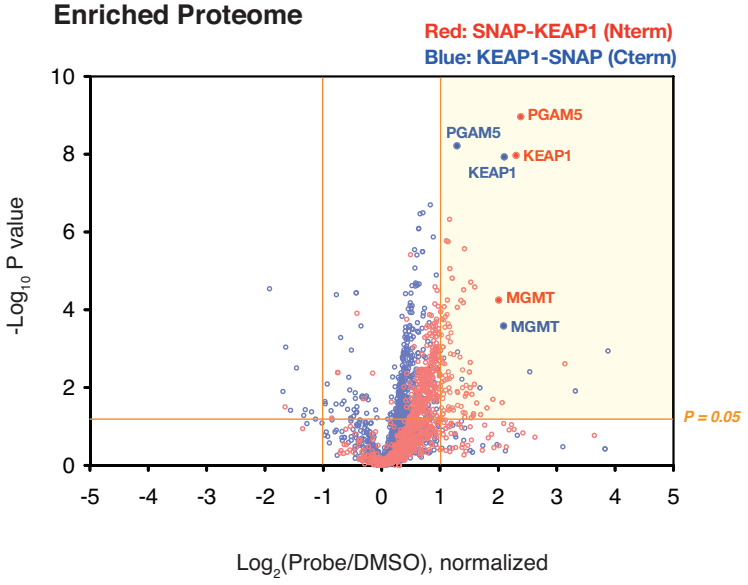
Figure 3.4 *In situ* proximity labeling and mass spectrometric analysis workflow for P3 profiling system

a. *In situ* P3 profiling workflow. SNAP-POI fusion construct is expressed in cells, and P3 probe is treated to these cells to make a covalent tag with expressed SNAP-POI. After UV light irradiation, proximal proteins of the POI are labeled, enriched with streptavidin pull down, and analyzed by high resolution mass spectrometry (MS). **b.** Indicated doses of PP1 were treated to HEK293T cells expressing SNAP fused FLAG-KEAP1 to validate optimal probe concentration for *in situ* tests. **c.** Schematic of the sample preparation and data work up process for quantitative proteomic analysis using SILAC cells. Enriched proteins by P3 probes were identified and listed by normalized SILAC values (Probe/DMSO) and *P* values.

a



b



c

List of Enriched Proteins

(Probe/DMSO > 2, P < 0.05)

SNAP-KEAP1		KEAP1-SNAP
PUF60	NUP133	SLC29A1
PGAM5	MRTO4	RAF1, ARAF, BRAF
KEAP1	RALY	C2orf44
GPAT4	RPL15	KEAP1
MGMT	MRPS24, 21	MGMT
SOAT1	EBNA1BP2	PLK1
PRPF38A	ACIN1	MLLT4
LUC7L2	WDR46	PGAM5
NOL6	FCF1	HLA-A,B,C
DHX15	HIST1H1-C,D,E	RAB7A
CTBP2	SLC25A12	NDUFV1
HK2	GNL2	
ALDH4A1	NDUFA9	
C11orf98	URB1	
RPF2	MKI67	
SCD	SLC25A4	
SNRPD2	DAP3	
RAB5C	PUM3	
MRPL44	RBMX	
CYP5A1	MRPS9	
UTP20	DDX56	
MRPL49	MRPL22	
DHCR7	RPL36	
RRS1	NOC2L	
NOP2	COIL	
ELOVL5	SLC25A19	
CHTOP	SUN1	
URB2	RAB1B	
FAR1	MRPL15	
BRIX1	SNU13	
EPHA2	MICU2	
RAB1C	ATAD3C	
STOML2	PRPF4B	
SRSF6	POLRMT	
NDUFS7	SLC25A10	
IMP4	TIMM23, TIMM23B	
TARS2	HEATR1	
SRSF4	BAZ2A	
KDELR1	HNRNPL	
HSDL2	ABCD3	
AQR	NOC4L	
ZC3HAV1	PBRM1	
TMEM33		

(Figure 3.5, continued)

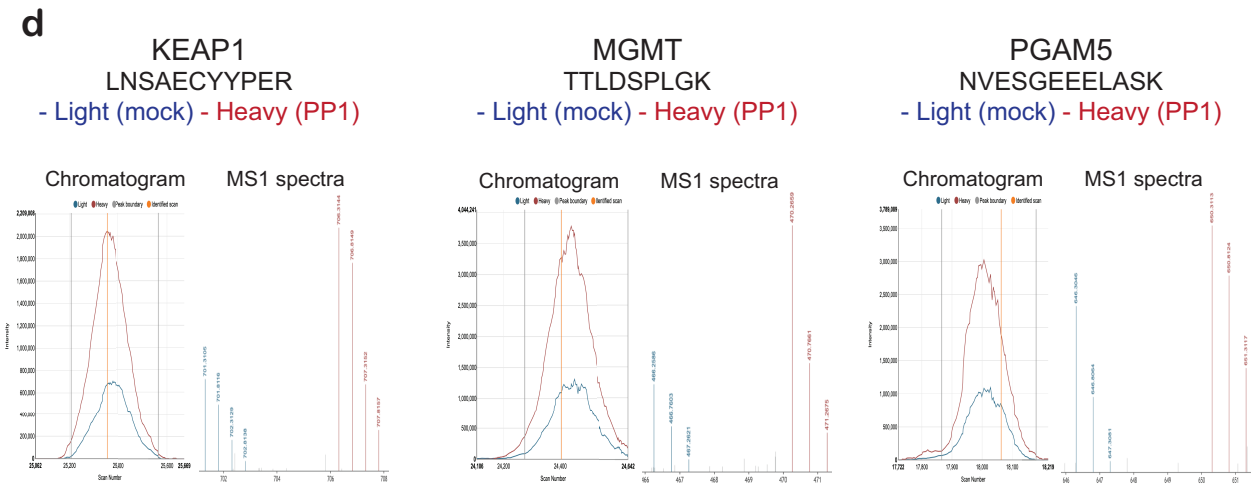


Figure 3.5 Application P3 profiling platform identified PPI partners of KEAP1
a. **b.** Volcano plots of bulk proteins (**a**) and enriched proteins (**b**). The logarithmic ratios of protein SILAC value (Probe/DMSO) were normalized and plotted against negative logarithmic *P* values from univariate two-sided t-test. Data from bulk proteins were analyzed with 3 biologically independent samples (n=3), and enriched proteins were analyzed with 3 biologically independent samples and 3 replicate MS runs per each sample (n=9). **c.** Enriched proteins from KEAP1-SNAP and SNAP-KEAP1 HEK293T cells with *P* values of < 0.05 and Probe/DMSO ratios of >2 were listed in the box. **d.** Chromatograms and MS1 spectrum of the peptide from three representative proximity labeling targets, both self-labeled bait proteins and well-characterized interaction partner.

3.3 Conclusion and Discussion

SNAP is a relatively small protein (19.4 kDa) and is ideal for the P3 profiling studies due to its unique strengths. 1) SNAP protein forms an irreversible covalent bond with BnG derivatives and the efficiency the reaction is not affected by the chemical structure of the BnG synthetic probes, allowing us to design the diverse P3 probes without any structural limitations. 2) BnG substrate is chemically inert but only labels SNAP protein, therefore it avoids the non-specific labeling in the complex biological system. 3) Fusion of SNAP protein to POI is not restricted with respect to cellular localization and expression host. Hence, the P3 profiling technology is versatile and universally applicable regardless of any specific types of POIs in live cells.

Our validation experiments revealed that P3 probe is cell permeable, is covalently liked to the intracellular SNAP protein in live cells, supplies photocrosslinkable biotin substrates to the neighborhoods to label proximal proteins *in vitro* and in live cells. Application of the P3 profiling technology to the representative redox sensitive sensor protein KEAP1 revealed that our strategy successfully identified PPI partners in live cells, as evidenced by the identification of well characterized target PGAM5. While PGAM5 was detected from both SNAP-KEAP1 and KEAP1-SNAP expressed HEK293T cells, the list of highly enriched proteins from two different constructs expressed cells are independent and not completely overlapped. The results may imply that the P3 profiling platform, with the fusion of SNAP in the specific locations, highly compartmentalizes the proximal proteins near the SNAP fusion that leads to the high-resolution mapping of direct protein communications.

A half-life of the reactive functionalities, such as carbenes and radicals generated by a light activation of diverse photoreactive groups[221-224], could be readily adjusted by the combination and arrangement of various chemical moieties nearby. Therefore, the P3 profiling technology

potentially enables the proximal profiling with a desired labeling radius. Our first application of P3 profiling platform to KEAP1 construct was able to identify the promising target PGAM5, yet other well-known interaction partners of KEAP1, such as CUL3, RBX1, and NRF2, were not detected in our attempts with PP1 probe. As actual KEAP1 behaviors and temporal communication network in live cells are not completely unveiled, our future studies will focus on the development of next generation P3 probes and application of these tools to collect the information of spatial distance. Indeed, the novel proximity profiling technology will be utilized for studying other intriguing cellular regulatory programs.

3.4 Materials and Methods

Abbreviations

HATU: 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate

NHS: N-Hydroxysuccinimide

EDC•HCl: N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride

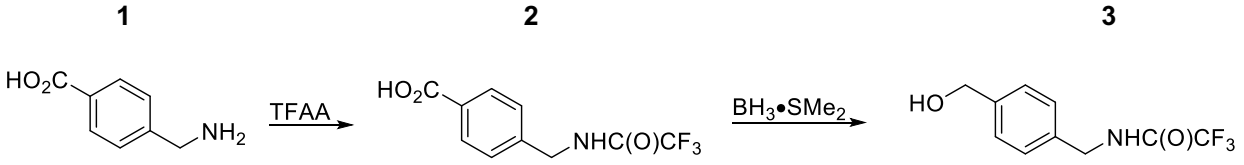
DSC: *N,N′*-Disuccinimidyl carbonate

General synthetic methods

Reagents purchased from commercial suppliers were analytical grade and used without further purification. All reactions were carried out in oven dried flasks using anhydrous solvents (Acros) unless otherwise specified. Reaction progress was monitored by thin-layer chromatography on Macherey-Nagel SIL G-25 UV254 TLC plates, visualized with UV light, ceric ammonium molybdate (CAM), *p*-anisidine, bromophenol blue, 2,4-dinitrophenyl hydrazine (DNP), or KMnO4 TLC stains. Nuclear magnetic resonance spectra were acquired using either a Bruker AVANCE II+ 500; 11.7 Tesla NMR or Bruker DRX 400; 9.3 Tesla NMR instrument. Accurate mass measurements were obtained using an Agilent 6224 Tof-MS instrument. When unavoidableCompounds were purified via flash column chromatography using Siliaflash F60 60 Å, 230-400 mesh silica gel (Silicycle).

Spectroline spectrolinker XL-1500A used for UVA irradiation.

Chemical synthesis of photoproximity profiling (P3)-probes



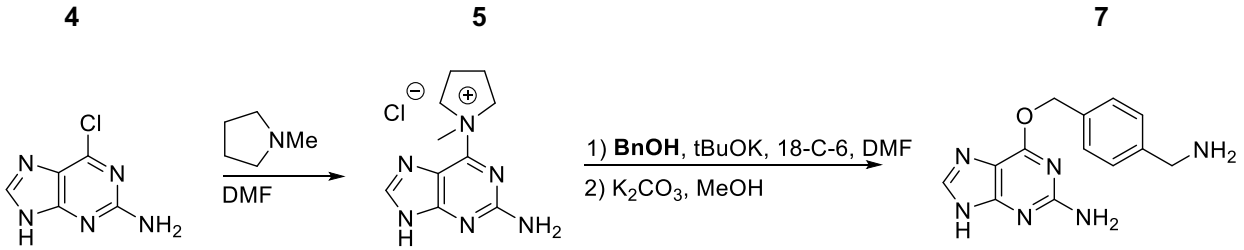
4-[[(2,2,2-trifluoroacetyl)amino]methyl]benzoic acid (2)

Solid 4-(aminomethyl)benzoic acid (1) (15.1 g, 100 mmol) was dissolved in TFAA (42 mL) cooled to 0 °C. Once dissolved, the ice bath was removed and the reaction was allowed to stir at room temperature until starting material was consumed, ~2 hr. Upon completion, the reaction was quenched with H_2O (100 mL) and precipitate collected via vacuum filtration. The product was dried under suction then collected to afford benzoic acid (2) (24.0 g 97%) as a white solid. ^{1}H NMR (500 MHz, DMSO-d^6) δ 7.93 (d, J = 8.3 Hz, 2H), 7.39 (d, J = 8.3 Hz, 2H), 4.47 (d, J = 6.0 Hz, 2H). ^{13}C NMR (125 MHz, DMSO-d^6) δ 167.12, 156.58 (q, J = 36.3 Hz), 142.50, 129.89, 129.65, 127.44, 116.05 (q, J = 228.1 Hz), 42.41.

2,2,2-trifluoro-*N*-[[4-(hydroxymethyl)phenyl]methyl]acetamide (3)

Borane dimethylsulfide (13.8 mL, 145 mmol) was added dropwise to an anhydrous THF (483 mL) solution of 4-[[(2,2,2-trifluoroacetyl)amino]methyl]benzoic acid (2) (11.9 g, 48.3 mmol) while maintaining an internal temperature of 0 °C. After complete addition the ice bath was removed and the mixture was stirred at room temperature overnight. The reaction was quenched with MeOH (100 mL) and stirred at room temperature for an additional 1 hr. The volatiles were removed and the residue taken up in EtOAc. Impurities were removed by successive washes with 1M NaOH, H_2O, and brine. The organics were then dried over Na_2SO_4, filtered, and concentrated

in vacuo. Subsequent column chromatographic purification of the residue, eluting with 20:1 (CH_2Cl_2:MeOH), provided amide product (3) (10.4 g, 92%) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$-d) δ 7.35 (d, J = 8.2 Hz, 1H), 7.27 (d, J = 8.0 Hz, 1H), 4.67 (s, 1H), 4.50 (d, J = 5.8 Hz, 1H). ^{13}C NMR (125 MHz, CDCl3-d) δ 157.32 (q, J = 36.7 Hz), 141.11, 135.33, 128.29, 127.67, 115.97 (q, J = 287.8 Hz), 64.90, 43.75.



6-(1-methylpyrrolidin-1-ium-1-yl)-7H-purin-2-amine chloride (5)

Neat *N*-methylpyrollidine (4) (7.80 mL, 73.7 mmol) was added to an anhydrous DMF (144 mL) solution of 6-chloro-7H-purin-2-amine (5.00 g, 29.5 mmol) and stirred at 40 °C overnight. The resultant chloride salt (5) (5.45 g, 73%) was collected via vacuum filtration, dried under suction, and used without further purification.

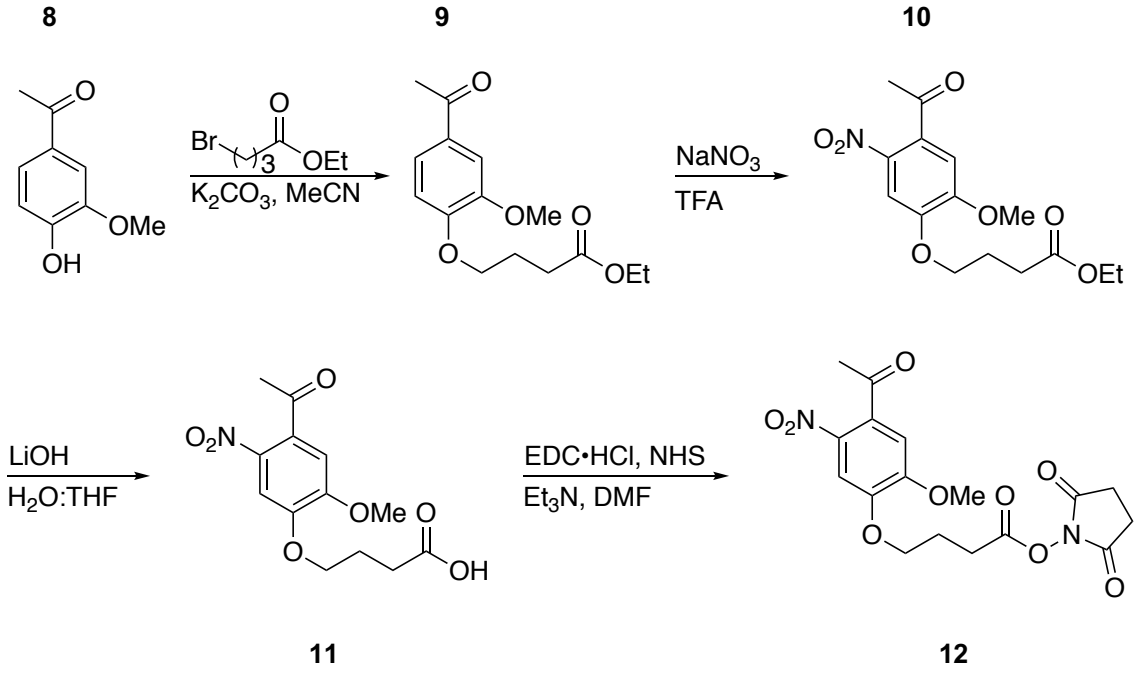
***N*-[[4-[(2-amino-7H-purin-6-yl)oxymethyl]phenyl]methyl]-2,2,2-trifluoroacetamide (6)**

An oven dried round-bottom flask containing 6-(1-methylpyrrolidin-1-ium-1-yl)-7H-purin-2-amine chloride (5) (6.90 g, 27.1 mmol), 2,2,2-trifluoro-*N*-[[4-(hydroxymethyl)phenyl]methyl] acetamide (3) (12.6 g, 54.2 mmol), potassium tert-butoxide (12.2 g, 108 mmol), and 18-c-6 (1.07 g, 4.07 mmol) in 54 mL DMF was stirred for 6 hrs at 50 °C. Upon completion the solvent was evaporated and the crude residue absorbed to silica. Purification was achieved using column

chromatography eluting with a gradient of MeOH in CH_2Cl_2 (2-10 %) to give acetamide product (6) (8.94 g, 90%) as a white solid.

6-[[4-(aminomethyl)phenyl]methoxy]-7H-purin-2-amine (7)

N-[[4-[(2-amino-7H-purin-6-yl)oxymethyl]phenyl]methyl]-2,2,2-trifluoroacetamide (6) (2.59 g, 7.06 mmol) was added to a suspension of K_2CO_3 (4.84 g, 35.0 mmol) in 21 mL of MeOH:H2O (20:1) and stirred vigorously overnight at 50 °C. Upon consumption of the starting material the mixture was filtered through a pad of celite, washing with MeOH. The filtrate was concentrated *in vacuo* and the residue taken up in 10 mL H_2O. While cooling, the pH was adjusted to ~7 with HCl. The resultant precipitate was isolated via suction filtration and washed with cold water to yield amine product (7) (1.77 g, 93%) as a white solid. ^{13}C NMR (12d MHz, DMSO-d^6) δ 159.65, 157.73, 143.72, 140.45, 135.00, 134.67, 128.47, 127.14, 126.98, 66.66, 45.27.



Ethyl 4-(4-acetyl-2-methoxyphenoxy)butanoate (9)

Solid 1-(4-hydroxy-3-methoxyphenyl)ethenone (8) (8.31 g, 50 mmol) was added to a suspension of K_2CO_3 (69.1 g, 500 mmol) in 100 mL of anhydrous MeCN. Ethyl 4-bromobutanoate (14.3 mL, 100 mmol) was added and the mixture stirred overnight at 60 °C. Upon consumption of starting material the mixture was filtered over a pad of celite, washing with cold MeCN. Volatiles were evaporated and the residue recrystallized from Et_2O to provide ester product (9) (13.2 g, 94%) as a white powder.

Ethyl 4-(4-acetyl-2-methoxy-5-nitrophenoxy)butanoate (10)

Using an ice bath, 50 mL of trifluoroacetic acid (50 mL) was cooled to 0 °C prior to the addition of ethyl 4-(4-acetyl-2-methoxyphenoxy)butanoate (9) (12.8 g, 45.7 mmol). Solid $NaNO_3$ (11.6 g, 136 mmol) was added in portions to the stirred mixture maintaining 0 °C. Upon completion the reaction was quenched with 200 mL of H_2O, the resulting precipitate filtered, then dried under suction to afford the nitrated compound (10) (13.8 g, 93%) as a yellow powder. ^{1}H NMR (400 MHz, CDCl3-d) δ 7.61 (s, 1H), 6.74 (s, 1H), 4.21 – 4.12 (m, 4H), 3.95 (s, 3H), 2.54 (t, J = 7.2 Hz, 2H), 2.49 (s, 3H), 2.25 – 2.16 (m, 2H), 1.27 (t, J = 7.2 Hz, 3H).

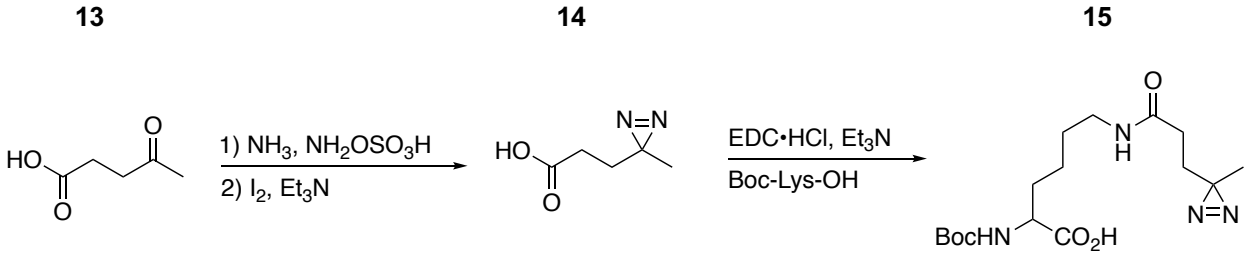
4-(4-acetyl-2-methoxy-5-nitrophenoxy)butanoic acid (11)

A solution of ethyl 4-(4-acetyl-2-methoxy-5-nitrophenoxy)butanoate (10) (2.52 g, 7.76 mmol) in THF (76 mL) was combined with 39 mL of 2M LiOH. The reaction was allowed to stir vigorously for 3 hrs at room temperature. Upon starting material consumption, the reaction was quenched with 1M $NaHSO_4$ (100 mL) and extracted with EtOAc. The combined organic extracts were dried over $NaSO_4$ and filtered. Precipitate that formed upon solvent removal *in vacuo* was

filtered and dried under suction to provide the title acid (12) (1.86 g, 81%) as a pale yellow powder. ^{1}H NMR (400 MHz, $CDCl_3$-d) δ 7.62 (s, 1H), 6.75 (s, 1H), 4.17 (t, J = 6.2 Hz, 2H), 3.95 (s, 3H), 2.63 (t, J = 7.1 Hz, 2H), 2.49 (s, 3H), 2.27 – 2.13 (m, 2H).

(2,5-dioxopyrrolidin-1-yl) 4-(4-acetyl-2-methoxyphenoxy)butanoate (12)

4-(4-acetyl-2-methoxy-5-nitrophenoxy)butanoic acid (11) (4.65 g, 15.6 mmol) was added to an anhydrous DMF (30 mL) solution containing EDC•HCl (4.48 g, 23.4 mmol) and *N*-hydroxysuccinimide (NHS; 2.69 g, 23.4 mmol). The mixture was stirred overnight at room temperature. Addition of chilled Et_2O to the bulk solution resulted in precipitate, which was collected via vacuum filtration and dried under suction to yield NHS-ester (12) (5.47 g, 89%) as a yellow powder. ^{1}H NMR (500 MHz, $CDCl_3$-d) δ 7.63 (s, 1H), 6.75 (s, 1H), 4.21 (t, J = 6.0 Hz, 2H), 3.96 (s, 3H), 2.89 (t, J = 7.3 Hz, 2H), 2.85 (bs, 4H), 2.49 (s, 3H), 2.35 – 2.28 (m, 2H). ^{13}C NMR (125 MHz, CDCl3) δ 200.28, 169.20, 168.18, 154.54, 148.71, 133.29, 108.88, 108.44, 67.66, 56.72, 30.58, 27.62, 25.73, 24.21.



3-(3-methyldiazirin-3-yl)propanoic acid (14)

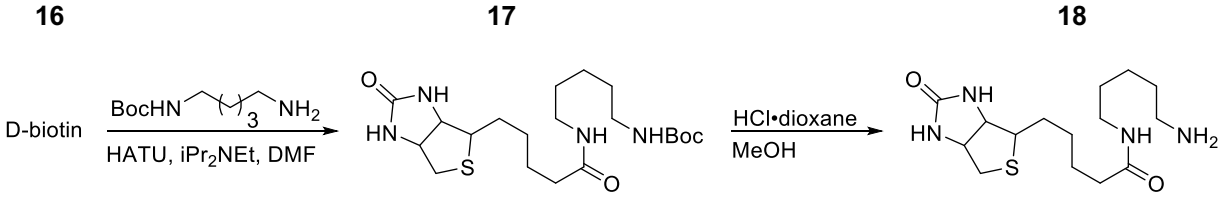
Gaseous ammonia (*ca.* 200 mL) was condensed at -78 °C into an oven dried two-neck round-bottom flask, 4-oxopentanoic acid (13) (11.6 g, 100 mmol) added, then the mixture refluxed at 0 °C for 4 hrs. A suspension of amino hydrogen sulfate (14.0 g, 123 mmol) in anhydrous MeOH

(150 mL) was added via addition funnel over a 45 min period maintaining 0 °C. The heterogenous reaction was then vigorously stirred overnight, allowing the ammonia to slowly evaporate as the temperature rose to room temperature. The resulting slurry was filtered over a pad of celite, washing the solids with MeOH. The solvent was reduced under vacuum (50 mL) to ensure removal of residual ammonia. The crude residue was then diluted with MeOH (100 mL) and cooled to 0 °C before the addition of Et_3N (20.8 mL, 150 mmol). Solid I_2 (25.5 g, 100 mmol) was added in portions until the color of iodine persisted and the reaction stirred for 2 hrs allowing the temperature to rise to room temperature. Upon completion, the volatiles were removed in vacuo and the residue diluted with EtOAc (150 mL). The organics were successively washed with 1 M NaHSO4 (50 mL x 2), 0.5 M $Na_2S_2O_3$ (50 mL), brine (50 mL) then dried over $NaSO_4$. The combined organics were filtered and absorbed to silica then subjected to chromatographic purification eluting with a gradient of MeOH in CH_2Cl_2 (2 – 5%). The title acid (14) (7.29 g, 57%) was isolated as a thin red oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.18 (bs, 1H), 2.18 – 2.14 (m, 2H), 1.68 – 1.60 (m, 2H), 0.95 (s, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 178.66, 29.24, 28.48, 25.05, 19.55.

(2S)-6-[[3-(3-methyldiazirin-3-yl)propanoyl]amino]-2-[(2-methylpropan-2 yl) oxycarbonylamino]hexanoic acid (15)

HATU (910 mg, 2.39 mmol) was added to a 20 mL vial charged with 3-(3-methyldiazirin-3-yl)propanoic acid (14) (323 mg, 2.52 mmol), *i*Pr_2NEt (1.30 mL, 7.87 mmol) in anhydrous DMF (12 mL). The mixture was stirred for 1 hr at room temperature after which Boc-Lys-OH (621 mg, 2.52 mmol) was added in one portion and stirring continued overnight. Upon completion, the reaction was diluted with EtOAc and washed successively with 1 M NaHSO4 (20 mL x 2), H_2O

(20 mL x 2), and brine (20 mL). The organics were filtered and concentrated before being purified by reverse phase HPLC, eluting with a gradient of MeOH in H_2O (0 – 95%). The desired acid product (15) (381 mg, 42%) was obtained in modest yield as a beige solid. ^{1}H NMR (400 MHz, DMSO-d^6) δ 12.4 (s, 1H), 7.83 (t, J = 5.6 Hz, 1H), 7.01 (d, J = 8.0 Hz, 1H), 3.81 (ddd, J = 9.5, 7.9, 4.7 Hz, 1H), 3.00 (q, J = 6.4 Hz, 2H), 1.93 (dd, J = 8.5, 6.9 Hz, 2H), 1.67 – 1.59 (m, 1H), 1.59 – 1.52 (m, 4H), 1.37 – 1.27 (m, 12H), 0.97 (s, 3H). ^{13}C NMR (125 MHz, DMSO-d^6) δ 174.28, 170.50, 155.61, 77.95, 53.43, 38.24, 30.41, 29.88, 29.79, 28.68, 28.23, 25.85, 23.07, 19.34.



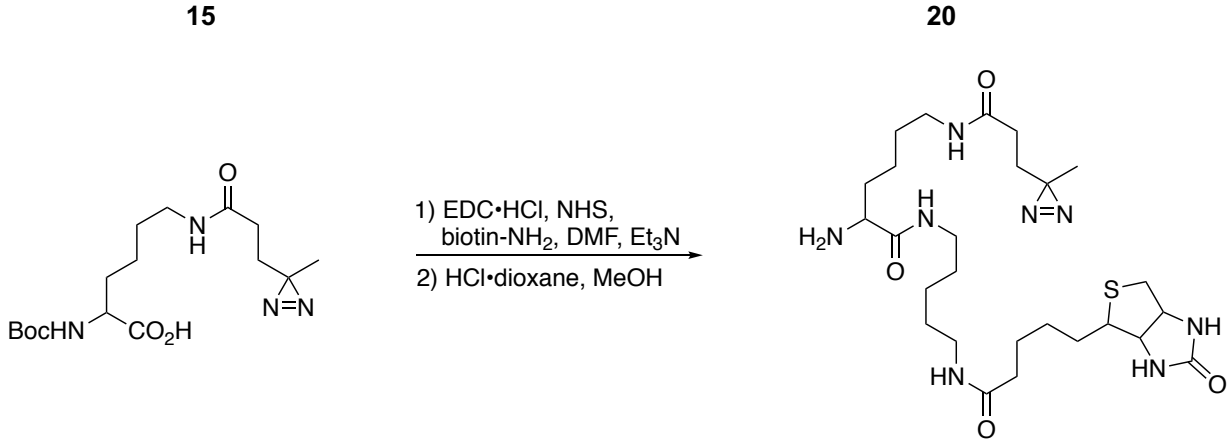
(3aS,4S,6aR)-*N*-[5-[(1,1-dimethylethoxy)carbonyl]aminopentyl]hexahydro-2-oxo-1H-thieno[3,4-d]imidazole-4-pentanamide (17)

A solution of D-biotin (16) (1.95 g, 7.98 mmol), *i*Pr$_2$NEt (4.20 mL, 24.1 mmol) and HATU (6.08 g, 16.0 mmol) in anhydrous DMF (20 mL) was stirred for 1 hr at room temperature. Neat tert-butyl *N*-(5-aminopentyl)carbamate (1.78 g, 8.78 mmol) was added and the reaction stirred overnight. Upon completion Et_2O was flowed in and the resultant precipitate filtered under vacuum. The crude solid was purified by iterative recrystallization from acetone and hexanes then H_2O and acetone. Amide product (17) (2.95 g, 80%) was isolated by vacuum filtration as a white powder. ^{1}H NMR (500 MHz, DMSO-d^6) δ 7.73 (t, J = 5.6 Hz, 1H), 6.76 (t, J = 5.7 Hz, 1H), 6.43 (s, 1H), 6.36 (s, 1H), 4.30 (dd, J = 7.6, 5.0 Hz, 1H), 4.12 (ddd, J = 7.7, 4.4, 1.9 Hz, 1H), 3.09 (ddd, J = 8.7, 6.1, 4.4 Hz, 1H), 2.99 (q, J = 6.6 Hz, 2H), 2.87 (q, J = 6.7 Hz, 2H), 2.82 (dd, J = 12.4, 5.1 Hz, 1H),

2.57 (d, J = 12.4 Hz, 1H), 2.03 (t, J = 7.4 Hz, 2H), 1.60 – 1.57 (m, 1H), 1.54 – 1.41 (m, 3H), 1.36 – 1.17 (m, 20H). ^{13}C NMR (125 MHz, DMSO-d^6) δ 171.78, 162.71, 155.58, 77.32, 61.05, 59.19, 55.46, 38.35, 35.23, 29.20, 28.89, 28.30 (4C), 28.25, 28.06, 25.36, 23.74.

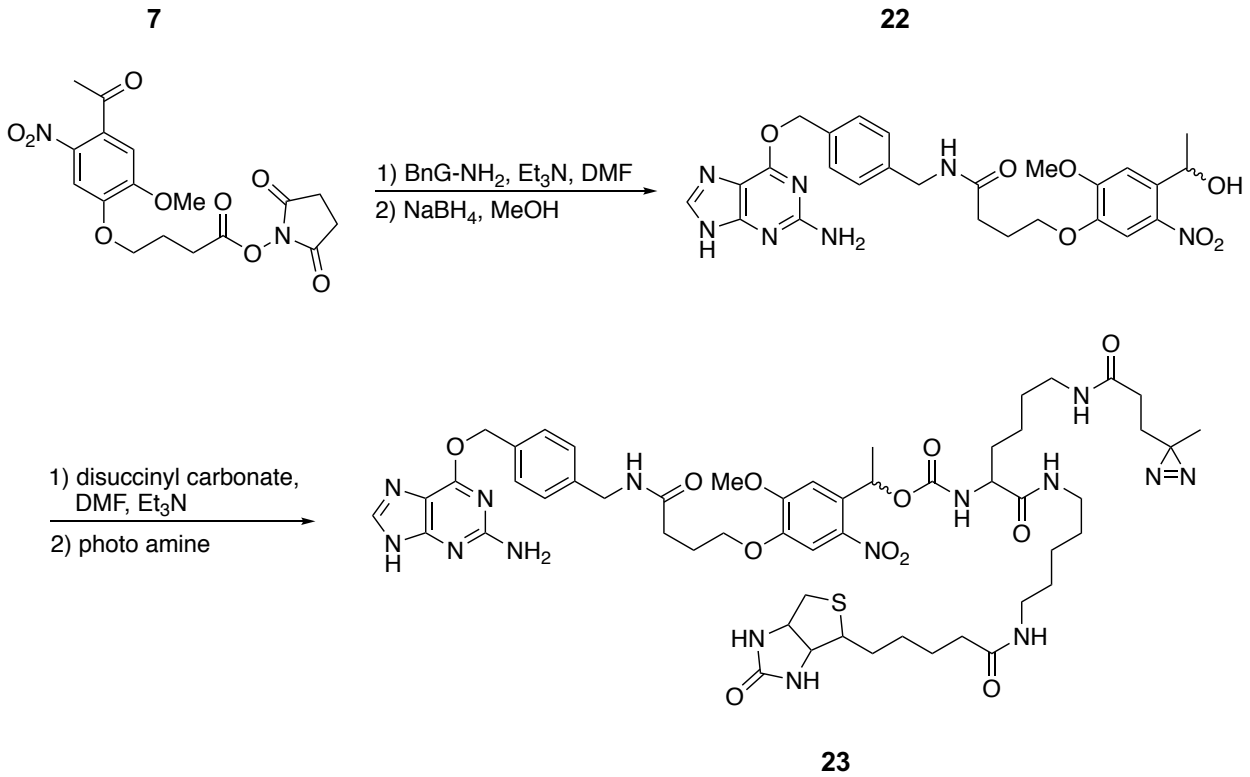
_N_-(5-aminopentyl)-5-(2-oxo-1,3,3a,4,6,6a-hexahydrothieno[3,4-d]imidazol-4-yl) pentanamide (18)

(3aS,4S,6aR)-_N_-[5-[(1,1-dimethylethoxy)carbonyl]aminopentyl]hexahydro-2-oxo-1H-thieno[3,4-d]imidazole-4-pentanamide (17) (1.80 g, 4.20 mmol) was dissolved with 4 M HCl in 1,4-dioxane (21 mL) and MeOH (1 mL). The mixture stirred at room temperature for 1 hr after which the volatiles were evaporated. The crude residue was taken up in 20 mL of MeOH and chilled with an ice bath before the addition of 7 M ammonia in MeOH (6 mL). The organics were reduced under vacuum and filtered to remove solids. The filtrate was concentrated and the residue recrystallized from MeOH and Et_2O to afford amine product (18) in quantitative yield, which was used without further purification.



Tert-butyl *N*-[6-[[3-(3-methyldiazirin-3-yl)propanoyl]amino]-1-oxo-1-[5-[5-(2-oxo-1,3,3a,4,6,6a-hexahydrothieno[3,4-d]imidazol-4-yl)pentanoylamino]pentylamino]hexan-2-yl]carbamate (19)

A separate round-bottom flask containing (2S)-6-[[3-(3-methyldiazirin-3-yl)propanoyl] amino]-2-[(2-methylpropan-2 yl) oxycarbonylamino]hexanoic acid (15) (712 mg, 2.00 mmol), HATU (912 mg, 2.40 mmol), and Et_3N (417 μL, 3.0 mmol) in anhydrous MeCN (10 mL) was stirred at room temperature for 4 hrs. Afterward, an anhydrous DMSO (7 mL) solution of *N*-(5-aminopentyl)-5-(2-oxo-1,3,3a,4,6,6a-hexahydrothieno[3,4-d]imidazol-4-yl)pentanamide (18) (788 mg, 2.40 mmol) was flowed in and the mixture stirred overnight. Upon completion, the reaction was diluted with EtOAc and washed successively with 1 M $NaHSO_4$ (20 mL), H_2O (20 mL x 2), brine (20 mL) then dried over $NaSO_4$. The crude residue was purified by preparatory reverse phase HPLC, eluting with a gradient of MeOH in H_2O (50 – 95%). Carbamate product (19) (453 mg, 34%) was isolated as a waxy solid. ^{1}H NMR (500 MHz, CD3OD) δ 4.50 (dd, J = 7.9, 4.8 Hz, 1H), 4.31 (dd, J = 7.9, 4.4 Hz, 1H), 3.95 (dd, J = 8.9, 5.2 Hz, 1H), 3.24 – 3.13 (m, 7H), 2.93 (dd, J = 12.7, 5.0 Hz, 1H), 2.71 (d, J = 12.7 Hz, 1H), 2.20 (t, J = 7.4 Hz, 2H), 2.07 (dd, J = 8.5, 6.9 Hz, 2H), 1.79 – 1.30 (m, 29H), 1.01 (s, 3H). ^{13}C NMR (125 MHz, CD3OD) δ 175.80, 175.03, 174.29, 165.98, 157.68, 80.47, 63.31, 61.54, 56.99, 56.13, 41.05, 40.17, 40.13, 40.06, 36.78, 33.06, 31.49, 31.31, 29.97, 29.93, 29.76, 29.46, 28.74, 26.89, 26.33, 25.12, 24.26, 19.76.



Ketone product (21)

A DMF (16 mL) solution of (2,5-dioxopyrrolidin-1-yl) 4-(4-acetyl-2-methoxyphenoxy) butanoate (12) (2.30 g, 5.83 mmol) was added to a suspension of 6-[[4-(aminomethyl)phenyl] methoxy]-7H-purin-2-amine (7) (1.50 g, 5.55 mmol) and *i*Pr_2NEt (2.75 mL, 16.6 mmol) in anhydrous DMF (25 mL). The resulting homogenous mixture was stirred at room temperature for 5 hrs. Upon consumption of starting material, chilled Et_2O was flowed into the reaction and supernatant decanted from resultant oil. The crude residue was taken up in CH_2Cl_2 and sonicated until formation of precipitate, which was isolated by filtration. The solids were dried under suction to yield pure ketone (21) (1.32 g, 43%) as a beige powder. ^{1}H NMR (500 MHz, DMSO-d^6) δ 8.42 (t, J = 5.9 Hz, 1H), 8.19 (s, 1H), 7.62 (s, 1H), 7.46 (d, J = 7.7 Hz, 2H), 7.27 (d, J = 7.8 Hz, 2H), 7.23 (s, 1H), 5.48 (s, 2H), 4.28 (d, J = 5.9 Hz, 2H), 4.12 (t, J = 6.4 Hz, 2H), 3.92 (s, 3H), 2.51 (s, 3H), 2.33 (t, J = 7.4 Hz, 2H), 2.00 (p, J = 6.9 Hz, 2H). ^{13}C NMR (125 MHz, DMSO) δ 199.35,

171.42, 159.13, 158.88, 154.76, 154.19, 153.28, 148.58, 139.83, 138.36, 134.59, 131.11, 128.77, 127.31, 109.84, 107.96, 68.55, 67.38, 56.67, 41.89, 31.47, 30.73, 30.05, 24.58.

Alcohol product (22)

Solid $NaBH_4$ (113 mg, 3.00 mmol) was added in portions to a room temperature methanolic solution (10 mL) of ketone product (21) (550 mg, 1.00 mmol) with vigorous stirring. Upon completion, the reaction was passed through a plug of silica to remove inorganics. The filtrate was concentrated in vacuo and the residue sonicated in Et_2O. The resultant precipitate was isolated via vacuum filtration then dried under suction to yield alcohol product (22) (395 mg, 72%) as a yellow powder. ^1H NMR (400 MHz, CD_3OD) δ 7.84 (bs, 1H), 7.52 (s, 1H), 7.42 (d, J = 7.9 Hz, 2H), 7.37 (s, 1H), 7.26 (d, J = 8.0 Hz, 2H), 5.49 (s, 2H), 5.44 (q, J = 6.2 Hz, 1H), 4.36 (s, 2H), 4.04 (t, J = 6.1 Hz, 2H), 3.91 (s, 3H), 2.46 (t, J = 7.3 Hz, 2H), 2.12 (p, J = 6.6 Hz, 2H), 1.45 (d, J = 6.3 Hz, 3H).

PP1 (23)

A solution of tert-butyl *N*-[6-[[3-(3-methyldiazirin-3-yl)propanoyl]amino]-1-oxo-1-[5-[5-(2-oxo-1,3,3a,4,6,6a-hexahydrothieno[3,4-d]imidazol-4-yl)pentanoylamino]pentylamino]hexan-2-yl]carbamate (19) (270 mg, 405 μmol) in MeOH (1 mL) was added to 4 M HCl•dioxane (2 mL) at 0 °C. The temperature was allowed to come to room temperature while stirring for 2 hrs. Upon disappearance of starting material, volatiles were evaporated and the residue left to dry under vacuum overnight. The crude hydrochloride (20) was used without further purification.

In a separate flask, alcohol product (22) (57.9 mg, 105 μmol) and Et_3N (41.0 μL, 294 μmol) were stirred at room temperature in anhydrous MeCN (10 mL) while solid disuccinyl carbonate

(DSC; 75.3 mg, 294 μmol) was added. The mixture was stirred overnight, after which additional Et_3N (22 μL, 157 μmol) was added. A solution of amine product (20) in DMSO (1 mL) was added and the mixture stirred until consumption of starting material. The volatiles were removed under vacuum and the residue purified by reverse phase HPLC eluting with a gradient of MeOH in H_2O (50 – 95%). The proximity probe PP1 (23) (7.60 mg, 6%) was isolated as a yellow solid and characterized as mixture of diastereomers.

Cell culture

HEK293T and HeLa cell lines were purchased from ATCC. HeLa, HEK293T, and stably expressed HEK293T cells were propagated in DMEM (Corning) or in RPMI 1640 with 2 mM L-glutamine (HyClone) supplemented with 10% fetal bovine serum (FBS, Corning) and 1% penicillin-streptomycin (Thermo Fisher Scientific). All cell lines were grown at 37 ˚C in a 5% CO_2 humid incubator

Western blot

Cells were harvested by scraping, pelleted by centrifugation, washed twice with PBS and lysed in 8 M urea, 50 mM NH_4HCO_3 and EDTA-free complete protease inhibitor (Roche), pH 8.0, at 4 ˚C. Cells were sonicated (Fisher Scientific FB-505), insoluble debris cleared by centrifugation, and the supernatant was diluted into 4X Laemmli buffer containing 50 mM dithiothreitol (DTT) or 6% beta-mercaptoethanol (βMe) as reducing agents. Samples were prepared for SDS-PAGE by heating to 95 ˚C for 5 minutes, cooled to room temperature, resolved on NuPAGE Novex 4-12% Bis-Tris Protein Gels (Invitrogen) or 10% SDS-PAGE gel, and transferred onto nitrocellulose

membranes by standard western blotting methods. Membranes were blocked in 2% BSA in TBS containing 0.1% tween-20 (TBST) and probed with primary and secondary antibodies. Primary antibodies used in this study include: anti-FLAG-M2 (1:1000, F1804, Sigma Aldrich), anti-ACTB (1:1000, 4790, Cell Signaling), and Streptavidin (1:10000, 92632230, LI-COR). Secondary donkey anti-rabbit, donkey anti-goat, and donkey anti-mouse (Licor), were used at 1:10,000 dilution in 2% BSA-containing TBST and incubated for 1 hour prior to washing and imaging on a Licor infrared scanner. Densitometry measurements were performed with ImageJ software.

Cell viability measurements

$5x10^3$ HEK293T cells stably expressing FLAG-SNAP were plated in 50 μL of DMEM growth medium in white 96-well plates. After 24 hours of growing, 50 μL of PP1 diluted to the indicated concentrations (0, 0.5, 1, 5, 15, 45 μM in sextuplicate) were added to each well in a total volume of 100 μL DMEM. After 2 hours incubation, cell viability measurements were recorded on an Envision plate reader after the addition of 100 μL of a Cell Titer Glo solution (Promega, diluted 1:4 in PBS). Plate was imaged using Synergy Neo HST plate reader (BioTek). Relative viabilities are reported as a fraction relative to the same dose of compound treatment without PP1.

Circular polymerase extension cloning (CPEC) construction of mammalian plasmids

All PCR reactions were performed using NEB Q5 high-fidelity polymerase (M0491S) and Promega dNTP mix (U1515). Initial constructs were generated using the pSnapf vector from NEB (#N9183S), the pFlag-Keap1 vector from Addgene (#28023) along with the following CPEC primers purchased from IDT (Table 3.1).

The constructs described above were subcloned into the pLenti6/V5-p53_R273H Addgene (#22934) vector for lentiviral transduction using the following CPEC primers purchased from IDT (Table 3.2).

Generation of stable cell lines

Mammalian cells stably expressing the Keap1 SNAP-Tag fusions were obtained by co-transforming a 6 cm plate of HEK293T cells with 0.1 µg pCMV-VSV-G (Addgene #8454), 0.9 µg pCMV delta R8.2 (Addgene #12263), and 1.0 µg of each primer (Table 3.2). The resultant viral media was collected at 24 and 48 hours, passed through a 0.45-micron filter, and diluted with serum-free DMEM containing 8 µg/mL polybrene (Sigma), final concentration. Viral transduction was achieved by culturing a separate population of HEK293T in the diluted viral media for 24 hours. Afterward, the viral media was removed and the transduced cells grown in full DMEM containing 8 µg/mL blasticidin (Gibco). Stable incorporation of the transgene was confirmed by western blotting using monoclonal anti-FLAG M2 antibody (Sigma). Stable expression of SnapFlag, derived from pSnapf vector (NEB #N9183S), was achieved through chemical selection of transiently transfect HEK293T cells with 500 µM G418 (Gibco).

Photocleavage assay

Recombinant purified His_6-SNAP-Tag (20 µg) was incubated in 200 µL DPBS containing 5 µM **PP1** for 1 hour at 37 °C. Afterward, the reaction was divided into six 30 µL aliquots, placed on ice, and irradiated with 365 nm light using a spectrolinker XL-1500a (Spectroline). Samples were successively removed from irradiation after 0, 0.5, 1, 2.5, 5, 10 minutes of irradiation, diluted

with loading buffer, and run on SDS-PAGE gel. Photocleavage was judged by in-gel fluorescence visualized using a ChemiDoc imaging system (Bio-Rad) scanning for fluorescein.

P3-probe dose response assay

Each well of a 6-well plate was seeded with 300,000 HEK293T cells stably expressing either Snap-Flag or Keap1-Snap. After reaching ~90% confluency the growth media was removed, cells washed with DPBS, and treated with varying concentrations (0, 0.5, 1, 5, 15, 45 μM) of photoproximity probe **PP1** in 500 μL serum-free DMEM for 2 hours at 37 °C. Post treatment, media was aspirated and non-reacted probe washed out with 3 mL full DMEM over 40 minutes, changing the media twice. Cells were harvested, washed, and lysed in DPBS containing protease inhibitor using a tip sonicator. Lysate was normalized to 1 mg/mL then diluted with loading buffer and run on SDS-PAGE gel that was transferred by western blot to nitrocellulose. The nitrocellulose membrane was stained for biotin and Flag-Tag using streptavidin IR dye (Li-cor #926-32230) and anti-Flag (Sigma # F3165, Li-cor #926-68072). Fluorescence images were collected using an Odyssey infrared imager (Li-cor).

Proximity-labeling assays

Clarified cell lysate from HEK293T cells either transiently or stably expressing SNAP-Tag (NEB #N9183S) or Snap-Flag was normalized to 1 mg/mL and 250 μL aliquots incubated in a microcentrifuge tube with either 500 nM of **PP1**, **PP2**, or DMSO for 1 hour at 37 °C. Washed anti-FLAG M2 affinity gel, 40 μL of a 50% slurry, was transferred to each reaction in 750 μL of DPBS and the resultant suspension left to rotate overnight at 4 °C (Sigma #A2220). Samples were then spun down at 4000 x g, supernatant aspirated, and the resin washed with 1M urea in DPBS

(1 mL x 8) followed by DPBS (1 mL). After being resuspended in 100 µL of DPBS, the samples were placed on ice where they were irradiated with 365 nm light for 10 minutes (Spectrolinker XL-1500a). Once irradiated, the beads were washed with 100 mM pH 3.5 glycine buffer (100 µL x 2) and DPBS (1 mL). To elute the remaining SnapFlag and resin-bound FLAG-antibody the beads were boiled for 5 minutes at 95 °C in 20 µL 4x-loading buffer containing 8% SDS and 400 mM DTT. After which, 60 µL of DPBS was added and the suspension boiled at 95 °C for an additional 5 minutes. The samples were then run on SDS-PAGE gel and transferred to nitrocellulose where they were stained with anti-mouse IR dye (Li-cor #926-68072) and streptavidin IR dye (Li-cor #926-32230). Bands corresponding to biotinylated proteins and the FLAG antibody were visualized using an Odyssey infrared imager (Li-cor).

SILAC cell culture methods and proteomic sample preparation

SILAC labeling was performed by growing cells for at least five passages in lysine- and arginine-free SILAC medium (RPMI, Invitrogen) supplemented with 10% dialyzed fetal calf serum, 2 mM L-glutamine and 1% P/S. “Light” and “heavy” media were supplemented with natural lysine and arginine (0.1 mg/mL), and ^{13}C-, ^{15}N-labeled lysine and arginine (0.1 mg/mL), respectively.

General protein digestion for LC-MS/MS analysis was performed by diluting protein (e.g. whole lysate or enriched proteins) in digestion buffer (8 M urea, 50 mM NH_4HCO_3, pH 8.0), followed by disulfide reduction with DTT (10 mM, 40 minutes, 50 ˚C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5 mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin

(~1:100 enzyme/protein ratio; Promega) at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted on ZipTip C18 tips (100 µL, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized. Lyophilized peptides were dissolved in LC-MS/MS Buffer A (H_2O with 0.1% formic acid, LC-MS grade, Sigma Aldrich) for proteomic analysis.

Sample preparation and streptavidin enrichment of proteins biotinylated by P3 profiling

Quantitative proximity labeling study with SILAC quantitative proteomics was performed with "heavy" and "light" labeled HEK293T cells expressing KEAP SNAP fusion constructs. SILAC-labeled cells, grown to 80-90% confluency in 10 cm cell-culture treated plates (Denville) each, were incubated with DMSO alone (light cells) or PP1 probe (15 µM, heavy cells) for 2 hours in serum-free SILAC RPMI. After incubation, excess probes in "heavy" cells were pulled out by a replacement to "heavy" growth media in every 10 minutes for 2 times. All cells were then incubated in 2 mL cold PBS, UV irradiated using a Spectroline XL-1500A instrument for 15 minutes, scraped, washed with cold PBS (2x), and tiny aliquots of cells (20 µL out of 500 µL resuspended cells in PBS) from each cell plate were taken for the analysis of bulk proteome. These "heavy" and "light" cell aliquots were combined and digested by general protein digestion protocol described above.

The rest of cells (480 µL out of 500 µL resuspended cells in PBS) were pelleted and then lysed in RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 1% Triton X-100, 0.5% deoxycholate, pH 7.4) supplemented with EDTA-free complete protease inhibitor (Roche) and 1 mM DTT, at 4 ˚C. After sonication, insoluble debris was cleared by centrifugation (17,000 *g*, 10 min). Streptavidin

C1 magnetic beads (30 µL slurry, 65001, Invitrogen) were washed twice with RIPA buffer, and each cell lysate was separately incubated with the magnetic beads with rotation overnight at 4 ˚C. The beads were subsequently washed five times with 0.5 mL of RIPA lysis buffer containing 1 mM DTT, combined together, then washed once with 1 mL of 1 M KCl, four times with 0.5 mL PBS, and two times with 2 M Urea in 25 mM ammonium bicarbonate. 500 µL of 6 M Urea in 50 mM ammonium bicarbonate was then added to the beads, and samples were reduced on resin by TCEP (10 mM final), with orbital shaking, for 20 minutes at 65 ˚C. Samples were then alkylated by adding iodoacetamide (20 mM final), covered from the light and with orbital shaking, for 40 minutes at 37 ˚C. The streptavidin magnetic beads were collected, washed once with 2 M Urea in 25 mM ammonium bicarbonate, and the buffer exchanged to 2 M Urea in 25 mM ammonium bicarbonate supplemented with 1 mM $CaCl_2$. Enriched proteins were digested on bead by the incubation of 2 µg sequencing grade trypsin overnight at 37 ˚C. Following trypsinization, supernatant was collected, acidified with HPLC grade formic acid (2% final, pH 2-3), and peptides were desalted as indicated above.

Proteomic LC-MS/MS and data analysis

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra high-pressure LC system (ThermoFisher) using a PepMap RSLC C18 column heated to 45˚C (column: 75 µm x 50 cm; 2 µm, 100 Å) coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). Digested peptides (500 ng - 1 µg) in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% Formic acid acetonitrile) at 300 nL/min: 2-2% buffer B over 5 minutes, 2-25% buffer B over 170 minutes, 25-40% buffer B over 40 minutes, 40-90% buffer B over 10 minutes, 90-90% buffer B over 5 minutes, 90-2% buffer

B over 5 minutes, 2-2% buffer B over 5 minutes, 2-90% buffer B over 5 minutes, 90-90% buffer B over 3 minutes, 90-2% buffer B over 5 minutes, 2-2% buffer B over 3 minutes, 2-90% buffer B over 5 minutes, 90-90% buffer B over 5 minutes, 90-2% buffer B over 5 minutes, and 2-2% buffer B over 3 minutes. MS/MS spectra were collected from 0 to 240 minutes using a data-dependent, top 10 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 375-1500 m/z, maximum IT of 60 ms, AGC target of 1e6, and data collection in profile mode. MS2 scans was performed by HCD fragmentation with a resolution of 30,000, AGC target of 1e5, maximum IT of 60 ms, NCE of 27, MSX count of 1, and data type in centroid mode. Isolation window for precursor ions was set to 2.0 m/*z* with isolation offset of 0.0 m/z. Peptides with charge state 1 and undefined were excluded and dynamic exclusion was set to twenty seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.20kV and ionization chamber temperature of 275 ˚C.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database (UniProt_Human_reviewed_04-10-2017.fasta). Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation); two total differential modification sites per peptide, including oxidized methionine (+15.9949); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting

statistics were performed with a Δmass cutoff = 15 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%, peptide modification requirement (-m) was set to 1, and spectra display mode (-t) was set to 1. SILAC searchers were performed as above with "light" and "heavy" database searches of MS1 and MS2 files by including static modification of +8.014168 for lysine and +10.0083 for arginine in a parallel heavy search. SILAC quantification was performed using the QuantCompare algorithm, with a mass tolerance of 10 p.p.m. or less in cases where co-eluting peptide interfere. In general all quantified peptides has mass error within 3 p.p.m..

Quantitative proteomic data analyses of enriched proteomic samples

The SILAC ratios of the proteins from enriched proteomic samples were normalized by the median SILAC ratio of the corresponding bulk proteomic sample. The overall normalized SILAC data, from three biologically independent batches and three technical replication LC-MS/MS runs of each batch, were combined. The mean SILAC ratios of each protein were converted to Log_2 values, and *P* values were calculated by univariate two-sided t-test with a group of unnormalized SILAC ratios of the protein from enriched samples and a group of median SILAC ratios of the bulk samples followed by the conversion of *P* values to $-Log_{10}$ values. The volcano plots were plotted with x-axis of Log_2 SILAC values y-axis of $-Log_{10}$ *P* values, and the proteins which passed the filter (Probe-to-DMSO ratio > 2, $P < 0.05$) were considered as actual enriched proteins with the P3 profiling platform.

3.5 Tables

Plasmid	Primers
Keap1-SNAP-Flag3x	**F1** = ATG GAC AAA GAC TGC GAA ATG AAG CGC ACC ACCC **R1** = ACC CAG CCC AGG CTT GCC CA **F2** = GGC AAG CCT GGG CTG GGT GAC TAC AAA GAC CAT GAC GGG GAT TAT AAA GAT CAT GAC AT **R2** = GTG CGC TTC ATT TCG CAG TCT TTG TCC ATG CTT CCG CCG CCG CGG CCG CCA CAG GTA CA
SNAP-Keap1-Flag3x	**F1** = ATG CAG CCA GAT CCC AGG CCT AGC **R1** = GAT ATC TGC AGA ATT CCA CCA CAC TGG ACT AGT GGA TCC **F2** = CCAGTG CCA GTG TGG TGG AAT TCT GCA GAT ATC ATG GAC AAA GAC TGC GAA ATG AAG CGC ACC AC **R2** = GCC GCC TGG GAT CTG GCT GCA TGC TCC CTC CGC CGC CAC CCA GCC CAG GCT TGC CC

Table 3.1 Primers used for generation of pSNAP-KEAP1 plasmids

Plasmid	Primers
Keap1-SNAP-Flag3x	**F1** = TAG TAA TGA GTT TGG ATT TAA TTC TGT GGA ATG TGT GTC AGT TAG GG **R1** = GGT GAA GGG ATC AAT TCC ACC ACA CTG G **F2** = GGT GGA ATT GAT CCC TTC ACC ATG CAG CCA GAT CCC AGG **R2** = CAC ATT CCA CAG AAT TAA TTC CAA ACT CAT TAC TAC TTG TCA TCG TCA TCC TTG TAG TCG
SNAP-Keap1-Flag3x	**F1** = TAG TAA TGA GTT TGG AAT TAA TTC TGT GGA ATG TGT GTC AGT TAG GG **R1** = GGT GAA GGG ATC AAT TCC ACC ACA CTG G **F2** = GGT GGA ATT GAT CCC TTC ACC ATG GAC AAA GAC TGC GAA ATG AAG C **R2** = CAC ATT CCA CAG AAT TAA TTC CAA ACT CAT TAC TAC TTG TCA TCG TCA TCC TTG TAG TCG

Table 3.2 Primers used for generation of pLenti6/SNAP-KEAP1 plasmids

CHAPTER 4

SILAC surface map: profiling of protein-ligand interactions in live cells*

We are preparing for the publication of the work described in this chapter.

4.1 Introduction

The small molecules that were designed to regulate target proteins involved in critical cell signaling pathways, such as intervening oncogene expressions and rewiring metabolic fluxes, have been widely utilized for studying functions and behaviors of the proteins of interest (POIs) in cells and pharmaceutically applied for human diseases. The advance of cell-based high throughput screening (HTS) assays[225] with library compounds and structure-activity-relationship (SAR) studies[226] has led the identification of the natural ligands and artificial inhibitors that potently regulate the target cell signaling pathways.

Understanding the structural and behavioral alterations of the POI by protein-ligand interactions necessitates our structural insights into the binding sites of the ligand to its interaction partner. The traditional, primary, and predominant approach to identify the non-covalent ligand

* The author's contribution to the studies presented in this chapter: The author conducted all experiments described in this chapter. The author designed, synthesized, and characterized the chemical probes. The author designed, performed, and analyzed biochemical assays, cell-based biological experiments, and proteomic experiments. The author wrote the manuscript with Raymond E. Moellering for the publication of this work.

binding site of the POI is X-ray crystallography. Since the first crystal structure of the protein was successfully solved in 1958 by this approach[227], numerous protein structures, as well as protein-ligand bindings[228], have been revealed. Another basic approach to address this issue is NMR spectroscopy[229] which is capable of providing information of the structure and dynamics of proteins in solution. Thereby, studies of the protein-ligand interactions have also been performed with this method[230]. Recently, various membrane protein-ligand binding structures have been identified with the cryogenic electron microscopy (Cryo-EM)[231]. Each structural characterization means possesses its unique strengths and disadvantages. However, *in vitro* studies of protein-ligand interactions potentially bias the actual protein-ligand interactions in live cells. Indeed, applications of each method to depict a structure are largely restricted by the nature of the POI and usually takes long period of time to collect data.

The combination of high-resolution mass spectrometry, complete map of human genome database, and unique chemical functionalities opened the new field 'chemical proteomics' (or called 'chemoproteomics') which has been introducing diverse technologies to address the fundamental biological questions into protein natures in live cells, as well as discovery of the actual targets of novel xenobiotic molecules. Various label- and label free technologies have been developed, demonstrating that both approaches successfully identified the targets of inhibitors in live cells by global manner[153,154,232]. The common 'label' chemoproteomic strategy for studying non-covalent protein-ligand interactions is structural modification of the ligand to generate a ligand analogue probe which contains photoreactive (e.g. diazirine) and tag (e.g. biotin, alkyne) substrates[233]. Thus, the ligand analogue probe specifically enriches the targets *in situ* after light activation followed by biotin-streptavidin pulldown. Indeed, 'label' approach in chemoproteomics provides a promising clue to imply a protein-ligand binding site, as photoreactive substrate of the

ligand analogue probe forms a covalent crosslink to the proximal amino-acid residues of the ligand binding site in the protein and these crosslinked sites are identified by high-resolution LC-MS/MS and following proteomic analyses. Several published studies have shown to identify the ligand binding sites of both natural compounds[234] and artificially designed small molecules[235] by photoreactive probes.

Nonetheless, technical challenges are still remained with the current photoaffinity-chemoproteomics workflow to identify the ligand binding sites: 1) covalent modifications of the amino acid residues by photoreactive groups in the protein are randomly generated, thus identification of the photocrosslinked sites necessitates the differential modification (diff. mod.) searches of all 21 proteinogenic amino acids and 2) intrinsic nature of the ligand analogue probe potentially affects the downstream LC-MS/MS analysis, as covalent adduct formation by a crosslinked probe to the peptides alters the whole size and hydrophilicity of these modified peptides and the vulnerability of the chemical modifications (i.e. fragmentation of the probe-derived modifications under high collision energy in MS) that restrict the diff. mod. searches to find the sites of crosslink.

We previously identified an irreversibly crosslinked site in covalent KEAP1 homodimer, which is formed by a non-enzymatic modification with endogenous glycolytic adduct methylglyoxal, by a SILAC-based quantitative proteomic method based on changes in tryptic peptide rations across the entire protein[149]. The apparent merit of this approach, termed 'SILAC Surface Mapping', is to identify the modified sites of the entire protein in unbiased and global manners, without potentially bringing the technical issues of the current photoaffinity-based site identification approach. Hence, we applied the SILAC surface mapping strategy to the current photoaffinity-MS approach and investigated non-covalent protein-ligand interactions in live cells.

4.2 Results

The key of SILAC surface mapping strategy is to differentiate between unmodified- and modified regions of the entire protein by quantifying the relative population of unmodified, each tryptic peptide from 'heavy' crosslinked-protein and 'light' unmodified peptide digests (Figure 4.1). We wondered if application of the strategy successfully identifies the potential ligand binding sites, thereby we validate the method with two representative probe-ligand interactions: c-ABL/GNF-2 and PGK1/CBR-470-1 bindings.

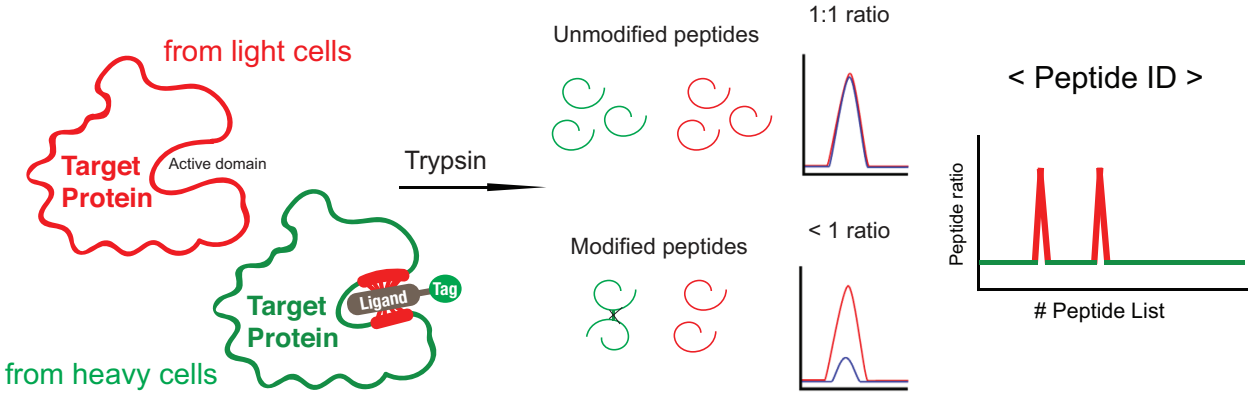


Figure 4.1 SILAC surface mapping assay
Schematic of the SILAC surface mapping assay to identify the ligand interaction sites on the protein with photoaffinity probe and SILAC quantitative proteomics.

4.2.1 Identification of GNF-2 binding site in c-ABL

A Src family non-receptor tyrosine kinase c-ABL, which forms a protein complex with BCR in chronic myeloid leukemia (CML) characterized by the Philadelphia (Ph^+) chromosome and BCR-ABL fusion gene, is a significant substrate of BCR-ABL oncoprotein that has been reported to be involved in abnormal circulation, activation of the RAS oncogene, and proliferation of the

host cells[236-238]. To therapeutically target Ph^+-CML, as well as other c-ABL associated human diseases, small molecule inhibitors of c-ABL have been developed, including Imatinib (trade name Gleevec)[239] which is binding to the pocket of ATP binding site[240] and GNF-2 which binds to the C-terminal myristate pocket enabling the regulation of c-ABL allosterically[241-244]. As binding sites of these synthetic ligands in kinase domain of c-ABL were well studied, thus, we chose c-ABL/GNF-2 binding as a proof-of-principle model of our method.

4.2.1.1 Design of GNF-PAP probe

To design a photoreactive analogue probe of GNF-2, termed 'GNF-PAP', we first determined the site of substitution which employs diazirine and alkyne tag but not significantly interferes its binding to c-ABL. c-ABL/GNF-2 binding was originally depicted by several published studies utilizing both 2D-NMR spectroscopy and X-ray crystallography with recombinant, isolated kinase domain of c-ABL (Figure 4.2a). These results suggested a minimized interruption of c-ABL/GNF-PAP binding by a substitution in amide NH substrate of GNF-2 which was also implied by the study of c-ABL affinity pulldown with GNF-2, an GNF-2 analogue possessing PEG linker connected to the ester moiety of GNF-2 and a resin. Thus, we substituted the tail of GNF-2 with a diazirine-alkyne tag (Figure 4.2b).

To synthesize the GNF-PAP probe, we first synthesized the two distinct parts of the compound – the active GNF substrate and diazirine-alkyne tag – and covalently linked them by amide coupling reaction. GNF-PAP was synthesized by total 10 synthetic steps (3 steps of GNF precursor synthesis (Figure 4.2c), 5 steps of diazirine-alkyne tag synthesis, 2 steps of deprotection - amide coupling reactions) and final product was validated by ^{1}H-NMR and LC-MS measurements.

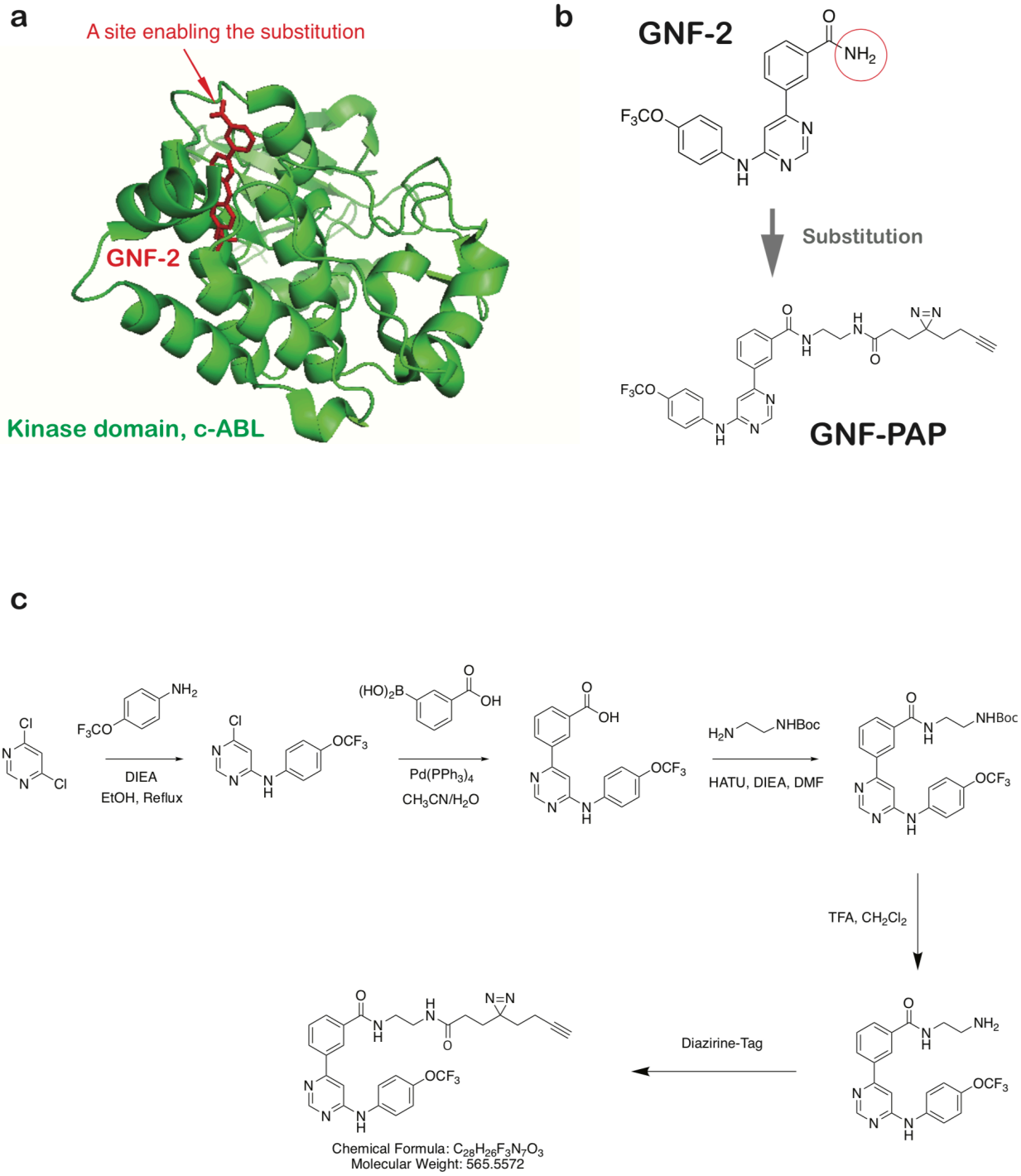


Figure 4.2 Design and synthesis of GNF-PAP
a. c-ABL/GNF-2 crystal structure. Amide NH substrate of GNF-2 is open to the outer space of c-ABL kinase domain myristate pocket (PDB: 3K5V). **b**. Design of GNF-PAP probe. Photoreactive substrate (diazirine) and alkyne tags were employed with the substitution of Amide NH site. **c.** Synthetic scheme of GNF-PAP photo affinity probe. Diazirine-alkyne tag was synthesized separately (see more details in '4.4 Materials and method' section).

4.2.1.2 Validation of GNF-PAP in live cells

We wondered if the substitution to diazirine-alkyne tag affects the potency of GNF-PAP to BCR-ABL expressed cells, may implying a binding efficiency of GNF-PAP to c-ABL protein in cells. To evaluate, we performed a cell viability test with K562 cell, a CML cell-line, treated with increasing doses of GNF-PAP. As confirmed by CTG readout, IC_{50} of the probe was 3.3 μM in K562 cells, less potent than GNF-2 in BaF3 cells (IC_{50} = 340 nM) but still displayed to retain a similar potency (Figure 4.3a and 4.3b).

Next, we sought to test if GNF-PAP successfully generates UV light-activated irreversible crosslinks to its reversible binding partner c-ABL in live cells. To investigate, 10 μM GNF-PAP was treated to FLAG-c-ABL expressed HEK293T cells followed by UV irradiation to the cells to activate the diazirine substrate of the probe. FLAG-c-ABL proteins were mostly observed in the insoluble fractions of PBS lysates, and click chemistry was performed to tag the biotin to the GNF-PAP target proteins. We observed an apparent biotin band near 150 kDa, which was completely overlapped with the FLAG band, in FLAG- and biotin Western blots. Biotin band was not shown in the sample from DMSO treated cells, implying that FLAG-c-ABL protein and is a target of CBR-470-PAP in live cells (Figure 4.3c).

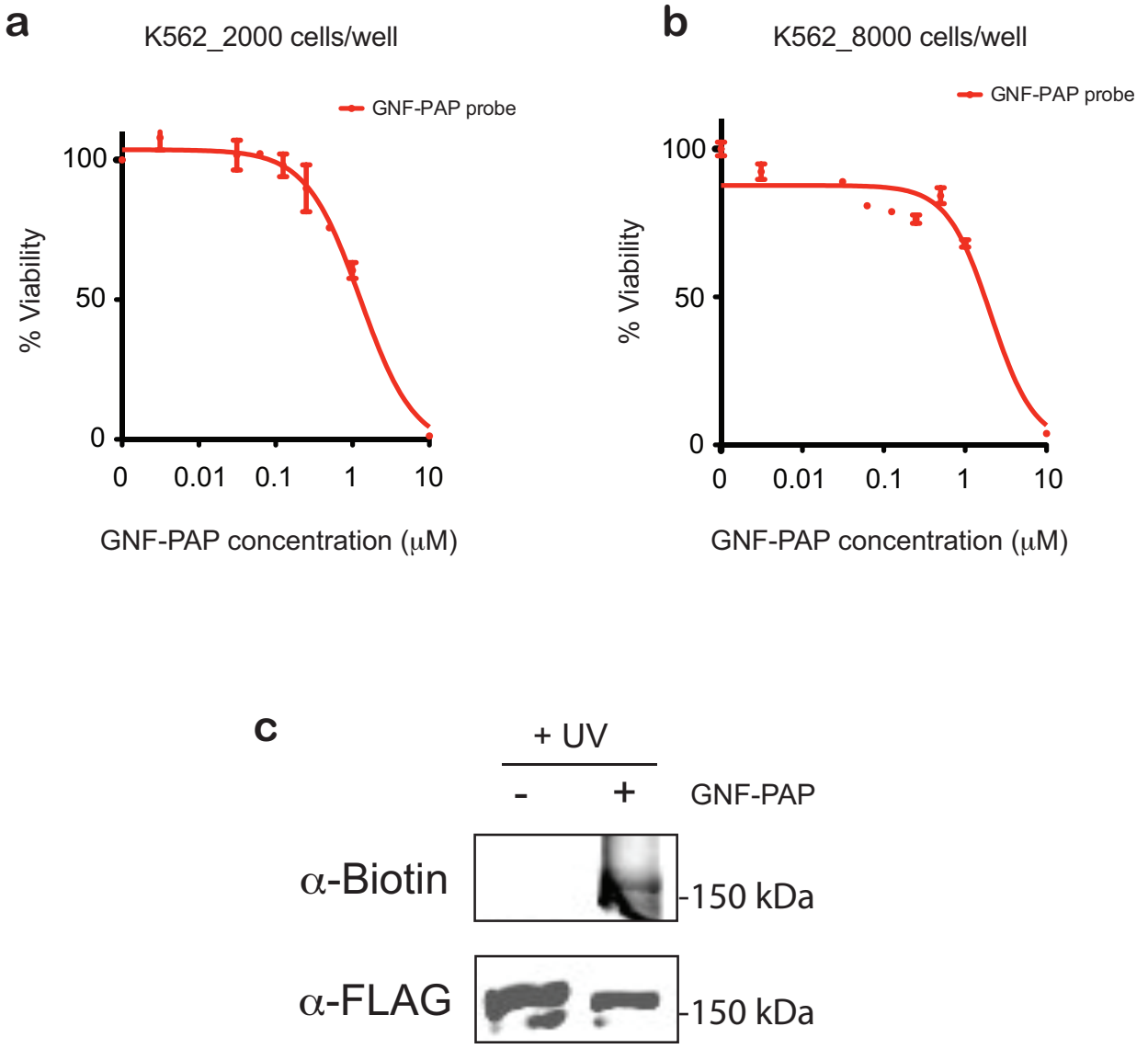


Figure 4.3 *In situ* validation of GNF-PAP
a. b. Viability measurement of K562 cells treated with indicated doses of GNF-PAP photoaffinity probe in 72 hours, each K562 cells were initially plated with 2,000 cells/well **(a)** and 8,000 cells/well **(b)** (n=3 each, mean ± SEM). **c.** Anti-biotin and anti-FLAG Western blot analyses of insoluble proteins of FLAG-c-ABL transfected HEK293T cells treated with GNF-PAP or not, followed by UV irradiation of cells.

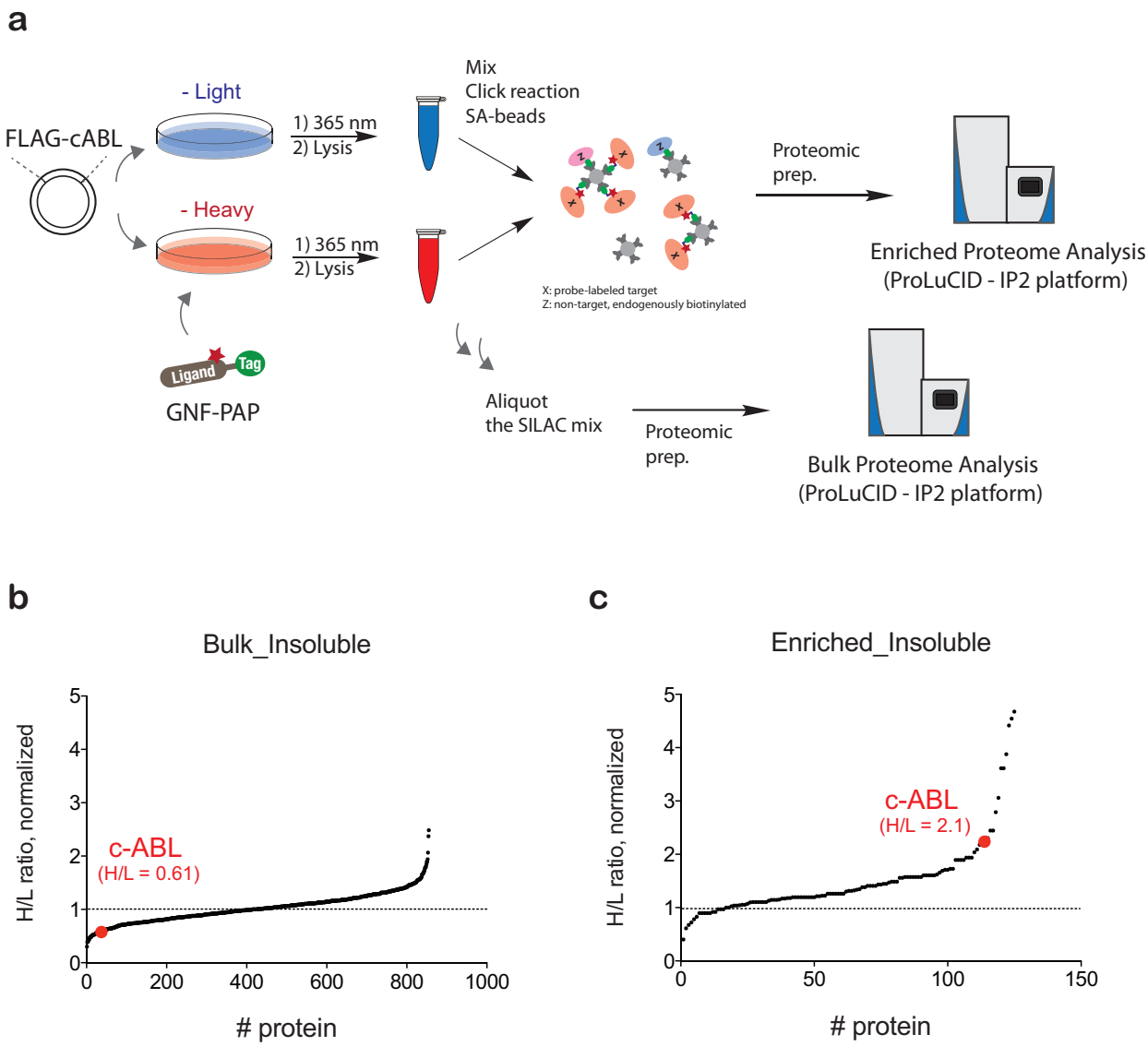


Figure 4.4 Target identification of GNF-PAP with *in situ* photoaffinity-MS
a. Workflow of the CBR-470-PAP global target identification test with SILAC quantitative proteomics. **b.** Global quantification of protein levels in Bulk-insoluble SILAC proteome. Transient transfection of FLAG-c-ABL induced slight fluctuations of the relative protein expressions in cells. **c.** Global quantification of CBR-470-PAP enriched proteins

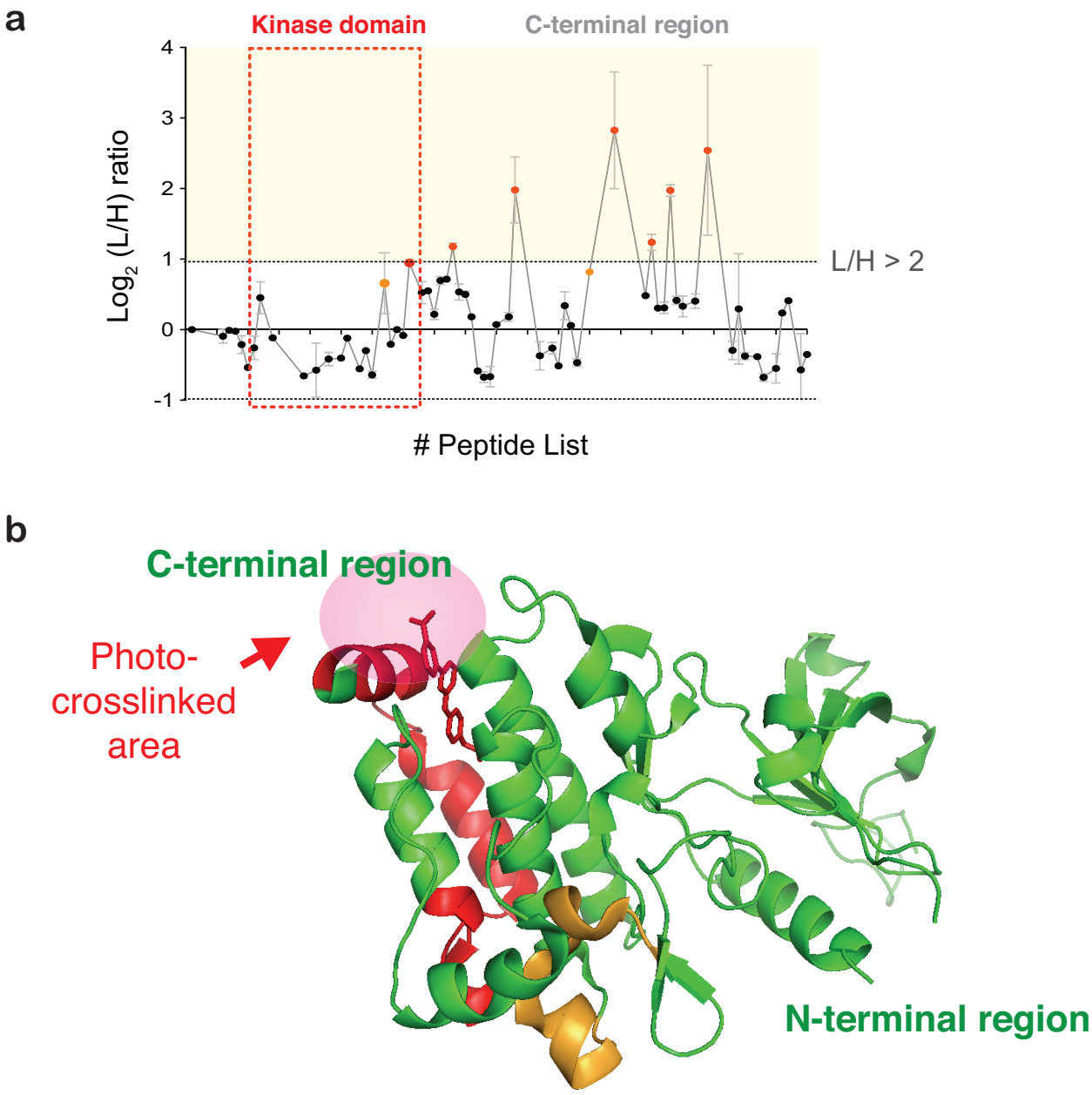


Figure 4.5 *In situ* SILAC surface mapping of c-ABL with GNF-PAP, preliminary test
a. SILAC ratios for individual tryptic peptides of FLAG-c-ABL from DMSO treated 'light' cells and GNF-PAP treated 'heavy' cells, relative to the median ratio of all c-ABL peptides. Tryptic peptides which were reduced by 2-fold or more upon relative to c-ABL median (L/H > 2) were marked with red dots, indicative of possible covalent crosslinks. **b.** Structural depiction of potentially modified stretches of human c-ABL (red) using published x-ray crystal structure of Kinase domain (PDB: 3K5V). Poorly characterized C-terminal regions c-ABL are depicted on their position near myristate pocket of Kinase domain.

In order to identify the targets of GNF-PAP in full-scale proteomics, we employed quantitative proteomics using SILAC HEK293T cells expressed with FLAG-c-ABL. GNF-PAP was treated to the heavy cells and photocrosslinked targets were enriched from the SILAC mixture of insoluble fractions (Figure 4.4a). Among the list of enriched proteins which passed the filter (normalized H/L ratio > 2, total peptide counts > 2), c-ABL was shown as one of most abundant proteins (total peptide counts = 21). Indeed, H/L ratio of the c-ABL in the enriched proteomic sample was significantly increased than H/L ratio of the c-ABL in the bulk proteomic sample, more than 3 folds increase (H/L in enriched = 2.1, H/L in bulk = 0.61), suggesting that c-ABL is a target of GNF-PAP in live cells and successfully identified by quantitative proteomic analyses (Figure 4.4b and 4.4c).

4.2.1.3 SILAC surface mapping to identify GNF-PAP binding in c-ABL

We next employed a SILAC surface mapping assay to investigate the potentially labeled sites in c-ABL by GNF-PAP, thus we treated 10 μM GNF-PAP to heavy cells and made SILAC mixtures of insoluble fractions. Among the list of proteins in the SILAC mix, FLAG-c-ABL was one of the most abundant proteins, therefore, we ran bulk SILAC digests of the insoluble fractions to get preliminary data.

c-ABL is a relatively big protein (mass of 123 kDa, total 1130 amino acids) and structural depiction of the entire protein has not been completed, only the crystal structure of isolated domains (SH2 and SH3 domains: 32-240, Kinase domain: 240-512) of c-ABL has been resolved. Preliminary tests of the SILAC surface mapping approach to c-ABL/GNF-PAP binding resulted in identification of the several peptides distinguished by skewed SILAC ratios (normalized L/H ratio > 2). Most of these peptides (555-574, 632-674, 789-820, 821-856, 895-909, and 957-1007)

were from C-terminal region (Figure 4.5a). C-terminal region of c-ABL has been demonstrated to regulate the localization of c-ABL to the nuclei and interact with DNA, but the structure of this region has not been characterized. Intriguingly, one peptide (480-512) in Kinase domain was also identified by the surface mapping assay, and we confirmed that the peptide is located in GNF-2 binding site of c-ABL (Figure 4.5b). The arrangement of GNF-2 in the crystal structure of isolated Kinase domain may imply that diazirine substrate of GNF-PAP is likely to crosslink with the identified peptide in the Kinase domain and few other peptides in the C-terminal region of c-ABL (Figure 4.5b).

Collectively, the results from the preliminary tests may indicate that SILAC surface mapping enable to find the protein-ligand binding sites in unbiased way. Further optimization of the assay with c-ABL/GNF-PAP model will be performed in the future.

4.2.2 Identification of CBR-470-1 binding site in PGK1

The original purpose of developing a SILAC surface mapping assay was to elucidate the actual interactions between glycolytic enzyme PGK1 with its novel inhibitor CBR-470-1. We previously utilized CBR-470-1 to discover the direct link between glucose metabolism and KEAP1-NRF2 pathway.

In our previous studies, we discovered a target of CBR-470-1 by utilizing various label- and label free technologies[149]. Target of CBR-470-1 was initially identified with a photoaffinity analogue probe CBR-470-PAP. Treatment of CBR-470-PAP to the live cells resulted in apparent biotin band formation near 45 kDa in Western blots, and LC-MS/MS analyses of gel-isolated samples revealed a PGK1 as a target of CBR-470-PAP. The results from other label free assays, such as *in vitro* and cellular thermal shift assays, *in vitro* isothermal dose response assay, and *in*

vitro enzymatic assay, also suggested that target of CBR-470-1 is PGK, however, our attempts to get a crystal structure of PGK1/CBR-470-1 binding were failed. Therefore, we employed a 'SILAC Surface Mapping' assay to identify photocrosslinked sites of the entire PGK1.

We performed on-bead digestions of the enriched targets of CBR-470-PAP in HEK293T cells or HeLa cells, and these digests were spiked-in by the unmodified, entire heavy PGK1 digests to prepare SILAC samples. Photocrosslinked peptides of light PGK1 were still remained on the beads and the rest of PGK1 peptides were collected in the following process with mixing entire heavy PGK1 digests, thus, photocrosslinked peptides of PGK1 by CBR-470-PAP were expected to be well distinguished with the extremely skewed SILAC ratios. We performed several biological replicates, including label-swapped sample preparations to reduce the noise of the data, and combined the SILAC data of PGK1 peptides. From both normal and label-swapped SILAC experiments, we identified several peptides which were apparently differentiated by the CBR470-PAP treatments (23-39, 87-97, 97-106, 92-106, 124-139, 124-141, and 140-151) (Figure 4.6a). These peptides were from N-terminal domain of PGK1 and located near the substrate binding site (1,3-BPG and 3-PG binding site) (Figure 4.6b). Our data indicates that CBR-470-1 may bind to the substrate binding site of PGK1.

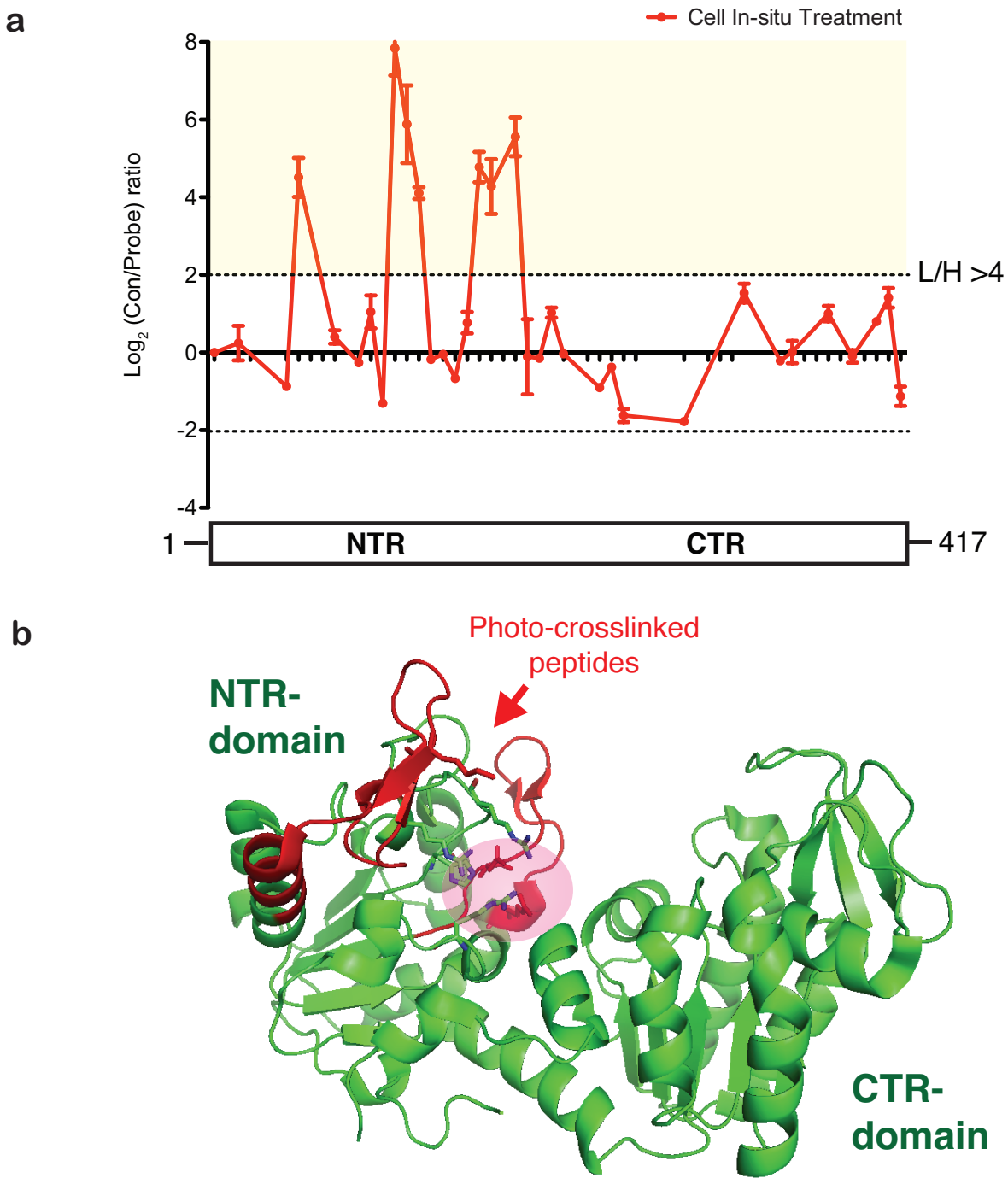


Figure 4.6 *In situ* SILAC surface mapping of PGK1 with CBR-470-PAP
a. SILAC ratios for individual tryptic peptides of FLAG-PGK1 enriched from DMSO treated 'heavy' cells and SA-pulldown endogenous PGK1 from GNF-PAP treated 'light' cells, relative to the median ratio of all PGK1 peptides. Tryptic peptides which were reduced by 4-fold or more upon relative to c-ABL median (Con/Probe > 4) were marked, indicative of photocrosslinking with photoaffinity probe. **b.** Structural depiction of modified stretches of human PGK1 (red) using published x-ray crystal structure of PGK1 (PDB: 2XE6).

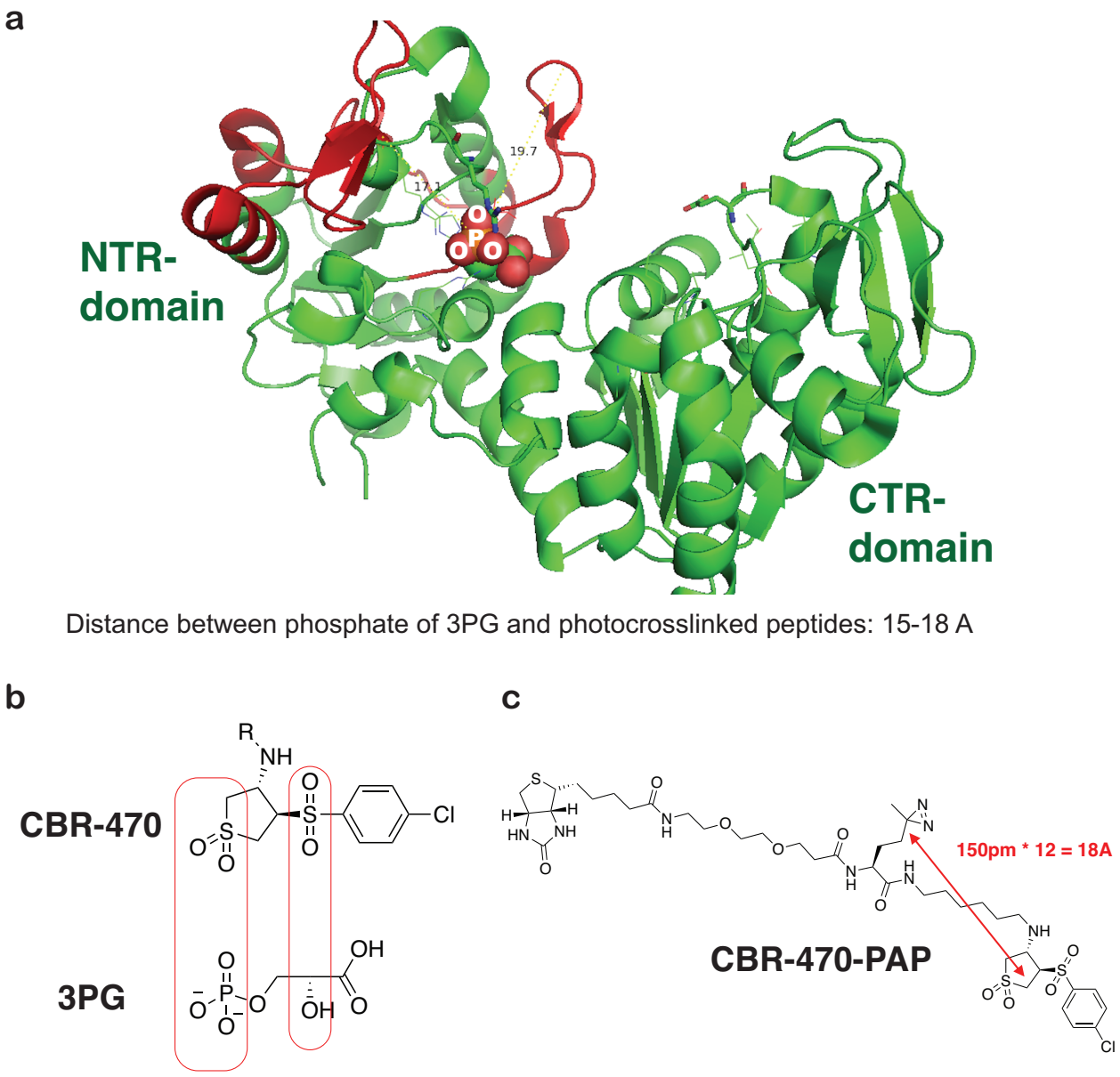


Figure 4.7 Structural resemblance between CBR-470 and 3PG
a. Distance between the phosphate of 3PB and photocrosslinked stretches of PGK1 (~15-20 Å). **b.** Structural resemblance between CBR-470 and 3PG. Conformation of oxygen atoms of CBR-470 potentially forms hydrogen bonds with PGK1 substrate binding site. **c.** Distance between CBR-470 and diazirine in CBR-470-1 (~18 Å).

4.3 Conclusion and Discussion

Study of the roles of target proteins in biological system by regulating its functions and behaviors with non-covalent protein-ligand interactions, thereby addressing the fundamental biological questions and developing novel therapeutic strategies, necessitates the identification of the binding sites of the ligands to their target protein. Indeed, structural depiction of the protein-ligand binding lead to the development of the next-generation small molecule inhibitors with better therapeutic efficacy.

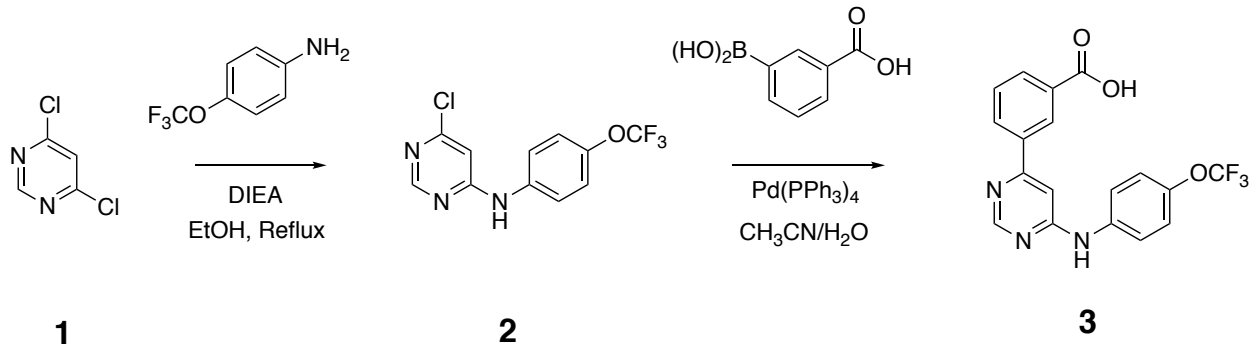
We proposed a novel chemical proteomic technology which is capable of identifying the ligand binding sites with SILAC quantification of the entire protein peptides. We first applied the method to the c-ABL/GNF-compound binding, which is a well characterized protein-ligand interaction model, and the preliminary tests resulted in identification of the GNF-PAP binding site in Kinase domain of c-ABL. The newer model PGK1/CBR-470-series interaction, which was recently discovered but the structural characterization of this protein-ligand binding is not accomplished, were also studied with the surface mapping strategy and peptides located near the metabolite binding sites of PGK1 were successfully distinguished. The results propose an inhibitory mechanism of PGK1 by CBR-470-1; CBR-470-1 blocks the substrate binding site of PGK1. Structural similarity between CBR-470-1 and the substrates (3-PG or 1,3-BPG) of PGK1, which potentially lead to form similar conformational hydrogen bonds with PGK1, also supports the proposed inhibitory mechanism of PGK1 (Figure 4.7a, Figure 4.7b).

Successful depiction of protein-ligand binding by SILAC surface mapping assay requires the near-complete coverages of the entire protein and low SILAC noise of each peptide. These technical issues could be solved by the overexpression of target protein to increase its cellular abundance and the generation of combined SILAC data from both normal and label-swapped

SILAC experiments to reduce the noise. In summary, the SILAC surface mapping assay will provide the unbiased information of protein-ligand interactions which affect the regulatory signaling pathways in cellular system.

4.4 Materials and Methods

Synthesis of GNF-PAP



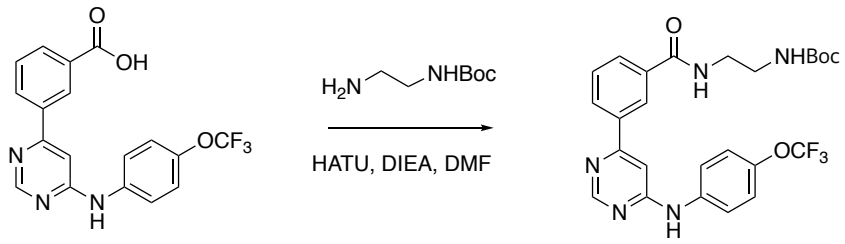
(6-Chloro-pyrimidin-4-yl)-(4-trifluoromethoxy-phenyl)-amine (2)

4,6-dichloropyrimidine (1) (1.006 g, 6.52 mmol) and p-trifluoromethoxy aniline (0.9 mL, 6.57 mmol) were dissolved in ethanol (20 mL) followed by the addition of diisopropylethylamine (1.75 mL). The resulting solution was heated to reflux in pressure tube for overnight, cooled to room temperature, after which the solvent was removed *in vacuo*, and the crude product purified by the flash chromatography (10% → 25% → 50% ethyl acetate in hexane) to give (6-Chloro-pyrimidin-4-yl)-(4-trifluoromethoxy-phenyl)-amine (2) as a white solid (1.533g, 81.2% yield). Rf = 0.8 (EA/Hx = 1:1, v/v, UV 254 nm).

3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzoic acid (3)

(6-Chloro-pyrimidin-4-yl)-(4-trifluoromethoxy-phenyl)-amine (2) (0.2214 g, 0.7644 mmol), 3-carboxyphenylboronic acid (0.1280g, 2.836 mmol), palladium tetrakis triphenylphosphine (0.0424 g, 0.0367 mmol), and sodium carbonate (0.3233g, 3.05 mmol) were dissolved in acetonitrile/water (v/v, 1:1, 6 mL). The pressure tube was purged with nitrogen and

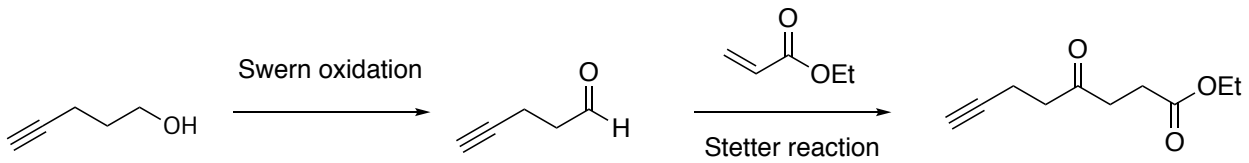
stirred at 90 ℃ for overnight after which the insoluble were removed by filtering the hot reaction solution. A solution of aqueous hydrochloric acid (6 N) was added until the pH was less than 5, while observing the precipitate formation. The solid product (3) was collected by filtration and washed twice with water (5 mL) and used without further purification (0.2465 g, 85.9% yleid). Rf = 0.1 (EA/Hx = 1:1, v/v, UV 254 nm)



3 **4**

***N*-Boc-aminoethyl-3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzamide (4)**

To a solution of 3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzoic acid (3) (0.3727 g, 0.993 mmol), N-Boc-ethylenediamine (0.19 mL, 1.20 mmol), and diisopropylethylamine (0.61 mL, 3.502 mmol) in anhydrous dimethylformamide (3.9 mL) was added 2-(1H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluranium "HATU" (0.453 g, 1.191 mmol) at room temperature. The reaction mixture was stirred for 7 hours at room temperature under nitrogen and the solvent was removed *in vacuo*. The crude product purified by the flash chromatography (50% ethyl acetate in hexane → 100% ethyl acetate) to give N-Boc-aminoethyl-3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzamide (4) (0.62g, over 100%). MS (ESI^+): m/z 518.4 ($M+H^+$). Rf = 0.55 (EA, UV 254 nm).



5 **6**

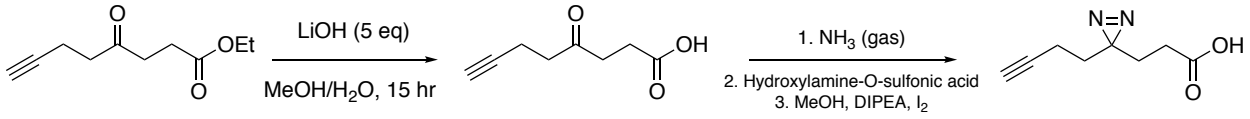
Pent-4-ynal (5)

Oxalyl chloride (7.2 mL, 85.05 mmol) was dissolved in dry dichloromethane (120 mL) and cooled to -78 ℃, and a solution of DMSO (12.8 mL, 180.2 mmol) in dry dichloromethane (12 mL) was added within 5 minutes. The mixture was stirred at this temperature for 30 minutes before a solution of 4-pentynol (6.2 g, 73.7 mmol) in dry dichloromethane (10 mL) was added dropwise. After being stirred for 1 hour, the mixture was treated with dry triethylamine (34 mL, 243.94 mmol), and stirring was continued at -78 ℃ for 30 minutes before the mixture was warmed slowly to room temperature. Then water (50 mL) was added, the phases were separated, and the aqueous phase was extracted with 3 x 30 mL dichloromethane. The combined organic layers were washed with brine (2 x 30 mL) and dried over $MgSO_4$, and the solvent was rotary evaporated carefully under slightly reduced pressure (400 mbar, 30 ℃).

Ethyl 4-oxooct-7-ynoate (6)

A solution of crude pent-4-ynal (freshly synthesized, 73.7 mmol) and ethyl acrylate (16 mL, 146.7 mmol) in 1,4-dioxane (44 mL) was added dropwise over a period of overnight to a suspension of 3-benzyl-5-(2-hydroxyethyl)-4-methylthiaolium chloride (2.05 g, 7.6 mmol), triethylamine (7.5 mL, 53.81 mmol) and ethyl acrylate (16 mL) in1,4- dioxane (56 mL) at 80 ℃ under an atmosphere of nitrogen. The mixture was stirred and heated at 80 ℃ for 72 hours and then volatiles removed by rotary evaporation. Residue resuspended in 100 mL of dichloromethane and washed with 10% H_2SO_4 (40 mL), saturated $NaHCO_3$ (60 mL), and brine (50 mL), then dried

over anhydrous $MgSO_4$ and volatiles removed by rotary evaporation. Crude was purified by flash column chromatography (100% hexanes → 10% → 12.5% → 25% ethyl acetate in hexanes), resulting in ethyl 4-oxooct-7-ynoate (6) as a light brown oil (1.421 g, 21.2% yield in two steps). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 4.14 (q, J = 7.1 Hz, 2H), 2.75 (q, J = 6.7 Hz, 4H), 2.61 (t, J = 6.6 Hz, 2H), 2.47 (td, J = 2.7, 6.5 Hz, 2H), 1.97 (t, J = 2.7 Hz, 1H), 1,25 (t, J = 7.0 Hz, 4H).



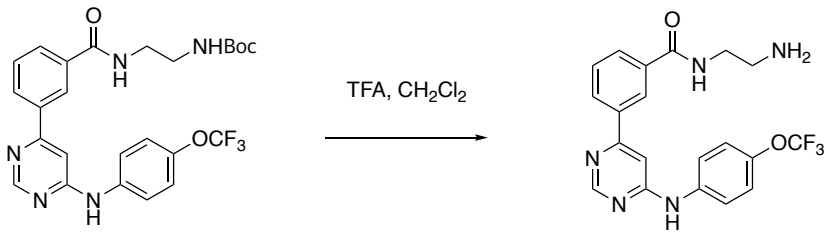
6 **7** **8**

4-oxooct-7-ynoic acid (7)

To a solution of ethyl 4-oxooct-7-ynoate (6) (3.56 g, 19.54 mmol) in methanol (75 mL), added LiOH (4.08 g, 170.35 mmol, 5 equiv) and water (2.7 mL, 5equiv) and let resulting solution stir at room temperature for 22 hours when TLC (1:3 ethyl acetate/hexanes) indicated the complete consumption of starting material. The solution was carefully acidified with aqueous HCl (6 M) until a pH of ~3 was achieved. The resulting solution was then extracted with dichloromethane and the combined organic layers were dried over anhydrous $MgSO_4$ and volaties were removed by rotary evaporation, resulting in 4-oxooct-7-ynoic acid (7) as a brown solid, which was used without further purification (3.0962g, 100% yield). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 2.78 – 2.65 (m, 6H), 2.47 (td, J = 2.6, 6.5 Hz, 2H), 1.96 (t, J = 2.7 Hz, 1H).

3-(3-but-3-yn-1-yl)-3H-diazirin-3-yl) propanoic acid (8)

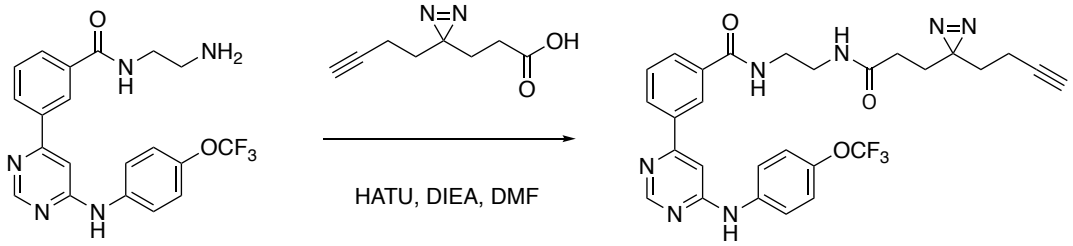
A dried 2-neck round bottom flask (side neck capped with plastic septa, top neck connected to Dewar condenser) containing 4-oxooct-7-ynoic $_{ac22id}$ (7) (0.346 g, 2.24 mmol) and anhydrous methanol (2 mL) mixture was cooled to -78 ℃ and then liquid NH_3 (~10mL) was added dropwise. A resulting solution was stirred at -78 ℃ for 5 hours. At this time, a solution of hydroxylamine-*O*-sulfonic acid (0.31 g, 2.74 mmol) in anhydrous methanol (5 mL) was added dropwise via using dropping funnel at -78 ℃. The resulting solution was stirred at -78 ℃ for additional 3 hours and then allowed to warm to room temperature overnight. Resulting suspension was evaporated to dryness and resuspended in methanol (30 mL) and solid was filtered washed few times with methanol. The combined filtrate was evaporated and resuspended in anhydrous methanol (5 mL), then cooled to 0 ℃, while protecting from light. Triethylamine (0.46 mL, 3.3l mmol) was added, followed by iodine (potion-wise), until a dark brown color persisted for more than 30 minutes. The solution was then diluted with ethyl acetate (~30 mL) and washed with aq. 1N HCl, saturated aqueous $Na_2S_2O_3$ (until organic layer clarified), and brine. Combined aqueous phases were washed once with ethyl acetate and all organic layers were combined, then dried over anhydrous Na_2SO_4 and volatiles removed by rotary evaporation. Crude was purified by flash column chromatography (100% dichloromethane $\rightarrow$ 20% $\rightarrow$ 50% ethyl acetate in dichloromethane), resulting in 3-(3-but-3-yn-1-yl)-3H-diazirin-3-yl) propanoic acid (8) as a colorless oil (24 mg). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 2.18 (t, J = 7.5 Hz, 2H), 2.04 – 1.99 (m, 3H), 1.82 (t, J = 7.1, 2H), 1.67 (t, J = 7.2 Hz, 2H).



4 **9**

***N*-aminoethyl-3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzamide (9)**

N-Boc-aminoethyl-3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzamide (4) (0.0413 g, 0.0798 mmol) was dissolved in methanol (3 mL) and trifluoroacetic acid (1mL), followed by stirred at room temperature for overnight. The volatiles in the mixture were removed by rotary evaporation and remained trifluoroacetic acid was removed by three times of rotary evaporation after diluting dried sample with methanol (5 mL). (0.0308g, 92.3% yleid).



9 **10**

GNF-PAP (10)

To a solution of *N*-aminoethyl-3-[6-(4-Trifluoromethoxy-phenylamino)-pyrimidin-4-yl]-benzamide (9) (0.0308 g, 0.074 mmol), 3-(3-but-3-yn-1-yl)-3H-diazirin-3-yl) propanoic acid (8) (0.0126 g, 0.076 mmol), and diisopropylethylamine (0.05 mL, 0.287 mmol) in anhydrous dimethylformamide (5 mL) was added 2-(1H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluranium

"HATU" (0.035 g, 0.092 mmol) at room temperature. The reaction mixture was stirred for 72 hours at room temperature under nitrogen and the solvent was removed *in vacuo*. The crude product was dissolved in ethyl acetate (20 mL) and washed with brine (20 mL x 4), then dried over anhydrous $MgSO_4$ and volatiles removed by rotary evaporation. The aliquot of crude was finally purified by reverse phase HPLC eluting with a gradient of MeOH in H_2O (60 – 100% MeOH). The GNF-PAP (3.3 mg) was isolated as a yellow solid. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.70 (d, J = 1 Hz, 1H), 8.43 (t, J = 1.7 Hz, 1H), 8.14 (dt, J = 1.6, 7.9 Hz, 1H), 7.94 (dt, J = 1.6, 7.8 Hz, 1H), 7.80 (d, J = 9.1 Hz, 2H), 7.62 (t, J = 7.8 Hz, 1H), 7.26 (d, J = 8.7 Hz, 2H), 7.21 (d, J = 1.2 Hz, 1H), 4.61 (s, 2H), 3.53 (t, J = 6.3 Hz, 2H), 3.44 (m, 4H), 3.16 (m, 2H), 2.22 (t, J = 2.7 Hz, 1H), 2.02 (t, J = 7.5 Hz, 2H), 1.95 (td, J = 2.7, 7.5 Hz, 2H), 1.73 (t, 7.3 Hz, 2H), 1.55 (t, 7.6 Hz, 2H). MS (ESI^+): m/z 566.46 ($M+H^+$).

Cell culture

HeLa and HEK293T cells were propagated in DMEM (Corning) or in RPMI 1640 with 2 mM L-glutamine (HyClone) supplemented with 10% fetal bovine serum (FBS, Corning) and 1% penicillin-streptomycin (Thermo Fisher Scientific). All cell lines were grown at 37 ˚C in a 5% CO_2 humid incubator

FLAG-tagged protein expression and Western blot

Transient protein expression of full-length, human c-ABL (HG11199-NF, Sino Biological) was performed in confluent 10 cm plates of HEK293T cells by transfection of 1 μg plasmid with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol.

Cells were harvested by scraping, pelleted by centrifugation, washed twice with cold PBS and lysed in PBS (Corning) supplemented with EDTA-free complete protease inhibitor (Roche), pH 7.4, at 4 ˚C. Cells were sonicated (Fisher Scientific FB-505), insoluble debris separated by centrifugation (17,000 g, 20 minutes, 4 ˚C), and the insoluble fraction was resuspended in PBS followed by diluted into 4X Laemmli buffer containing 50 mM dithiothreitol (DTT) or 6% beta-mercaptoethanol (βMe) as reducing agents. Samples were prepared for SDS-PAGE by heating to 95 ˚C for 5 minutes, cooled to room temperature, resolved on NuPAGE Novex 4-12% Bis-Tris Protein Gels (Invitrogen) or 10% SDS-PAGE gel, and transferred onto nitrocellulose membranes by standard western blotting methods. Membranes were blocked in 2% BSA in TBS containing 0.1% tween-20 (TBST) and probed with primary and secondary antibodies. Primary antibodies used in this study include: anti-FLAG-M2 (1:1000, F1804, Sigma Aldrich), anti-ACTB (1:1000, 4790, Cell Signaling), and Streptavidin (1:10000, 92632230, LI-COR). Secondary donkey anti-rabbit, donkey anti-goat, and donkey anti-mouse (Licor), were used at 1:10,000 dilution in 2% BSA-containing TBST and incubated for 1 hour prior to washing and imaging on a Licor infrared scanner. Densitometry measurements were performed with ImageJ software.

Cell viability measurements

$2x10^3$ K562 were plated in 50 μL of RPMI 1640 growth medium in white 96-well plates and 50 μL of GNF-PAP diluted to the indicated concentrations was added to each well. After 72 hours incubation, cell viability measurements were recorded on an Envision plate reader after the addition of 100 μL of a Cell Titer Glo solution (Promega, diluted 1:4 in PBS). Relative viabilities are reported as a fraction relative to the same dose of compound treatment without GNF-PAP.

Light-activated target crosslink with GNF-PAP

10 cm dishes of confluent HEK293T cells transiently expressed with FLAG-cABL were exposed to 10 μM GNF-PAP in serum-free RPMI 1640 for 2 hours at 37 °C. Cell dishes were then UV irradiated using a Spectroline XL-1500A instrument for 20 minutes, scraped, washed with cold PBS (2x), and lysed in DPBS supplemented with protease inhibitors.

SILAC cell culture methods and proteomic sample preparation

SILAC labeling was performed by growing cells for at least five passages in lysine- and arginine-free SILAC medium (RPMI, Invitrogen) supplemented with 10% dialyzed fetal calf serum, 2 mM L-glutamine and 1% P/S. "Light" and "heavy" media were supplemented with natural lysine and arginine (0.1 mg/mL), and ^{13}C-, ^{15}N-labeled lysine and arginine (0.1 mg/mL), respectively.

General protein digestion for LC-MS/MS analysis was performed by the denaturation of proteins in PBS lysate with Urea to make 8 M final concentration followed by disulfide reduction with DTT (10 mM, 40 minutes, 50 ˚C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Promega) at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted on ZipTip C18 tips (100 μL, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized. Lyophilized peptides were dissolved in LC-MS/MS Buffer A (H_2O with 0.1% formic acid, LC-MS grade, Sigma Aldrich) for proteomic analysis.

Preparation of proteomic samples for MS-based target identification analysis (GNF-PAP)

Proteome-wide identification of the target of GNF-PAP with SILAC quantitative proteomics was performed with "heavy" and "light" labeled HEK293T cells transiently expressing FLAG-cABL constructs. SILAC-labeled cells, grown to 80-90% confluency in 10 cm cell-culture treated plates (Denville) each, were incubated with DMSO alone (light cells) or GNF-PAP probe (10 μM, heavy cells) for 2 hours in serum-free SILAC RPMI. After incubation, cells were UV irradiated using a Spectroline XL-1500A instrument for 20 minutes, scraped, washed with cold PBS (2x), and lysed in PBS (Corning) supplemented with EDTA-free complete protease inhibitor (Roche), pH 7.4, at 4 ˚C. Cells were sonicated, and the insoluble fraction was resuspended in PBS after the centrifugation (17,000 g, 20 minutes, 4 ˚C). To the combined mixture of heavy and light insoluble proteomes (1 mg/mL) in 1 mL PBS, TBTA (60 μL, 1.67 mM in 1:4 DMSO:t-BuOH), $CuSO_4$ (20 μL, 50 mM in H_2O), TCEP (20 μL, 50 mM in H_2O), and Biotin-PEG5-N_3 (10 μL, 10 mM in DMSO) was added and each sample was rotated at room temperature for 1 hour, The mixture was transferred to 8 mL glass vial and 2 mL of methanol (MeOH), 0.5 mL of chloroform ($CHCl_3$), and 1 mL of H_2O was added to the vial. After vortex, the resulting cloudy mixture was centrifuged (3,200 g, 20 minutes, 4 ˚C) to fractionate the protein interphase from the organic and aqueous solvent layers. After carefully collecting protein disc and and transferring the protein to 1.6 mL Eppendorf tube with 0.9 mL of MeOH, the precipitate was pelleted by centrifugation (9,000 g, 5 minutes, 4 ˚C). The pellet was resuspended in 800 μL 6 M Urea in 25 mM ammonium bicarbonate then dissolved by tip sonication. Disulfides were reduced by adding DTT (10 mM final concentration) for for 15 minutes at 65 ˚C, and reduced thiols were then alkylated by addition of iodoacetamide (40 mM final concentration) for 40 minutes at room temperature protected from

light. Each sample was then transferred to 15 mL falcon tube and sequentially diluted with 25 mM ammonium bicarbonate solution (5 mL final, applied tip sonication in the dilution step) and incubate with pre-equilibrated streptavidin agarose resin (80 µL, 1:1 slurry, Pierce) for 3 hours at room temperature on a rotator. The streptavidin beads were collected by centrifugation (500 g, 2 minutes) and sequentially washed with 1 M Urea in PBS (7 x 1 mL) and PBS (3 x 1 mL) to remove nonspecifically attached proteins on beads. The resin was transferred to a Protein LoBind tube and biotinylated, enriched proteins were digested on-bead overnight at 37 ˚C in 200 µL total volume of 2 M Urea in 25 mM ammonium bicarbonate supplemented with 1 mM $CaCl_2$ and sequencing grade trypsin (2 µg, Thermo Fisher Scientific). The proteolyzed supernatant was transferred to a fresh Protein LoBind tube, acidified with formic acid (2% final), and desalted as indicated above.

After LC-MS/MS analysis, the SILAC ratios of the proteins from enriched proteomic sample were normalized by the median SILAC ratio of the corresponding bulk proteomic sample.

Proteomic LC-MS/MS and data analysis

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra high-pressure LC system (ThermoFisher) using a PepMap RSLC C18 column heated to 45˚C (column: 75 µm x 15 cm; 3 µm, 100 Å) coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). Digested peptides (500 ng - 1 µg) in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% Formic acid acetonitrile) at 300 nL/min: 2-2% buffer B over 5 minutes, 2-25% buffer B over 170 minutes, 25-40% buffer B over 40 minutes, 40-90% buffer B over 10 minutes, 90-90% buffer B over 5 minutes, 90-2% buffer B over 5 minutes, 2-2% buffer B over 5 minutes, 2-90% buffer B over 5 minutes, 90-90% buffer B over 3 minutes, 90-2% buffer B over 5 minutes, 2-2% buffer B over 3 minutes, 2-90% buffer B

over 5 minutes, 90-90% buffer B over 5 minutes, 90-2% buffer B over 5 minutes, and 2-2% buffer B over 3 minutes. MS/MS spectra were collected from 0 to 240 minutes using a data-dependent, top 10 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 375-1500 m/z, maximum IT of 60 ms, AGC target of 1e6, and data collection in profile mode. MS2 scans was performed by HCD fragmentation with a resolution of 30,000, AGC target of 1e5, maximum IT of 60 ms, NCE of 27, MSX count of 1, and data type in centroid mode. Isolation window for precursor ions was set to 2.0 m/*z* with isolation offset of 0.0 m/z. Peptides with charge state 1 and undefined were excluded and dynamic exclusion was set to twenty seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.20kV and ionization chamber temperature of 275 ˚C.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database (UniProt_Human_reviewed_04-10-2017.fasta). Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation); two total differential modification sites per peptide, including oxidized methionine (+15.9949); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting statistics were performed with a Δmass cutoff = 15 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%, peptide modification requirement (-m) was

set to 1, and spectra display mode (-t) was set to 1. SILAC searchers were performed as above with “light” and “heavy” database searches of MS1 and MS2 files by including static modification of +8.014168 for lysine and +10.0083 for arginine in a parallel heavy search. SILAC quantification was performed using the QuantCompare algorithm, with a mass tolerance of 10 p.p.m. or less in cases where co-eluting peptide interfere. In general all quantified peptides has mass error within 3 p.p.m..

Quantitative proteomic detection of potential PGK1/CBR-470 binding site

Quantitative surface mapping with SILAC quantitative proteomics was performed with “heavy” and “light” labeled HEK293T cells or HeLa cells. Cells were incubated with CBR-470-1 (20 μM, light cells) for 1 hours followed by UV irradiated using a Spectroline XL-1500A instrument for 20 minutes. Then, cells were scraped, washed with cold PBS (2x), and lysed in PBS (Corning) supplemented with EDTA-free complete protease inhibitor (Roche), pH 7.4, at 4 ˚C. After sonication, insoluble debris was cleared by centrifugation (17,000 *g*, 10 min). Streptavidin C1 magnetic beads (30 μL slurry, 65001, Invitrogen) were washed twice with PBS, and each cell lysate was incubated with the magnetic beads with rotation overnight at 4 ˚C. The beads were washed 8 times with 1 mL of PBS, and 500 μL of 6 M Urea in 50 mM ammonium bicarbonate was then added to the beads. Samples were reduced on resin by TCEP (10 mM final), with orbital shaking, for 20 minutes at 65 ˚C. Samples were then alkylated by adding iodoacetamide (20 mM final), covered from the light and with orbital shaking, for 40 minutes at 37 ˚C. The streptavidin magnetic beads were collected, washed once with 2 M Urea in 25 mM ammonium bicarbonate, and the buffer exchanged to 2 M Urea in 25 mM ammonium bicarbonate supplemented with 1 mM $CaCl_2$. Enriched proteins were digested on bead by the incubation of 2 μg sequencing grade trypsin

overnight at 37 ˚C. Following trypsinization, supernatant was collected, acidified with HPLC grade formic acid (2% final, pH 2-3), and peptides were desalted as indicated above.

Heavy, unmodified digests of entire PGK1 was prepared by the transient expression of FLAG-PGK1 into heavy HEK293T cells followed by the cell lysis in Urea lysis buffer [8 M Urea, 50 mM NH_4HCO_3, nicotinamide (1 mM), phosphatase inhibitor cocktail (Sigma Aldrich), and EDTA-free complete protease inhibitor (Roche), pH 8.0] by sonication at 4 ˚C. After sonication, insoluble debris was cleared by centrifugation (17,000 *g*, 10 min), diluted with Milli-Q water to give 1 M urea, and lysate was incubated with anti-FLAG M2 resin (100 µL slurry, A2220, Sigma Aldrich) at 4 ˚C overnight while rotating. Then, anti-FLAG resin was washed with PBS (7x1 mL), FLAG-KEAP1 protein eluted with glycine-HCl buffer (0.1 M glycine, pH 3.5, 2x500 µL), followed by 8 M urea (2x100 µL). The combined eluent was brought up to 8 M urea total concentration and processed for trypsin digestion and LC-MS/MS analysis as indicated above. Heavy digests were then spiked-in to the light enriched digests and analyzed by LC-MS/MS.

For SILAC label-swap experiments, CBR-470-PAP was treated to heavy cells and prepared by spike-in of light entire PGK1 digests. The SILAC maps were generated by comparing SILAC ratios for each peptide, relative to the median value for all PGK1 peptides. SILAC ratios were converted to Log_2 values and plotted to visualize peptides that are significantly perturbed, for example by modification, relative to the rest of the protein. A minimum of three SILAC ratios for each peptide was required for inclusion in PGK1 surface maps, which allowed for ~80% coverage of the PGK1 protein. Missing sequences were caused by the lack or close spacing of tryptic sites, resulting in inadequate peptides for MS/MS detection.

CHAPTER 5

Summary and Perspective

5.1 Development and application of novel chemical proteomic technologies

Sustaining the living system is supported by the numerous and dynamic interactions of functional proteins and other cellular components that are involved in gene expressions and the process to generate the energy and building blocks for the system. Therefore, studies of the protein regulatory mechanisms and their dynamic communication networks are necessary to find the answers to the fundamental biological questions and following establishment of the new therapeutic and pharmaceutic approaches. To address these challenges, development and application of new chemical proteomic technologies are required to study the dynamic behaviors, interactions, and communications of the proteome in live cells. We recently developed the novel technologies to gain high resolution protein-protein interaction maps and actual protein-ligand interactions in live cells.

5.1.1 Protein-protein interactions in rmPTMs

Understanding the mechanisms by which reactive metabolites are delivered to their target proteome and induces non-enzymatic protein modifications is one of remained challenges to elucidate the cycle of rmPTM signaling cascades. As non-enzymatic protein modifications are completely relied on the intrinsic chemical reactivity of the reactive metabolites and active residues of the protein, half-life of the reactive functionalities, and the diffusion radius from the source of generation (i.e. metabolic enzymes), the proximity information between the source and targets is a critical clue to elucidate the protein networks communicated by the specific metabolic signals.

The targets of rmPTMs commonly possess sensitive amino acid residues (i.e. nucleophilicity or electrophilicity) located in the functional sites of these proteins. For example, sensitive electrophile sensor proteins (e.g. KEAP1) are vulnerable toward the exposure of highly reactive molecules, such as peroxide. Indeed, the level of rmPTMs in the proteome is significantly influenced by the level of corresponding reactive metabolites generated from the mainstream metabolic fluxes, thus, rmPTMs directly reflect the metabolic status in the cell. The advanced technology that enables to take a snapshot of protein interactions regulated by dynamic metabolic alterations and fast kinetics of non-enzymatic modifications is necessitated for the study of rmPTM networks. However, the current proximity labeling technologies, of both enzymatic ligations and peroxidase systems, are incapable of being applied for the study of rmPTM networks.

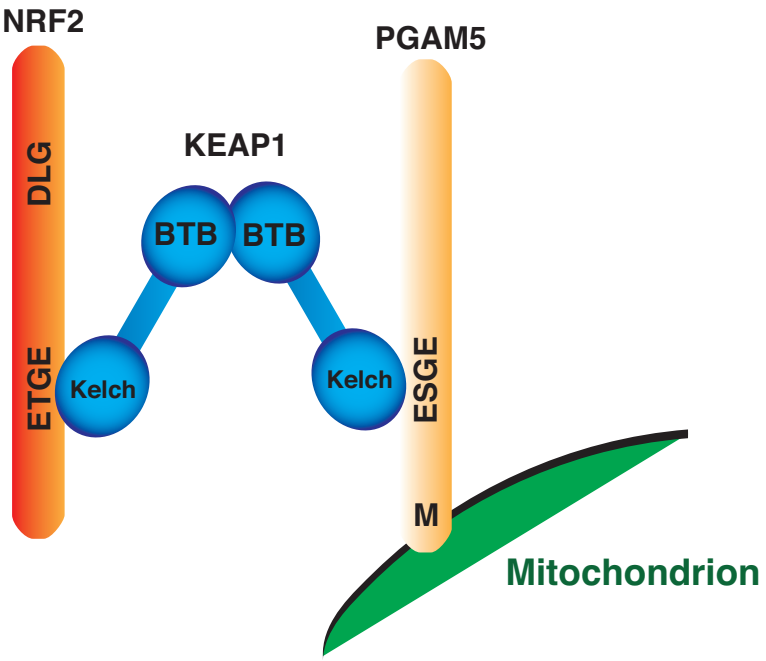


Figure 5.1 The "quaternary complex" model
PGAM5 has been reported as a strong substrate for KEAP1 that forms the PGAM5-KEAP1-NRF2 complex localized to the cytosolic surface of the outer mitochondrial membrane, yet actual communication and behaviors of KEAP1 still need further investigation.

We recently developed the novel proximity biotin tagging system which is activated by the short-period light exposure. This new technology, termed P3 profiling, enables the study of rmPTM networks as biotin labeling of the prey proteins occurs in the mild and fast reaction. We applied P3 profiling system to KEAP1 construct fused with SNAP protein and we successfully identified the well-known target PGAM5[217-219,245] that is highly enriched with the P3 probe treatment in live cells (Figure 5.1). We also identified the list of other proteins enriched with the probes, some of these are shown in the public databases of KEAP1 communication network.

After the completion of optimizing P3 system, with the various options of designing next generation P3 probes, it will be widely applicable for the study of numerous protein-protein interactions and especially the networks of rmPTMs. Followed by our previous studies of glycolysis-KEAP1-NRF2 signaling axis, the next questions are still remained: 1) How MGx is delivered to KEAP1? 2) How HMW-KEAP1 formation alters its behavior and structural conformation? 3) What is the actual model that the modification of KEAP1 stabilizes cellular NRF2? 4) Does the modification of KEAP1 by MGx alter its location and communication partners in cells? 5) What is the mechanism that cleans up the HMW-KEAP1 and bring it back to monomeric KEAP1 in cells? We previously observed the gradual decrease of HMW-KEAP1 levels in cells after shutting off the CBR-470-1 treatment (Figure 5.2), but the mechanisms beyond the cleaning up temporal modifications[116,246] and the cleaners of the modification have not been revealed.

Indeed, the application of P3 technology to other defined rmPTM targets will lead us to discover the factors that regulate the localization of the metabolic enzymes followed by the delivery of reactive species to the targets.

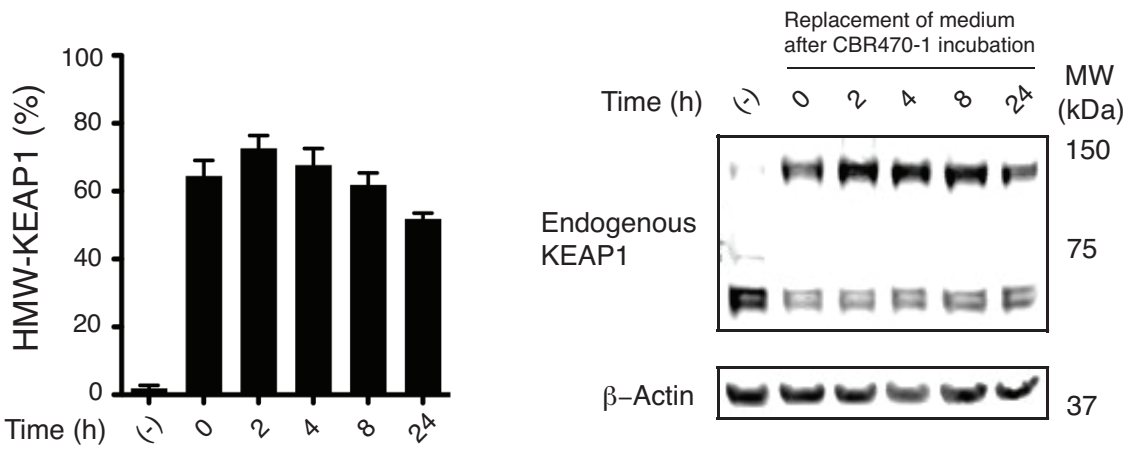


Figure 5.2 Cleaning up MICA modification of HMW-KEAP1 in cells
Relative level of HMW-KEAP1 is decreased over time after stopping the treatment of CBR-470-1. Next challenge is to elucidate the clean-up mechanism of HMW-KEAP1.

5.1.2 Protein-ligand interactions in rmPTMs

rmPTMs are novel signaling mechanisms that directly link the metabolism to the responsive signaling cascades. Thus, the level of rmPTMs in target proteome reflects the population of these specific metabolites generated from the corresponding metabolic fluxes. The direct link between glucose metabolism and the KEAP1-NRF2 pathway by MGx was discovered by the treatment of novel PGK1 inhibitor CBR-470-1. In our previous studies, glycolytic enzyme PGK1 was identified as a target of CBR-470-1 and its analogues, validated by the diverse label- and label free assays. However, we couldn't solve one question: how CBR-470-1 regulate the PGK1 activity?

To address, we previously attempted to get a crystal structure of PGK/CBR-470-1 binding and characterize the binding conformation, but we failed. Indeed, elucidation of the mechanism by which CBR-470-1 regulates PGK1 function in the cell necessitates the new methods to study actual

PGK1/CBR-470-1 interactions. As the application of SILAC surface mapping assay successfully led to the identification of modification sites in HMW-KEAP1, we wondered if we can identify the PGK1/CBR-470-1 binding sites by photoaffinity analogue probe CBR-470-PAP and SILAC peptide quantitation approach. The data from our experiments suggested that CBR-470-PAP bound to the substrate binding site of PGK1 in N-terminal domain, which potentially implies that CBR-470-1 inhibits PGK1 by blocking 3-PG (or 1,3-BPG) binding sites. This assumption is also supported by the structural resemblance between 3-PG (or 1,3-BPG) with the potent substrate of CBR-470-1. In order to generalize our new approach with another protein-ligand interactions, we are currently applying the strategy to well-characterized c-ABL/GNF-2 binding model.

To overcome the restriction of *in vitro* structural characterization methods, we suggested a new chemical proteomic approach which possesses many advantages, including relatively simple- and fast rate of sample preparation and data analysis. The future applications of this tool in the study of rmPTMs will be the elucidation of the regulatory mechanisms of novel metabolic enzyme inhibitors with a whole, unbiased sight.

5.2 Future directions: elucidation of a signaling cycle of reactive metabolite-induced protein modifications

Recently published studies, including our recent discovery of direct link between glucose metabolism and KEAP1-NRF2 pathway, have implicated that the non-enzymatic posttranslational modifications by intrinsically reactive species may play a key role in cell signaling. We need further investigations to understand a signaling cycle of rmPTMs in cellular system and its roles in human metabolic diseases.

5.2.1 Generation of cellular reactive metabolites

To control the specific reactive metabolites in the cell, understanding the mechanisms by which cellular levels of reactive metabolites are regulated is primarily required. For example, 1,3-BPG is directly generated by the glycolytic enzyme GAPDH in the catabolic process and PGK1 in the anabolic process of glucose metabolism. Cellular MGx is mostly generated in glycolysis, from the intermediate of the glycolytic enzyme triosephosphate isomerase 1 (TPI1) which keeps the balance between the level of two triosephosphate isomers GAP and DHAP in the glucose metabolism.

Several strategies to regulate the generation of cellular reactive metabolites could be suggested: 1) Developing the novel inhibitors directly targeting the enzymes that produce reactive metabolites, as we previously observed the 3-fold elevation of 1,3-BPG and following pgK modifications with the inhibition of PGK1. 2) Systematic approaches such as indirect regulation of reactive metabolite levels by the altering metabolic fluxes, as we obtained the accumulation of cellular MGx by PGK1 inhibition. 3) Potential regulatory factors of the reactive metabolite generators, such as specific posttranslational modifications on the active sites of these enzymes.

5.2.2 Non-enzymatic modifications to target proteome

After the generation of cellular reactive metabolites, the next process of the rmPTM signaling cascade is to deliver these metabolites to the targets. Efficiency of the delivery to targets is mainly affected by the distance between the source of reactive metabolites and the target proteins. Indeed, the amounts of reactive metabolites will be also critical for the efficiency of delivery.

The specificity of the rmPTMs on targets will be determined by the nature of modification sites, such as the pocket capacity and the reactivity of these active residues. The effect of rmPTMs on target protein activity will be estimated by the identification of modification sites and the structural insight into the protein: Do rmPTMs allosterically regulate the target protein or directly block the active sites? Do rmPTMs induce the translocalization of the target to specific locations?

5.2.3 Activation of the downstream signaling cascades and cleaning up processes

To discover the roles of rmPTMs in cell signaling, as well as unveiling the direct links between metabolism and responsive pathways, global identification of the modification targets and their sites should be preceded. For example, MGx targets the specific sites of histones to regulate the gene expressions and C151 site of KEAP1 to activate the NRF2 signaling in cells. Modification targets of other non-enzymatic modifications, such as 1,3-BPG[98] and itaconate[247], have also identified. The powerful approach to identify the targets and sites of the modifications in full-scale proteomics is activity-based protein profiling (ABPP) which induces covalent modifications of the targets with the reactive metabolite analogue probes and enriches these targets with following proteomics workflow[248]. Through chemical proteomics, future works will be the identification of rmPTM targets in mammalian cells and other organisms.

The last process of the rmPTM signaling cycle is the activation of responsive pathways and following downstream process, such as regulating target gene expressions and reprogramming the cellular system. Indeed, cleaning process could be applied to the specific rmPTMs on target proteins. We previously observed the dephosphorylation and deacetylation of pgK modification in cell lysates and gradual decrease of HMW-KEAP1 levels in cells after shutting off the CBR-470-1 treatment, but the mechanisms beyond the cleaning up temporal modifications[246] and the cleaners (e.g. HDAC, phosphatases) of these modifications have not been revealed. Future plan is to develop and apply the novel proximity profiling technology to identify the dynamic interactions between the target of rmPTMs and proteins involving in cleaning up processes.

List of References

1. Pasteur, L. Mémoire sur la fermentation alcoolique. *Annales de Chimie et de Physique* **58**, 323-426 (1860).

2. Buchner, E. Alkoholische Gärung ohne Hefezellen (Vorläufige Mittheilung). *Berichte der Deutschen Chemischen Gesellschaft* **30**, 117-124 (1897).

3. Buchner, E. Alkoholische Gärung ohne Hefezellen (Zweite Mittheilung). *Berichte der Deutschen Chemischen Gesellschaft* **30**, 1110-1113 (1897).

4. Barnett, J.A. A history of research on yeasts 5: the fermentation pathway. *Yeast* **20**, 509-543 (2003).

5. Kresge, N., Simoni, R.D. & Hill, R.L. Otto Fritz Meyerhof and the elucidation of the glycolytic pathway. *J Biol Chem* **280**, e3 (2005).

6. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033 (2009).

7. Dang, C.V., Lewis, B.C., Dolde, C., Dang, G. & Shim, H. Oncogenes in tumor metabolism, tumorigenesis, and apoptosis. *J Bioenerg Biomembr* **29**, 345-354 (1997).

8. Warburg, O., Wind, F. & Negelein, E. The Metabolism of Tumors in the Body. *J Gen Physiol* **8**, 519-530 (1927).

9. Faubert, B.*, et al.* Lactate Metabolism in Human Lung Tumors. *Cell* **171**, 358-371 e359 (2017).

10. DeFronzo, R.A., Gunnarsson, R., Bjorkman, O., Olsson, M. & Wahren, J. Effects of insulin on peripheral and splanchnic glucose metabolism in noninsulin-dependent (type II) diabetes mellitus. *J Clin Invest* **76**, 149-155 (1985).

11. Rossetti, L. & Giaccari, A. Relative contribution of glycogen synthesis and glycolysis to insulin-mediated glucose uptake. A dose-response euglycemic clamp study in normal and diabetic rats. *J Clin Invest* **85**, 1785-1792 (1990).

12. Spiegel, K., Tasali, E., Leproult, R. & Van Cauter, E. Effects of poor and short sleep on glucose metabolism and obesity risk. *Nat Rev Endocrinol* **5**, 253-261 (2009).

13. Peiris, A.N., Struve, M.F., Mueller, R.A., Lee, M.B. & Kissebah, A.H. Glucose metabolism in obesity: influence of body fat distribution. *J Clin Endocrinol Metab* **67**, 760-767 (1988).

14. Mosconi, L. Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD. *Eur J Nucl Med Mol Imaging* **32**, 486-510 (2005).

15. Fukuda, M.*, et al.* Networks mediating the clinical effects of pallidal brain stimulation for Parkinson's disease: a PET study of resting-state glucose metabolism. *Brain* **124**, 1601-1609 (2001).

16. Suji, G. & Sivakami, S. Glucose, glycation and aging. *Biogerontology* **5**, 365-373 (2004).

17. Warburg, O. On respiratory impairment in cancer cells. *Science* **124**, 269-270 (1956).

18. Warburg, O. On the origin of cancer cells. *Science* **123**, 309-314 (1956).

19. Weinhouse, S. Summation: molecular mechanisms of gene regulation--session 3. *Cancer Res* **36**, 4330-4331 (1976).

20. Fantin, V.R., St-Pierre, J. & Leder, P. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. *Cancer Cell* **9**, 425-434 (2006).

21. Moreno-Sanchez, R., Rodriguez-Enriquez, S., Marin-Hernandez, A. & Saavedra, E. Energy metabolism in tumor cells. *FEBS J* **274**, 1393-1418 (2007).

22. Hawkins, R.A. & Phelps, M.E. PET in clinical oncology. *Cancer Metastasis Rev* **7**, 119-142 (1988).

23. Wahl, R.L.*, et al.* Proceedings of a National Cancer Institute workshop: PET in oncology--a clinical research agenda. *Radiology* **193**, 604-606 (1994).

24. Weber, W.A., Avril, N. & Schwaiger, M. Relevance of positron emission tomography (PET) in oncology. *Strahlenther Onkol* **175**, 356-373 (1999).

25. Gambhir, S.S. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer* **2**, 683-693 (2002).

26. Fan, T.W.*, et al.* Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). *Mol Cancer* **8**, 41 (2009).

27. Lane, A.N.*, et al.* Stable isotope-resolved metabolomics (SIRM) in cancer research with clinical application to nonsmall cell lung cancer. *OMICS* **15**, 173-182 (2011).

28. Jang, C., Chen, L. & Rabinowitz, J.D. Metabolomics and Isotope Tracing. *Cell* **173**, 822-837 (2018).

29. Gatenby, R.A. & Gillies, R.J. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer* **4**, 891-899 (2004).

30. Kunkel, M., *et al.* Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral squamous cell carcinoma. *Cancer* **97**, 1015-1024 (2003).

31. Mochiki, E., *et al.* Evaluation of 18F-2-deoxy-2-fluoro-D-glucose positron emission tomography for gastric cancer. *World J Surg* **28**, 247-253 (2004).

32. Graeber, T.G., *et al.* Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88-91 (1996).

33. Gatenby, R.A. & Gawlinski, E.T. A reaction-diffusion model of cancer invasion. *Cancer Res* **56**, 5745-5753 (1996).

34. Hatzivassiliou, G., *et al.* ATP citrate lyase inhibition can suppress tumor cell growth. *Cancer Cell* **8**, 311-321 (2005).

35. DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G. & Thompson, C.B. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab* **7**, 11-20 (2008).

36. Buzzai, M., *et al.* The glucose dependence of Akt-transformed cells can be reversed by pharmacologic activation of fatty acid beta-oxidation. *Oncogene* **24**, 4165-4173 (2005).

37. Osthus, R.C., *et al.* Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. *J Biol Chem* **275**, 21797-21800 (2000).

38. Noguchi, Y., *et al.* Expression of facilitative glucose transporter 1 mRNA in colon cancer was not regulated by k-ras. *Cancer Lett* **154**, 137-142 (2000).

39. Iozzo, P., *et al.* Insulin stimulates liver glucose uptake in humans: an 18F-FDG PET Study. *J Nucl Med* **44**, 682-689 (2003).

40. Saltiel, A.R. & Kahn, C.R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799-806 (2001).

41. Sonksen, P. & Sonksen, J. Insulin: understanding its action in health and disease. *Br J Anaesth* **85**, 69-79 (2000).

42. Goodyear, L.J. & Kahn, B.B. Exercise, glucose transport, and insulin sensitivity. *Annu Rev Med* **49**, 235-261 (1998).

43. Shepherd, P.R. & Kahn, B.B. Glucose transporters and insulin action--implications for insulin resistance and diabetes mellitus. *N Engl J Med* **341**, 248-257 (1999).

44. Abel, E.D., *et al.* Adipose-selective targeting of the GLUT4 gene impairs insulin action in muscle and liver. *Nature* **409**, 729-733 (2001).

45. Michael, M.D., *et al.* Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell* **6**, 87-97 (2000).

46. Bergman, R.N. New concepts in extracellular signaling for insulin action: the single gateway hypothesis. *Recent Prog Horm Res* **52**, 359-385; discussion 385-357 (1997).

47. Bergman, R.N. & Ader, M. Free fatty acids and pathogenesis of type 2 diabetes mellitus. *Trends Endocrinol Metab* **11**, 351-356 (2000).

48. Pilkis, S.J. & Granner, D.K. Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis. *Annu Rev Physiol* **54**, 885-909 (1992).

49. Sutherland, C., O'Brien, R.M. & Granner, D.K. New connections in the regulation of PEPCK gene expression by insulin. *Philos Trans R Soc Lond B Biol Sci* **351**, 191-199 (1996).

50. Petersen, K.F., Price, T., Cline, G.W., Rothman, D.L. & Shulman, G.I. Contribution of net hepatic glycogenolysis to glucose production during the early postprandial period. *Am J Physiol* **270**, E186-191 (1996).

51. Hellerstein, M.K., *et al.* Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study. *J Clin Invest* **100**, 1305-1319 (1997).

52. Miller, S.P., *et al.* Characterization of glucokinase mutations associated with maturity-onset diabetes of the young type 2 (MODY-2): different glucokinase defects lead to a common phenotype. *Diabetes* **48**, 1645-1651 (1999).

53. Matschinsky, F., *et al.* Glucokinase as pancreatic beta cell glucose sensor and diabetes gene. *J Clin Invest* **92**, 2092-2098 (1993).

54. Matschinsky, F.M., Glaser, B. & Magnuson, M.A. Pancreatic beta-cell glucokinase: closing the gap between theoretical concepts and experimental realities. *Diabetes* **47**, 307-315 (1998).

55. Kahn, B.B. & Flier, J.S. Obesity and insulin resistance. *J Clin Invest* **106**, 473-481 (2000).

56. Ye, J. Mechanisms of insulin resistance in obesity. *Front Med* **7**, 14-24 (2013).

57. Goyal, M.S., *et al.* Loss of Brain Aerobic Glycolysis in Normal Human Aging. *Cell Metab* **26**, 353-360 e353 (2017).

58. Kalpouzos, G., *et al.* Voxel-based mapping of brain gray matter volume and glucose metabolism profiles in normal aging. *Neurobiol Aging* **30**, 112-124 (2009).

59. Gkogkolou, P. & Bohm, M. Advanced glycation end products: Key players in skin aging? *Dermatoendocrinol* **4**, 259-270 (2012).

60. Kalyani, R.R. & Egan, J.M. Diabetes and altered glucose metabolism with aging. *Endocrinol Metab Clin North Am* **42**, 333-347 (2013).

61. Barzilai, N., Huffman, D.M., Muzumdar, R.H. & Bartke, A. The critical role of metabolic pathways in aging. *Diabetes* **61**, 1315-1322 (2012).

62. Defronzo, R.A. Glucose intolerance and aging: evidence for tissue insensitivity to insulin. *Diabetes* **28**, 1095-1101 (1979).

63. Fu, W., Shi, D., Westaway, D. & Jhamandas, J.H. Bioenergetic mechanisms in astrocytes may contribute to amyloid plaque deposition and toxicity. *J Biol Chem* **290**, 12504-12513 (2015).

64. Dunn, L.*, et al.* Dysregulation of glucose metabolism is an early event in sporadic Parkinson's disease. *Neurobiol Aging* **35**, 1111-1115 (2014).

65. Peppard, R.F.*, et al.* Cerebral glucose metabolism in Parkinson's disease with and without dementia. *Arch Neurol* **49**, 1262-1268 (1992).

66. Goodey, N.M. & Benkovic, S.J. Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* **4**, 474-482 (2008).

67. Walsh, C. Posttranslational modification of proteins : expanding nature's inventory. *Roberts and Co. Publishers, Englewood, Colo.* (2006).

68. Lin, H., Su, X. & He, B. Protein lysine acylation and cysteine succination by intermediates of energy metabolism. *ACS Chem Biol* **7**, 947-960 (2012).

69. Wang, L., Tang, Y., Cole, P.A. & Marmorstein, R. Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. *Curr Opin Struct Biol* **18**, 741-747 (2008).

70. Du, J.*, et al.* Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334**, 806-809 (2011).

71. Zhang, Z.*, et al.* Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol* **7**, 58-63 (2011).

72. Frizzell, N., Lima, M. & Baynes, J.W. Succination of proteins in diabetes. *Free Radic Res* **45**, 101-109 (2011).

73. Kim, S.C.*, et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**, 607-618 (2006).

74. Peng, C.*, et al.* The first identification of lysine malonylation substrates and its regulatory enzyme. *Mol Cell Proteomics* **10**, M111 012658 (2011).

75. Wang, Q.*, et al.* Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* **327**, 1004-1007 (2010).

76. Zhao, S.*, et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-1004 (2010).

77. Nakagawa, T., Lomb, D.J., Haigis, M.C. & Guarente, L. SIRT5 Deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle. *Cell* **137**, 560-570 (2009).

78. Choudhary, C.*, et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840 (2009).

79. Taunton, J., Hassig, C.A. & Schreiber, S.L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411 (1996).

80. Haberland, M., Montgomery, R.L. & Olson, E.N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* **10**, 32-42 (2009).

81. Yang, X.J. & Seto, E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol* **9**, 206-218 (2008).

82. Hirschey, M.D.*, et al.* SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation. *Nature* **464**, 121-125 (2010).

83. Yu, W., Dittenhafer-Reed, K.E. & Denu, J.M. SIRT3 protein deacetylates isocitrate dehydrogenase 2 (IDH2) and regulates mitochondrial redox status. *J Biol Chem* **287**, 14078-14086 (2012).

84. Jiang, H.*, et al.* SIRT6 regulates TNF-alpha secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **496**, 110-113 (2013).

85. Wagner, G.R.*, et al.* A Class of Reactive Acyl-CoA Species Reveals the Non-enzymatic Origins of Protein Acylation. *Cell Metab* **25**, 823-837 e828 (2017).

86. Paik, W.K., Pearson, D., Lee, H.W. & Kim, S. Nonenzymatic acetylation of histones with acetyl-CoA. *Biochim Biophys Acta* **213**, 513-522 (1970).

87. Wagner, G.R. & Payne, R.M. Widespread and enzyme-independent Nepsilon-acetylation and Nepsilon-succinylation of proteins in the chemical conditions of the mitochondrial matrix. *J Biol Chem* **288**, 29036-29045 (2013).

88. Weinert, B.T.*, et al.* Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. *Mol Syst Biol* **10**, 716 (2014).

89. Wagner, G.R. & Hirschey, M.D. Nonenzymatic protein acylation as a carbon stress regulated by sirtuin deacylases. *Mol Cell* **54**, 5-16 (2014).

90. Tanner, K.G.*, et al.* Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J Biol Chem* **274**, 18157-18160 (1999).

91. Weinert, B.T.*, et al.* Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli. *Mol Cell* **51**, 265-272 (2013).

92. Casey, J.R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat Rev Mol Cell Biol* **11**, 50-61 (2010).

93. Hansford, R.G. & Johnson, R.N. The steady state concentrations of coenzyme A-SH and coenzyme A thioester, citrate, and isocitrate during tricarboxylate cycle oxidations in rabbit heart mitochondria. *J Biol Chem* **250**, 8361-8375 (1975).

94. Garland, P.B., Shepherd, D. & Yates, D.W. Steady-state concentrations of coenzyme A, acetyl-coenzyme A and long-chain fatty acyl-coenzyme A in rat-liver mitochondria oxidizing palmitate. *Biochem J* **97**, 587-594 (1965).

95. Cai, L., Sutter, B.M., Li, B. & Tu, B.P. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell* **42**, 426-437 (2011).

96. Baeza, J., Smallegan, M.J. & Denu, J.M. Site-specific reactivity of nonenzymatic lysine acetylation. *ACS Chem Biol* **10**, 122-128 (2015).

97. Kulkarni, R.A.*, et al.* Discovering Targets of Non-enzymatic Acylation by Thioester Reactivity Profiling. *Cell Chem Biol* **24**, 231-242 (2017).

98. Moellering, R.E. & Cravatt, B.F. Functional lysine modification by an intrinsically reactive primary glycolytic metabolite. *Science* **341**, 549-553 (2013).

99. Schiff, H. Mittheilungen aus dem Universitäts-laboratorium in Pisa: 2. Eine neue Reihe organischer Basen. *Annalen der Chemie und Pharmacie* **131**, 118-119 (1864).

100. Schiff, H. Eine neue Reihe organischer Diamine. Zweite Abtheilung. *Annalen der Chemie und Pharmacie* **140**, 92-137 (1866).

101. Koenig, R.J. & Cerami, A. Hemoglobin A Ic and diabetes mellitus. *Annu Rev Med* **31**, 29-34 (1980).

102. Grandhee, S.K. & Monnier, V.M. Mechanism of formation of the Maillard protein cross-link pentosidine. Glucose, fructose, and ascorbate as pentosidine precursors. *J Biol Chem* **266**, 11649-11653 (1991).

103. Singh, R., Barden, A., Mori, T. & Beilin, L. Advanced glycation end-products: a review. *Diabetologia* **44**, 129-146 (2001).

104. Lo, T.W., Westwood, M.E., McLellan, A.C., Selwood, T. & Thornalley, P.J. Binding and modification of proteins by methylglyoxal under physiological conditions. A kinetic and mechanistic study with N alpha-acetylarginine, N alpha-acetylcysteine, and N alpha-acetyllysine, and bovine serum albumin. *J Biol Chem* **269**, 32299-32305 (1994).

105. Rabbani, N. & Thornalley, P.J. Measurement of methylglyoxal by stable isotopic dilution analysis LC-MS/MS with corroborative prediction in physiological samples. *Nat Protoc* **9**, 1969-1979 (2014).

106. Zender, L.*, et al.* An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135**, 852-864 (2008).

107. Guo, Y.*, et al.* Effects of methylglyoxal and glyoxalase I inhibition on breast cancer cells proliferation, invasion, and apoptosis through modulation of MAPKs, MMP9, and Bcl-2. *Cancer Biol Ther* **17**, 169-180 (2016).

108. Hosoda, F.*, et al.* Integrated genomic and functional analyses reveal glyoxalase I as a novel metabolic oncogene in human gastric cancer. *Oncogene* **34**, 1196-1206 (2015).

109. Sullivan, L.B., Gui, D.Y. & Vander Heiden, M.G. Altered metabolite levels in cancer: implications for tumour biology and cancer therapy. *Nat Rev Cancer* **16**, 680-693 (2016).

110. Kulkarni, R.A.*, et al.* A chemoproteomic portrait of the oncometabolite fumarate. *Nat Chem Biol* **15**, 391-400 (2019).

111. Alderson, N.L.*, et al.* S-(2-Succinyl)cysteine: a novel chemical modification of tissue proteins by a Krebs cycle intermediate. *Arch Biochem Biophys* **450**, 1-8 (2006).

112. Blatnik, M., Thorpe, S.R. & Baynes, J.W. Succination of proteins by fumarate: mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes. *Ann N Y Acad Sci* **1126**, 272-275 (2008).

113. Yang, M., Soga, T., Pollard, P.J. & Adam, J. The emerging role of fumarate as an oncometabolite. *Front Oncol* **2**, 85 (2012).

114. Yun, J.*, et al.* Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. *Science* **350**, 1391-1396 (2015).

115. Mustafa, A.K.*, et al.* H2S signals through protein S-sulfhydration. *Sci Signal* **2**, ra72 (2009).

116. Richarme, G.*, et al.* Guanine glycation repair by DJ-1/Park7 and its bacterial homologs. *Science* **357**, 208-211 (2017).

117. Zheng, Q.*, et al.* Reversible histone glycation is associated with disease-related changes in chromatin architecture. *Nat Commun* **10**, 1289 (2019).

118. Kansanen, E., Kuosmanen, S.M., Leinonen, H. & Levonen, A.L. The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biol* **1**, 45-49 (2013).

119. Taguchi, K., Motohashi, H. & Yamamoto, M. Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution. *Genes Cells* **16**, 123-140 (2011).

120. Zhang, D.D., Lo, S.C., Cross, J.V., Templeton, D.J. & Hannink, M. Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. *Mol Cell Biol* **24**, 10941-10953 (2004).

121. Itoh, K.*, et al.* Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* **13**, 76-86 (1999).

122. Kobayashi, A.*, et al.* Oxidative and electrophilic stresses activate Nrf2 through inhibition of ubiquitination activity of Keap1. *Mol Cell Biol* **26**, 221-229 (2006).

123. Lo, S.C., Li, X., Henzl, M.T., Beamer, L.J. & Hannink, M. Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* **25**, 3605-3617 (2006).

124. Baird, L., Lleres, D., Swift, S. & Dinkova-Kostova, A.T. Regulatory flexibility in the Nrf2-mediated stress response is conferred by conformational cycling of the Keap1-Nrf2 protein complex. *Proc Natl Acad Sci U S A* **110**, 15259-15264 (2013).

125. Nguyen, T., Sherratt, P.J. & Pickett, C.B. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu Rev Pharmacol Toxicol* **43**, 233-260 (2003).

126. Jaramillo, M.C. & Zhang, D.D. The emerging role of the Nrf2-Keap1 signaling pathway in cancer. *Genes Dev* **27**, 2179-2191 (2013).

127. Uruno, A.*, et al.* The Keap1-Nrf2 system prevents onset of diabetes mellitus. *Mol Cell Biol* **33**, 2996-3010 (2013).

128. Long, M.*, et al.* An Essential Role of NRF2 in Diabetic Wound Healing. *Diabetes* **65**, 780-793 (2016).

129. Khor, T.O.*, et al.* Nrf2-deficient mice have an increased susceptibility to dextran sulfate sodium-induced colitis. *Cancer Res* **66**, 11580-11584 (2006).

130. Scannevin, R.H., *et al.* Fumarates promote cytoprotection of central nervous system cells against oxidative stress via the nuclear factor (erythroid-derived 2)-like 2 pathway. *J Pharmacol Exp Ther* **341**, 274-284 (2012).

131. Sykiotis, G.P. & Bohmann, D. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. *Dev Cell* **14**, 76-85 (2008).

132. Kansanen, E., Jyrkkanen, H.K. & Levonen, A.L. Activation of stress signaling pathways by electrophilic oxidized and nitrated lipids. *Free Radic Biol Med* **52**, 973-982 (2012).

133. Kansanen, E., Kivela, A.M. & Levonen, A.L. Regulation of Nrf2-dependent gene expression by 15-deoxy-Delta12,14-prostaglandin J2. *Free Radic Biol Med* **47**, 1310-1317 (2009).

134. Magesh, S., Chen, Y. & Hu, L. Small molecule modulators of Keap1-Nrf2-ARE pathway as potential preventive and therapeutic agents. *Med Res Rev* **32**, 687-726 (2012).

135. Boyland, E. & Chasseaud, L.F. Enzyme-catalysed conjugations of glutathione with unsaturated compounds. *Biochem J* **104**, 95-102 (1967).

136. Kubal, G., Meyer, D.J., Norman, R.E. & Sadler, P.J. Investigations of glutathione conjugation in vitro by 1H NMR spectroscopy. Uncatalyzed and glutathione transferase-catalyzed reactions. *Chem Res Toxicol* **8**, 780-791 (1995).

137. Schmidt, T.J., Ak, M. & Mrowietz, U. Reactivity of dimethyl fumarate and methylhydrogen fumarate towards glutathione and N-acetyl-L-cysteine--preparation of S-substituted thiosuccinic acid esters. *Bioorg Med Chem* **15**, 333-342 (2007).

138. Cleasby, A., *et al.* Structure of the BTB domain of Keap1 and its interaction with the triterpenoid antagonist CDDO. *PLoS One* **9**, e98896 (2014).

139. de Zeeuw, D., *et al.* Bardoxolone methyl in type 2 diabetes and stage 4 chronic kidney disease. *N Engl J Med* **369**, 2492-2503 (2013).

140. Tayek, J.A. & Kalantar-Zadeh, K. The extinguished BEACON of bardoxolone: not a Monday morning quarterback story. *Am J Nephrol* **37**, 208-211 (2013).

141. Baird, L. & Dinkova-Kostova, A.T. The cytoprotective role of the Keap1-Nrf2 pathway. *Arch Toxicol* **85**, 241-272 (2011).

142. Hur, W., *et al.* A small-molecule inducer of the antioxidant response element. *Chem Biol* **17**, 537-547 (2010).

143. Prochaska, H.J., Bregman, H.S., De Long, M.J. & Talalay, P. Specificity of induction of cancer protective enzymes by analogues of tert-butyl-4-hydroxyanisole (BHA). *Biochem Pharmacol* **34**, 3909-3914 (1985).

144. Li, Y., *et al.* Copper redox-dependent activation of 2-tert-butyl(1,4)hydroquinone: formation of reactive oxygen species and induction of oxidative DNA damage in isolated DNA and cultured rat hepatocytes. *Mutat Res* **518**, 123-133 (2002).

145. Brader, G., Mikkelsen, M.D., Halkier, B.A. & Tapio Palva, E. Altering glucosinolate profiles modulates disease resistance in plants. *Plant J* **46**, 758-767 (2006).

146. Weerapana, E., *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790-795 (2010).

147. Abo, M., Bak, D.W. & Weerapana, E. Optimization of Caged Electrophiles for Improved Monitoring of Cysteine Reactivity in Living Cells. *Chembiochem* **18**, 81-84 (2017).

148. Abo, M. & Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J Am Chem Soc* **137**, 7087-7090 (2015).

149. Bollong, M.J., *et al.* A metabolite-derived protein modification integrates glycolysis with KEAP1-NRF2 signalling. *Nature* **562**, 600-604 (2018).

150. Saw, C.L., *et al.* Impact of Nrf2 on UVB-induced skin inflammation/photoprotection and photoprotective effect of sulforaphane. *Mol Carcinog* **50**, 479-486 (2011).

151. Tao, S., Justiniano, R., Zhang, D.D. & Wondrak, G.T. The Nrf2-inducers tanshinone I and dihydrotanshinone protect human skin cells and reconstructed human skin against solar simulated UV. *Redox Biol* **1**, 532-541 (2013).

152. El-Abaseri, T.B., Putta, S. & Hansen, L.A. Ultraviolet irradiation induces keratinocyte proliferation and epidermal hyperplasia through the activation of the epidermal growth factor receptor. *Carcinogenesis* **27**, 225-231 (2006).

153. Martinez Molina, D., *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84-87 (2013).

154. Savitski, M.M., *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).

155. Huber, K.V., *et al.* Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. *Nat Methods* **12**, 1055-1057 (2015).

156. Ong, S.E., *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

157. Bucher, T. Phosphoglycerate Kinase from Brewer's Yeast. *Methods in enzymology* 415-422 (1955).

158. Vopel, T. & Makhatadze, G.I. Enzyme activity in the crowded milieu. *PLoS One* **7**, e39418 (2012).

159. Chang, J.W., Lee, G., Coukos, J.S. & Moellering, R.E. Profiling Reactive Metabolites via Chemical Trapping and Targeted Mass Spectrometry. *Anal Chem* **88**, 6658-6661 (2016).

160. Ogura, T.*, et al.* Keap1 is a forked-stem dimer structure with two large spheres enclosing the intervening, double glycine repeat, and C-terminal domains. *Proc Natl Acad Sci U S A* **107**, 2842-2847 (2010).

161. Wakabayashi, N.*, et al.* Protection against electrophile and oxidant stress by induction of the phase 2 response: fate of cysteines of the Keap1 sensor modified by inducers. *Proc Natl Acad Sci U S A* **101**, 2040-2045 (2004).

162. Zhang, D.D. & Hannink, M. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol Cell Biol* **23**, 8137-8151 (2003).

163. Rachakonda, G.*, et al.* Covalent modification at Cys151 dissociates the electrophile sensor Keap1 from the ubiquitin ligase CUL3. *Chem Res Toxicol* **21**, 705-710 (2008).

164. Fourquet, S., Guerois, R., Biard, D. & Toledano, M.B. Activation of NRF2 by nitrosative agents and H2O2 involves KEAP1 disulfide formation. *J Biol Chem* **285**, 8463-8471 (2010).

165. Rabbani, N. & Thornalley, P.J. Dicarbonyl proteome and genome damage in metabolic and vascular disease. *Biochem Soc Trans* **42**, 425-432 (2014).

166. Takahashi, K. The reaction of phenylglyoxal with arginine residues in proteins. *J Biol Chem* **243**, 6171-6179 (1968).

167. Yankeelov, J.A., Jr. [52] Modification of arginine by diketones. *Methods Enzymol* **25**, 566-579 (1972).

168. Ravichandran, M.*, et al.* Impairing L-Threonine Catabolism Promotes Healthspan through Methylglyoxal-Mediated Proteohormesis. *Cell Metab* **27**, 914-925 e915 (2018).

169. Moraru, A.*, et al.* Elevated Levels of the Reactive Metabolite Methylglyoxal Recapitulate Progression of Type 2 Diabetes. *Cell Metab* **27**, 926-934 e928 (2018).

170. Canning, P., Sorrell, F.J. & Bullock, A.N. Structural basis of Keap1 interactions with Nrf2. *Free Radic Biol Med* **88**, 101-107 (2015).

171. Eggler, A.L., Liu, G., Pezzuto, J.M., van Breemen, R.B. & Mesecar, A.D. Modifying specific cysteines of the electrophile-sensing human Keap1 protein is insufficient to disrupt binding to the Nrf2 domain Neh2. *Proc Natl Acad Sci U S A* **102**, 10070-10075 (2005).

172. Galligan, J.J., *et al.* Methylglyoxal-derived posttranslational arginine modifications are abundant histone marks. *Proc Natl Acad Sci U S A* **115**, 9228-9233 (2018).

173. Nishimoto, S., Koike, S., Inoue, N., Suzuki, T. & Ogasawara, Y. Activation of Nrf2 attenuates carbonyl stress induced by methylglyoxal in human neuroblastoma cells: Increase in GSH levels is a critical event for the detoxification mechanism. *Biochem Biophys Res Commun* **483**, 874-879 (2017).

174. Mitsuishi, Y., *et al.* Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. *Cancer Cell* **22**, 66-79 (2012).

175. Hur, W. & Gray, N.S. Small molecule modulators of antioxidant response pathway. *Curr Opin Chem Biol* **15**, 162-173 (2011).

176. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-2860 (2006).

177. Ruifrok, A.C. & Johnston, D.A. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* **23**, 291-299 (2001).

178. Ideker, T. & Krogan, N.J. Differential network biology. *Mol Syst Biol* **8**, 565 (2012).

179. Jager, S., *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365-370 (2011).

180. Krogan, N.J., *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-643 (2006).

181. Gavin, A.C., *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-636 (2006).

182. Scott, J.D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220-1224 (2009).

183. Foltz, D.R., *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* **8**, 458-469 (2006).

184. Lavallee-Adam, M., *et al.* Discovery of cell compartment specific protein-protein interactions using affinity purification combined with tandem mass spectrometry. *J Proteome Res* **12**, 272-281 (2013).

185. Bisson, N., *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol* **29**, 653-658 (2011).

186. Collins, B.C., *et al.* Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods* **10**, 1246-1253 (2013).

187. 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* **196**, 801-810 (2012).

188. Gupta, G.D., *et al.* A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **163**, 1484-1499 (2015).

189. Kim, D.I., *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A* **111**, E2453-2461 (2014).

190. Uezu, A., *et al.* Identification of an elaborate complex mediating postsynaptic inhibition. *Science* **353**, 1123-1129 (2016).

191. Opitz, N., *et al.* Capturing the Asc1p/Receptor for Activated C Kinase 1 (RACK1) Microenvironment at the Head Region of the 40S Ribosome with Quantitative BioID in Yeast. *Mol Cell Proteomics* **16**, 2199-2218 (2017).

192. Lin, Q., *et al.* Screening of Proximal and Interacting Proteins in Rice Protoplasts by Proximity-Dependent Biotinylation. *Front Plant Sci* **8**, 749 (2017).

193. Han, S., Li, J. & Ting, A.Y. Proximity labeling: spatially resolved proteomic mapping for neurobiology. *Curr Opin Neurobiol* **50**, 17-23 (2018).

194. Branon, T.C., *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880-887 (2018).

195. Hill, Z.B., Pollock, S.B., Zhuang, M. & Wells, J.A. Direct Proximity Tagging of Small Molecule Protein Targets Using an Engineered NEDD8 Ligase. *J Am Chem Soc* **138**, 13123-13126 (2016).

196. Liu, Q., *et al.* A proximity-tagging system to identify membrane protein-protein interactions. *Nat Methods* **15**, 715-722 (2018).

197. Lam, S.S., *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51-54 (2015).

198. Martell, J.D., *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143-1148 (2012).

199. Paek, J., *et al.* Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**, 338-349 e311 (2017).

200. Chu, Q.*, et al.* Identification of Microprotein-Protein Interactions via APEX Tagging. *Biochemistry* **56**, 3299-3306 (2017).

201. Han, S.*, et al.* Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem Biol* **24**, 404-414 (2017).

202. Myers, S.A.*, et al.* Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat Methods* **15**, 437-439 (2018).

203. Hung, V.*, et al.* Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell* **55**, 332-341 (2014).

204. Lee, S.Y.*, et al.* APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell Rep* **15**, 1837-1847 (2016).

205. Reinke, A.W., Balla, K.M., Bennett, E.J. & Troemel, E.R. Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. *Nat Commun* **8**, 14023 (2017).

206. Reinke, A.W., Mak, R., Troemel, E.R. & Bennett, E.J. In vivo mapping of tissue- and subcellular-specific proteomes in Caenorhabditis elegans. *Sci Adv* **3**, e1602426 (2017).

207. Lobingier, B.T.*, et al.* An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **169**, 350-360 e312 (2017).

208. Honke, K. & Kotani, N. The enzyme-mediated activation of radical source reaction: a new approach to identify partners of a given molecule in membrane microdomains. *J Neurochem* **116**, 690-695 (2011).

209. Li, X.W.*, et al.* New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. *J Biol Chem* **289**, 14434-14447 (2014).

210. Mollwitz, B.*, et al.* Directed evolution of the suicide protein O(6)-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986-994 (2012).

211. Juillerat, A.*, et al.* Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem Biol* **10**, 313-317 (2003).

212. Crivat, G. & Taraska, J.W. Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol* **30**, 8-16 (2012).

213. Mie, M., Naoki, T., Uchida, K. & Kobatake, E. Development of a split SNAP-tag protein complementation assay for visualization of protein-protein interactions in living cells. *Analyst* **137**, 4760-4765 (2012).

214. Maurel, D.*, et al.* Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat Methods* **5**, 561-567 (2008).

215. Bodor, D.L., Rodriguez, M.G., Moreno, N. & Jansen, L.E. Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. *Curr Protoc Cell Biol* **Chapter 8**, Unit8 8 (2012).

216. Chidley, C., Haruki, H., Pedersen, M.G., Muller, E. & Johnsson, K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol* **7**, 375-383 (2011).

217. Holze, C.*, et al.* Oxeiptosis, a ROS-induced caspase-independent apoptosis-like cell-death pathway. *Nat Immunol* **19**, 130-140 (2018).

218. Lo, S.C. & Hannink, M. PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* **281**, 37893-37903 (2006).

219. O'Mealey, G.B.*, et al.* A PGAM5-KEAP1-Nrf2 complex is required for stress-induced mitochondrial retrograde trafficking. *J Cell Sci* **130**, 3467-3480 (2017).

220. Sykiotis, G.P. & Bohmann, D. Stress-activated cap'n'collar transcription factors in aging and human disease. *Sci Signal* **3**, re3 (2010).

221. Zhang, Y., Burdzinski, G., Kubicki, J. & Platz, M.S. Direct observation of carbene and diazo formation from aryldiazirines by ultrafast infrared spectroscopy. *J Am Chem Soc* **130**, 16134-16135 (2008).

222. Noller, B.*, et al.* Ultrafast dynamics of isolated phenylcarbenes followed by femtosecond time-resolved velocity map imaging. *J Phys Chem A* **113**, 3041-3050 (2009).

223. Noller, B.*, et al.* The photoionisation of two phenylcarbenes and their diazirine precursors investigated using synchrotron radiation. *Phys Chem Chem Phys* **11**, 5384-5391 (2009).

224. Brunner, J., Senn, H. & Richards, F.M. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J Biol Chem* **255**, 3313-3318 (1980).

225. An, W.F. & Tolliday, N. Cell-based assays for high-throughput screening. *Mol Biotechnol* **45**, 180-186 (2010).

226. Srinivasan, B., Tonddast-Navaei, S. & Skolnick, J. Ligand binding studies, preliminary structure-activity relationship and detailed mechanistic characterization of 1-phenyl-6,6-dimethyl-1,3,5-triazine-2,4-diamine derivatives as inhibitors of Escherichia coli dihydrofolate reductase. *Eur J Med Chem* **103**, 600-614 (2015).

227. Kendrew, J.C., *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).

228. Scapin, G. Structural biology and drug discovery. *Curr Pharm Des* **12**, 2087-2097 (2006).

229. Wuthrich, K. The way to NMR structures of proteins. *Nat Struct Biol* **8**, 923-925 (2001).

230. Laskowski, R.A. Structural quality assurance. *Methods Biochem Anal* **44**, 273-303 (2003).

231. Renaud, J.P., *et al.* Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov* **17**, 471-492 (2018).

232. Moellering, R.E. & Cravatt, B.F. How chemoproteomics can enable drug discovery and development. *Chem Biol* **19**, 11-22 (2012).

233. Parker, C.G., *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541 e529 (2017).

234. Yip, G.M., *et al.* A propofol binding site on mammalian GABAA receptors identified by photolabeling. *Nat Chem Biol* **9**, 715-720 (2013).

235. Gao, J., Mfuh, A., Amako, Y. & Woo, C.M. Small Molecule Interactome Mapping by Photoaffinity Labeling Reveals Binding Site Hotspots for the NSAIDs. *J Am Chem Soc* **140**, 4259-4268 (2018).

236. Uphoff, C.C., Habig, S., Fombonne, S., Matsuo, Y. & Drexler, H.G. ABL-BCR expression in BCR-ABL-positive human leukemia cell lines. *Leuk Res* **23**, 1055-1060 (1999).

237. Shaul, Y. & Ben-Yehoyada, M. Role of c-Abl in the DNA damage stress response. *Cell Res* **15**, 33-35 (2005).

238. Yuan, Z.M., *et al.* Regulation of DNA damage-induced apoptosis by the c-Abl tyrosine kinase. *Proc Natl Acad Sci U S A* **94**, 1437-1440 (1997).

239. Gambacorti-Passerini, C. Part I: Milestones in personalised medicine--imatinib. *Lancet Oncol* **9**, 600 (2008).

240. Nagar, B., *et al.* Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res* **62**, 4236-4243 (2002).

241. Clark, M.J., *et al.* GNF-2 Inhibits Dengue Virus by Targeting Abl Kinases and the Viral E Protein. *Cell Chem Biol* **23**, 443-452 (2016).

242. Zhang, J., *et al.* Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501-506 (2010).

243. Choi, Y., *et al.* N-myristoylated c-Abl tyrosine kinase localizes to the endoplasmic reticulum upon binding to an allosteric inhibitor. *J Biol Chem* **284**, 29005-29014 (2009).

244. Adrian, F.J., *et al.* Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* **2**, 95-102 (2006).

245. Zhuang, M., Guan, S., Wang, H., Burlingame, A.L. & Wells, J.A. Substrates of IAP ubiquitin ligases identified with a designed orthogonal E3 ligase, the NEDDylator. *Mol Cell* **49**, 273-282 (2013).

246. Gan, L., Johnson, D.A. & Johnson, J.A. Keap1-Nrf2 activation in the presence and absence of DJ-1. *Eur J Neurosci* **31**, 967-977 (2010).

247. Mills, E.L., *et al.* Itaconate is an anti-inflammatory metabolite that activates Nrf2 via alkylation of KEAP1. *Nature* **556**, 113-117 (2018).

248. Sibbersen, C., *et al.* Profiling of Methylglyoxal Blood Metabolism and Advanced Glycation End-Product Proteome Using a Chemical Probe. *ACS Chem Biol* **13**, 3294-3305 (2018).