Title
Bioreactive Unnatural Amino Acids as Tools for Probing Protein-Protein Interactions

Author
Berdan, Viktoriya Y

Publication Date
2022

Peer reviewed|Thesis/dissertation

Bioreactive Unnatural Amino Acids as Tools for Probing Protein-Protein Interactions

by
Viktoriya Berdan

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
EA8EE3D104C542D...
Lei Wang
Chair

DocuSigned by:
Danica Galonic Fujimori
Danica Galonic Fujimori

DocuSigned by:
Natalia Jura
E438674A382B42F...
Natalia Jura

Committee Members

Acknowledgements

I am grateful to the people who made it possible for me to get to where I am today. I want to thank my lab for all their scientific support and for making our work environment a pleasant one to be in day in and day out. Thank you for all the fun that we managed to have together as a group, despite COVID trying its hardest to keep us from going to any more hot pots together. Thank you to my PI, Prof. Lei Wang, for allowing me to pursue a project that interested me scientifically, in a field that I have been interested in since I first learned about chemical biology, since I first knew I wanted to get my Ph.D. To my thesis committee members, Prof. Danica Fujimori and Prof. Natalia Jura, who gave me the confidence to believe in myself as a scientist, specifically as a female scientist, when I did not have any other female mentors to inspire me. Thank you especially to Prof. Fujimori: rotating in your lab was the first time I felt like I could handle graduate school, like I actually had what it took to become a good scientist.

To the mentors I had before UCSF, specifically my undergraduate PI, Prof. Bobby Arora, for taking a chance on me and allowing me to conduct research in his lab, and for telling me I had to apply to graduate school, even though I still had no clue where my future was headed. Your advice and encouragement helped guide me. To Lt. Col. Wint, the OG mentor. Everyone needs mentors who believe in you more than you believe in yourself, and you were the first of many to do so.

To my family, for supporting me despite having no clue what I have been doing with my life for the past many, many years. Thank you to my parents, Yuriy and Yelena, who, despite some reluctances towards letting me move to the opposite end of the country, have always looked out for me and tried their best to support me. I will always be grateful for every sacrifice you have

ever made for me. To my sister, Tanya, thank you for existing and preventing me from developing Only Child Syndrome. Thank you for being my mirror image and also my polar opposite, and for always going first in every new thing we had to figure out in this country.

To my friends, for keeping me sane throughout the whole Ph.D. process. Thank you to the new friends I made during my time in the program, who understood exactly what I was going through. Thank you to my old friends, especially Laurel, who remind me of who I used to be and still like who I've become. Lastly, thank you to my wonderful partner and best friend, P.C.K., for his endless support both at the bench and at home. I really couldn't have made it this far without you. Thank you.

Contributions

The work presented in chapter 1 of this thesis was conducted primarily by V.Y.B., with the assistance of Prof. Lei Wang in analyzing data and planning experiments. Initial construction of human growth hormone plasmid is attributed to Dr. Jun Liu. Ba/F3 GHR+ cells were provided by Prof. Jim Wells' lab.

Chapter 2 of this thesis is a reprint of the material as it appears in:

Berdan, V.Y.; Klauser, P.C.; Wang, L. Covalent Peptides and Proteins for Therapeutics *Bioorganic and Medical Chemistry* 2021. DOI: 10.1016/j.bmc.2020.115896

The work presented in chapter 3 of this thesis was conducted primarily by V.Y.B., with the assistance of Prof. Lei Wang in analyzing data and planning experiments. Analysis of trypsin-digested proteins on Orbi-Nano was conducted by the University of Delaware Mass Spec Core. Analysis of mass spectrometry data was assisted by Prof. Bing Yang.

Bioreactive Unnatural Amino Acids as Tools for Probing Protein-Protein Interactions

By

Viktoriya Berdan

Abstract

The study of protein-protein interactions is essential to our understanding of cell biology and our development of therapeutics for human diseases. Bioreactive unnatural amino acids (Uaas), which can selectively crosslink with specific nucleophilic canonical amino acids, have the potential to change the way we study protein-protein interactions. Bioreactive Uaas can allow us to study and manipulate protein-protein interactions without fear of drastically altering native protein structure, function, or binding capabilities. By engineering a bioreactive Uaa into a protein in order to target another protein, or proteins, of interest, we make a relatively small mutation but endows the mutant protein with an incredibly powerful novel capability to chemically crosslink nucleophilic residues on other proteins. The potential of bioreactive Uaas, though vast, has yet to be proven. The work described in this thesis attempts to showcase some of the potential of bioreactive Uaas as both tools to develop potential bio-therapeutics and as tools to study and identify novel protein-protein interactions.

Chapter 1 describes efforts to elucidate how chemical crosslinking affects the biologics of receptor-ligand interactions. Using human growth hormone as a model system, we investigate whether a native ligand can be engineered to successfully bind and crosslink with its receptor, and how this crosslinking affects the downstream signaling of the receptor-ligand pair.

Chapter 2 is a review of current literature pertaining to the use of chemical crosslinking, including bioreactive unnatural amino acids, in the development of peptide and protein

therapeutics. We compile known examples in the literature of peptide inhibitors that have been developed to crosslink proteins of interest and surmise their therapeutic potential. We also introduce the relatively novel concept of covalent protein therapeutics, particularly proteins containing bioreactive unnatural amino acids which would crosslink the proteins to their binding partners and effect a therapeutic outcome.

Chapter 3 details the major efforts of this doctoral degree in the study of the interactome of mammalian histidine kinases. Through the use of various bioreactive unnatural amino acids, we generate a library of mutants for the proteins NDPKA and NDPKB. We investigate whether these mutants crosslink proteins in mammalian cells and mammalian cell lysates, and, upon finding mutants with crosslinking potential, we utilize different purification techniques to prepare samples for analysis via mass spectrometry. The proteins identified through mass spectrometry analysis have the potential to be substrates or binding partners of the mutant proteins.

Table of Contents

1. Elucidating the biological effects of a covalent human Growth Hormone

1.1 Abstract 1

1.2 Introduction 2

1.3 Results 4

1.4 Discussion 10

1.5 Materials and Methods 11

1.6 Supplemental Figures 15

1.7 References 18

2. Covalent peptides and proteins for therapeutics

2.1 Abstract 21

2.2 Introduction 21

2.3 Advances in Covalent Peptide Drugs 24

2.4 Advances in Covalent Protein Drugs 30

2.5 Conclusions and perspectives 36

2.6 References 37

3. Utilizing Genetically Encoded Chemical Crosslinking for the Discovery of Novel Binding Partners of Mammalian Histidine Protein Kinases

3.1 Abstract 47

3.2 Introduction 48

3.3 Results 51

3.4 Discussion 60

3.5 Materials and Methods 51

3.6 Significant Figures 68

3.7 References 80

List of Figures

Figure 1.1: Crosslinking of hGH with growth hormone receptor. 6

Figure 1.2: Downstream signaling assays comparing WT and mutant hGH proteins. 7

Figure S1.1: FPLC traces for the purification of hGH-WT and hGH-Q68FSY. 15

Figure S1.2: Expression of hGH-WT, hGH-Y103FSY, and hGH-Q68FSY proteins. 15

Figure S1.3: Testing different concentrations of hGH-WT and hGH-Q68FSY. 16

Figure S1.4: hGH-Q68FSY crosslinking on cell surface. 17

Figure 2.1: Covalent crosslinking between the target protein and peptide or protein drug. 23

Figure 2.2: Electrophilic warheads and amino acids utilized in covalent peptide design. 26

Figure 2.3: mSF-SAH crosslinks with MDM2/4. 28

Figure 2.4: Bioreactive Uaas. 33

Figure 2.5: Covalent PD-1 drug. 35

Figure 3.1: Crosslinking of NDPKB mutants. 52

Figure 3.2: Mass spectrometry results of NDPKB-F60FSY mutant. 56

Figure 3.3: Expression of new mutants of NDPKB. 57

Figure 3.4: Expression and purification of NDPKB-N115FSY mutant. 58

Figure 3.5: NDPKA expression in E. coli. 59

Figure S3.1: Crystal structure of NDPKB with mutated residues highlighted. 68

Figure S3.2: Expression of NDPKA mutants in HEK293T cells. 68

Figure S3.3: Time course assay of NDPKB-F60FSY expression in HEK293T cells. 69

Figure S3.4: Purifying NDPKB-F60FSY mutant using IP pull-down. 69

Figure S3.5: Initial mass spectrometry results for NPDKB-F60FSY. 70

Figure S3.6: Mass spectrometry data for NDPKB-F60FSY. 71

Figure S3.7: Co-expression of RNASEL and NDPKB-F60FSY in HEK293T cells. 72

Figure S3.8: Crystal structure of NDPKB showing Y67 and K128 residues. 72

Figure S3.9: NDPKB-F60FSY/Y67F/K128R mutant expression. 73

Figure S3.10: Final mutagenesis campaign on NDPKB. 73

Figure S3.11: Expression of NDPKB-N115FSY and NDPKB-F60FSY in E. coli. 74

Figure S3.12: NDPKB-N115FSY crosslinked peptides. 74

Figure S3.13: Self-crosslinking data for N115FSY mutant. 76

Chapter 1: Elucidating the biological effects of a covalent human Growth Hormone

1.1 Abstract

Biological functions are carried out through complex networks of signaling pathways. Many of these intracellular pathways are initiated by the binding of extracellular ligands to cell-surface receptors, leading to downstream signaling cascades. Tight regulation of signaling networks is vital for cell health, and aberrant signaling caused by certain ligand-receptor interactions can lead to the development of various diseases, including cancer. Therefore, there is a pressing need for the development of more precise and accurate methods to study ligand-receptor interactions. Specifically crosslinking ligands to their receptors would provide a novel approach to the study of downstream signaling cascades and other biological properties of ligand-receptor binding, and could potentially lead to the development of better therapeutics that target ligand-receptor pairs. By utilizing amber stop codon suppression, we can incorporate a latent bioreactive unnatural amino acid (Uaa) into a ligand of interest and, through proximity-enabled crosslinking, covalently bind the ligand with its specific receptor. Bioreactive Uaa incorporation has the potential to be a powerful tool for studying a wide range of protein-protein interactions. Here, we report crosslinking between human growth hormone (hGH) and its receptor (hGHR) utilizing a novel proximity-enabled bioreactive Uaa. We attempt to study how this covalent interaction affects downstream signaling, and try to discern if there is any difference in downstream signaling cascades between mutant crosslinking hGH and wild-type protein.

1.2 Introduction

Unnatural amino acids (Uaas) have been incorporated into proteins for over two decades.[1,2] Until recently, however, the vast majority of Uaas being utilized have contained inert or biorthogonal side chains. Bioreactive Uaas, which could potentially chemically interact with canonical amino acids, were not developed as often due to fear that the Uaas would be over-reactive and lead to toxicity in the cells, or would simply be too reactive to allow for the selection of orthogonal synthetase/tRNA pairs. However, recent developments have shown that it is possible to develop bioreactive Uaas that have proximity-enabled reactivity, meaning that they only reactive with nucleophilic residues when in close proximity, such as during a protein-protein interaction.[3,4] Taking a page from the development of covalent drugs, which use electrophilic warheads that are specific for certain nucleophilic groups found on only a handful of native amino acids, the genetic code expansion field has progressed towards the development of various Uaas that selectively target nucleophilic residues such as Cys, Lys, and/or His.[5,6]

Our lab has recently developed one such Uaa, fluorosulfate-L-tyrosine (FSY), which we have shown to react selectively with Lys, His, and Tyr residues in close proximity.[7] This makes FSY particularly unique, in that it can react with more than one nucleophilic residue, when previous bioreactive Uaas have generally targeted only Cys residues or only Lys residues.[4] This allows FSY to target a wider range of protein-protein interactions (PPIs) by not limiting it to PPIs with one specific nucleophilic residue at the interface. This makes FSY a potentially powerful tool, and worthy of further study.

Previous bioreactive Uaa studies have only shown that the Uaas have the ability to crosslink two proteins together, but do not delve into what the biological effects of the crosslink

would be.[7,8] We are interested in exploring the potential of FSY to crosslink an endogenous ligand with its receptor. Receptor-ligand interactions are a particular class of PPIs that could potentially be greatly affected by the introduction of a permanent chemical crosslink between the two proteins. Receptor-ligand interactions are paramount in the activation of downstream signaling cascades that form the basis for the biological function of cells. The affinity of the ligand and the structural changes that occur upon ligand binding all play an important role in the biological response of the receptor and other downstream proteins. Because any changes in binding affinity or protein structure upon ligand binding could drastically change the biological response of the receptor, the introduction of a permanent chemical crosslink in a receptor-ligand pair would be an interesting system to study in order to further understand how bioreactive Uaas can alter biology.

To begin our study of how bioreactive Uaas affect receptor-ligand interactions, we decided to focus on a well-known model system for our biological assays. Therefore, we decided to study the human growth hormone (hGH) and growth hormone receptor (hGHR) pair. The kinetics of the hGH-hGHR interaction have been well studied.[9,10] The downstream signaling pathways of hGH-hGHR are well known.[9,10,11] Previous studies have shown the biological effects of increasing the binding affinity of hGH; there is an upper limit for the biological signaling of the hGHR that, once reached, will no longer be affected by increasing the binding affinity of hGH.[12] There is a point at which increasing the affinity of hGH no longer has any biological effect on the receptor. Using a system with this known limitation to study the effects of introducing a bioreactive Uaa into the interaction is therefore interesting because 1) we can attempt to elucidate the difference between having a ligand with 50-fold higher affinity compared to the original ligand versus a ligand with an effective off rate that is infinitely small, which is the case for a nonreversible crosslinking ligand, and 2) we can try to determine if altering the interaction from noncovalent to covalent

affects the structure or biological function of the PPI without necessarily connecting those changes to a change in affinity. Furthermore, a long term goal of studying the hGH-hGHR interaction and designing hGH mutants with crosslinking ability would be to determine if the Uaa-containing mutants have a biological effect on the downstream signaling pathway of hGHR that could potentially be therapeutic. Growth hormone deficiency affects 1 out of every 4000 kids in the US and can lead to stunted growth and hypoglycemia;[9,13] in recent years, excess hGH has also been implicated in the development, progression, and metastasis of breast cancer.[14] Further studying the hGH-hGHR interaction using bioreactive Uaas, and attempting to alter the downstream signaling of this interaction, would not only give us further essential insight into the effects of bioreactive Uaas on PPIs, but could also lead to the development of a type of mutant hGH that could be used for 1) further study in academic setting of the hGH-hGHR interactions and 2) therapeutic treatment of growth hormone deficiency or even potentially breast cancer treatment.

Thus, having decided to use hGH-hGHR as a model system, we move forward with designing mutant hGH proteins containing our Uaa of choice, FSY, for crosslinking with the growth hormone receptor, testing the crosslinking capabilities of these mutants, then utilizing successful mutants for further cellular assays to study the effects of crosslinking on downstream signaling, in order to gain more insight into the potential bioreactive Uaas might have as tools in the study of PPIs and as tools for the development of potential bio-therapeutics.

1.3 Results

Design, expression, and crosslinking of mutant hGH proteins. Human growth hormone receptor forms a homodimer on the cell surface. The hGH ligand then binds to this dimer to activate downstream signaling. The hGH ligand, however, is asymmetrical, thus it binds to the receptor

first with site 1, which has a higher affinity for the receptor, then with site 2, which has a lower affinity.[10] We decided to make two hGH mutants, with a bioreactive Uaa on either site 1 or site 2, in order to see which site would afford more efficient crosslinking, though we postulated that a bioreactive Uaa in site 1, which binds first and more tightly, would be more likely to crosslink the receptor. We looked at the interaction interface of hGH-hGHR to identify a nucleophilic residue on the receptor which was in close proximity to residues on site 1 and residues on site 2. We identified site Lys166 on the hGHR monomer which was in close proximity to residues Gln68 on site 1 and Tyr103 on site 2 (**Fig 1.1a**). By mutating this residue to our bioreactive Uaa FSY, we hoped to see crosslinking between the mutant hGH proteins and the Lys166 residue of the hGHR. Plasmids containing TAG codon mutations at either site Gln68 or Tyr103 were constructed and co-transformed with the plasmid containing the FSY synthetase/tRNA pair (FSYRS). Mutant hGH proteins were expressed in E. coli cells and purified under previously published conditions (**Fig S1.1 and S1.2**).[15] After purification, in order to test the crosslinking capabilities of the mutant hGH proteins (denoted at hGH-Q68FSY or hGH-Y103FSY), the mutant proteins were incubated overnight at 37C with purified extracellular domain of hGHR. *In vitro* crosslinking results were visualized via Western blotting, which showed that, as previously hypothesized, the hGH-Q68FSY mutant, which has FSY incorporated at the higher affinity site 1, crosslinked with the extracellular domain of hGHR, while the hGH-Y103FSY mutant, which has weaker binding affinity for hGHR, did not show any crosslinking *in vitro* (**Fig 1.1b**). Based upon this initial result, we decided to move forward with the hGH-Q68FSY mutant for further biological study of the effect crosslinking has on the hGH-hGHR interaction. We conducted an *in vitro* time course assay to determine optimal incubation time between hGH mutant and the receptor for crosslinking to occur, and saw that crosslinking was seen as early as after 20 minutes of incubation at 37C but was most robust at

around 1-2 hours of incubation (**Fig 1.1c**). Moving forward, we would continue optimizing incubation times for further cellular studies of downstream signaling.

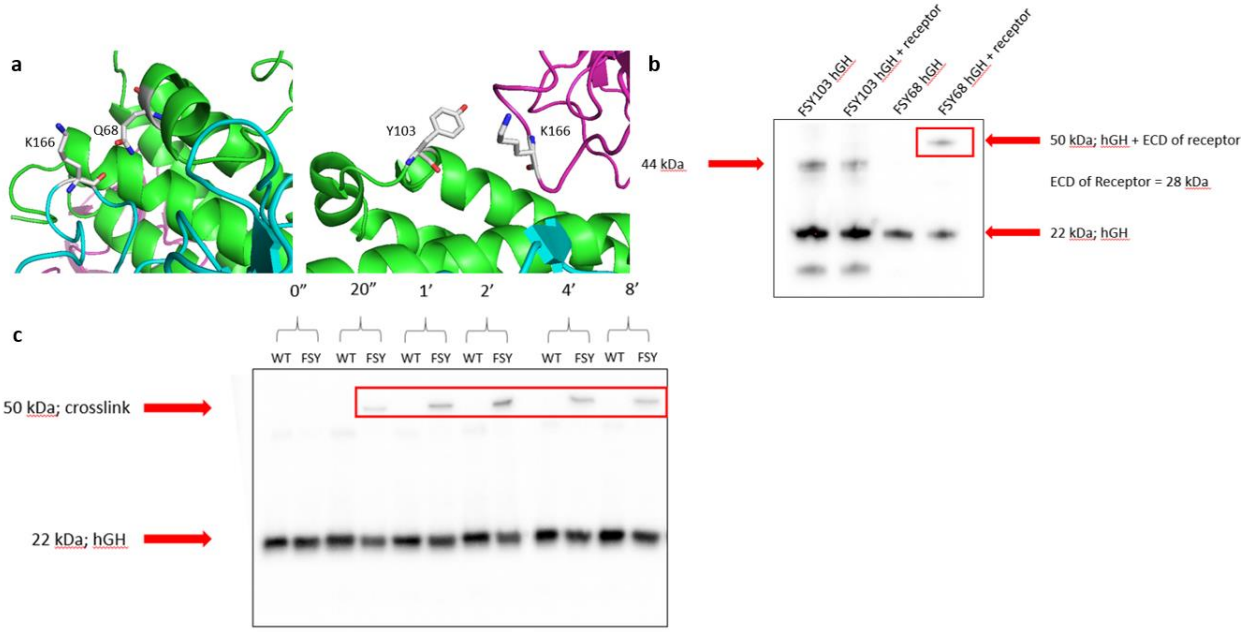


Fig 1.1: Crosslinking of hGH with growth hormone receptor. a) Location of mutant sites Gln68 (left) and Tyr103 (right) in proximity to site Lys166 on hGHR. Human growth hormone ligand is depicted in green, and human growth hormone receptor is shown in either cyan or magenta. PDB: 3HHR. **b)** *In vitro* crosslinking test of hGH-Q68FSY and hGH-Y103FSY. The purified mutant hGH proteins were incubated O/N at 37C either with or without the extracellular domain of the hGHR. Results show crosslinking of hGH-Q68FSY mutant but no crosslinking of hGH-Y103FSY mutant. Anti-His tag Western blot. **c)** *In vitro* time course assay for crosslinking of hGH-Q68FSY mutant and extracellular domain of hGHR. Assay shows crosslinking after 20 minutes, with most robust crosslinking after 2 hours. No crosslinking is seen with the wild type (WT) control of hGH. Anti-His tag Western blot.

Studying the downstream signaling effects of hGH-Q68FSY. Having shown that we are capable of crosslinking a mutant hGH protein containing a bioreactive Uaa to its target receptor *in vitro*, we then moved forward to study how this crosslinking might affect the downstream signaling of the hGHR. Upon binding of hGH to hGHR, several pathways can be activated; one of the most well-studied is the JAK/STAT pathway leading to the transcription of genes important for cellular immunity, proliferation, and apoptosis.[11] Various other studies on hGH-hGHR signaling have used the phosphorylation of STAT5 as a measure of downstream signaling activity.[16-18] Thus, we wanted to initially study how the crosslinking of hGH to hGHR on cell surfaces might alter the

signaling of STAT5/pSTAT5. As a model system, we decided to use the Ba/F3 GHR+ cell line, which is a murine pro B cell line engineered to express hGHR. Ba/F3 cells are IL-3 dependent, so cells have to be serum starved 24 hours before treatment with hGH to de-convolute IL-3 activation of the STAT5 pathway from activation with hGH.[19] Ba/F3 cells were grown up according to previously established protocols. First, we wanted to see if my mutant hGH would even elicit any downstream signaling, or if the crosslinking potentially negates activation of the hGHR. We added hGH-WT or hGH-Q68FSY to serum-starved Ba/F3 cells at a concentration of 100 ng/mL and harvested the cells for Western blot analysis after 1 hour. We saw that our initial study showed similar activation of pSTAT5 by both the WT and the mutant hGH protein, when compared to the negative control cells that had nothing added to activate the pSTAT5 pathway (**Fig 1.2a**).

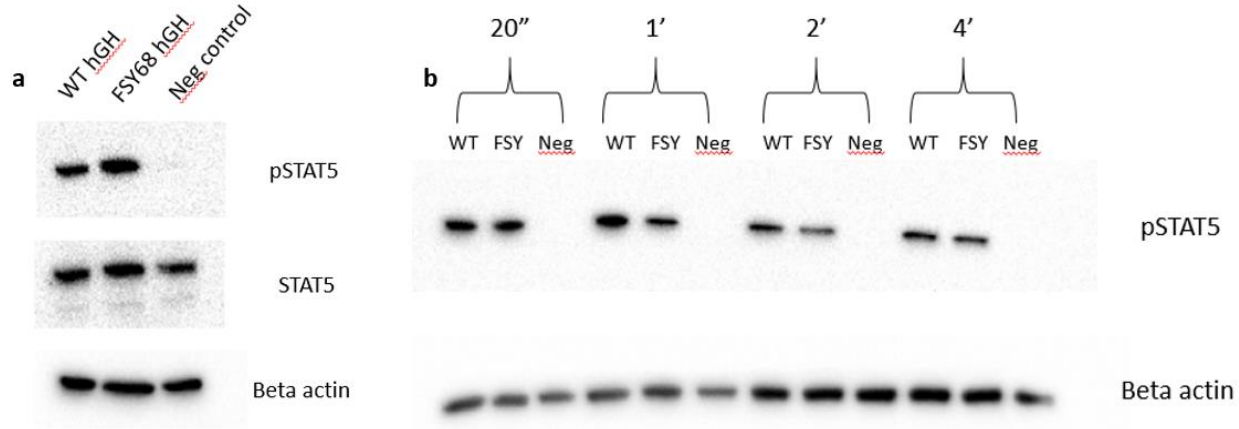


Fig 1.2: Downstream signaling assays comparing WT and mutant hGH proteins. a) Initial study to see if hGH-Q68FSY elicits activation of pSTAT5 pathway. 100 ng/mL of either WT or hGH-Q68FSY mutant was incubated with serum starved Ba/F3 cells for 1 hour. After 1 hour, cells were harvested, lysed, and run on Western. Blocking with anti-pSTAT5 antibody shows similar activation of pSTAT5 pathway for WT and mutant hGH proteins. No activation of pathway is seen in negative control where PBS was added instead of hGH protein. Beta actin is shown as loading control. Blotting with anti-STAT5 antibody shows similar expression of STAT5 across samples. **b)** Time course assay for downstream signaling shows no significant difference between the mutant hGH-Q68FSY protein and the hGH-WT protein at any time point. Western blot analysis with anti-pSTAT5 antibody. Beta actin shown as loading control.

Once it was established that our mutant hGH protein could activate downstream signaling of the pSTAT5 pathway, we decided to conduct a time course cellular assay to determine if there was differences between hGH-WT and hGH-Q68FSY after incubation for longer periods of time. Because our mutant hGH would bind once, crosslink, and remain bound to the receptor, while the WT protein would bind and unbind with the receptor, we postulated that our mutant protein would have a stronger effect on downstream signaling as it might activate the receptor more by binding indefinitely. We incubated the WT protein and the FSY mutant protein for various time points, up to 4 hours, to see if pSTAT5 activation was altered in some way between the two proteins. However, our results showed that there was minimal, if any, difference between the two proteins' effect on pSTAT5 signaling (**Fig 1.2b**). Activation of pSTAT5 is fast and prolonged, regardless of the length of time the two different proteins are incubated.

Once we learned that incubating the same concentration of hGH-Q68FSY and WT for varying lengths of time did not have any noticeable effect on downstream signaling, we wanted to test whether incubating different concentrations of hGH-Q68FSY and hGH-WT for the same length of time would change the downstream signaling. We hypothesized that, due to the crosslinking that occurs between hGH-Q68FSY and the hGHR, a lower concentration of hGH-Q68FSY could potentially elicit the same level of downstream signaling as a higher concentration of hGH-WT. If this is true, lower doses of hGH-Q68FSY could be used to treat growth hormone deficiency disorders, which would be therapeutically beneficial. However, several attempts were made to deduce if different concentrations of hGH-Q68FSY and hGH-WT led to differences in downstream signaling, but all were inconclusive (**Fig. S1.3**). The initial study, which looked at three different concentrations (1 ng/mL, 10 ng/mL, and 100 ng/mL) of each protein incubated with serum starved Ba/F3 cells for 1 hour showed no pSTAT5 activation, despite previous studies

showing that 100 ng/mL of both hGH-WT and hGH-Q68FSY incubated at 1 hour did lead to noticeable activation of pSTAT5 (**Fig S1.3a**). Another study was conducted in parallel with the concentration assays: one set of serum starved Ba/F3 cells were treated with 100 ng/mL of hGH-Q68FSY and hGH-WT for different periods of time, including 1 hour. Another set of serum starved Ba/F3 cells were simulatenously treated with different concentrations of hGH-Q68FSY and hGH-WT (1 ng/mL, 10 ng/mL, and 100 ng/mL) for 1 hour. Both sets of cells were harvested and analyzed in the same manner. However, the data once again showed pSTAT5 activation in the set of cells that were treated with 100 ng/mL of either protein for 1 hour as part of the time course assay, but no activation of pSTAT5 in the cells that were treated under the same condition in the concentration assay (**Fig. S1.3b**). The last attempt to study how different concentrations of mutant hGH might affect downstream signaling compared to WT hGH showed some difference at higher concentrations (500 ng/mL); however, the negative control falsely showed pSTAT5 activation (**Fig. S1.3c**), ultimately making the data gathered from this investigation inconclusive and unreliable. Thus, we cannot conclude that engineering chemical crosslinking onto hGH ligand has any different effect on downstream signaling of the STAT pathway through the activation of hGHR.

The last study we attempted during our investigation of the effects of crosslinking on the hGH-hGHR pair was to try and show chemical crosslinking between the two proteins on the cell surface. After incubation of the hGH-Q68FSY protein with serum starved Ba/F3 cells, we lysed the cells and attempted to pull down and purify out the crosslinked complex using His tag purification. However, despite trying different concentrations of hGH-Q68FSY, we were unable to see any crosslinked complexes after pull down (**Fig. S1.4**). This might be due to the relatively quick internalization and turnover of the hGHR upon ligand binding. Ultimately, though several

attempts were made, we could not definitively prove mutant hGH-hGHR crosslinking on cell surface.

1.4 Discussion

We were able to show that it was possible to achieve crosslinking between a native ligand and its receptor if a bioreactive Uaa was engineered into the binding site of the ligand-receptor interaction in close proximity to a nucleophilic residue. Our mutant hGH-Q68FSY was able to target the extracellular domain of hGHR and crosslink with the receptor in a matter of minutes during *in vitro* studies. We were also able to show that our mutant hGH-Q68FSY, despite having a mutation in the binding site, was still able to elicit prolonged downstream signaling, as seen via time course assay, through the STAT5 pathway in a manner similar to the WT hGH. We were unable to show any conclusive differences in signaling between hGH-WT and hGH-Q68FSY at different concentrations, however, nor were we able to purify out a crosslinked complex between hGH-Q68FSY and hGHR after treating cells with the mutant.

Many of the issues we faced in the study of the hGH-hGHR interaction might have to do with specific special properties of this receptor-ligand interaction. We know that simply increasing binding affinity of hGH does not in turn increase downstream signaling. Though we hoped to see some definitive difference in signaling between our mutant and WT, it is unsurprising that making the mutant covalent, in essence, increasing the binding affinity infinitely, did not have an effect on downstream signaling. Furthermore, the fast internalization and turnover of the receptor upon ligand binding might also explain why we were not able to show crosslinked complex on cell surfaces. It is also worth noting that though we did not see any *in vitro* crosslinking between hGH-Y103FSY and the hGHR extracellular domain, this might be due to site 2, where Y103 is located,

being unable to bind to the hGHR monomer. The hGHR extracellular domain protein we used in this initial study did not contain the necessary domain for the receptor to dimerize. If tested on full-length hGHR, the hGH-Y103FSY mutant has a potential to crosslink successfully with hGHR. Further study should focus on the second mutant, as improving binding at site 2 might have more biological effect than site 1.

Ultimately, the hGH-hGHR model system might not be best for further studying the effects of chemical crosslinking on PPIs and signal activation. There are other ligand-receptor pairs which might serve as better model systems and which can be used to study the questions posed in this proposal. The utility of bioreactive Uaas in the study of PPIs and biological signaling processes is still to be shown, but we were able to make promising progress with the results that we gathered.

1.5 Materials and Methods

Full length hGH amino acid sequence

FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPT
PSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEE
GIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFL
RIVQCRSVEGSCGF

Q68 and Y103 highlighted in red

Codon optimized hGH sequence

TTCCCGACCATTCCGCTGAGCCGTCTGTTTGATAATGCCATGCTGCGCGCCCATC
GCCTGCATCAGCTGGCTTTTGATACCTATCAGGAATTTGAAGAAGCATATATTCCG
AAAGAACAGAAATATAGTTTCCTGCAGAATCCGCAGACCAGCCTGTGTTTTAGTGA
AAGCATTCCGACCCCGAGCAATCGTGAAGAAACCCAGCAGAAAAGTAATCTGGAA
CTGCTGCGTATTAGCCTGCTGCTGATTCAGAGTTGGCTGGAACCGGTGCAGTTTCT
GCGTAGCGTGTTTGCCAATAGCCTGGTGTATGGTGCCAGTGATAGTAATGTTTATG
ATCTGCTGAAAGATCTGGAAGAAGGCATTCAGACCCTGATGGGCCGTCTGGAAGA
TGGCAGCCCGCGCACAGGTCAGATTTTTAAACAGACCTATAGTAAATTCGACACCA
ATAGCCATAATGATGATGCACTGCTGAAAAATTATGGCCTGCTGTATTGTTTTCGCA

AAGATATGGATAAAGTGGAAACCTTTCTGCGTATTGTTCAGTGCCGCAGTGTGGAA
GGCAGTTGTGGTTTT

Q68TAG primers

Forward: GAGCAATCGTGAAGAAACCtagCAGAAAAGTAATCTGGAAC

Reverse: GTTCCAGATTACTTTTCTGctaGGTTTCTTCACGATTGCTC

Y103TAG primers

Forward: GTAGCGTGTTTGCCAATAGCCTGGTGtagGGTGCCAGTGATAGTAATGTTTATG

Reverse: CATAAACATTACTATCACTGGCACCctaCACCAGGCTATTGGCAAACACGCTAC

Expression and purification of hGH proteins

Mutant and WT hGH proteins were cloned into pBad plasmid with arabinose promoter. 5 uL of approx. 1ng/uL pBad plasmid containing hGH protein was transformed into DH10b cells along with 5 uL approx. 1 ng/uL pEvol plasmid containing FSYRS/3xtRNA^{FSY}. After transformation, cells were spread onto +Amp/+Cam plates and incubated at 37C O/N. Colonies were picked and grown to OD600 in 100 mL 2XYT media +Amp/+Cam, shaking at 37C. Once culture reached OD, 1 mL of 100mM stock of FSY was added to culture containing mutant proteins. 1 mL 20% arabinose (by weight) was added to all cultures. Cultures were transferred to 18C shaker and grown O/N. Purification of proteins followed established protocol. In brief, the O/N cultures were spun down at 4000xg for 30 min. The pellets were re-suspended in wash buffer containing 1x PBS, 150 mM NaCL, 10% glycerol, 10 mM imidazole, and 0.1% Triton X-100. Lysozyme, protease inhibitor, and DNase were added to re-suspended pellets. Pellets were sonicated at 20% amplitude, 10 sec on, 10 sec off, for 5 min, then spun down for 30 min at 30000xg. The cleared supernatants were either purified using the AKTA FPLC with a Ni-NTA column or using Ni-NTA resin. To

purify using resin, 200 mL of beads per sample were equilibrated with wash buffer and incubated with cleared supernatant in 50mL tube at 4C for 1 hour while rotating. Samples were spun down at 500xg for 5 min, supernatant decanted, and resin washed with 3 mL wash buffer. Resin for each sample was transferred to small filtered column and washed an additional 3 times, 3x column volume. Samples were eluted with elution buffer consisting of 1x PBS, 150 mM NaCL, 10% glycerol, 500 mM imidazole, and 0.1% Triton X-100. Samples were buffer exchanged to storage buffer (50mM Tris-HCl, pH8.0, 150mM NaCl) and concentrated using spin columns.

***In vitro* crosslinking**

Purified WT hGH and hGH mutant proteins were diluted to 0.1 mg/mL. 5 uL of diluted protein was added to tubes with 5 uL 0.2 mg/mL hGHR extracellular domain. Samples were incubated, shaking, at 37C for either set time points, as in the time course assay (0min, 20min, 1h, 2h, or 4h), or O/N (approx.. 16h), as in initial crosslinking study. After incubation, samples were boiled and run on SDS-PAGE gel and transferred to Western blot to identify crosslinking.

Western blotting

For western blotting, all samples were mixed with 4x laemmli buffer and boiled at 95C for 5 minutes. Boiled samples were loaded onto 12% SDS-PAGE gels and run at 90V for 10-15 min, then at 170V for 40-50 min. Membrane for Western blot transfer was soaked in methanol. Transfer for western blot was run at 25V for 45 min. Membrane was covered in 5% blocking buffer for 1 hr, then with relevant antibody (anti-His antibody conjugated with HRP, beta actin antibody, pSTAT5 antibody, or STAT5 antibody) was added with fresh blocking buffer and incubated for another 1 hr. In necessary, a secondary antibody with HRP conjugated was added with fresh

blocking buffer for 1 hr after the initial antibody. The Western was washed 3x for 5 min with TBST. HRP substrate was added and Western blot was imaged for up to 3 min.

Time course cell assay protocol

Ba/F3 GHR+ cells, at confluency in 6 well plates, were serum starved the night before experiment – RPMI 1640 media (10% FBS) with containing IL-3 was removed and RPMI 1640 media (10% FBS) without IL-3 was added. WT hGH and hGH-Q68FSY protein were thawed and diluted to 100 ng/mL. The 100 ng/mL protein stocks were added to wells: hGH-Q68FSY was added to 4 wells and incubated for 20 min, 1 hr, 2 hr, or 4 hr at 37C before harvesting; hGH-WT was also added to 4 wells and incubated for the same time points. At each time point, cells were harvested, spun down, lysed with RIPA buffer containing protease inhibitor and a phosphatase inhibitor cocktail. Lysed cells were analyzed via Western blot as previously described for pSTAT5 activation.

Cell assay for testing different concentrations

As for the time course cell assay, Ba/F3 GHR+ cells were grown to confluency then serum starved O/N before the experiment in RPMI 1640 + 10% FBS. hGH-Q68FSY and hGH-WT protein stocks were thawed and diluted to various concentrations: 100 ng/mL, 10 ng/mL, or 1 ng/mL (or 500 ng/mL). Cells were incubated 1 hr, then harvested and lysed with RIPA buffer (+ protease inhibitor, + phosphatase inhibitor) and analyzed via Western blot with anti-pSTAT5 antibody.

Crosslink purification protocol

Ba/F3 GHR+ cells, serum starved O/N, were incubated with various concentrations (100 ng/mL or 500 ng/mL) of hGH-WT or hGH-Q68FSY for 20 minutes. After 20 minutes, cells were collected, spun down, and washed with cold PBS 3x. Cells were then lysed with 40 uL of 6M

guanidine-HCl solution and vortexed vigorously to disrupt membrane and allow for isolation of potential hGH-Q68FSY-hGHR crosslinked complex. Ni-NTA beads, equilibrated with wash buffer previously described, were added to the samples and incubated at rt for 1-2 hrs. Samples were washed 3x with wash buffer; 4x laemmli buffer was added to the samples and they were boiled to remove His-tagged complex from Ni-NTA beads. Boiled samples were briefly spun down, then analyzed via Western blot to attempt to identify crosslinked complex from cell surfaces.

1.6 Supplemental Figures

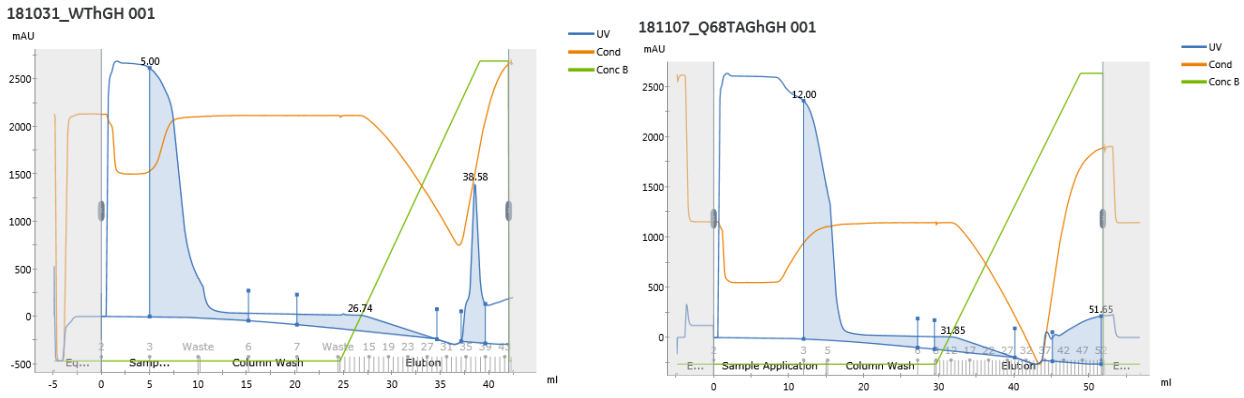


Fig S1.1: FPLC traces for the purification of hGH-WT and hGH-Q68FSY. hGH-WT is on the left and hGH-Q68FSY is on the right.

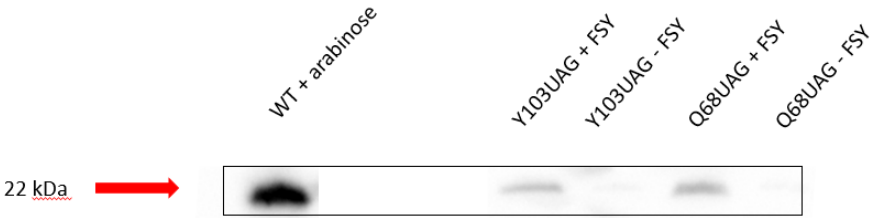


Fig S1.2: Expression of hGH-WT, hGH-Y103FSY, and hGH-Q68FSY proteins.

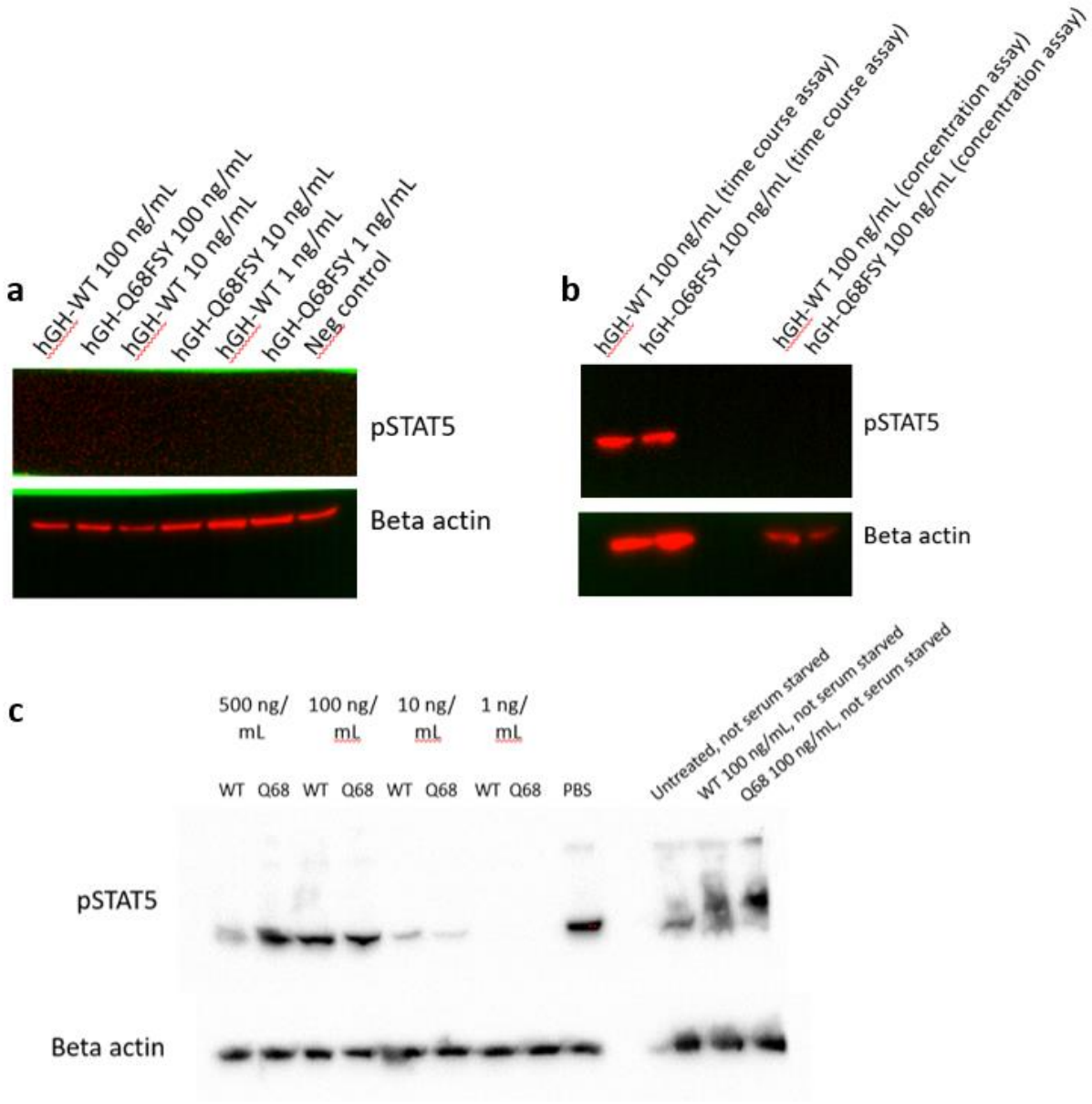


Fig S1.3: Testing different concentrations of hGH-WT and hGH-Q68FSY. a) Initial attempt to test the downstream signaling effects of different concentrations of hGH-WT and hGH-Q68FSY. No pSTAT5 signaling was detected for either WT or mutant protein at varying concentrations of 100 ng/mL, 10 ng/mL, or 1 ng/mL. **b)** Time course assay using 100 ng/mL of hGH-WT and hGH-Q68FSY was conducted in parallel with the concentration study. Despite observing downstream signaling when adding 100 ng/mL of ligand for 1 hour during the time course assay, no pSTAT5 signaling was detected for the 100 ng/mL samples from the concentration assay. **c)** A third attempt to elucidate potential differences in signaling between different concentrations of hGH-WT and hGH-Q68FSY shows that hGH-Q68FSY might signal better at higher concentrations (500 ng/mL) than hGH-WT, while hGH-WT might signal better at lower concentrations (10 ng/mL). These results are inconclusive due to a false positive in the negative control sample.

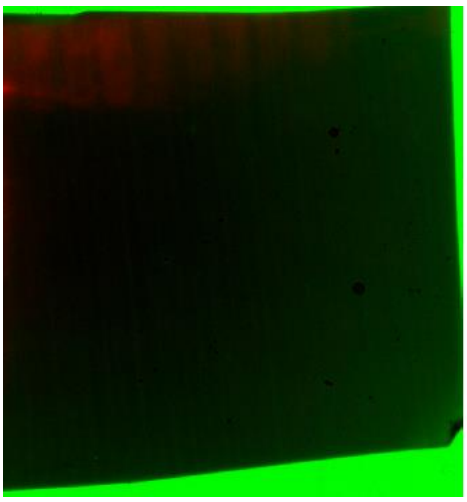


Fig S1.4: hGH-Q68FSY crosslinking on cell surface. Attempt at purifying crosslinked ligand-receptor after on cell incubation shows no detected crosslinking. This is potentially due to the quick internalization of hGH-receptor upon activation.

1.7 References

1) Wang, L. **Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts.** Acc. Chem. Res., 50 (2017), pp. 2767-2775, 10.1021/acs.accounts.7b00376

2) Wang, L., Brock, A., Herberich, B., Schultz, P.G. **Expanding the Genetic Code of Escherichia coli.** Science, 292 (2001), pp. 498-500, 10.1126/science.1060077

3) Nguyen, T.A., Cigler, M., Lang, K. **Expanding the Genetic Code to Study Protein-Protein Interactions.** Angew. Chem. Int. Ed., 57 (2018), pp. 14350-14361, 10.1002/anie.201805869

4) Berdan, V.Y., Klauser, P.C., Wang, L. **Covalent peptides and proteins for therapeutics**. Bioorg. & Med. Chem, 29 (2021), 10.1016/j.bmc.2020.115896

5) Bauer, RA. **Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies**. Drug Discov Toda, 20 (2015), pp. 1061–1073, 10.1016/j.drudis.2015.05.005

6) Baillie, TA. **Targeted Covalent Inhibitors for Drug Design**. Angew. Chem. Int. Ed., 55 (2016), pp. 13408–13421, 10.1002/anie.201601091

7) Wang, N., Yang, B., Fu, C., *et al*. **Genetically Encoding Fluorosulfate- l -tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo**. J. Am. Chem. Soc. 140 (2018), pp. 4995–4999, 10.1021/jacs.8b01087

8) Yang, B., Tang, S., Ma, C., *et al*. **Spontaneous and specific chemical crosslinking in live cells to capture and identify protein interactions**. Nat. Comm., 8 (2017), pp. 1-10, 10.1038/s41467-017-02409-z

9) Ranke, M.B., Wit, J.M. **Growth hormone – past, present, and future**. Nature Reviews Endocrinology, 14 (2018), pp. 1-16, 10.1038/nrendo.2018.22

10) Clackson, T., Ultsch, M.H., Wells, J.A., de Vos, A.M. **Structural and Functional Analysis of the 1:1 Growth Hormone:Receptor Complex Reveals the Molecular Basis for Receptor Affinity**. J. Mol. Biol., 277 (1998), pp. 1111-1128, 10.1006/jmbi.1998.1669

11) Waters, M.J., Brooks, A.J. **Growth Hormone and Cell Growth**. Endocr Dev., 23 (2012), pp. 86-95, 10.1159/000341761

12) Pearce, K.H., Cunningham, B.C., Fuh, G., Teeri, T., Wells, J.A. **Growth Hormone Binding Affinity for Its Receptor Surpasses the Requirements for Cellular Activity.** Biochemistry, 38 (1999), pp.81-89, 10.1021/bi9817008

13) Kopchick, J.J. **Discovery and mechanism of action of pegvisomant**. Eur. J. Endocrinol., 148 (2003), pp. S21-S25, 10.1530/eje.0.148s021

14) Subramani, R., Nandy, S.B., Pedroza, D.A., Lakshmanaswamy, R. **Role of Growth Hormone in Breast Cancer**. Endocrinology, 158 (2017), pp. 1543-1555, 10.1210/en.2016-1928

15) Kim, M.J., Park, H.S., Sea, K.H., Yang, H.J., Kim, S.K., Choi, J.H. **Complete Solubilization and Purification of Recombinant Human Growth Hormone Produced in Escherichia coli**. PLOS one, 8 (2013), pp. 1-8, 10.1371/journal.pone.0056168

16) Ortiz. B.L., Sanchez-Gomez, M., Norstedt, G. **Comparison of STAT5 mRNA levels in GH-treated male and female rats analysed by a solution hybridization assay.** Growth Horm IGF Res., 10 (2000), pp. 236-241, 10.1054/ghir.2000.0159

17) Silva, C.M., Lu, H., Day, R.N. **Characterization and cloning of STAT5 from IM-9 cells and its activation by growth hormone.** Mol. Endocrinol., 10 (1996), pp. 508-518, 10.1210/mend.10.5.8732682

18) Herrington, J., Smit. L.S., Schwartz, J., Carter-Su, C. **The role of STAT proteins in growth hormone signaling.** Oncogene, 19 (2000), pp. 2585-2597, 10.1038/sj.onc.1203526

19) Baixeras, E., Jeay, S., Kelly, P.A., Postel-Vinay, M.C. **The proliferative and antiapoptotic actions of growth hormone and insulin-like growth factor-1 are mediated through distinct signaling pathways in the Pro-B Ba/F3 cell line.** Endocrinology, 142 (2001), pp. 2968-2977, 10.1210/endo.142.7.8242

Chapter 2: Covalent peptides and proteins for therapeutics

2.1 Abstract

Drugs with a covalent mechanism of action benefit from enhanced potency, selectivity, and *in vivo* efficacy. Historically, the only covalent drugs on the market have been covalent small molecules. However, many proteins and protein-protein interactions cannot be targeted by small molecules due to their lack of small molecule binding pockets, and are thus deemed "undruggable." In order to drug the undruggable, peptide and protein therapeutics that can better bind to flat protein surfaces have been developed. Until recently, peptide and protein therapeutics have had noncovalent mechanisms of action. The recent advancement of unnatural amino acid chemistry, along with the development of better and more specific electrophilic warheads, has allowed for the application of covalent mechanisms to peptide and protein drugs. Covalent peptide and protein therapeutics have the potential to benefit from the same advantages that covalent small molecules have over their noncovalent counterparts. Here we provide a brief overview of the chemistry that makes this advancement possible, as well as examples of covalent peptides and the first covalent protein drug. These examples successfully crosslink their target proteins and have beneficial therapeutic effects.

2.2 Introduction

In recent years, there have been major efforts in the fields of drug discovery and chemical biology to develop covalent small molecule drugs.[1,2,3] These drugs provide a unique mechanism of action that gives them certain added advantages when compared to conventional noncovalent

binders. Covalent drugs possess a two-step binding mechanism: first, the drug binds to the target through reversible, noncovalent interactions; then, the drug reacts with a residue on the target and forms a covalent bond, resulting in the final covalent complex.[4] This additional step of covalent bond formation results in an increased residence time of the drug and a possible off-rate of zero, driving *in vivo* efficacy and more desirable pharmacodynamics.[5] Furthermore, compared to their noncovalent counterparts, covalent drugs are typically more potent, have higher selectivity, and can potentially avoid certain resistance mechanisms.[6,7]

Traditionally, only small molecules have been developed into covalent drugs due to their potential for infinite chemical diversity.[8,9] Strategies to covalently target a protein with a small molecule are only limited by the reactivity of the protein, not the small molecule. Most covalent small molecule drugs target cysteine or lysine residues that are buried in protein pockets or active sites.[10,11] However, small molecule drugs are generally not suited for binding to smooth and flat protein surfaces or for inhibiting protein–protein interactions (PPIs), which account for the majority of biological activity and interactions in cells. Protein-protein interactions occur across interfaces that can span hundreds of angstroms in area. Small molecule drugs are more apt for binding specifically in distinct pockets than to these vast surfaces.[9,12] In order to address this gap in the field, many have started to develop peptide and protein drugs instead. Both peptide and protein drugs are able to bind specifically to large surface areas.[13,14,15] Antibodies are perfect examples of clinically effective protein drugs,[16] and recent years have seen a gradual shift towards the development of more peptide and peptidomimetic drugs as well.[14,17] While progress continues to be made in the fields of protein and peptide drug development, both fields could benefit from the potential advantages of developing covalent peptide and protein therapeutics. Peptide and

protein therapeutics that can bind covalently to their target of interest could have improved pharmacodynamic properties that make them desirable drug candidates (Fig. 2.1)

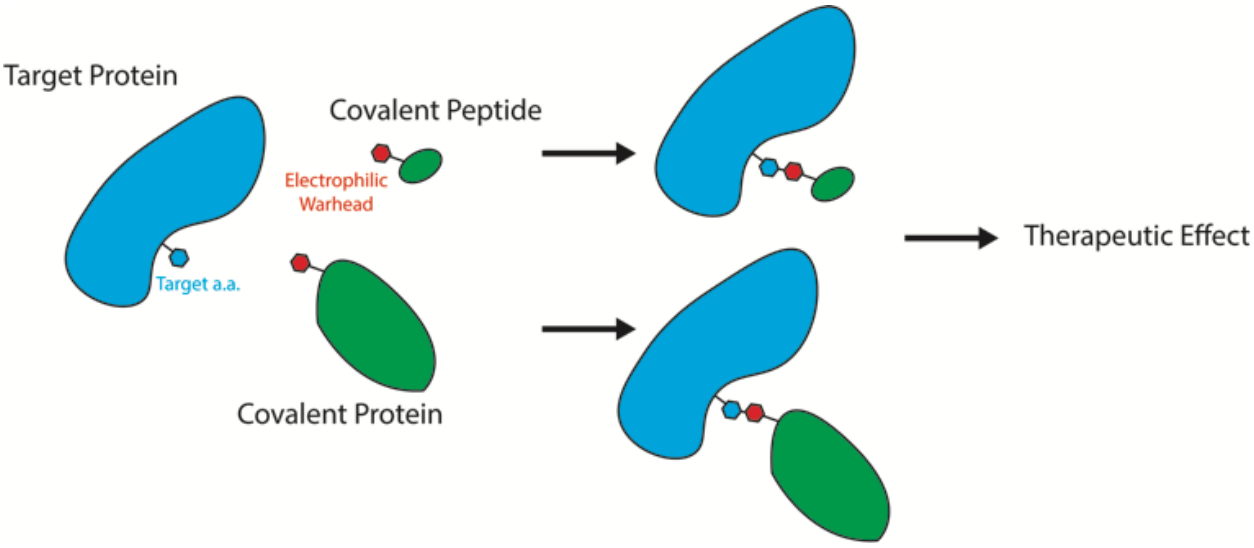


Figure 2.1: Covalent crosslinking between the target protein and peptide or protein drug. By installing electrophilic warheads that target specific amino acids onto peptides and proteins, covalent crosslinking between the target protein and peptide or protein drug can be achieved. A covalent mechanism of action has the potential to improve the therapeutic effect of the peptide or protein drug.

Unfortunately, unlike small molecule drugs, peptide and protein drugs are significantly limited by the chemical reactivity of canonical amino acids. Of the twenty canonical amino acids, only one, cysteine, can form a covalent bond with another amino acid – itself – without catalytic assistance from an enzyme. However, disulfide bonds are weak, reversible, and redox sensitive, making cysteine-cysteine crosslinking unsuitable for use in developing covalent peptide and protein drugs.[18] While other nucleophilic amino acid residues, such as lysine and tyrosine, exist, there are no canonical electrophilic amino acids; this fundamental issue has prevented advancement in the development of covalent peptide and protein drugs.

This lack of chemical diversity, however, can be solved by the introduction of unnatural amino acids (Uaas) into peptides and proteins through either solid phase peptide synthesis,[19,20]

post-synthesis chemical modification,[21] or genetic code expansion methods.[22,23,24] Many Uaas have been developed for a wide variety of purposes, including photoswitchable,[25] photocrosslinkable,[26] fluorescent,[27] and bioorthogonal Uaas.[28,29] However, only in recent years have latent bioreactive Uaas been developed to covalently target natural amino acids.[30] These latent bioreactive Uaas do not react with free amino acids or other biomolecules inside cells; rather, they only react with specific natural amino acid residues when they are in close proximity and in the correct orientation in regards to each other, e.g., during a protein–protein interaction. Unlike photocrosslinkable Uaas, they do not require UV light for activation, making them ideal for crosslinking in *in vivo* systems. Further innovation of these latent bioreactive Uaas allows for the development of covalent peptides and proteins.

Herein, we briefly review the current development of covalent peptides and proteins as therapeutics, including developing new latent bioreactive Uaas to react with protein targets.

2.3 Advances in Covalent Peptide Drugs

Targeting protein–protein interactions, which can span hundreds of angstroms in area, with small molecules has been a difficult task.[12] The development of peptide drugs, which can better target, bind, and block these PPI interfaces, has allowed us to go after proteins and interactions that were previously deemed "undruggable".[31,32] Peptide drugs are often derived from mimicking – and optimizing – one half of a protein–protein interaction in order to inhibit the two proteins from binding, though some peptide drugs have also been designed *de novo.*[32,33,34,35,36,37] Unlike protein drugs such as antibodies, certain peptides can cross the cell membrane, allowing for the targeting of intracellular proteins as well as membrane-bound ones.[38] Peptide and peptidomimetic drugs have real potential for being an effective and important class of drugs.

The issue with peptides is that mimicking a small portion of a PPI interface leads to a significant decrease in binding affinity, which can be a major problem as many proteins already naturally have weak affinities for their interaction partners. Peptide drugs may also have a relatively short half-life because of instability, degradation by proteases, or quick clearance due to their small size.[39] Given that small molecule drugs are able to overcome similar problems via introduction of a covalent mechanism of action, making peptide drugs covalent is an attractive strategy to design and develop better and more effective peptide drugs.

Early attempts to develop covalent peptide drugs relied on a photoreactive moiety,[19,26] which is inapt for *in vivo* applications. Incorporating warheads utilized in covalent small molecule drugs into peptides is a simple and straight-forward strategy to develop peptide inhibitors that can covalently bind with native amino acid residues such as cysteine. In one study, after identifying a 13-mer peptide (BI-107D1) that targets the E3 ubiquitin ligase Siah – which is key in regulating cellular events that lead to cancer development and progression – Stebbins *et al.* converted BI-107D1 into a covalent peptide by synthesizing a Lys-acrylamide (Fig. 2.2 **compound 1**) warhead onto a residue in the peptide.[40] The Lys-acrylamide crosslinked an endogenous Cys residue next to the binding pocket of the peptide. Cellular studies showed that the covalent peptide more efficiently inhibited Siah than the noncovalent peptide. A later study installed multiple electrophilic warheads into the native peptide ligand BIM to target a Cys residue of the oncogenic protein Bcl2A1.[21] When acrylamide was used, the modified BIM peptide bound covalently and irreversibly to one Cys residue within the helix-binding groove of Bcl2A1, but not to two other exposed Cys residues on its surface. When the more reactive chloroacetamide or propiolamide warheads were used, nonspecific reactions with the two other surface Cys residues of Blc2A1 were observed. These results suggest the importance of achieving reaction specificity and predictability.

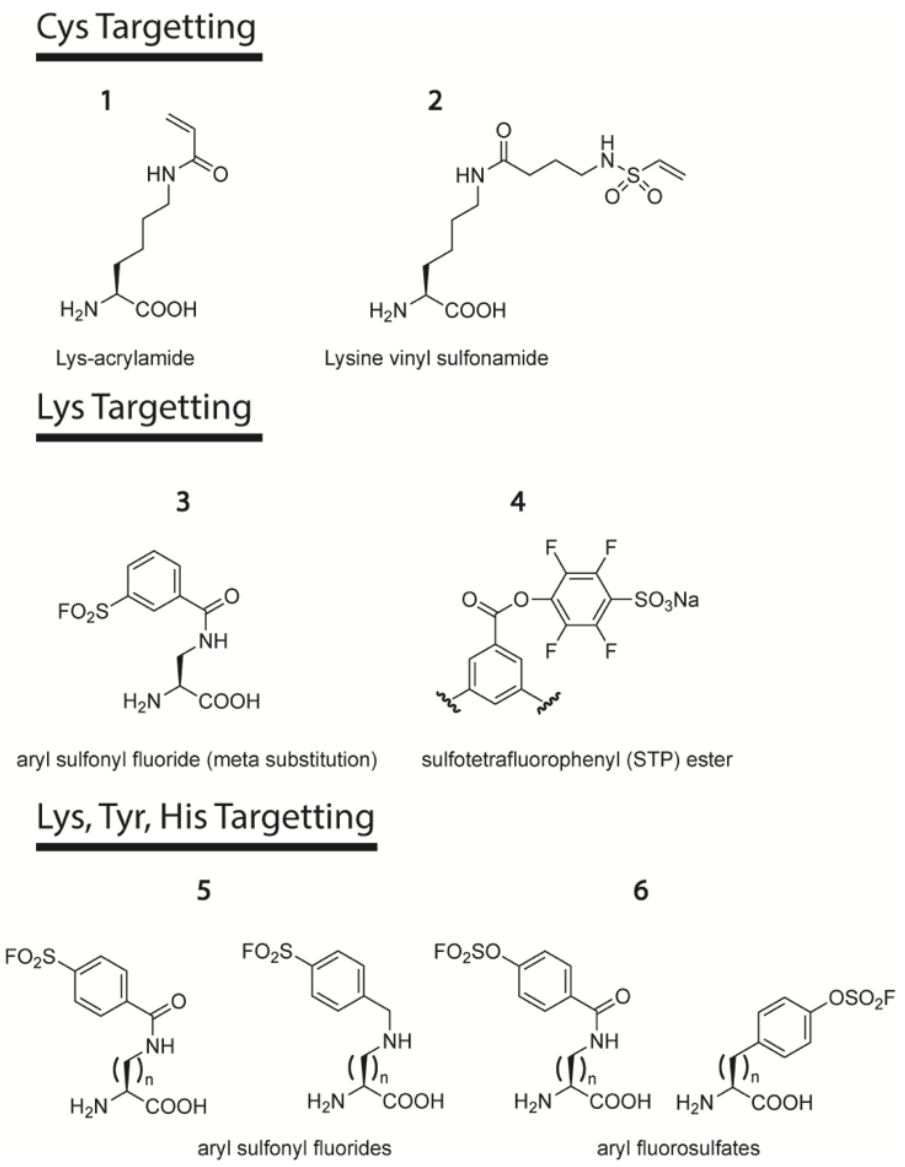


Fig. 2.2: Electrophilic warheads and amino acids utilized in covalent peptide design. Used to target nucleophilic amino acids.

While unpaired Cys residues can be found in active sites or binding pockets, they are rare at protein–protein interaction interfaces. The ability to covalently target other residues such as Lys, Tyr, and His, which are often present on protein surfaces and PPI interfaces, as well as a general mechanism to ensure reaction specificity, are highly desired. The introduction of proximity-enabled bioreactive Uaas targeting various natural residues has effectively advanced the field.[30] In

a 2016 paper,[41] Hoppmann *et al.* initially put forward the concept of proximity-enabled bioreactivity to generate covalent peptide binders: an otherwise chemically stable Uaa – the latent bioreactive Uaa – is incorporated into a peptide; this Uaa then reacts with the target natural residue of the protein only when the peptide binds with the protein and brings the Uaa into proximity of the target natural residue (Fig. 2.3). The latent reactivity of the bioreactive Uaa prevents nonspecific reactions and any potential cytotoxicity from over-reactivity; the proximity-enabled reactivity of the Uaa offers high target specificity. To prove the concept, they designed an aryl sulfonyl fluoride-containing Uaa (Fig. 2.2 **compound 3**) that could target either a nucleophilic Lys or His residue and introduced the Uaa into the SAH_{p53-8} peptide that inhibits p53-Mdm2/4 interactions. Sulfonyl fluoride participates in the Sulfur(VI) Fluoride Exchange (SuFEx) reaction, a new generation of click chemistry pioneered by Sharpless.[42,43] As designed, the resultant mSF-SAH peptide efficiently crosslinked with MDM2 and MDM4 under mild conditions. The covalent binding increased the inhibition activity against MDM4 by > 10-fold. Interestingly, covalent binding of the SAH peptide to MDM2/4 occurred when sulfonyl fluoride was installed on the *meta*-, but not the *para*-position of the bioreactive Uaa, indicating that the covalent reaction is dependent on side chain orientation, which can afford high specificity. In addition, unlike the noncovalent peptide SAH_{p53-8}, which binds to MDM2 and MDM4 with similar affinity, the mSF-SAH peptide was more selective for MDM4 by a 6-fold difference. Selectivity for MDM4 over MDM2 has been difficult to achieve in the field of MDM2/4 inhibition.[44] This work establishes the use of latent bioreactive Uaas to covalently target noncatalytic, non-cysteine residues via a defined mechanism of proximity-enabled bioreactivity, and also demonstrates the potential benefits that covalence might have for peptide drug efficacy and specificity.

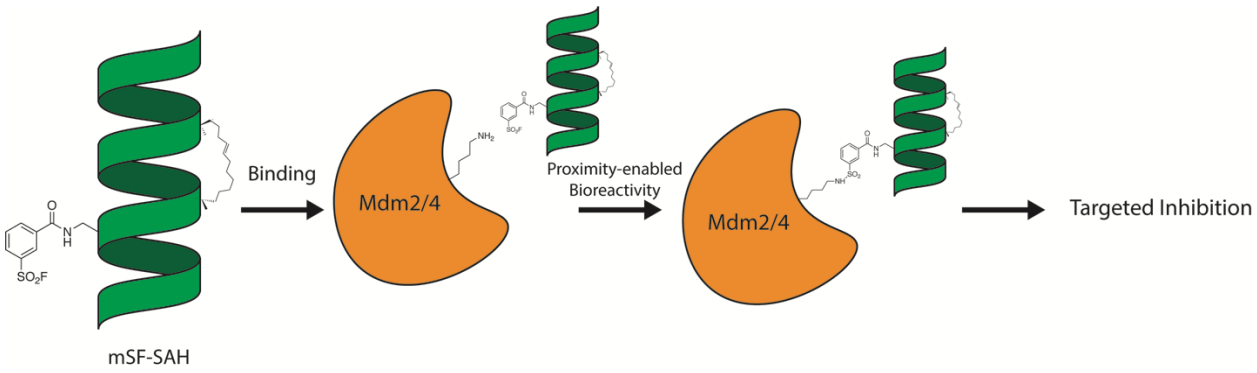


Fig. 2.3: mSF-SAH crosslinks with MDM2/4. mSF-SAH is a stapled peptide with an electrophilic aryl sulfonyl fluoride warhead, which can covalently react with nucleophilic residues, such as Lys and His, placed in proximity. The binding interface of MDM2/4 has both a Lys and His residue that can react with the warhead. Incubation of the covalent peptide with MDM2/4 (both purified and in cell lysate) showed efficient binding and crosslinking of the peptide with the protein(s), leading to target inhibition.

Inspired by the work above,[41] Charoenpattarapreeda *et al*. took on a slightly different approach towards the synthesis of covalent stapled peptides.[45,46] Instead of directly incorporating the warhead onto a residue in the peptide, the warhead was placed into a staple core, which bridges two residues of the peptide to form the staple. This "two-component" peptide stapling eases the synthesis of stapled peptides containing electrophiles via a "one-size fits all" approach. They used an activated ester (Fig. 2.2 **compound 4**) installed on the staple of peptide *P*1 to target a Lys residue of MDM2. The resultant peptide covalently bound to MDM2 and showed an improved dissociation constant with time when compared with the noncovalent peptide.

The Pellecchia group has extensively studied the enthalpy and entropy of binding of peptide ligands to members of the inhibitors of apoptosis proteins (IAPs) family of proteins, and has been able to identify pan-specific noncovalent peptide inhibitors, as well as some more selective peptide inhibitors.[47] By incorporating a sulfonyl fluoride warhead (Fig. 2.2 **compound 5**) into one of these inhibitors, they were able to specifically target a Lys residue on the X-linked

IAP (XIAP) protein. This Lys residue is a Glu residue in cellular IAP1 and IAP2, two other members of the IAP family. When compared to the lead noncovalent peptides, the covalent peptides selectively targeted XIAP, showed enhanced cell killing of various cancer cell lines, and were also effective in cell lines insensitive to the lead noncovalent peptide. As the SuFEx applications progress, aryl sulfonyl fluoride and aryl fluorosulfate are increasingly appreciated as better warheads for targeting other nucleophilic residues including Lys, Tyr, His, Thr, and Ser in a protein context.[43,48,49,50] In a follow up study,[51] the Pellecchia group paired different warheads (Fig. 2.2 **compounds 5**–**8**) against a panel of XIAP proteins with different nucleophilic amino acid mutations at the target Lys site. They showed that both benzyl sulfonyl fluoride and benzyl fluorosulfate can covalently target Lys, Tyr, and His residues, which is consistent with the earlier reported reactivity of benzyl fluorosulfate genetically incorporated in proteins.[50] They further altered the short tetramer peptide to be more "drug-like" by methylating the *N*-terminal alanine residue and by replacing the C-terminal phenylalanine residue with different amino-indanes, which improved peptide stability in serum and also cell permeability. While this work showed modest therapeutic potential, a wider screen of cell lines is underway to better ascertain the therapeutic potential of these covalent inhibitors.

While stapled peptides have long been the standard for synthesizing short peptides with proper secondary structure,[38,52] alternative methods, such as the hydrogen bond surrogate (HBS) approach, are available.[53] The advantage of using the HBS approach, where the *N*-terminal backbone hydrogen bond is replaced with an isosteric, covalent carbon–carbon bond to initiate nucleation of the peptide helix, is that, unlike stapled peptides where one face of the helix is the large, often hydrophobic staple, all side chain residues in the HBS peptide are open and able to make proper contacts for molecular recognition.[53] Yoo *et al*. combined the HBS platform for

peptide design and synthesis with covalent peptide technology.[54] This paper targeted the G12C mutation of Ras, one of the first enzymes to be successfully targeted with a covalent small molecule drug, via a Sos-mimic peptide.[55,56] Yoo *et al*. conjugated various electrophilic warheads to the HBS Soc-mimic to create covalent peptide inhibitors. A vinyl sulfonamide (Fig. 2.2 **compound 2**) was shown to have the best efficacy. When compared against control HBS peptides, the HBS peptide containing compound 2 showed promising therapeutic potential in cell viability assays and an ERK downstream signalling assay. This work showcases the potential of covalent peptides on a different scaffold, as well as their ability to target the catalytic sites of enzymes.

2.4 Advances in Covalent Protein Drugs

The development of covalent protein therapeutics can, in theory, provide the same advantages as covalent small molecule and peptide drugs. These potential advantages include 'infinitely high' affinity, better selectivity, and enhanced pharmacodynamic properties.[4] Antibodies can form covalent bonds with small molecule antigens when the latter contain or are modified with warheads to react with natural residues of the antibody,[57,58] yet this approach cannot enable covalent binding of protein antigens, because target proteins usually cannot be modified *in vivo* in therapeutics. In a reverse manner, Holm *et al*. were able to conjugate a weakly electrophilic acrylamide onto an affibody,[59] and showed that the affibody forms a covalent complex with its substrate protein ZSPA through the reaction of the attached acrylamide with either a Cys, His, or Lys residue on ZSPA. The covalent affibody has good specificity towards the target protein both *in vitro* and on mammalian cell surfaces expressing the target protein, improving the detection sensitivity of ELISA and stability in cell surface imaging. Unfortunately, the covalent bond formation with Lys and His has a slow rate and a low yield, and introduction of the electrophilic

warhead via Cys conjugation is not generally applicable to other proteins. Moreover, the specificity and performance of the covalent affibody *in vivo*, which are critical to drug development, have not been evaluated.

Unlike peptides, in which various chemical functionalities can be readily installed through chemical synthesis or post-synthesis modification, the introduction of bioreactive functionalities into proteins imposes a huge challenge. A general site-specific method for incorporating Uaas into proteins in live systems is genetic code expansion, which uses an orthogonal pair of tRNA and aminoacyl-tRNA synthetase (aaRS) to translationally incorporate the desired Uaa into proteins.[22,23,24] To be compatible with live cells or systems, Uaas genetically incorporated in the past have contained functional groups that are either chemically inert, bio-orthogonal, or caged, that is, they are non-bioreactive.[60,61,62] Genetically encoding a bioreactive Uaa itself is a conundrum:[30] to form a new covalent bond between the side chains of a Uaa and a natural amino acid, the Uaa must be reactive toward the target natural amino acid. However, the ubiquitous presence of the natural amino acid in proteins and inside cells could result in nonspecific linkages, which may trap the Uaa during translation and/or cause cytotoxicity.

This impasse of bioreactivity and selectivity has been overcome elegantly through the concept of proximity-enabled bioreactivity initially proposed and demonstrated by Xiang *et al*.[63] Specifically, the reactivity of the Uaa is fine-tuned so that it does not react with free natural amino acids and other biomolecules under physiological conditions, thereby permitting genetic incorporation *in vivo* via endogenous translational machinery. When the Uaa comes into proximity to its target natural residue in proteins while in the appropriate orientation, the increased local effective concentration then facilitates the Uaa to selectively react with the side chain of the target

natural residue to create a covalent bond. Xiang *et al.* designed and incorporated the first bioreactive Uaa, *p*-2′-fluoroacetyl-phenylalanine (F_{fact}), to target Cys.[63] When introduced at the binding interface of an affibody and its substrate protein, F_{fact} specifically reacts with a proximal Cys to generate a covalent complex of the two proteins. F_{fact} has also been incorporated into fluorescent proteins to generate covalent bonds with Cys intramolecularly in live cells, improving the photon output and photostability of the proteins, and into a GPCR expressed on live mammalian cell surfaces in order to pinpoint receptor-peptide ligand interactions.

The work of Xiang *et al*. opens the gate to the incorporation of bioreactive Uaas into proteins via genetic code expansion. A range of bioreactive Uaas have subsequently been incorporated to covalently target different natural residues, such as Uaas containing haloalkanes (haloalkane Uaa, EB3), perfluoro-benzene (F-PSCaa), or fluoroacetamide (FAcK) to target Cys,[25,64,65,66,67,68] vinyl sulphonamide (VSF), acrylamide (AcrK, AcrF), or phenyl isothiocyanate (pNCSF) to target Lys,[69,70] alkyl bromide (BrC6K) to target Lys/Cys/His,[66] and aryl carbamate (FPheK) to target Lys/Cys/Tyr (Fig. 2.4)[71]. These bioreactive Uaas have largely been used to study protein–protein interactions *in vitro* and in cells.

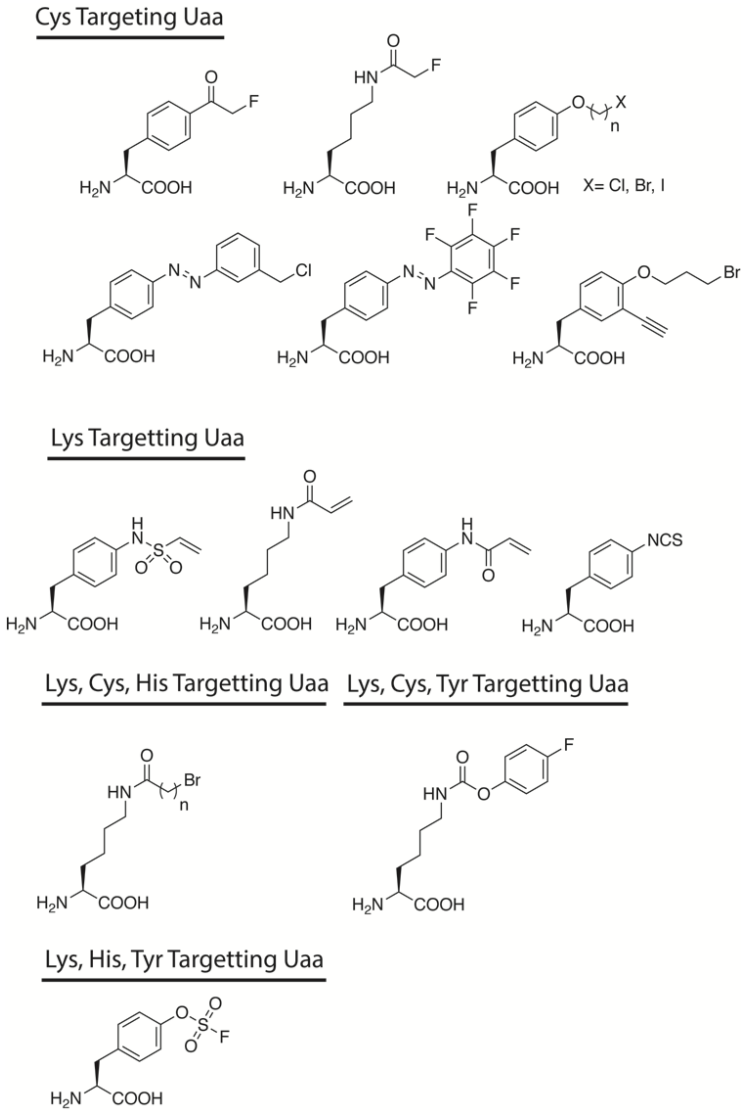


Fig 2.4: Bioreactive Uaas. These Uaas have been successfully incorporated into proteins via genetic code expansion and used to target specific nucleophilic natural residues.

To develop covalent protein drugs, the bioreactive Uaa needs to meet more stringent requirements: these Uaas must be chemically inert *in vivo*; they must efficiently react with the target residue before target dissociation; and they must have no cytotoxicity and no cytotoxic reaction products or metabolic byproducts. A milestone toward this goal was the design and genetic incorporation of the latent bioreactive Uaa flourosulfate-l-tyrosine (FSY) by Wang *et al.*[50] Aryl fluorosulfonate installed on small molecules has been shown to be essentially unreactive

toward the proteome.[48] Wang *et al.* showed that FSY is nontoxic to mammalian cells, remains inert in proteins inside cells, and is able to react with residues Tyr, His, or Lys in proximity efficiently via SuFEx, generating stable linkages resistant to hydrolysis. In addition, the byproduct of the reaction, the fluoride ion, is found in fluoridated salt, milk, and toothpaste, and is excreted through the kidneys, which means it should be nontoxic *in vivo*. Interestingly, the reaction of FSY with Ser and Thr in proximity results in the conversion of the latter into dehyroalanine and dehydrobutyrine, respectively.[72]

Recently, Li *et al*. demonstrated the first successful example of a covalent protein drug by incorporating FSY into the human programmed cell death protein-1 (PD-1) through a general platform technology termed PERx (Proximity-Enabled Reactive Therapeutics).[73] The latent bioreactive Uaa FSY was genetically incorporated into protein PD-1, and, upon PD-1 binding to PD-L1, the Uaa reacted with a natural His residue of the target PD-L1 via proximity-enabled reactivity. This highly specific reactivity resulted in irreversible binding of the PD-1(FSY) drug to the target PD-L1 selectively *in vitro*, on cancer cell surface, and in tumors *in vivo*. The PD-1/PD-L1 interaction inhibited *T*-lymphocyte proliferation and activity, resulting in exhaustion and apoptosis of tumor-specific T cells. As PD-L1 is often overexpressed in different tumors, Li *et al.* used PD-1(FSY) to block the PD-1/PD-L1 interaction as cancer immunotherapy. Compared with the noncovalent PD-1(WT), the covalent PD-1(FSY) significantly enhanced the functional activities of human T cells and CAR-T cells *in vitro*. When administered in tumor xenograft mouse models immune-humanized with human peripheral blood mononuclear cells (PBMCs), human CAR-T cells, or with the advanced systematic bone marrow-liver-thymus humanization, in all cases the covalent protein drug PD-1(FSY) showed an antitumor effect that is markedly more potent than PD-1(WT), with a therapeutic effect equivalent to or better than the anti-PD-L1

monoclonal antibody atezolizumab (Fig. 2.5). In addition, the covalent binding of PD-1(FSY) to PD-L1 showcased that covalent reactivity in PERx necessitates both drug-target binding and Uaa-natural residue pairing, thus uniquely affording unusual specificity and target selectivity. Moreover, irreversible binding through PERx also enabled the direct use of small proteins *in vivo* as protein drugs without worrying about their half-lives. This seminal paper on the benefits of covalent protein therapeutics and the use of latent bioreactive Uaas for their development opens the field of drug discovery up to an innovative and potentially advantageous approach to protein drug development that has not been accessible before.

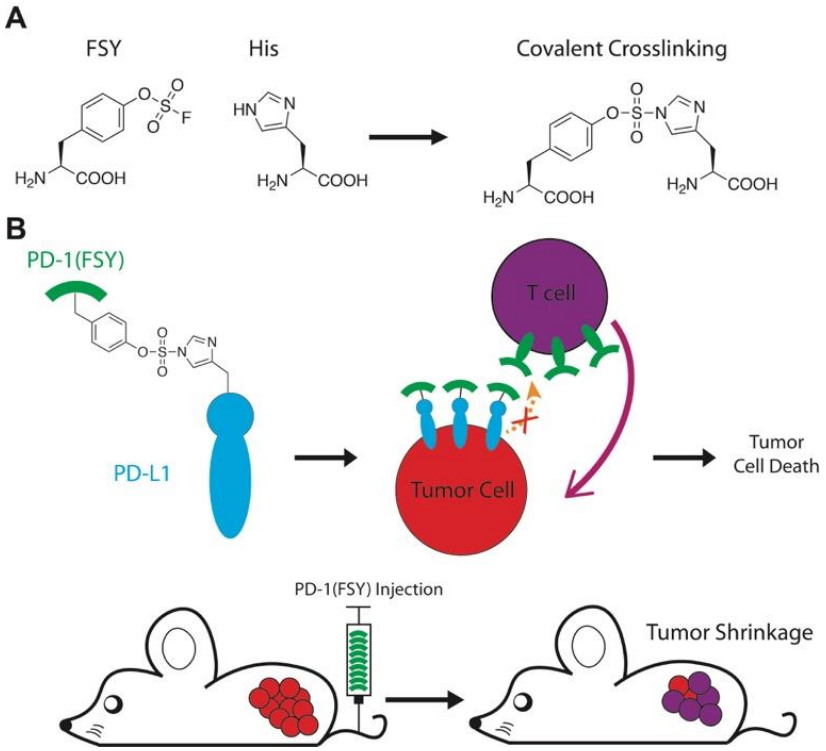


Fig 2.5: Covalent PD-1 drug. A) The latent bioreactive Uaa FSY reacts with His in proximity via SuFEx. **B**) Upon binding, PD-1(FSY) crosslinks with PD-L1 covalently and specifically on tumor cells *in vivo*. This irreversible binding efficiently blocks the native interaction between PD-1 (on T cells) and PD-L1 (on tumor cells), reversing the dampened antitumor immune response to kill tumor cells.

2.5 Conclusions and perspectives

Developing covalent peptide and protein drugs has the potential to improve on existing therapeutics and lead to a new generation of biological therapeutics working in the covalent mode. Here we focus on the Uaa chemistry that has been developed to allow for the design and generation of covalent peptide and protein drugs, and also look into several successful examples of covalent peptide and protein drug development. The field of covalent peptide and protein drugs is currently in its infancy and more work needs to be done to investigate the full therapeutic impact of these new drug classes. Uaa incorporation via stop codon suppression often leads to lower protein yields when compared to the wild-type protein. Release factor-knockout bacterial strains and Uaa incorporation via unnatural base pairs are mitigating this issue.[74,75] Furthermore, there are still many questions to be asked about the impact of covalent binding on the target biology and drug efficacy. While covalent peptide and protein drugs might show significant improvement over noncovalent, weak-binding peptide and protein drugs, the difference between high affinity binders and covalent binders might be more subtle. Investigation into how a covalent moiety affects the efficacy of a tight-binding antibody protein might, for example, give us insight into this question. In addition, more investigation into the *in vivo* efficacy of these drugs is needed, as many studies so far have not extended past cellular assays. Exploring the efficacy of covalent peptide and protein drugs in mouse models and studies in other higher order organisms, including immunogenicity tests, will speak better to the potential of these drugs as viable therapeutic options for the treatment of cancers and other diseases.

2.6 References

1) R.A. Bauer. **Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies**. Drug Discov. Today, 20 (2015), pp. 1061-1073, 10.1016/j.drudis.2015.05.005

2) D.S. Johnson, E. Weerapana, B.F. Cravatt. **Strategies for discovering and derisking covalent, irreversible enzyme inhibitors**. Future Med. Chem., 2 (2010), pp. 949-964, 10.4155/fmc.10.21

3) T.A. Baillie. **Targeted Covalent Inhibitors for Drug Design**. Angew. Chem. Int. Ed., 55 (2016), pp. 13408-13421, 10.1002/anie.201601091

4) J. Singh, J., Petter, T.A. Baillie, *et al.* **The resurgence of covalent drugs**. Nat. Rev. Drug Discov., 10 (2011), pp. 307-317, 10.1038/nrd3410

5) R.A. Copeland. **The drug–target residence time model: a 10-year retrospective**. Nat. Rev. Drug Discov., 15 (2016), pp. 87-95, 10.1038/nrd.2015.18

6) A.J.T. Smith, X. Zhang, A.G. Leach, *et al.* **Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins**. J. Med. Chem., 52 (2009), pp. 225-233, 10.1021/jm800498e

7) M. Hagel, D. Niu, T.S. Martin, *et al.* **Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine**. Nat. Chem. Biol., 7 (2011), pp. 22-24, 10.1038/nchembio.492

8) A. Cuesta, J. Taunton. **Lysine-Targeted Inhibitors and Chemoproteomic Probes**. Annu. Rev. Biochem., 88 (2016), pp. 1-17, 10.1146/annurev-biochem-061516-044805

9) M. Gehringer, S.A. Laufer. **Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology**. J. Med. Chem., 62 (2018), pp. 5673-5724, 10.1021/acs.jmedchem.8b01153

10) R.A. Laskowski, N.M. Luscombe, M.B. Swindells, *et al.* **Protein clefts in molecular recognition and function**. Protein Sci. Publ. Protein Soc., 5 (1996), pp. 2438-2452

11) M.H. Potashman, M.E. Duggan. **Covalent Modifiers: An Orthogonal Approach to Drug Design**. J. Med. Chem., 52 (2009), pp. 1231-1246, 10.1021/jm8008597

12) S. Jones, J.M. Thornton. **Principles of protein-protein interactions**. Proc. National Acad. Sci., 93 (1996), pp. 13-20, 10.1073/pnas.93.1.13

13) L.-G. Milroy, T.N. Grossmann, S. Hennig, *et al.* **Modulators of Protein-Protein Interactions**. Chem. Rev., 114 (2014), pp. 4695-4748, 10.1021/cr400698c

14) D.J. Craik, D.P. Fairlie, S. Liras, *et al.* **The Future of Peptide-based Drugs**. Chem. Biol. Drug Des., 81 (2013), pp. 136-147, 10.1111/cbdd.12055

15) B. Leader, Q.J. Baca, D.E. Golan. **Protein therapeutics: a summary and pharmacological classification**. Nat. Rev. Drug Discov., 7 (2008), pp. 21-39, 10.1038/nrd2399

16) K.M. Hargadon, C.E. Johnson, C.J. Williams. **Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors**. Int. Immunopharmacol., 62 (2018), pp. 29-39, 10.1016/j.intimp.2018.06.001

17) A.A. Vinogradov, Y. Yin, H. Suga. **Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges**. J. Am. Chem. Soc., 141 (2019), pp. 4167-4181, 10.1021/jacs.8b13178

18) C.S. Sevier, C.A. Kaiser. **Formation and transfer of disulphide bonds in living cells**. Nat. Rev. Mol. Cell Bio., 3 (2002), pp. 836-847, 10.1038/nrm954

19) C.R. Braun, J. Mintseris, E. Gavathiotis, *et al.* **Photoreactive Stapled BH3 Peptides to Dissect the BCL-2 Family Interactome**. Chem. Biol., 17 (2010), pp. 1325-1333, 10.1016/j.chembiol.2010.09.015

20) W. Chen, J. Dong, S. Li, *et al.* **Synthesis of Sulfotyrosine-Containing Peptides by Incorporating Fluorosulfated Tyrosine Using an Fmoc-Based Solid-Phase Strategy**. Angew. Chem. Int. Ed., 128 (2015), pp. 1867-1870,

21) A.D. de Araujo, J. Lim, A.C. Good, *et al.* **Electrophilic Helical Peptides That Bond Covalently, Irreversibly, and Selectively in a Protein-Protein Interaction Site**. ACS Med. Chem. Lett., 8 (2016), pp. 22-26, 10.1021/acsmedchemlett.6b00395

22) L. Wang, A. Brock, B. Herberich, P.G. Schultz. **Expanding the Genetic Code of *Escherichia coli***. Science, 292 (2001), pp. 498-500, 10.1126/science.1060077

23) L. Wang, P.G. Schultz. **Expanding the Genetic Code**. Angew. Chem. Int. Ed., 44 (2004), pp. 34-66, 10.1002/anie.200460627

24) L. Wang. **Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts**. Accounts Chem. Res., 50 (2017), pp. 2767-2775, 10.1021/acs.accounts.7b00376

25) C. Hoppmann, V.K. Lacey, G.V. Louie, *et al.* **Genetically Encoding Photoswitchable Click Amino Acids in Escherichia coli and Mammalian Cells**. Angew. Chem. Int. Ed., 53 (2014), pp. 3932-3936, 10.1002/anie.201400001

26) I. Coin, M.H. Perrin, W.W. Vale, *et al.* **Photo-Cross-Linkers Incorporated into G-Protein-Coupled Receptors in Mammalian Cells: A Ligand Comparison**. Angew. Chem. Int. Ed., 123 (2011), pp. 8227-8231, 10.1002/ange.201102646

27) H.S. Lee, J. Guo, E.A. Lemke, *et al.* **Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces cerevisiae**. J. Am. Chem. Soc., 131 (2009), pp. 12921-12923, 10.1021/ja904896s

28) L. Wang, Z. Zhang, A. Brock, *et al.* **Addition of the keto functional group to the genetic code of Escherichia coli**. Proc. National Acad. Sci., 100 (2003), pp. 56-61, 10.1073/pnas.0234824100

29) A. Deiters, P.G. Schultz. **In vivo incorporation of an alkyne into proteins in Escherichia coli**. Bioorg. Med. Chem. Lett., 15 (2005), pp. 1521-1524, 10.1016/j.bmcl.2004.12.065

30) L. Wang. **Genetically encoding new bioreactivity**. N. Biotechnol., 38 (2017), pp. 16-25, 10.1016/j.nbt.2016.10.003

31) L.N. Makley, J.E. Gestwicki. **Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein-Protein Interactions**. Chem. Biol. Drug Des., 81 (2013), pp. 22-32, 10.1111/cbdd.12066

32) M. Raj, B.N. Bullock, P.S. Arora. **Plucking the high hanging fruit: A systematic approach for targeting protein–protein interactions**, Bioorgan. Med. Chem., 21 (2013), pp. 4051-4057, 10.1016/j.bmc.2012.11.023

33) J. Garner, M.M. Harding. **Design and synthesis of α-helical peptides and mimetics**. Org. Biomol. Chem., 5 (2007), pp. 3577-3585, 10.1039/b710425a

34) J.A. Kritzer, O.M. Stephens, D.A. Guarracino, *et al.* **β-Peptides as inhibitors of protein–protein interactions**. Bioorgan. Med. Chem., 13 (2005), pp. 11-16, 10.1016/j.bmc.2004.09.009

35) K.D. Stigers, M.J. Soth, J.S. Nowick. **Designed molecules that fold to mimic protein secondary structures**. Curr. Opin. Chem. Biol., 3 (1999), pp. 714-723, 10.1016/s1367-5931(99)00030-7

36) G. Bhardwaj, V.K. Mulligan, C.D. Bahl, *et al.* **Accurate de novo design of hyperstable constrained peptides**. Nature, 538 (2016), pp. 329-335, 10.1038/nature19791

37) A.M. Watkins, R. Bonneau, P.S. Arora. **Modeling Peptide-Protein Interactions, Methods and Protocols**. Methods Mol. Biology Clifton N J, 1561 (2017), pp. 291-307, 10.1007/978-1-4939-6798-8_17

38) P.M. Cromm, J. Spiegel, T.N. Grossmann. **Hydrocarbon Stapled Peptides as Modulators of Biological Function**. ACS Chem. Biol., 10 (2015), pp. 1362-1375, 10.1021/cb501020r

39) L.D. Walensky, G.H. Bird. **Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress**. J. Med. Chem., 57 (2014), pp. 6275-6288, 10.1021/jm4011675

40) J.L. Stebbins, E. Santelli, Y. Feng, *et al.* **Structure-Based Design of Covalent Siah Inhibitors**. Chem. Biol., 20 (2013), pp. 973-982, 10.1016/j.chembiol.2013.06.008

41) C. Hoppmann, L. Wang. **Proximity-enabled bioreactivity to generate covalent peptide inhibitors of p53–Mdm4**. Chem. Commun., 52 (2016), pp. 5140-5143, 10.1039/c6cc01226d

42) J. Dong, L. Krasnova, M.G. Finn, *et al.* **Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry**. Angew. Chem. Int. Ed., 53 (2014), pp. 9430-9448, 10.1002/anie.201309399

43) A. Narayanan, L.H. Jones. **Sulfonyl fluorides as privileged warheads in chemical biology**. Chem. Sci., 6 (2015), pp. 2650-2659, 10.1039/c5sc00408j

44) Q. Li, G. Lozano. **Molecular Pathways: Targeting Mdm2 and Mdm4 in Cancer Therapy**. Clin. Cancer Res., 19 (2013), pp. 34-41, 10.1158/1078-0432.ccr-12-0053

45) J. Charoenpattarapreeda, Y.S. Tan, J. Iegre, *et al.* **Targeted covalent inhibitors of MDM2 using electrophile-bearing stapled peptides**. Chem. Commun., 55 (2019), pp. 7914-7917, 10.1039/c9cc04022f

46) Y.H. Lau, Y. Wu, P. de Andrade, *et al.* **A two-component "double-click" approach to peptide stapling**. Nat. Protoc., 10 (2015), pp. 585-594, 10.1038/nprot.2015.033

47) C. Baggio, L. Gambini, P. Udompholkul, *et al.* **Design of Potent pan-IAP and Lys-Covalent XIAP Selective Inhibitors Using a Thermodynamics Driven Approach**. J. Med. Chem., 61 (2018), pp. 6350-6363, 10.1021/acs.jmedchem.8b00810

48) W. Chen, J. Dong, L. Plate, *et al.* **Arylfluorosulfates Inactivate Intracellular Lipid Binding Protein(s) through Chemoselective SuFEx Reaction with a Binding Site Tyr Residue**. J. Am. Chem. Soc., 138 (2016), pp. 7353-7364, 10.1021/jacs.6b02960

49) D.E. Mortenson, G.J. Brighty, L. Plate, *et al.* **"Inverse Drug Discovery" Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates**. J. Am. Chem. Soc., 140 (2017), pp. 200-210, 10.1021/jacs.7b08366

50) N. Wang, B. Yang, C. Fu, *et al.* **Genetically Encoding Fluorosulfate- l -tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo**. J. Am. Chem. Soc., 140 (2018), pp. 4995-4999, 10.1021/jacs.8b01087

51) L. Gambini, C. Baggio, P. Udompholkul, *et al.* **Covalent Inhibitors of Protein-Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues**. J. Med. Chem., 62 (2019), pp. 5616-5627, 10.1021/acs.jmedchem.9b00561

52) M. Moiola, M.G. Memeo, P. Quadrelli. **Stapled Peptides—A Useful Improvement for Peptide-Based Drugs**. Molecules, 24 (2019), p. 3654, 10.3390/molecules24203654

53) A. Patgiri, A.L. Jochim, P.S. Arora. **A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in α-Helical Conformation**. Accounts Chem. Res., 41 (2008), pp. 1289-1300, 10.1021/ar700264k

54) D.Y. Yoo, A.D. Hauser, S.T. Joy, *et al.* **Covalent Targeting of Ras G12C by Rationally Designed Peptidomimetics**. ACS Chem. Biol., 15 (2020), pp. 1604-1612, 10.1021/acschembio.0c00204

55) D.J. Maly, J.A. Allen, K.M. Shokat. **A Mechanism-Based Cross-Linker for the Identification of Kinase−Substrate Pairs**. J. Am. Chem. Soc., 126 (2004), pp. 9160-9161, 10.1021/ja048659i

56) J.M. Ostrem, U. Peters, M.L. Sos, *et al.* **K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions**. Nature, 503 (2013), pp. 548-551, 10.1038/nature12796

57) C.F. Barbas, A. Heine, G. Zhong, *et al.* **Immune Versus Natural Selection: Antibody Aldolases with Enzymic Rates But Broader Scope**. Science, 278 (1997), pp. 2085-2092, 10.1126/science.278.5346.2085

58) A.J. Chmura, M.S. Orton, C.F. Meares. **Antibodies with infinite affinity**. Proc. National Acad. Sci., 98 (2001), pp. 8480-8484, 10.1073/pnas.151260298

59) L. Holm, P. Moody, M. Howarth. **Electrophilic Affibodies Forming Covalent Bonds to Protein Targets**. J. Biol. Chem., 284 (2009), pp. 32906-32913, 10.1074/jbc.m109.034322

60) C.C. Liu, P.G. Schultz. **Adding New Chemistries to the Genetic Code**. Annu. Rev. Biochem., 79 (2010), pp. 413-444, 10.1146/annurev.biochem.052308.105824

61) J. Liu, F. Zheng, R. Cheng, *et al.* **Site-Specific Incorporation of Selenocysteine Using an Expanded Genetic Code and Palladium-Mediated Chemical Deprotection**. J. Am. Chem. Soc., 140 (2018), pp. 8807-8816, 10.1021/jacs.8b04603

62) J. Liu, S. Li, N.A. Aslam, *et al.* **Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo**. J. Am. Chem. Soc., 141 (2019), pp. 9458-9462, 10.1021/jacs.9b01738

63) Z. Xiang, H. Ren, Y.S. Hu, *et al.* **Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity**. Nat. Methods, 10 (2013), pp. 885-888, 10.1038/nmeth.2595

64) Z. Xiang, V.K. Lacey, H. Ren, *et al.* **Proximity-Enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids**. Angew. Chem. Int. Ed., 53 (2014), pp. 2190-2193, 10.1002/anie.201308794

65) B. Yang, S. Tang, C. Ma, *et al.* **Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions**. Nat. Commun., 8 (2017), p. 2240, 10.1038/s41467-017-02409-z

66) X.-H. Chen, Z. Xiang, Y.S. Hu, *et al.* **Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors**. ACS Chem. Biol., 9 (2014), pp. 1956-1961, 10.1021/cb500453a

67) C. Hoppmann, I. Maslennikov, S. Choe, *et al.* **In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light**. J. Am. Chem. Soc., 137 (2015), pp. 11218-11221, 10.1021/jacs.5b06234

68) T. Kobayashi, C. Hoppmann, B. Yang, *et al.* **Using Protein-Confined Proximity To Determine Chemical Reactivity**. J. Am. Chem. Soc., 138 (2016), pp. 14832-14835, 10.1021/jacs.6b08656

69) J.L. Furman, M. Kang, S. Choi, *et al.* **A Genetically Encoded aza-Michael Acceptor for Covalent Cross-Linking of Protein-Receptor Complexes**. J. Am. Chem. Soc., 136 (2014), pp. 8411-8417, 10.1021/ja502851h

70) W. Xuan, J. Li, X. Luo, P.G. Schultz. **Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins**. Angew. Chem. Int. Ed., 55 (2016), pp. 10065-10068, 10.1002/anie.201604891

71) W. Xuan, S. Shao, P.G. Schultz. **Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains**. Angew. Chem. Int. Ed., 56 (2017), pp. 5096-5100, 10.1002/anie.201611841

72) B. Yang, N. Wang, P.D. Schnier, *et al.* **Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins**. J. Am. Chem. Soc., 141 (2019), pp. 7698-7703, 10.1021/jacs.9b02611

73) Q. Li, Q. Chen, P.C. Klauser, *et al.* **Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics**. Cell, 182 (2020), pp. 85-97.e16, 10.1016/j.cell.2020.05.028

74) D.B.F. Johnson, J. Xu, Z. Shen, *et al.* **RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites**. Nat. Chem. Biol., 7 (2011), pp. 779-786, 10.1038/nchembio.657

75) Y. Zhang, J.L. Ptacin, E.C. Fischer, *et al.* **A Semi-synthetic Organism that Stores and Retrieves Increased Genetic Information**. Nature, 551 (2017), pp. 644-647, 10.1038/nature24659

Chapter 3: Utilizing Genetically Encoded Chemical Crosslinking for the Discovery of Novel Binding Partners of Mammalian Histidine Protein Kinases

3.1 Abstract

Histidine phosphorylation is an understudied post-translational modification in mammalian cells that has recently been linked with important signaling pathways. The two known mammalian histidine protein kinases, NDPKA and NDPKB, have few confirmed substrates. In order to further understand the role of phosphohistidine in protein structure and function, we need to develop a method to robustly capture and identify substrate proteins of NDPKA and NDPKB. We hypothesize that unnatural amino acid (Uaa) incorporation via genetic code expansion will allow us to elucidate the functional roles of phosphohistidine in mammalian cell biology. We have previously developed bioreactive Uaas to capture and identify protein-protein interactions in live cells. We will explore using bioreactive Uaas to capture unknown interacting partners of the two mammalian histidine kinases across multiple cell lines, which we will then identify and attempt to validate. The findings will give us further insight into these histidine kinases and the identity of their substrates and other interacting proteins, and how they might contribute to cell health and function. Our proposed research will not only provide new technologies enabling previously infeasible studies of phosphohistidine, but also advance our understanding of histidine phosphorylation in mammalian cell biology

3.2 Introduction

Phosphorylation is a post-translational modification that plays an important role in pathways controlling cell proliferation, apoptosis, and survival. Misregulation of these pathways is linked with various human diseases. It is therefore vital for therapeutic development to fully understand this modification and its related pathways.[1-3] Histidine phosphorylation, which plays an important role in bacterial gene expression and metabolism, has recently been implicated in important mammalian cellular processes, including epigenetic signaling, ion channel conduction, and metastasis suppression.[4-7] These studies suggest phosphohistidine plays a far more important role in mammalian cell biology than previously thought. However, this modification remains understudied due to the labile nature of the phosphoramidate bond, which gets readily cleaved under the acidic conditions typically used to study phosphoproteins.[8,9] Consequently, we lack the tools necessary to identify sites of phosphohistidine modification or characterize the functional roles this modification plays in proteins and signaling pathways. This has led to controversy over the physiological relevance and biological importance of histidine phosphorylation. It is imperative that we develop new and reliable methods to identify novel phosphohistidine-containing proteins so that we may further study the functional role of phosphohistidine and better understand how this post-translational modification contributes to cellular health and function.

There are two known mammalian histidine kinases: nucleoside-diphosphate kinase (NDPK) A and B. These multifunctional proteins are most commonly known for their roles as house-keeping proteins that catalyze the transfer of γ-phosphate from nucleoside-triphosphates to nucleoside-diphosphates, a mechanism that occurs via formation of a high-energy phosphohistidine intermediate on a conserved histidine residue in the catalytic domain.[10] These two proteins also contribute to metastasis suppression in mammalian cancer cells, though their

exact roles in this complex process are relatively unknown.[11] To add to their growing repertoire of functions, NDPKA and NDPKB have also been proven to act as histidine protein kinases, however, only a handful of substrates have so far been identified.[12,13] Some of those substrates include the G protein beta subunit and the potassium ion channel KCa3.1.[14,15] How histidine phosphorylation affects the function of these substrates is still under investigation, as is the question of how histidine phosphorylation affects cellular health in general. Some studies have suggested that the protein kinase activities of NDPKA and NDPKB are linked with their roles as metastasis suppressers; other studies show that global histidine phosphorylation levels differ between healthy and cancerous cells.[7,16] Identifying novel substrates and interacting partners of NDPKA and NDPKB would give us further insight into how histidine phosphorylation affects cell health.

Advances in the field of genetic code expansion allow us to incorporate unnatural amino acids (Uaas) at specific sites in proteins-of-interest, making it a powerful tool for the cellular study of protein function and interactions.[17,18] Genetic code expansion has recently been utilized to study enzyme-substrate interactions through the use of proximity-enabled bioreactive Uaas.[19,20] Bioreactive Uaas are generally electrophiles that can crosslink protein-protein interactions by targeting nucleophilic residues near interaction sites. Crosslinking occurs upon binding of the two proteins, and these crosslinked proteins can then be identified via mass spectrometry analysis as the crosslinked peptides of the two interacting proteins remain intact. Bioreactive Uaas have been used to capture and identify known and unknown enzyme-substrate interactions. Our lab has crosslinked the enzyme UBE2D3 to its substrate PCNA directly in *E. coli* cells by genetically encoding BprY, a bioreactive Uaa that targets cysteine residues, at three sites on PCNA that interact with the catalytic site of UBE2D3. We have also utilized BprY to capture and identify unknown substrates of the enzyme thioredoxin (Trx) in live *E. coli* cells by incorporating BprY at

a site adjacent to the catalytic residue of Trx.[20] Other studies have also shown that enzyme-substrate interactions, known and unknown, can be captured in mammalian cells.[21] These studies together confirm that bioreactive Uaas have the potential to be powerful tools for the study of protein-protein and, more specifically, enzyme-substrate interactions.

Many bioreactive Uaas only target cysteine residues, greatly limiting the range of substrates that can be covalently trapped to ones with cysteine at or near the interaction site. In order to expanding our crosslinking capabilities beyond one amino acid and be able to target a wider range of the proteome, our lab has recently developed a Uaa, fluorosulfate-L-tyrosine (FSY), that can crosslink with lysine, tyrosine, or histidine residues.[22] FSY reacts via sulfur fluoride exchange, or SuFEx click chemistry, which is robust, stable, and compatible with aqueous and physiological conditions.[22,23] FSY has been shown to crosslink known interacting proteins *in vitro*, and in cells when both proteins are overexpressed. Thus far, FSY has not been used to identify novel interactions or to crosslink proteins that are expressed endogenously. We believe FSY has the potential to be a really powerful tool for capturing unknown interactions.

We hypothesize that unnatural amino acid incorporation via genetic code expansion would create innovative and reliable tools in the study of phosphohistidine and phosphohistidine-containing proteins**.** We plan to use our bioreactive Uaa FSY, alongside other Uaas, to capture unknown proteins that interact with the histidine kinases NDPKA and NDPKB in live cells. We will identify these proteins via mass spectrometry analysis and attempt to validate the interactions in order to bring further insight into the types of proteins interacting with NDPKA and NDPKB and how these proteins might contribute to the complex biological functions of these proteins.

3.3 Results

Designing mutant kinases to capture substrate proteins. Previous studies of protein-protein interactions using bioreactive Uaas have relied heavily on crystal structures to determine where best to place the unnatural amino acids so that it is in proximity to a nucleophilic residue on the protein of interest one is trying to capture. When attempting to study or capture unknown interactions, determining where to engineer in the Uaa is less straightforward. When approaching the study of NDPKA and NDPKB, we first began by looking at their active site residues, as we determined that any substrate that binds to these kinases would have to bind in close proximity to the catalytic histidine residue 118, which is the residue responsible for phosphate transfer during the phosphorylation of NDPs, and which is also believed to be responsible for the phosphorylation of protein substrates as well. Based on the crystal structure of both NDPKA and NDPKB (**Fig. S3.1**), we decided to pick six residues initially for mutation to FSY: L55, F60, L64, R88, T94, and R105. These residues sit primarily to the front of the active site, where the bases of NDPs and NTPs bind. We hoped that any protein or substrate that binds would be near that area of the active site, and would thus be captured by our FSY Uaa, potentially even at the histidine position which would be phosphorylated by the catalytic histidine. After initially testing these mutants, we decided to make four more mutants at the K66, Y67, S125, and K128 positions. We expressed these mutant NDPKA and NDPKB proteins in HEK293T cells and were excited to see that for a number of the NDPKB mutants, there were potential crosslinking bands visible after Western analysis of cell lysates (**Fig 3.1a**). Unfortunately, NDPKA mutants did not express well in HEK293T cells (**Fig. S3.2**), thus the majority of our mammalian cell work was conducted with NDPKB. To compare FSY to other unnatural amino acids we had in our repertoire, we also tested some of the mutant NDPKB proteins with FmnbY, our photocaged quinone methide Uaa, or with Azi, a

photocrosslinking Uaa. The results showed consistent, reproducible crosslinking patterns across several of the mutants(**Fig 3.1b**). These results ultimately suggest that FSY as these positions is capturing one or two very major interactions which dominate over any other interaction. Because FSY is less reactive than photoactivated Uaas, we expect to see it capture major or tightly binding interactions since it requires more time to react with a nucleophilic residue in close proximity. The results with the photoactivated Uaas should have shown more bands if any other proteins were captured that bound less tightly or less frequently. However, since the results all looked the same, the sites we chose to incorporate FSY into are not allowing us to capture a wider range of interacting partners because they are being biased towards a specific interacting partner or two. We expect that mass spectrometry analysis, which is more sensitive than Western blot analysis, might show more crosslinks in the sample that we cannot visualize but can potentially purify out.

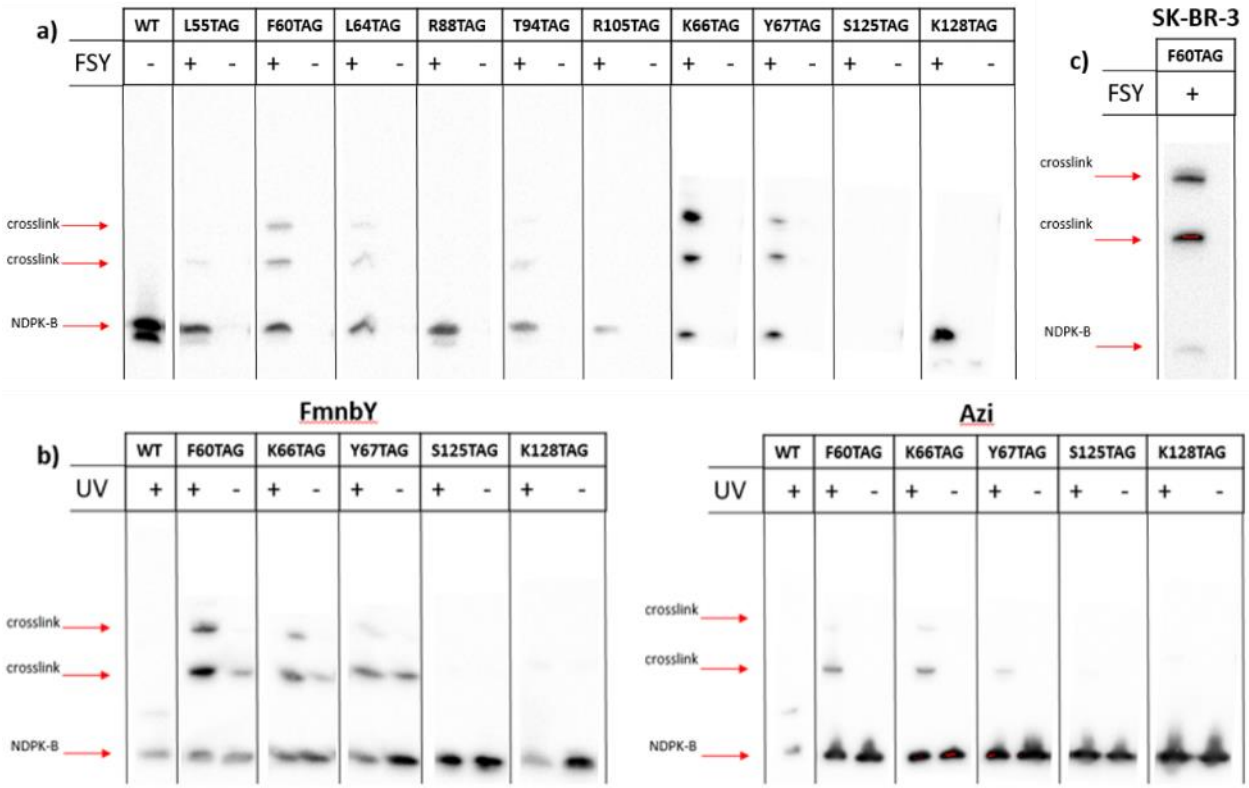


Fig 3.1: Crosslinking of NDPKB mutants. a) Mutant protein incorporating FSY expressed in

HEK293T cells. Certain mutants show potential crosslinking bands. **b**) Mutant proteins incorporating FmnbY expressed in HEK293T cells. Similar band patter as with FSY crosslinking, even though FmnbY can crosslink a wider array of amino acids, even though it is photoactivatable. Mutant proteins incorporating Azi. Similar crosslinking pattern as FSY. **c**) Expression of F60 mutant in SK-BR-3 cells shows same crosslinking pattern as in HEK293T cells, suggesting no new proteins captured.

We also briefly looked into whether we could capture different proteins in a different cell line, particularly a breast cancer model cell line like SK-BR-3. After transiently transfecting one of our most robustly expressing mutants, NDPKB-F60FSY, into SK-BR-3 cells, we saw a similar crosslinking pattern as in our HEK293T studies (**Fig 3.1c**).

In order to identify these potentially crosslinked proteins, we turned to mass spectrometric analysis. We decided to focus first on the NDPKB-F60FSY mutant, as it expressed robustly in HEK293T cells and showed very strong upper bands in Western analysis. Before our initial mass spectrometry study, we had not yet made the K66FSY or Y67FSY mutants, which also show robust expression and crosslinking. First, we scaled up the expression of NDPKB-F60FSY in HEK293T cells so that we could express enough for purification from cell lysate and mass spectrometry analysis. We conducted a time course assay to determine the best time to harvest our cells for optimal concentration of crosslinked complexes (**Fig. S3.3**). In order to purify the crosslinked protein complexes, we initially decided to use immunoprecipitation (IP) pull down on the Flag tag at the N terminus of our NDPKB mutant constructs. Though overall yield of NDPKB-F60FSY protein and its crosslinked complexes was low using this method (**Fig. S3.4**), and we were unable to visualize bands on SDS-PAGE gel, we ultimately purified enough for mass spectrometric analysis. Samples were trypsin digested and further prepared using an established protocol, then sent off for analysis via Orbitrap Mass Spec.

Mass spectrometry data was analyze using either MaxQuant or OpenUaa data analyzing software. The initial mass spec data for the NDPKB-F60FSY mutant, which consisted of two

biological replicates, which was analyzed via MaxQuant, showed that the proteins with the largest sequence coverage were NDPKA and NDPKB (**Fig. S3.5**). Due to the over 90% sequence homology between NDPKA and NDPKB, we postulated that the high sequence coverage for NDPKA in our sample was most likely due to the analysis software assigning NDPKB peptides to NDPKA as well. Most of the other proteins identified in these samples appeared to be contaminations, such as members of the keratin family of proteins, or members of the heat shock protein (HSP) family. We were unable to identify any crosslinked peptides using MaxQuant.

We conducted further mass spectrometry studies, using the same IP pulldown protocol, but this time further scaling up our expression of NDPKB-F60FSY. We conducted analysis of four biological replicates of purified NDPKB-F60FSY and crosslinked complexes using OpenUaa software, which is specifically designed to identify crosslinked peptides in mass spectrometric data. The results of this analysis showed NDPKB-F60FSY peptides crosslinked with four potential proteins of interest: 2-5A-dependent ribonuclease (RNASEL), E3 SUMO2 ligase KIAA1586, zinc finger protein 568, and SET-binding protein (**Fig. S3.6**). None of the identified proteins matched the speculated molecular weights as seen in the crosslinked bands of the Westerns. While some of the proteins, notably the E3 SUMO2 ligase KIAA1586 and SET-binding protein, show some literature precedent for potentially interacting with NDPKA and NDPKB, we have not yet been able to establish a valid connection between the identified proteins and NDPKB. Attempts to validate in cell crosslinking between one of the proteins of interest, 2-5A-dependent ribonuclease, and NDPKB-F60FSY by overexpressing both proteins in HEK293T cells was unsuccessful (**Fig. S3.7**).

One interesting discovery in the mass spectrometry data was that NDPKB-F60FSY crosslinked with residues Y67 and K128 of NDPKB protein (**Fig 3.2**). This might suggest that this

crosslinking is occurring in an intramolecular manner. The F60 residue is relatively close to the Y67 and K128 residues in the active site of NDPKB. However, when studying the crystal structure closely, we see that the distance between the Cα of F60 and the O of Y67 is around 12.1 Å, and the distance between the Cα of F60 and the Nε of K128 is 16.7 Å (**Fig. S3.8**). The ideal reactive distance for FSY is 9Å or less. In order for intramolecular crosslinking to occur, the residues would have to be brought into closer proximity, thought the F60 residue is at the beginning of an unstructured loop of the protein, which might also bring it closer to residues Y67 and K128. However, the formation of intramolecular crosslinking bonds would not explain away the bands we see in the Western blot. The residues Y67 and K128, due to their proximity, would act like chemical sponges, absorbing the majority of available F60FSY. Instead of crosslinking with any potential interacting proteins, F60FSY is more likely to crosslink with its neighboring residues instead. This might explain why we do not see a lot of crosslinking bands, but the very prominent bands we do see would have to be proteins that are out-competing F60FSY crosslinking with Y67 or K128.

This then suggests that NDPKB-F60FSY is potentially crosslinking in an intermolecular manner, thus forming crosslinked dimers or other oligomers. This would explain the crosslinking bands seen in Western blot analysis, as they appear to be around the correct MW to be dimer and trimer formations. However, mutating the residues Y67 and K128 to phenylalanine and arginine, respectfully, does not alter the band pattern we see in Western blot analysis of NDPKB-F60FSY expressed in HEK293T cells (**Fig. S3.9a**). If these residues were causing the formation of intermolecular bonds which we then see as the bands on the Western blot, mutating them to inert residues should eliminate or significantly reduce the bands we see. This is not the case. Furthermore, the F60 residue does not sit at the interface of dimer/oligomer formation for the

NDPKB protein, and thus should not be crosslinking the protein in an intermolecular manner. Crystal structure analysis shows that F60 is very distal from Y67 and K128 residues on other NDPKB proteins when in an oligomeric form (**Fig. S3.9b**). Therefore, we are still unsure about the manner of crosslinking between F60FSY and the Y67 and K128 residues.

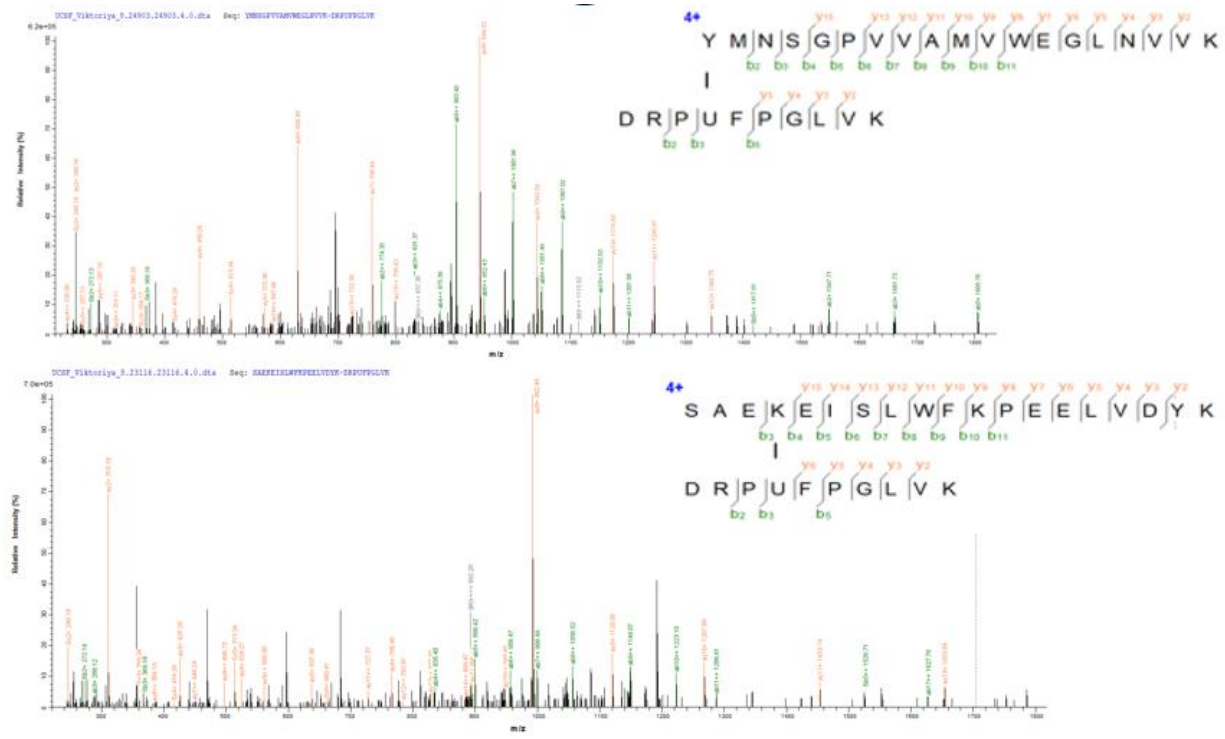


Fig 3.2: Mass spectrometry results of NDPKB-F60FSY mutant. Results show self-crosslinking. Ion map of NDPKB-F60FSY peptide crosslinked with Y67-containing peptide (top) and ion map of NDPKB-F60FSY peptide crosslinked with K128-containing peptide (bottom).

Due to the inconclusive results of the mass spectrometry data of the previously studied mutant, NDPKB-F60FSY, we decided to embark upon another mutagenesis of NDPKB. We selected an additional 25 residues dispersed around NDPKB, on top of the already studied 10 mutants, and engineered FSY into each new position (Fig. S10). Of the new 25 mutants, only one, N115FSY, showed any new crosslinking band pattern upon Western blot analysis (**Fig. 3.3**)

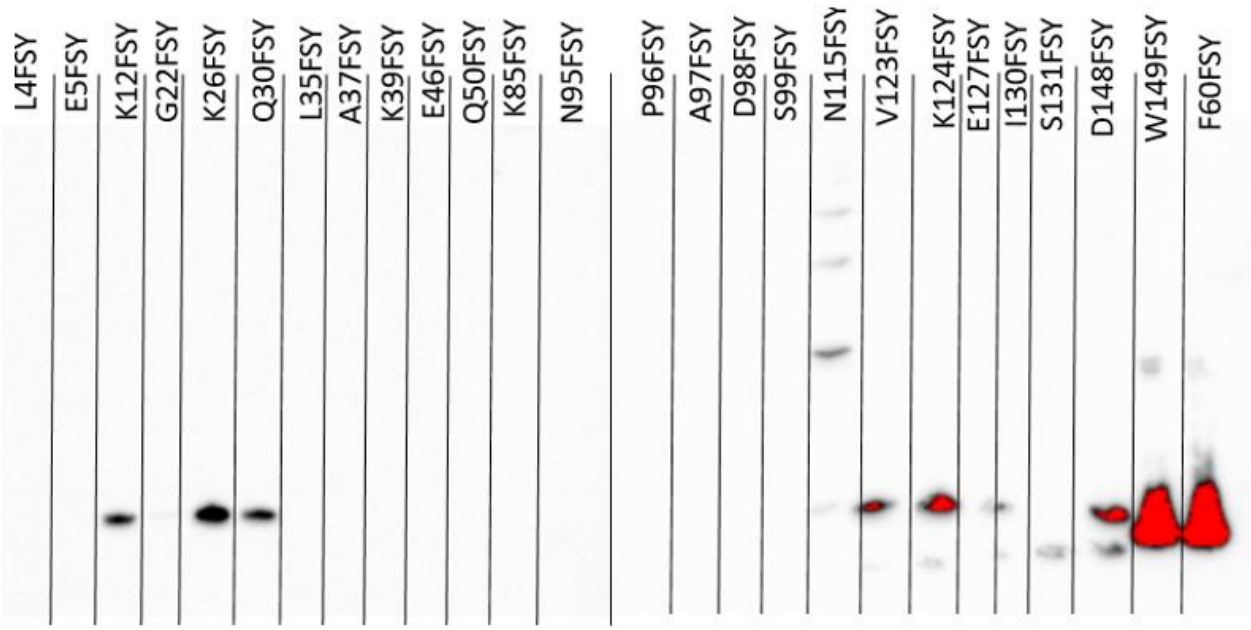


Fig 3.3: Expression of new mutants of NDPKB. New mutants show lack of potential crosslinking bands except for N115FSY mutant, which shows multiple bands, and has a different banding pattern than the previously studied F60FSY mutant.

Expressing NDPKB-N115FSY in HEK293T cells once again, we are able to reproduce the binding pattern we saw initially (**Fig. 3.4a**). This time, in order to purify out the N115FSY containing protein and its crosslinked complexes for mass spec analysis, we first express NDPKB-N115FSY in E. coli. When expressed in E. coli, NDPKB-N115FSY shows minimal bands in Western blot analysis, suggesting that most of the bands seen in the Western blot of the protein expression in HEK293T are uniquely mammalian proteins. When NDPKB-F60FSY protein is expressed in E. coli, the same crosslinking pattern appears in Western blot analysis as when NDPKB-F60FSY is expressed in HEK293T cells. (**Fig. S3.11**). After expressing the mutant protein in E coli, we purify it out using anti-Flag beads, which still use an anti-Flag antibody to purify out the protein, but because of better protein yield from E coli expression, we are able to purify out more NDPKB-N115FSY protein. Next, we incubate the purified mutant protein, still

attached to Flag beads, with mammalian cell lysate. After incubation, we wash and elute. The banding pattern of NDPKB-N115FSY when purified this way is the same as when expressed directly in HEK293T cells, suggesting that this method is as effective at capturing binding partners of NDPKB-N115FSY as directly expressing the mutant protein in HEK293T cells (**Fig. 3.4b**).

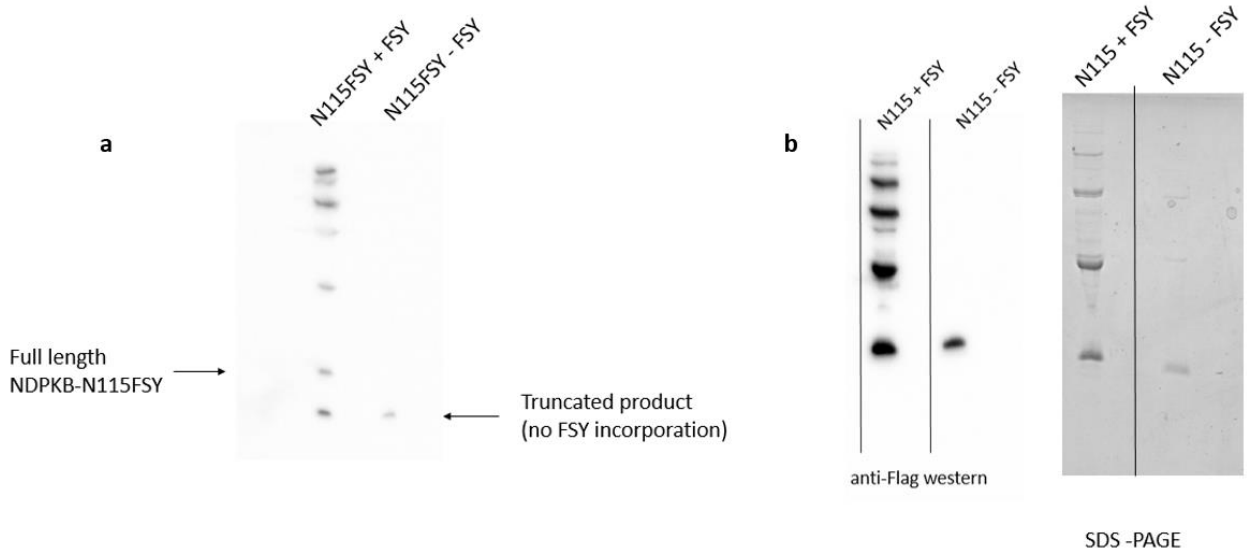


Fig 3.4: Expression and purification of NDPKB-N115FSY mutant. a) Expression of NDPKB-N115FSY in HEK293T cells. **b**) Western blot analysis of NDPKB-N115FSY after purification from E coli and incubation with mammalian cell lysate. The band pattern in this Western matches the band pattern for the protein after direct expression in HEK293T cells. SDS-PAGE gel shows similar band patter, suggesting a higher yield than in previous methods of purification.

The purified samples of NDPKB-N115FSY protein and crosslinked complexes were prepared for mass spectrometry analysis in the same method as previously used to analyze the NDPKB-F60FSY mutant. The results were analyzed using OpenUaa. OpenUaa software once again identified a number of potential crosslinked proteins (**Fig. S3.12**), however, these proteins were not identified in a second biological replicate of NDPKB-N115FSY, suggesting that this data might not be conclusive. The analysis did identify more crosslinks between NDPKB-N115FSY

and other residues of NDPKB, including 8 new crosslinked site on top of residues Y67 and K128 (Fig. S3.13). Due to N115 being situated in the active site, we believe that most of these crosslinks are intramolecular, and that N115FSY is capable of interacting and crosslinking with all of these residues.

We have submitted more samples for mass spectrometry analysis in hopes of getting reproducible results. Data for that is still unavailable.

We have also attempted to express NDPKA-N115FSY in E. coli, incubate it with mammalian cell lysate, and look for any potential crosslinks after purification using Flag resin. For the first time, we see successful expression of an NDPKA mutant, which we were never able to see in mammalian cells (Fig. 3.5). Furthermore, we see that NDPKA-N115FSY, when incubated with mammalian cell lysate, also captures some proteins from the lysate. Samples of NDPKA-N115FSY has been sent out for mass spectrometry analysis and the results are pending.

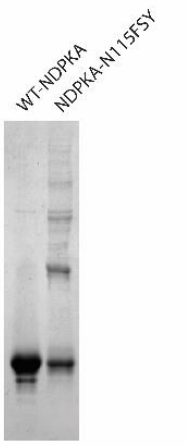


Fig. 3.5: NDPKA expression in E. coli. SDS-PAGE gel showing NDPKA-WT and NDPKA-N115FSY after expression in E. coli and incubation with mammalian cell lysate. NDPKA mutant expresses robustly in E. coli compared with previous HEK293T expression data, and NDPKA-N115FSY captures proteins from mammalian cell lysate.

3.4 Discussion

Our attempts to utilize bioreactive Uaas to capture novel interacting proteins of the kinases NDPKA and NDPKB have shown some of the strengths and drawbacks of this new method for studying unknown and undiscovered protein-protein interactions. First, not every mutant protein will express well, as seen with NDPKA mutants which did not express well in mammalian cells. Furthermore, identifying the interactions can be difficult and data intensive, as the more mutants one wants to study, and the more biological replicates, the greater the mass spectrometry data set is. Validating the mass spectrometry data is also crucial to ensure that the data is accurate.

Some of the benefits of utilizing this method to look for new PPIs is the relative easy to make a large mutant library. Furthermore, when studying mammalian protein interactions, one can scale up the expression of the mutant proteins by expression them in high volume in E coli, then incubating the purified protein with mammalian cell lysate. The bioreactive Uaa is still active and the proteins can still crosslink, as long as one is not conducting the experiment in reducing conditions and the mammalian cells have been lysed gently.

There is still much work to be done to prove the utility of bioreactive Uaas in capturing and elucidating novel protein interactions. Enzymes in particular are difficult proteins to study due to the transient nature of their substrate binding. But with constitutively active Uaas, which can bind without the need for UV activation, one is also more likely to capture a wider range of proteins, not just proteins that interact with the mutant protein at one instance in time upon UV irradiation.

3.5 Materials and Methods

NDPKA amino acid sequence

MANCERTFIAIKPDGVQRGLVGEIIKRFEQKGFRLVGLKFMQASEDLLKEHYVDLKDRPFFAGLVKYMHSGPVVAMVW
EGLNVVKTGRVMLGETNPADSKPGTIRGDFCIQVGRNIIHGSDSVESAEKEIGLWFHPEELVDYTSCAQNWIYE

NDPKB amino acid sequence

MANLERTFIAIKPDGVQRGLVGEIIKRFEQKGFRLVAMKFLRASEEHLKQHYIDLKDRPFFPGLVKYMNSGPVVAMVW
EGLNVVKTGRVMLGETNPADSKPGTIRGDFCIQVGRNIIHGSDSVKSAEKEISLWFKPEELVDYKSCAHDWVYE

NDPKA codon sequence

ATGGCCAACTGTGAGCGTACCTTCATTGCGATCAAACCAGATGGGGTCCAGCGGGGTCTTGTGGGAGAGATTATC
AAGCGTTTTGAGCAGAAAGGATTCCGCCTTGTTGGTCTGAAATTCATGCAAGCTTCCGAAGATCTTCTCAAGGAAC
ACTACGTTGACCTGAAGGACCGTCCATTCTTTGCCGGCCTGGTGAAATACATGCACTCAGGGCCGGTAGTTGCCAT
GGTCTGGGAGGGGCTGAATGTGGTGAAGACGGGCCGAGTCATGCTCGGGGAGACCAACCCTGCAGACTCCAAGC
CTGGGACCATCCGTGGAGACTTCTGCATACAAGTTGGCAGGAACATTATACATGGCAGTGATTCTGTGGAGAGTG
CAGAGAAGGAGATCGGCTTGTGGTTTCACCCTGAGGAACTGGTAGATTACACGAGCTGTGCTCAGAACTGGATCT
ATGAATGA

NDPKB codon sequence

ATGGCCAACCTGGAGCGCACCTTCATCGCCATCAAGCCGGACGGCGTGCAGCGCGGCCTGGTGGGCGAGATCAT
CAAGCGCTTCGAGCAGAAGGGATTCCGCCTCGTGGCCATGAAGTTCCTCCGGGCCTCTGAAGAACACCTGAAGCA
GCACTACATTGACCTGAAAGACCGACCATTCTTCCCTGGGCTGGTGAAGTACATGAACTCAGGGCCGGTTGTGGC
CATGGTCTGGGAGGGGCTGAACGTGGTGAAGACAGGCCGAGTGATGCTTGGGGAGACCAATCCAGCAGATTCAA
AGCCAGGCACCATTCGTGGGGACTTCTGCATTCAGGTTGGCAGGAACATCATTCATGGCAGTGATTCAGTAAAAA
GTGCTGAAAAAGAAATCAGCCTATGGTTTAAGCCTGAAGAACTGGTTGACTACAAGTCTTGTGCTCATGACTGGGT
CTATGAATGA

Construction of NDPKA and NDPKB mutants

NDPKB Forward Primer for pBad plasmid	AAG AAG GAG ATA TAC ATC ATG GCC AAC CTG GAG CGC ACC T
NDPKB Reverse Primer for pBad plasmid (His tag)	CAT CCG CCA AAA CAG CCA AGC TTT TAG TGA TGG TGA TGG TGA TGT TCA TAG ACC CAG TCA TGA GCA CAA G
NDPKA Forward Primer for pBad plasmid	AAG AAG GAG ATA TAC ATC ATG GCC AAC TGT GAG CGT ACC TTC
NDPKA Reverse Primer for pBad plasmid (His tag)	CAT CCG CCA AAA CAG CCA AGC TTT TAG TGA TGG TGA TGG TGA TGT TCA TAG ATC CAG TTC TGA GCA CAG CTC
NDPKB Forward Primer for pcDNA plasmid (Flag tag)	GCT GGC TAG CGT TTA AAC TTA GCC ACC ATG GAC TAC AAG GAC GAC GAT GAC AAG GGT GGC GGA GGT AGC ATG GCC AAC CTG GAG CGC ACC
NDPKB Reverse Primer for pcDNA plasmid (His tag)	TTC CAC CAC ACT GGA CTA GTG GAT CTT AGT GAT GGT GAT GGT GAT GTT CAT AGA CCC AGT CAT GAG CAC AAG

NDPKA Forward Primer for pcDNA	GCT GGC TAG CGT TTA AAC TTA GCC ACC ATG GAC TAC AAG GAC GAC GAT GAC AAG GGT GGC GGA GGT AGC ATG GCC AAC TGT GAG CGT ACC
NDPKA Reverse Primer for pcDNA	TTC CAC CAC ACT GGA CTA GTG GAT CTC ATT CAT AGA TCC AGT TCT GAG CAC AGC TC
NDPKB L55TAG Forward Primer	CTA CAT TGA CTA GAA AGA CCG ACC ATT C
NDPKB L55TAG Reverse Primer	TGC TGC TTC AGG TGT TCT
NDPKB F60TAG Forward Primer	AGA CCG ACC ATA GTT CCC TGG GC
NDPKB F60TAG Reverse Primer	TTC AGG TCA ATG TAG TGC
NDPKB L64TAG Forward Primer	CTT CCC TGG GTA GGT GAA GTA CAT G
NDPKB L64TAG Reverse Primer	AAT GGT CGG TCT TTC AGG
NDPKB R88TAG Forward Primer	GAA GAC AGG CTA GGT GAT GCT TGG GGA GAC
NDPKB R88TAG Reverse Primer	ACC ACG TTC AGC CCC TCC
NDPKB T94TAG Forward Primer	GCT TGG GGA GTA GAA TCC AGC AGA TTC
NDPKB T94TAG Reverse Primer	ATC ACT CGG CCT GTC TTC
NDPKB R105TAG Forward Primer	AGG CAC CAT TTA GGG GGA CTT CTG
NDPKB R105TAG Reverse Primer	GGC TTT GAA TCT GCT GGA TTG
NDPKA L55TAG Forward Primer	CTA CGT TGA CTA GAA GGA CCG TCC ATT CT
NDPKA L55TAG Reverse Primer	TGT TCC TTG AGA AGA TCT TCC TTG AGA AGA TCT TC
NDPKA F60TAG Forward Primer	GGA CCG TCC ATA GTT TGC CGG CC
NDPKA F60TAG Reverse Primer	TTC AGG TCA ACG TAG TGT TCC TTG AGA AGA TCT TCG GA
NDPKA L64TAG Forward Primer	CTT TGC CGG CTA GGT GAA ATA CAT GCA CTC AGG
NDPKA L64TAG Reverse Primer	AAT GGA CGG TCC TTC AGG TCA ACG TAG T
NDPKA R88TAG Forward Primer	GAA GAC GGG CTA GGT CAT GCT CGG G
NDPKA R88TAG Reverse Primer	ACC ACA TTC AGC CCC TCC CAG ACC
NDPKA T94TAG Forward Primer	GCT CGG GGA GTA GAA CCC TGC AG
NDPKA T94TAG Reverse Primer	ATG ACT CGG CCC GTC TTC ACC ACA
NDPKA R105TAG Forward Primer	TGG GAC CAT CTA GGG AGA CTT CTG CAT ACA AG
NDPKA R105TAG Reverse Primer	GGC TTG GAG TCT GCA GGG TTG GT
NDPKA K66TAG Forward Primer	CCT GGT GTA GTA CAT GCA CTC AGG GCC GGT AGT TGC
NDPKA K66TAG Reverse Primer	AGT GCA TGT ACT ACA CCA GGC CGG CAA AGA ATG GAC
NDPKA Y67TAG Forward Primer	TGG TGA AAT AGA TGC ACT CAG GGC CGG TAG TTG CCA
NDPKA Y67TAG Reverse Primer	CTG AGT GCA TCT ATT TCA CCA GGC CGG CAA AGA ATG
NDPKA S125TAG Forward Primer	CTG TGG AGT AGG CAG AGA AGG AGA TCG GCT TGT GGT TT
NDPKA S125TAG Reverse Primer	CTT CTC TGC CTA CTC CAC AGA ATC ACT GCC ATG TAT AAT GTT CC
NDPKA K128TAG Forward Primer	TGC AGA GTA GGA GAT CGG CTT GTG GTT TCA CCC TGA G
NDPKA K128TAG Reverse Primer	CCG ATC TCC TAC TCT GCA CTC TCC ACA GAA TCA CTG CC
NDPKB K66TAG Forward Primer	GGG CTG GTG TAG TAC ATG AAC TCA GGG CCG GTT GT
NDPKB K66TAG Reverse Primer	TCA TGT ACT ACA CCA GCC CAG GGA AGA ATG GTC GGT C
NDPKB Y67TAG Forward Primer	GCT GGT GAA GTA GAT GAA CTC AGG GCC GGT TGT GG
NDPKB Y67TAG Reverse Primer	TGA GTT CAT CTA CTT CAC CAG CCC AGG GAA GAA TGG TCG

NDPKB S125TAG Forward Primer	ATT CAG TAA AAT AGG CTG AAA AAG AAA TCA GCC TAT GGT TTA AGC C
NDPKB S125TAG Reverse Primer	TCT TTT TCA GCC TAT TTT ACT GAA TCA CTG CCA TGA ATG ATG TTC
NDPKB K128TAG Forward Primer	GTG CTG AAT AGG AAA TCA GCC TAT GGT TTA AGC CTG AAG AAC TG
NDPKB K128TAG Reverse Primer	GGC TGA TTT CCT ATT CAG CAC TTT TTA CTG AAT CAC TGC CAT G
NDPKB Y67F Forward Primer	GCT GGT GAA GTT CAT GAA CTC AGG GCC GGT TGT GG
NDPKB Y67F Reverse Primer	GAG TTC ATG AAC TTC ACC AGC CCA GGG AAG AAT GGT CG
NDPKB K128R Forward Primer	AAG TGC TGA AAG AGA AAT CAG CCT ATG GTT TAA GCC TGA AGA ACT G
NDPKB K128R Reverse Primer	GCT GAT TTC TCT TTC AGC ACT TTT TAC TGA ATC ACT GCC ATG AAT
NDPKB L4TAG Forward Primer	AAC TAG GAG CGC ACC TTC ATC GCC ATC AAG CC
NDPKB L4TAG Reverse Primer	AAG GTG CGC TCC TAG TTG GCC ATG GTG GCA AGC
NDPKB E5TAG Forward Primer	AAC CTG TAG CGC ACC TTC ATC GCC ATC AAG CC
NDPKB E5TAG Reverse Primer	AAG GTG CGC TAC AGG TTG GCC ATG GTG GCA AG
NDPKB K12TAG Forward Primer	TTC ATC GCC ATC TAG CCG GAC GGC GTG CAG CGC
NDPKB K12TAG Reverse Primer	GGC TAG ATG GCG ATG AAG GTG CGC TCC AGG TT
NDPKB G22TAG Forward Primer	CCT GGT GTA GGA GAT CAT CAA GCG CTT CGA GC
NDPKB G22TAG Reverse Primer	TGA TCT CCT ACA CCA GGC CGC GCT GCA CGC CG
NDPKB K26TAG Forward Primer	GAT CAT CTA GCG CTT CGA GCA GAA GGG ATT CC
NDPKB K26TAG Reverse Primer	CGA AGC GCT AGA TGA TCT CGC CCA CCA GGC CG
NDPKB Q30TAG Forward Primer	CTT CGA GTA GAA GGG ATT CCG CCT CGT GGC CA
NDPKB Q30TAG Reverse Primer	ATC CCT TCT ACT CGA AGC GCT TGA TGA TCT CG
NDPKB L35TAG Forward Primer	AAG GGA TTC CGC TAG GTG GCC ATG AAG TTC CTC CG
NDPKB L35TAG Reverse Primer	ACC TAG CGG AAT CCC TTC TGC TCG AAG CGC TT
NDPKB A37TAG Forward Primer	CCT CGT GTA GAT GAA GTT CCT CCG GGC CTC TG
NDPKB A37TAG Reverse Primer	ACT TCA TCT ACA CGA GGC GGA ATC CCT TCT GC
NDPKB K39TAG Forward Primer	TGG CCA TGT AGT TCC TCC GGG CCT CTG AAG AA
NDPKB K39TAG Reverse Primer	GAG GAA CTA CAT GGC CAC GAG GCG GAA TCC CT
NDPKB E46TAG Forward Primer	CTC TGA ATA GCA CCT GAA GCA GCA CTA CAT TGA
NDPKB E46TAG Reverse Primer	TCA GGT GCT ATT CAG AGG CCC GGA GGA ACT TC
NDPKB Q50TAG Forward Primer	CCT GAA GTA GCA CTA CAT TGA CCT GAA AGA CCG A
NDPKB Q50TAG Reverse Primer	TGT AGT GCT ACT TCA GGT GTT CTT CAG AGG CC
NDPKB K85TAG Forward Primer	AAC GTG GTG TAG ACA GGC CGA GTG ATG CTT GG
NDPKB K85TAG Reverse Primer	CCT GTC TAC ACC ACG TTC AGC CCC TCC CAG AC
NDPKB N95TAG Forward Primer	AGA CCT AGC CAG CAG ATT CAA AGC CAG GCA CC
NDPKB N95TAG Reverse Primer	ATC TGC TGG CTA GGT CTC CCC AAG CAT CAC TCG
NDPKB P96TAG Forward Primer	GAG ACC AAT TAG GCA GAT TCA AAG CCA GGC AC
NDPKB P96TAG Reverse Primer	TCT GCC TAA TTG GTC TCC CCA AGC ATC ACT CG
NDPKB A97TAG Forward Primer	CCA TAG GAT TCA AAG CCA GGC ACC ATT CGT GG
NDPKB A97TAG Reverse Primer	GGC TTT GAA TCC TAT GGA TTG GTC TCC CCA AGC
NDPKB D98TAG Forward Primer	TCC AGC ATA GTC AAA GCC AGG CAC CAT TCG TG
NDPKB D98TAG Reverse Primer	GCT TTG ACT ATG CTG GAT TGG TCT CCC CAA GC
NDPKB S99TAG Forward Primer	AGC AGA TTA GAA GCC AGG CAC CAT TCG TGG GG

NDPKB S99TAG Reverse Primer	CTG GCT TCT AAT CTG CTG GAT TGG TCT CCC CA
NDPKB N115TAG Forward Primer	TGG CAG GTA GAT CAT TCA TGG CAG TGA TTC AGT AA
NDPKB N115TAG Reverse Primer	GAA TGA TCT ACC TGC CAA CCT GAA TGC AGA AG
NDPKB V123TAG Forward Primer	GGC AGT GAT TCA TAG AAA AGT GCT GAA AAA GAA ATC AGC C
NDPKB V123TAG Reverse Primer	TTC TAT GAA TCA CTG CCA TGA ATG ATG TTC CT
NDPKB K124TAG Forward Primer	TTC AGT ATA GAG TGC TGA AAA AGA AAT CAG CCT AT
NDPKB K124TAG Reverse Primer	CAG CAC TCT ATA CTG AAT CAC TGC CAT GAA TGA TG
NDPKB E127TAG Forward Primer	GTG CTT AGA AAG AAA TCA GCC TAT GGT TTA AGC
NDPKB E127TAG Reverse Primer	GAT TTC TTT CTA AGC ACT TTT TAC TGA ATC ACT GCC
NDPKB I130TAG Forward Primer	AAG AAT AGA GCC TAT GGT TTA AGC CTG AAG AA
NDPKB I130TAG Reverse Primer	CCA TAG GCT CTA TTC TTT TTC AGC ACT TTT TAC TGA ATC
NDPKB S131TAG Forward Primer	CTA GCT ATG GTT TAA GCC TGA AGA ACT GGT TG
NDPKB S131TAG Reverse Primer	GCT TAA ACC ATA GCT AGA TTT CTT TTT CAG CAC TTT TTA CTG A
NDPKB D148TAG Forward Primer	TGC TCA TTA GTG GGT CTA TGA ACA TCA CCA TCA
NDPKB D148TAG Reverse Primer	AGA CCC ACT AAT GAG CAC AAG ACT TGT AGT CAA CCA
NDPKB W149TAG Forward Primer	GCT CAT GAC TAG GTC TAT GAA CAT CAC CAT CAC CAT C
NDPKB W149TAG Reverse Primer	TAG ACC TAG TCA TGA GCA CAA GAC TTG TAG TCA A
NDPKA N115TAG Forward Primer	CAA GTT GGC AGG TAG ATT ATA CAT GGC AGT GAT TCT GTG G
NDPKA N115TAG Reverse Primer	ATC TAC CTG CCA ACT TGT ATG CAG AAG TCT CC
RNASEL Forward Primer for pcDNA	TTC GCC AGA ACC AGC AGC GGA GCC AGC GGA TCC GCA CCC AGG GCT GGC
RNASEL Reverse Primer for pcDNA (HA tag)	AAA CTT AAG CTT GCC ACC ATG GAG AGC AGG GAT CAT AAC AAC

Transient transfection protocol for HEK293T cells

In 6 well plates, HEK293T cells were grown up in DMEM media + 10% FBA. When cells were around 70% confluent, they were washed with 1 mL of PBS and given fresh DMEM (+ 10% FBS) media. After adding fresh media, cell were returned to 37C incubator while transfection media was set up. 2 ug of each plasmid was added to 150 uL of opt-DMEM in an Eppendorf tube (2 ug of pcDNA-NDPKB-WT plasmid for transfection of NDPKB-WT, and 2 ug of pcDNA-NDPKB-mutant + 2 ug pMP-FSYRS/3xtRNA for transfection of mutant NDPKB). 9 uL of PEI reagent was

added to another 150 uL of opt-DMEM in another Eppendorf. The solutions were mixed via pipetting and left at RT for 5 min. After 5 min, the solutions (2 ug DNA/opt-DMEM and 9 uL PEI/opt-DMEM) were mixed together via vigorous pipetting, then left at rt for 15 min. After 15 min, the solutions were added drop-wise to their respective wells containing HEK293T cells. The cells were placed back in the incubator for 6 hr. After 6 hr, 20 uL of 100 mM FSY was added to wells containing mutant NDPKB plasmids. Cells were left to incubate at 37C for 36-48 hrs, then harvested for further analysis.

IP Pull down protocol

Lyse cells using NP-40 buffer with protease inhibitor for 1 hr on ice. Determine the total protein concentration of samples using BCA assay. Add 1-4 ug of primary antibody to cell lysate (whole cell lysate, not pre-cleared); incubate while gently rocking at 4C for 2-4 hours (4 uL of antibody should be ~ 4 ug). Add 50 uL of Protein G beads per 0.5-1 mg total protein in lysate; gently rock mixture at 4C for 1-4 hours. Centrifuge mixture at 1,000 rpm for 30 seconds at 4C; discard supernatant. Wash beads 3-4 X with 1mL 1X TBST with Protease inhibitor; centrifuge and discard supernatant as in step 4. To elute, add 50uL of 2mg/mL Flag peptide (if just running gel to check purity, can directly add 4x laemmli buffer to sample and boil). Remove elution, buffer exchange and concentrate in necessary.

Flag bead purification protocol

5 uL of approx. 1 ng pBad plasmid containing NDPK protein was transformed into DH10b cells. 5 uL og approx. 1 ng pEvol plasmid with FSYRS/3xtRNA was also transformed along with pBad plasmids containing mutant NDPK proteins. Transformed cells were spread on plates with +Amp for WT or +Amp/+Cam for mutant NDPK. Plates were incubated O/N at 37C. Colonies were

picked and E. coli was grown in 50 mL 2XYT +Amp or +Amp/+Cam, shaking, at 37C until OD600. Once at OD, 500 uL of 20% arabinose was added to all cultures; 500 uL of 100 mM FSY was added to cultures containing NDPK mutants. Proteins were expressed O/N, shaking at 18C. Expression cultures were spun down at 4000xg for 30min at 4C. Pellets were re-suspended in 4mL lysis buffer (4mL PBS + 40 uL stock lysozyme solution + 20 uL stock DNase I solution + 40 uL stock protease inhibitor solution). Pellets were sonicated in ice bath (7.5min, 1 sec on, 1 sec off, 40% amp, or until clear). After sonication, samples were spun down at 30,000 g for 30 min at 4C; supernatant was decanted into 15mL tube. Approx 50 uL of Flag resin per sample was washed with 500 uL PBS 3x, then added to tubes with supernatant, and incubated at 4C (rotating) for 2 hours. During the last hour of resin/supernatant incubation, frozen EXPI cell pellets were thawed and lysed with 1 mL RIPA buffer (1mL per pellet) on ice for an hour. Samples incubating with Flag resin were spun down for 3 min at 500g, the supernatant decanted. FLAG resin was washed 3x with 500 uL PBS, then transferred to Eppendorf tubes. The lysed EXPI cells were spun down, 3min at 500g, the supernatant was collected and 500 uL added to each Eppendorf tube with FLAG resin. The FLAG resin was incubated with EXPI cell lysate at room temp for 3-4 hours. After, the samples were spun down, the supernatant removed. Resin was washed with 5x500 uL PBS. 50uL of 2mg/mL FLAG peptide was added to samples, incubated at 4C for 1 hour before eluting. The samples were all spun down, the supernatant was collected as the elution. All samples were buffer exchanged supernatants to storage buffer and concentrated down to ~50 uL using spin columns.

Mass spectrometry work up protocol

To purified crosslinking reaction, cold acetone in a 6:1 ratio and incubate at -20 C for 1hr after mixing well. Spin the samples at -10C 22,000g for 30 min. While samples spin, prepare fresh 8 M Urea (240 mg Urea, 270 uL water, 50 uL 1M Tris pH 8.5). Decant and discard the supernatant

from the reaction and let sit on table for 20 min to complete dry. Add 100 uL of 8M urea to completely dry sample. Sonicate for 10 min. Reduce with 5 mM of TCEP for 20 min at RT (1 uL of 500 mM TCEP). Add alkylation buffer (IAA) 10 mM and incubate at RT in the Dark for 15 min (stock IAA 500 mM, 2 uL). Add 300 uL of 100 mM Tris and 4 uL of 0.5 ug/uL Trypsin (total 2 ug). Incubate at 37 C O/N. After O/N incubation, add Formic Acid until 5% and incubate in 4C while preparing the desalting tips. Use a metal rod to pierce the desalting filter. Push filter into a 200 uL tip until it's tight but not too tight. Peirce the tip onto the top of an Eppendorf tube that is cut up on top with a razor and opened with a 1 mL tip. Activate filter with 20 uL of methanol added to the top of the tip, run at 3000 rpm for 20 sec (note: never let the filter dry). Repeat with buffer B (ACN + Formic acid) and twice with buffer A, making sure to check on samples regularly and don't spin for too long so filter does not run dry. Add your sample to the tip and spin all of it down. Wash with buffer A two times. Elute the sample with buffer B into a fresh tube. Dry the sample in the speedVac.

3.6 Significant Figures

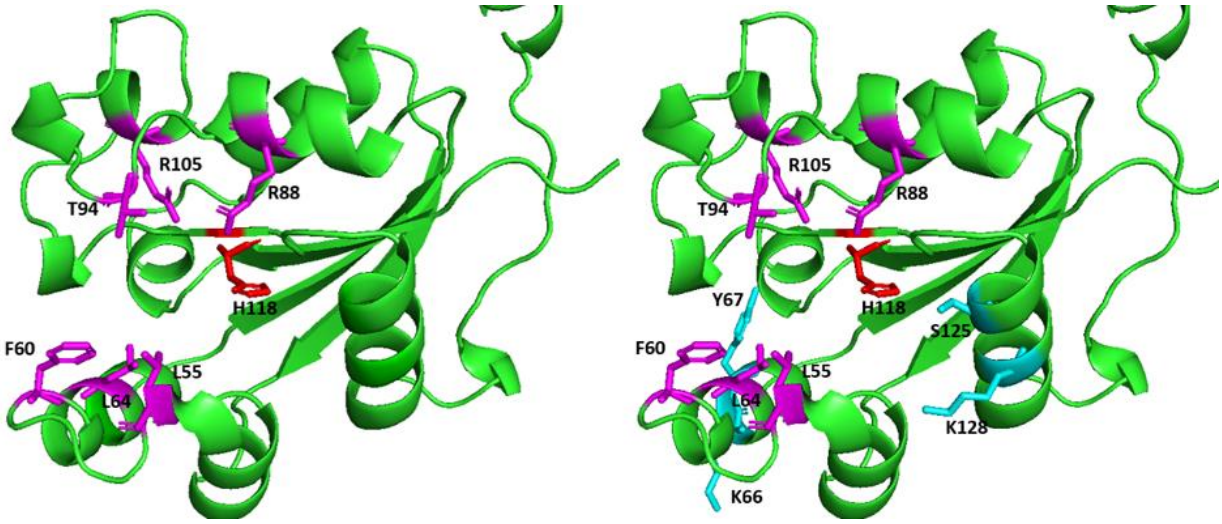


Fig S3.1: Crystal structure of NDPKB with mutated residues highlighted. Residues in magenta (left) were part of the initial screen of mutants made; residues in cyan (right) were part of the second screen of mutants looking for crosslinking.

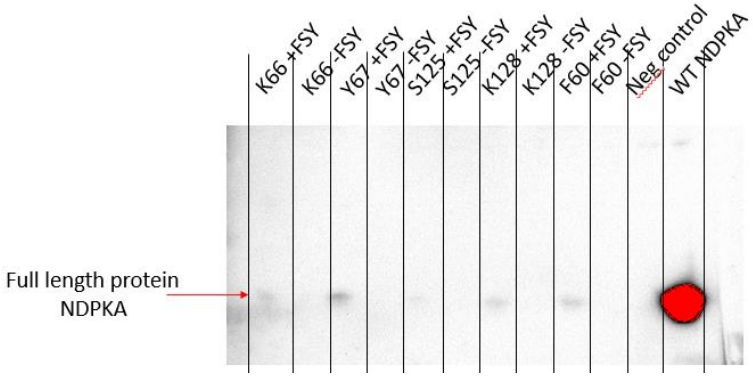


Fig. S3.2: Expression of NDPKA mutants in HEK293T cells. Lanes depict cells transiently transfected with a mutant, with or without FSY added. Very faint expression is seen in samples with FSY added. No upper bands indicating potential crosslinking are seen. WT NDPKA expressed well.

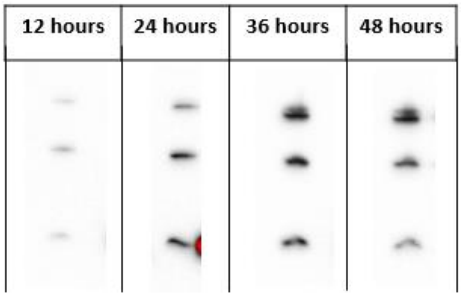


Fig S3.3: Time course assay of NDPKB-F60FSY expression in HEK293T cells. Assay performed to determine optimal cell harvesting time to further study the crosslinked complexes. Data shows that harvesting between 36 and 48 hours would lead to capture of the greatest concentration of crosslinked complexes.

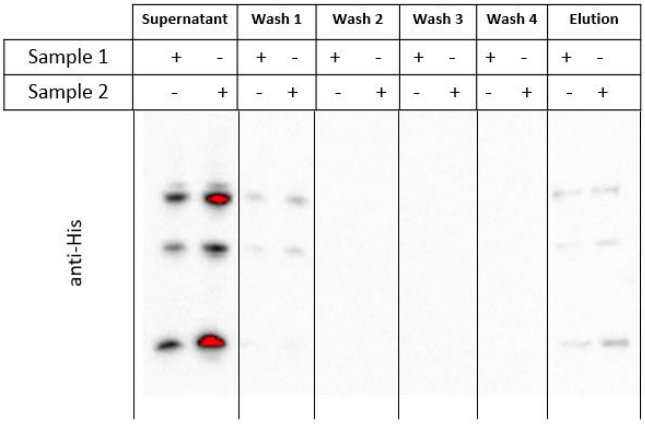


Fig S3.4: Purifying NDPKB-F60FSY mutant using IP pull-down. After transiently transfecting cells and waiting 36 hours, we harvested cells, then lysed before doing IP pull down. Anti-Flag antibody used for IP pulldown.

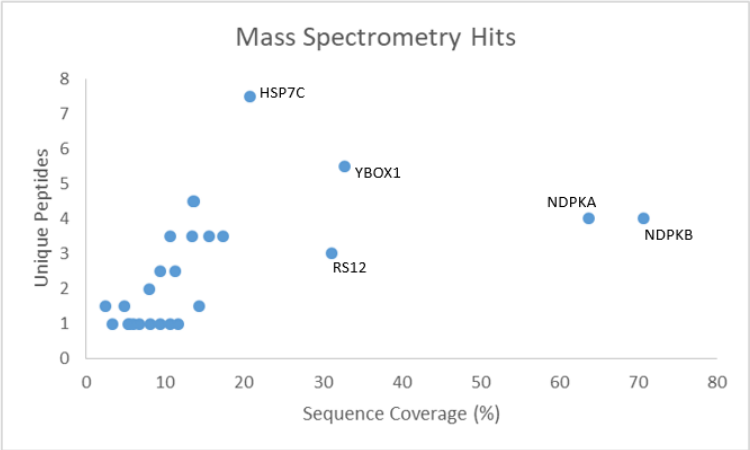


Fig S3.5: Initial mass spectrometry results for NDPKB-F60FSY. Analysis with MaxQuant showed high abundance of NPDKB and NDPKA protein (based on sequence coverage of peptides in mass spec analysis**)**

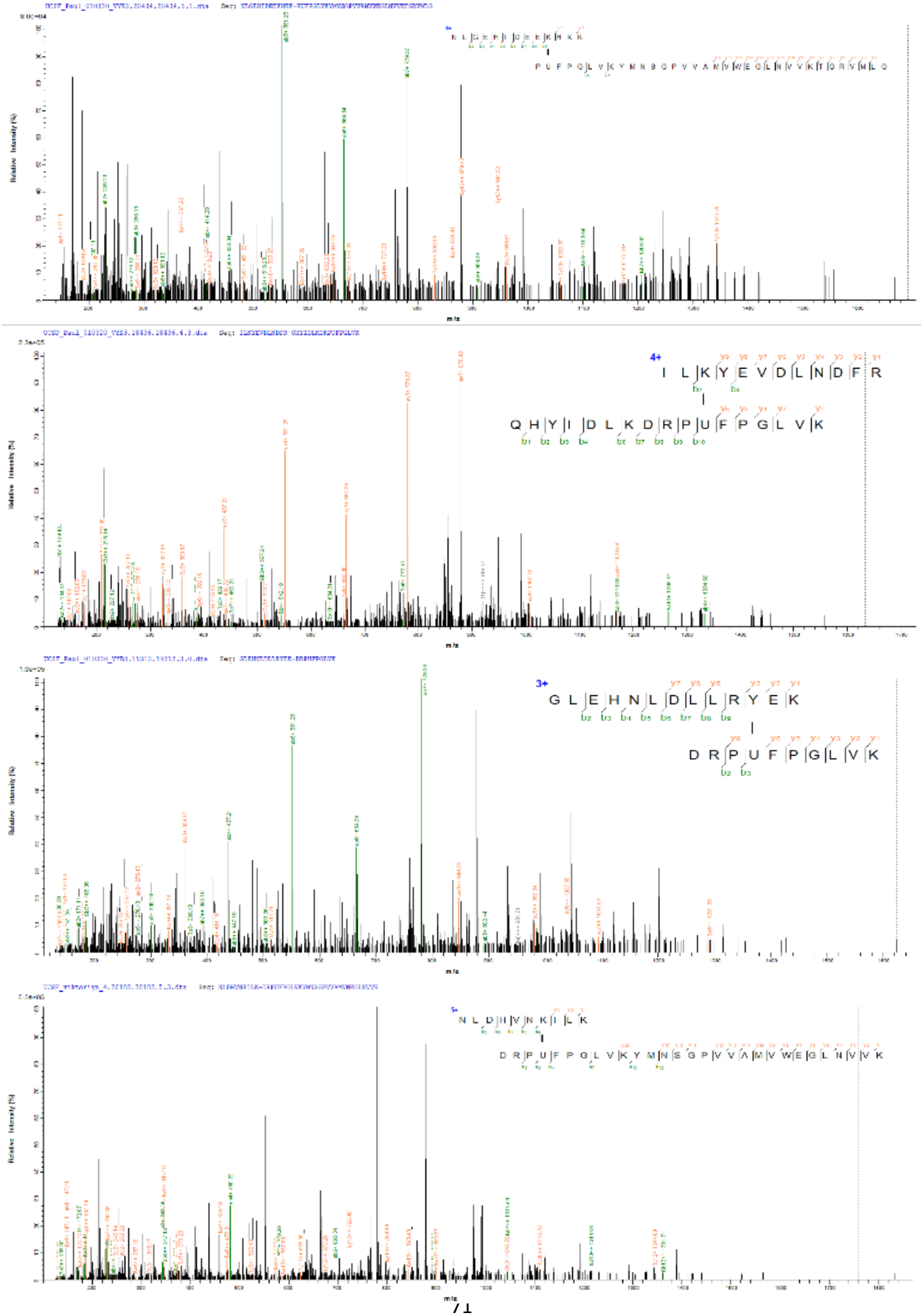

Relative Intensity (%)
m/z
4+
I L K Y E V D L N D F R
Q H Y I D L K D R P U F P G L V K
3+
G L E H N L D L L R Y E K
D R P U F P G L V K
5+
N L D H V N K I L K
D R P U F P G L V K Y M N S G P V V A M V W E G L N V V K

Fig S3.6: Mass spectrometry data for NDPKB-F60FSY. Second round of analysis, done on OpenUaa software, identified crosslinked peptides corresponding to NDPKB-F60FSY and peptides from other proteins. In order from top to bottom: NDPKB-F60FSY peptide crosslinked to 2-5A-dependent ribonuclease; NDPKB-F60FSY peptide crosslinked to E3 ligase KIAA1586; NDPKB-F60FSY peptide crosslinked to zinc finger protein 568; NDPKB-F60FSY peptide crosslinked to SET-binding protein.

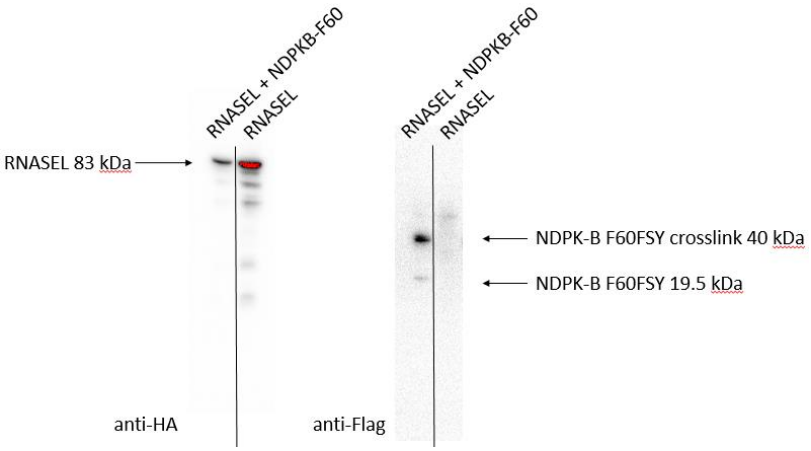


Fig S3.7: Co-expression of RNASEL and NDPKB-F60FSY in HEK293T cells. 2-5A-dependent ribonuclease, also known as RNASEL, was transiently transfected into HEK293T cells along with NDPKB-F60FSY and FSYRS. Western analysis of cell lysates shows no overlap between bands containing RNASEL (anti-HA tag) and NDPKB-F60FSY (anti-Flag tag). This suggests these proteins are not crosslinking when overexpressed in cells.

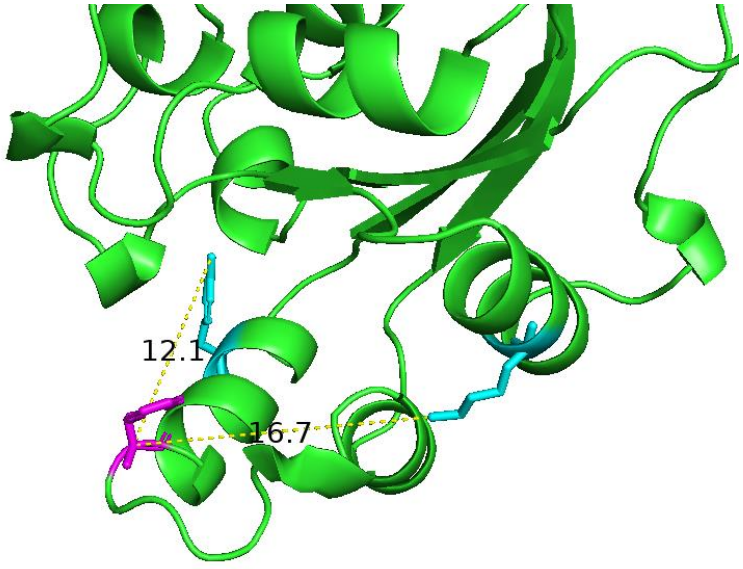


Fig S3.8: Crystal structure of NDPKB showing Y67 and K128 residues. Distances between F60 residue and the Y67 and K128 residues are depicted. Distances calculated from crystal

structures fall outside of the known reaction radius of FSY, though F60 is on a flexible loop region of the protein. PDB: 1NUE.

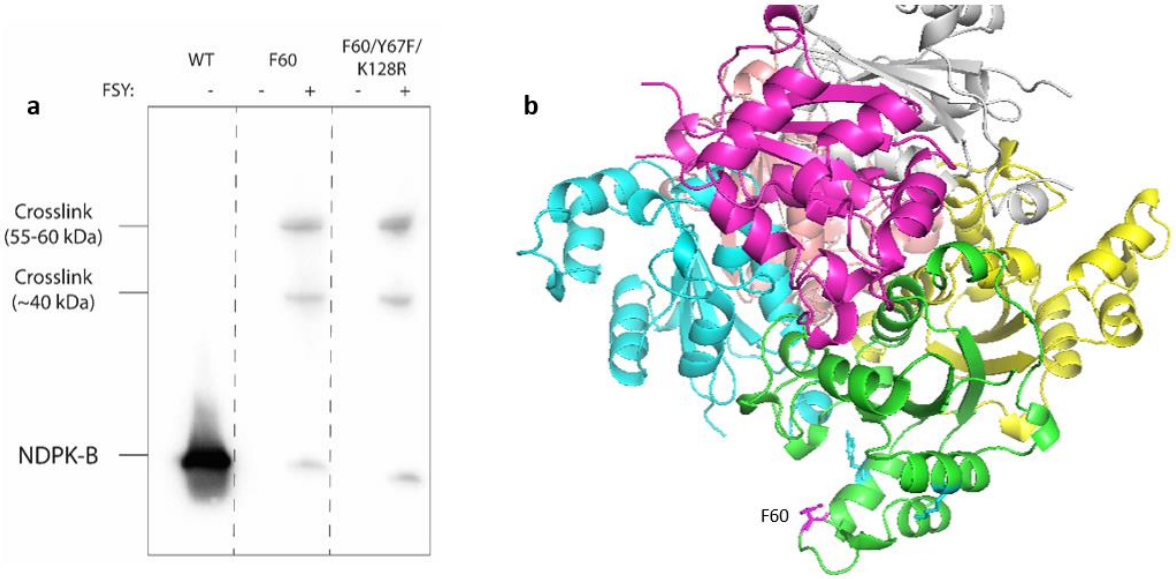


Fig. S3.9: NDPKB-F60FSY/Y67F/K128R mutant expression. a) Western blot analysis of NDPKB-F60FSY crosslinking in HEK293T cells compared with NDPKB-F60FSY-Y67F-K128R mutant, which has Y67 and K128 mutated to non-nucleophilic residues. The same crosslinking patter is seen for both NDPKB-F60FSY and the other mutant protein. Anti-Flag tag Western blot. **b**) Crystal structure showing F60 in reference to the oligomerization interface of NDPKB; F60 sits very distal from the other proteins, suggesting that significant changes would occur in dimerization/oligomerization of NDPKB for intermolecular bonds to form. PDB: 1NUE.

L4FSY	**Q30FSY**	**Q50FSY**	**D98FSY**	**E127FSY**
E5FSY	**L35FSY**	**K85FSY**	**S99FSY**	**I130FSY**
K12FSY	**A37FSY**	**N95FSY**	**N115FSY**	**S131FSY**
G22FSY	**K39FSY**	**P96FSY**	**V123FSY**	**D148FSY**
K26FSY	**E46FSY**	**A97FSY**	**K124FSY**	**W149FSY**

Fig S3.10: Final mutagenesis campaign on NDPKB. 25 residues were selected for mutation to FSY.

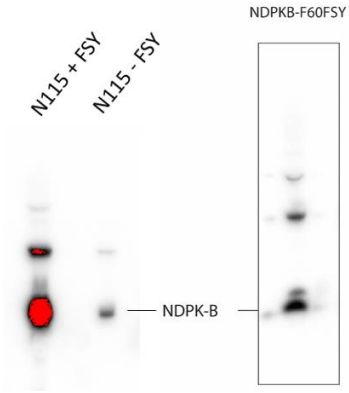
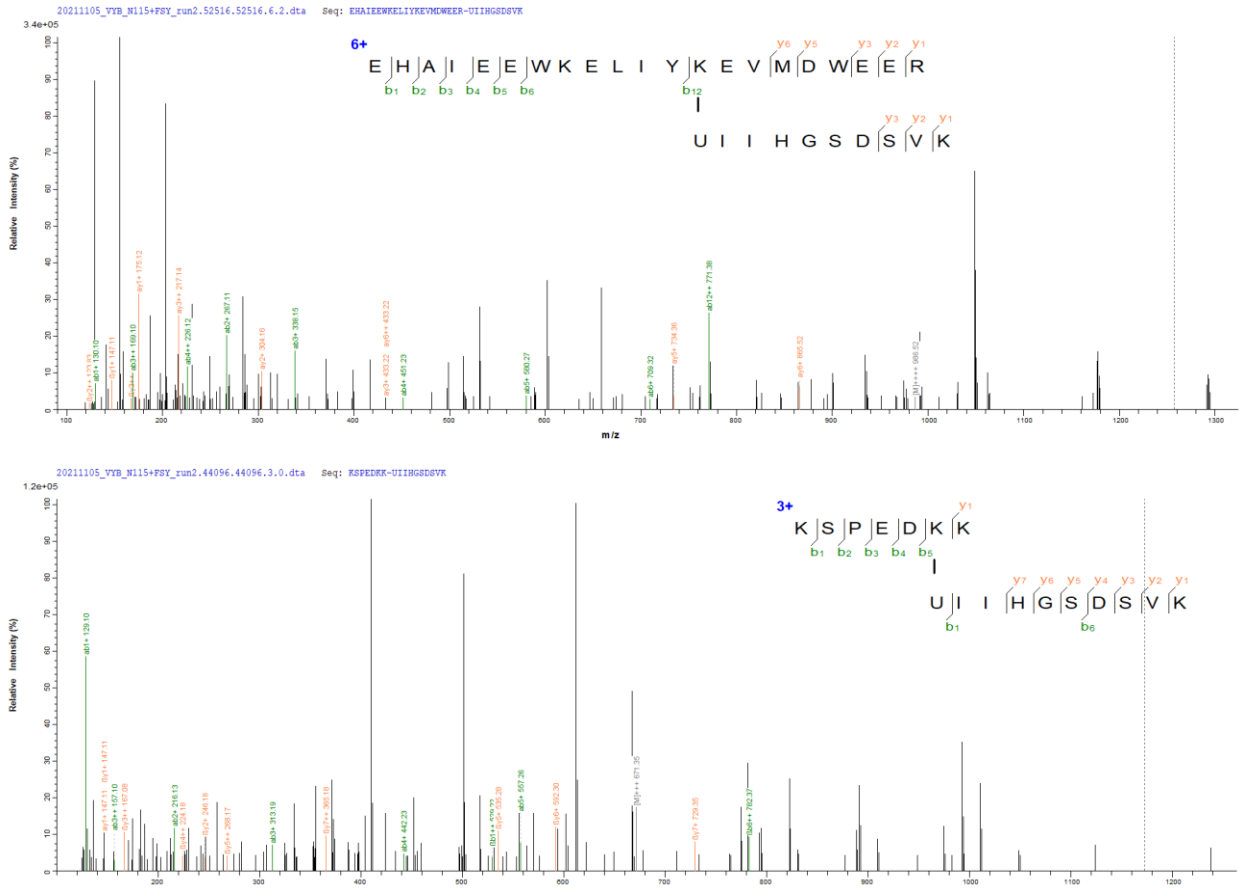
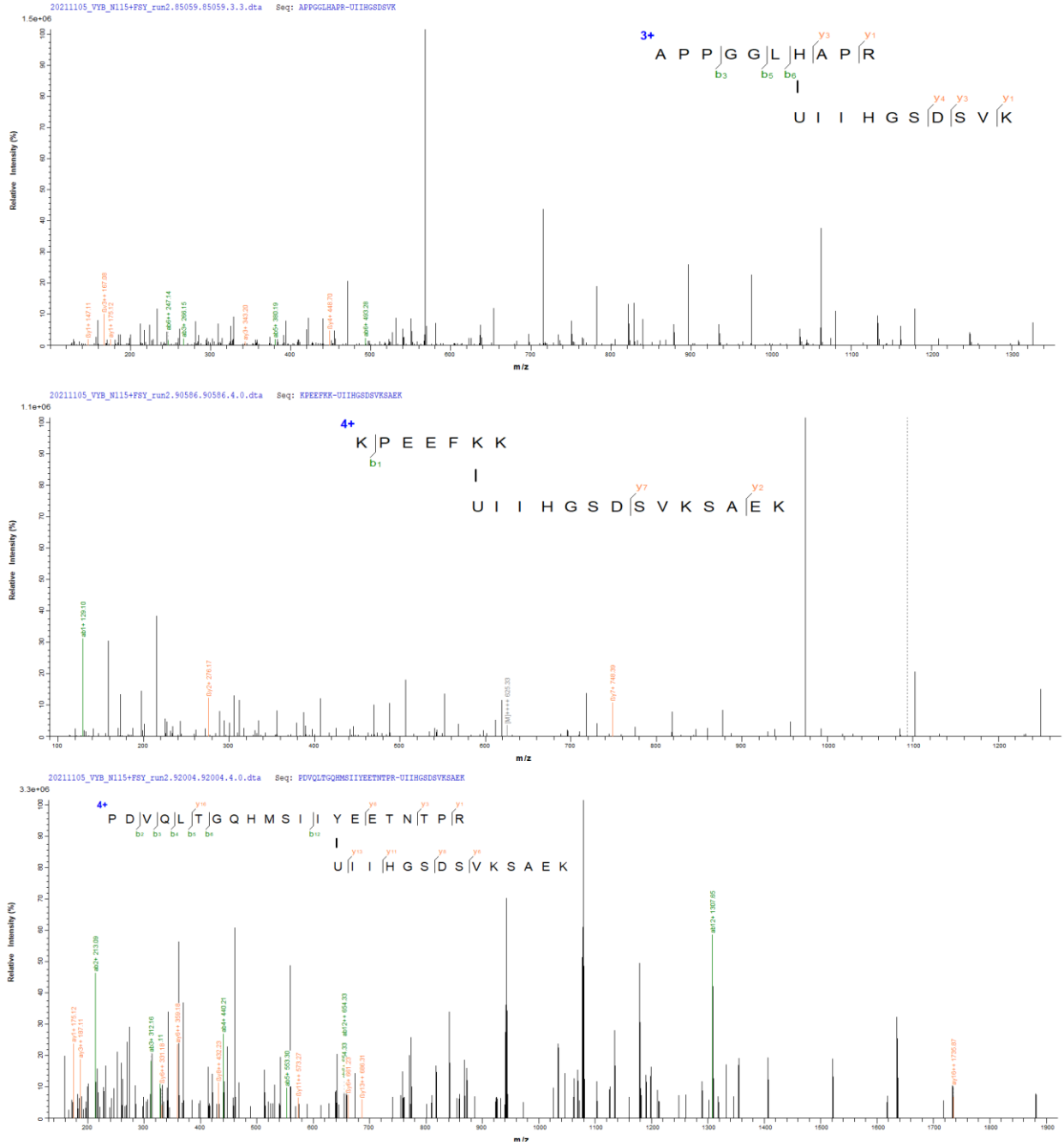


Fig S3.11: Expression of NDPKB-N115FSY and NDPKB-F60FSY in E. coli. NDPKB-N115FSY only two upper bands when expressed in E. coli, as compared with the multiple bands seen when expressed in HEK293T cells. The bands seen in the HEK293T Western blots are mammalian specific proteins. NDPKB-F60FSY has the same crosslinking pattern in E. coli as in HEK293T cells, suggesting the proteins captured are not mammalian specific.




20211105_VYB_N115+FSY_run2.85059.85059.3.3.dta Seq: APPGGLHAPR-UIIHGSDSVK
3+
A P P G G L H A P R
U I I H G S D S V K
Relative Intensity (%)
m/z
20211105_VYB_N115+FSY_run2.90586.90586.4.0.dta Seq: KPEEFKK-UIIHGSDSVKSAEK
4+
K P E E F K K
U I I H G S D S V K S A E K
Relative Intensity (%)
m/z
20211105_VYB_N115+FSY_run2.92004.92004.4.0.dta Seq: PDVQLTGQHMSIIYEETNTPR-UIIHGSDSVKSAEK
4+
P D V Q L T G Q H M S I I Y E E T N T P R
U I I H G S D S V K S A E K
Relative Intensity (%)
m/z

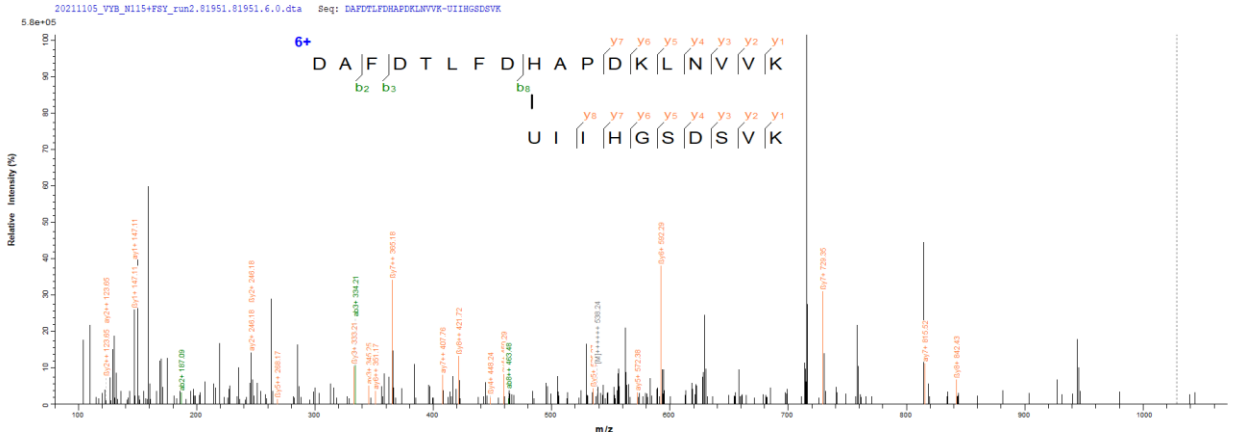
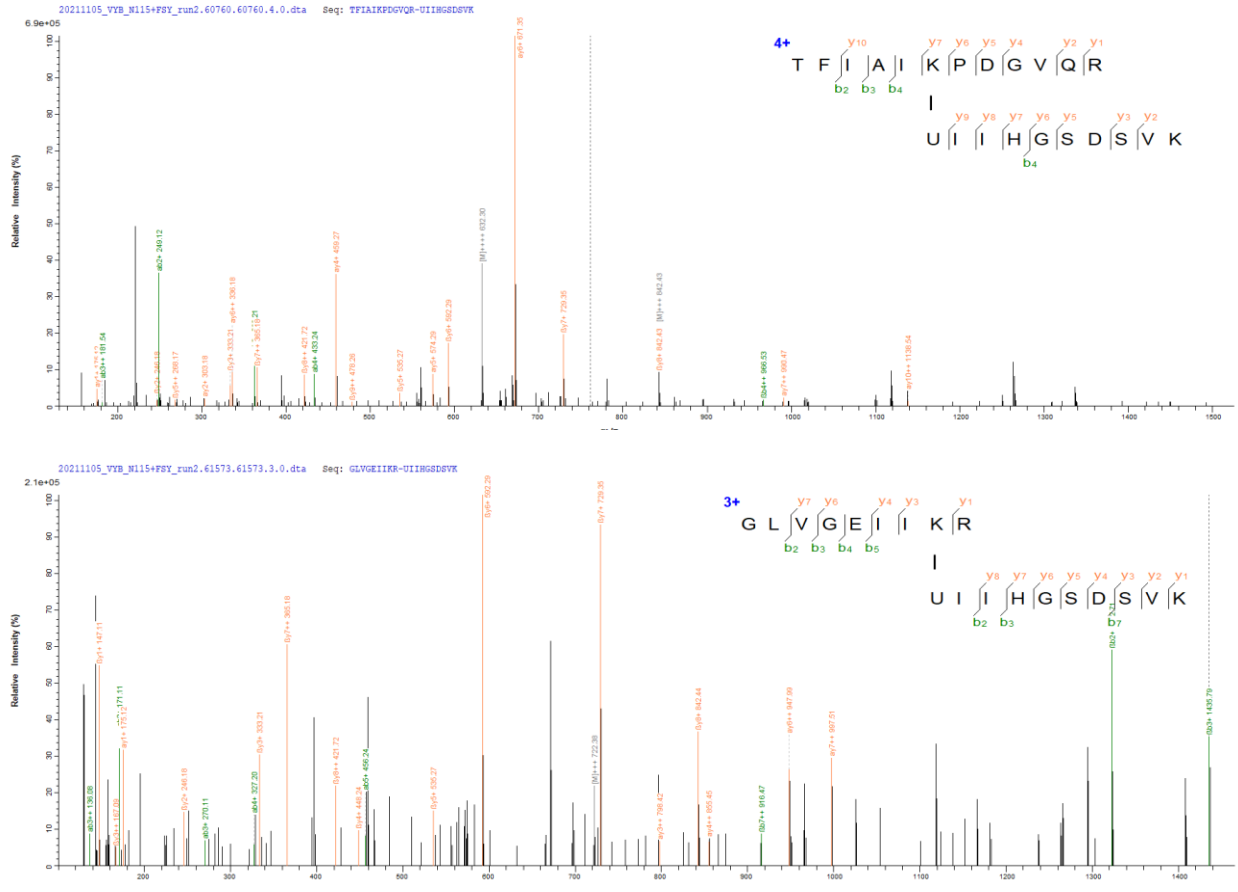
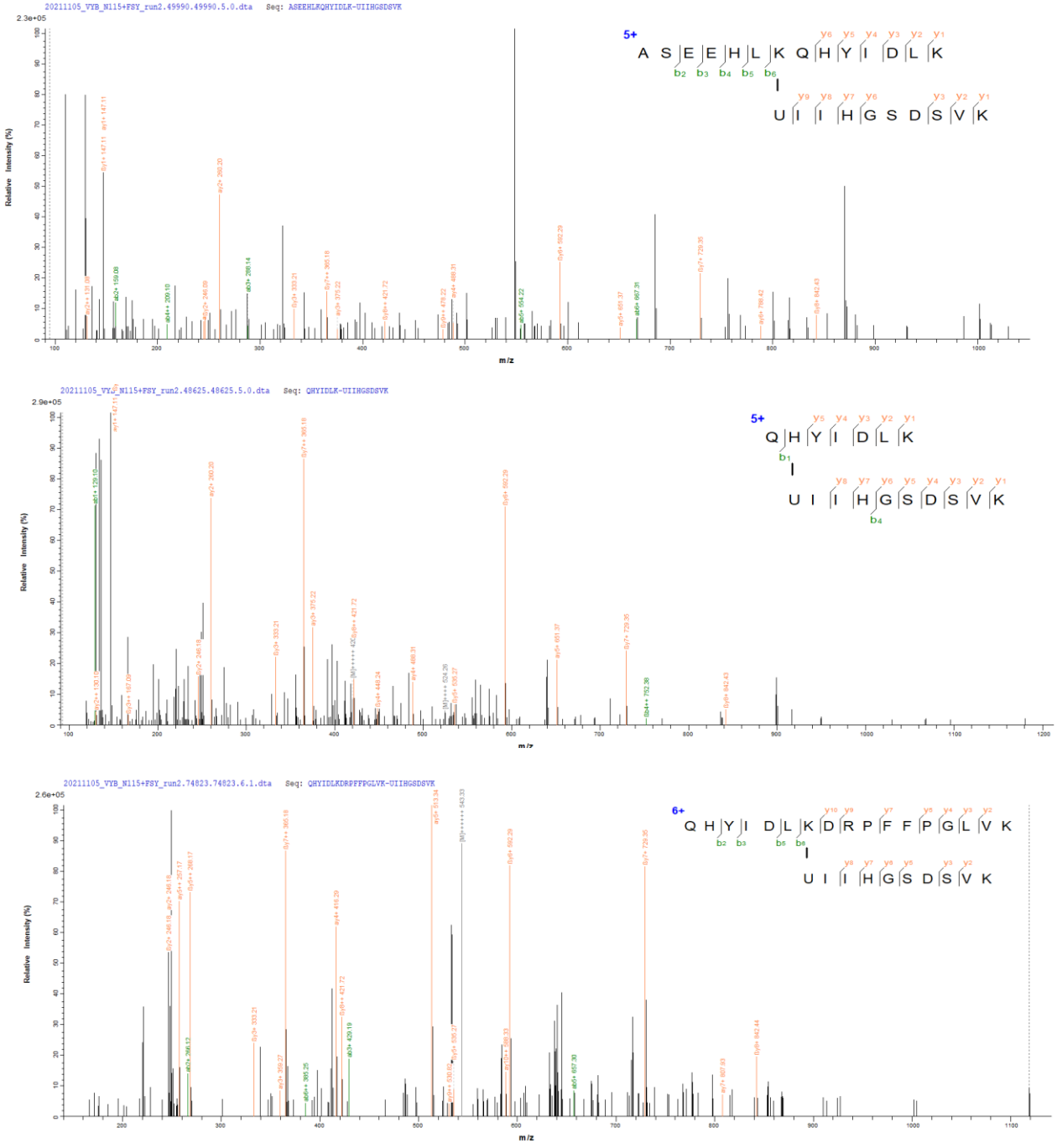
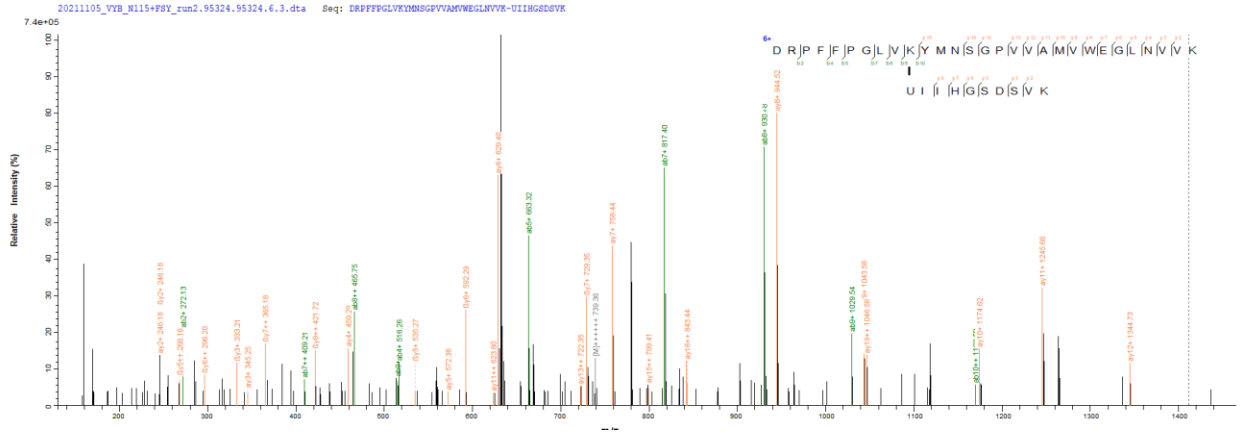
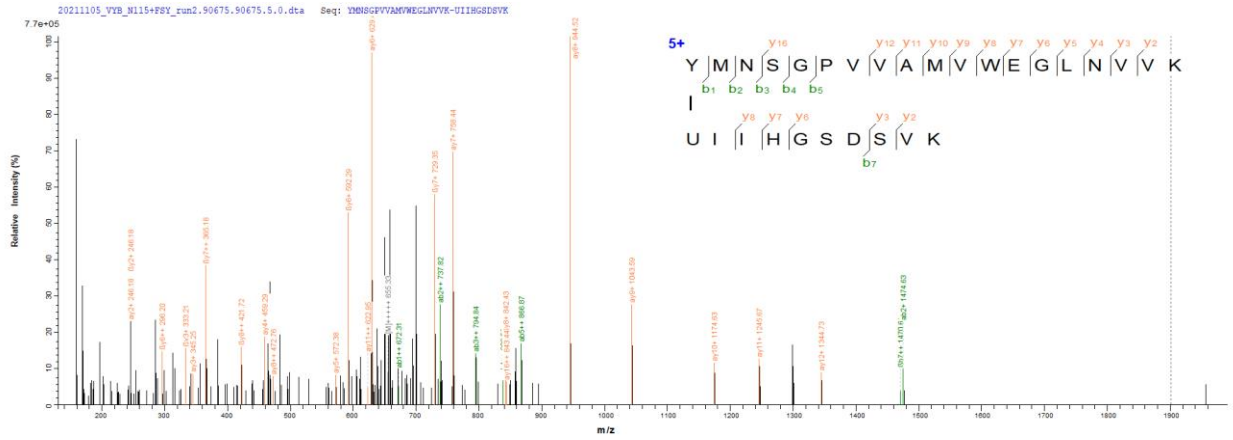


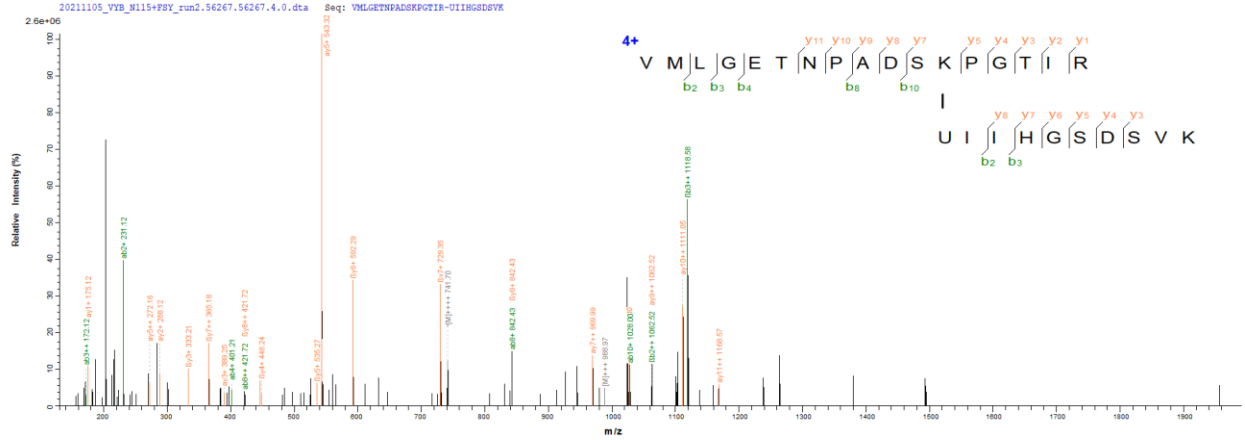
Fig. S3.12: NDPKB-N115FSY crosslinked peptides. Ion maps of crosslinked peptides identified on OpenUaa of NDPKB-N115FSY and unknown proteins. From top to bottom, the proteins identified in each ion map are: MAPK9, mitogen-activated protein kinase; SHANK3, SH3 and multiple ankyrin repeat domains protein 3; uncharacterized protein C6orf132; DNAH7, dynein axonemal heavy chain 7; laminin subunit alpha-3; and alpha-parvin.




20211105_VYB_N115+FSY_run2.49990.49990.5.0.dta Seq: ASEEHLKQHYIDLK-UIIHGSDSVK
Relative Intensity (%)
m/z
5+
A S E E H L K Q H Y I D L K
U I I H G S D S V K
20211105_VYB_N115+FSY_run2.48625.48625.5.0.dta Seq: QHYIDLK-UIIHGSDSVK
Relative Intensity (%)
m/z
5+
Q H Y I D L K
U I I H G S D S V K
20211105_VYB_N115+FSY_run2.74823.74823.6.1.dta Seq: QHYIDLKDRPFFPGLVK-UIIHGSDSVK
Relative Intensity (%)
m/z
6+
Q H Y I D L K D R P F F P G L V K
U I I H G S D S V K


20211105_VYB_N115+FSY_run2.95324.95324.6.3.dta Seq: DRPFFPGLVKYMNSGPVVAMVWEGLNVVK-UIIHGSDSVK
Relative Intensity (%)
m/z


20211105_VYB_N115+FSY_run2.90675.90675.5.0.dta Seq: YMNSGPVVAMVWEGLNVVK-UIIHGSDSVK
Relative Intensity (%)
m/z


20211105_VYB_N115+FSY_run2.56267.56267.4.0.dta Seq: VMLGETNPADSKPGTIR-UIIHGSDSVK
Relative Intensity (%)
m/z

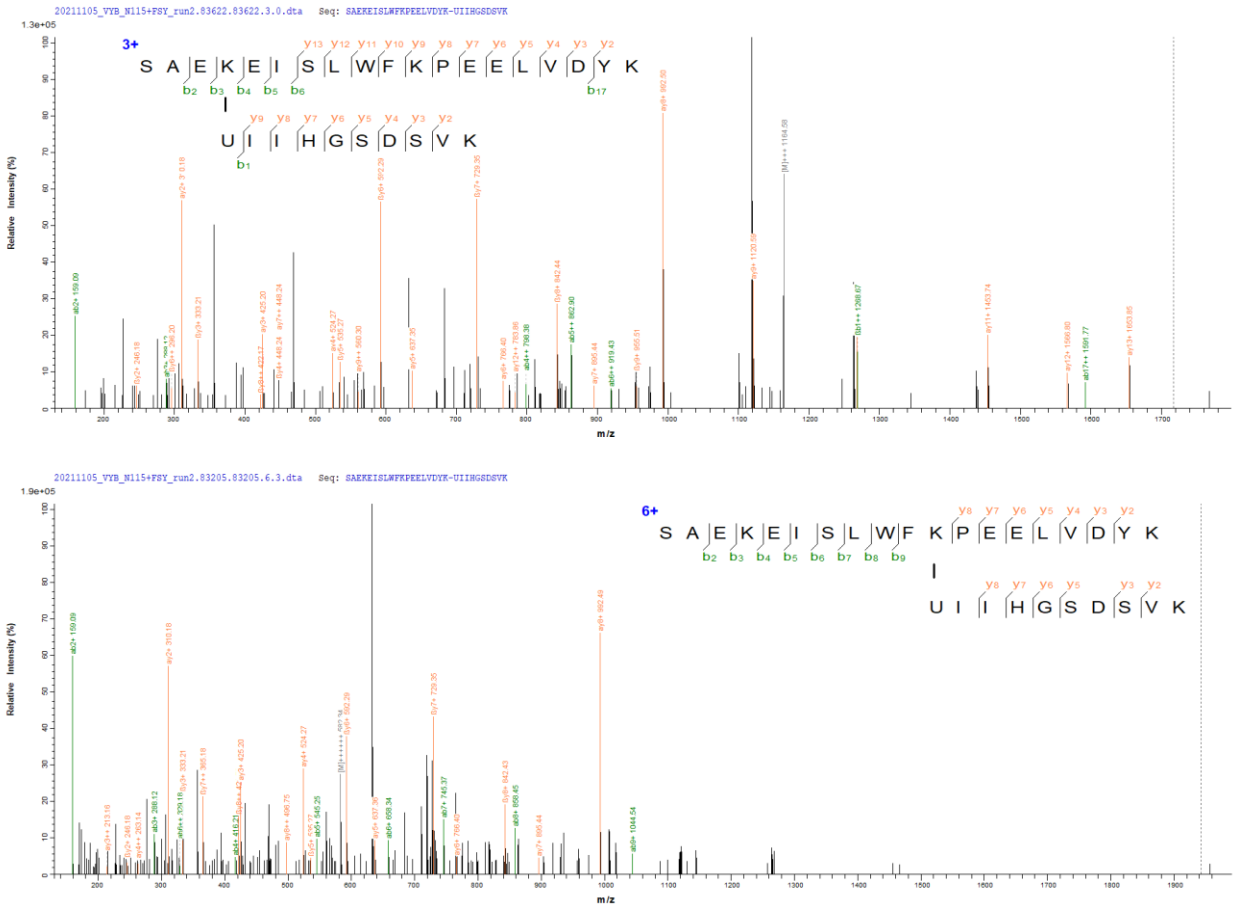


Fig S3.13: Self-crosslinking data for N115FSY mutant. Ion maps of all identified self-crosslinked peptides in NDPKB-N115FSY mutant mass spectrometry data. From top to bottom, residues crosslinked: K12; K26; K49; H51; K56; K66; Y67; K100; K128; and K135.

3.7 References

1) M.K. Tarrant, P.A. Cole. **The chemical biology of protein phosphorylation**. Annu. Rev. Biochem., 78 (2009), pp. 797–825, 10.1146/annurev.biochem.78.070907.103047

2) S.J. Humphrey, D.E. James, M. Mann. **Protein phosphorylation: a major switch mechanism for metabolic regulation**. Trends Endocrinol. Metab., 26 (2015), pp. 676–687, 10.1016/j.tem.2015.09.013

3) F. Ardito, *et al.* **The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy.** Int. J. Mol. Med., 40 (2017), pp. 271–280, 10.3892/ijmm.2017.3036

4) K. Adam, T. Hunter **Histidine kinases and the missing phosphoproteome from prokaryotes to eukaryotes.** Lab. Investig., 98 (2018), pp. 233–247, 10.1038/labinvest.2017.118

5) P.V. Attwood. **Histidine kinases from bacteria to humans**. Biochem. Soc. Trans., 41 (2013), pp. 1023–1028, 10.1042/BST20130019

6) S.R. Fuhs, T. Hunter. **pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification**. Curr. Opin. Cell Biol., 45 (2017), pp. 8–16, 10.1016/j.ceb.2016.12.010

7) I. Khan, P.S. Steeg. **The relationship of NM23 (NME) metastasis suppressor histidine phosphorylation to its nucleoside disphosphate kinase, histdine protein kinase, and motility suppression activities**. Oncotarget, 9 (2018), pp. 10185-10202, 10.18632/oncotarget.23796

8) M.-B. Gonzalez-Sanchez, *et al.* **Attempting to rewrite His tory: challenges with the analysis of histidine-phosphorylated peptides**. Biochem. Soc. Trans., 41 (2013), pp. 1089–1095, 10.1042/BST20130072

9) P.G. Besant, P.V. Attwood. **Detection and analysis of protein histidine phosphorylation**. Mol. Cell. Biochem., 329 (2009), pp. 93–106, 10.1007/s11010-009-0117-2

10) S. Moréra, *et al.* **Mechanism of phosphate transfer by nucleoside diphosphate kinase: X-ray structures of the phosphohistidine intermediate of the enzymes from drosophila and dictyostelium.** Biochemistry, 34 (1995), pp. 11062–11070, 10.1021/bi00035a011

11) Y.-T. Tee, *et al*. **Nm23-H1: A metastasis-associated gene**. Taiwanese J. Obstet. Gynecol., 45 (2006), pp. 107-113, 10.1016/S1028-4559(09)60206-0

12) T. Wieland, *et al.* **Reversible histidine phosphorylation in mammalian cells: a teeter-totter formed by nucleoside disphosphate kinase and protein histidine phosphatase 1.** Methods Enzymol.**,** 471 (2010), pp. 379–402, 10.1016/S0076-6879(10)71020-X

13) P.V. Attwood, R. Muimo. **The actions of NME1/NDPK-A and NME2/NDPK-B as protein kinases.** Lab. Investig., 98 (2018), pp. 283–290, 10.1038/labinvest.2017.125

14) S. Lutz, *et al.* **Nucleoside diphosphate kinase-mediated activation of heretotrimeric G proteins**. Methods Enzymol.**,** 390 (2004), pp. 403-418, 10.1016/S0076-6879(04)90025-0

15) S. Srivastava, *et al.* **Histidine phosphorylation of the potassium channel KCa3.1 by nucleoside diphosphate kinase B is required for activation of KCa3.1 and CD4 T cells**. Mol. Cell**,** 24 (2006), pp. 665–675, 10.1016/j.molcel.2006.11.012

16) S.K. Hindupur, *et al.* **The protein histidine phosphatase LHPP is a tumour suppressor**. Nature, 555 (2018), pp. 678–682, 10.1038/nature26140

17) J.K. Takimoto, *et al.* **Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids.**

ACS Chem. Biol., 6 (2011), pp. 733–743, 10.1021/cb200057a

18) W. Wan, J. Tharp, W. Liu. **Pyrrolysyl-tRNA Synthetase: an ordinary enzyme but an outstanding genetic code expansion tool**. Biochim. Biophys. Acta., 1844 (2014), pp. 1059–1070, 10.1016/j.bbapap.2014.03.002

19) L. Wang. **Genetically encoding new bioreactivity**. N. Biotechnol., 38 (2014), pp. 16–25, 10.1016/j.nbt.2016.10.003

20) B. Yang, *et al.* **Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions.** Nat. Commun., 8 (2017), pp. 1–10, 10.1038/s41467-017-02409-z

21) H. Tang, *et al.* **Proteomic identification of protein tyrosine phosphatase and substrate interactions in living mammalian cells by genetic encoding of irreversible enzyme inhibitors.** J. Am. Chem. Soc., 140 (2018), pp. 13253–13259, 10.1021/jacs.8b06922

22) N. Wang, *et al.* **Genetically Encoding Fluorosulfate- L -tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo**. J. Am. Chem. Soc., 140 (2018), pp. 4995–4999, 10.1021/jacs.8b01087

23) J. Dong, *et al.* **Sulfur(VI) fluoride exchange (SuFEx): Another good reaction for click chemistry.** Angew. Chem. Int. Ed., 53 (2014), pp. 9430–9448, 10.1002/anie.201309399

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Berdan, Viktoriya
40052240FC6B40C...
Author Signature

6/1/2022
Date