Claremont Colleges

Scholarship @ Claremont

2022

Here, There, and Everywhere: Probing Ubiquitin-Cavitand Binding via 15N-1H HSQC

Julia Didziulis

Recommended Citation

Here, There, and Everywhere:

Probing Ubiquitin-Cavitand Binding via ^{15}N-^{1}H HSQC

A Thesis Presented

by

Julia Didziulis

To the Keck Science Department

Of Claremont McKenna, Pitzer, and Scripps Colleges

In partial fulfillment of

The degree of Bachelor of Arts

Senior Thesis in Chemistry

December 13, 2021

Table of Contents

1   Abstract

Ubiquitination and other post-translational modifications play a major role in the proliferation of many neurodegenerative diseases, developmental disorders, and cancers, and as such are subjects of recently increased biochemical interest. Expanding upon such research, this study confirmed a robust method of ubiquitin expression and purification, then used ^{15}N-^{1}H HSQC to analyze ubiquitin samples containing varying concentrations of a deep cavitand with affinity for lysine and arginine side chains. In the pure ubiquitin spectrum, arginine side chain chemical shifts were detected in the nitrogen 80-85 ppm range and lysine side chain signals were undetected. However, upon introduction of the cavitand, collection of chemical shift perturbation data was prohibited by the cavitand precipitating out of solution. To overcome this obstacle, future experiments may benefit by refining the purification protocol via size-exclusion chromatography or adjusting sample properties such as pH, salt content, and more. Once complete, these experiments may indicate the cavitand's viability as a biosensor for ubiquitination and other modifications, possibly accelerating diagnostics and disease research as a result.

2 Abbreviation Glossary

APB – activity-based probe

APC/C – anaphase promoting complex

BSA – bovine serum albumin

COSY – correlation spectroscopy

DUB – deubiquitinase

FID – free induction decay

FRET – fluorescence resonance energy transfer

HECT – Homologous to E6-AP Carboxyl Terminus

HMQC – heteronuclear multiple quantum coherence

HSQC – heteronuclear single quantum coherence

IBR – In-Between-RING

LBD – ligand binding domain

NMR – nuclear magnetic resonance

PCNA – proliferating cell nuclear antigen

PTM – post-translational modification

RBR – RING-between-RING

RING – Really Interesting New Gene

SUMO – small ubiquitin-like modifier

SW – spectral width

TEV – tobacco etch virus

TUBE – tandem ubiquitin binding entity

UB – ubiquitin

UBD – ubiquitin-binding domain

UBL – ubiquitin-like

UIM – ubiquitin-interacting motif

UPS – ubiquitin-proteasome system

3   Introduction

3.1   Post-translational Modifications

The term 'post-translational modification' (PTM) encompasses a vast array of protein alterations, including covalent side chain additions of small chemical motifs or entire proteins, autocleavage, peptide rearrangement, and even proteolysis.[1–3] Given the limited combinatorial diversity of proteins comprised of just twenty amino acids, PTMs effectively expand the complexity of the proteome by several orders of magnitude beyond the coding capabilities of the genome.[4] Such modifications, especially covalent additions, can be executed and reversed relatively quickly, and significantly faster than the time scale of protein transcription. This environmental sensitivity establishes PTMs as dynamic tools which, when applied and managed properly, enable rapid proteomic responses to the evolving needs of a biochemical system.

Side chain modifications often regulate biochemical pathways by inducing conformational changes, altering surface charge, or blocking key residues on a given protein – these effects tend to dramatically impact enzyme function and molecular recognition in order to provide a timely response to altered cellular needs or stimuli.[5,6] For example, phosphorylation is a common PTM characterized by the addition of a phosphate group to an amino acid side chain, typically serine. The steric bulk and strongly negative charge imparted by this addition is often used to induce dramatic changes in its substrates, and phosphorylation is utilized to modulate enzyme activity in countless key biological pathways, including glycolysis. Beyond phosphorylation, other common PTMs include acetylation, carbonylation, glycosylation, hydroxylation, methylation, nitration, SUMOylation, sulfation, and ubiquitination.[7] The final of these modifications involves the substrate ligation of ubiquitin, an entire protein, as a complex code of chains whose varying length and linkage patterns denote specific functional and regulatory signals.[8]

Most PTMs require a large number of proteins to oversee their addition and removal from substrates; while this regulatory feat is responsible for the importance of protein modifications in many major biochemical pathways, it significantly complicates their study. With so many complex modification pathways and substrates in existence, PTM dysfunction is inevitable and has therefore been implicated in a number of disorders, ranging from cancer to neurodegeneration.[4] Such discoveries have piqued interest in PTMs, leading to increased efforts to understand the still-obscure breadth of modification pathways and target them for new therapies. In addition to their complex regulation, the transience of PTMs that proves so biologically useful also increases the difficulty of their detection and monitoring. Modifications are unpredictable and relatively short-lived in the biochemical system – as such, there remains constant effort in the field to develop improved PTM-monitoring probes with the ability to extract maximum information while keeping up with the pace of real-time biological processes.[3] It is only with these observation methods that research will potentially improve our understanding of post-translational modification pathways and effectively target them for therapy and prevention of disease.

This study focuses specifically on ubiquitin, a protein modification motif with enormously complex implications; the product of complex execution and extensive regulation, ubiquitination remains relatively misunderstood despite its significance in many major cellular pathways and diseases.[9] Here, the interactions of ubiquitin with a synthetic deep cavitand molecule are observed via ^{15}N-^{1}H HSQC. The particular cavitand being used herein is the negative cavitand, described as such due to the four carboxylates lining its rim. These deep cavitands, synthesized in the Hooley lab at UC Riverside, have already been employed as fluorescent biosensors to monitor other PTMs in real time, such as lysine methylation in histones.[10–13] Given this precedence, it is believed that these cavitands may also serve effectively as probes for ubiquitination or other PTMs – ultimately,

such a development could aid current understandings of PTM functions while providing a simple means of targeting ubiquitination in disease therapy and prevention.

3.2 Ubiquitin and the Ubiquitination Machinery

Ubiquitin (UB) is a small protein with such immense biological consequence that it is represented in every eukaryotic cell – from fruit flies to humans, ubiquitin is consistently expressed and almost perfectly conserved across all species.[14] Encoded along with the additional ubiquitin machinery by four different genes (UBB, UBC, RPS27, UBA52) which account for several percent of the human genome, ubiquitin is synthesized as a chain of ubiquitin moieties which is then broken down for ligation of substrates.[9,14,15] By expanding ubiquitin coding across multiple genes and expression types, biological systems enable sensitive maintenance of ubiquitin levels to suit a variety of cellular conditions. With primary use as a motif for post-translational protein modification, this cellular omnipresence of ubiquitin demonstrates the key role of ubiquitination in countless processes; as such, malfunctions in UB-writing, recognizing, and erasing enzymes may engender serious disruptions and cellular stresses which elevate risks of cancer, neurodegenerative disease, or developmental disorders.[16–19]
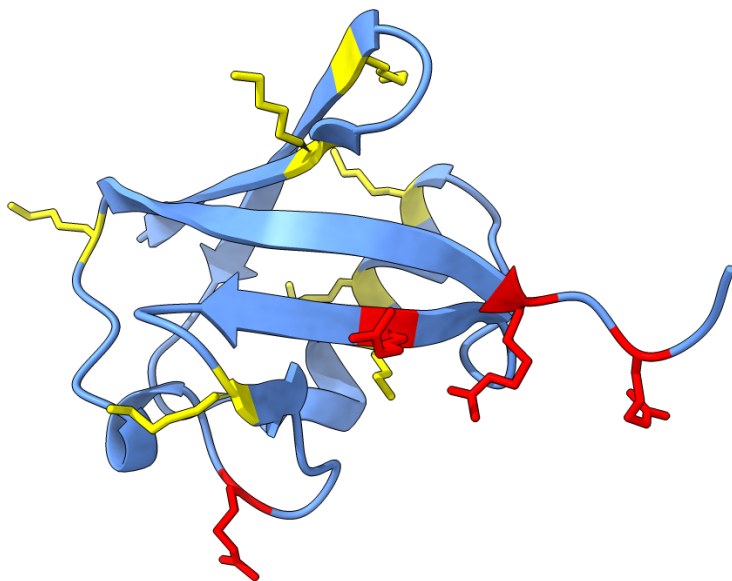

Figure 1: Protein structure of ubiquitin. Ubiquitin's seven lysine residues (K4, K11, K27, K29, K33, K48, K63) are highlighted in yellow, and the four arginines (R42, R54, R72, R74) in red.[22]

Compact and supremely kinetically stable, the 76 residues that make up the UB protein adopt a well-characterized ß-grasp fold structure that manifests as a five-strand ß-sheet appearing to 'grasp' its single alpha-helical segment; this structure appears in a number of ubiquitin-like proteins as well, and is known to demonstrate a striking range of signaling and recognition properties across the proteome (Figure 1).[8,20] Ubiquitin also contains a six-residue, glycine-rich tail at its C-terminus – this flexible tail is key in many ubiquitin-recognizing interactions, though it is just one of the many surfaces targeted by the current library of structurally and functionally diverse ubiquitin binding domains (UBDs).[21,22]

Despite the simplicity of a single ubiquitin moiety, the entire ubiquitin code is enormously complex and remains the subject of continued study. Hundreds of proteins have been found to attach, recognize, and edit substrate-bound ubiquitin, and a variety of different ubiquitination types exist as well. For instance, ubiquitination can occur with an individual moiety in monoubiquitination, multiple individual moieties in multimonoubiquitination, or as a chain in polyubiquitination. Polyubiquitination is quite varied in itself: chains can range from two to ten or more moieties in length, with increasing complexity in topology, linkage, and branching as they grow (Figure 2).[23,24] Each of these seemingly minor details is consequential to the ultimate role of the modification in enzyme function and signaling, further underscoring the need for continued study of ubiquitination and development of sensitive ubiquitination probes. The vast library of enzymes involved in the execution of ubiquitin coding can afford few malfunctions – even slight aberrations in the editing or recognition of ubiquitin chains can disrupt vital processes, as each of the countless possible ubiquitin modifications encodes a unique, biologically significant message.
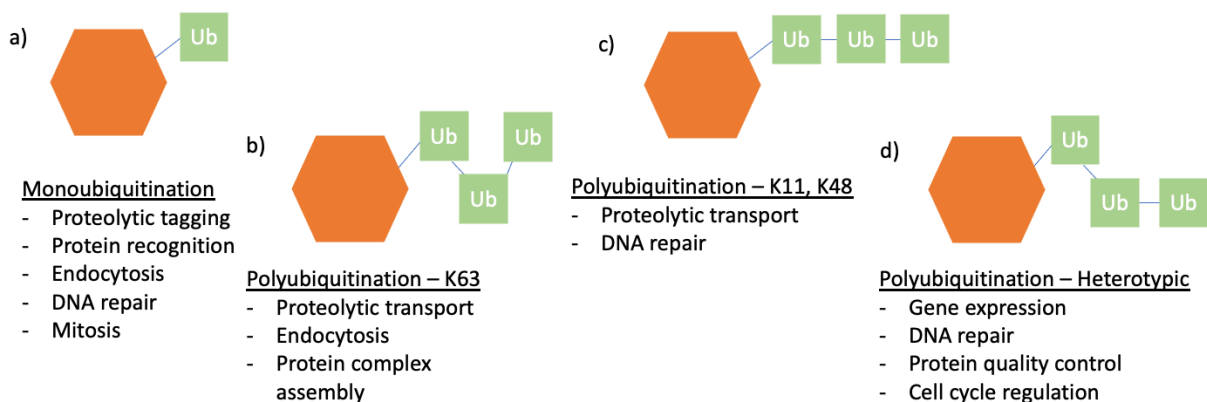
Figure 2: Visualization of various ubiquitin modifications and their functions, including a) monoubiquitination. Monoubiquitination appears as a linkage between a substrate lysine and the ubiquitin C-terminus, and is used to tag proteins for degradation, alter protein recognition patterns, and regulate endocytosis, the cell cycle, and more. Monoubiquitin modifications also signal errors in DNA for recognition and repair by DNA polymerases. Polyubiquitinations also regulate many of these same pathways, though each with their distinct functions. For instance, b) K63 chains, c) K48 chains, K11 chains, and more all act as transport systems for proteolytic pathways in addition to their control of protein complex assembly, mitosis, and signaling of error-free DNA. These modifications, as well as d) heterotypic chains, also play major roles in apoptosis and gene expression.[9]

Ubiquitination begins with a single moiety that is bonded to a substrate protein; this initial attachment occurs between a lysine on the substrate and the C-terminus of the first ubiquitin. In turn, the chain is then extended as needed via attachment of additional ubiquitin C-termini to one of the seven lysines on the target ubiquitin substrate. Some attachments have also been observed at the N-terminus of a ubiquitin substrate, a methionine.[8] These chains may demonstrate consistent linkage types, one example being attachment of each moiety to the K27 residue of the ubiquitin before it. These chains are thus defined as homotypic, while heterotypic chains demonstrate a mixture of linkage types. Depending on the electrostatic and dynamic features imparted by this linkage pattern, ubiquitin chains will adopt a variety of conformations that have been observed in numerous ^{19}F NMR experiments: some are dynamic and flexible, others pack into extremely

compact shapes, and more still are known to adopt multiple conformations in the chain or the substrate protein itself (Figure 3). [21,25,26] Such unique topologies are, of course, critical to the ubiquitin code, with their shapes and surface characteristics acting as the primary features sensed by UBDs.
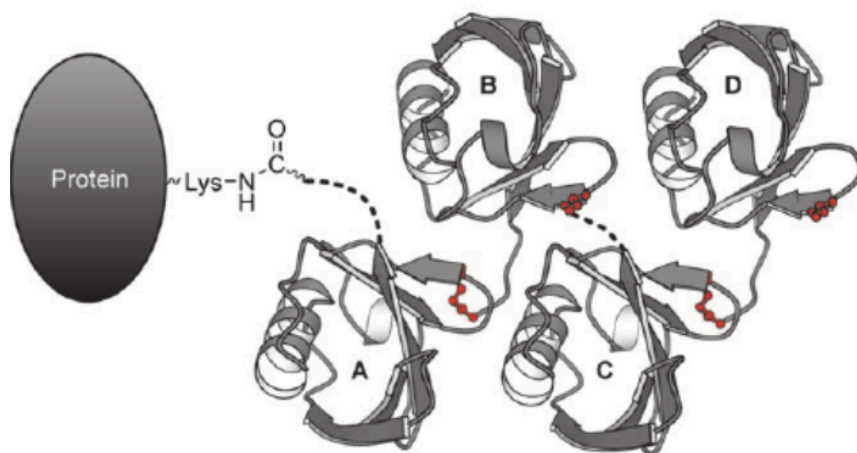

Figure 3: Rendering of a ubiquitin chain. This linkage pattern causes the chain to adopt a coiled topology which can be bound by substrates containing ubiquitin-binding domains. UB-readers which successfully bind and 'read' this chain may play a role in such processes as proteolysis, DNA repair, and cell cycle regulation.[26]

While other PTMs are often executed in one step, ubiquitination occurs through a cascade of three enzymes – the vast library of these ubiquitin writers, referred to as E1, E2, and E3s, add another level of complexity to the ubiquitin code (and inevitably increase opportunities for its malfunction as well). In any one ubiquitination event, the exact enzymes involved are highly consequential in indicating the type of modification being performed, whether it be a simple monoubiquitination or the construction of a complex, lengthy chain.[27] Of these ubiquitin writers, E1s are the least diverse; only two have been identified in humans, with most ubiquitin activation performed by the UBA1 enzyme. In the first step of the cascade, one of the E1s uses energetic input from ATP hydrolysis to activate ubiquitin and form a thioester bond between the ubiquitin C-terminus and the E1 catalytic cysteine.[8] This high-energy ubiquitin intermediate is then

transferred to a cysteine on the E2 conjugator before attachment to the target substrate by an E3 ligase mediator (Figure 4). These two writers are much more varied – there about 40 E2 enzymes in humans, while analyses estimate that over 600 E3 ligases are encoded by the genome.[9,27] Such enzymatic diversity exists to in order to fully encompass the wide range of ubiquitination functions and targets, though this complexity significantly prolonged ubiquitination's obscurity and complicated its study as a major regulator of cellular homeostasis.[7]
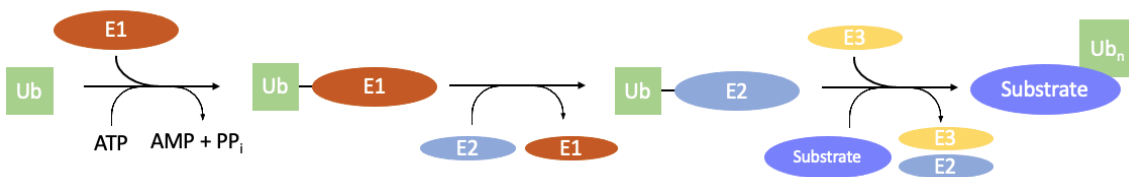

Figure 4: Mechanism of ubiquitination. The first step requires ATP to activate ubiquitin by creating a high energy ubiquitin-E1 thioester. The protein is then conjugated onto the E2 to form another ubiquitin-enzyme complex. In the final step of the cascade, E3 binds and positions the desired substrate and transfers ubiquitin from the UB-E2 complex to a lysine on the substrate. This final step may require either one or two steps depending on the classification of the E3.[9]

The final enzyme in the ubiquitin cascade, the E3, plays the largest role in controlling the unique biological implications ascribed to each possible permutation of chain lengths, linkages, branching patterns, and substrates. This diversity is expected of such a complex protein modification code – some E3s target only one substrate via highly specific recognition pathways, others bind and ubiquitinate all proteins containing a given sequence motif, and others are simply recruited to certain areas of the cell to ubiquitinate any protein they encounter.[27] Much of this variation depends on the cellular pathway being targeted for regulation, as well as the actual ubiquitination pattern required. For example, monoubiquitination and multimonoubiquitination are executed by chain-initiation E3s, which must be able to properly bind and orient target proteins for modification at a specific lysine residue. In contrast, polyubiquitination is made possible by chain-elongation E3s, which demonstrate the necessary flexibility to negotiate and conjugate a

growing substrate chain.[9,27] Still, despite wide variations in their method of ubiquitin transfer, all known human E3s are divided into just three general classes: RING (Really Interesting New Gene), HECT (Homologous to E6-AP Carboxyl Terminus), and RBR (RING-between-RING) E3s.[8] Each of these classes is further subdivided into as many as seven families, often distinguished based on their employment of distinct ubiquitination domains and mechanisms.

With a library of around 600 enzymes, the RING class encompasses the vast majority of known E3s. Despite their great diversity in size and target, RING E3s are consistent in their use of a characteristic RING or U-box domain to recruit and activate E2-linked ubiquitin for attachment to the target substrate. These ligases are also distinguished by their use of a simple one-step mechanism, in which ubiquitin is directly transferred to its target after cleavage from the E2 conjugator.[27] However, despite their overwhelming role in controlling the majority of ubiquitination events, much still remains to be discovered about the exact mechanisms and substrates of RING E3s. Recent studies in this area indicate that proximity to a RING E3 may induce allosteric or conformational changes in certain E2-ubiquitin complexes, further encouraging their reactivity and likelihood of attack by the ligase.[27] In such cases, the E2-ubiquitin complex remains relatively stable until its C-terminus is properly positioned near the correct E3 to allow for successful substrate conjugation. In general, though, this large class of E3s remains a subject of continued study, especially due to its significant role in the specificity of a major post-translational modification.

The two other classes of E3 are significantly smaller (the HECT class contains about 30 enzymes, the RBR class only 12) and similarly obscure, though they are known to utilize a different process of ubiquitin transfer than RING E3s. Both HECT and RBR E3s employ a two-step mechanism in which they first covalently recruit E2-linked ubiquitin via a covalent cysteine

linkage, then subsequently transfer ubiquitin to the substrate. In order to do so, the HECT E3s employ a characteristic HECT domain while RBRs contain two distinct RING1 and RING2 domains separated by an In-Between-RING (IBR) domain.[27,28] Still, regardless of its categorization within these classes, each E3 remains extremely unique in its conformational and surface complementarity requirements used to target each substrate lysine with high specificity. Even further, while the role of E3s in mediating ubiquitination cannot be overstated, the function of E2s is still significant: despite being notably less varied and containing a conserved core, these ubiquitin-conjugators are known to define the actual biological outcome of any one ubiquitination event, while E3s simply collaborate by recruiting and orienting the necessary substrate to achieve this desired outcome.[29] Working in tandem, the E2-E3 module ubiquitinates a wide variety of substrate enzymes, binding and positioning target lysines for highly specific chain elongation.

Given the specificity in UB-writer functionality required to precisely ligate countless substrates, these enzymes are subject to multiple levels of close regulation. The most common form of E3 regulation is simply activity control – when not actively binding or ligating a substrate, E3s most always occupy inactive, autoinhibited conformations.[27,28] In this state, the enzyme is unable to bind its E2-UB substrate due to a buried recruitment motif or additional steric hindrance, only to be activated when necessary by a small binding peptide or adaptor protein. Many ubiquitinating enzymes are also subject to varying post-translational modifications and allostery of their own, all of which aid in fine-tuning ubiquitination and ubiquitin homeostasis within the cell.[7] For example, the autoinhibited RBR E3 named PARKIN undergoes a conformational change upon phosphorylation at Ser65, which in turn exposes the vital RING1 domain and positions it closely to its RING2 stretch. Even more, the same kinase that phosphorylates PARKIN also phosphorylates the substrate's ubiquitin chain to accelerate ligation of the two – with this level of

specificity available, PARKIN effectively monitors mitochondrial activity and induces their repair via mitophagy to ultimately maintain cellular homeostasis.[30] PARKIN is just one of nearly 600 E3s subject to these complex regulatory systems – additional forms of enzymatic oversight include allostery, molecular interactions, signaling, and additional post-translational modifications. Many E3s have even been found to be regulated via ubiquitination itself, the very modification they are designed to execute.[31] Though seemingly redundant or cyclical, this discovery perfectly encapsulates the importance of ubiquitination in biological systems: a regulatory network as complex and intricate as the ubiquitin code requires constant precision and variety, which is afforded by ubiquitination far more than many other regulatory mechanisms.

Perhaps just as important as ubiquitinating enzymes are its erasers and editors, known as deubiquitinases (DUBs). This species, of which 79 are known, take on the momentous task of completing or correcting signaling pathways and maintaining cellular ubiquitin levels.[9,32] The deubiquitinases are divided into five functional families (C-terminal hydrolases, UB-specific proteases, ovarian tumor proteases, Josephins, and metalloenzymes), but most share similar structural features which include ubiquitin-like (UBL) folds, as well as multiple ubiquitin binding domains (UBDs) and interacting motifs (UIMs).[32] Regardless of their mechanistic features, all DUBs generally fall into three main functional roles; these include generation of free cellular ubiquitin, removal from enzymes at the end of a signaling cascade, and editing of incorrect or defunct ubiquitin chains. This first function is vital to the maintenance of cellular homeostasis – ubiquitin is synthesized in long, free chains which need to be broken down before attachment to substrates.[31] Without deubiquitinases constantly digesting these chains and producing free ubiquitin, vital signaling pathways and cell cycle processes would halt and throw the cell into damaging ubiquitin stress. On top of this, DUBs also complete regulatory cycles by modifying or

completely amputating obsolete ubiquitin chains from their substrates.[8,31,32] This function highlights the significance of DUBs in maintaining proper cellular environments: with improper ubiquitination implicated in many human diseases and disorders, the enzymes designed to fix these errors and complete the modification cycle have emerged as fruitful research subjects and possible therapeutic targets.[18]

Regardless of mechanistic classification, most DUBs share a unique binding method in comparison to other UB-reading enzymes. As the sole editors of ubiquitin coding, they have been found to utilize extremely precise techniques, interacting with 20-40% of the UB surface where other UB readers only cover around 10%.[32] Additionally, DUBs are able to distinguish ubiquitin from other ubiquitin-like species, such as SUMO, by recognizing its unique C-terminal sequence (LRLRGG). This motif is key to the function of ubiquitination, as it allows for robust and highly precise control of the editing of ubiquitin chains – SUMO (Small Ubiquitin-like Modifier), a ubiquitin-like protein which is employed in a parallel post-translational modification called SUMOylation, lacks any arginine residues near its C-terminus in a feature which bolsters the ability of DUBs to distinguish these species.[1] Even more significant, deubiquitinases themselves have been characterized as quite ubiquitin-like in their structures and sequences; in order to protect these enzymes from instead binding one another, they lack ubiquitin's characteristic double glycine terminus and therefore deter binding by other ubiquitin machinery. These details highlight the importance of certain residues in keeping the ubiquitin code functioning properly, some of the most notable being the lysines involved in various chains and the arginines required for precise ubiquitin recognition by UDBs.

Even amongst DUBs, recognition varies based on the specific functions required of each enzyme: for instance, some DUBs are considered 'promiscuous,' relying on more general

recognition patterns in order to edit a variety of chain types on many different substrate enzymes.[11] In contrast, other deubiquitinases carry out highly distinct functions, editing only certain linkage types or modifying specific substrate enzymes. For instance, the DUBs which remove monoubiquitinations are often highly specific due to the binding events required of them: while polyubiquitin chains can often be bound by the same enzyme regardless of the substrate to which they are linked, monoubiquitin DUBs must bind both to the UB modification and the attached substrate protein. As a result, these DUBs contain domains designed to recognize the substrate with specificity, therefore disallowing its modification of other enzyme substrates.[32] As for the deubiquitinases which edit polyubiquitin chains, these enzymes can be defined as either endo- or exo-DUBs: endo-DUBs cleave chains internally, while exo-deubiquitinases edit solely at the at the distal end of such chains.[32] This extensive network of enzymes, designed uniquely for the editing of ubiquitin signaling, also requires stringent specificity – there are entire biological pathways which rely on ubiquitination for the majority of their oversight, so UB editors are subject to extensive regulation in the form of allostery, cellular localization, and post-translational modifications (including ubiquitination itself).[7] Despite this highly focused oversight, DUBs remain quite vulnerable to aberration, and, given their major role in fine-tuning the UB code, such malfunctions reflect significantly in disease and development pathways.[9]

Despite their functional diversity, this vast universe of ubiquitin writers, editors, and erasers all share a common identity as ubiquitin 'readers.' In order to bind and recognize the ubiquitin protein, all of these species must contain one or more domains characterized as ubiquitin-binding domains (UBDs). There exists great diversity amongst these domains based on their recognition mechanism and function; this is unsurprising, as NMR studies have characterized the ubiquitin surface as highly diverse, with its conformation varying widely across different chain

linkages, environmental contexts, and binding partners.[21] For instance, homotypic Lys48 chains are known to exist in closely packed conformations, while Lys63 chains exist in a purely extended form with no surface contact between ubiquitin units.[11,21] These unique topological patterns call for a library of distinct UBDs to exist at the disposal of the ubiquitin machinery and carry out the PTM's biological functions – with this, researchers have identified twenty distinct classes of UBD types.[33] Across these groups, it is clear that UBDs quite often adopt alpha-helical conformations and bind near the UB Ile44 residue, though the causes for such details are ill-understood given historical difficulties in identifying and predicting ubiquitin-binding domains.[21]

Of primary importance to the function of UBD-containing enzymes is the nature of the binding event: with the exception of E3 ligases which bind the E2-UB conjugate, enzyme activity studies have deduced that only a single entity may bind to a ubiquitin chain at a time. This consequence follows from the need for UBDs to cover a comparatively high percentage of the ubiquitin surface, especially when all possible binders are in competition to bind to adjacent regions on the surface of ubiquitin.[21] Since only a single enzyme may bind the modification at a time, incorrect signaling is prevented by negating the stimulation of multiple pathways, especially if the chain is not yet complete. Additionally, as the complexity and length of certain ubiquitin chains increases, their corresponding readers must contain UBDs more numerous and complex.[33] For modifications as intricate as to contain branching and heterotypic linkages, reader enzymes must bind with the aid of multiple UBDs covering the chains' surface with numerous ubiquitin-interacting motifs (UIMs), all spatially oriented to interpret the biological code contained in their length and flexibility.[11] The ceiling of this great complexity has still not been identified, as new domains are constantly being identified in connection with various UB-based pathways (i.e. double-sided UBDs), so this area remains a fruitful region of biochemical research.[21] Furthermore,

the prevalence of these domains across the ubiquitination machinery demonstrates their functional importance, which lies at the crux of many essential biological pathways in eukaryotic life.

3.3 Regulation of Ubiquitin

Many intricacies of ubiquitination are further encapsulated by the extensive crosstalk existing between ubiquitin and various other post-translational modifications; examples of this regulatory communication include promotion or suppression of another PTM, overlapping functional cascades, and multi-level modifications.[34] While such communication exists with most major pathways (for example, acetylation of ubiquitin itself has been found to act as a 'cap' that prevents further chain elongation), ubiquitination is most closely linked to phosphorylation and SUMOylation pathways.[15] For instance, phosphorylation acts as the primary regulator of E3 ligase activity, regardless of their structural and mechanistic diversity.[9,14,27] Likewise, many kinases (phosphorylating enzymes) are directed by ubiquitination events which often serve to downregulate persistently active kinases.[35] Given the major role of protein kinases in vital processes like cell respiration, ubiquitination occupies a similarly important role, not only in its own pathways but in proximity to phosphorylation as well.

Alternatively, SUMOylation is a modification which acts somewhat antiparallel to ubiquitination, as SUMO is a similarly compact protein used to modify enzymes like ubiquitin. Through NMR studies, SUMO has been found to adopt an analogous structure to ubiquitin with 18% sequence consistency, and is even ligated using a three-enzyme cascade as well (Figure 5).[34] However, this cascade is mechanistically distinct from ubiquitination, and in contrast to the diverse library of ubiquitin E3s, few SUMO ligases are known to exist. This mechanistic dichotomy underscores the distinction between the two regulatory proteins: despite similar structures, their electronic surfaces differ greatly, along with a few key residues.[32,34] For instance, the previously

discussed C-terminal arginines on ubiquitin, especially the Arg74 crucial to UBD recognition, are notably absent in SUMO – these features and more underscore key divergences in both modification pathways and their 'readers.'[10] Moreover, SUMOylations exist primarily as monomers, a feature further distinguishing them from ubiquitin signaling. Nevertheless, the similarities between these modification pathways is indicative of their close association – ubiquitin and SUMO are implemented in many of the same biological circumstances, yet often with contradictory effects.
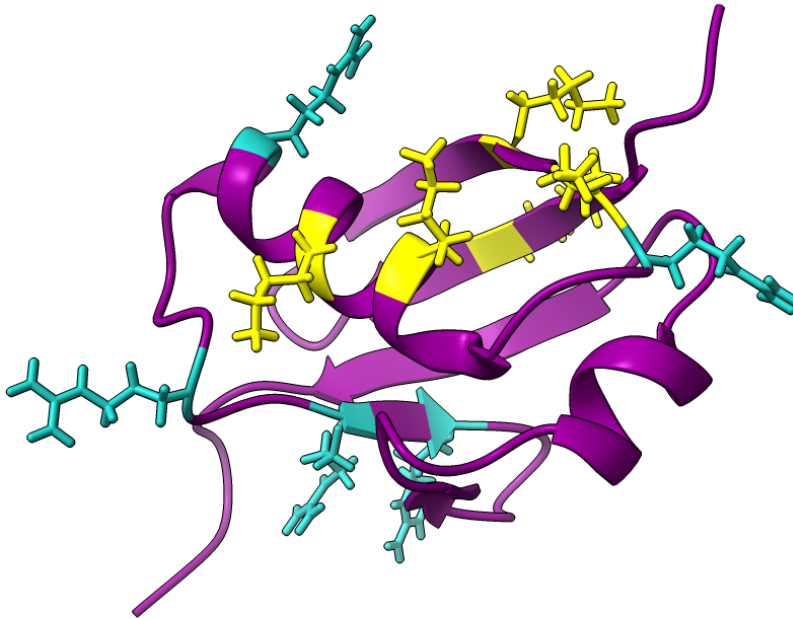

Figure 5: Protein structure of SUMO (Small Ubiquitin-like Modifier). Lysine residues are shown in yellow, and arginines in blue. SUMO and ubiquitin are both utilized in post-translational modifications, with opposite biological effects. Notably, SUMO lacks two arginines near its C-terminus that are present in the ubiquitin C-terminus. This site therefore plays an important role in ubiquitin recognition, so biosensors with the ability to distinguish UB and SUMO termini are necessary for continued research of these pathways.[22]

Functionally, ubiquitin and SUMO modifications both appear in such processes as cell cycle progression, intracellular signaling, DNA repair, and protein localization.[34] One example of these intertwined pathways lies in their regulation of gene expression, in which both ubiquitin and SUMO have been found to alter chromatin structure through modification of histones. However, these parallel modifications appear to manifest in opposing ways: ubiquitination of histones increases gene expression, while

SUMOylation suppresses it.[14] This antithetical mode of regulation, or negative crosstalk, is consistent throughout much of their shared biological functions, and as a result, SUMOylation increases the complexity of UB signaling and amplifies its effects, whether intended or not.[35] Still, despite SUMO's comparable simplicity to ubiquitin, this modification has been more extensively studied, possibly as a result of its characterization as the fastest evolving post-translational modification.[36]

3.4　　Biological Roles of Ubiquitination

Independent of its shared functions with SUMO, ubiquitin fulfills a tremendous, ever-increasing number of biological roles. The function typically ascribed to ubiquitination is mediation of proteolysis, and for good reason: monoubiquitination tags and sorts enzymes destined for degradation (often dysfunctional or misfolded), while multiple polyubiquitinations allow for these species to be transported and bound by their respective proteases at the 26S proteasome.[14] However, ubiquitin coding contributes to countless other processes. For instance, another major pathway requiring multiple forms of UB modification is DNA repair: here, the proliferating cell nuclear antigen (PCNA), which aids DNA polymerase function, uses monoubiquitination to signal errors in the genetic code and K164 polyubiquitination to indicate error-free sections of DNA.[37] Additionally, monoubiquitin also adopts unique roles in the regulation of protein activity, DNA repair, meiosis, transcription, proteolysis, endocytosis, and more – this modification also plays a major part in the cell cycle, as it directs progression from mitosis to anaphase via the anaphase promoting complex (APC/C), an E3 ligase.[14,15,21,33]

In further regulation of the cell cycle, ubiquitination is known to trigger apoptosis via modification (and activation) of necrosis binding factors and caspases.[29] And because the APC/C also determines cell fate, ubiquitination is a key regulator of tissue specialization, brain

development, and early cell communication in organ growth.[9] In fact, the first known sample of a ubiquitinated protein was a monoubiquitinated histone, in which the UB tag was found to promote gene expression via transcription by the NF-κB transcription factor.[3,14,38] Overall, monoubiquitinations exists as the most common form of ubiquitination, pervading the cell in their oversight of its constant growth and maintenance.

Polyubiquitination, given the wide variety of unique chains that exists, is much more varied and biologically obscure in its functions. Like monoubiquitination, these chains are vital to protein degradation at the 26S proteasome, where K11, K48, K29, and K63 homotypic chains are all utilized for distinct purposes in the tagging of substrates at various proteolytic stages.[16,39,40] The complexity grows from there: homotypic linkage via K63 or M1 regulates endocytosis and the assembly of protein complexes, while heterotypic chains of mixed M1/K63 linkage are vital to the expression of NF-κBs, a family of transcription factors with a vital role in immune response.[9] About 10-20% of ubiquitin chains also contain branching, forming a subset of heterotypic modifications that primarily regulate some of the major functions already discussed like proteolysis and NF-κB activation.[23,24] Additionally, some miscellaneous pathways involving ubiquitination include protein synthesis and localization (not just degradation), cellular communication and stress responses, and maintenance of circadian rhythms.[15,41] Overall, the known functions of ubiquitination are numerous and ever-expanding; with further study and more advanced techniques, even more biological roles are sure to be deduced.

3.5   Ubiquitin in Disease

With all of its efforts to maintain cellular homeostasis, ubiquitination occupies a vital position in maintaining the well-being of the organism, human or otherwise. In humans specifically, aberrant ubiquitination has come to light as a cause or symptom in major disease

pathways, including cancers, autoimmune and developmental disorders, and neurodegenerative, pathogenic, and even oral disease.[9,17–19,42] Many of these consequences become obvious when one considers the constant necessity of ubiquitination in the cell: central to cell cycle progression at multiple stages, this modification oversees early development and specialization, mitosis, and apoptosis and proteolysis as well.[29] When any one of these processes falters in its function, disease is therefore likely to take hold.

The enormous complexity of the ubiquitin code and associated machinery further magnifies its contribution to disease pathways, for aberration in even just one step has been shown to manifest poorly in the cell and organism as a whole. These ubiquitination steps typically include maintenance of proper UB levels, the three-enzyme cascade (often repeated for UB chains), chain editing, chain reading and functional output, and removal of the modification. Each of these events has been linked to disease in its own way – for instance, improper maintenance of available cellular UB (a responsibility of DUBs) is known to cause adult-onset neurodegeneration, while malfunctioning E2s are connected to improper mitosis, a driver of developmental issues.[5,9,14] Pathogens are also known to target or even pose as components of the UB system, taking advantage of its important role in order to gain control of cellular processes.[15] As such, ubiquitination remains a vital research subject, both for therapeutic targeting and scientific advancement.

Cancer and ubiquitin are particularly intertwined, as cancer itself simply involves loss of cell cycle control. Tumor cells may arise from a number of errors, but some of their major features include rapid and uncontrolled mitosis and the overriding of extracellular apoptosis signals. As such, E3 ligases stand just behind protein kinases as the second most prevalent gene family involved in cancer, with DUBs acting as a major contributor as well.[18] Dysfunction, over- or underexpression, or mutation of these actors often proves deleterious, especially since UB is used

almost exclusively to regulate the localization and function of p53, the foremost tumor suppressor gene.[19] The role of UB in developing and sustaining cancerous cells is therefore especially notable to ongoing cancer research, and thus many enzymes in the UB machinery are current therapeutic targets.

In contrast, the role played by ubiquitin in neurodegeneration is less tied to its roles in the cell cycle, and instead often follows from dysfunctional proteolysis. While eventual degradation is the fate of all proteins, some suffer from poor function or misfolding – these features may render a protein completely inhibited or severely debilitated beyond the help of a chaperone, at which point it will be tagged with a ubiquitin chain and incorporated into the proteasome. However, if such aberrations manifest in any of the ubiquitin readers, writers, or editors required for this process, misfolded proteins are likely to build up as plaques in the cell; this protein aggregation, especially in brain tissue, is strongly associated with neurodegenerative diseases like Alzheimer's, Parkinson's, and Huntington's.[14,42] Additionally, ubiquitin's effect on apoptosis is also known to play a role in such decline – while deficient apoptosis signals cancer pathways, uncontrolled cell death (especially in the brain) can lead to neurodegeneration and other conditions, such as cancers and developmental disorders.[29] With such a role in proliferating many widespread diseases, ubiquitin is therefore gaining attention as a significant component of the healthy eukaryotic cell.

3.6   PTM Biosensors

The significance of ubiquitin in the context of a living system necessitates further attention than is currently being paid. However, research on ubiquitination and other PTMs has historically been limited by the transience of the subject matter: PTMs are short-lived by design, so productive research on these enzymatic features requires advanced biosensing tools to probe samples in real time.[15] A number have already been implemented successfully and thus advanced scientific

understanding on the world of PTMs, but much still clearly remains to be discovered on their extent, function, and overall significance.

One of the main proteomic techniques used to study PTMs in the past has been mass spectrometry (MS); this method can be quite reliable for identifying and studying modified proteins and has aided greatly in expanding ubiquitin research.[43] However, the use of MS alone places constraints on the systems that can be analyzed and the information that can be gleaned. For instance, MS experiments require stringent sample preparation, and therefore severely slow progress while negating any possible research on more complex samples.[15] These samples often do not mimic the biological systems and time scale to which they are most relevant, so new methods are required which may withstand more natural sample conditions while still providing valuable data and diagnostic power.[3]

Supplementing MS, biosensors of all kinds have come to light as extremely valuable tools in a number of settings – for application to PTMs, they are most useful as advanced research methods and diagnostic tools.[44] These probes are often constructed by linking some ligand-binding domain (LBD) to a detectable protein (oftentimes this protein detection occurs through some transcription activation or fluorescence).[45] Specifically, fluorescent biosensors have practically revolutionized the study of post-translational modifications, as they are largely able to integrate into a number of samples and provide nearly instantaneous information. Typically, they operate by binding via the LBD to the desired modification, therefore inducing a conformational change and subsequently fluorescing – the dynamics of the targeted modification then become detectable via fluorescence resonance energy transfer spectroscopy (FRET).[3] Using these techniques, we have gained valuable insight into intracellular PTM execution and regulation, but there is more still to be found. Biosensors are under constant development, and through this progress researchers

are honing in on various biological and organic LBDs with improved substrate affinity and visualization.

Fortunately, biosensors and activity-based probes (ABPs) have proven pivotal to our understanding of even the most basic processes in the UB code – both ubiquitinating and deubiquitinating enzymes have been mechanistically characterized using these biosensors, with more details constantly being discovered.[46] Nonetheless, the complexity of ubiquitination poses a particularly challenging biosensing subject. In developing UBD-based probes, for instance, the vast diversity among UB chains and UBDs alike demands extensive research to even identify possible substrate-LBD combinations for further study. And, like MS, these tools require high purity of complexed proteins, significantly slowing the overall process of ubiquitin research. Most glaring, however, is the lack of chain specificity: neither MS nor LBD biosensors can conclusively distinguish specific chains from one another.[15] There is a great need for the ability to differentiate such modifications with high specificity due to their extremely precise implications, so this gap in research is especially egregious.[2] These and other obstacles in UB-sensing highlight the need for continued work in this area, as improved biosensors will surely expand our understanding of this modification, its biological roles, and PTMs in general.

3.7   The Deep Cavitand

One such PTM biosensor is the deep cavitand synthesized in the Hooley lab at University of California, Riverside. This organic macromolecule is symmetrically functionalized with four carboxylates along its rim, resulting in an overall negatively charged region. Intramolecular hydrogen bonds cause the cavitand to self-fold into a deep well shape, which promotes its binding to various functional groups, primarily positively-charged motifs (Figure 6).[47,48]

The cavitand is synthesized in three steps from an octanitrocavitand: the starting material is reduced to an octaamine with SnCl$_2$, heated with an imidate to yield a tetrabenzimidazole tetraester, then hydrolyzed with NaOH to produce the negative cavitand as a tetrasodium salt.[49] A single molecule of THF from the final synthetic step is positioned within the well of the cavitand – attempts to remove it have proven unsuccessful, and therefore only a sufficient guest, such as a lysine side chain, may displace the THF.[50,51] Crucially, the cavitand is also water-soluble over pH 7, and therefore ideal for a number of aqueous biochemical applications.[12]
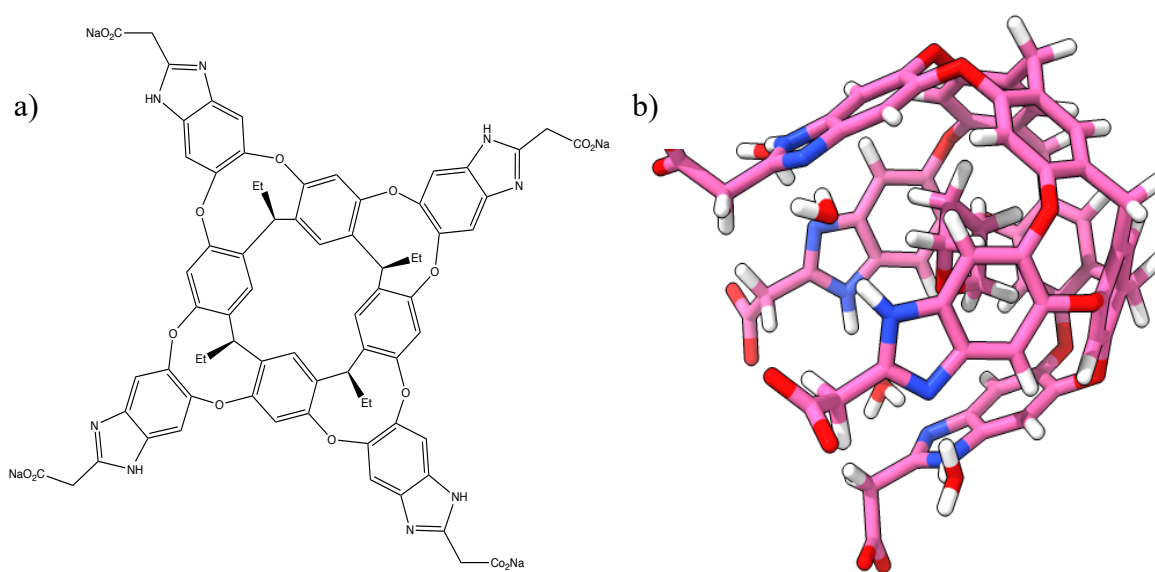

Figure 6: a) 2D and b) 3D structure and of the cavitand synthesized by the Hooley lab. Molecule is functionalized around the edge by carboxylate groups, which allow it to self-fold into a well shape via intramolecular hydrogen bonding. This molecule has been used successfully as a receptor in biomimetic membranes, an endocytosis and drug delivery agent, and as a fluorescent biosensor of PTMs—as such, these applications may be expanded into biosensing of ubiquitination and other modifications.

Thus far, the cavitand has proven successful as a receptor in biomimetic membranes, an endocytosis and drug delivery agent, and most notably as a fluorescent biosensor of post-translational modifications.[10–13,52] Perhaps most relevant to this project is this viability as a PTM biosensor – the cavitand used here has a documented affinity for lysine and arginine side chains,

27

and thus its incorporation into FRET biosensors of lysine-based modifications (such as methylations of histones) has proven quite successful.[53,54] Expansion into the sensing of additional PTMs is therefore a natural next step, especially for ubiquitination. And furthermore, the unique functions and modes of ligation characteristic of ubiquitination make even the more peripheral cavitand applications highly relevant to this modification. For example, ubiquitin's role in endocytosis is crucial to pharmaceutical development and drug delivery, since delivering medication across the cell membrane has proven historically difficult. The documented ability of the cavitand to permeate this barrier demonstrates promise as a possible vehicle through which to monitor ubiquitin's role in both intra- and extracellular communication or to help deliver drugs targeting the ubiquitin machinery.[10]
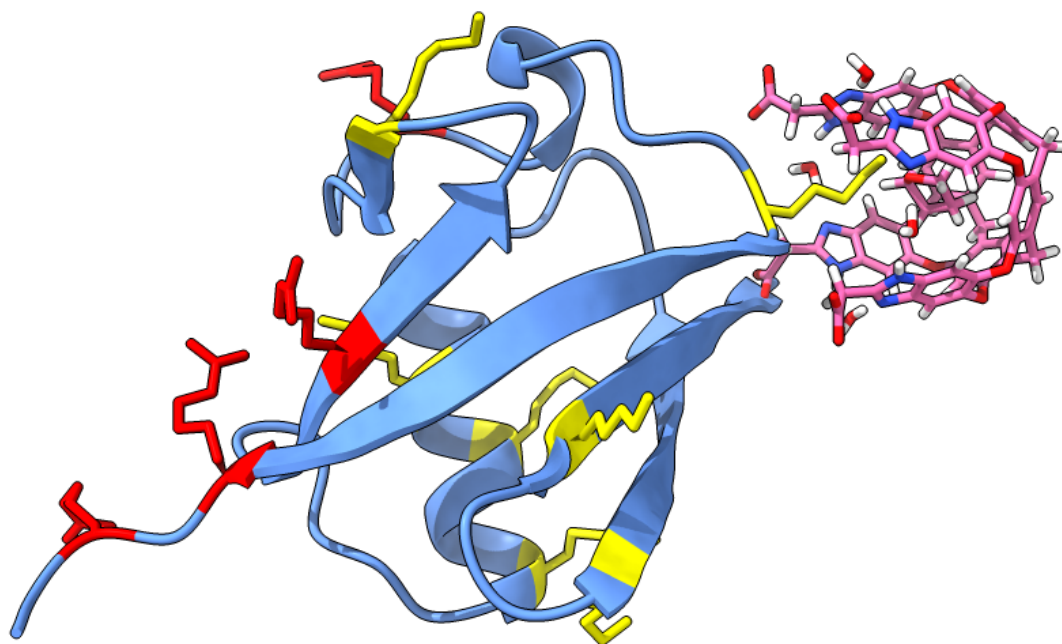

Figure 7: Projection of the ubiquitin-cavitand binding event. The cavitand has affinity for the amino groups on lysine and arginine side chains, and is therefore shown in complex with the K63 side chain.[22]

In general, depending on how effectively the cavitand binds to lysines and arginines, it may enter the library of viable LBDs for incorporation into various biosensors or tandem ubiquitin binding entities (TUBEs).[43] Lysines, of course, are of utmost importance in ubiquitin coding as the most prevalent chain linkage site – as such, it is possible that these chains may be distinguished from one another by sensing free lysines versus those used in dipeptide bonds. The small cavitand is therefore ideal for this role, a viable biosensor to scour topologically complex ubiquitin modifications and determine their linkage via selective lysine binding (Figure 7).

The ability to bind to arginine side chains is also of great use in ubiquitin sensing. As previously mentioned, UBDs often distinguish ubiquitin from similar proteins like SUMO via recognition of its C-terminus, which contains two arginine residues. These amino acids, Arg72 and Arg74, are notably absent in SUMO, and thus biosensing at these residues could also prove valuable in differentiating ubiquitination modifications from SUMOylations.[32] The structural and functional similarities in these two pathways pose a research challenge, one that is solvable through the targeting of unique sequence motifs like the ubiquitin C-terminus – this tail is often free and flexible, so it is possible that the cavitand may easily bind this characteristic LRLRGG region on ubiquitin and aid in distinguishing it from SUMO.[32] Such abilities, therefore, may prove applicable to the study of ubiquitin/SUMO crosstalk and these modifications' opposing roles in complex biological environments.

Overall, the cavitand's affinity for lysine and arginine side chains is an especially attractive feature in the area of biosensing. If made into a fluorescent or tandem probe, this molecule may prove successful in identifying the extent and specific type of ubiquitination in samples, without risk of binding adjacent SUMO modifications as well. Therefore, a study of the cavitand's binding abilities to the ubiquitin protein is quite warranted. Even beyond its applications to ubiquitin, the

cavitand may also prove valuable as a general PTM biosensor, particularly for lysine-based modifications. Nevertheless, ubiquitin's use as a model protein is key here, for through this function the patterns of cavitand/UB binding may be extrapolated and applied to cationic proteins in general.[55]

3.8   Protein NMR and HSQC

One technique which has proven quite useful in a number of areas of chemistry, including biophysical chemistry, is NMR spectroscopy. On a basic level, this method uses a magnet to excite particular nuclei in a sample and translates the feedback from their relaxation to deduce molecular structure, internuclear distances, and more.[56] Protein NMR in particular is often conducted in two dimensions (i.e. exciting two isotopes) because such large biomolecules can often return extremely complex one-dimensional spectra. However, even with this improvement, proteins above 30 kDa must often be analyzed via solid-state NMR rather than in the solution-state due to their increased complexity.[57] Solution and solid-state NMR both have their advantages in protein NMR: for one, solution NMR provides more sensitivity, but is inhibited by increasing spectral complexity which manifests as messily overlapping peaks. Multidimensional solution NMR helps to improve resolution, spreading the signals over multiple axes so peaks may be more readily identified. Alternatively, solid-state NMR, despite its inferior sensitivity and broader peaks, provides a wealth of information about the sample partially due to its inherent anisotropy. Nevertheless, ubiquitin's light and compact structure (8.6 kDa), along with the small cavitand, falls into the range of molecules amenable to analysis via two-dimensional solution-state NMR.[14]

The primary 2D NMR methods used to analyze protein samples can be either homonuclear (i.e. ^{1}H-^{1}H) or heteronuclear (i.e. ^{15}N-^{1}H) and often yield information about through-bond or through-space coupling of nuclei in the sample. Homonuclear correlation spectroscopy (COSY),

for instance, excites a single isotope (often ^{1}H) to yield a spectrum with crosspeaks indicating J-coupling of indirectly-bonded nuclei.

In contrast, the most common *hetero*nuclear correlation experiments include heteronuclear multiple quantum coherence (HMQC) and heteronuclear single quantum coherence (HSQC). These are quite useful in protein NMR because the J-coupling of insensitive nuclei directly bonded to protons reduces peak overlap, simplifies relaxation data, allows for bond orientation calculations from dipolar couplings, and enables protein assignments via amide N-H couplings.[56] In general, these experiments work in a similar way: both nuclei are excited by a signal from the NMR probe, after which transverse proton magnetization is generated and evolved through a 90° pulse as proton polarization is transferred to the observed heteroatom. The chemical shift of the heteroatom is recorded during this evolution period, after which it receives another 90° pulse prior to final detection of the sample's chemical shifts and J-couplings (Figure 8).[56]
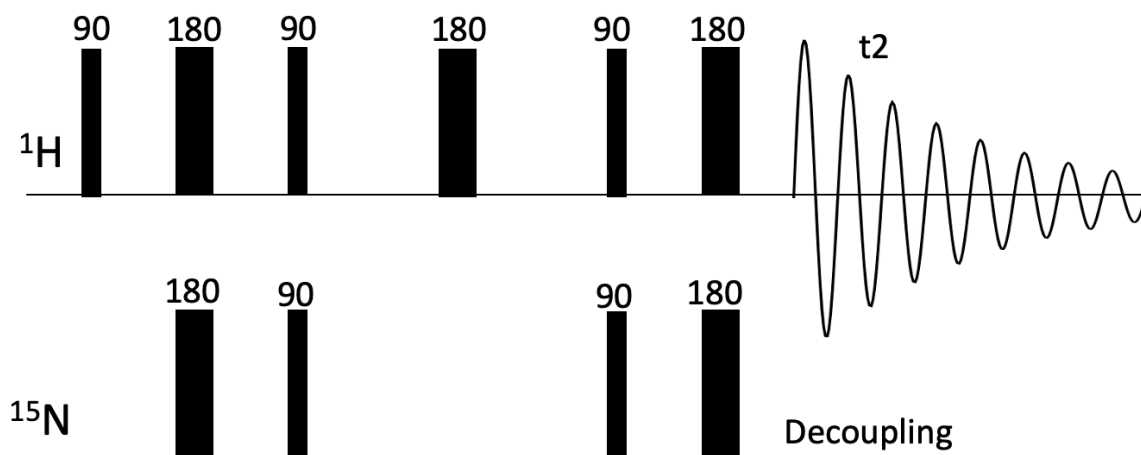

Figure 8: Pulse sequence of an ^{15}N-^{1}H HSQC experiment. This experiment requires a series of 90° pulses to excite the two observation nuclei and allow their signals to evolve. The resulting spectrum yields information on the N-H correlations in the sample, and as such has proven extremely useful in protein NMR.

The primary difference between these experiments lies in the evolution period – during this time, an HMQC allows for the magnetization of both types of nuclei to evolve, whereas an HSQC only allows for evolution of the heteroatom magnetization. This distinction primarily affects peak width, as the homonuclear J-coupling seen in HMQC experiments tends to broaden peaks in the spectrum.[56]

The narrower linewidths of HSQC spectra often aid in boosting sensitivity. This comparative sensitivity makes HSQC an especially good candidate for protein NMR – for a small, 10 kDa protein, for example, signal loss from relaxation is only 7%, whereas other protein experiments often experience over twice as much magnetization loss due to variation in the protein's T_2 relaxation time.[56] The T_2 of a sample designates the time it takes for transverse magnetization to decrease by 37% as nuclear spins fall out of phase, while T_1 relaxation concerns the return of longitudinal magnetization to its equilibrium magnitude. As molecular weight increases and molecular tumbling rate decreases, T_2 times shrink as well; these effects then increase linewidths to yield complex spectra for large proteins.[58] This phenomenon is what renders solution NMR fairly unhelpful for large proteins compared to solid-state, despite the wealth of information it can provide for smaller molecules and peptides.

These features have been demonstrated repeatedly in benchmarking of protein NMR experiments, which have found that small protein experiments achieve the best signal-to-noise (S/N) using HSQC.[59] Additional methods to boost this sensitivity include the use of cryogenic probes and larger magnets, and these have been essential and advancing the study of larger proteins via NMR.[59] Given the data and sensitivity afforded by HSQC experiments, this method is ideal for probing the interactions and binding of the deep cavitand to ubiquitin in solution. In particular, an ^{15}N-^{1}H HSQC allows one to understand the interbond J-coupling of N-H motifs in ubiquitin, which

are prevalent both along the backbone and in nitrogen-containing side chains such as lysine and arginine. The cavitand is predicted to bind to these side chains, so this alteration of the side chains' electronic environment should manifest in the spectrum as a change in the chemical shifts of the N-H signals. As such, analysis of the lysine and arginine chemical shifts in a UB-cavitand sample can yield valuable information on the location, strength, and prevalence of various cavitand binding events on the surface of ubiquitin.

3.9   Precedent

Despite the importance of lysine and arginine side chains in recognizing and executing the ubiquitin code, these residues have been studied very little via NMR. Most published research featuring ubiquitin ^{15}N-^{1}H HSQC is designed to focus on N-H pairs in the amide backbone rather than in side chains, and thus these chemical shifts have not been thoroughly characterized. This distinct lack of precedent in the literature highlights another reason to study ubiquitin side chains via HSQC.

Previous UB NMR inquiries focusing on its lysine and arginine side chains are minimal – of the few that exist, one study by Lu et al. utilized ^{15}N-^{1}H HSQC in tandem with mass spectrometry in order to assign arginine side chain resonances in natural-abundance human ubiquitin. The chemical shifts obtained for ^{1}H and ^{15}N varied between the four arginine residues present in UB, but all fell within the region of ~7 ppm and ~85 ppm, respectively (Figure 9a).[60] Surprisingly, 2D NMR studies of ubiquitin's lysine side chains are even more rare – the only spectrum available in the literature is a one-dimensional ^{15}N experiment, the work of Igumenova et al. to assign carbon resonances in microcrystalline ubiquitin using solid-state NMR. This spectrum provides much less information than a ^{15}N-^{1}H HSQC, but details that the ^{15}N chemical shifts of the lysine side chains fall around 30 ppm, as expected for an amine nitrogen (Figure 9b).[61]

Despite this gap in the literature, these studies have provided valuable information on the chemical shifts of ubiquitin's key residues, providing additional data to benchmark the ^{15}N-^{1}H HSQC experiments detailed here, both with and without the cavitand.
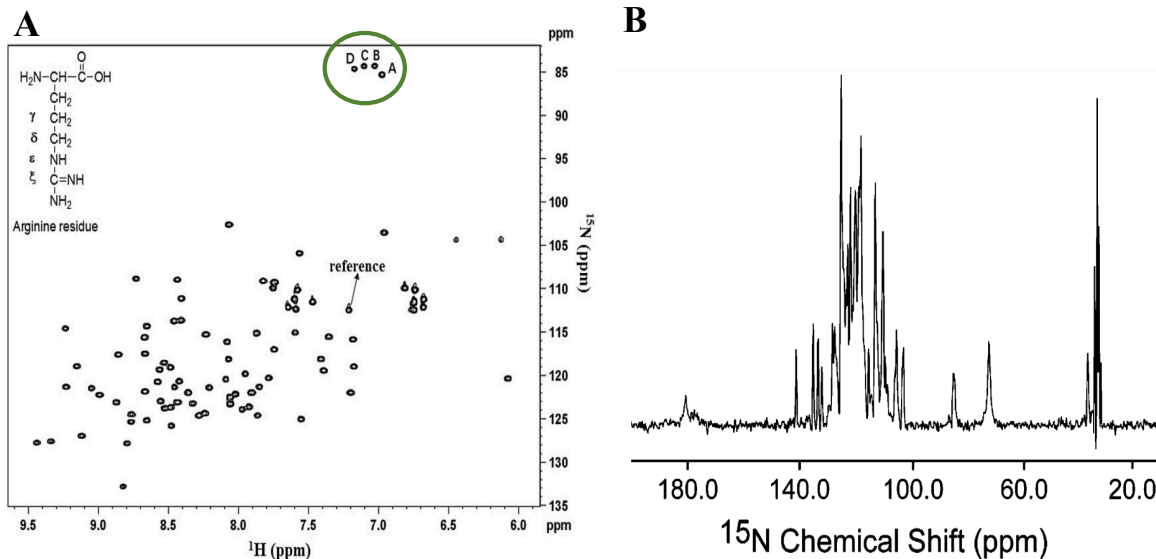

Figure 9: a) ^{15}N-^{1}H HSQC spectrum of ubiquitin, from Lu et al.[60] The chemical shifts of the four arginine side chains are indicated in the green circle. B) ^{15}N spectrum of ubiquitin, obtained by Iguemova et al.[61] Lysine side chain chemical shifts fall approximately around 30 ppm. These two regions of the spectrum have been identified as possible sites of cavitand binding, so this study focuses on perturbations in these chemical shifts, which may ultimately indicate a binding event.

This inquiry, a solution HSQC experiment of a protein with a small biosensor, also has precedent in the literature. McGovern et al. conducted a similar analysis to study the binding of small calixarene molecules to cytochrome C – these calixarenes resemble the cavitand both in their sulfate-functionalized rim and well-shaped conformation.[62] By using ^{15}N-^{1}H HSQC to analyze these molecules in complex with cytochrome C at various concentrations, McGovern et al. were able to identify the residues with significant chemical shift perturbations as the likely sites of calixarene binding. Though this research was largely conducted with the intention of adopting

calixarenes as possible protein camouflage agents, this work nonetheless provides an excellent model for similar ubiquitin-cavitand inquiries.

This study adopts an analogous method to McGovern et al., utilizing ^{15}N-^{1}H HSQC to analyze the chemical shifts of lysine and arginine side chains in ubiquitin titrated with the negative cavitand. As such, herein lies the groundwork for possible ubiquitination and general PTM biosensing by the cavitand, which may aid in furthering the study of post-translational modifications and assisting ubiquitination biosensing and diagnostics.

4   Materials and Methods

4.1   Ubiquitin Expression

Despite the prevalence of ubiquitin research and available expression protocols, the predominant method of ubiquitin expression cited in the literature relied on relatively outdated techniques for induction, lysing, and more. As such, unlabeled and ^{15}N-labeled ubiquitin was expressed using an updated version of this protocol.[63–66]

Human ubiquitin was expressed in *E. coli* Rosetta cells with a His-TEV tag. This tag, a sequence of six histidine residues ending with a unique TEV (tobacco etch virus) linkage site, was incorporated into the protein in order to aid later on in the purification step. After inoculation of the bacteria with a human UB glycerol stock, cells were grown in 1 L sterilized Luria broth (1% tryptone, 1% sodium chloride, 0.5% yeast extract, 0.01% kanamycin) for 5 hours at 37 °C, shaking at 225 rpm. After reaching OD 1 at 600 nm, media were centrifuged for 10 minutes at 4000 rpm, room temperature. Supernatant was discarded, and cells were transferred in the flame to sterilized culture flasks containing 1 L M9 Minimal Media (50 mM Na_2HPO_4, 50 mM KH_2PO_4, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2% mixed salt V/V, 0.5% 2 mg/mL biotin, 0.03 mM thiamine hydrochloride, 33 mM glucose, 55 mM $^{15}NH_4Cl$ (unlabeled for the unlabeled samples), 1% BME

vitamin solution, 4% ^{15}N-labeled BioExpress (omitted in the unlabeled samples), 0.005% kanamycin). Once pellets were dissolved, 0.2% IPTG was added to induce ubiquitin production and flasks were left to shake at 25 °C for 18 hours. Cells were then collected and weighed after centrifuging media for 30 minutes at 8000 rpm, 4 °C.

4.2   Ubiquitin Purification

To begin the purification, His-TEV-UB cells were resuspended in 30 mL Buffer A (50 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, 5% glycerol, 1 mM DTT, pH 8.0) with 1 mM PMSF as a protease inhibitor. This mixture was lysed by sonication over ice in 10 s, 80% amplitude pulses, for a total of 5 minutes. During sonication, a Ni-NTA affinity gel column was equilibrated with 50 mL Buffer A; this column was chosen based on its use of Ni^{2+} ions in the stationary phase, which bind the His-TEV tag on ubiquitin and enable its separation from impurities in the solution.

Following sonication, the lysed cells were centrifuged at 10,000 rpm, 4 °C, for 20 minutes. The resulting crude supernatant was then poured on the column and digested on a turn table at room temperature for one hour to ensure adequate mixing of the mobile and stationary phases; the pellet of cellular solids was discarded. After one hour of turning, the column's stopcock was opened to collect the flow-through. Column was then washed with 100 mL Buffer A, after which labeled His-TEV-UB was collected by eluting with 50 mL Buffer B (50 mM Tris-HCl, 500 mM NaCl, 400mM imidazole, pH 8.0). Elution of the protein was made possible by the high imidazole concentration in Buffer B, which facilitated displacement of the ubiquitin on the column. The protein was then precipitated from the elution by adding $(NH_4)_2SO_4$ to a final concentration of 4 mM, and the mixture was subsequently centrifuged for 20 minutes at 10,000 rpm, 4 °C. At this point, the supernatant was discarded and the remaining solid protein was resuspended in approximately 10 mL Buffer C (25 mM Tris-HCl, 200 mM NaCl, 5% glycerol, 1 mM DTT).

Ubiquitin sample concentration was then determined using a standard Bradford method assay; this technique requires the construction of a standard absorbance curve of bovine serum albumin (BSA) in solution with the Bradford reagent (0.12 mM Coomassie Brilliant Blue G-250, 10% V/V 85% H_3PO_4, 5% V/V 95% EtOH). The curve was constructed by mixing a series of BSA solutions (125-1500 µg/mL) with 3 mL Bradford reagent and allowing the mixtures to sit 5 minutes before reading their absorbances via UV-Vis (595 nm). The purified ubiquitin solution was then prepared at three different concentrations (undiluted, 1:1, 1:10), mixed with 3 mL Bradford reagent, and similarly analyzed via UV-Vis.

Finally, in order to cleave off the His-TEV tag, 100 µL TEV protease was added to the solution and allowed to digest overnight at room temperature. This digested sample was then loaded onto a Ni-NTA affinity gel column equilibrated with Buffer D (25 mM Tris-HCl, 200 mM NaCl, 30 mM imidazole, 5% glycerol, pH 8.0) and turned on the turn table for one hour at room temperature, after which the stopcock was opened to collect pure digested ubiquitin in the flow-through. The column was also washed with an additional 25 mL Buffer D and collected with the ubiquitin sample to maximize yield. A second Bradford method assay was conducted using this final digested sample in order to evaluate the effect of the third column on overall ubiquitin concentration and determine the concentration of the final, purified sample.

4.3   Gel Electrophoresis

The effectiveness of this overall purification protocol was consistently evaluated throughout the process in multiple rounds of gel electrophoresis. In this procedure, gels were poured using 5 mL 14% resolving gel (25% 1.5 M Tris-HCl, 35% bis/acrylamide, 1% SDS, 1% APS, 0.1% TEMED, pH 8.8) and 1.5 mL 4% stacking gel (25% 0.5 M Tris-HCl, 10% bis/acrylamide, 1% SDS, 1% APS, 0.1% TEMED, pH 6.8). Meanwhile, samples of ubiquitin from

throughout the purification process were prepared with 25% sample loading buffer and heated to 95 °C for 10 minutes to denature. The first gel was run using the crude lysis supernatant, flow-through, wash, and elution samples from two separate expressions. The second run simply involved samples of ^{15}N-labeled ubiquitin from before and after TEV protease digestion. Once the gel was set and samples were denatured, wells were loaded with a protein ladder (3 μL) and the prepared ubiquitin solutions (10 μL). The gel apparatus was then filled with 1.2 L TG-SDS running buffer, electrodes were connected, and the gel was allowed to run for 1.75 hours at a constant voltage of 225 V. When complete, the resulting gel was stained for about 2 hours in the prepared gel stain (0.29 mM Coomassie Brilliant Blue G-250, 10% CH_3COOH). Gel was then destained in water prior to analysis.

4.4   NMR Sample Preparation

Purified ubiquitin samples were divided amongst 4-mL centrifuge filtration tubes and spun down for 35 minutes at 4000xg, 4 °C, to reach a concentrated volume of around 200 μL. Each tube was then filled with 3.5 mL exchange buffer (25 mM Tris-HCl, 50 mM NaCl, pH 7.1) and centrifuged for 45 minutes at 4000xg, 4 °C. Fractions were pooled and spun down in the same conditions for an additional 10 minutes to reach a final volume of 500 μL. This process was repeated twice to yield two 500-μL aliquots of concentrated, desalted ubiquitin.

A number of ubiquitin samples were planned for HSQC analysis in varying cavitand : ubiquitin concentrations: 0:1, 0.25:1, 0.5:1, 0.75:1, 1:1, and 5:1. Cavitand was added as a 10 mM D_2O solution to provide a lock signal for the NMR. Finally, each sample was ready for analysis after its addition to a clean, dry NMR tube.

4.5 NMR Experiments

All samples were subjected to ^{15}N-^{1}H HSQC. These data were acquired on a Bruker Avance Neo spectrometer operating at 9.4 T (400.13 MHz-^{1}H) and 297 K using a TRX solution-state cryoprobe. The FID size for F2 was 2048 and 256 for F1, and each experiment consisted of 4 scans with an acquisition time of 0.164 s for proton and 0.0789 s for nitrogen. Spectra were processed using TopSpin 4.0.3. The spectrum obtained for the ubiquitin sample containing no cavitand was used to assign backbone resonances (SW 40 ppm, O1P 117 ppm) based on prior HSQC experiments.[67] A spectrum including some of the arginine side chain resonances was also obtained (SW 120 ppm, O1P 80 ppm), while signals for the lysine side chains could not be detected.

Completion of the NMR experiments using cavitand was prevented by precipitation of the cavitand from the sample solution. This was confirmed by proceeding with data collection, which demonstrated that protein signal was unchanged and unperturbed, indicating that the precipitate was cavitand only and that ubiquitin remained in solution. Sample pH was confirmed as 7.1, as pH above 7 is required for dissolution of the cavitand.

5 Results and Discussion

5.1 Gel Electrophoresis

Gel electrophoresis is designed to demonstrate the purity of a sample and identities of its components by separating the sample out based on each compound's molecular weight. Therefore, a sample with the desired purity should appear on the gel as a single band located adjacent to its correct molecular weight mark on the protein ladder. As such, this method proves useful in evaluating the efficacy of a given expression and purification protocol by demonstrating the components in a resulting sample.
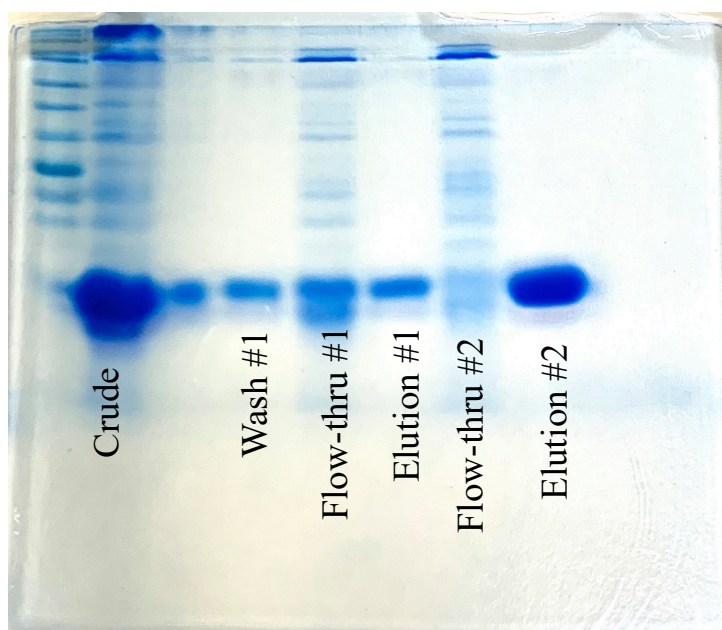
Figure 10: Results from gel electrophoresis run using samples from the purification of two ^{15}N-His-TEV-UB expressions. The numerous lines seen in the columns corresponding to crude supernatant and flow-through samples indicate extensive impurity in the starting samples. In contrast, one thick band is seen in the elution samples around the 8 kDa mark, demonstrating their relative purity and significant concentration of ubiquitin.

Results from gel electrophoresis runs demonstrated the robustness of the above ubiquitin expression and purification protocol, with the first gel specifically testing samples from two expressions of labeled ubiquitin (Figure 10). The wells containing crude supernatant and flow-through samples indicated numerous impurities at a range of different molecular weights, though by the end of purification, just one thick band was seen in the elution samples. It is unknown why the ubiquitin band from elution #2 was considerably larger than the other, however one explanation for this discrepancy may be unintended dilution of the sample during this step. Nevertheless, these elution bands illustrate the reliability of the detailed protocol, for they fall around the 8 kDa mark which is consistent with the molecular weight of ubiquitin and show no extraneous lines indicating impurities in the sample.

Additionally, the second gel was run to evaluate the effects of the TEV protease on a labeled ubiquitin sample, as it was unknown whether the enzyme spontaneously degraded or filtered out in the nickel column (Figure 11). It was also unclear how this additional purification step impacted the ubiquitin yield. As shown, the wells corresponding to digested samples yielded smaller ubiquitin bands than the undigested samples, indicating slightly lower concentrations of protein. Digestion byproduct was also present at a number of different molecular weights after using the TEV protease – this step therefore introduced 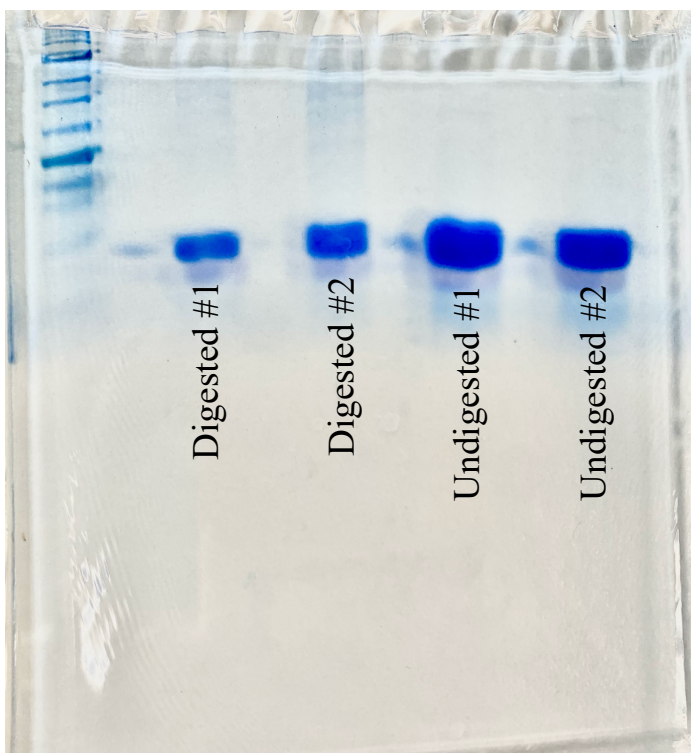some impurities that remained in the sample, indicating that a size-exclusion column may ultimately be necessary to obtain perfectly pure samples.

Figure 11: Results from the second gel electrophoresis run comparing TEV protease-digested and undigested ^{15}N-ubiquitin from two expressions. Both digested columns contain smaller bands, showing some loss in protein yield, as well as a subtle gradient indicating additional digestion byproduct remaining in the sample.

It is thought that keeping the His-TEV tag on ubiquitin may not significantly alter NMR experiments run with the sample, so future expressions may benefit by forgoing the digestion step in order to increase ubiquitin yields, improve purity of the sample, and simplify the protocol. However, before adoption of this altered method, future inquiries comparing the behavior of UB and His-TEV-UB in NMR experiments are warranted.

5.2   Bradford Assays

The Bradford assay yielded a standard absorbance curve for BSA, which was used to determine the concentrations of both digested and undigested ubiquitin samples (Figure 12). Undiluted sample 1 (His-TEV-UB) yielded a protein concentration of 67 μg/mL (7.4 μM), while undiluted sample 2 (untagged UB) was found to be 64 μg/mL (7.2 μM). After concentration via centrifugation, the final concentration of the sample was about 230 μM, which is on par which previous UB HSQC experiments.[38]

Given the smaller volume obtained after digestion, these results indicate that post-digestion purification decreased the overall ubiquitin yield, with negligible improvements to sample purity – however, as previously stated, NMR experiments must be used to confirm that eliminating the role of the TEV protease would not significantly alter the behavior and dynamics of the protein.
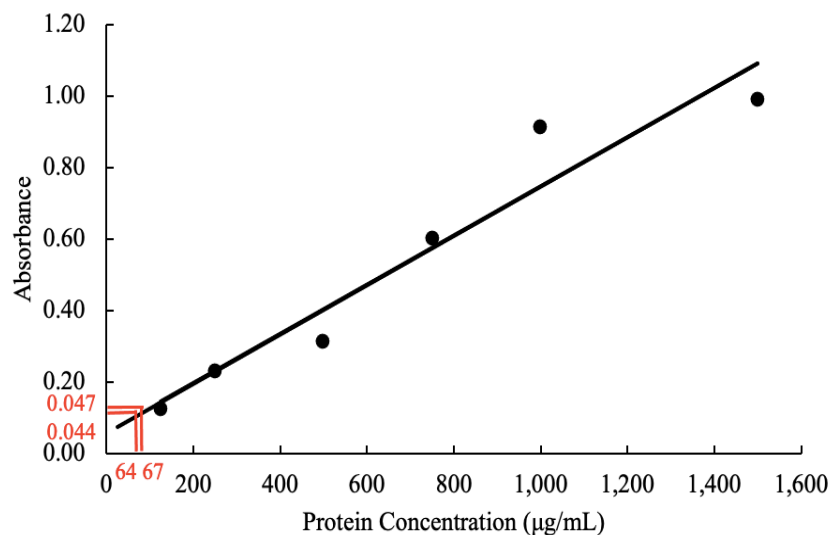

Figure 12: Standard absorbance curve of bovine serum albumin (BSA) [$y = 0.000688x + 9.51 \times 10^{-5}$, $R^2 = 0.93$]. This curve was used to determine the concentrations of six samples of digested or undigested ubiquitin (undiluted, 1:1, 10:1 dilution). Each sample was mixed with 3 mL Bradford reagent, allowed to sit 5 minutes, then analyzed via UV-Vis at 595 nm. Undiluted sample concentrations are indicated in red (undigested 67 μg/mL, digested 64 μg/mL).

5.3   NMR experiments

The ^{15}N-^{1}H HSQC backbone spectrum of pure labeled ubiquitin was relatively consistent with literature examples (Figure 13).[67] However, the desired side chain resonances were less visible for pure ubiquitin – only two of the four possible arginine side chain peaks were seen in the expected chemical shift region (Figure 14). This observation has made assignment of these residues relatively difficult, though it is possible that multiple arginine side chains exhibited identical chemical shifts and therefore overlapped in the spectrum.

Additionally, no peaks were detected for any of the seven lysine side chains on the protein. This is likely a result of proton exchange with the sample solvent, as it is quite common for protein amino groups to rapidly trade protons with surrounding protic solvent. Depending on the rate of this exchange, the timescale of an NMR experiment may be such that no proton is present on the
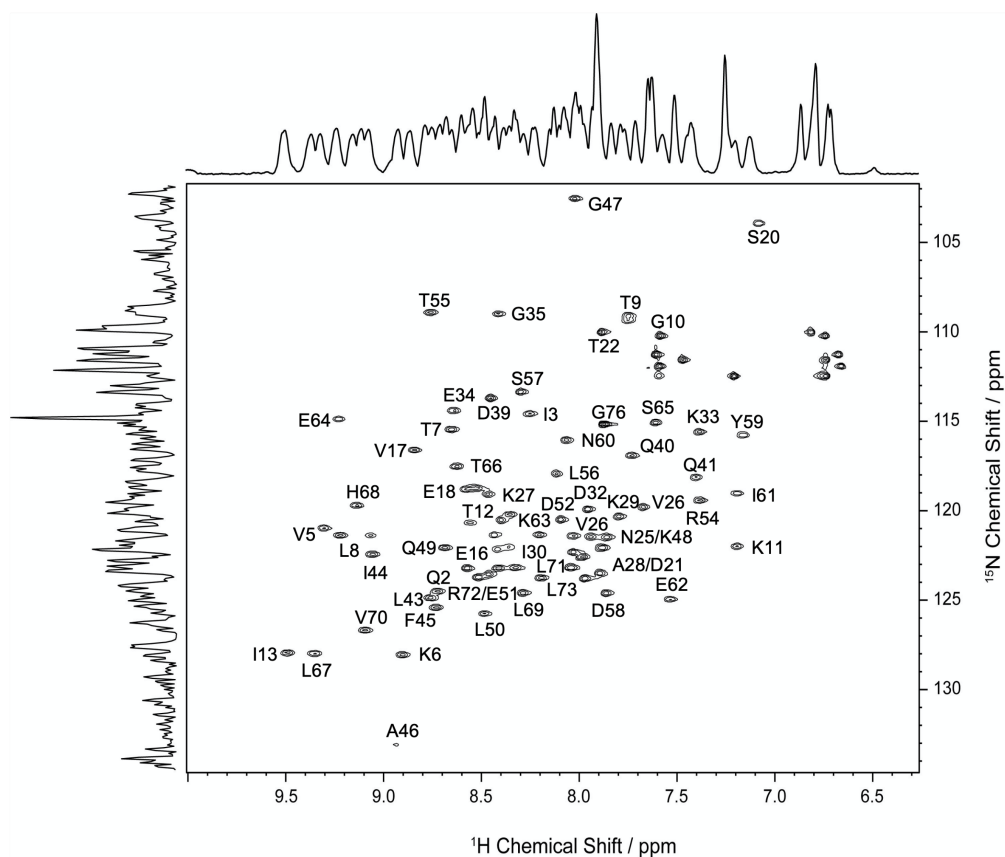

Figure 13: ^{15}N-^{1}H HSQC spectrum of pure, unperturbed ubiquitin backbone. Signals are assigned according to previously published spectra.[67]

protein long enough for correlation with nitrogen to occur and yield a crosspeak in the spectrum.[68] For this reason, lysine side chains have proven to be particularly challenging to detect via correlation spectroscopy, especially because the nature and rate of the exchange event varies widely based on pH of the sample solution.[69] For instance, a recent NMR study on lysine side chains in calmodulin led to the conclusion that proton exchange in these residues is highly pH-dependent (exchange is slowest around pH 6-7), and likely follows a complex 12-step exchange mechanism.[70]
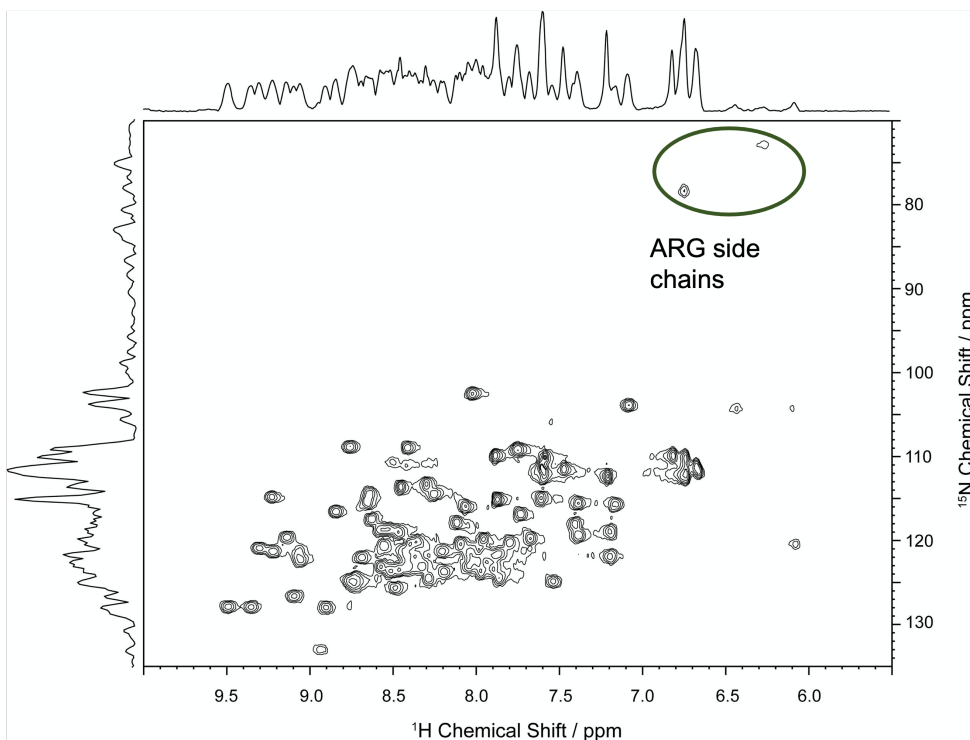

Figure 14: ^{15}N-^{1}H HSQC spectrum of pure, unperturbed ubiquitin backbone with side chains. Arginine side chain signals are designated by the green circle and fall around 80 – 85 ppm on the nitrogen axis, as is consistent with previously published spectra.[60] Only two signals can be seen, but it is possible that the of the four arginine residues, multiple have overlapping chemical shifts. Signals from the lysine side chains were undetected, likely due to proton exchange with the solvent.

Signals in such samples may only be studied via NMR when the exchange rate is much lower than the frequency difference of the two observe nuclei – even then, proper study of exchange processes is only possible through the use of equilibration or magnetization transfer.[68,71]

Without such methods, these studies are unlikely to properly capture the correlation of side chains prone to proton exchange. This outcome is consistent with the distinct lack of HSQC studies focused on ubiquitin's lysine side chains: it is very likely that observation of these peaks is difficult, so few papers have attempted such studies. However, it has been predicted that the cavitand binding event may inhibit lysine proton exchange by involving the side chain in a strong hydrogen bond and preventing exchange with the surrounding solvent molecules, therefore making correlation and the occurrence of observable peaks more possible with increasing cavitand concentration. Given this possibility, ^{15}N-^{1}H HSQC was still deemed to be a valid method for this study despite the absence of lysine side chain signals in the pure ubiquitin spectrum.

All experiments featuring the cavitand were prohibited by unexpected behavior of the cavitand upon sample preparation: prior to the experiments, the cavitand precipitated out of solution. The precipitate's identity was confirmed by continuing data collection, which demonstrated a protein signal consistent with previous experiments and no chemical shifts corresponding to the cavitand, confirming that only the cavitand had come out of solution. This result is unexpected in these sample conditions, so further investigation is warranted into the cause of the precipitation event and how sample preparation might be altered to encourage dissolution of the cavitand while simultaneously maintaining protein concentration.

One possible cause of cavitand precipitation was the prescence of digestion product in the final sample, as indicated by gel electrophoresis (Figure 11). The sample was reasonably pure, but it was found that after the final column, some byproduct of high molecular weight remained in the sample. This was likely some degraded form of the TEV protease and may have contributed to precipitation of the cavitand. To remedy this issue, future purifications may improve sample purity by introducing size-exclusion chromatography as the final step. Another option is sucrose-gradient

centrifugation, in which a sample is fractionated based on molecular size and density. Either of these methods may improve sample conditions by taking advantage of the size disparity between ubiquitin and the TEV protease, allowing for collection of lower-weight fractions containing pure ubiquitin in brine.

If the cavitand remains insoluble in the sample even after this additional purification step, some additional parameters may need to be altered. For example, increasing the sample's salt content may improve cavitand solubility; however increased conductivity can also hamper tuning and matching, and therefore the RF signal in NMR experiments.[56] With this, an ideal salt content may be found that minimizes NMR signal interference while enabling cavitand dissolution. Additional properties that may be tested include pH and buffer content. Overall, optimization of these properties may help to improve cavitand solubility in order for HSQC experiments to proceed. Once this is complete, ubiquitin-cavitand interactions may finally be characterized for the eventual development of a ubiquitination or general PTM biosensor.

6   Conclusion

6.1   Initial Results

This study aimed to expand cavitand and ubiquitin research by characterizing the nature of UB-cavitand interactions via ^{15}N-^{1}H HSQC. These interactions hold relevance to the future of many biochemical applications: for example, ubiquitin is a supremely important protein in the maintenance of cellular homeostasis, especially given its role in regulating cell cycles, protein degradation, and DNA repair. Currently, these biochemical pathways are viewed as major research frontiers due to their well-documented involvement in the manifestation and proliferation of disease, including multiple types cancer and neurodegeneration. Even further, the cavitand synthesized in the Hooley lab serves as a useful and promising tool to advance biochemical

discovery, not only in research applications but in biosensing and diagnostics as well. Since the cavitand has proven successful as a FRET biosensor of PTMs like lysine methylation, it has been thought that these functions may extend to additional modifications such as ubiquitination, SUMOylation, acetylation, and more.

Given such promise, this study on the UB-cavitand binding event was highly warranted, especially due to the increased complexity of the ubiquitin surface in comparison to other PTM substrates like methylated lysines. This type of experiment has precedence in the study of other small molecule receptors and proteins, but this particular NMR-based investigation of ubiquitin's lysine and arginine side chains is novel. Its relevance is further extended due to ubiquitin's dominance in the literature as a model for cationic proteins – this designation indicates that the characteristics of UB-cavitand binding likely extend to many other biologically relevant proteins, which only further expands the possible cavitand biosensing subjects and applications. To this end, ^{15}N-labeled ubiquitin was expressed, purified, and studied in the presence of cavitand via ^{15}N-^{1}H HSQC, a common protein NMR experiment which allows one to observe the correlations of directly-bonded N-H motifs in the sample. The cavitand has previously demonstrated affinity for lysine and arginine side chains, both of which contain N-H motifs, so it was hypothesized that the UB-cavitand binding event may manifest in the HSQC spectra as statistically significant chemical shift perturbations. It was also predicted that these perturbations would possibly increase in number or intensity with increasing cavitand concentration.

In practice, it was found that the NMR sample conditions were inhibitory to the dissolution of the cavitand, which prevented characterization of its interactions with ubiquitin. Spectra of pure protein were obtained in which arginine side chain signals were detected while lysine side chains were not, likely due to proton exchange with the sample solvent inhibiting correlation. This

spectrum also confirmed the success of our ubiquitin expression and purification, which is currently unpublished. However, the precipitation of the cavitand that inhibited further NMR experiments may indicate the need for improvements to the protocol, such as the addition of size-exclusion chromatography or adjustment of sample pH or buffer. To complete the inquiry laid out in this study, additional experiments are required which may only be conducted after adjusting the sample preparation process to improve the solubility of the cavitand.

6.2   Future Work and Applications

Given the ineffectiveness of this experiment in characterizing UB-cavitand interactions, further research is required before the cavitand is ready for application as a ubiquitination biosensor in a natural environment. Immediate next steps have already been outlined, which will mainly focus on refining the sample preparation protocol to enable dissolution of the cavitand. Certain properties that may be altered include the use of size-exclusion chromatography or modification of sample pH and buffer content. These investigations will hopefully allow for the eventual characterization of UB-cavitand interactions – spectra of pure ubiquitin have already been obtained and yielded information on unperturbed arginine side chains, and upcoming experiments will be able to complete such work by identifying sites of cavitand binding via analysis of chemical shift perturbations.

Beyond the research proposed here, the cavitand remains a promising subject for a variety of relevant, and increasingly complex, applications. For instance, this study focused on the binding of a single free ubiquitin, whereas natural environments are often further complicated by the proximity of ubiquitinated substrates and the existence of complex UB chains. As such, similar HSQC inquiries of UB-cavitand interactions using samples of ubiquitinated protein or ubiquitin chains may shed light on the potential of the cavitand as a UB biosensor in more realistic

environments. From there, comparisons of cavitand binding on the existing array of ubiquitin modifications may reveal how the molecule interacts with specific chains based on their specific topologies and surfaces. Since chain-specific UB sensing has proven difficult in the past, sufficiently unique cavitand binding events with different chains may indicate great promise for this molecule in future ubiquitination research.

For biosensing applications, characterization of the cavitand's interactions with other proteins or modifications is also warranted. In essence, this may shed light on its affinity and sensing potential for a variety of biochemical species, as preferential binding of the sensing target is, of course, a desired attribute of any biosensor. For example, lysozyme is a model substrate for many E3 ligases, and as such is relatively similar to ubiquitin in its surface electrostatics.[14] In the same vein, SUMO has been discussed previously as a protein with remarkable structural similarities to ubiquitin—therefore, HSQC studies of these enzymes in complex with the cavitand may reveal the extent to which it can distinguish structurally and electronically similar substrates. This experiment is of particular relevance due to the intertwining of UB and SUMO modifications: since these pathways are often functionally opposed, distinct sensing of both could therefore double the diagnostic power of this cavitand in ongoing cancer and neurodegeneration research.

The previously documented applications of this cavitand, one example being the sensing of lysine methylation, also inform any future directions for this research. Given this precedence and the role of ubiquitin as a model protein, the sensing of additional modifications like acetylation is likely possible as well. Additionally, the cavitand has also been used as an endocytosis agent, capable of transporting drugs or other small molecules across the cell membrane. There are currently no known cell-permeable inhibitors of ubiquitin, so a host molecule like the cavitand

may serve this purpose effectively, helping a given inhibitor guest across the membrane or even delivering it directly to the ubiquitin target using its affinity for the protein.[14]

These possible applications are just a few of the many possibilities given the results documented here. Still, despite its importance in so many vital cell processes, characterization of ubiquitin function, behavior, and regulation remains quite incomplete. This knowledge gap warrants immediate attention, for advancing research continues to shed light on the ever-increasing biological importance of ubiquitin in countless pathways, many of which are plagued by errors and implicated in major diseases and developmental disorders. As shown here, the Hooley lab's cavitand serves as one possible avenue out of this obscurity; with its biosensing potential, future biochemical research may more effectively study ubiquitination pathways and untangle the intricacies of this vital and omnipresent protein.

7   Acknowledgements

I owe endless thanks to my advisor, Dr. Bethany Caulkins, for all of her contributions to my research and growth as a scientist. Through the process of writing this thesis, she taught me how to feel capable in the lab, and even enabled the completion of this research by conducting the NMR experiments at UC Riverside. I am also grateful to the rest of my lab group and Dr. Mary Hatcher-Skeers for their support and comments throughout the semester. Without these contributions, this paper would not be of the same quality.

8   References

(1)   Hoffman, M. D.; Sniatynski, M. J.; Kast, J. Current Approaches for Global Post-Translational Modification Discovery and Mass Spectrometric Analysis. *Anal. Chim. Acta* **2008**, *627* (1), 50–61. https://doi.org/10.1016/j.aca.2008.03.032.

(2)   Conibear, A. C. Deciphering Protein Post-Translational Modifications Using Chemical Biology Tools. *Nat. Rev. Chem.* **2020**, *4* (12), 674–695. https://doi.org/10.1038/s41570-020-00223-8.

(3)   Aye-Han, N.-N.; Ni, Q.; Zhang, J. Fluorescent Biosensors for Real-Time Tracking of Post-Translational Modification Dynamics. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 392–397. https://doi.org/10.1016/j.cbpa.2009.07.009.

(4)   Millar, A. H.; Heazlewood, J. L.; Giglione, C.; Holdsworth, M. J.; Bachmair, A.; Schulze, W. X. The Scope, Functions, and Dynamics of Posttranslational Protein Modifications. *Annu. Rev. Plant Biol.* **2019**, *70* (1), 119–151. https://doi.org/10.1146/annurev-arplant-050718-100211.

(5)   Deribe, Y. L.; Pawson, T.; Dikic, I. Post-Translational Modifications in Signal Integration. *Nat. Struct. Mol. Biol.* **2010**, *17* (6), 666–672. https://doi.org/10.1038/nsmb.1842.

(6)   Kumar, A.; Narayanan, V.; Sekhar, A. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. *Biochemistry* **2020**, *59* (1), 57–73. https://doi.org/10.1021/acs.biochem.9b00827.

(7)   Karve, T. M.; Cheema, A. K. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *J. Amino Acids* **2011**, *2011*, e207691. https://doi.org/10.4061/2011/207691.

(8)   Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203–229. https://doi.org/10.1146/annurev-biochem-060310-170328.

(9)   Rape, M. Ubiquitylation at the Crossroads of Development and Disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (1), 59–70. https://doi.org/10.1038/nrm.2017.83.

(10)   Ghang, Y.-J.; Schramm, M. P.; Zhang, F.; Acey, R. A.; David, C. N.; Wilson, E. H.; Wang, Y.; Cheng, Q.; Hooley, R. J. Selective Cavitand-Mediated Endocytosis of Targeted Imaging Agents into Live Cells. *J. Am. Chem. Soc.* **2013**, *135* (19), 7090–7093. https://doi.org/10.1021/ja401273g.

(11)   Gill, A. D.; Hickey, B. L.; Wang, S.; Xue, M.; Zhong, W.; Hooley, R. J. Sensing of Citrulline Modifications in Histone Peptides by Deep Cavitand Hosts. *Chem. Commun.* **2019**, *55* (88), 13259–13262. https://doi.org/10.1039/C9CC07002H.

(12)   Perez, L.; Mettry, M.; Hinman, S. S.; Byers, S. R.; McKeating, K. S.; Caulkins, B. G.; Cheng, Q.; Hooley, R. J. Selective Protein Recognition in Supported Lipid Bilayer Arrays by Tailored, Dual-Mode Deep Cavitand Hosts. *Soft Matter* **2017**, *13* (21), 3966–3974. https://doi.org/10.1039/C7SM00192D.

(13)   Liu, Y.; Perez, L.; Mettry, M.; Easley, C. J.; Hooley, R. J.; Zhong, W. Self-Aggregating Deep Cavitand Acts as a Fluorescence Displacement Sensor for Lysine Methylation. *J. Am. Chem. Soc.* **2016**, *138* (34), 10746–10749. https://doi.org/10.1021/jacs.6b05897.

(14)   Pickart, C. M. Back to the Future with Ubiquitin. *Cell* **2004**, *116* (2), 181–190. https://doi.org/10.1016/s0092-8674(03)01074-2.

(15)   Kliza, K.; Husnjak, K. Resolving the Complexity of Ubiquitin Networks. *Front. Mol. Biosci.* **2020**, *7*. https://doi.org/10.3389/fmolb.2020.00021.

(16) Hagai, T.; Azia, A.; Tóth-Petróczy, Á.; Levy, Y. Intrinsic Disorder in Ubiquitination Substrates. *J. Mol. Biol.* **2011**, *412* (3), 319–324. https://doi.org/10.1016/j.jmb.2011.07.024.

(17) Tsuchida, S.; Nakayama, T. Ubiquitination and Deubiquitination in Oral Disease. *Int. J. Mol. Sci.* **2021**, *22* (11), 5488. https://doi.org/10.3390/ijms22115488.

(18) Fraile, J. M.; Quesada, V.; Rodríguez, D.; Freije, J. M. P.; López-Otín, C. Deubiquitinases in Cancer: New Functions and Therapeutic Options. *Oncogene* **2012**, *31* (19), 2373–2388. https://doi.org/10.1038/onc.2011.443.

(19) Shi, D.; Grossman, S. R. Ubiquitin Becomes Ubiquitous in Cancer. *Cancer Biol. Ther.* **2010**, *10* (8), 737–747. https://doi.org/10.4161/cbt.10.8.13417.

(20) Burroughs, A. M.; Balaji, S.; Iyer, L. M.; Aravind, L. Small but Versatile: The Extraordinary Functional and Structural Diversity of the β-Grasp Fold. *Biol. Direct* **2007**, *2* (1), 18. https://doi.org/10.1186/1745-6150-2-18.

(21) Dikic, I.; Wakatsuki, S.; Walters, K. J. Ubiquitin Binding Domains — from Structures to Functions. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (10), 659–671. https://doi.org/10.1038/nrm2767.

(22) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(23) Wang, Y.-S.; Wu, K.-P.; Jiang, H.-K.; Kurkute, P.; Chen, R.-H. Branched Ubiquitination: Detection Methods, Biological Functions and Chemical Synthesis. *Molecules* **2020**, *25* (21), 5200. https://doi.org/10.3390/molecules25215200.

(24) Dittmar, G.; Winklhofer, K. F. Linear Ubiquitin Chains: Cellular Functions and Strategies for Detection and Quantification. *Front. Chem.* **2020**, *0*. https://doi.org/10.3389/fchem.2019.00915.

(25) Shekhawat, S. S.; Pham, G. H.; Prabakaran, J.; Strieter, E. R. Simultaneous Detection of Distinct Ubiquitin Chain Topologies by 19F NMR. *ACS Chem. Biol.* **2014**, *9* (10), 2229–2236. https://doi.org/10.1021/cb500589c.

(26) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed.* **2005**, *44* (45), 7342–7372. https://doi.org/10.1002/anie.200501023.

(27) Buetow, L.; Huang, D. T. Structural Insights into the Catalysis and Regulation of E3 Ubiquitin Ligases. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (10), 626–642. https://doi.org/10.1038/nrm.2016.91.

(28) Ardley, H. C.; Robinson, P. A. E3 Ubiquitin Ligases. *Essays Biochem.* **2005**, *41*, 15–30. https://doi.org/10.1042/EB0410015.

(29) Vucic, D.; Dixit, V. M.; Wertz, I. E. Ubiquitylation in Apoptosis: A Post-Translational Modification at the Edge of Life and Death. *Nat. Rev. Mol. Cell Biol.* **2011**, *12* (7), 439–452. https://doi.org/10.1038/nrm3143.

(30) Seirafi, M.; Kozlov, G.; Gehring, K. Parkin Structure and Function. *Febs J.* **2015**, *282* (11), 2076–2088. https://doi.org/10.1111/febs.13249.

(31) Callis, J. The Ubiquitination Machinery of the Ubiquitin System. *Arab. Book Am. Soc. Plant Biol.* **2014**, *12*. https://doi.org/10.1199/tab.0174.

(32) Komander, D.; Clague, M. J.; Urbé, S. Breaking the Chains: Structure and Function of the Deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (8), 550–563. https://doi.org/10.1038/nrm2731.

(33) Hurley, J. H.; Lee, S.; Prag, G. Ubiquitin-Binding Domains. *Biochem. J.* **2006**, *399* (3), 361–372. https://doi.org/10.1042/BJ20061138.

(34) Gill, G. SUMO and Ubiquitin in the Nucleus: Different Functions, Similar Mechanisms? *Genes Dev.* **2004**, *18* (17), 2046–2059. https://doi.org/10.1101/gad.1214604.

(35) Hunter, T. The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Mol. Cell* **2007**, *28* (5), 730–738. https://doi.org/10.1016/j.molcel.2007.11.019.

(36) Minguez, P.; Parca, L.; Diella, F.; Mende, D. R.; Kumar, R.; Helmer-Citterich, M.; Gavin, A.-C.; van Noort, V.; Bork, P. Deciphering a Global Network of Functionally Associated Post-Translational Modifications. *Mol. Syst. Biol.* **2012**, *8* (1), 599. https://doi.org/10.1038/msb.2012.31.

(37) Chen, J.; Bozza, W.; Zhuang, Z. Ubiquitination of PCNA and Its Essential Role in Eukaryotic Translesion Synthesis. *Cell Biochem. Biophys.* **2011**, *60* (1), 47–60. https://doi.org/10.1007/s12013-011-9187-3.

(38) McKenna, S.; Moraes, T.; Pastushok, L.; Ptak, C.; Xiao, W.; Spyracopoulos, L.; Ellison, M. J. An NMR-Based Model of the Ubiquitin-Bound Human Ubiquitin Conjugation Complex Mms2·Ubc13: THE STRUCTURAL BASIS FOR LYSINE 63 CHAIN CATALYSIS * 210. *J. Biol. Chem.* **2003**, *278* (15), 13151–13158. https://doi.org/10.1074/jbc.M212353200.

(39) Wilkinson, K. D. The Discovery of Ubiquitin-Dependent Proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15280–15282. https://doi.org/10.1073/pnas.0504842102.

(40)   Beal, R.; Deveraux, Q.; Xia, G.; Rechsteiner, M.; Pickart, C. Surface Hydrophobic Residues of Multiubiquitin Chains Essential for Proteolytic Targeting. *Proc. Natl. Acad. Sci.* **1996**, *93* (2), 861–866. https://doi.org/10.1073/pnas.93.2.861.

(41)   Kerscher, O.; Felberbaum, R.; Hochstrasser, M. Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins. *Annu. Rev. Cell Dev. Biol.* **2006**, *22* (1), 159–180. https://doi.org/10.1146/annurev.cellbio.22.010605.093503.

(42)   Kumar, D.; Ambasta, R. K.; Kumar, P. Ubiquitin Biology in Neurodegenerative Disorders: From Impairment to Therapeutic Strategies. *Ageing Res. Rev.* **2020**, *61*, 101078. https://doi.org/10.1016/j.arr.2020.101078.

(43)   Beaudette, P.; Popp, O.; Dittmar, G. Proteomic Techniques to Probe the Ubiquitin Landscape. *PROTEOMICS* **2016**, *16* (2), 273–287. https://doi.org/10.1002/pmic.201500290.

(44)   Mehrotra, P. Biosensors and Their Applications – A Review. *J. Oral Biol. Craniofacial Res.* **2016**, *6* (2), 153–159. https://doi.org/10.1016/j.jobcr.2015.12.002.

(45)   Feng, J.; Jester, B. W.; Tinberg, C. E.; Mandell, D. J.; Antunes, M. S.; Chari, R.; Morey, K. J.; Rios, X.; Medford, J. I.; Church, G. M.; Fields, S.; Baker, D. A General Strategy to Construct Small Molecule Biosensors in Eukaryotes. *eLife* **2015**, *4*, e10606. https://doi.org/10.7554/eLife.10606.

(46)   Witting, K. F.; Mulder, M. P. C.; Ovaa, H. Advancing Our Understanding of Ubiquitination Using the Ub-Toolkit. *J. Mol. Biol.* **2017**, *429* (22), 3388–3394. https://doi.org/10.1016/j.jmb.2017.04.002.

(47)   Hooley, R. J.; Shenoy, S. R.; Rebek, J. Electronic and Steric Effects in Binding of Deep Cavitands. *Org. Lett.* **2008**, *10* (23), 5397–5400. https://doi.org/10.1021/ol8022775.

(48)   Liu, Y.; Perez, L.; Mettry, M.; Gill, A. D.; Byers, S. R.; Easley, C. J.; Bardeen, C. J.; Zhong, W.; Hooley, R. J. Site Selective Reading of Epigenetic Markers by a Dual-Mode Synthetic Receptor Array. *Chem. Sci.* **2017**, *8* (5), 3960–3970. https://doi.org/10.1039/C7SC00865A.

(49)   Hof, F.; Trembleau, L.; Ullrich, E. C.; Rebek, J., Jr. Acetylcholine Recognition by a Deep, Biomimetic Pocket. *Angew. Chem. Int. Ed.* **2003**, *42* (27), 3150–3153. https://doi.org/10.1002/anie.200351174.

(50)   Biros, S. M.; Ullrich, E. C.; Hof, F.; Trembleau, L.; Rebek, J. Kinetically Stable Complexes in Water: The Role of Hydration and Hydrophobicity. *J. Am. Chem. Soc.* **2004**, *126* (9), 2870–2876. https://doi.org/10.1021/ja038823m.

(51)   Hooley, R. J.; Van Anda, H. J.; Rebek, J. Extraction of Hydrophobic Species into a Water-Soluble Synthetic Receptor. *J. Am. Chem. Soc.* **2007**, *129* (44), 13464–13473. https://doi.org/10.1021/ja0727058.

(52)   Liu, Y.; Lee, J.; Perez, L.; Gill, A. D.; Hooley, R. J.; Zhong, W. Selective Sensing of Phosphorylated Peptides and Monitoring Kinase and Phosphatase Activity with a Supramolecular Tandem Assay. *J. Am. Chem. Soc.* **2018**, *140* (42), 13869–13877. https://doi.org/10.1021/jacs.8b08693.

(53)   Liu, Y.; Perez, L.; Gill, A. D.; Mettry, M.; Li, L.; Wang, Y.; Hooley, R. J.; Zhong, W. Site-Selective Sensing of Histone Methylation Enzyme Activity via an Arrayed Supramolecular Tandem Assay. *J. Am. Chem. Soc.* **2017**, *139* (32), 10964–10967. https://doi.org/10.1021/jacs.7b05002.

(54)   Lee, J.; Perez, L.; Liu, Y.; Wang, H.; Hooley, R. J.; Zhong, W. Separation of Methylated Histone Peptides via Host-Assisted Capillary Electrophoresis. *Anal. Chem.* **2018**, *90* (3), 1881–1888. https://doi.org/10.1021/acs.analchem.7b03969.

(55) Pickart, C. M.; Eddins, M. J. Ubiquitin: Structures, Functions, Mechanisms. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2004**, *1695* (1), 55–72. https://doi.org/10.1016/j.bbamcr.2004.09.019.

(56) Rule, G. S.; Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*; Focus on Structural Biology; Springer Netherlands, 2006. https://doi.org/10.1007/1-4020-3500-4.

(57) Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y.; Lee, H.-W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Determination of Solution Structures of Proteins up to 40 KDa Using CS-Rosetta with Sparse NMR Data from Deuterated Samples. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (27), 10873–10878. https://doi.org/10.1073/pnas.1203013109.

(58) Grzesiek, S. Notes on Relaxation and Dynamics. *EMBO Pract. Course NMR* **2003**, 41.

(59) Senthamarai, R. R. P.; Kuprov, I.; Pervushin, K. Benchmarking NMR Experiments: A Relational Database of Protein Pulse Sequences. *J. Magn. Reson.* **2010**, *203* (1), 129–137. https://doi.org/10.1016/j.jmr.2009.12.008.

(60) Lu, J.; Zhou, F.; Liu, W.; Yu, F. Mass Spectrometry Assisted Arginine Side Chains Assignment of NMR Resonances in Natural Abundance Proteins. *J. Biomol. NMR* **2020**, *74* (2), 173–181. https://doi.org/10.1007/s10858-020-00302-4.

(61) Igumenova, T. I.; McDermott, A. E.; Zilm, K. W.; Martin, R. W.; Paulson, E. K.; Wand, A. J. Assignments of Carbon NMR Resonances for Microcrystalline Ubiquitin. *J. Am. Chem. Soc.* **2004**, *126* (21), 6720–6727. https://doi.org/10.1021/ja030547o.

(62) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. Protein Camouflage in Cytochrome c –Calixarene Complexes. *Nat. Chem.* **2012**, *4* (7), 527–533. https://doi.org/10.1038/nchem.1342.

(63)   Chemical Synthesis and Expression of a Cassette Adapted Ubiquitin Gene. *J. Biol. Chem.* **1987**, *262* (8), 3524–3527. https://doi.org/10.1016/S0021-9258(18)61382-1.

(64)   Gene Synthesis, Expression, Structures, and Functional Activities of Site-Specific Mutants of Ubiquitin. *J. Biol. Chem.* **1987**, *262* (29), 14213–14221. https://doi.org/10.1016/S0021-9258(18)47925-2.

(65)   Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. Internal Dynamics of Human Ubiquitin Revealed by 13C-Relaxation Studies of Randomly Fractionally Labeled Protein. *Biochemistry* **1996**, *35* (19), 6116–6125. https://doi.org/10.1021/bi9530144.

(66)   Catanzariti, A.-M.; Soboleva, T. A.; Jans, D. A.; Board, P. G.; Baker, R. T. An Efficient System for High-Level Expression and Easy Purification of Authentic Recombinant Proteins. *Protein Sci. Publ. Protein Soc.* **2004**, *13* (5), 1331–1339. https://doi.org/10.1110/ps.04618904.

(67)   Bougault, C.; Feng, L.; Glushka, J.; Kupce, E.; Prestegard, J. H. Quantitation of Rapid Proton-Deuteron Amide Exchange Using Hadamard Spectroscopy. *J. Biomol. NMR* **2004**, *28* (4), 385–390. https://doi.org/10.1023/B:JNMR.0000015406.66725.30.

(68)   Wójcik, J.; Ruszczyńska, K.; Zhukov, I.; Ejchart, A. NMR Measurements of Proton Exchange between Solvent and Peptides and Proteins. *Acta Biochim. Pol.* **1999**, *46* (3), 651–663.

(69)   Weininger, U.; Modig, K.; Ishida, H.; Vogel, H. J.; Akke, M. Rotamer Jumps, Proton Exchange, and Amine Inversion Dynamics of Dimethylated Lysine Residues in Proteins Resolved by PH-Dependent 1H and 13C NMR Relaxation Dispersion. *J. Phys. Chem. B* **2019**, *123* (46), 9742–9750. https://doi.org/10.1021/acs.jpcb.9b06408.

(70)   Dynamics of Lysine Side-Chain Amino Groups in a Protein Studied by Heteronuclear 1H−15N NMR Spectroscopy | Journal of the American Chemical Society https://pubs-acs-org.ccl.idm.oclc.org/doi/10.1021/ja107847d (accessed 2021 -12 -10).

(71)   Zhang, Y. Z.; Paterson, Y.; Roder, H. Rapid Amide Proton Exchange Rates in Peptides and Proteins Measured by Solvent Quenching and Two-Dimensional NMR. *Protein Sci. Publ. Protein Soc.* **1995**, *4* (4), 804–814. https://doi.org/10.1002/pro.5560040420.