Examination of Surface Residues of Proteins using Cucurbit[6]uril and Ultraviolet Photodissociation Mass Spectrometry

Presented by Vishnu Srinivasa

in partial fulfillment of the requirements for graduation with the degree of

Bachelor of Science in Chemistry (Dean's Scholars Honors)

The University of Texas at Austin

Dr. Jennifer S. Brodbelt
Supervising Professor; Chair, Department of Chemistry

Dr. Graeme Henkelman
Honors Advisor in Chemistry

Abstract

Cucurbit[6]uril (CB[6]) is a large macrocyclic molecule that binds electrostatically to lysine, and has previously showed promise as a probe to tag the most surface-accessible regions of proteins. In order to validate this, in conjunction with tandem mass spectrometry, as an effective method to study protein surface structure, 193 nm ultraviolet photodissociation (UVPD) mass spectrometry was used to characterize the binding interactions between CB[6] and lysine residues of ubiquitin, and a data analysis protocol for this data was developed to figure out the most effective way to locate the binding sites. UVPD was successfully able to produce fragment ions that retained the non-covalent interactions between CB[6] and ubiquitin, and analysis of the pattern of N-terminal and C-terminal fragments that retained CB[6] demonstrated that Lys48 was a possible binding site. This residue is strongly involved in polyubiquitination and other protein interactions, demonstrating that UVPD in conjunction with CB[6] was successful in identifying residues that are highly accessible for interactions. As well, observation of species with multiple CB[6] adducts showed that a common method for determination of binding sites, subtracting fragment abundances of the uncomplexed species from the complexed species, is not necessarily useful for analysis of CB[6] binding due to the bulk of the molecule causing suppressed fragmentation elsewhere. However, preliminary results from this method also suggest potential binding to lysines around residue 30, another surface-accessible region, although this binding region is not confirmed and will require further investigation to determine whether fragmentation suppression is due to binding or interference from elsewhere in the sequence. This study overall provides a foundation for the integration of supramolecular electrostatic tagging and tandem mass spectrometry to locate structural

features of interest without relying on covalent tags or complex data analysis. As well, it demonstrates the utility of UVPD in opening up additional pathways for protein structural analysis, through its ability to retain non-covalent interactions even after fragmentation and its highly efficient and effective sequence coverage.

Introduction and Background

Many large biomolecules have very complex structures, involving multiple levels of large-scale organization. Proteins are essential to a variety of biological functions, and their structure essentially determines their function and effectiveness, but probing this structure is very challenging due to their complex folded secondary, tertiary, and quaternary structures. In addition, many techniques for studying protein structures *in vitro* require sample preparation that may change the observed structures, as the structure of a protein is highly sensitive to even small changes in its environment[1]. Despite these challenges, several techniques have emerged as important tools in the structural biology of proteins. The most common of these are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, which both allow for very direct and detailed analysis of structures at extremely high resolution[2]. However, both of these methods are limited by technical and cost challenges; some proteins cannot be crystallized effectively with X-ray crystallography, and the complex sample preparation can cause the protein to adopt a different conformation than its native solution-state conformation[2,3]. In addition, this method only provides a glimpse of static structure, without preserving solution-phase stabilizing interactions. While NMR spectroscopy can resolve many of

these issues, particularly with recent advances in high-resolution solid-state and solution NMR correlation methods, the data needed to gain atomic-scale or similarly high resolution is often very difficult to interpret, due to the massive number of signals[2,3]. Furthermore, many NMR methods have a functional upper mass limit of around 30-40 kDa, making it very difficult to measure large proteins and protein complexes in their native states without extensive sample preparation[2,3]. Recent advances in bioanalytical mass spectrometry and electron microscopy have made mass spectrometry-based methods a highly attractive alternative to X-ray crystallography and NMR spectroscopy for native protein structure analysis.

Bioanalytical mass spectrometry approaches most frequently employ tandem mass spectrometry strategies, in which the molecules in question are initially ionized through an ionization source into the spectrometer and sorted by their mass-to-charge ratio in the instrument's mass analyzer. Following this, individual ions are isolated and fragmented using various ion activation methods, and the fragment ions are analyzed by mass-to-charge ratios and quantified. For proteins, the characteristic fragmentation pattern occurs at certain bonds along the backbone (**Figure 2**), and the masses and relative abundances of the fragments can elucidate the amino acid sequence and structural motifs in the protein[1,2].

Most of the original studies in using tandem mass spectrometry for structural biology analysis and proteomics involved a bottom-up workflow. These workflows involve the enzymatic digestion into peptides of covalently-tagged or crosslinked proteins or protein complexes, followed by separation through chromatography or electrophoresis, mass spectral analysis, and individual peptide identification and sequencing[3]. The observed locations of the tags and crosslinks would thus reveal information about the conformation of the protein, and regions of

the protein involved in interactions with other molecules in solution or with the solvent[1-32]. These bottom-up methods only involve mass spectrometry of fairly small peptides, allowing for easy identification and analysis, but they provide an indirect picture of the protein's native environment. Another disadvantage of this is that the sample preparation process may not preserve some post-translational modifications, or they may be scrambled through the protein, resulting in loss of native structural information[3]. However, in recent years, it has become possible, with the improvement of mass spectrometry technology, to measure intact proteins this way without the need for digestion into peptides[4-25]. These top-down methods more closely approach direct examination of native structures, and collisional activation and photoactivation methods have made it possible to gain equal sequence coverage to bottom-up methods, particularly for smaller proteins[4-6,8,14].

Recent advances in electrospray ionization of proteins and high-resolution mass analyzers have made it possible to ionize intact, native-like proteins into the gas phase and directly analyze molecules in a native or native-like configuration, using patterns in the fragmentation of these native conformations to elucidate structural information and surface interactions. These new methods use electrospray ionization to take proteins from a native-like buffered solution into gas-phase ions, whose structures are presumed to approximate the solution-phase native structures due to the observed lower charge states when compared to denatured species[11,33]. The fragmentation patterns of these native-like proteins and protein complexes correlate with the surface accessibility and degrees of freedom and disorder in various parts of the protein[10,11,15,34-39], and can therefore be used to obtain structural information – for instance, the rigid residues near a bound ligand may result in suppressed fragmentation around the

ligand, revealing the binding site, or regions of increased fragmentation may be highly solvent-accessible[11, 20-21]. Some of the most common ion activation methods for fragmentation of intact proteins are collisional methods, such as collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD)[22,26-32]. These involve the collision of the protein molecules with an inert molecule, converting the kinetic energy of the ion to internal energy and promoting fragmentation through a stepwise dissociation mechanism. HCD has been used in a variety of native and top-down methods and results in significant sequence coverage for smaller proteins and protein complexes, but especially in large proteins the dissociation mechanism often results in much higher sequence coverage near the protein termini rather than in the inner portions of the sequence[22]. As such, new photoactivation methods, including ultraviolet photodissociation (UVPD), have been developed in recent years to aid in efficient characterization throughout the entire sequence[2,15,26,27]. In UVPD, the protein molecules are dissociated through being struck with a beam of UV laser light, which transfers energy to the molecule and allows for dissociation of peptide bonds throughout the sequence. UVPD allows access to a much larger range of fragmentation pathways than HCD, and produces backbone fragments of many more types – using the common nomenclature (**Figure 2**), UVPD often produces a, b, c, x, y, and z type ions, allowing for increased possibilities of fragmentation and thus more detailed coverage than HCD, which produces predominantly b/y type ions. This increase in the available fragmentation pathways through even a single laser pulse, in contrast to the stepwise dissociation that is common in collisional activation methods, allows UVPD to provide sequence coverage throughout the sequence, rather than primarily near the termini[15,26,27]. Furthermore, UVPD can be used to study ligand binding and other non-covalent

binding interactions, as prior studies have shown that UVPD is able to retain non-covalent interactions even after fragmentation[15]. Analysis of the fragment ions that retain non-covalent interactions (holo ions) can yield information about the binding site of various non-covalently bound molecules, as only fragments that contain the residues at the binding site are able to retain the non-covalent interaction, so the pattern of which holo ions are obtained yields information about binding sites by determining where holo ions stop appearing[15]. As well, the fragmentation patterns generated by UVPD have been previously shown to reflect conformational changes associated with non-covalent interactions in a variety of protein-ligand interactions[15], allowing for localization of binding sites using UVPD. Due to these advantages, UVPD combined with native mass spectrometry is an ideal alternative and complement to NMR spectroscopy and X-ray crystallography for the efficient characterization of a wide variety of native proteins and protein complexes, as well as to study the conformational changes associated with ligand binding[15,26,27].

Native mass spectrometry has also led to the development of a variety of methods to more efficiently probe the structures of proteins through non-covalent interactions with a variety of molecular probes, followed by native mass spectrometry of the non-covalent complexes and location of the binding sites to indirectly determine information about the structure of the original protein[15,33]. Indirect methods like these provide a quick and convenient way to gain structural information about specific portions of a protein's structure and orientation without relying on fragmentation intensities. These are especially relevant for locating the most surface-accessible side chains, as these are frequently involved in protein-protein interactions and are typically easy to non-covalently tag due to their prominence[40].

One such supramolecular probe is cucurbit[6]uril (CB[6]), a large (~ 1 kDa) macrocyclic molecule consisting of six glycoluril subunits linked through methylene bridges (**Figure 1**). The orientation of the carbonyl oxygens on both sides of the ring results in a highly electronegative inner cavity, which has allowed CB[6] to be used as a host in host-guest complexes for supramolecular catalysis, as well as to electrostatically bind positively charged species such as metal and ammonium cations[41]. This electronegativity also makes it a prime candidate for tagging the side chains of highly accessible positively charged residues (lysine and arginine) to act as a probe for analysis of protein surface structure through tandem mass spectrometry[42]. Previous studies involving top-down, CID tandem mass spectrometry of CB[6] and ubiquitin – a small (8.5 kDa) protein that occurs in almost all eukaryotic cells[43] – have demonstrated that CB[6] selectively binds to lysine side chains over arginine side chains, and also preferentially binds to Lys48 and, sometimes, Lys63 of ubiquitin[43]. This result is encouraging for the usage of CB[6] as a surface probe for mass spectrometry, as these two residues are the most prominent and are the most involved in protein-protein interactions. In particular, these are the two possible residues involved in polyubiquitination, in which ubiquitin molecules covalently bond to each other and the resultant complex covalently bonds to another target protein in order to tag it for a specific fate in the cell[43,44]. Lys48-bound polyubiquitin chains are far more common than Lys63-bound chains, so Lys48 can be considered to be the most highly accessible residue of ubiquitin for protein interactions, and tagging with CB[6] has previously been shown to identify this residue[42].

In this study, a method was developed to study the mechanism and location of the binding of CB[6] to native-state ubiquitin using UVPD, as a first step towards the development of a UVPD

and supramolecular chemistry-based method to efficiently locate regions of protein surfaces that are likely to be involved in protein-protein or protein-ligand interactions, and to efficiently characterize the surface structure of native proteins. It was observed that native electrospray produced ions containing a single CB[6] adduct as well as ions containing two CB[6] adducts, and that native UVPD was able to recover the same Lys48 binding site that wasreported previously using CID, demonstrating this method's utility for locating the primary surface-accessible residues. As well, UVPD mass spectrometry was used to probe the previously mentioned species containing multiple CB[6] adducts.

Methods

Sample Preparation: Ubiquitin and cucurbit[6]uril were obtained commercially. Cucurbit[6]uril was dissolved in 1 M ammonium acetate in DI H_2O to a final concentration of 1 mM. The final samples were prepared by adding ubiquitin to 50 mM ammonium acetate (for native conditions) and 50/50 MeOH/H_2O with 1% formic acid (for denaturing conditions) for a final concentration of 15 µM ubiquitin. To form the Ub-CB[6] complex, CB[6] solution, as prepared previously, was added in 1x molar equivalence and 10x excess to the 15 µM ubiquitin solutions and allowed to incubate at room temperature for 1 hour prior to measurement.

Instrumentation: All mass spectra were acquired on a Thermo Scientific Instruments Orbitrap Elite mass spectrometer (Bremen, Germany) in positive mode with a 193 nm Coherent Excistar excimer laser (Santa Cruz, CA), to enable UVPD in the HCD cell at 10 mTorr He pressure (**Figure 3**). HCD was performed using the instrument's normal HCD capacity at various energies ranging

from 15 to 35% normalized collision energy (NCE). UVPD was performed using a single laser pulse at 2.5 mJ and 3.0 mJ depending on the sample. For native species, the 6+ charge state of both unbound ubiquitin and the ubiquitin complex was selected using an isolation window of 10 m/z and AGC target of 1e5. For denatured species, charge states from 7+ to 13+ were selected using the same isolation parameters. In addition, for native species the observed peak corresponding to a complex with a 1:2 stoichiometric ratio of ubiquitin: CB[6] was selected using the same parameters. For each experiment, 100 scans were averaged using dynamic signal averaging at a resolution of 240,000.

Data Analysis: Mass spectra were deconvoluted using the Thermo Xtract algorithm accessed through Xcalibur with a S/N ratio of 3. The resulting deconvoluted spectra were processed using UV-POSIT algorithm[47], searching for common ion types in HCD (a, b, y) and UVPD (a, b, c, x, y, z) to fully characterize the fragmentation. Furthermore, ligand-containing fragment ions (holo ions) were searched by looking for the characteristic mass shift of CB[6] bound to the protein. To determine the binding sites, for each residue all N-terminal fragments and C-terminal fragments were summed, and the resulting fragment abundances from the complex were subtracted from the fragment abundances of uncomplexed ubiquitin in order to observe regions of fragmentation suppression or increase. As well, the N-terminal and C-terminal holo ions were studied to determine the pattern of holo fragmentation and elucidate the locations of bound CB[6].

Figures:

Note: figures are referenced throughout the text above and below.

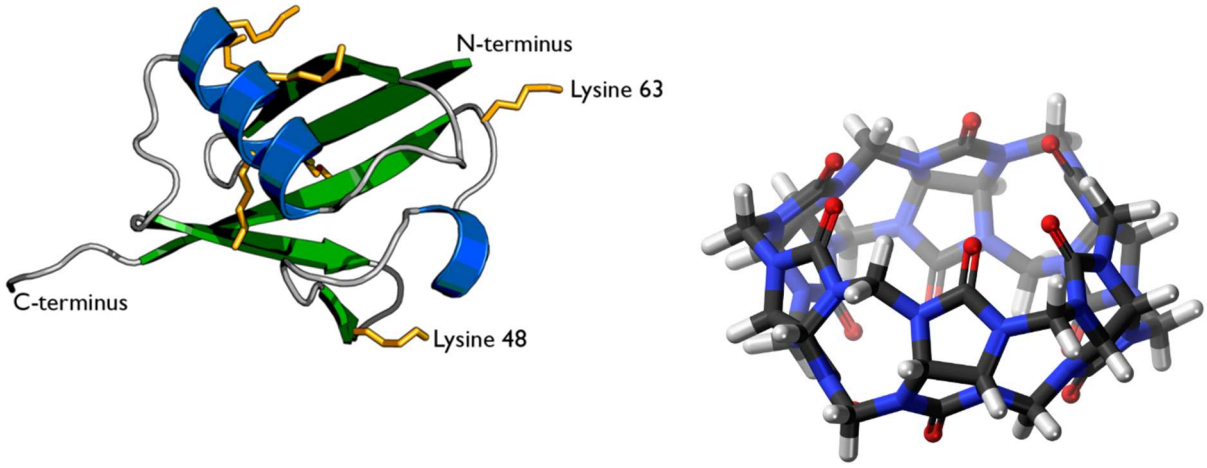


Figure 1. Crystal structure of ubiquitin (PDB code 1UBQ) (left) and molecular structure of cucurbit[6]uril (right).

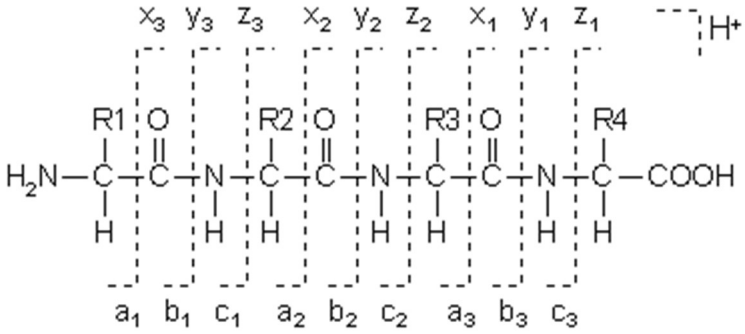


Figure 2. Backbone fragment nomenclature for tandem mass spectrometry of proteins and peptides, where each R group represents the side chains[46].

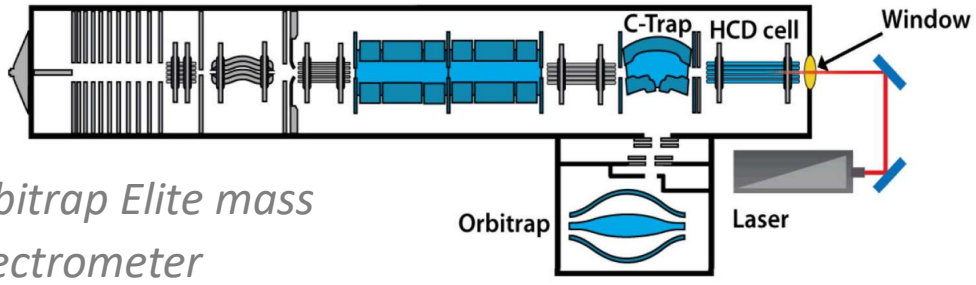


Figure 3. Schematic of Orbitrap Elite mass spectrometer with Coherent Excistar laser in HCD cell.

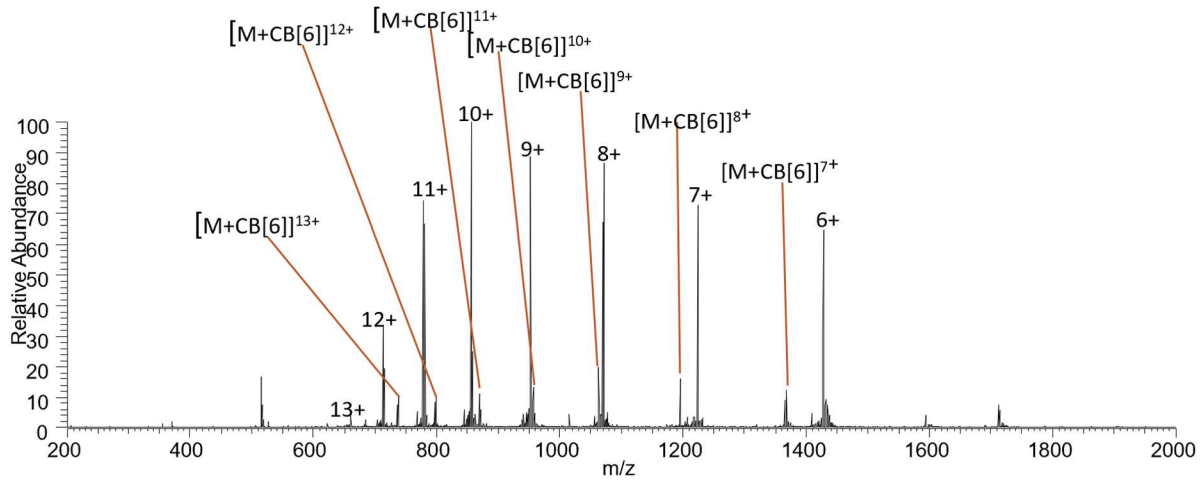


Figure 4. Representative mass spectrum of denatured ubiquitin in complex with CB[6], showing retained CB[6] adducts.

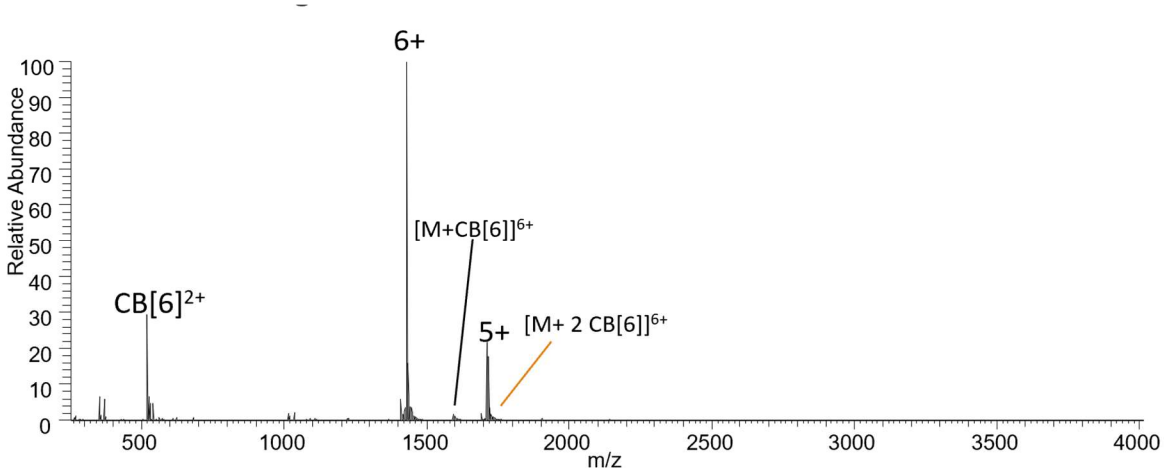


Figure 5. Representative mass spectrum of native ubiquitin in complex with CB[6], showing retained CB[6] adducts and a double-adducted species.

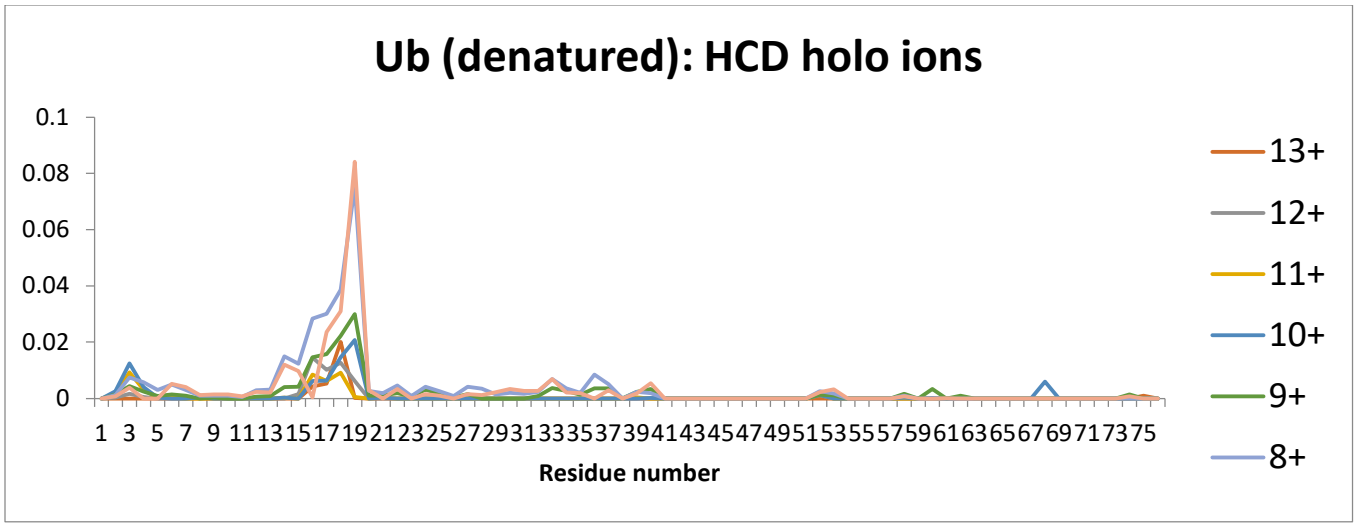


Figure 6. Holo ions produced by HCD of denatured ubiquitin complexed with CB[6].

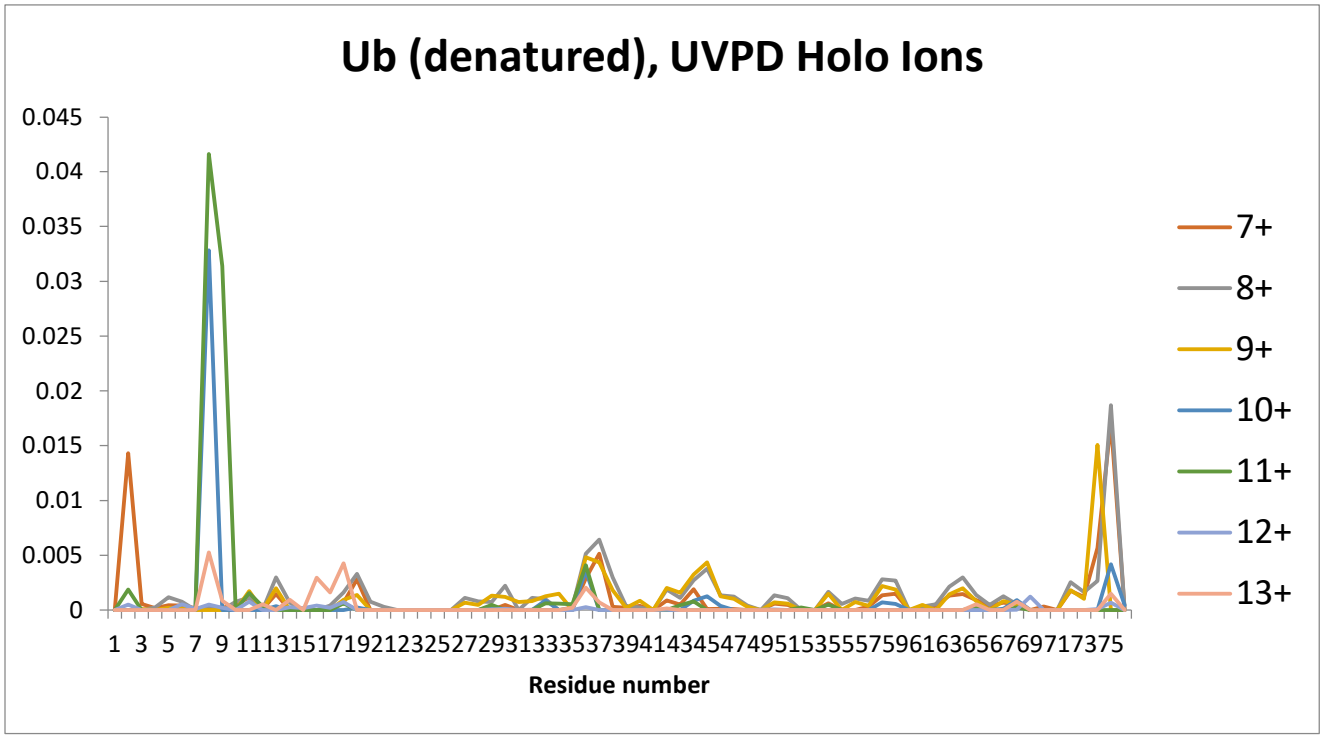


Figure 7. Holo ions produced by UVPD of denatured ubiquitin complexed with CB[6]. UVPD provides more holo ions than HCD.

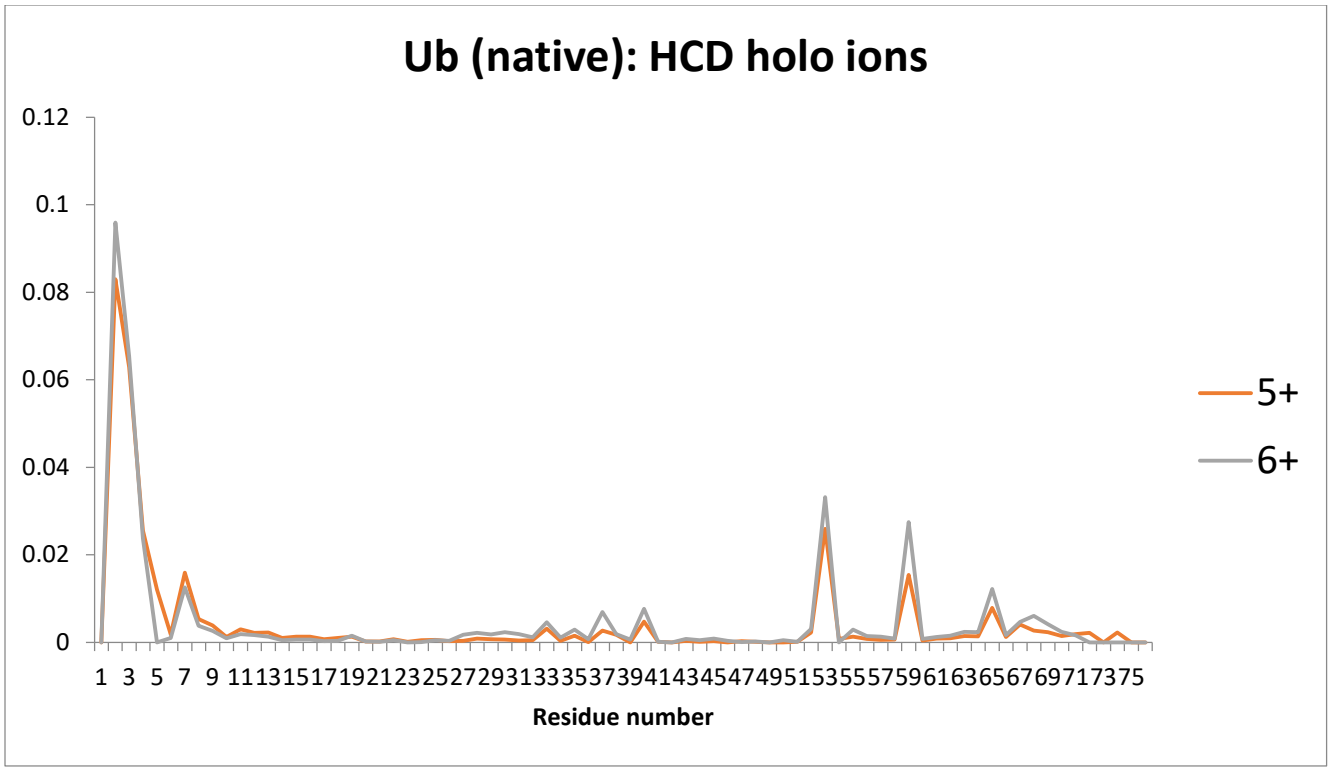


Figure 8. All ions produced by HCD of native ubiquitin complexed with CB[6].

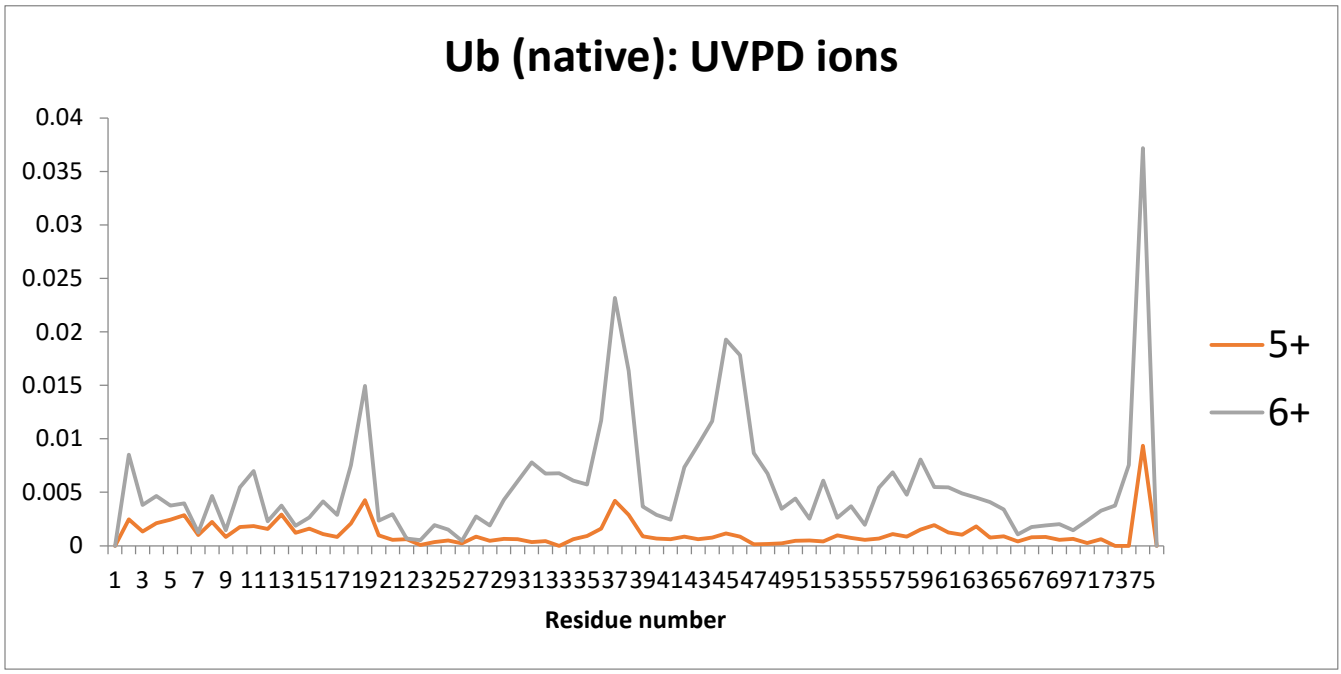


Figure 9. All ions produced by UVPD of native ubiquitin complexed with CB[6]. UVPD produces a more even fragment distribution than HCD.

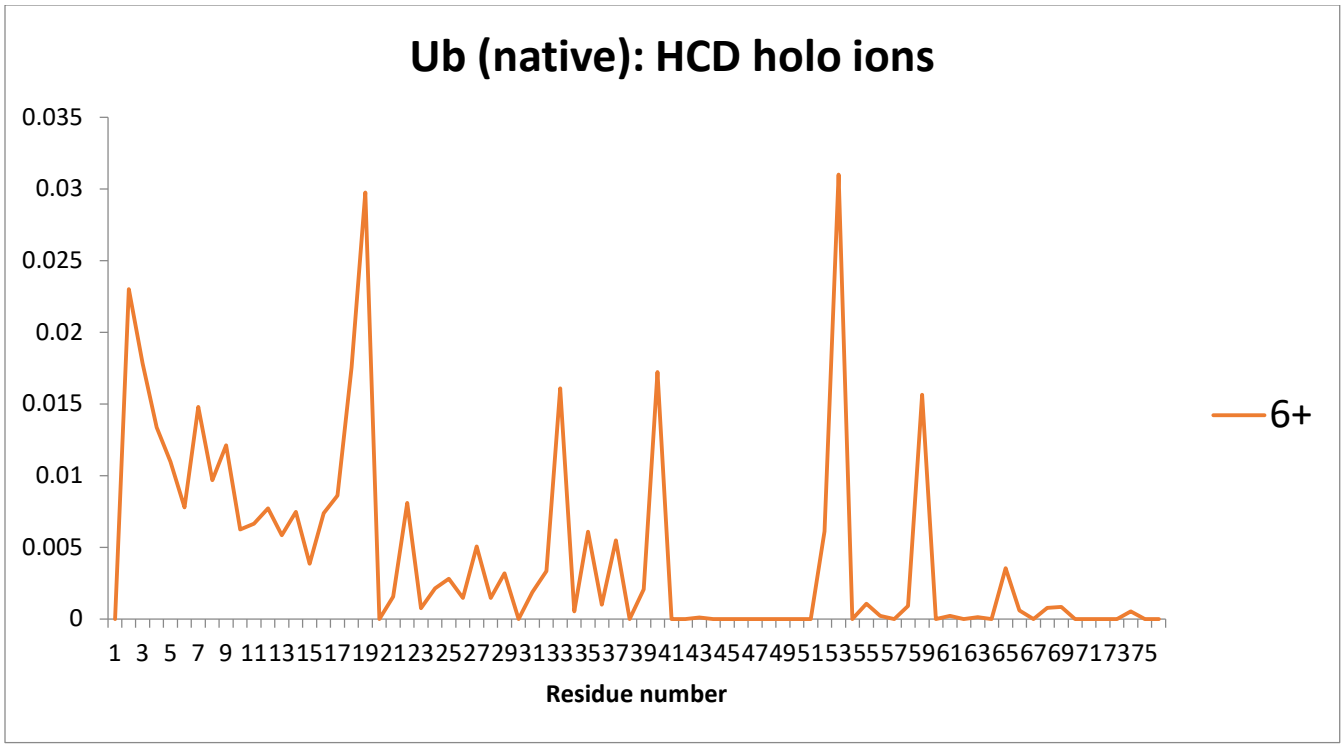


Figure 10. Holo ions produced by HCD of native ubiquitin complexed with CB[6].

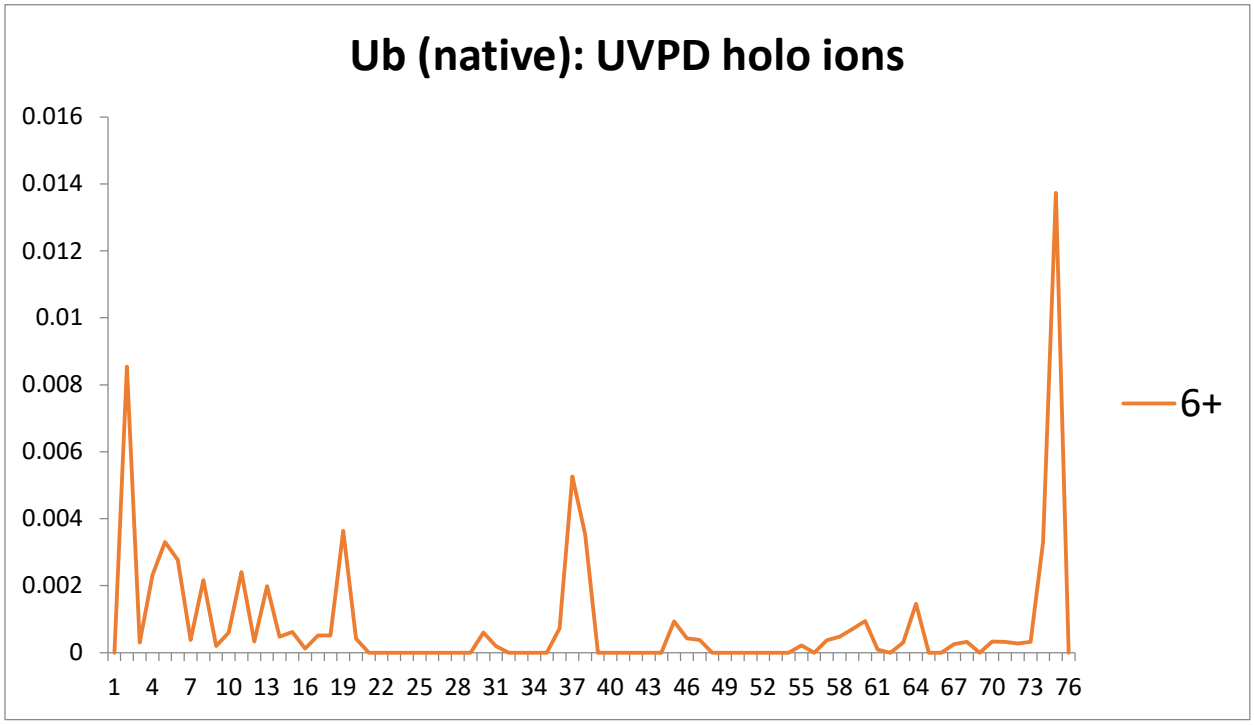


Figure 11. Holo ions produced by UVPD of native ubiquitin complexed with CB[6]. UVPD yields more holo ions than HCD, and has more noticeable peaks in the holo ions when compared to denatured samples, which have more consistent, but lower, holo fragment distribution.

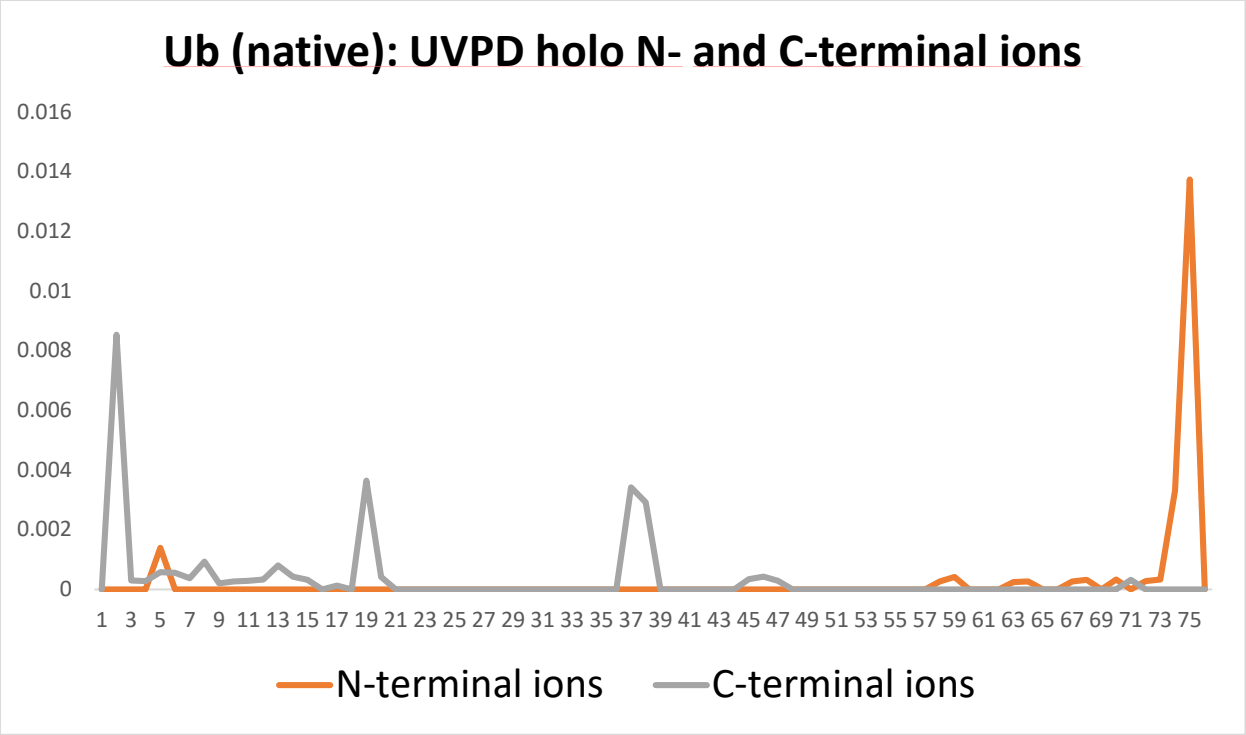


Figure 12. UVPD holo ions from native ubiquitin complexed with CB[6], split into N-terminal and C-terminal ions.

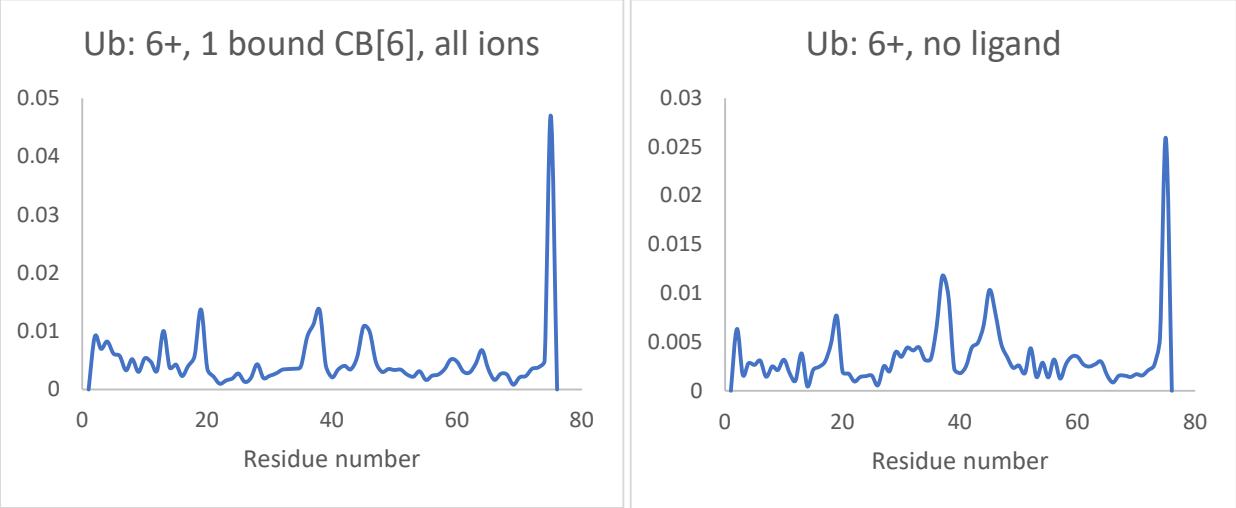


Figure 13-14. Plot of all UVPD fragment ions (including apo and holo ions) for native ubiquitin in 6+ charge state with bound CB[6] (left) and without bound CB[6] (right) vs. residue number.

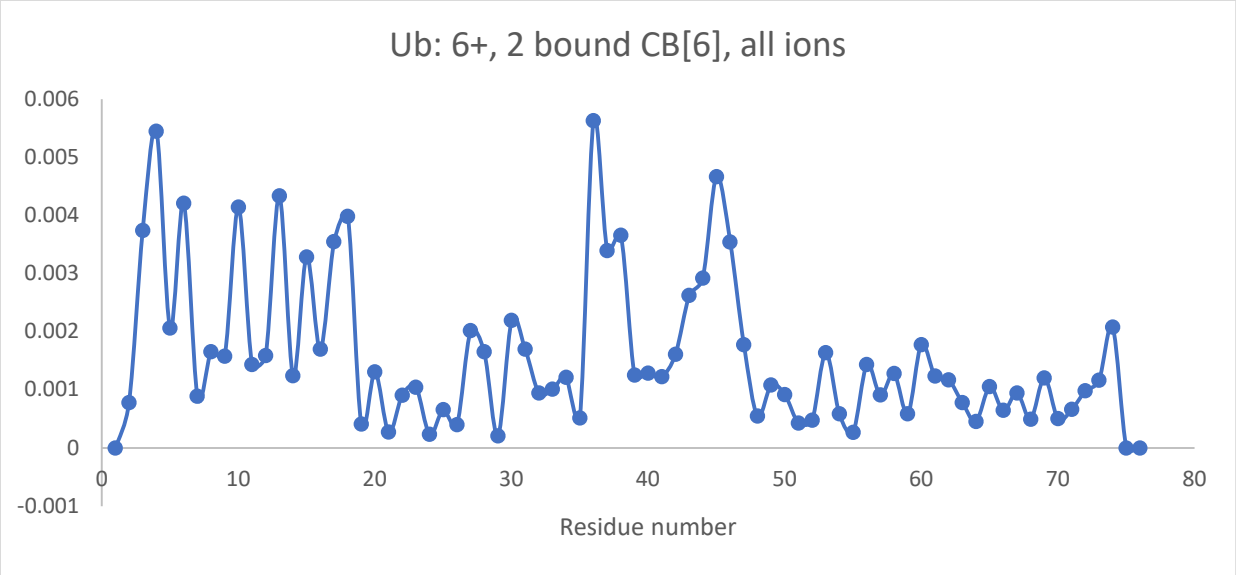
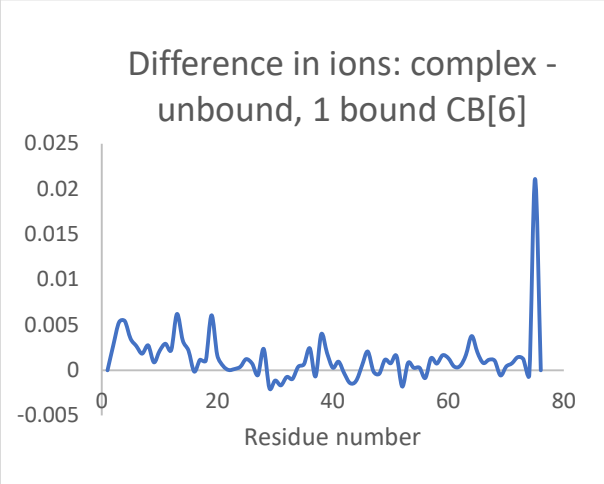


Figure 15. Plot of all UVPD fragment ions (including apo and holo ions) for native ubiquitin in 6+ charge state with two bound CB[6] vs. residue number.



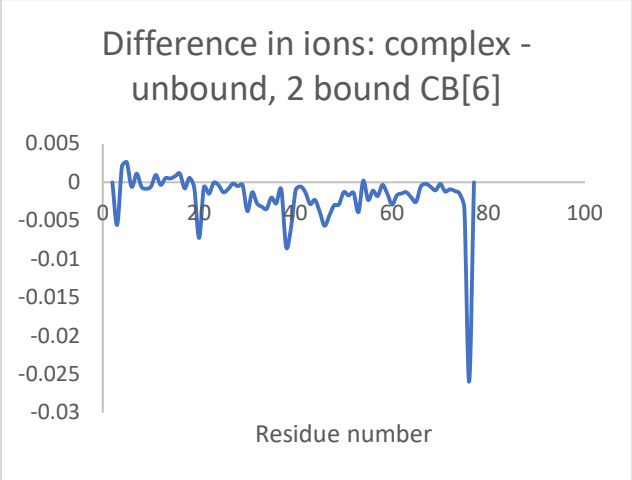


Figure 16-17. Plot of the difference in ion abundance between complexed and uncomplexed species, for species with one bound CB[6] and two bound CB[6], vs. residue number.

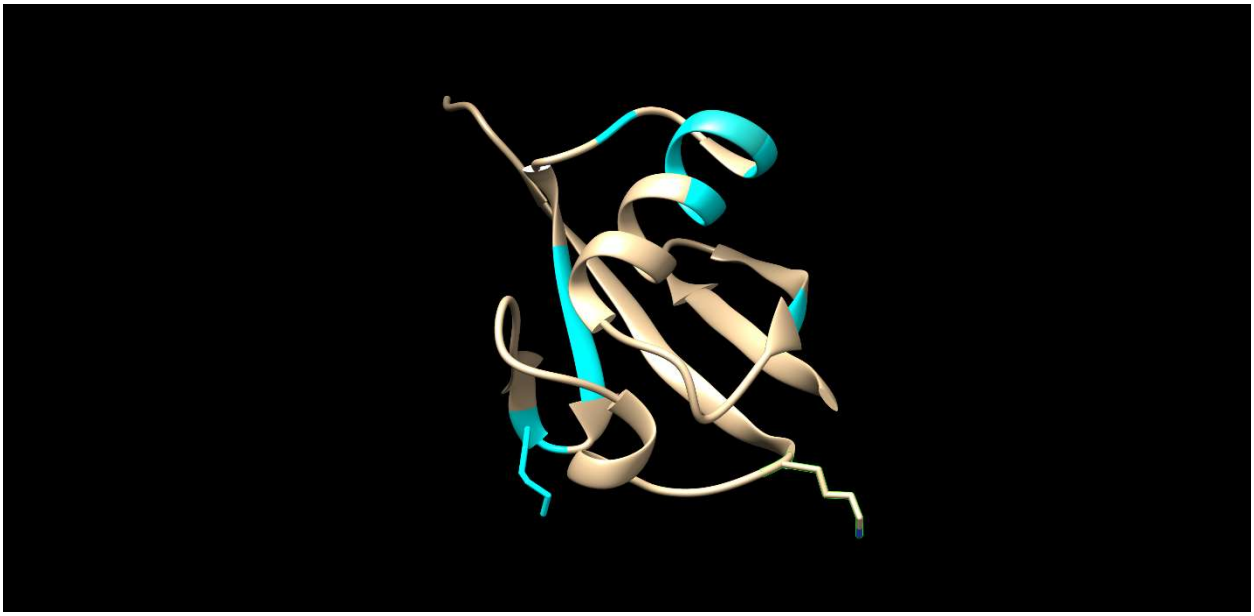


Figure 18[45]. Crystal structure of ubiquitin (1UBQ) with regions with suppressed fragmentation highlighted in blue. The region around residue 30 is shown in the upper alpha helix, while the suppressed region near Lys48 is shown in the bottom left. Lys48 (bottom left) and Lys63 (bottom right), the residues involved in polyubiquitination, have their side chains shown.

Results and Discussion

In this study, UVPD and HCD were used to characterize the binding locations of CB[6] onto surface lysine side chains of ubiquitin, as a first step towards using this method to characterize additional, larger and less well-characterized proteins. Denatured and native species were studied in order to compare the two and determine locations with different abundances in the native species; the denatured species was used as a control to determine whether the site-specific binding was characteristic only of the native species or if it occurred as well in denatured species. This process was repeated using two different methods to determine the binding site: for the first trial the binding site of CB[6] to native ubiquitin was located by analyzing the pattern of holo fragment ions and determining which fragments were able to

retain CB[6], while for the second trial the binding site was located by comparing the fragmentation of the complex to the fragmentation of uncomplexed ubiquitin and determining which portions of the sequence had suppressed or enhanced fragmentation. This second data analysis method was also used to study complexes containing two CB[6] molecules bound to each ubiquitin molecule. Electrospray ionization of a mixture of ubiquitin and CB[6] from ammonium acetate buffer was used to produce the 6+ and 5+ charge state of ubiquitin, demonstrating that the produced ions retain native-like conformations (representative spectra for both native and denatured samples are shown above in **Figures 4 and 5**). As well, non-covalent complexes, preserving the interactions between CB[6] and ubiquitin, were also produced in the 6+ charge state, with a species with two CB[6] adducts observed in the 6+ charge state. The most abundant charge state for both the complex and the protein was the 6+ charge state, so this charge state was selected for isolation and tandem mass spectrometry. In the denatured spectra, a full charge envelope from 7+ to 13+ was observed in both the complex and the uncomplexed protein, so all of these charge states were selected for isolation and tandem mass spectrometry. **Figures 6-9** show the distribution of fragments throughout the sequence for both UVPD and HCD. The UVPD spectra exhibit rich fragmentation containing multiple charges of each ion type, while the HCD spectra (with higher abundance than the UVPD spectra) demonstrate many b/y type ions, as expected from HCD. While HCD fragmentation was much richer near the N- and C- termini, UVPD demonstrated rich fragmentation throughout the sequence, and a complete array of holo ions (plotted separately in **Figures 10-13**). The holo ions discovered demonstrate the increased utility of UVPD over HCD, and the additional ways in which UVPD can be used to study binding interactions and

protein structure. The combination of mass spectrometry and supramolecular chemistry described in this approach, therefore, is particularly suited to UVPD, as UVPD produces many more holo fragment ions and significantly higher sequence coverage. UVPD of the denatured uncomplexed ubiquitin demonstrated sequence coverage of up to 100%, demonstrating the utility of UVPD for efficient characterization throughout the sequence of the protein. In addition, UVPD of both the denatured and native complexes yielded holo fragment ions (i.e. ions containing bound CB[6]) throughout the sequence. In the denatured complexes, the holo ions were distributed throughout the sequence, suggesting that fragments throughout the sequence retained CB[6], while the holo ions from the native complexes were somewhat localized, suggesting that only fragments that contain the specific binding residue were able to retain CB[6] and therefore suggesting that CB[6] binds preferentially to certain residues. This shows that there is a noticeable difference between the fragmentation pattern and the location of holo fragment ions between the native species and the denatured species, suggesting that site-specific binding of CB[6] is characteristic only of the native species, and that the structure of the protein is confirmed to be important for the binding environment and binding site. In addition, this demonstrates that the fragmentation pattern and the pattern of the identified holo ions (as used to study some of the experimental results) reveals information about the locations of the bound CB[6] and about the conformational state of the protein (denatured or native).

To process the fragmentation data, all spectra were run through the UV-POSIT algorithm[47], an algorithm designed specifically to process UVPD mass spectral data, after deconvolution, which assigned each fragment to its type and the fragmentation site. UV-POSIT also normalizes

fragment abundance to the total ion current intensity, so samples with different total ion current can be directly compared as their abundances are normalized to relative abundances. To evaluate the data, all types of ions at each fragmentation site were summed and the resulting total ions were plotted against the fragmentation site to gain an overview of the relative abundance of fragmentation at each location and to determine the relative fragmentation propensity between each pair of adjacent amino acids. This was also repeated for the ubiquitin-CB[6] complexes. In addition, for those complexes, the apo and holo ions were separated and also plotted individually in the same way in order to observe any possible patterns. For the first set of experiments, the binding site was located by observing the holo ions and plotting them against fragmentation site. Comparing the native samples with the denatured samples (**Figures 6-11**), the holo fragments obtained from denatured samples tend to be more evenly spread throughout the sequence, suggesting that fragments throughout the sequence retain bound CB[6] as there are no noticeable gaps in holo fragmentation where observed holo ions are absent. In contrast, the holo ions observed from the native samples demonstrated much more localized holo fragmentation with noticeable gaps in observed holo ions in various parts of the sequence, suggesting that the binding of CB[6] to native samples only occurs at specific, highly accessible regions. To locate these regions, the N-terminal holo ions (i.e. all holo a, b, c ions) were summed and plotted, and the C-terminal holo ions (i.e. all holo x, y, z ions) were separately summed and plotted on the same graph (**Figure 12**). It is possible to see a large gap between where C-terminal holo ions stop and N-terminal holo ions begin, from approximately residue 40 to 58. Only C-terminal ions prior to residue 40 contain bound CB[6], and only N-terminal ions after residue 58 contain bound CB[6]. This implies that

the binding site is somewhere within the residue range of 40 to 58, as N-terminal ions beyond residue 58 contain all residues 1 to 58 while C-terminal ions prior to residue 40 contain all residues 40 to 76, and this is the region in which these two overlap. Between residues 40 to 58, there is only one lysine residue, residue 48, showing that this is the binding site of CB[6] in the native-like complexes (as CB[6] has been previously shown to only bind to lysine residues). This result is encouraging, as Lys48 is one of the most surface-accessible residues, and is also the residues involved in polyubiquitination. Previous results have also demonstrated that this is one of the known binding sites. This result shows that UVPD was successful in identifying regions of the ubiquitin sequence that are highly active in interactions. Thus, it is possible to see the utility of UVPD for this type of measurement, and that a combination of supramolecular chemistry and mass spectrometry can be used to locate highly accessible regions of the protein structure that are involved in important interactions.

To study this binding site location with a different data analysis method, and to examine the previously identified species that contained two CB[6] adducts bound to one molecule of ubiquitin, this experiment was repeated and the binding site was located through comparison with all fragments, rather than just holo fragments. For these experiments, the mixture of ubiquitin and CB[6] in a 1:10 stoichiometric ratio was sprayed into the mass spectrometer, and the 6+ charge state of uncomplexed ubiquitin, ubiquitin with one bound CB[6], and ubiquitin with two bound CB[6] were selected and isolated for UVPD. To determine the binding site, all UVPD fragment ions (including both holo and apo ions) at each fragmentation residue were summed for each sample and plotted against their residue position to determine the relative fragmentation propensity at each residue (**Figures 13-15**). For each of the complexed samples

(with one or two adducts), regions with increased or suppressed fragmentation in comparison to the uncomplexed species were determined by subtracting the relative fragment abundances of the uncomplexed sample from the relative fragment abundances of the complexed sample and plotting the resulting difference against the fragmentation site (**Figures 16 and 17**). Regions where this plot is below 0 represent regions of the sequence where fragmentation propensity is suppressed in the complexed species, while regions where this plot is above 0 represent regions of the sequence where fragmentation propensity is enhanced in the complexed species. Previous research in UVPD fragmentation of native-like proteins and protein complexes has shown that suppression and enhancement of fragmentation in non-covalent complexes is dependent on two main effects: increased conformational flexibility in some regions resulting in higher fragmentation, and non-covalent interactions decreasing conformational flexibility and requiring additional disruption of interactions in order to yield fragments, resulting in reduced fragmentation. As such, regions with suppressed fragmentation are expected to be involved in more of the relevant non-covalent interactions, as these interactions make it more difficult for UVPD to generate fragments[10,11,15,34-39]. **Figure 18** shows the crystal structure of ubiquitin with approximate regions highlighted in blue for suppressed fragmentation in the complexed species. In the native samples with a single bound CB[6], there is suppressed fragmentation at various points between residues 40 and 50, as well as around residue 30. Between residues 40 and 50, there is only one lysine residue, Lys48, reconfirming that this is a possible binding site of CB[6]. The erratic suppression around this residue is likely due to the bulk of CB[6] affecting fragmentation several positions away from it, but as Lys48 is the only possible binding site in this region, it is possible to confirm this as a binding site. However, the region of suppressed

fragmentation around residue 30 is immediately adjacent to a very lysine-rich region of the protein, suggesting that there does exist some binding of CB[6] to some or all of these lysines. Since this region is so rich in lysines, the suppression of fragmentation due to small amounts of binding to each of the lysines in this region likely would result in suppressed fragmentation throughout the region, as observed here. As such, the binding to each of these sites, if it is occurring, is likely not as significant or as abundant as binding to Lys48. This region is located on a portion of an alpha helix and is also highly surface-accessible, as seen in the crystal structure, suggesting that CB[6] can also bind to other portions of the protein surface besides the expected highly important residues. This result would be encouraging for further surface analysis using this method, as it demonstrates the utility of CB[6] to tag lysines all over the surface of a protein, which is critical in analysis of the structure and surface-accessible regions of the sequences of larger proteins. However, this is not certain; the suppression of fragmentation here (and elsewhere in the sequence) may also partially be due to bulk of CB[6] molecules bound at different sites; CB[6], at nearly 1 kDa in mass and forming a large ring structure, is large enough to potentially interact with these residues even when bound to other portions of the protein, so this fragmentation suppression may or may not be due to binding in this region of the protein. When observing the complexes with two bound CB[6], fragmentation is suppressed in general throughout the sequence, without any particularly strong pattern in important regions. This suggests that the bulk from two or more CB[6] molecules is enough to suppress fragmentation throughout the sequence, rather than at the specific sites near the binding site. One potential explanation for this is that there may exist multiple different combinations of two binding sites, and that there are a variety of species present with different

binding sites. However, this is not currently possible to confirm, and this result may potentially be due only to bulk. As such, this complex reveals that observing regions of fragmentation suppression to locate binding of CB[6] may be less informative than observing the pattern of UVPD holo ions, as CB[6] can cause nonspecific suppression of fragmentation. This is also supported by the action of CB[6] as a charge-reducing agent, which can prevent the dissociation of bonds in various places throughout the sequence. Thus, observing these species demonstrates that the best method for analysis of binding locations of CB[6] using UVPD is the location of holo fragment ions and the determination of which regions lack holo fragmentation.

Conclusion

This study confirms the utility of 193nm UVPD for the analysis of the binding of CB[6] to surface-accessible lysine residues of native-like proteins by using ubiquitin as a model system for method development. UVPD is able to retain the non-covalent interactions between CB[6] and ubiquitin, allowing for location of the binding site by analysis of the location and abundance of N- and C-terminal holo fragment ions. This result also shows that, while evaluation of fragmentation suppression when compared to the uncomplexed species is often used to probe non-covalent interactions, this method is not always useful for CB[6], as it is a bulky molecule that can suppress fragmentation in various places throughout the sequence. Studying the binding of CB[6] to ubiquitin using the pattern of holo ions revealed that CB[6] binds to Lys48, the primary and most common residue involved in polyubiquitination. Analyzing regions with fragmentation suppression also yielded a potential binding site in the lysine-rich

region near residue 30, which is also surface-accessible. This result, while unconfirmed due to potential interference from the bulk of CB[6], suggests that other surface-accessible regions beyond the most important ones in the protein may also bind to CB[6], further confirming its utility for protein surface analysis. This is an encouraging result towards the use of this hybrid method for identifying accessible regions of the surfaces of uncharacterized proteins. Future work on this project to ensure the robustness of this method would include further probing of the binding interactions and binding kinetics of CB[6] to lysine, and testing this method on other model proteins. Overall, this method shows the utility of UVPD for surface structure analysis and makes UVPD a compelling approach for a variety of applications in structural biochemistry.

References

1. Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2015**, *137* (28), 9128–9135.

2. Cammarata, M. B.; Brodbelt, J. S. *Chem. Sci.* **2015**, 6, 1324.

3. Konermann, L.; Vahidi, S.; Sowole, M. A. *Analyt. Chem.*, **2014**, 86, 213-232.

4. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc.* **2009**, 131, 12801–12808.

5. Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. *J. Mass Spectrom.* **2004**, 39, 322–328.

6. Wang, G.; Kaltashov, I. A. *Anal. Chem*. **2014**, 86, 7293–7298.

7. Stefanowicz, P.; Kijewska, M.; Szewczuk, Z. *Anal. Chem*. **2014**, 86, 7247–7251.

8. Bobst, C. E.; Kaltashov, I. A. *Anal. Chem.* **2014**, 86, 5225–5231.

9. Nagy, K.; Redeuil, K.; Rezzi, S. *Anal. Chem.* **2009**, 81, 9365–9371.

10. Robinson, E. W.; Leib, R. D.; Williams, E. R. *J. Am. Soc. Mass Spectrom*. **2006**, 17, 1470–1480.

11. Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem*. **2007**, 76, 167–193.

12. Hall, Z.; Hernández, H.; Marsh, J. A.; Teichmann, S. A.; Robinson, C. V. *Structure* **2013**, 21, 1325– 1337.

13. Zhang, Z.; Browne, S. J.; Vachet, R. W. *J. Am. Soc. Mass Spectrom*. **2014**, 25, 604–613.

14. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc*. **2008**, 130, 11574–11575.

15. O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2014**, 136, 12920–12928.

16. Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009**, 20, 1851–1858.

17. Simmons, D. A.; Dunn, S. D.; Konermann, L. *Biochemistry* **2003**, 42, 5896–5905.

18. Wang, F.; Tang, X. *Biochemistry* **1996**, 35, 4069–4078.

19. Wright, P. J.; Zhang, J.; Douglas, D. J. *J. Am. Soc. Mass Spectrom.* **2008**, 19, 1906–1913.

20. Zhou, M.; Jones, C. M.; Wysocki, V. H. *Anal. Chem.* **2013**, 85, 8262–8267.

21. Breuker, K.; Brüschweiler, S.; Tollinger, M. *Angew. Chem. Int. Ed.* **2011**, 50, 873–877.

22. Yin, S.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2010**, 21, 899–907.

23. Ly, T.; Julian, R. R. *J. Am. Chem. Soc.* **2010**, 132, 8602–8609.

24. Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F. *J. Am. Soc. Mass Spectrom.* **2014**, 25, 343–350.

25. Modzel, M.; Stefanowicz, P.; Szewczuk, Z. *Rapid Commun. Mass Spectrom.* **2012**, 26, 2739–2744.

26. Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *Anal. Chem.* **2014**, 8, 2185–2192.

27. Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc*. **2013**, 135, 12646–12651.

28. Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L. *Mol. Cell. Proteomics* **2013**, 12, 3465–3473.

29. Skinner, O. S.; Catherman, A. D.; Early, B. P.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. *Anal. Chem.* **2014**, 86, 4627-4634.

30. Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. *Anal. Chem.* **2013**, 85, 11163–11173.

31. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 9528–9533.

32. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, 120, 3265–3266.

33. Heck, A. J. R. *Nat. Methods* **2008**, 5, 927–933.

34. Cammarata, M. B.; Brodbelt, J. S. *Chem. Sci.* **2015**, 6, 1324–1333.

35. Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci.* **2008**, 105, 18145–18152.

36. Li, H.; Wongkongkathep, P.; Orden, S. L. V.; Loo, R. R. O.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2014**, 25, 1–9.

37. Zhang, H.; Cui, W.; Gross, M. L.; Blankenship, R. E. *FEBS Lett.* **2013**, 587, 1012–1020.

38. Cui, W.; Rohrs, H. W.; Gross, M. L. *Analyst* **2011**, 136, 3854–3864.

39. Warnke, S.; Baldauf, C.; Bowers, M. T.; Pagel, K.; von Helden, G. *J. Am. Chem. Soc*. **2014**, 136, 10308-10314.

40. Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. *J. Comput. Chem.* **2004**, 25(5), 712-724.

41. Assaf, K. I.; Nau, W. M. *Chem. Soc. Rev.,* **2015**, 44, 394.

42. Lee, J.W.; Heo, S. W.; Lee, S. J.; Ko, J. Y.; Kim, H.; Kim, H.I. *J. Am. Soc. Mass Spectrom.*, **2013**, 24(1), 21-29.

43. Pickart, C. M. *Annu. Rev. Biochem*., **2001**, 70, 503-533.

44. Vijaykumar, S..; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.,* **1987**, 194, 536-544.

45. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, 25(13), 1605-12.

46. Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal. Chem.* **1987**, 59(21), 2621-2625.

47. Rosenberg, J.; Parker, W. R.; Cammarata, M. B.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2018**, 29(6), 1323-1326.