

SANDIA REPORT  
SAND2005-1698  
Unlimited Release  
Printed March 2005

# Intra-molecular Cross-linking of Acidic Residues for Protein Structure Studies

Petr Novak, Joe Schoeniger, Malin Young, and Gary H. Kruppa

Prepared by  
Sandia National Laboratories  
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,  
a Lockheed Martin Company, for the United States Department of Energy's  
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.



**Sandia National Laboratories**

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

**NOTICE:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from  
U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831

Telephone: (865) 576-8401  
Facsimile: (865) 576-5728  
E-Mail: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)  
Online ordering: <http://www.doe.gov/bridge>

Available to the public from  
U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Rd  
Springfield, VA 22161

Telephone: (800) 553-6847  
Facsimile: (703) 605-6900  
E-Mail: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



## **Intra-molecular Cross-linking of Acidic Residues for Protein Structure Studies**

**Petr Novak, Joe Schoeniger, Malin Young, and Gary H. Kruppa**  
Sandia National Laboratories  
PO Box 969  
Livermore, CA 94551-0969

### **Abstract**

Intra-molecular cross-linking has been suggested as a method of obtaining distance constraints that would be useful in developing structural models of proteins. Recent work published on intra-molecular cross-linking for protein structural studies has employed commercially available primary amine selective reagents that can cross-link lysine residues to other lysine residues or the amino terminus. Previous work using these cross-linkers has shown that for several proteins of known structure, the number of cross-links that can be obtained experimentally may be small compared to what would be expected from the known structure, due to the relative reactivity, distribution, and solvent accessibility of the lysines in the protein sequence. To overcome these limitations we have investigated the use of cross-linking reagents that can react with other reactive sidechains in proteins. We used 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) to activate the carboxylic acid containing residues, aspartic acid (D), glutamic acid (E), and the carboxy terminus (O), for cross-linking reactions. Once activated, the DEO sidechains can react to form "zero-length" cross-links with nearby primary amine containing residues, lysines (K) and the amino terminus (X), via the formation of a new amide bond. We also show that the EDC-activated DEO sidechains can be cross-linked to each other using dihydrazides, two hydrazide moieties connected by an alkyl cross-linker arm of variable length. Using these reagents, we have found three new "zero-length" cross-links in ubiquitin consistent with its known structure (M1-E16, M1-E18, and K63-E64). Using the dihydrazide cross-linkers, we have identified 2 new cross-links (D21-D32 and E24-D32) unambiguously. Using a library of dihydrazide cross-linkers with varying arm length, we have shown that there is a minimum arm length required for the DEO-DEO cross-links of 5.8 angstroms. These results show that additional structural information can be obtained by exploiting new cross-linker chemistry, increasing the probability that the protein target of choice will yield sufficient distance constraints to develop a structural model.

Key Words: Top-Down, FTMS, Solvent Accessibility, Proteins, Chemical Cross-linking

## **Acknowledgements**

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, which is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy under contract DE-AC04-94AL85000.

## Table of Contents

<b>Abstract</b> .....	3
<b>Key Words: Top-Down, FTMS, Solvent Accessibility, Proteins, Chemical Cross-linking</b>	
<b>Acknowledgements</b> .....	3
<b>Acknowledgements</b> .....	4
<b>Table of Contents</b> .....	5
<b>Table of Figures</b> .....	5
<b>Introduction</b> .....	7
<b>Experimental Materials</b> .....	8
Zero length (DEO-KX) cross-linking .....	8
Carboxy-carboxy (DEO-DEO) cross-linking and purification of cross-linked products	9
Sample preparation and mass spectrometry.....	9
<b>Results and discussion</b> .....	10
General Cross-linking Reaction Scheme. ....	10
Zero-length cross-links. ....	10
DEO-DEO Cross-linking. ....	12
<b>Conclusions</b> .....	13
<b>Figures</b> .....	145
<b>References</b> .....	23
<b>Distribution</b> .....	24

## Table of Figures

Figure 1: The sequence and FTMS spectrum of native ubiquitin and ubiquitin with different degrees of zero-length (KX) cross-linking by EDC.....	15
Figure 2: MS/MS data from ubiquitin with different numbers of zero-length cross-links, showing degree of water loss in fragmentation ions correlates with number of cross-links in the unfragmented protein. ....	16
Figure 3: Localization of cross-links in b or y ions is performed by examining nested b or y ions or internal fragments. ....	17
Figure 4: Mass spectra of adducts formed upon reaction of three different dihydrazidic cross-linkers with EDC-activated ubiquitin, showing effects of affinity purification to remove dangling cross-linkers. ....	18
Figure 5: MS/MS spectrum of ADH-cross-linked ubiquitin, showing the effect of affinity purification on the mass distribution of fragmentation products. ....	19
Figure 6: Mass spectra of ubiquitin fragments b18, 19-52, and Y24, the the presence and absence of ADH cross-linker. ....	20
Figure 7: Mass spectra of fragments 12-32, 19-25, 24-32 and 19-23 enable further localization of the cross-links. ....	21
Figure 8: Ubiquitin structure showing previously measured lysine-lysine cross-links and cross-links identified in the new work described here. ....	22

**This page is intentionally left blank**

## Introduction

Inter-molecular chemical cross-linking has a long history as a tool for the study of the quaternary structure of protein complexes,<sup>1,2</sup> and many years ago it was suggested that intra-molecular cross-linking could be used as a method of obtaining distance constraints that would be useful in developing structural models of proteins.<sup>3</sup> However, the promise of intra-molecular cross-linking for structural modeling has only recently been enabled by state of the art mass spectrometric methods that make determination of the cross-linked residues in a protein practical on a reasonable time scale and with small quantities of protein. The first study to derive a sufficient number of intra-molecular distance constraints to develop a structural model for a protein used fibroblast growth factor 2 (FGF-2) as a test case.<sup>4</sup> FGF-2 is a 17 kDa protein, which has 14 lysines evenly dispersed in its 155 residue sequence. Using only the commercially available primary amine reactive cross-linker bis(sulfosuccinimidyl) suberate, 18 cross-links were found in FGF2 in this study, of which 15 provided useful distance constraints for determining the fold family of the protein. Using these 15 constraints it was possible to assign the FGF2 to the correct  $\beta$ -trefoil fold family, using threading techniques.

All of the subsequent recent work published on intra-molecular cross-linking for protein structural studies has employed similar, commercially available primary amine selective reagents that can cross-link lysine residues to other lysine residues or the amino terminus.<sup>5-8</sup> In our own recent work, we showed that a top down approach to the localization of cross-links using FTMS could be applied to localize 3 cross-links in ubiquitin<sup>5</sup>, and in a more recent publication we showed that the distance constraints can be improved by using a series of cross-linkers of different lengths.<sup>6</sup> However, ubiquitin has 8 primary amino groups (7 on lysines and one at the amino terminus) that should be reactive with disuccinimidyl esters, the reagents that were used in the preliminary studies described above, yet only three cross-links were observed. Examination of the known structure of ubiquitin showed that there are many geometrically allowed cross-links between primary amino groups that were not observed experimentally. To determine whether our method could not localize these cross-links or if the cross-links were not formed, we probed the reactivity of the primary amino groups of ubiquitin by reaction with N-hydroxysuccinimidyl acetate (NHSAc), which specifically acetylates primary amines by the elimination of N-hydroxy succinimid; this is the same elimination reaction that occurs twice in the cross-linking reaction with the disuccinimidyl esters.<sup>9</sup> As the stoichiometric ratio of NHSAc:protein increased, identification of the fragments from native protein and protein with successively increasing modification allowed the assignment of the complete order of reactivity of the primary amino groups in ubiquitin (Met1  $\approx$  Lys 6  $\approx$  Lys 48  $\approx$  Lys 63 > Lys 33 > Lys11 > Lys 27, Lys 29). These results are in excellent agreement with the reactivity expected from other studies and predicted from the known crystal structure of ubiquitin, which shows that the least reactive lysines are involved in strong hydrogen bonding interactions. The reactivity order observed also correlates well with the observed cross-links K6-K11, K1-K6, and K48-K63, which all involve one of the most reactive lysines with another nearby, reactive lysine. Clearly the number of cross-links observed when starting from the singly cross-linked parent ion is limited by the lysine reactivity.

One could hope to overcome this limitation by starting with parent ions containing multiple cross-links, but the abundance of multiply cross-linked species is often quite low, and heavily modifying the protein may perturb its native state structure. In a recent study by another group using proteolytic digests to localize the cross-links in cross-linked cytochrome-C, only one cross-link was observed with lysine reactive cross-linkers when the reaction was optimized to yield a single cross-link per protein molecule.<sup>7</sup> At higher cross-linker concentration optimized to yield two cross-links per protein molecule, the digest yielded only 4 additional cross-links. Only two of these 5 total cross-links in cytochrome-C were structurally significant, the others were between lysines that were very close together in the sequence and would have been expected to

cross-link. This study combined with our own work and discussions with others working in the area provides support for our hypothesis that in order to obtain a large number (10-100) of structurally significant cross-links it is necessary to take advantage of reactive side-chains other than lysine. There are several other advantages to examining other reactive sidechains, one is that these may lead to better C $\alpha$ -C $\alpha$  distance constraints: the lysine side chain is quite long and flexible, about 6 Å from C $\alpha$  to N $\epsilon$ , so a cross-link between two lysines gives a minimum 16 Å C $\alpha$ -C $\alpha$  distance constraint even with a short cross-linker arm containing 2 methylene groups. The distribution of lysines in a protein can also be problematic, e.g. the cytochrome-C sequence contains one set of three adjacent lysines, and three pairs of adjacent lysines. Cross-links involving one lysine at these positions would be hard to resolve by proteolytic digestion and/or MS/MS, and cross-links between lysines that are adjacent in the sequence provide no new structural information. Having a variety of cross-linkers with different sidechain specificities to choose from allows a cross-linking regimen to be chosen for a given protein based on the abundance and sequence distribution of its reactive amino acid sidechains, improving the information yield and making resolution of the cross-links via a top down MS/MS approach and/or proteolytic digestion much easier.

1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) is known to activate carboxylic acid groups for reaction with primary amines to form new amide bonds.<sup>10</sup> In this work, we present two cross-linking schemes to take advantage of the activation of carboxylic acid side chains with EDC to form intra-molecular cross-links. First, EDC by itself can be used to activate the carboxylic side chains, which can then react with nearby primary amines on lysine or the amino terminus. In this reaction, there is no cross-linker molecule, the reactive sidechains must approach and react directly with each other, and we refer to the product of this type of reaction as a “zero-length cross-link”. Previous work has also shown that hydrazides can specifically react with EDC-activated carboxylic acid sidechains.<sup>11</sup> This reactivity led us to investigate the possibility of using commercially available dihydrazides, with alkyl chains of varying length between the hydrazide moieties, as homobifunctional cross-linking reagents for EDC-activated sidechains in a protein.

Throughout this paper, aspartic acid and glutamic acid will be referred to by their single letter residue codes, D and E respectively, the carboxy terminus will be represented by the letter O, and carboxylic sidechain residues in general will be referred to by the abbreviation DEO. Likewise, the primary amine containing residue lysine will be referred to by its one letter code, K, the amino terminus will be represented by X, and primary amine containing residues in general will be referred to by the abbreviation KX.

## **Experimental Materials**

### **Zero length (DEO-KX) cross-linking**

Ubiquitin was dissolved in a 50mM pyridine/hydrochloride (pH 6.0) buffer at a concentration of 0.5mg/ml. Protein solutions were split into the three independent reactions, one control, and two to which cross-linking reagent was added. Each reaction contained 50ug of protein. Ubiquitin was cross-linked by two concentrations of EDC/HCl (N-Ethyl-N’-(3-dimethylaminopropyl) carbodiimide hydrochloride) (1:100 and 1:200, ubiquitin:EDC) at room temperature for eight hours. The control sample consisting of protein in buffer but without the EDC was also left at room temperature for eight hours.

## **Carboxy-carboxy (DEO-DEO) cross-linking and purification of cross-linked products**

Ubiquitin was dissolved in a 50mM pyridine/hydrochloride (pH 5.5) buffer at a concentration of 0.5mg/ml. Protein solutions were split into the six independent reactions, one control, and five with cross-linking reagents. Each of the 5 cross-linking reactions contained 50ug of protein, cross-linked with 5 different dihydrazide cross-linking reagents. Ubiquitin was reacted with a fifty molar excess of EDC/HCl (N-Ethyl-N'-(3-dimethylaminopropyl) carbodiimide hydrochloride) and a 25 molar excess of cross-linking reagents (1:50:25, ubiquitin:EDC:cross-linker) at room temperature for eight hours. Adipic dihydrazide (ADH), succinic dihydrazide (SDH), oxalic dihydrazide (ODH), carbonic dihydrazide (CDH) and hydrazine (HH) were used as cross-linking reagents. Two thirds of the reaction mixture was incubated with 100 $\mu$ l of periodate-activated sepharose slurry for four hours after the cross-linking reactions were complete and the sample was recovered by pipetting the supernatant after centrifugation. The reason for the extra purification step in the DEO-DEO cross-linking sample workup will be presented in the results and discussion section. The complete reaction and purification was performed on ubiquitin in the absence of EDC and the cross-linking reagent as a control.

### **Sample preparation and mass spectrometry**

All samples, including the controls, were purified using a one-step protein trap cartridge cleanup (Macro Traps, Michrom BioResources, Auburn, CA) and were diluted to a 2 $\mu$ M final protein concentration in aqueous solution of 6% acetic acid and 50% methanol (EM Science, Gibbstown, NJ) for subsequent mass spectrometric analysis.

All MS and MS/MS experiments were performed on a commercial prototype APEX Q-FTMS instrument equipped with a 7.0 T superconducting magnet and an Apollo ESI ion source (Bruker Daltonics, Billerica MA). The Q front-end consists of a quadrupole mass filter followed by a hexapole collision cell. By switching the potentials on the exit lenses appropriately under the control of the data acquisition computer, ions could be accumulated either in the hexapole of the Apollo ESI source, or in the hexapole collision cell of the Q front end, prior to transfer to the FTMS analyzer cell. Mass spectra of the reaction product mixtures were obtained by accumulating ions in the ESI source hexapole and running the quadrupole mass filter in non mass-selective (Rf-only) mode so that ions of a broad  $m/z$  range (200-2000) were passed to the FTMS analyzer cell.

Ubiquitin with a single internal cross-link was isolated from the mixture by setting the quadrupole mass filter to pass the  $m/z$  of one of its charge states, with a resolution of approximately  $\pm 2 m/z$ . In all cases, the +11 charge state was selected by the quadrupole mass filter and fragmented to give MS/MS spectra. The quadrupole mass filter had sufficient mass resolution to selectively pass only the species of interest, as determined by accumulating the selected ions in the collision cell at low collision energy and then transferring those ions to the FTMS analyzer cell to obtain a precursor ion mass spectrum. After the clean selection of the desired precursor ion was confirmed, the DC offset on the collision cell was dropped, providing approximately 19 V of collision energy to the ions, which induced fragmentation. The collision energy was adjusted for each species to give extensive fragmentation, and the fragments were simultaneously produced and accumulated in the collision cell. Since the precursor ion contained only a single internal cross-link, many unmodified ubiquitin fragment ions were typically observed in the MS/MS spectra. The unmodified fragments were identified and used to internally calibrate the spectra resulting in typical average mass accuracy below 1 ppm.

## Results and discussion

**General Cross-linking Reaction Scheme.** The reactions important for the cross-linking studies presented here are shown in Scheme 1. Species I depicts a polypeptide backbone with three side-chains, the middle one being a lysine (which has a terminal primary amine), and two outer side-chains representing either aspartic or glutamic acid. The first step in all of the cross-linking reactions relevant to this work is the activation of the carboxylic acid groups by EDC to form species II. This activated intermediate can undergo several subsequent reactions, depending on the other reagents present. For zero-length cross-linking of KX residues to DEO residues, EDC is the only reagent used, and the only possible outcome is the reaction of the activated DEO residues with a KX residue to form a new amide bond directly between the two residues, species III. If there is no KX residue near enough to react with the EDC activated carboxylic acid residue, then no further reaction occurs, and the activated residue is hydrolyzed back to the unmodified carboxylic acid residue during the reaction workup.

In the presence of a bifunctional cross-linking reagent several more products are possible (species III, IV and V). The presence of cross-linking reagent does not prevent the formation of zero-length cross-links, as shown in species IV. Also shown (species IV) is the formation of a “hanging” cross-link (also referred to as a dead-end cross-link by some groups), where the dihydrazide reagent has reacted at one end with an activated carboxylic acid, but there is no other activated carboxylic acid group sufficiently nearby to react with. Another possible outcome of the reaction in the presence of dihydrazide cross-linkers is the cross-linked species V, where both ends of the dihydrazide have reacted to form an internal cross-link. Two other possible outcomes of the reaction, if only one or two modifications are considered, not shown in Scheme I, are species VI, a single hanging cross-link, and species VII, a zero-length cross-link together with a DEO-DEO cross-link.

**Zero-length cross-links.** For reference the sequence of ubiquitin is shown in Figure 1, with the primary amine containing residues in bold font, and the acidic residues in bold italic font. For zero-length cross-linking a new amide bond is formed directly between an activated carboxylic acid residue and a primary amine, resulting in the net loss of a water molecule from the starting protein. Figure 1 also shows the deconvoluted ESI-FTMS spectrum of ubiquitin after reaction with 100x and 200x EDC for eight hours. The control spectrum shows no change after standing in buffer at room temperature for 8 hours, however the reaction mixtures with ubiquitin show abundant water losses indicating the formation of multiple zero-length cross-links. To localize the cross-links MS/MS was performed on the unmodified protein from the control sample and the protein containing 1 and 2 internal cross-links by isolating the protein from the reaction mixtures with 1 and 2 water losses. The analysis of the MS/MS spectra is complicated by the fact that water loss from peptides and proteins is commonly caused by the CID process. However, by analyzing trends in the relative intensity of the water loss fragments as a function of the number of internal cross-links, we found it possible to assign fragments containing zero-length cross-links with confidence. Water loss is most commonly observed from Ser and Thr residues, and since we are cross-linking KX and DEO residues, we do not expect the cross-linked species to show an appreciable change in the extent of water loss that is caused by the fragmentation process alone.

An example of the MS/MS data obtained is shown in Figure 2. Our strategy for interpreting the spectra typically involves identifying large and abundant fragments that help determine which parts of the protein sequence are involved in cross-links. The b18 and y58 are complementary fragments covering the entire sequence. As shown in Figure 2a, the intensity of the water-loss fragment for the b18 ion shows a clear trend of increasing intensity in the MS/MS spectrum of the singly and doubly cross-linked precursor ions compared to MS/MS of the control, unmodified protein. In addition double water loss peaks appear with a trend of increasing intensity in the singly and doubly cross-linked precursor ions, and the double water loss is completely absent in the control spectrum. The increase in the water loss intensity is clear evidence for the formation of at least one zero-length cross-link in the part of the sequence covered by the b18 ion. The double water loss peaks provide further evidence for this, since they are almost certainly due to water loss from the cross-linked fragment or possibly the b18 fragment with two zero-length cross-links. The intensity of the double water loss peak relative to the single water loss peak in both cases is similar to the relative intensity of the water loss peak to the unmodified protein in the control spectrum indicating that much of the signal in the double water loss peak is due to water loss from the b18 ion with a single internal cross-link. In contrast, for the y58 ion shown in Figure 2b, the trend in water loss patterns is very different, the intensity of the first water loss peak increases in the singly cross-linked MS/MS spectrum, but then decreases in the doubly cross-linked spectrum where the double water loss peak becomes the most abundant. The relative abundances of the double and triple water loss peaks in the cross-linked spectra also support the conclusion that the most abundant y58 ion derived by MS/MS from the singly cross-linked precursor contains a single cross-link, while the most abundant y58 ion from the doubly cross-linked precursor contains two zero-length cross-links.

There are still many residues that could form zero-length cross-links in the parts of the sequence covered by the b18 and y58 ions. To further localize the cross-links, we now search for sequence ions that can confirm or eliminate the participation of particular residues. As shown in Figure 3a and 3b, there are two additional sequence ions that allow us to localize the cross-link in the sequence covered by the b18 ion. The fact that the b16 ion is observed with a similar trend in water loss pattern as the b18 ion as a function of the number of cross-links in the precursor ion indicates that the glutamic acid E16 is participating in cross-link formation to an appreciable extent, as this is the only acidic residue in this part of the sequence. The 4-16 internal ion does not display any significant change in its water loss pattern as a function of the number of cross-links in the precursor ion, which eliminates K6 and K11 from being the lysine partners in the cross-link with E16. This shows that there must be a cross-link between E16 and the amino terminus M1, the only remaining primary amine containing residue in the part of the sequence covered by b16. The fact that the increase in the water loss peaks relative to the unmodified fragment is less for b16 than for b18 indicates that E18 is also involved in cross-link formation. Similar to the discussion above for the E16 to M1 cross-link, there is an unmodified 4-18 fragment in the MS/MS spectrum of the singly cross-linked precursor that shows that E18 also forms a zero-length cross-link to the amino terminus.

Clear evidence that at least one of the zero-length cross-links in the part of the sequence covered by the y58 ion involves K63 and E64 is shown by the peak intensities in Figure 3c. While the 63-74 internal ion is a weak fragment in the control spectrum,

there is a clear trend of increasing intensity in the 63-74 with water loss fragment relative to a decreasing intensity in the 63-74 fragment. There are many other fragments from the part of the sequence covered by y58 that show trends of increasing water loss, however none of these fragments allow unambiguous assignments of other zero-length cross-linked pairs of acidic and primary amine containing residues. For example there is an abundant y40 ion that clearly shows the presence of a cross-link from the trend in water loss peak intensity, however this fragment contains 5 acidic residues and 2 primary amine containing residues. We hope to use MS<sup>3</sup> and also Electron Capture Dissociation in the future to improve our ability to localize these cross-links.

**DEO-DEO Cross-linking.** In the case of DEO-DEO cross-linking an extra purification step is performed by incubating the products of the cross-linking reaction with periodate-activated sepharose beads in slurry for four hours. The reason for doing this is that the species IV in scheme 1, the protein with a hanging cross-link and a single zero-length cross-link, that can form in the DEO-DEO cross-linking reaction has exactly the same molecular weight as the species V, the desired species containing a single internal DEO-DEO cross-link. Incubating the sample with periodate-activated sepharose traps all species containing a hanging cross-link on the beads, including species IV, allowing it to be removed easily. The spectra in Figure 4 show the products from single modification of the protein by the dihydrazide reagent before and after purification, with the species labeled according to Scheme I. The efficiency of cross-link formation using the shortest dihydrazide reagent, ODH, is clearly shown to be much lower than the longer dihydrazides ADH and SDH, after the purification step. ADH and SDH show significant amounts of the internally cross-linked species V after the purification, while most of the intensity at the correct mass for species IV and V with ODH largely disappears after purification, indicating that it is mostly due to product IV and not the internally cross-linked product.

The efficiency of the purification is shown by the MS/MS spectra in Figure 5. When the precursor ion corresponding to the m/z of species IV and V are isolated from the crude reaction mixture of ADH with ubiquitin, the MS/MS spectra in the region of the y24 ion shown in Figure 5 contain peaks that correspond exactly to the masses of the y24 ion with a hanging cross-link. In the MS/MS spectra of the reaction products purified using periodate-activated sepharose, all evidence for fragment ions with hanging cross-links disappears. Similar results are observed for the other cross-linkers (data not shown). Eliminating these species by chemical purification makes possible the isolation of only the singly internally cross-linked ion and greatly simplifies the data analysis.

Regions of the MS/MS spectra resulting from isolation of the ubiquitin ion that has the m/z corresponding to a single internal cross-link with ADH are shown in Figures 6 and 7. Figure 6 shows that the DEO-DEO cross-links do not form in residues 1-18 (b18 ion, Fig. 6a) or 52-76 (y24 ion, Fig. 6c), but only in residues 19-52 (19-52 internal ion, Fig. 6b). The 19-32 internal ion appears only with an internal cross-link as shown in Figure 7a, with nearly equal intensity as the unmodified fragment in the MS/MS spectrum of the control. Figure 7b shows that the 19-25 internal ion does not appear in the MS/MS spectrum of the ADH cross-linked ubiquitin, but also does not appear with an internal cross-link. This implies that one or both of the acidic residues D21 or E24 are participating in a cross-link that inhibits the formation of the 19-25 internal fragment, but not with each other. Figure 7c shows that E24 cross-links to D32 since the 24-32 internal

ion appears with a cross-link. However, while the 24-32 internal ion completely disappears in the MS/MS of the cross-linked spectrum, the intensity of the cross-linked peak does not correspond to the intensity of the unmodified peak in the control spectrum, which indicates that at least one other cross-link is contributing to the loss of the 24-32 internal ion signal. The 19-23 and 27-32 internal ions (Figure 7d) also decrease in intensity in the MS/MS spectra of the cross-linked species indicating that D21 and D32 participate in cross-linking, and the only remaining possibility is that they cross-link to each other, and this contributes to the loss of the 24-32 internal ion signal in Figure 7c.

## Conclusions

We have shown that cross-linking reactions between primary amine containing side chains and acidic side-chains can be used to obtain 'zero-length' distance constraints, and that these cross-links can be localized in ubiquitin using the top-down method. Similar results for cross-linking of carboxylic acid side chains to each other using dihydrazides as the cross-linking reagents have also been shown. The value of these new reagents is clear Figure 8 is considered. Using only primary amine reactive cross-linking reagents we could obtain only three distance constraints in our previous work on ubiquitin. Figure 8 shows that by extending the cross-linking chemistry to include zero-length and acidic side-chain cross-links, 5 new cross-links are observed for a total of 8 distance constraints. One of these is a cross-link between adjacent residues, which does not provide any useful structural information. However, 7 distance constraints is close to our goal of obtaining 1 cross-link per 10 residues, which is a number that should be useful to the protein structure modeling community for filtering the correct structure from those predicted by state of the art computational methods. Furthermore, our data shows that there are other cross-links that we could not fully localize to date using our top-down MS/MS methods, so that future developments using MS<sup>3</sup> or a combination of top-down and bottom-up methods should yield more constraints.

## Figures

Scheme 1: Reaction of EDC with acidic side-chains and possible outcomes of reactions of activated acid groups with dihydrazidic cross-linkers.

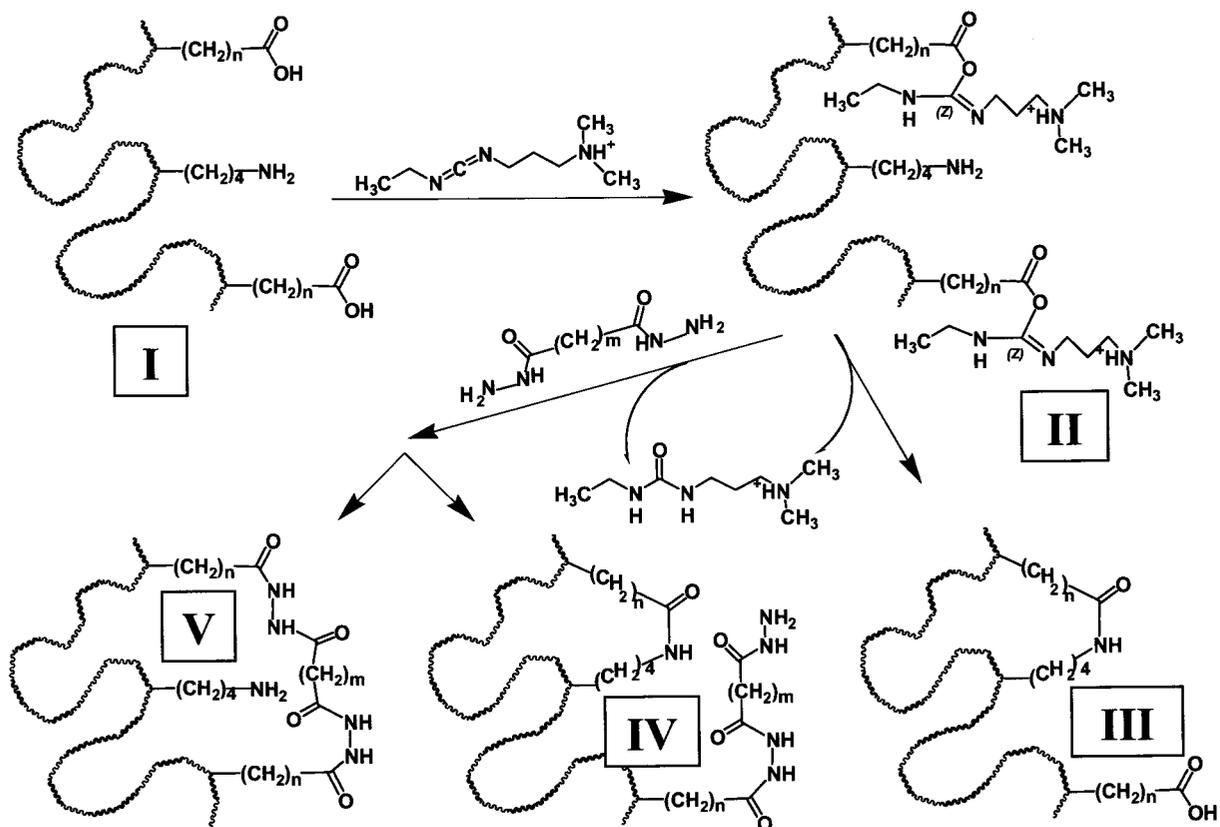
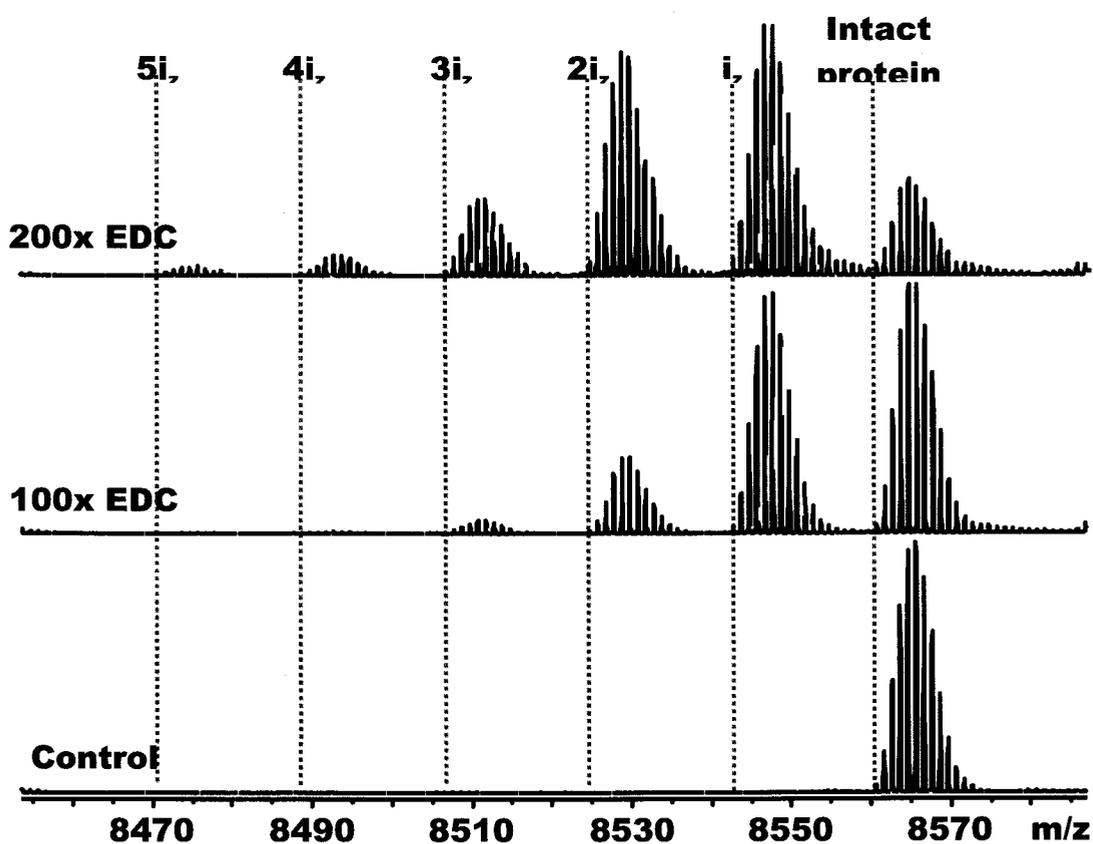


Figure 1: The sequence and FTMS spectrum of native ubiquitin and ubiquitin with different degrees of zero-length (KX) cross-linking by EDC.



MQIFVKTLTG<sup>11</sup>KTITEVEPS<sup>21</sup>DTIENVKAKI<sup>31</sup>QDKEGIPPDQ  
<sup>41</sup>QRLIFAGKQL<sup>51</sup>EDGRTLSDYN<sup>61</sup>IQKESTLHLV<sup>71</sup>LRLRGG

Figure 2: MS/MS data from ubiquitin with different numbers of zero-length cross-links, showing degree of water loss in fragment ions correlates with number of cross-links in the unfragmented protein.

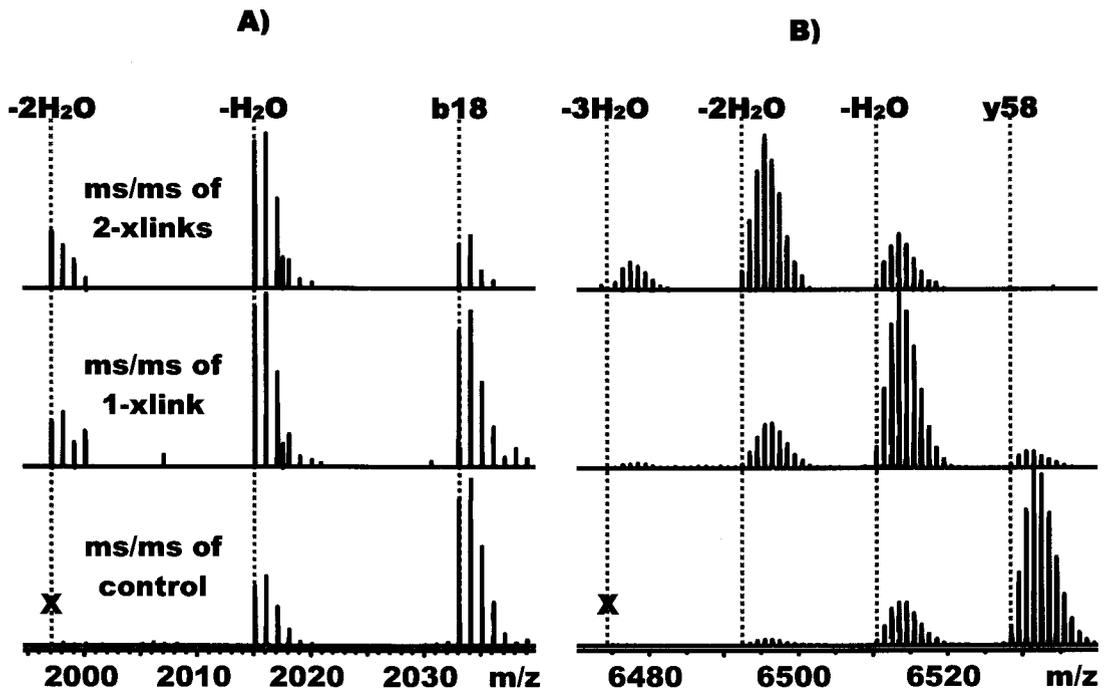


Figure 3: Localization of cross-links in b or y ions is performed by examining nested b or y ions or internal fragments. When a cross-link is never present in a nested fragment, there is no correlation of water loss intensity and the number of cross-links in the original protein that was fragmented, allowing the residues in that fragment to be excluded. Fig. 3A and 3B localize the cross-link within b18. Fig. 3C localizes the cross-link within y58.

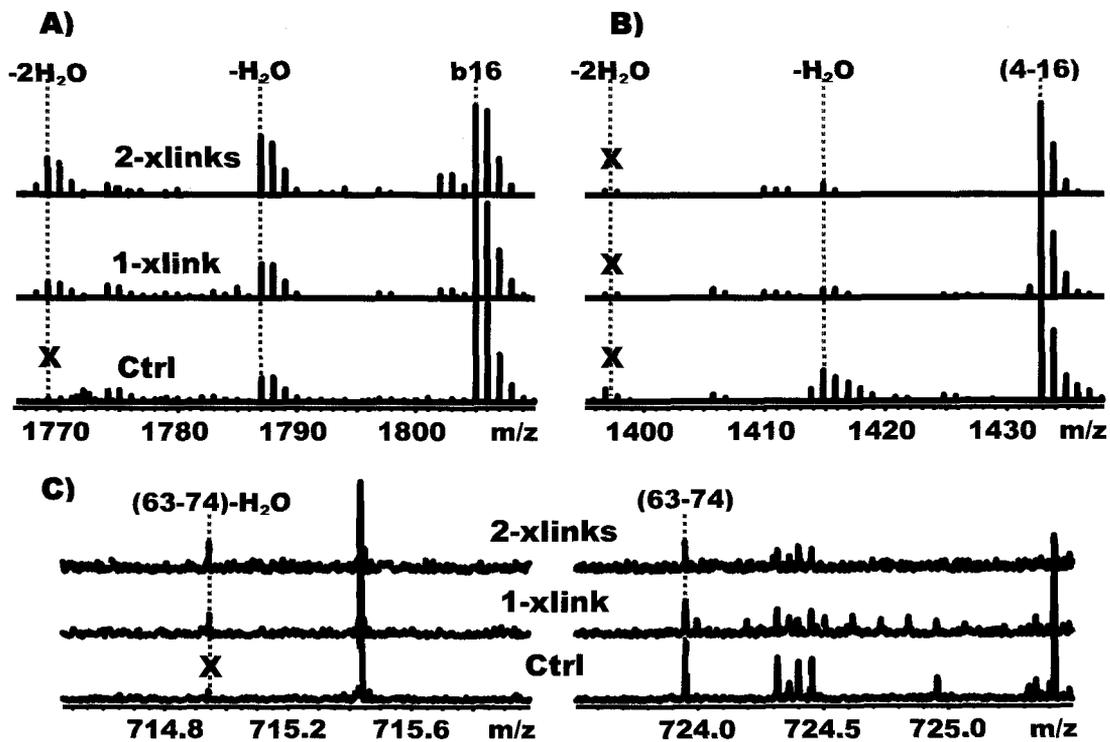


Figure 4: Mass spectra of adducts formed upon reaction of three different dihydrazidic cross-linkers with EDC-activated ubiquitin, showing effects of affinity purification to remove dangling cross-linkers.

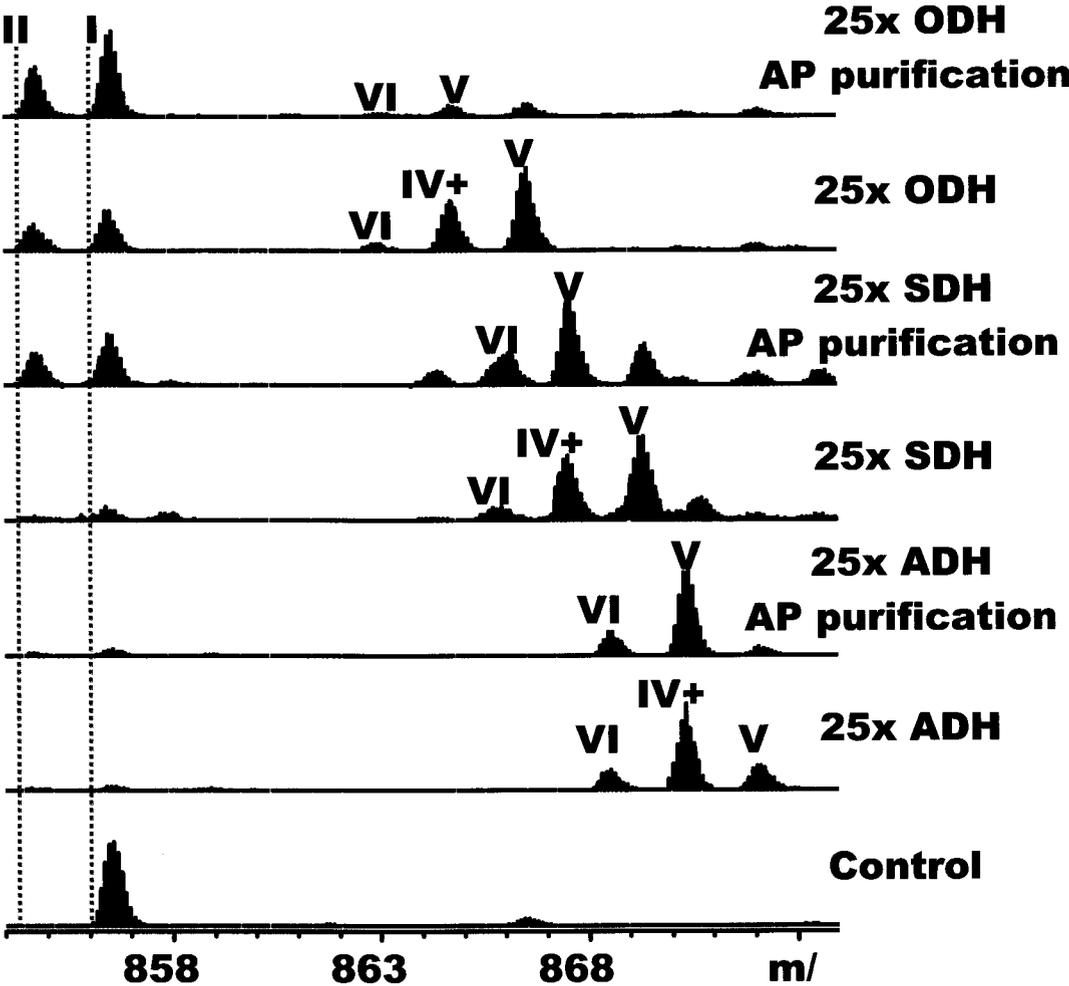


Figure 5: MS/MS spectrum of ADH-cross-linked ubiquitin, showing the effect of affinity purification on the mass distribution of fragmentation products. The hanging ADH species is completely removed.

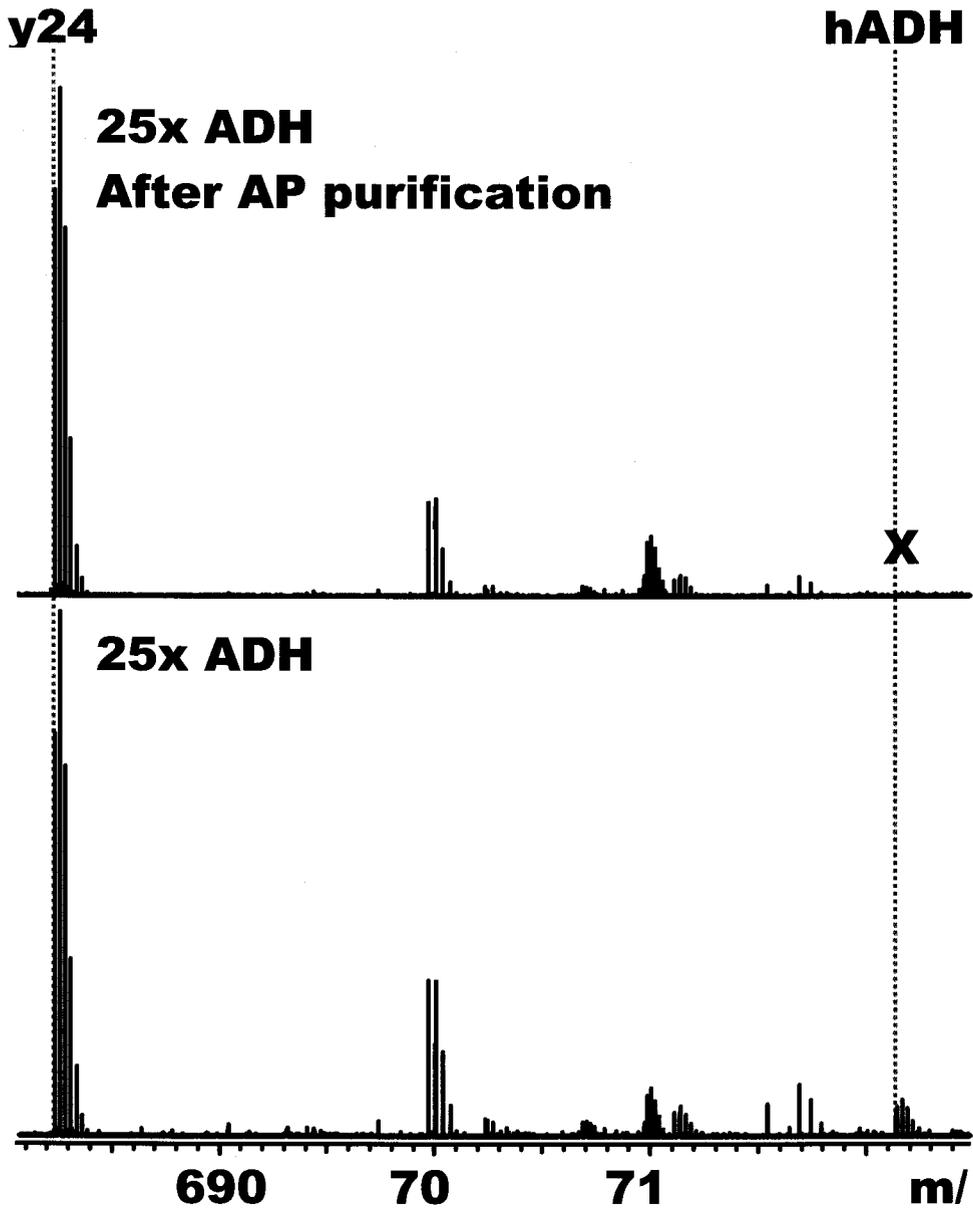


Figure 6: Mass spectra of ubiquitin fragments b18, 19-52, and Y24, the the presence and absence of ADH cross-linker. An internal ADH cross-link is observed only in 19-52.

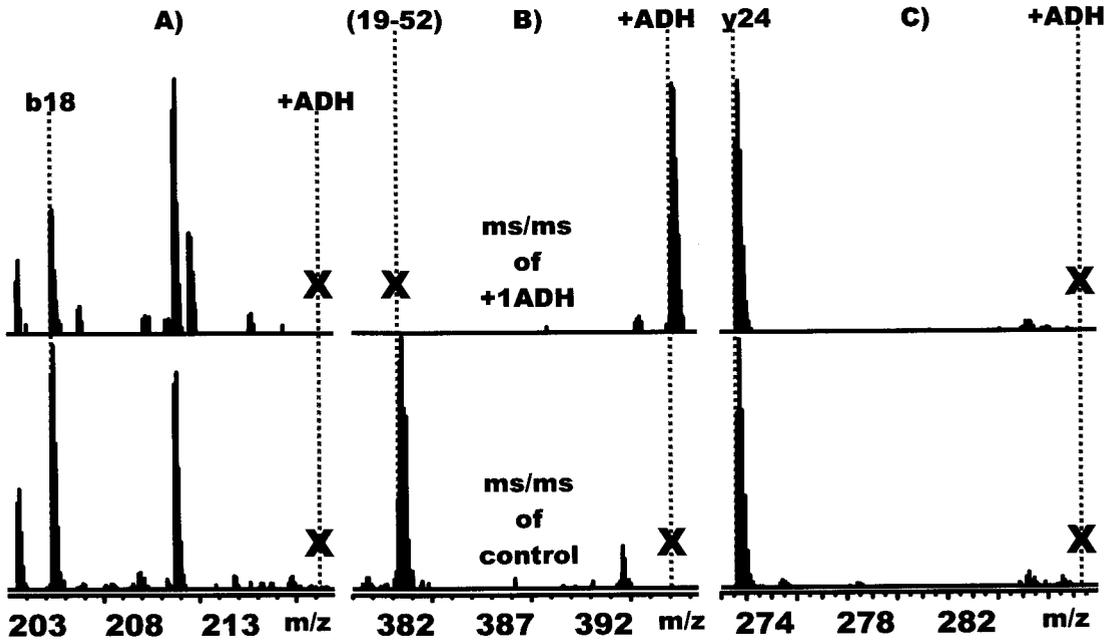


Figure 7: Mass spectra of fragments 12-32, 19-25, 24-32 and 19-23 enable further localization of the cross-links.

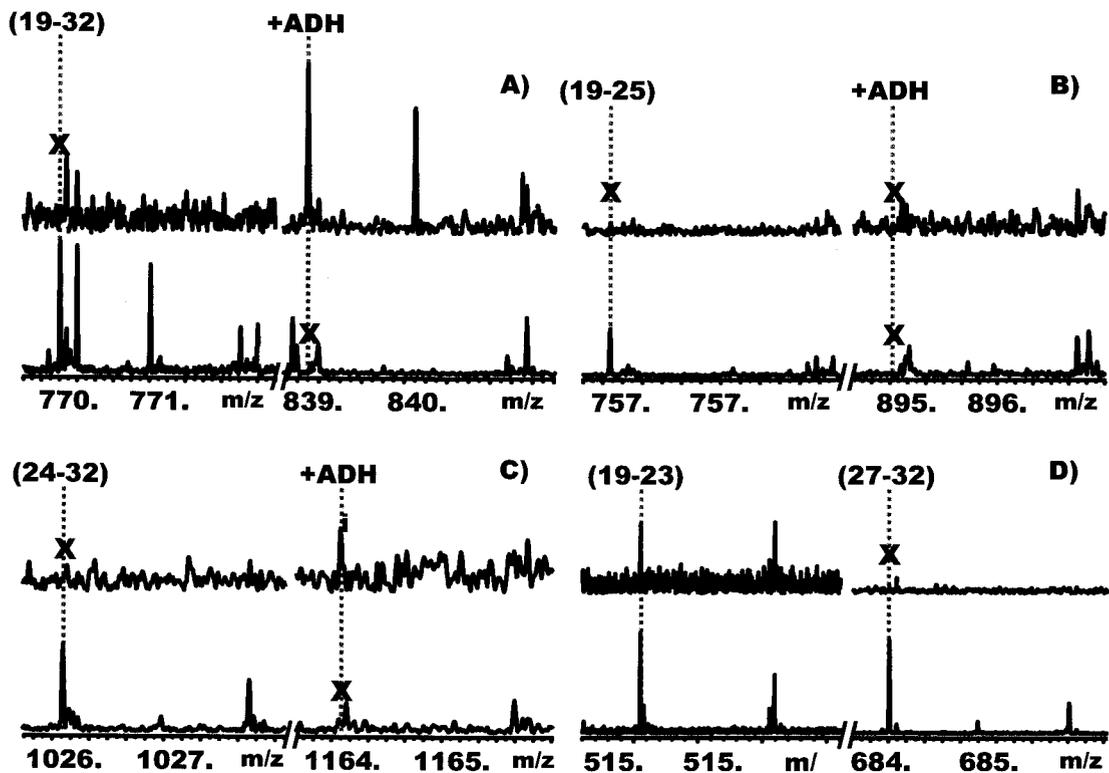
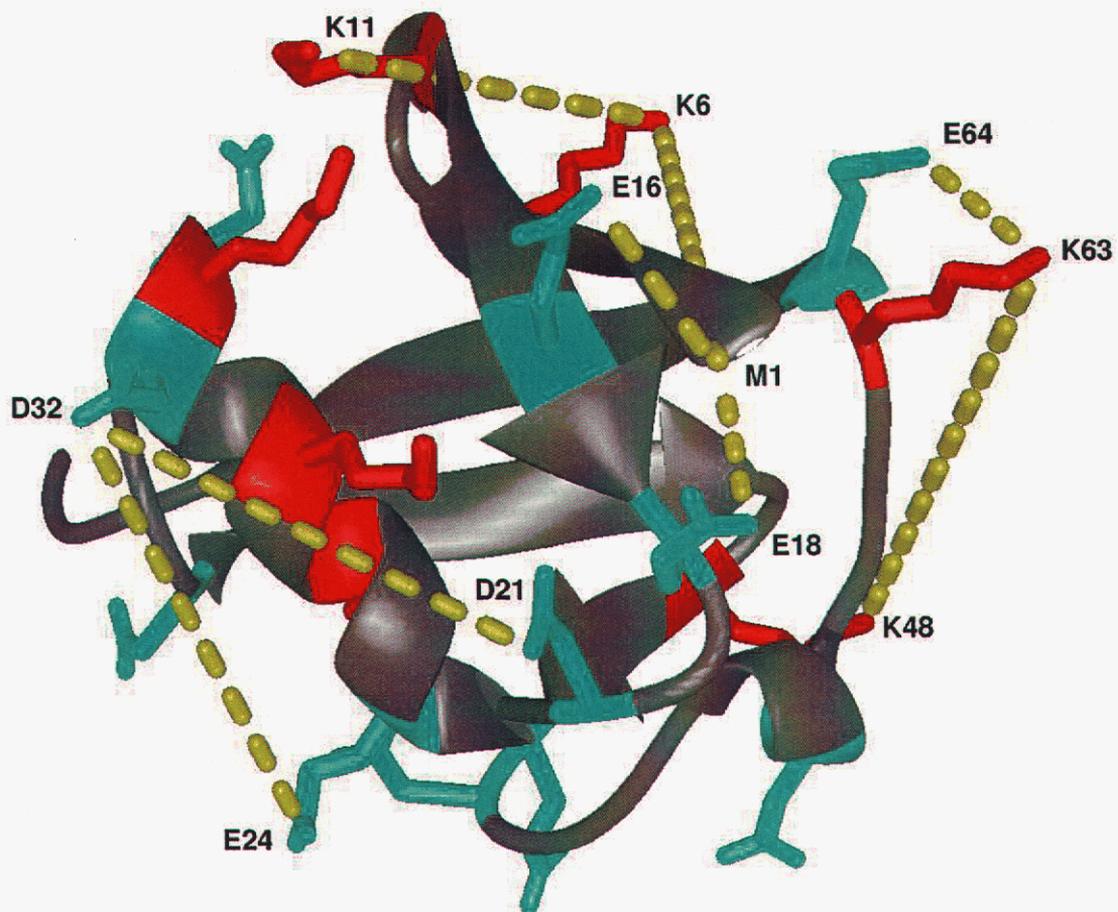


Figure 8: Ubiquitin structure showing previously measured lysine-lysine cross-links and cross-links identified in the new work described here.



## References

1. Peters K, Richards FM. Chemical cross-linking: Reagents and problems in studies of membrane structure. *Ann. Rev. Biochem.* 1977; **46**: 523-551.
2. Brunner J. New photolabeling and cross-linking methods. *Ann. Rev. of Biochem.* 1993; **62**: 483-514.
3. Wold F. *Methods in Enzymology.* 1972; **25**: 623.
4. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High-throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Nat. Acad. Sci.* 2000; **97**: 5802-6.
5. Kruppa G, Schoeniger J, Young M. A top down approach to protein structural studies using chemical cross-linking and fourier transform mass spectrometry. *RAPID COMMUNICATIONS IN MASS SPECTROMETRY.* 2003; **17**: 155-162.
6. Novak P, Young, M. M., Schoeniger, J., Kruppa, G. H. A top down approach to protein structure studies using chemical cross-linking and fourier transform mass spectrometry. *European Journal of Mass Spectrometry.* 2003; **9**: 623-631.
7. Pearson KM, Pannell LK, Fales HM. Intramolecular cross-linking experiments on cytochrome c and ribonuclease a using an isotope multiplet method. *Rapid Communications In Mass Spectrometry.* 2002; **16**: 149-159.
8. Dihazi GH, Sinz A. Mapping low-resolution three-dimensional protein structures using chemical cross-linking and fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom.* 2003; **17**: 2005-14.
9. Novak P, Kruppa, G. H., Young, M. M., Schoeniger, J. A top down method for the determination of residue specific accessibility in proteins. *Journal of Mass Spectrometry.* 2003; **In Press**:
10. Yamada H, Imoto H, Fujita K, Okazaki K, Motomura M. *Biochemistry.* 1981; **18**: 4836-42.
11. Wilchek M, Bayer EA. Labeling glycoconjugates with hydrazide reagents. *Methods Enzymol.* 1987; **138**: 429-42.

## **Distribution**

1	MS9292	J. S. Schoeniger, 08321
1	MS9291	L. Napolitano, 08320
1	MS9292	M. Young, 08321
1	MS1413	T. Michalske, 08300
3	MS9018	Central Technical Files, 8945-1
1	MS0899	Technical Library, 9616
1	MS9021	Classification Office, 8511/Technical Library, MS 0899, 9616
1	MS9021	Classification Office, 8511 for DOE/OSTI
1	MS0188	D. Chavez, LDRD Office, 1030