

SANDIA REPORT

SAND2005-1699

Unlimited Release

Printed March 2005

Chemical Crosslinking and Mass Spectrometry Studies of the Structure and Dynamics of Membrane Proteins and Receptors

M.J. Ayson, W.E. Haskins, J. Hong, R.B. Jacobsen,
G.H. Kruppa, M.D. Leavell, P. Lane, P. Novak, K.L. Sale,
N.L. Wood, M.M. Young, & Joseph S. Schoeniger

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
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
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



SAND2005-1699
Unlimited Release
March 2005

Chemical Crosslinking and Mass Spectrometry Studies of the Structure and Dynamics of Membrane Proteins and Receptors

Marites J. Ayson, William E. Haskins, Joohee Hong, Richard B. Jacobsen, Gary H. Kruppa, Michael D. Leavell, Pamela Lane, Petr Novak, Kenneth L. Sale, Nichole L. Wood, Malin M. Young, and Joseph S. Schoeniger

Sandia National Laboratories
PO Box 969
Livermore, CA 94551-0969

Abstract

Membrane proteins make up a diverse and important subset of proteins for which structural information is limited. In this study, chemical cross-linking and mass spectrometry were used to explore the structure of the G-protein-coupled photoreceptor bovine rhodopsin in the dark-state conformation. All experiments were performed in rod outer segment membranes using amino acid "handles" in the native protein sequence and thus minimizing perturbations to the native protein structure. Cysteine and lysine residues were covalently cross-linked using commercially available reagents with a range of linker arm lengths. Following chemical digestion of cross-linked protein, cross-linked peptides were identified by accurate mass measurement using liquid chromatography-fourier transform mass spectrometry and an automated data analysis pipeline. Assignments were confirmed and, if necessary, resolved, by tandem MS. The relative reactivity of lysine residues participating in cross-links was evaluated by labeling with NHS-esters. A distinct pattern of cross-link formation within the C-terminal domain, and between loop I and the C-terminal domain, emerged. Theoretical distances based on cross-linking were compared to inter-atomic distances determined from the energy-minimized X-ray crystal structure and Monte Carlo conformational search procedures. In general, the observed cross-links can be explained by re-positioning participating side-chains without significantly altering backbone structure. One exception, between C316 and K325, requires backbone motion to bring the reactive atoms into sufficient proximity for cross-linking. Evidence from other studies suggests that residues around K325 form a region of high backbone mobility. These findings show that cross-linking studies can provide insight into the structural dynamics of membrane proteins in their native environment.

Keywords: bovine rhodopsin, chemical cross-linking, Q-Tof, FT-ICR, tandem mass spectrometry

This page is intentionally left blank

Table of Contents

<i>Section I: Chemical Cross-linking and Mass Spectrometry Applied to Determination of Bovine Rhodopsin Structure and Dynamics</i>	9
<i>Abbreviations</i>	9
<i>Introduction</i>	9
<i>Experimental Procedures</i>	11
<i>Molecular modeling methods</i>	14
<i>Results</i>	15
<i>Identification of Cross-Linked Products</i>	17
<i>Computational Analysis</i>	20
<i>Discussion</i>	21
<i>Acknowledgements</i>	26
<i>References</i>	27
<i>Tables</i>	32
<i>Figures</i>	35
<i>Section II: Advanced Mass Spectrometry Techniques</i>	45
<i>Introduction</i>	46
<i>Experimental</i>	47
<i>Results and Discussion</i>	48
<i>Conclusions</i>	51
<i>Acknowledgments</i>	51
<i>References</i>	52
<i>Figures</i>	53

List of Tables

Table 1. Observed $M + H^+$ (Da) and experimental error (ppm) of cysteine-lysine cross-linked rhodopsin peptides. Protonated masses correspond to the monoisotopic peak unless indicated..... 32

Table 2. Observed $M + H^+$ (Da) and experimental error (ppm) of lysine-lysine cross-linked rhodopsin CNBr digest products. Protonated masses correspond to the monoisotopic peak unless indicated..... 33

Table 3. Comparisons of minimum cross-linker length and theoretical minimum approach distances for theoretical (*italics*) and observed (**bold**) cross-linked atoms. 34

List of Figures

Section 1

Figure 2. Total ion chromatograms (TICs) of the reverse phase HPLC separation on a PRLP-S column and ESI-FTMS of CNBr-digested (a) rhodopsin and (b) EMCS cross-linked rhodopsin. Selected ion chromatograms (SICs) from the EMCS cross-linked spectrum of the most abundant isotope peak for (c) peptide 50—86 cross-linked to peptide 310—317 (expanded vertically 4-fold) and (d) the unmodified CNBr peptide 50—86. The SICs are shown to scale overlaid on the TIC. All post-translational modifications were observed; free cysteines were pyridylethylated. 36

Figure 3. Mass spectra of rhodopsin CNBr digest peptides 50 – 86 and 310 – 317 cross-linked with Cys-Lys cross-linkers of different linker arm length. A single spectrum from an LC-MS experiment is shown for each cross-linker and the monoisotopic peak is indicated for (a) SIA, m/z 1055.3881⁺⁵, signal/noise (S/N) = 241.4, (b) GMBS, m/z 1080.38319⁺⁵, S/N = 717.7, (c) EMCS, m/z 1085.9978⁺⁵, S/N = 462.7, (d) LC-SMCC, m/z 1113.8155⁺⁵, S/N = 344.6. S/N is based on most abundant isotope peak. See Table 1 for $M + H^+$ (Da) and experimental error (Δ ppm). 37

Figure 4. ESI-Q-Tof-MS2 of rhodopsin CNBr digest fragments 310 – 317 and 318 – 348 cross-linked with the hetero-bifunctional Cys/Lys-specific reagent LC-SMCC. A cross-link was selectively formed between Cys316 of peptide 310 – 317 and Lys325 of peptide 318 – 348. The spectrum is labeled with major b-, y-, and internal ion series; a cross-link between fragments is indicated by an 'x'. Cys322 and Cys323 are palmitoylated and Met317 is modified to homoserine lactone by CNBr cleavage. The largest peaks have been truncated. 38

Figure 5. ESI-Q-Tof-MS2 of rhodopsin CNBr digest fragments 310 – 317 and 318 – 348 cross-linked by the homo-bifunctional Lys/Lys-specific reagent DSG. The cross-link was selectively formed between between Lys311 of peptide 310 – 317 and Lys339 of peptide 318 – 348. The spectrum is labeled with major b-, y-, and internal ion series; a cross-link between fragments is indicated by an 'x'. Cys322 and Cys323 are palmitoylated, Cys316 is pyridylethylated, and Met317 is modified to homoserine lactone by CNBr cleavage. The largest peaks have been truncated. 39

Figure 6. Cytoplasmic face of rhodopsin displaying chemically reactive Cys and Lys side chains. Solid lines indicate un-ambiguous cross-links, including C316xK67, C316xK325, C316xK311, K67xK339, K67xK325, K66x67, and K325xK339. Dashed lines represent ambiguous cross-link assignments, including C316xK66/67 and a cross-link involving K231/K245/K248. Helical domains I – VIII and loops CI – CIII are labeled, arrows indicate the C-terminal direction of the loops. 40

Figure 7. The simulated distance of closest approach (DCA, see Methods) between for cross-linked atom pairs is plotted against the minimum cross-linker arm length observed to form a cross-link between each pair. Horizontal error bars correspond to the vectorial sum of C alpha RMSDs for the cross-linked residues taken from the crystal structure B factors. The DCA is less than the cross-linker arm length, except for K325 cross-linked to C316, suggesting that additional movement beyond that considered in the DCA simulation is required for this cross-link to form... 41

Supplemental Figure 1. Lysine-Lysine Cross-linkers	42
Supplemental Figure 2. Cysteine-Lysine Cross-linkers	43

Section 2

Fig 2B: IRMPD of α-EMCS-β1	55
----------------------------------------------------------------------------	----

Figure 3. ECD of the $[M+EMCS+6H]6+$ ion yielded abundant c- and z-type ions with few internal fragment ions (Figure 3A). Likewise, ECD of the sextuply charged cation $[M+SIA+6H]6+$ at 879.6533 m/z ($\Delta m = 0.7$ ppm) from the cross-linked peptide α -SIA- β 1 (5271.8797 Da) provided extensive backbone fragmentation (Figure 3B). ECD of the pentuply charged cation $[M+5H]5+$ for the control peptide α at 849.0904 m/z ($\Delta m = 0.7$ ppm, 4240.4185 Da) is shown in Figure 3C. Cleavage of the bond between K66 and K67 was observed in every case.....

Figure 3. ECD of the $[M+EMCS+6H]6+$ ion yielded abundant c- and z-type ions with few internal fragment ions (Figure 3A). Likewise, ECD of the sextuply charged cation $[M+SIA+6H]6+$ at 879.6533 m/z ($\Delta m = 0.7$ ppm) from the cross-linked peptide α -SIA- β 1 (5271.8797 Da) provided extensive backbone fragmentation (Figure 3B). ECD of the pentuply charged cation $[M+5H]5+$ for the control peptide α at 849.0904 m/z ($\Delta m = 0.7$ ppm, 4240.4185 Da) is shown in Figure 3C. Cleavage of the bond between K66 and K67 was observed in every case.....

Fig 3A: ECD of α-EMCS-β1	56
--------------------------------------------------------------------------	----

Fig 3B: ECD of α-SIA-β1	57
-------------------------------------------------------------------------	----

Fig 3C: ECD of control α	57
---------------------------------------------------------	----

Figure 4. ECD of the septuply charged cation $[M+DSS+7H]7+$ at 1140.6378 m/z ($\Delta m = 1.0$ ppm) from the cross-linked peptide α -DSS- β 2 (7977.4214 Da) revealed full palmitoylation of adjacent cysteines (C₃₂₂ and C₃₂₃) and cross-linking of K₆₇ (and not K₆₆) to K₃₂₅ and K₃₃₉.....

Figure 5. Bond cleavages observed by ECD for (A) α -EMCS- β 1 and (B) α -SIA- β 1. The frequency of cleavage at sites I-III, deduced from the relative abundance of ions formed from cleavage at each site, is indicated by the size of the lightning bolt (not to scale).....

Section I: Chemical Cross-linking and Mass Spectrometry Applied to Determination of Bovine Rhodopsin Structure and Dynamics

Mobility in the C-terminal region of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry.

Abbreviations

MS, mass spectrometry; FTICR, fourier transform infra-red coupled resonance; Q-ToF, quadrupole time-of-flight; LC-MS, liquid chromatography – mass spectrometry; MS/MS, tandem mass spectrometry; CID, collision-induced dissociation; ECD, electron capture dissociation; ROS, rod outer segment; HPLC, high-pressure liquid chromatography; TFA, trifluoroacetic acid; ACN, acetonitrile; TCEP, tris (2-carboxyethyl) phosphine; NHS, N-hydroxysuccinimide, MD, molecular dynamics; ALS, ammonium lauryl sulfate; BME, β -mercaptoethanol

Introduction

Recent advances in genomic and proteomic approaches to identifying novel proteins highlight a need for methods of evaluating the structure and conformational dynamics of proteins and protein complexes. Membrane proteins, encoded by ~20% of the genes in most organisms, play a central role in intracellular signaling, energy and material transport, cell intoxication and pathogenesis, and cell recognition and motility. However, relative to soluble proteins few membrane protein structures have been resolved by X-ray crystallography and/or NMR spectroscopy.⁽¹⁾ Progress in the structural analysis of eukaryotic membrane proteins in particular has been slowed by the instability of these proteins in environments lacking phospholipids, their tendency toward aggregation and precipitation, low protein abundance, difficulties in expressing functional protein, and sample purity issues which have hindered the application of standard structural determination methods.

Such technical challenges have necessitated the use of a variety of techniques to gather information about the structure and conformational dynamics of membrane proteins. Disulfide mapping, photoaffinity labeling, metal-ion binding, solvent accessibility, and site-directed spin labeling studies combined with electron paramagnetic resonance

(SDSL-EPR) have been used to obtain low to moderate resolution distance constraints in membrane proteins.(2-12) These methods typically require the construction and expression of site-specific histidine or cysteine mutants of the native protein. Several groups have recently reported an approach to the analysis of protein structure which combines chemical cross-linking, mass spectrometry, and computer modeling to obtain moderate-resolution structural information (13-16). Chemical cross-linking utilizing the amino acid ‘handles’ available in the native protein sequence offers the advantage of minimally perturbing protein structure and allowing studies to be carried out in the native membranes.

We chose the intensively studied G-protein-coupled receptor (GPCR) rhodopsin as a model system for developing a cross-linking methodology for membrane proteins. Bovine rhodopsin has the major advantages that it can be purified in large quantities in its native rod outer segment (ROS) membrane and it has a known crystal structure (17) against which we can validate our results. The study of native membrane proteins by chemical cross-linking requires significant optimization of a) sample preparation methods (including reaction conditions, separation of monomeric protein, and proteolysis), b) LC-MS conditions for separating and identifying complex mixtures of hydrophobic peptides, and c) data analysis software for the assignment of complex MS spectra. We show that the specific sites of cross-linking in bovine rhodopsin can be identified using this strategy.

We also considered the best way to relate the formation of inter-residue cross-links to the structure and dynamics of a protein, an issue that has great impact on the utility of cross-linking data in structure determination. In principal, upper bounds on the distance of closest approach (DCA) between the involved atoms can be inferred from the formation of the cross-link, provided the cross-linking reaction does not somehow distort the protein structure. Considered in isolation, flexible cross-links between atoms on flexible amino acid side chains produce constraints with large distance uncertainties, limiting the accuracy of structural models which depend on them, whether the protein is soluble(16) or a membrane protein.(18) An alternative approach is to perform a series of cross-linking reactions with several reagents with nested lengths and identical end-group reactivity and determine whether there is a minimum length that allows for the formation of the cross-link.(19) This approach sets upper *and* lower bounds on the DCA between the two tethered atoms, potentially resulting in a distance measurement associated with a small uncertainty (~1-2 Å), albeit a measurement of the dynamic DCA between two atoms that is achieved over a time scale of minutes. This approach is similar to the use of nested-length covalently-tethered inhibitors on transmembrane channels as molecular “tape measures” to measure the radial distance of residues from the pore.(20) In order to validate chemical cross-linking as a method for deriving bounds on inter-residue distances, we have extensively compared experimentally observed cross-links in bovine rhodopsin with those predicted from the known structure of the protein using a variety of modeling approaches. As a baseline, we compared cross-link derived distances to the side-chain distances determined from the energy minimized X-ray crystal structure.(17) The crystal structure, however, represents a single static conformation of rhodopsin, and therefore conformational search procedures and the results of a 40 ns molecular dynamics simulation(21) were also used to identify side-chain and/or backbone dynamics that could lead to the observed cross-links.

In general, the observed cross-links are consistent with the structural information available for bovine rhodopsin when evaluated using the modeling techniques described above, although the data implies significant side-chain motion relative to the crystal structure coordinates. The results indicate that cross-linking is a useful technique for probing membrane structure and, in particular, assessing the dynamic features of a structure – features that may be important in the conformational changes that accompany protein activation and protein-protein interactions.

Experimental Procedures

Materials

Frozen bovine retinas were purchased from Schenk Packing Company, Inc. (Stanwood, WA). Tris(2-carboxyethyl) phosphine hydrochloride (TCEP-HCl), and the cross-linking reagents, including disuccinimidyl tartarate (DST), disuccinimidyl glutarate (DSG), disuccinimidyl suberate (DSS), N-(γ -Maleimidobutyryloxy) succinimide ester (GMBS), N- ϵ -Maleimidocaproyloxy]succinimide ester (EMCS), succinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxy-(6-amidocaproate) (LC-SMCC), and N-Succinimidyl iodoacetate (SIA), were obtained from Pierce Biotechnology, Inc. (Rockford, IL). Dimethyl sulfoxide (DMSO), N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 4-vinyl-pyridine, trifluoroacetic acid (TFA), ammonium lauryl sulfate (ALS), and cyanogen bromide (CNBr), were purchased from Sigma-Aldrich (St. Louis, MO). The detergent, *n*-Nonyl- β -D-glucoside, was purchased from Anatrace, Inc. (Maumee, OH). N-hydroxysuccinimidyl acetate (NHS-Ac), was purchased from ICN Biomedical (Irvine, CA).

ROS membrane purification. Rod outer segments (ROS) from bovine retinas were prepared under dim red light (> 650 nm) as previously described.(22) Briefly, ROS fragments were isolated using the sucrose flotation method.(23) and purified via a sucrose step gradient centrifugation. ROS membranes were washed with hypotonic buffer (10 mM HEPES, 1 mM DTT, 100 μ M EDTA, and 100 μ M PMSF) to remove loosely bound polypeptides and subsequently stored in isotonic buffer (100 mM NaCl, 10 mM HEPES, 5 mM MgCl₂·6H₂O, 1 mM DTT, and 100 μ M PMSF) at -80°C until needed.

Rhodopsin quantification. Extraction of rhodopsin from purified ROS membranes was achieved by centrifugation (16,000 g, 5 min) and subsequent resuspension in 50 mM pyridine-HCl (pH 6.5) with 25 mM ZnCl₂ and 0.6% *n*-nonyl- β -D-glucoside.(24) Samples were incubated at room temperature for 30 min. to solubilize, and the concentration of rhodopsin was determined based on the absorbance difference at 500 nm before and after illumination ($\epsilon = 42,700$ M⁻¹ cm⁻¹). (25)

Cross-linking. All reactions were performed under dim red light unless otherwise indicated. Lysine-lysine cross-linkers were DST, DSG, DSS; cysteine-lysine cross-linkers were SIA, GMBS, EMCS, and LC-SMCC (refer to the Supplemental Section for cross-linker chemical structures). Stock solutions were prepared by dissolving cross-linkers in DMSO at a concentration of 300 mM. Lysine-lysine cross-linking was performed on rhodopsin in purified ROS membranes (10 μ M rhodopsin) in 50 mM

HEPES buffer, with 100 mM NaCl, pH 7.5, containing 10-200 fold molar excess of cross-linker. The reaction was performed at 37°C for 30 min, and quenched with 10 mM Tris, pH 7.5, for 15 min. at room temperature. Cysteine-lysine cross-linking was carried out in two steps. In the first step, cysteine coupling was favored by incubating cross-linker (typically at 200-fold molar excess over rhodopsin for 30 min. at 37°C) and ROS at low pH (50 mM pyridine, 100 mM NaCl, pH 6). The membrane was then pelleted and washed to remove unreacted cross-linker and incubated in 50 mM HEPES, 100 mM NaCl, pH 7.5 for one hour to facilitate the reaction of cysteine-bound cross-linker with lysine residues. Samples were quenched with 10 mM Tris-HCl, pH 7.5 and 0.01% β -mercaptoethanol (BME).

Prior to delipidation and digestion, cysteine residues were reduced and alkylated. Disulfides were reduced with 50 mM tris(2-carboxyethyl)phosphine (TCEP) in 50 mM HEPES buffer, 100 mM NaCl, pH 7.5, for 30 min. at 37°C. Samples were alkylated with 150 mM 4-vinylpyridine(26) at room temperature for 30 min and washed several times using the same buffer to remove the 4-vinylpyridine.

Monomer purification and cleavage with CNBr. Monomeric rhodopsin was separated from multimeric cross-linked protein and contaminants by preparative tris-glycine ALS-PAGE using a mini-prep cell (Bio-Rad) containing a column gel (11 % acrylamide resolving gel (5 cm long), 4 % acrylamide stacking gel (4 cm long), 1 cm diameter) equipped with a peristaltic pump and a fraction collector. Eluted protein fractions were collected and analyzed by SDS-PAGE and Coomassie staining, and fractions containing monomeric rhodopsin were pooled and concentrated using centrifugal filters (5-10K MWCO, Millipore). Purified rhodopsin was subsequently delipidated by chloroform/methanol/water extraction (27), and the resulting pellets were washed once with acetone before being solubilized for CNBr digestion. Failure to remove methanol by acetone wash following extraction resulted in substantial methylation of the protein. A stock solution of high purity (>99%) CNBr in acetonitrile (4M) was prepared. Rhodopsin samples were first dissolved in 100% TFA, diluted to 70% TFA with water, and then chemically digested with 100-200 mM CNBr. Generally, sample vials were flushed with nitrogen, covered with aluminum foil and shaken overnight at room temperature, although equivalent results were seen after 6 hrs of incubation. Digested samples were dried and washed once with acetonitrile to remove the residual TFA and CNBr. Dried samples were stored at -80°C until further analysis.

Lysine labeling and double digest. ROS membrane was pelleted and resuspended in 50mM pyridine, 100mM NaCl, pH 7 to a final rhodopsin concentration of 10 mM. For cysteine labeling, ROS membrane was incubated with 50 – 1000-fold excess of N-ethyl maleimide for 60 minutes at 37°C and quenched with 0.01 % BME. For lysine labeling, ROS membrane was incubated with a 50-fold (for straight labeling experiments) or 200-fold (for “pulse-chase” experiments) molar excess of NHS-acetate or NHS-propionate at 37°C for 60 minutes. For pulse-chase experiments, a 5600-fold molar excess of the appropriate NHS-ester was then added for 60 min. to acetylate unreacted lysines. Reactions were quenched with 100mM Tris, pH 7. Membrane was next washed twice with 100mM Ethyl Morpholine Acetate, pH 8.03 and resuspended to a rhodopsin concentration of 1 mg / mL in the same buffer with 5 % ACN and 0.01 % β -mercapthoethanol (by vol.) for digestion with trypsin. Then, 1:5 Trypsin was added to an enzyme:rhodopsin ratio of 1:5 (Trypsin Gold Mass Spect Grade, Promega) and incubated

at 60°C overnight (16 – 24 hours). Trypsin was added again at an enzyme:rhodopsin ratio of 1:10 and the samples were incubated at 37°C for another 3 hrs. The samples were washed once with 50mM pyridine, 100mM NaCl, pH 7 to remove trypsin, reduced and alkylated, and processed by CMW extraction and CNBr digestion, as described above.

LC-MS. Rhosopsin CNBr digest products were dissolved in formic acid, diluted with an equal volume of ACN and then water to a final ratio of 10:10:80 by volume. The typical concentration, based on quantification before CMW extraction, was 0.5–1 mg/mL. A portion (5-10 μ l) was loaded onto a polystyrene/divinylbenzene column (PLRP/S, 0.2x150mm, Michrom Bioresources, Auburn, CA) and separated using buffer and gradient conditions as follows: buffer A, 5% acetic acid, 2.5% ACN and 2.5% isopropanol; buffer B, 4% acetic acid, 40% ACN and 50% isopropanol; gradient (in % B buffer), 0 for 1 min, 0-15% in 4 min, 15-60 in 30 min, 60-100 in 15 min; flow rate, 5 μ l/min. All chromatographic separations were performed at 25°C using an Agilent 1100 capillary HPLC system (Agilent Technologies, Palo Alto, CA). The column was connected directly to the mass spectrometer with no split.

Mass spectrometry was performed using an APEX II FTMS equipped with a 7.0 T superconducting magnet and an Apollo ESI ion source (Bruker Daltonics, Billerica, MA), upgraded with a mass selective quadrupole front end. Mass spectra 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 (300-2000) were passed to the FTMS analyzer cell. All spectra were acquired in positive ion mode.

Preparative liquid chromatography and MS/MS. Preparative LC separation was performed on a Hewlett Packard 1100 series LC. The LC system was outfitted with a quaternary pump, a DAD (Diode Array Detector), an autosampler, and an in-line mobile phase vacuum degasser. Samples were purified using a Michrom BioResources, Inc. PLRP-S column (5 μ particle size, 300 Å 2.0x150mm) at 30°C. The gradient was (in % B buffer, see LC-MS section for description of buffers): 0-20 in 5 min, 20-70 in 50 min, 70-100 in 10 min. The flow rate was 0.25-0.3 mL/min. The samples were prepared by re-suspending the dried CNBr digested pellet with TFA or formic acid and diluting to 10 % with buffer A (final vol. 400 μ l or less). A typical prep used 200-400 μ g of starting material (based on quantification of rhodopsin in ROS membrane) for a final concentration of ca. 1 mg/ml or less. The effluent fractions were collected using a Bio-Rad 2110 fraction collector and their mass spectra were analyzed by direct infusion on an ESI-Q-ToF mass spectrometer (Micromass Ultima API, Waters) in order to determine the fractions containing the desired cross-linked species. MS/MS analysis was performed on the purified cross-linked peptides by either ESI-FT-MS (Apex II, Bruker Daltonics) or ESI-Q-ToF-MS (Micromass, Waters). For the former, mass spectra 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 (300-2000) were passed to the FTMS analyzer cell. After the precursor ion was isolated from the mixture by setting the quadrupole mass filter to pass the m/z of interest, the MS/MS spectra were obtained by dropping the entrance to the collision hexapole to approximately -20 V. All spectra were acquired in positive ion mode.

For MS/MS using Q-ToF-MS, direct infusion of LC fractions was at 1 μ l/min. The most abundant charge state was fragmented at collision energies ranging from 16 to 65

depending on the charge state, size and fragmentation pattern of the cross-linked peptides. Scans of one second duration were collected for up to 60 minutes in the case of low-abundance cross-links. The largest cross-linked peptides and any cross-link containing peptide 50 – 86 were most effectively analyzed by ECD using FT-MS because of the high resolution necessary to resolve complex mixtures of large fragments in the former case and the difficulty of fragmenting between two adjacent lysine residues in the center of peptide 50 – 86 in the latter case.

Data analysis. Reduction of MS spectra is automated by a macro developed in our laboratory and implemented within the Xmass software package (Bruker Daltonics) as described previously.(28) For all MS scans during an LC-MS experiment, the data reduction macro de-convolutes a list of m/z values into a single set of monoisotopic masses. The program output includes the monoisotopic mass, the observed charge states and corresponding m/z values, the relative abundance of the most abundant isotope peak for that mass, and an indicator of the certainty the observed mass is the monoisotopic mass. The list of monoisotopic masses are then matched to a theoretical library of cross-modified and/or unmodified proteolysis products generated by the program ASAP (Automated Spectrum Assignment Program), described previously (16). ASAP accepts user input on protein sequence, amino acid modifications, cross-linker specificities, proteolytic method, and mass error limits to generate a theoretical library and assign observed masses.

Tandem MS data was analyzed using the in-house program MS2Links, an updated software version of MS2Assign described previously.(29) MS2Assign and MS2Links were developed for the assignment of tandem mass spectra of unmodified, labeled, and/or cross-linked peptides on the same principle as ASAP (described above). The program generates a theoretical library of expected tandem MS fragments, accepting the same protein modifications as ASAP and considering a, b, c, x, y, z, internal, and immonium ions, and additional loss of H₂O, NH₃, CO, and CO₂. Web based versions of ASAP and MS2Assign are available at <http://roswell.ca.sandia.gov/~mmyoung>.

Molecular modeling methods

Maximum cross-linker length. A 3-D structure of the cross-linker was constructed using Chem3D.(30) The bond torsion angles in the molecule were varied to find the combination of torsions that resulted in the longest distance between the reactive groups on the cross-linker (referred to as the "maximum cross-linker length"). The resulting structure was optimized using MM3(31-33) as implemented in Chem3D to ensure that the molecule was at an energy minimum. If the resulting structure was not fully extended, the bond torsions were adjusted again to find the maximum cross-linker length for the optimized bond lengths and angles. The length of the MM3 optimized structure was generally within 0.1 Å of the maximum cross-linker length as described above.

Rhodopsin structure. The A chain of the bovine rhodopsin crystal structure (1F88.pdb) was used as the starting structure for all comparisons to experimental distances. Missing atoms were modeled using the internal coordinates tables in the CHARMM22 topology definitions file. Energy calculations were performed using the CHARMM22 all atom

force field.(34, 35) Both the CHARMM(34) and NAMD(36) molecular mechanics packages were used for energy minimization and conformational search.

Molecular dynamics trajectory. Paul Crozier kindly provided the results of a 40 ns molecular dynamics simulation of dark-adapted rhodopsin in an explicit lipid bilayer.(21) We sampled this MD trajectory every 20 ps to produce a more manageable trajectory of 2000 structures from which structural data was extracted.

Conformational search. Conformational searching was performed using a Monte Carlo sampling procedure in which 1 ps bursts of high-temperature molecular dynamics were used to randomize structure. Each 1ps burst of MD was followed by energy minimization using the conjugate gradients algorithm to relax the structure into the nearest local minimum. Monte Carlo sampling was performed at both 500° K and 750° K and 1000 structures were generated at each temperature. In order to determine whether side-chain reorientations were sufficient to account for the experimental cross-links, conformational searches were done with a fixed backbone. A script, written and run in NAMD, was used for this sampling procedure.

Constrained minimization. Constraints between cross-linked atoms were added to the CHARMM energy using the NOE constraints function built into CHARMM, which models distance constraints as a soft square well potential:

$$V_{dist} = k_{dist} \begin{cases} (r_{ij} - r_l)^2, & r_{ij} < r_l \\ 0, & r_l \leq r_{ij} \leq r_u \\ (r_{ij} - r_u)^2, & r_{ij} > r_u \end{cases}$$

where r_{ij} is the distance between atom i and atom j , r_l and r_u are the lower and upper bounds on the distance, respectively, and k_{dist} is a force constant, which was set to 10 kcal/mol \AA^2 . The well-width was defined with an upper bound on the cross-linker length as described earlier. The lower bound was set equal to the sum of the van der Waals radii of the two cross-linked atoms.

Constrained energy minimization was performed in three steps. In the first step, the structure was minimized with the distance constraint active and all non-bond interactions turned off. This allowed the side-chains to move freely within the structure to satisfy the distance constraint. In step two, non-bond interactions were turned on and the structure minimized to relieve bad contacts generated by the first minimization. In the final minimization the distance constraint was turned off to allow complete relaxation of the side-chains in their new local environment.

Solvent accessibility. The COOR SURF command in CHARMM was used to compute the solvent accessible surface area (SASA) of rhodopsin for each 20 ps time-step of the molecular dynamics trajectory. The COOR SURF command in CHARMM computes the SASA using the Lee and Richards methods.(37) A probe radius of 1.4 \AA was used for all SASA calculations.

Results

Sample preparation. Integral membrane proteins pose a significant analytical challenge due to their amphiphilic nature and their tendency to aggregate and adhere to

glass, plastic, and commonly used HPLC stationary phases. However, recent progress has been made in the preparation of integral membrane proteins, including rhodopsin, for mass spectrometry (38-45). In this work, methods were developed to maximize the recovery of cross-linked rhodopsin proteolysis products for subsequent LC-MS and tandem MS analysis.

Bovine rhodopsin in the rod outer segment (ROS) membrane was first treated with either a cysteine-lysine (C-K) or lysine-lysine (K-K) cross-linking reagent under dim red light. The hetero-bifunctional C-K reagents chosen have a primary-amine reactive N-hydroxysuccinimide (NHS) ester group on one end and a sulfhydryl-reactive group (iodoalkyl for SIA, maleimide for all others) on the other and a range of linker arm lengths (refer to the Materials section for the definition of cross-linker abbreviations and the Supplemental section for chemical structures). The homobifunctional K-K reagents used have NHS ester groups at both ends and also have a range of linker lengths.

During the cross-linking reaction, the cross-linker can react with two amino acid residues on single protein or between residues on different proteins to form a covalent dimer. Similarly, dimers can be linked to other rhodopsin molecules to form multimeric reaction products. All of these scenarios are indeed observed under the reaction conditions we used (Figure 1a). The molar excess of cross-linker to protein was optimized (typically 50:1) in order to minimize the formation of multimeric protein. However, separation of monomeric and multimeric protein was necessary to ensure that only intra-molecular cross-links were being identified. Separation was achieved by preparative tube gel electrophoresis and fractionation. Fractions containing monomeric rhodopsin were pooled and concentrated (Figure 1b), and detergent and residual lipids were then removed from the protein by chloroform/methanol/water (CMW) phase separation.(27) The pellet was then repeatedly washed with acetone (to avoid methylation due to residual methanol on the pellet) prior to being dissolved in 70% trifluoroacetic and subjected to CNBr cleavage.

LC-MS. The total ion chromatogram (TIC) of the reverse phase HPLC separation and ESI-MS of CNBr digested rhodopsin is shown in Figure 2. For each sample, a consistent and reproducible cleavage pattern was observed following treatment with CNBr, facilitating the mass spectral analysis and assignment of the rhodopsin peptide fragments (Figure 2a). Several groups have reported the use of C4, C8 or C18 columns for the RP-HPLC separation of digested peptides from integral membrane proteins including rhodopsin.(41, 42) We achieved optimal separation using a polystyrene/divinylbenzene column (PLRP-S, Michrom Bioresources) for the detection of hydrophobic and cross-linked rhodopsin fragments following CNBr cleavage. Using the LC-MS methods described in the Experimental section, 99% sequence coverage was obtained, not including the single methionine residues 1 and 309. Post-translational modifications of rhodopsin were also observed, including the glycosylation of peptide 2-39 and the palmitoylation of peptide 318-348. The retinal chromophore attached to Lys 296 and the disulfide bridge between Cys 110 and Cys 187, were not observed due to sample preparation conditions. Figure 2b illustrates the total ion chromatogram of EMCS cross-linked rhodopsin, with unique peaks due to the presence of un-modified peptides, singly modified peptides (i.e. “hanging” cross-linkers formed by the reaction of EMCS with a single residue within the peptide), internally cross-linked peptides, and cross-linked peptide pairs. It is important to note that, due to their lower concentrations, the selected

ion chromatograms (SICs) of cross-linked products (e.g. Figure 2d, 50-86x310-317) are typically of much lower intensity than the SICs of precursor peptides (e.g. Figure 2c, 50-86), making optimized sample recovery, chromatographic separation, and instrument sensitivity key factors in the detection of cross-linked species.

Identification of Cross-Linked Products

Criteria for identification. The fundamental basis for the MS3-D method is the identification of inter-residue chemical cross-links from which distance constraints are derived.(46, 47) We found it necessary to develop stringent guidelines for the identification of cross-linked peptides to rule out spurious assignments. From the present work, we determined that reliable identification of cross-linked peptides by mass spectrometry requires 1) reproducible chemical or enzymatic cleavage of the protein, thereby providing a consistent pattern of control peptides, 2) complete, or at least uniformly reproducible, sequence coverage in the control LC-MS spectra, 3) the presence of peaks corresponding to putative cross-linked species only in the LC-MS spectrum of the cross-linked sample and not in the control spectra, 4) the assignment of MS and tandem MS spectra to within a low experimental mass error (Δ ppm), and 5) the detection of mass-shifted peaks corresponding to cross-linked products following a change in the cross-linker molecular weight. All cross-linking experiments were checked following these guidelines, and experiments not satisfying each criterion were discarded.

Automated peptide assignments (ASAP). For each LC-MS experiment on CNBr-digested cross-linked rhodopsin, mass spectra were calibrated by identifying a MS spectrum in the LC-MS run which contained relatively intense peaks of unmodified rhodopsin peptides known to be present in the control spectrum. (These control peaks were initially identified by standard comparison of their measured masses to expected CNBr product masses.) The exact monoisotopic masses of these peaks were then used to generate calibration constants that were applied to all spectra in the run, generally resulting in mass accuracies better than 10 ppm across the run (see Tables 1 and 2). Reduction of the MS spectra is automated by a macro developed in our laboratory and implemented in the Xmass software package (Bruker Daltonics), resulting in an output list of monoisotopic masses which could then be matched to a theoretical library of unmodified, labeled, and/or cross-linked peptides generated by ASAP (Automated Spectrum Assignment Program) as described previously.(16) For each monoisotopic mass in the experimental results, ASAP searched the library for masses within \pm 5-10 ppm and reported all matching assignments.

Identification of cross-links. Peaks assigned to cross-linked peptides by ASAP were manually checked to confirm reasonable peak shapes, isotopic distributions, and high signal to noise ratios. In order to verify the presence of putative cross-linked species, the LC-MS run and the identified peaks were then checked against the criteria outlined above (see Criteria for identification). In addition, for each cross-linked species identified (e.g. 50-86x318-348), the mass spectrum was also checked for the presence of the precursor peptides (e.g. un-modified 50-86 and 318-348) and the presence of 'hanging' cross-linkers on each precursor peptide. Since these species were generally found in greater abundance (data not shown), their presence and relative intensities were checked for consistency when assigning the putative cross-linked products.

A summary of all confirmed cross-linked peptides observed from the native state of dark-adapted rhodopsin in ROS is shown (Tables 1 and 2). Minor, but significant, differences were observed in the cross-linking pattern produced by each of the four cysteine-lysine cross-linkers, SIA, GMBS, EMCS, and LC-SMCC. Based on the assignment criteria described above, cross-links were identified between the cysteine and lysine residues present on the cytoplasmic face of rhodopsin (Figure 3) in peptides 50-86, 310-317, 318-348, and 87-143. Cross-linked peptide pairs, 50-86x310-317 and 50-86x318-348 were observed in all cysteine-lysine cross-linking experiments, however the cross-links in peptide 87-143 (C140xK141) and peptide 310-317 (C316xK311) were not detected with the shortest cross-linker, SIA (Table 1). Cross-linking results with the three homobifunctional NHS ester cross-linking reagents, DST, DSG, and DSS are also reported (Table 2). Cross-links between the lysine residues on the cytoplasmic face of rhodopsin were identified in peptides 50-86, 310-317, and 318-348. An almost identical cross-linking pattern was observed with all three lysine-lysine cross-linkers, except in the case of K325xK339 which was not detected with the shortest cross-linker, DST, although the single ‘dangling’ hydrolyzed cross-link was observed (peptide 318-348 + 2 palmitoyl groups + hydrolyzed DST, observed $M+H^+ = 3731.9402$ Da, Δ ppm = 2). No cross-links were observed involving Lys 296, the retinal attachment site, or Lys 16, located on the extracellular side of the membrane.

The low experimental error (Δ ppm) between the theoretical and observed masses of the cross-linked species is due to the high mass accuracy and resolution of Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, as well as the use of unmodified peptides for internal calibration. This degree of mass accuracy is sufficiently high enough to provide unambiguous assignment of each cross-linked species without the use of isotopically labeled reagents or MS/MS. In addition, the appearance of new peptide species in the spectra of cross-linked rhodopsin relative to control, and that these unique species are observed for cross-linkers of the same specificity but different linker length with the expected mass shifts, provides a degree of redundancy to further confirm the identity of the cross-linked products. For example, in Figure 4 spectra show a cross-linked peptide pair, 50-86 and 310-317, that is formed with four different C-K cross-linkers: SIA, GMBS, EMCS, and LC-SMCC. In each spectrum, the peak is novel relative to the control spectrum of rhodopsin CNBr digest products (not shown), and the relative mass shift between spectra corresponds to the difference in mass of the cross-linkers. In this way, using cross-linkers of different mass provides further confirmation of cross-link assignments.

MS/MS. Three rhodopsin peptides resulting from CNBr digestion contain more than one lysine. To determine unambiguously which lysines were involved in cross-link assignments involving these peptides, we performed tandem MS experiments. Due to the high MW and low abundance of the cross-linked species, we were unable to perform data-dependent LC-MS/MS experiments. Instead, cross-linked products were purified by semi-preparative HPLC and analyzed by direct infusion. Fractions were collected and tested by direct infusion Q-Tof-MS to identify those containing cross-linked species. These fractions were then analyzed by direct infusion MS/MS as described below.

Several of the cross-linked peptides posed particular problems for analysis by tandem MS. The peptide 50 – 86 contains two adjacent lysines in the center of the peptide that could not be sequenced by FT-MS/MS using collision-induced dissociation (CID) or

infrared multi-photon dissociation (IRMPD) or by Q-ToF-MS/MS using CID. However, it was found that FT-MS/MS using electron capture dissociation (ECD) resulted in fragmentation through the center of the peptide. Thus, several cross-link assignments containing peptide 50 – 86, notably 50 – 86 x 310 – 317 (K66 or K67 cross-linked to C311) and 50 – 86 x 318 – 348 (residues K66 or K67 cross-linked to K325 or K339) were disambiguated using this technique. The FT-MS/MS studies involving CID, IRMPD, and ECD techniques is being prepared in a separate report and are only summarized here (see Tables). These results show that K67 and not K66 participate in cross-linking with C316, K325, and K339. One assignment, 50 – 86 x 310 – 317 (K66/67 x K311) occurred in too low abundance to perform MS/MS on. However, as discussed below, we believe the lysine participating in this cross-link can be inferred based on other data.

The peptide 318 – 348, which contains two well-separated lysines in position 325 and 339, was fragmented effectively by Q-ToF-MS/MS using CID. The results of two such experiments involving the peptide pair 318 – 348 and 310 – 317 cross-linked with either C – K or K – K cross-linkers are shown in Figures 4 and 5, respectively. An interesting cross-linking pattern is revealed by this data: C311 cross-links selectively to K325 while K316, on the same CNBr peptide as C311, is linked preferentially to K339. Since CID can fragment at the amide bond between the epsilon amino group of lysine and the carbonyl of the linker arm, it isn't possible to entirely rule out a cross-link based on its absence in a CID MS/MS spectrum. However, extensive y- and b- ion series directly confirming the cross-links described above were observed in the MS/MS spectra (see Figures 4 and 5), suggesting that these are the major products. In the case of cross-link C316 x K325, nearly identical spectra were observed following CID fragmentation of peptides cross-linked with three different reagents, SIA, GMBS, and LC-SMCC, which have a large range of linker arm lengths. All three spectra show indirect evidence that a very small portion of the cross-linking could occur between C316 and K339, in the form of an b ion fragment containing the unmodified sequence 318 to 326. However, this is a minor peak and could also be formed by a double fragmentation at both the amide bond in the cross-link and just before P327, which in control MS/MS experiments with peptide 318 – 348 is one of the major points of fragmentation.

A summary of cross-linked peptides and residue assignments is given for C – K and K – K cross-links in Tables I and II, respectively.

Lysine reactivity. Several factors likely contribute to the pattern of cross-linking observed. These include the distance between the residues, the overall mobility of the reactive atoms, the reactivity and accessibility of the residues, the stoichiometry of the reaction, and the relative abundance of the reaction products. To address the question of relative reactivity of residues involved in cross-linking, cysteine and lysine labeling experiments were performed. It was shown previously that only two cysteine residues, C140 and C316 are reactive under mild conditions (6, 48, 49) and results under our conditions using N-ethyl-maleimide were similar (data now shown). By contrast, 4-vinylpyridine at 50mM reacts with all reduced cysteines, as shown recently by (45).

We used a two-step labeling protocol to designed to allow quantitative analysis of relative lysine reactivity by tandem MS. Lysines were first labeled with NHS-propionate using stoichiometry and buffer conditions similar to those used for lysine cross-linking. Unlabeled lysines were then alkylated with NHS-acetate at > 5000-fold

molar excess to promote complete labeling. If the assumption is made the difference between acetyl and propionyl groups will not change the ionization and fragmentation of a peptide, we can then directly compare the relative amounts of propionated (ie. reactive) and acetylated (ie. unreacted) lysines in peptides containing one or more lysines. An example is shown in Figure 7. The MS/MS spectrum of CNBr peptide 318 – 348 containing one acetyl and one propionyl group contains a b-ion series through the first lysine, K325. This lysine can be seen in both propionated and acetylated form in approximately equal proportion.

We also used this method to compare the reactivities of K66 and K67. The CNBr peptide 50 – 86 proved resistant to CID fragmentation though the lysines, so we performed a trypsin / CNBr double digest and MS/MS on the fragment 50 – 69, which is cleaved at an N-terminal Met and a C-terminal Arg, containing one propionyl and one acetyl group. Based on the y3 ion, which contains only K67, the ratio of propionyl- to acetyl-lysine was about 80:20, indicating that K67 is the more reactive of the two lysines on this peptide (data not shown). This is consistent with MS/MS data of cross-linked peptides, where only cross-links involving K67 were detected.

Computational Analysis

Cross-linker distance constraints. In order to estimate the distances between reactive atoms joined by cross-links, the length of the cross-linker between the side-chain N or S atoms attached to the linker molecule was considered. Although recent studies have reported the average span and flexibility of commonly used cross-linkers,⁽⁵⁰⁾ the cross-linking experiment takes place on a minute-to-hour time scale. Extended conformations of the linker arm are accessible in these time-scales and therefore the maximum linker length should provide a useful upper bound on experimental cross-linked distances. Tables 1 and 2 report these distances, which were determined using a 3-D structure of the cross-linker constructed in Chem3D⁽³⁰⁾ and optimized using MM3⁽³¹⁻³³⁾ (refer to Methods section). Calculated maximum lengths are in good agreement with the upper range of N-N distances reported by Houk and co-workers⁽⁵⁰⁾ for the Lys-Lys cross-linkers, DST, DSG, and DSS (the Cys-Lys cross-linkers were not considered in their study).

In principal, the minimum length conformation of the cross-linking reagent can be also used to set a lower bound on the distance of greatest separation (DGS) between two reactive atoms in a protein. However, for the reagents we have employed, these minimum lengths are only in a few cases greater than the maximal lengths of the shorter cross-linkers for which cross-links were found (only in the case where LC-SMCC cross-links were found and SIA cross-links were also, see Tables 1 and 3). Simply given the conformational flexibility of side chains (especially lysine), the DGS will generally be many Angstroms greater than the DCA (e.g., at least 12-13 Å greater for any cross-link involving a lysine residue), for which we have been able to experimentally define absolute bounds to within < 6 Å in all cases, given that the minimum possible DCA is defined by the Van der Waals radii of the reacting atoms. Because no reagents are currently available that can effectively probe the appropriate distance range to set a useful bound on the DGS, we chose to computationally evaluate each cross-linker only in terms of the ability of its maximum length to bound the DCA. Thus, each distance constraint

used in our molecular modeling corresponds to the predicted maximum length of the shortest cross-linker for which a cross-link was detected.

Comparison of experimental and structure predicted distances. Comparison of cross-linked distances to those calculated from the known structure of bovine rhodopsin was achieved using a hierarchy of computational approaches, as described in the methods. A baseline comparison was first made between the cross-linked distances on the cytoplasmic face of rhodopsin and the side-chain distances (S-N or N-N distances) measured from the energy minimized X-ray structure.(17) However, since the crystal structure represents a single static picture of the protein structure, conformational search procedures were also used to investigate other energetically realistic side-chain and/or backbone orientations that could give rise to the experimentally observed cross-links not validated by direct measurement of the crystal structure. For each cross-linked peptide, we compared the candidate inter-atomic distances required to account for the observed cross-linked peptides to ensembles compiled from a) a 40 ns molecular dynamics simulation of dark-adapted bovine rhodopsin in an explicit lipid bilayer,(21) referred to as method MD, and b) constrained energy minimization of structures from an ensemble generated by Monte Carlo conformational searches with flexible side-chains and fixed backbone, referred to as method MC-SC-CM. During constrained energy minimization of structures from the Monte Carlo ensemble, the cross-linked side-chains are biased toward orientations that bring them within the cross-linker length by inclusion of a distance constraint potential in the CHARMM22 force field. Table 3 summarizes the results of these structural comparisons.

Discussion

Recent reports on the use of covalent cross-links to constrain structural models of proteins, stimulated in part by advances in high-resolution mass spectrometry, raise important issues regarding the general applicability of the MS3-D method and the interpretation of cross-links as valid inter-atomic and inter-residue distance constraints. In this work, we describe general methods that help extend the applicability of this approach to membrane proteins and protein complexes. We also present an initial set of computational methods for relating the observed cross-linked peptides to the structure and dynamics of our model membrane protein, bovine rhodopsin.

Protein cross-linking and mass spectrometric mapping of cross-linked peptides. The mass spectrometric mapping and analysis of chemical cross-links is a challenge, especially for membrane proteins in their native environment. For rhodopsin in the ROS membrane, it is not possible to control the local concentration of protein, and covalent homodimers and heterodimers are inevitably formed during the cross-linking reaction. As a result, a micro-preparative method for high-resolution protein separation is necessary in order to isolate intra-molecularly cross-linked monomers. We found preparative tube-gel electrophoresis followed by fraction collection provided excellent molecular weight resolution (compared to size-exclusion chromatography(16)) and sample recovery (compared to slab PAGE and electroelution). Complete reduction and alkylation of all cysteines was accomplished using TCEP and 4-vinylpyridine, and chloroform:methanol:water precipitation worked well for the removal of detergents (and other reagents such as the 4-VP). Tryptic digests in the absence or presence of detergent

gave poor cleavage efficiency and low sequence coverage compared with CNBr digestion in 70% TFA, which resulted in nearly complete cleavage and complete sequence coverage along with excellent reproducibility (see also (40, 41)). We found in-gel digests were of limited use in our study, since larger hydrophobic peptides poorly from the gel and in general yields were low (data not shown).

The overall framework we advocate for MS3-D experiments is to use multiple cross-linking reagents with different reactivities and to scan the length of the cross-linker. Previous reports with this method have focused primarily on the formation of lysine-lysine cross-links.(16, 19, 28, 51) We have demonstrated that, in addition to lysine-lysine cross-links, multiple lysine-cysteine cross-linked peptides can be identified from a protein such as rhodopsin with native exposed sulfhydryls (and presumably for proteins with cysteines introduced through site-specific mutation). As outlined in the Results section above, we have established criteria and methods for data analysis that allow identification of cross-linked peptides with a high degree of reliability. These methods rely not only on the high mass resolution of FTMS, but the use of a series of cross-linking reagents with nested backbone lengths, which allow the experiment, under favorable circumstances, to define inter-atomic distances, rather than just upper constraints on inter-residue distances.(19, 20) We believe that this approach is superior to using mixed isotopically-labeled reagents and pattern recognition for data analysis and cross-linked peptide identification(52, 53) for at least three reasons: 1) splitting a low-abundance peak into a labeled and un-labeled peak reduces its signal, making it hard to definitively identify the monoisotopic ^{12}C or single ^{13}C peak, 2) isotopically-labeled reagents are unnecessary: our results show that with sufficient mass accuracy, the peaks of cross-linked peptides can be identified without isotopically-labeled reagents, and 3) if desired, commercially available reagents differing by one of a few carbons in linker length can be mixed in two or more experiments, enabling pattern recognition for identification of modified peptides without the need for expensive labeled reagents, and giving additional information on distance constraints.

As always with MS-based analysis of complex samples, issues such as ionization suppression, peak overlap, and the low relative abundance of cross-linked products can all lead to low intensity and poor mass spectrometric detection. As a result, it is difficult to draw conclusions about cross-linked peptides that are not found in experiments with linkers of any length but might be expected to form based on the crystal structure. For example, no cross-linked peptides were identified in this study involving either the N-terminal face of rhodopsin or the cytoplasmic loop III (C-III) between helix V and VI. In addition, only one cross-link, C140xK141, was found in the cytoplasmic loop II (C-II). Dangling lysine modifications were found for the N-terminal peptide 2-39 and the C-III loop peptide 208-253, suggesting that a lack of reactivity cannot account for the absence of inter-peptide crosslinks. Labeling studies with N-hydroxysuccinimidyl acetate (NHS-Ac), which specifically acetylates primary amines, have shown that K231, K245, and K248 (located in or near the third cytoplasmic loop) are all reactive to some degree, as indicated by LC-MS detection of singly, doubly, and triply acetylated products of peptide 208-253 (see Supplementary section). Distance measurements taken from the X-ray structure and from the MD simulation also do not seem to offer explanation for the complete absence of cross-links. For example, such measurements suggest that it should be possible to form the K231xK245 cross-link using the cross-linker DSS, and the

K231xK248 and the K245xK248 cross-links using the cross-linkers DSS, DSG and DST. Further investigation will be required to determine whether cross-links involving this peptide are present and simply not detected by our LC-MS methods, or if intra-molecular cross-linking is prevented, for example due to rhodopsin dimer interactions.⁽⁵⁴⁾ With regard to the N-terminal face, with our reagent set, only one cross-link would even be possible on the N-terminal face (K16xC184), and if it were present, the multiple glycoforms of the N-terminal peptide make its identification problematic.

Computational analysis. Our computational analysis was directed at answering three questions:

- 1) Are the observed cross-linked peptides compatible with the crystal structure backbone if side-chain rearrangements are allowed?
- 2) Are the observed cross-links compatible with realistic side-chain and backbone dynamics?
- 3) Does the use of cross-linkers with nested lengths allow the determination of inter-atomic distances, rather than simply upper bounds on distances?

The first question has bearing on the issue of whether the cross-linking reactions distort or denature the protein. We deemed a cross-linked peptide compatible with the X-ray structure backbone if at least one of the possible inter-atomic cross-links that could account for the observed peptide was within the distance defined by the maximal length of the cross-linking reagent, as observed by either a) the energy minimized X-ray crystal structure or b) Monte Carlo sampling of side-chain conformations followed by constrained minimization and relaxation of the relevant inter-atomic distances (method MC-SC-CM). We found that at least one such cross-linkable atomic pair distance could in fact be found to account for each observed cross-linked peptide (see Table 3). Few cross-links, however, were directly consistent with the energy minimized X-ray structure: the unambiguous cross-links K311xC316 and K141xC140, and the potential cross-link K67xK339. Further experimental work to resolve the identities of the actual atom pairs (rather than just the peptides) involved in each of the cross-links is needed to determine if the atomic-pairs that achieved the minimum DCA upon MC side-chain sampling are in fact the ones involved in the formation of the cross-linked peptide. Based on the result presented here, though, none of the observed cross-linked peptides require major backbone rearrangements from the X-ray structure, and therefore there is no indication that the cross-linking reactions somehow force the protein into non-native conformations. It should be noted that these experiments proceed very differently than those involving introduction of site-specific cysteine mutations and assaying for formation of disulfide bonds in order to provide distance constraints. In these experiments, the redox potential of the reaction mixture may provide a source of free energy that drives the protein into unfavorable conformations, and there are possible effects involving disulfide bond formation or failure of bond formation during or after folding. In MS3-D, the cross-linking reagent first attaches one end to a reactive residue on the native protein, and then the other end of the reagent samples the surrounding environment, where it may or may not come across an appropriately reactive atom on another residue. The greatest potential source of cross-links that do not reflect the native structure appears to be from denatured protein that is either present in the preparation, or potentially induced by excessive

concentrations of the alkylating cross-linking reagent. These effects were controlled in our experiments by spectroscopic measurements on the ROS preparations that confirmed that the retinal was attached in its native environment, and by using limiting concentrations of the cross-linking reagent (see Methods).

One possible objection to using distances resulting from the constrained minimization of structures from the Monte Carlo sampled side-chain conformational ensemble is that we did not use explicit solvent or lipid (due to the high computational cost) and that the sampling is done at elevated temperature. As a result, it is not possible, even after minimization, to properly determine the energetic perturbation of the sampled structures relative to the crystal structure. This consideration prompts the second question above. Molecular dynamics simulation methods allow the sampling of thermal fluctuations about the crystal structure, including backbone motion, and we can use a MD simulation with explicit lipid and solvent to address the second question raised above. One limitation to this approach, however, is due to the fact that cross-linking experiments sample conformations on a time-scale of minutes or even hours, and the MD simulation of a membrane protein with explicit solvent and lipid takes one day per nanosecond even on a massively parallel computer.⁽²¹⁾ The sets of corresponding K-K and C-K distances extracted from the rhodopsin MD trajectory at 20 ps time-intervals account for all of the unambiguous cross-links on the internally cross-linked peptides 50-86, 87-143, 310-317, 318-348 (for all cross-linker lengths observed), but do not account for any of the cross-links formed between peptides on different loops. This result suggests that 40 ns is not sufficient to sample all of the relevant conformations. This is clearly seen in the case of peptide 50-86x318-348, where, partly due to fairly small initial backbone rearrangements that occur during equilibration of the MD ensemble, the DCA of K67xK339 reaches only 9.2 Å even though the minimized X-ray structure has a distance of 6.2 Å (Table 3). This result is perhaps indicative of why MD succeeds for internal cross-links, requiring only side-chain rearrangements, but fails for longer cross-links: it is much more difficult to sample large-scale relative backbone motions than to sample local motions. Conversely, this result suggests that cross-linking can be an experimental tool to probe backbone rearrangements that can occur on microsecond and longer time scales.

Regarding the third question, of the four cross-linked peptides observed that admit unambiguous assignment of the identity of the cross-linked atoms (C140xK141, C316xK311, K325xK339 and K66xK67), the first three display the following phenomenon: cross-links are found with all but the shortest of the reagents in the series. We cannot completely rule out a role for differences in reactivity of the shortest reagents (SIA has an iodomethyl rather than a maledimide sulfhydryl-reactive moiety and DST has bulky hydroxyl side chains), but based on the fact that these reagents readily form other cross-links, this seems unlikely. (Experiments are now underway to resolve this issue using strictly homologous cross-linking reagents differing only by the linker arm length.) Using the data at hand and employing the maximum length of the shortest cross-linker as a *minimum* bound on the DCA, the implied DCA for C140xK141 and C316xK311 is 4.0 - 9.7 Å, while for K325xK339 the DCA between the two reactive amines is 6.3 - 7.4 Å. For C140xK141 and for C316xK311 the DCA found by distance constrained energy minimization of structures from the side chain Monte Carlo ensemble was 3.3 and 3.2 Å, respectively (essentially corresponding to van der Waals contact), so that in any case the lower bound of 4.0 Å on the DCA derived from the maximally

extended conformation of SIA is accurate to within an Angstrom. For K325xK339, constrained side-chain MC finds a minimum distance of 3.9 Å, indicating a somewhat larger error in the experimental lower bound on the DCA. From the MD trajectory, however, we find that close distances of approach between K325 and K339 are accompanied by a large reduction in solvent accessibility of the reactive groups (data not shown), so that steric effects may be related to the inability of this cross-link to form. This observation suggests further simulation experiments are needed, since the relevant dynamics to be examined in cross-link formation are that of a protein *with one end of the cross-linker already attached to an amino acid side-chain*. In the native protein, compact side chain conformations may be achievable which are sterically or entropically unfavorable when one of the side chains has a dangling cross-linker attached. A systematic program of experiment and simulation across many proteins is needed to understand the impact of these effects on bounding the DCA, but could result in a calibrated molecular “tape measure”(20) for general use in protein structural studies.

Exact identification of the cross-linked residues in various observed cross-linked peptides will result in a more detailed structural analysis; however the results presented here are consistent with the known structure and dynamics of rhodopsin. Earlier studies have demonstrated the flexibility of rhodopsin’s cytoplasmic loops and C-terminus.(17) Khorana and co-workers {Cai, 1999 #53; Cai, 1997 #51} observed the formation of cysteine-cysteine disulfide bridges between residues on the cytoplasmic face of rhodopsin that are greater than 12 Å apart (C_{α} - C_{α}) in the X-ray crystal structure (see review by Meng *et al.*(55)). Strictly speaking, our modeling results did not require exceptional backbone flexibility to account for the experimental results, although the observation that cross-links form between all pairs of 50-86, 310-317, and 318-348 at DCAs that are generally several angstroms shorter than those in the X-ray structure or the MD trajectory, are consistent with the hypothesis that the C-terminal peptides have substantial backbone mobility. Although the use of a broader range of cross-linking reagents of varying length, rigidity, and reactivity will aid the structural modeling of cross-linked rhodopsin (e.g., by allowing us to derive upper- and lower-bounds on the distances of closest approach for cross-linked atoms), the results presented here demonstrate the potential of MS3-D in the identification of three-dimensional distance constraints for integral membrane proteins, and as a probe for local conformational mobility.

Chemical cross-linking based structure prediction. It is not likely that *de novo* modeling based purely on cross-linked distance constraints will be feasible in the near future, however it is already demonstrated that modeling approaches that incorporate cross-link-based distance constraints (either alone or in combination with other experimental distances) can provide a useful tool in protein structure prediction. The implementation of structural restraints into threading and fold recognition studies has recently been reported by several groups.(16, 56) Lengauer and co-workers demonstrated a 30% improvement in correct fold recognition rates when using cross-linking and NMR distance constraints with the 123D threading algorithm.(57, 58) Koster and co-workers have used chemical cross-linking and mass spectrometry to identify likely models for sequences with less than 30% homology to known structures.(59) In the case of membrane proteins, our group has recently reported a computational technique for correctly identifying helix-membrane protein folds matching a pre-defined set of distance constraints from NMR NOE, chemical cross-linking, dipolar EPR, and

FRET experiments.(18) Based on these modeling results, methods that would enable the rapid identification of experimental distance constraints by chemical cross-linking and mass spectrometry would contribute significantly to advances in membrane protein structure determination. We believe that we have overcome the principal difficulties in sample preparation and processing of membrane proteins required to effectively implement such methods. By using rigorous criteria in the identification of cross-linked species, and by validating cross-linked distances based on the known X-ray structure of rhodopsin, we have demonstrated the potential use of chemical cross-linking, mass spectrometry, and modeling (MS3-D) in the three-dimensional structural analysis of membrane proteins. Examination of the cross-linking results using computational models indicates that both structural and dynamical information may be recoverable from such experiments.

Acknowledgements

We would like to thank Helgi Adalsteinsson for helpful discussions and assistance in calculating the maximum cross-linker lengths. We are indebted to Mark Stevens and Paul Crozier for sharing the results of their molecular dynamics simulations of rhodopsin, and to Judith Klein-Seetharaman for helpful discussions regarding rhodopsin structure and dynamics. 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.

References

- (1) White, S. H. (2004) The progress of membrane protein structure determination, *Protein Sci.* 13, 1948-1949.
- (2) Nakanishi, K., Zhang, H., Lerro, K. A., Takekuma, S., Yamamoto, T., Lien, T. H., Sastry, L., Back, D. J., Moquin-Patthey, C., Boehm, M. F., and et al. (1995) Photoaffinity labeling of rhodopsin and bacteriorhodopsin, *Biophys. Chem.* 56, 13-22.
- (3) Nakayama, T. A., and Khorana, H. G. (1990) Orientation of retinal in bovine rhodopsin determined by cross-linking using a photoactivatable analog of 11-cis-retinal, *J. Biol. Chem.* 265, 15762-15769.
- (4) Yu, H., Kono, M., McKee, T., and Oprian, D. (1995) A general method for mapping tertiary contacts between amino-acid residues in membrane-embedded proteins., *Biochemistry* 34, 14963-14969.
- (5) Karnik, S. S., and Khorana, H. G. (1990) Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187, *J. Biol. Chem.* 265, 17520-17524.
- (6) Resek, J. F., Farahbakhsh, Z. T., Hubbell, W. L., and Khorana, H. G. (1993) Formation of the meta II photointermediate is accompanied by conformational changes in the cytoplasmic surface of rhodopsin, *Biochemistry* 32, 12025-12032.
- (7) Sheikh, S. P., Zvyaga, T. A., Lichtarge, O., Sakmar, T. P., and Bourne, H. R. (1996) Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F, *Nature* 383, 347-350.
- (8) Wu, J., and Kaback, H. R. (1996) A general method for determining helix packing in membrane proteins in situ: helices I and II are close to helix VII in the lactose permease of Escherichia coli, *Proc. Natl. Acad. Sci. U. S. A.* 93, 14498-14502.
- (9) Altenbach, C., Klein-Seetharaman, J., Cai, K., Khorana, H. G., and Hubbell, W. L. (2001) Structure and function in rhodopsin: mapping light-dependent changes in distance between residue 316 in helix 8 and residues in the sequence 60-75, covering the cytoplasmic end of helices TM1 and TM2 and their connection loop CL1, *Biochemistry* 40, 15493-15500.
- (10) Altenbach, C., Cai, K., Klein-Seetharaman, J., Khorana, H. G., and Hubbell, W. L. (2001) Structure and function in rhodopsin: mapping light-dependent changes in distance between residue 65 in helix TM1 and residues in the sequence 306-319 at the cytoplasmic end of helix TM7 and in helix H8, *Biochemistry* 40, 15483-15492.
- (11) Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996) Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin, *Science* 274, 768-770.
- (12) Hubbell, W. L., Altenbach, C., Hubbell, C. M., and Khorana, H. G. (2003) Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking, *Adv. Protein Chem.* 63, 243-290.
- (13) Davidson, W. S., and Hilliard, G. M. (2003) The spatial organization of apolipoprotein A-I on the edge of discoidal high density lipoprotein particles: a mass spectrometry study, *J Biol Chem* 278, 27199-27207.

- (14) Huang, B. X., Kim, H. Y., and Dass, C. (2004) Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry, *J. Am. Soc. Mass Spectrom.* 15, 1237-1247.
- (15) Silva, R. A., Hilliard, G. M., Fang, J., Macha, S., and Davidson, W. S. (2005) A Three-Dimensional Molecular Model of Lipid-Free Apolipoprotein A-I Determined by Cross-Linking/Mass Spectrometry and Sequence Threading, *Biochemistry* 44, 2759-2769.
- (16) Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.* 97, 5802-5806.
- (17) Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science* 289, 739-745.
- (18) Faulon, J. L., Sale, K., and Young, M. (2003) Exploring the conformational space of membrane protein folds matching distance constraints, *Protein Sci.* 12, 1750-1761.
- (19) Novak, P., Young, M. M., Schoeniger, J. S., and Kruppa, G. H. (2003) A top-down approach to protein structure studies using chemical cross-linking and Fourier transform mass spectrometry, *Eur. J. Mass Spectrom. (Chichester, Eng.)* 9, 623-631.
- (20) Blaustein, R. O., Cole, P. A., Williams, C., and Miller, C. (2000) Tethered blockers as molecular 'tape measures' for a voltage-gated K⁺ channel, *Nat. Struct. Biol.* 7, 309-311.
- (21) Crozier, P. S., Stevens, M. J., Forrest, L. R., and Woolf, T. B. (2003) Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change, *J. Mol. Biol.* 333, 493-514.
- (22) Okada, T., Matsuda, T., Kandori, H., Fukada, Y., Yoshizawa, T., and Shichida, Y. (1994) Circular dichroism of metaiodopsin II and its binding to transducin: a comparative study between meta II intermediates of iodopsin and rhodopsin, *Biochemistry* 33, 4940-4946.
- (23) Papermaster, D. S. (1982) Preparation of retinal rod outer segments, *Methods Enzymol.* 81, 48-52.
- (24) Okada, T., Takeda, K., and Kouyama, T. (1998) Highly selective separation of rhodopsin from bovine rod outer segment membranes using combination of divalent cation and alkyl(thio)glucoside, *Photochem. Photobiol.* 67, 495-499.
- (25) Hong, K., and Hubbell, W. L. (1973) Lipid requirements for Rhodopsin regenerability, *Biochemistry* 12, 4517-4523.
- (26) Friedman, M. (2001) Application of the S-pyridylethylation reaction to the elucidation of the structures and functions of proteins, *J. Protein Chem.* 20, 431-453.
- (27) Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, *Anal. Biochem.* 138, 141-143.

- (28) Kruppa, G. H., Schoeniger, J., and Young, M. M. (2003) A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry, *Rapid Commun. Mass Spectrom.* 17, 155-162.
- (29) Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, *J. Am. Soc. Mass Spectrom.* 14, 834-850.
- (30) CambridgeSoft Corporation, C. D., Cambridge, MA, 02140 USA.
- (31) Allinger, N. Y., YH; and Lii, JH. (1989) Molecular Mechanics- The MM3 Force-Field for Hydrocarbons. 1., *J. Am. Chem. Soc.* 111, 8551-8566.
- (32) Lii, J. H. A., N. L. (1989) Molecular Mechancis: The MM3 Force-Field for Hydrocarbons 2. Vibrational Frequencies and Thermodynamics., *J. Am. Chem. Soc.* 111, 8566-8575.
- (33) Lii, J. H. A., N. L. (1989) Molecular Mechancis: The MM3 Force-Field for Hydrocarbons 3. The Van der Waals Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons., *J. Am. Chem. Soc.* 111, 8576-8582.
- (34) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem.* 4, 187-217.
- (35) MacKerell, A. D. J., Brooks, B. R., Brooks, C. L. I., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, in *The Encyclopedia of Computational Chemistry* (al, P. V. R. S. e., Ed.) pp 271-277, John Wiley & Sons, Chichester.
- (36) Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics, *Journal of Comput Physics* 151, 283-312.
- (37) Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol.* 55, 379-400.
- (38) Ablonczy, Z., Goletz, P., Knapp, D. R., and Crouch, R. K. (2002) Mass spectrometric analysis of porcine rhodopsin, *Photochem. Photobiol.* 75, 316-321.
- (39) Ablonczy, Z., Knapp, D. R., Darrow, R., Organisciak, D. T., and Crouch, R. K. (2000) Mass spectrometric analysis of rhodopsin from light damaged rats, *Mol. Vis.* 6, 109-115.
- (40) Ablonczy, Z., Kono, M., Crouch, R. K., and Knapp, D. R. (2001) Mass spectrometric analysis of integral membrane proteins at the subnanomolar level: application to recombinant photopigments, *Anal. Chem.* 73, 4774-4779.
- (41) Ball, L. E., Oatis, J. E., Jr., Dharmasiri, K., Busman, M., Wang, J., Cowden, L. B., Galijatovic, A., Chen, N., Crouch, R. K., and Knapp, D. R. (1998) Mass spectrometric analysis of integral membrane proteins: application to complete mapping of bacteriorhodopsins and rhodopsin, *Protein Sci.* 7, 758-764.
- (42) Barnidge, D. R., Dratz, E. A., Sunner, J., and Jesaitis, A. J. (1997) Identification of transmembrane tryptic peptides of rhodopsin using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Protein Sci.* 6, 816-824.

- (43) Knapp, D. R., Crouch, R. K., Ball, L. E., Gelasco, A. K., and Ablonczy, Z. (2002) Mass spectrometric analysis of G protein-coupled receptors, *Methods Enzymol.* 343, 157-161.
- (44) Kraft, P., Mills, J., and Dratz, E. (2001) Mass spectrometric analysis of cyanogen bromide fragments of integral membrane proteins at the picomole level: application to rhodopsin, *Anal. Biochem.* 292, 76-86.
- (45) Wang, X., Kim, S. H., Ablonczy, Z., Crouch, R. K., and Knapp, D. R. (2004) Probing rhodopsin-transducin interactions by surface modification and mass spectrometry, *Biochemistry* 43, 11153-11162.
- (46) Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R., and Selvin, P. R. (1999) A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry, *Anal. Biochem.* 273, 73-80.
- (47) Lykkesfeldt, J. (2000) Determination of ascorbic acid and dehydroascorbic acid in biological samples by high-performance liquid chromatography using subtraction methods: reliable reduction with tris[2-carboxyethyl]phosphine hydrochloride, *Anal. Biochem.* 282, 89-93.
- (48) Findlay, J. B., Barclay, P. L., Brett, M., Davison, M., Pappin, D. J., and Thompson, P. (1984) The structure of mammalian rod opsins, *Vision Res* 24, 1501-1508.
- (49) Resek, J. F., Farrens, D., and Khorana, H. G. (1994) Structure and function in rhodopsin: covalent crosslinking of the rhodopsin (metarhodopsin II)-transducin complex--the rhodopsin cytoplasmic face links to the transducin alpha subunit, *Proc. Natl. Acad. Sci. U. S. A.* 91, 7643-7647.
- (50) Green, N. S., Reisler, E., and Houk, K. N. (2001) Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers, *Protein Sci.* 10, 1293-1304.
- (51) Novak, P., Kruppa, G. H., Young, M. M., and Schoeniger, J. (2004) A top-down method for the determination of residue-specific solvent accessibility in proteins, *J. Mass Spectrom.* 39, 322-328.
- (52) Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis, *Anal. Chem.* 73, 1927-1934.
- (53) Collins, C. J., Schilling, B., Young, M., Dollinger, G., and Guy, R. K. (2003) Isotopically labeled crosslinking reagents: resolution of mass degeneracy in the identification of crosslinked peptides, *Bioorg. Med. Chem. Lett.* 13, 4023-4026.
- (54) Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A., and Palczewski, K. (2003) Atomic-force microscopy: Rhodopsin dimers in native disc membranes, *Nature* 421, 127-128.
- (55) Meng, E. C. B., H. R. (2001) Receptor activation: what does the rhodopsin structure tell us? *Trends Pharmacol. Sci.* 22, 587-593.
- (56) Reva, B., Finkelstein, A., and Topiol, S. (2002) Threading with chemostructural restrictions method for predicting fold and functionally significant residues: application to dipeptidylpeptidase IV (DPP-IV), *Proteins* 47, 180-193.
- (57) Albrecht, M., Hanisch, D., Zimmer, R., and Lengauer, T. (2002) Improving fold recognition of protein threading by experimental distance constraints, *In Silico Biol.* 2, 325-337.

- (58) Sommer, I., Zien, A., von Ohlen, N., Zimmer, R., and Lengauer, T. (2002) Confidence measures for protein fold recognition, *Bioinformatics* 18, 802-812.
- (59) Back, J. W., Sanz, M. A., De Jong, L., De Koning, L. J., Nijtmans, L. G., De Koster, C. G., Grivell, L. A., Van Der Spek, H., and Muijsers, A. O. (2002) A structure for the yeast prohibitin complex: Structure prediction and evidence from chemical crosslinking and mass spectrometry, *Protein Sci.* 11, 2471-2478.

Tables

Table 1. Observed $M + H^+$ (Da) and experimental error (ppm) of cysteine-lysine cross-linked rhodopsin peptides. Protonated masses correspond to the monoisotopic peak unless indicated.

Cross-linked residues	Cross-linked peptides	Observed mass of Cys-Lys cross-linked peptides			
		SIA 4.0 Å ¹	GMBS 9.7 Å	EMCS 11.1 Å	LC-SMCC 15.0 Å
C316xK67	310-317, 50-86	5272.9097 3.7 ppm	5397.9458 7.2 ppm	5425.9624 0.8 ppm	5565.0869 4.0 ppm
C316xK325	310-317, 318-348 ²	4632.4380 7.0 ppm	4756.4448 1.1 ppm	4784.4668 3.0 ppm	4923.5869 1.2 ppm
C140xK141	87-143 ³	not observed	6414.2285 4.3 ppm ⁴	6442.227 9.4 ppm ⁴	6581.4228 5.4 ppm ⁴
C316xK311	310-317	not observed	1157.5200 3.4 ppm	1185.5441 2.9 ppm	1324.6490 1.3 ppm

¹ Maximum cross-linker length (see text for details).

² Contains 2 palmitoylated cys.

³ Contains 1 pyridylethylated cys.

⁴ Monoisotopic peak not observed, assignment based on second isotope peak.

Table 2. Observed M + H⁺ (Da) and experimental error (ppm) of lysine-lysine cross-linked rhodopsin CNBr digest products. Protonated masses correspond to the monoisotopic peak unless indicated.

Cross-linked residues	Cross-linked peptides	Observed mass of Lys-Lys cross-linked peptides		
		DST 6.3 Å ¹	DSG 7.4 Å	DSS 11.3 Å
K67xK325 and K67xK339	50-86, 318-348 ²	7955.3447 1.6 ppm ⁴	7937.4268 5.5 ppm ⁴	7979.4067 3.0 ppm ⁴
K311xK66/67	50-86, 310-317 ³	5451.9126 5.8 ppm	5434.9717 0.3 ppm ⁴	5475.9912 4.7 ppm
K311xK339	310-317 ³ , 318-348 ²	4811.4688 1.0 ppm ⁴	4793.5093 4.1 ppm ⁴	4835.541 0.9 ppm ⁴
K66xK67	50-86	4355.3784 9.6 ppm	4338.4404 2.1 ppm ⁴	4379.4756 4.0 ppm
K325xK339	318-348 ²	not observed	3696.9828 4.6 ppm ⁴	3737.9888 5.6 ppm

¹ Maximum cross-linker length (see text for details).

² Contains 2 palmitoylated cys.

³ Contains 1 pyridylethylated cys.

⁴ Monoisotopic peak not observed, assignment based on second isotope peak.

Table 3. Comparisons of minimum cross-linker length and theoretical minimum approach distances for theoretical (*italics*) and observed (bold**) cross-linked atoms.**

Cross-linked peptide(s)	Cross-linked residues	Max. reach of shortest observed cross-link ¹ (Å)	Distance of Closest Approach (DCA) ² (Å)		
			X-ray	MC-SC-CM	MD
87-143	K141xC140	9.7	9.5	3.3	7.3
310-317	K311xC316	9.7	9.3	3.2	8.8
50-86	K66xK67	6.3	13.4	3.1	5.7
318-348	K325xK339	7.4	16.1	3.9	4.9
208-257	K231xK245				
	K245xK248				
	K248xK231				
50-86, 310-317	<i>K66xC316</i>		<i>13.1</i>	<i>3.2</i>	<i>11.0</i>
	K67xC316	4.0	9.2	3.1	9.9
318-348, 310-317	<i>K339xC316</i>		<i>11.7</i>	<i>3.4</i>	<i>9.0</i>
	K325xC316	4.0	15.2	7.5	9.6
50-86, 310-317	K66xK311	6.3	22.3	11.1	23.3
	K67xK311	6.3	14.5	5.8	18.8
318-348, 310-317	<i>K311xK325</i>		<i>19.1</i>	<i>6.4</i>	<i>13.5</i>
	K311xK339	6.3 (7.4)	16.9	5.2	14.9
50-86, 318-348	<i>K66xK325</i>		<i>18.3</i>	<i>7.6</i>	<i>17.3</i>
	<i>K66xK339</i>		<i>9.9</i>	<i>2.9</i>	<i>7.1</i>
	K67xK325	6.3 (11.1)	18.5	6.4	10.7
	K67xK339	6.3 (7.4)	6.2	3.0	9.2

¹ Length between reactive atoms of the most extended structure of the shortest cross-linker for which a cross-link was detected. If the cross-link was ambiguous, the shortest cross-link for which ms/ms was performed to identify the exact residues involved is shown in parenthesis.

² The distance of closest approach for pairs of cross-linked atoms was determined as follows: X-ray, refers to energy (CHARMM22) minimized X-ray structure (1F88) (K339 side-chain atoms beyond the C β atom were modeled using the internal coordinates table in the CHARMM22 lysine topology); MC-SC-CM, refers to constrained energy minimization of structures from the Monte Carlo conformational search with fixed backbone; MD refers to the 40 ns MD simulation.

Figures

Figure 1. Cross-linking and purification of monomeric dark-adapted rhodopsin as visualized by Coomassie-stained 4-20% Tris-Glycine SDS-PAGE. (a) Lane 1, Mark 12™ Protein ladder (Invitrogen); lane 2, reduced and alkylated rhodopsin in ROS (control); lanes 3-7, reduced and alkylated rhodopsin in ROS following cross-linking; (b) Lanes 1 & 8, Mark 12™ Protein ladder (Invitrogen); lanes 2-7, Cross-linked monomeric rhodopsin fractions after purification by tube gel electrophoresis.

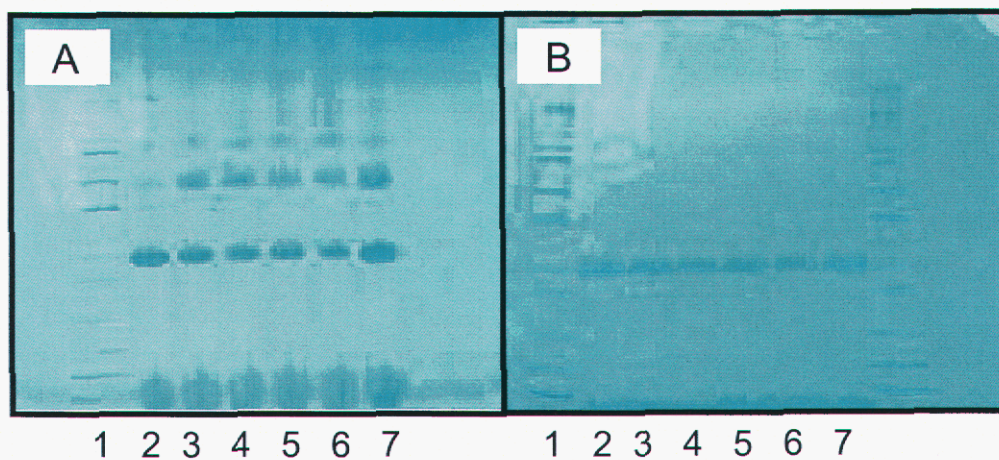


Figure 2. Total ion chromatograms (TICs) of the reverse phase HPLC separation on a PRLP-S column and ESI-FTMS of CNBr-digested (a) rhodopsin and (b) EMCS cross-linked rhodopsin. Selected ion chromatograms (SICs) from the EMCS cross-linked spectrum of the most abundant isotope peak for (c) peptide 50—86 cross-linked to peptide 310—317 (expanded vertically 4-fold) and (d) the unmodified CNBr peptide 50—86. The SICs are shown to scale overlaid on the TIC. All post-translational modifications were observed; free cysteines were pyridylethylated.

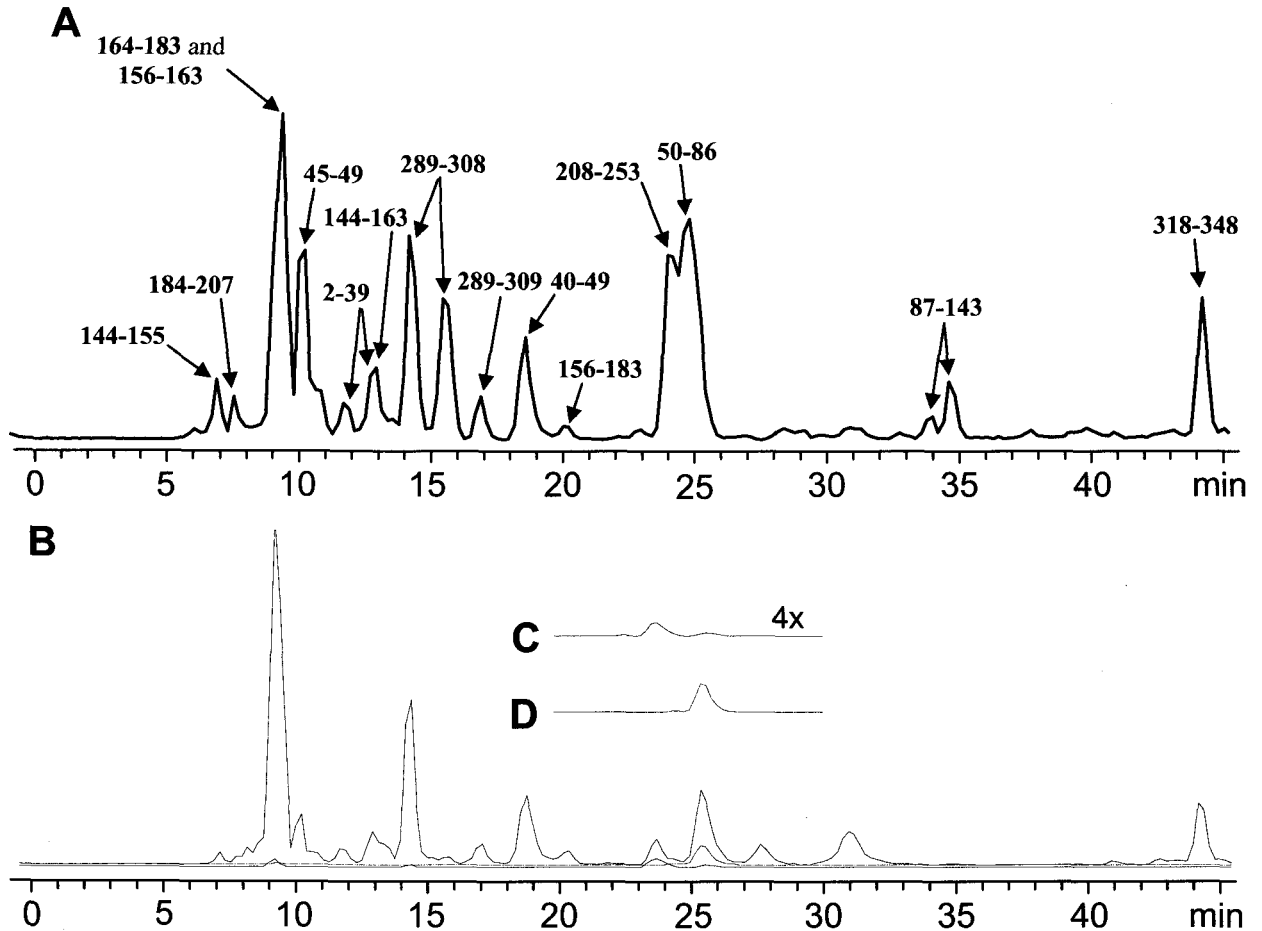


Figure 3. Mass spectra of rhodopsin CNBr digest peptides 50 – 86 and 310 – 317 cross-linked with Cys-Lys cross-linkers of different linker arm length. A single spectrum from an LC-MS experiment is shown for each cross-linker and the monoisotopic peak is indicated for (a) SIA, m/z 1055.3881⁺⁵, signal/noise (S/N) = 241.4, (b) GMBS, m/z 1080.38319⁺⁵, S/N = 717.7, (c) EMCS, m/z 1085.9978⁺⁵, S/N = 462.7, (d) LC-SMCC, m/z 1113.8155⁺⁵, S/N = 344.6. S/N is based on most abundant isotope peak. See Table 1 for $M + H^+$ (Da) and experimental error (Δ ppm).

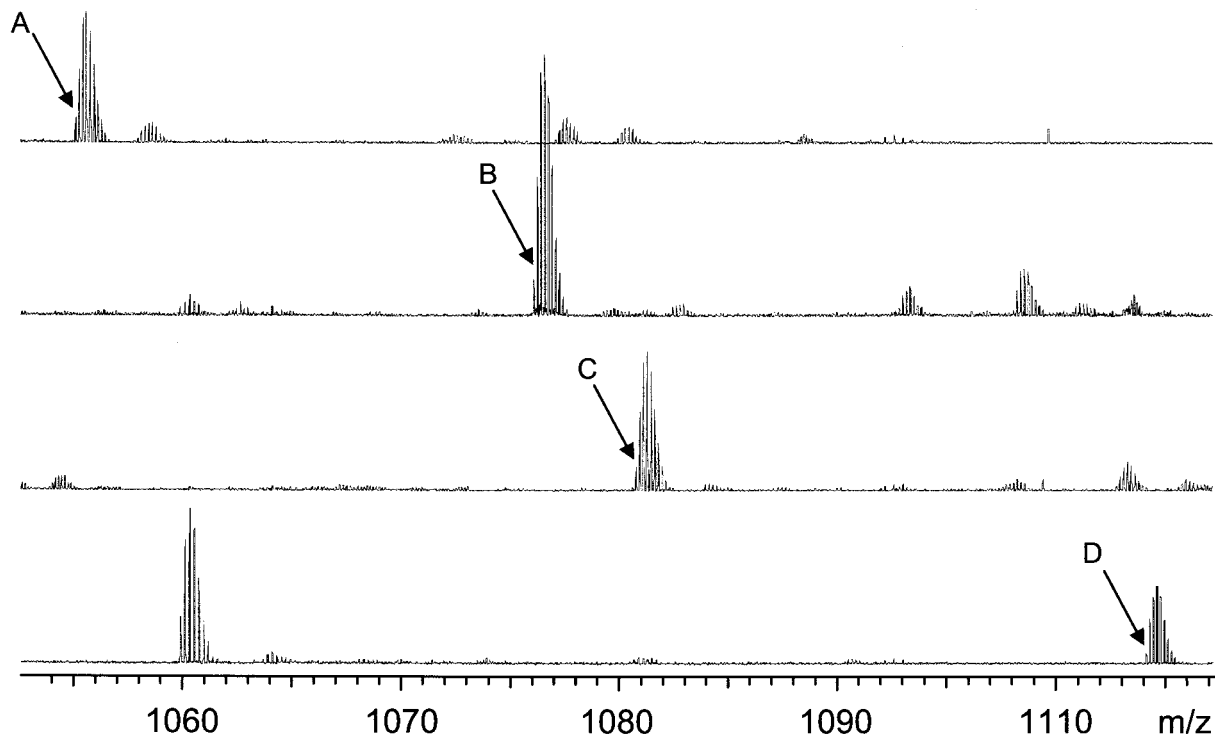


Figure 4. ESI-Q-ToF-MS2 of rhodopsin CNBr digest fragments 310 – 317 and 318 – 348 cross-linked with the hetero-bifunctional Cys/Lys-specific reagent LC-SMCC. A cross-link was selectively formed between Cys316 of peptide 310 – 317 and Lys325 of peptide 318 – 348. The spectrum is labeled with major b-, y-, and internal ion series; a cross-link between fragments is indicated by an ‘x’. Cys322 and Cys323 are palmitoylated and Met317 is modified to homoserine lactone by CNBr cleavage. The largest peaks have been truncated.

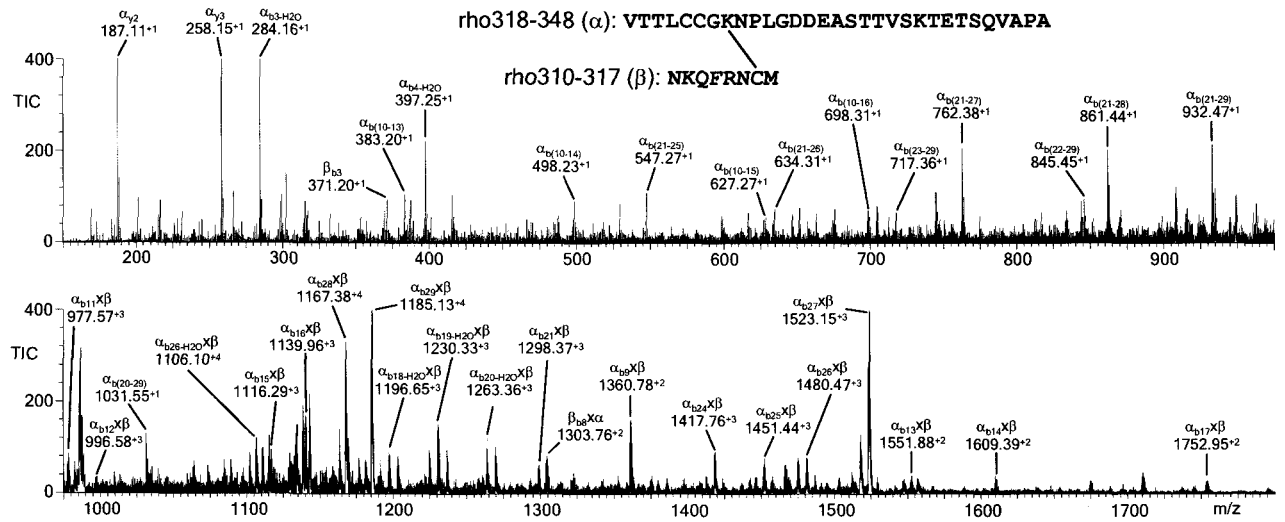


Figure 5. ESI-Q-ToF-MS2 of rhodopsin CNBr digest fragments 310 – 317 and 318 – 348 cross-linked by the homo-bifunctional Lys/Lys-specific reagent DSG. The cross-link was selectively formed between between Lys311 of peptide 310 – 317 and Lys339 of peptide 318 – 348. The spectrum is labeled with major b-, y-, and internal ion series; a cross-link between fragments is indicated by an 'x'. Cys322 and Cys323 are palmitoylated, Cys316 is pyridylethylated, and Met317 is modified to homoserine lactone by CNBr cleavage. The largest peaks have been truncated.

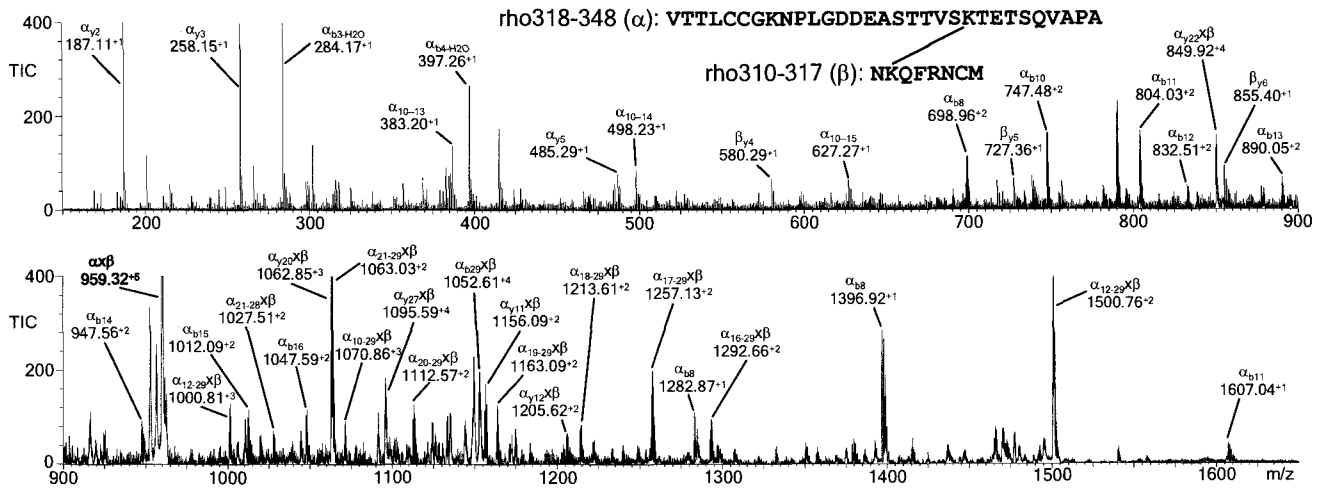


Figure 6. Cytoplasmic face of rhodopsin displaying chemically reactive Cys and Lys side chains. Solid lines indicate un-ambiguous cross-links, including C316xK67, C316xK325, C316xK311, K67xK339, K67xK325, K66x67, and K325xK339. Dashed lines represent ambiguous cross-link assignments, including C316xK66/67 and a cross-link involving K231/K245/K248. Helical domains I – VIII and loops CI – CIII are labeled, arrows indicate the C-terminal direction of the loops.

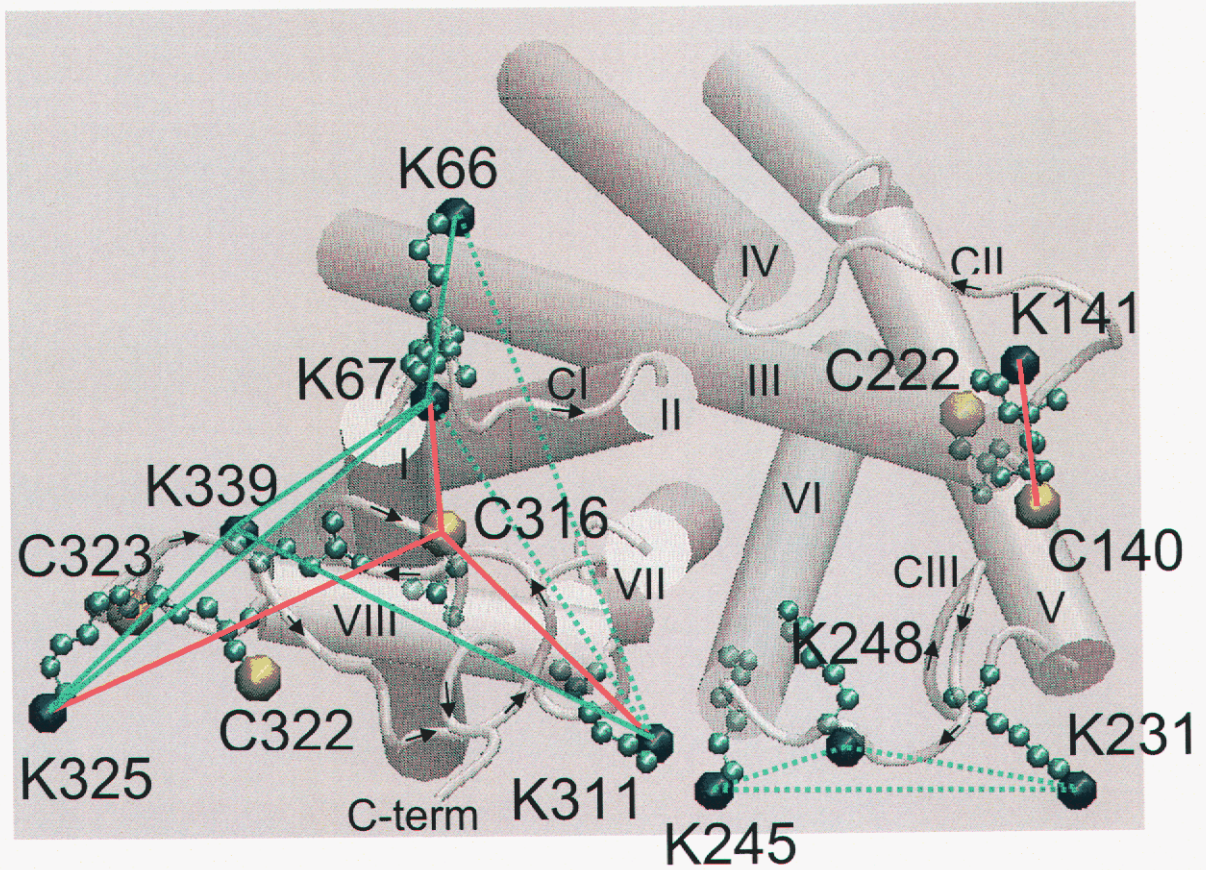
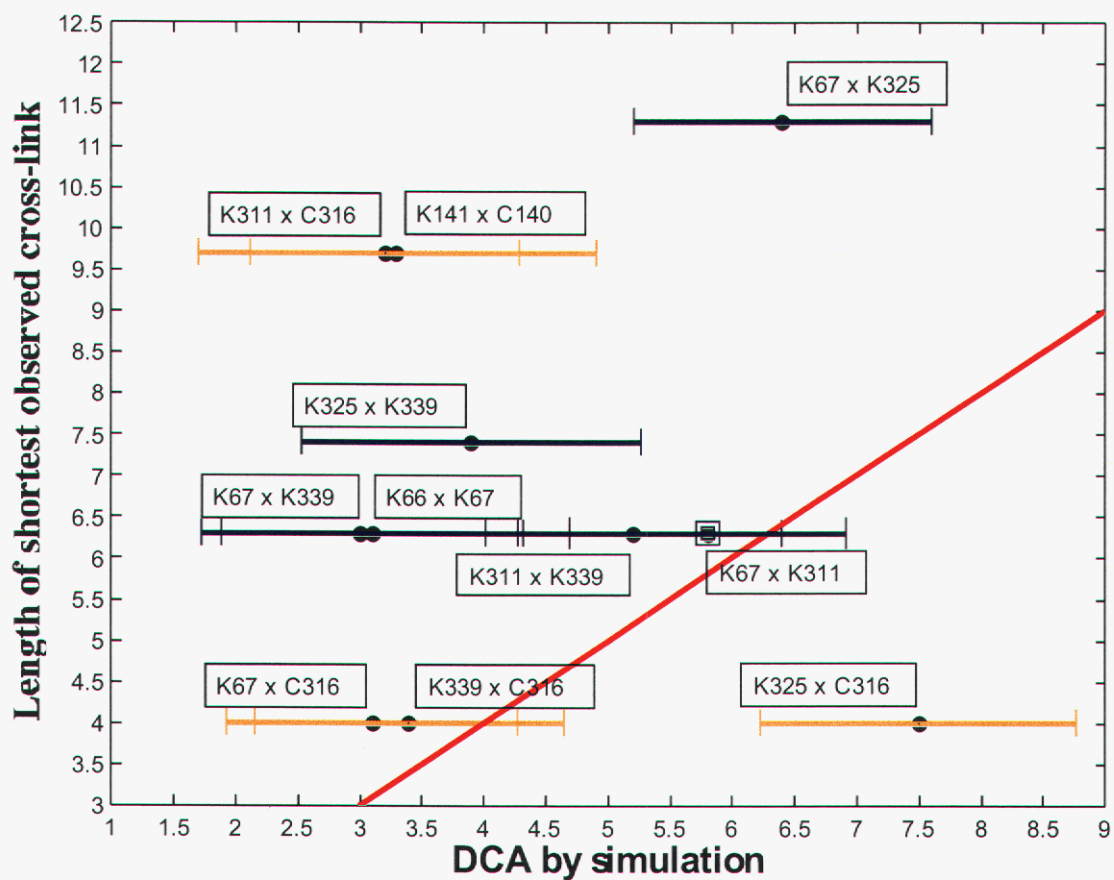
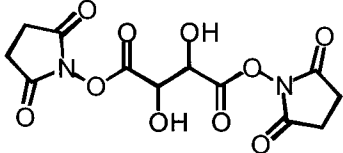
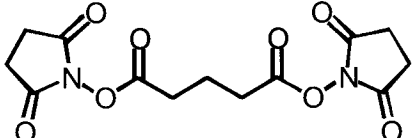
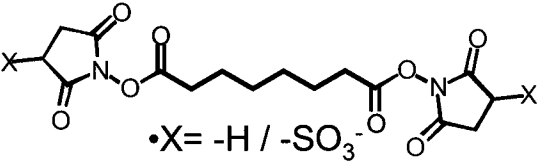


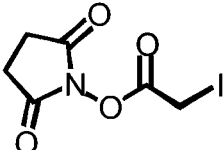
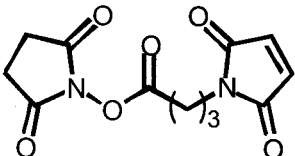
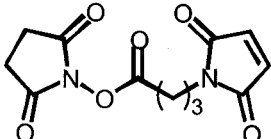
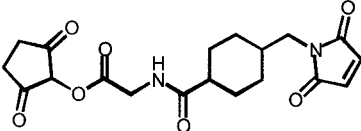
Figure 7. The simulated distance of closest approach (DCA, see Methods) between for cross-linked atom pairs is plotted against the minimum cross-linker arm length observed to form a cross-link between each pair. Horizontal error bars correspond to the vectorial sum of C alpha RMSDs for the cross-linked residues taken from the crystal structure B factors. The DCA is less than the cross-linker arm length, except for K325 cross-linked to C316, suggesting that additional movement beyond that considered in the DCA simulation is required for this cross-link to form.



Supplemental Figure 1. Lysine-Lysine Cross-linkers

Cross-linker	Structures	N-N distances (Å)
DST		6.32
DSG		7.40
DSS / BS ³	 <p data-bbox="628 886 842 922">•X= -H / -SO₃⁻</p>	11.25

Supplemental Figure 2. Cysteine-Lysine Cross-linkers

Cross-linker	Structures	S-N distances (Å)
SIA		4.0
GMBS		9.7
EMCS		11.1
LC-SMCC		15.0

This page is intentionally left blank

Section II: Advanced Mass Spectrometry Techniques

Unambiguous Assignment of Intra-Molecular Chemical Cross-Links in Modified Mammalian Membrane Proteins by Fourier Transform-Tandem Mass Spectrometry

Petr Novak^{1}, William E. Haskins^{1,2*}, Marites J. Ayson¹, Richard B. Jacobsen¹, Joseph S. Schoeniger¹, Michael D. Leavell¹, Malin M. Young¹, and Gary H. Kruppa^{1,3}*

Sandia National Laboratories¹, Livermore, CA 94551 USA

Current affiliations: Genentech², South San Francisco, CA 94080 USA; Bruker Daltonics³, Billerica, MA 01821 USA

*These authors contributed equally.

Unambiguous Assignment of Intra-Molecular Chemical Cross-links in Modified Mammalian Membrane Proteins by FT-MS/MS

Abstract

Fourier-transform tandem mass spectrometry (FT-MS/MS) can be used to unambiguously assign intra-molecular chemical cross-links to specific amino acid residue(s) when two or more possible even when cross-linking sites are adjacent in the cross-linked protein. Bovine Rhodopsin (Rho) in its dark-adapted state was intra-molecularly cross-linked with lysine-cysteine (K-C) or lysine-lysine (K-K) cross-linkers to obtain inter-atomic distance information. Large, multiply-charged, cross-linked peptide ions containing adjacent lysines, corresponding to Rho₅₀₋₈₆ (K₆₆ or K₆₇) cross-linked to Rho₃₁₀₋₃₁₇ (C₃₁₆) or Rho₃₁₈₋₃₄₈ (K₃₂₅ or K₃₃₉), were fragmented by collision-induced dissociation (CID), infrared multi-photon dissociation (IRMPD), and electron capture dissociation (ECD). Complementary sequence-specific information was obtained by combining cross-link assignments; however, only ECD revealed full palmitoylation of adjacent cysteines (C₃₂₂ and C₃₂₃) and cross-linking of K₆₇ (and not K₆₆) to C₃₁₆, K₃₂₅, and K₃₃₉. ECD spectra contained crucial c- and z-ions resulting from cleavage of the bond between K₆₆ and K₆₇. To our knowledge this is the first demonstration that ECD can be used to characterize S-linked fatty acid post translational modifications. The comprehensive fragmentation of large peptides by CID, IRMPD, and particularly ECD, in conjunction with the high resolution and mass accuracy of FT-MS/MS, is shown to be a valuable means of characterizing mammalian membrane proteins with both chemical- and post-translational modifications.

Introduction

Cross-linked peptides and proteins may be formed as part of natural or pathological post-translational modifications (PTMs) or during application of methods where covalent cross-linking reagents are used to define inter- or intra-molecular distance constraints. Inter-molecular cross-linking studies have focused on defining protein complex topology and protein-protein interactions, while intra-molecular cross-linking is primarily used for modeling the tertiary structure of proteins. Both types of experiments depend on deep extensive coverage of the cross-linked peptides and proteins which must be obtained despite the relatively low concentration of cross-linked species, missed cleavages, and various modifications.

A bottom-up approach, referred to as MS3D¹, was introduced in 1999 that combines the use of intra-molecular chemical cross-linking, proteolysis, and mass spectrometry to identify inter-residue cross-links in peptides, and then uses these cross-links to constrain three-dimensional structural models of proteins. Subsequent bottom-up MS3D studies²⁻⁵ and more recent top-down studies of intact cross-linked proteins^{6,7} indicate that these approaches have promise for defining a significant number of inter-atomic distance constraints (see reviews^{8,9}). However, it is particularly difficult to unambiguously assign intra-molecular chemical cross-links to specific amino acid residue(s) when two or more possible cross-linking sites are adjacent in the sequence of the cross-linked protein.

Tandem mass spectrometry (MS/MS) has been used throughout these studies to provide sequence-specific identification of cross-linked peptides and to improve confidence in cross-link assignments. However, in the bottom-up approach to MS3D, cross-linked peptide ions tend to be relatively large and of high mass-to-charge (m/z), making extensive fragmentation by MS/MS problematic. Large peptides are often formed from cross-linked proteins during proteolysis due to missed cleavages and the fact that cross-links can join what would normally be two individual proteolytic peptides. Low charge states are often formed during electrospray ionization due to cross-linking of basic amino acid residues because the most commonly employed cross-linking reagents react with primary amino groups at lysine and the amino terminus. Moreover, the number of peptides with the same nominal mass but different amino acid sequence increases exponentially with the number of amino acid residues in the peptide. Thus, the high resolution and mass accuracy of Fourier-transform mass spectrometry (FT-MS)¹⁰ and the ability to comprehensively fragment large peptides by FT-MS/MS¹¹ are critical for unambiguous assignment of intra-molecular cross-links^{6,12}.

The bottom-up MS3D method consists of several steps including: 1) cross-linking the protein in the native state, 2) purification and proteolysis of cross-linked monomers, 3) analysis of peptides using liquid chromatography (LC) and mass spectrometry (MS), 4) determination of cross-linked peptides by exact mass and cleavage specificity, 5) assignment of cross-links to specific amino acid residue(s) by MS/MS, 6) verification of cross-link assignments, and 7) calculation of inter-atomic distance constraints. Post-source decay (PSD)^{1,2} and collision-induced dissociation (CID)^{2,5} were successfully applied in previous studies to provide b- and y-type ions; however, these and other vibrational excitation techniques (e.g., infrared multi-photon dissociation-IRMPD¹³) often fail to discriminate the specific amino acids involved in a cross-link. Recently,

electron capture dissociation (ECD)¹⁴⁻¹⁶, a non-ergodic process, has proven to provide extensive backbone fragmentation for ions with labile PTMs^{17, 18}. Labile bonds to the modifying groups are usually the first to break during MS/MS with CID or IRMPD, making it difficult to localize these modifications, whereas ECD fragments the backbone without dissociation of these labile bonds. Therefore, we evaluated FT-MS/MS with CID, IRMPD, and ECD to determine how well they unambiguously assign intra-molecular chemical cross-links to specific amino acid residue(s) when two or more possible cross-linking sites are adjacent in the cross-linked protein.

The membrane-bound G-protein coupled receptor (GPCR) bovine Rhodopsin (Rho) proved to be a suitable target for these studies because it contains 6 unmodified cysteines and 10 unmodified lysines (K₆₆ and K₆₇ adjacent) that are potential sites for lysine-cysteine (K-C) and lysine-lysine (K-K) cross-linking. Rho is the active form of the ~40 kDa protein opsin (348 amino acids) that is modified by a retinal cofactor at K₂₉₆. Following light-induced isomerization of 11-*cis*-retinal to all-*trans*-retinal, Rho changes conformation and activates the vertebrate vision signaling pathway¹⁹. Membrane proteins, such as the GPCRs, are promising targets for MS3D because, despite their importance to human biology and the pharmaceutical industry, less than 90 membrane proteins have been resolved by X-ray crystallography and NMR spectroscopy²⁰. Non-traditional structural determination methods for membrane proteins, such as MS3D, that can overcome the challenges of resolving various PTMs, protein instability, and aggregation are therefore of great potential utility to current structural genomics efforts.

Experimental

Materials. Frozen bovine retinas were purchased from Schenk Packing Company, Inc. (Stanwood, WA). The cross-linking reagents, including disuccinimidyl suberate (DSS), N-succinimidyl iodoacetate (SIA), and N-(ϵ -maleimidocaproyloxy) succinimidyl ester (EMCS) were purchased from Pierce (Rockford, IL).

Sample preparation. Bovine Rhodopsin (Rho) in the rod outer segment was cross-linked in its native state under dim red light using K-C and K-K cross-linkers. K-K cross-linking was achieved by reaction with DSS, and K-C cross-linking was achieved by reaction with SIA or EMCS. The cross-linking reaction was performed by adding 100 fold molar excess of cross-linker in 50 mM pyridine buffer at pH 7 with 100 mM NaCl. Following the cross-linking reactions, samples were reduced with 50 mM tris(2-carboxyethyl)phosphine and alkylated with 150 mM 4-vinylpyridine. Monomeric Rho was separated from inter-molecular cross-linked species and contaminants by preparative PAGE using a miniprep cell (Bio-Rad, Hercules, CA). Purified Rho was precipitated and detergent was removed by chloroform/methanol/H₂O extraction. Rho was proteolyzed using 100-200 mM CNBr in 70% TFA under nitrogen overnight. Large cross-linked peptides containing adjacent lysines were purified by semi-preparative HPLC on a PLRP-S column (5 μ m, 300A, 2.0x150mm, Michrom Bioresources) using Hewlett Packard 1100 series LC system. The LC elution conditions were as follows (% v/v): solvent A = 2.5:2.5:5:90 isopropanol/acetonitrile/acetic acid/water; solvent B = 50:40:4:6 isopropanol/acetonitrile/acetic acid/water; the gradient method = 0 min - 100%A, 5 min -

80%A / 20%B, 50 min – 30%A / 70%B, 60 min – 0%A / 100%B, 70 min – 0%A / 100%B; flow rate = 0.3 mL/min. Eluent fractions were collected using a BioRad 2110 fraction collector.

FT-MS/MS. Fractions containing cross-linked peptides were analyzed by direct infusion on an APEX II FTMS instrument equipped with a 7.0 T superconducting magnet and an Apollo electrospray ionization (ESI) ion source (Bruker Daltonics, Billerica, MA) upgraded with a mass selective quadrupole (Q) front end²¹. IRMPD was performed with a 25 W infrared CO₂ laser (Synrad, Model No. 48-2-2, Mukilteo, WA). ECD was performed with an indirectly heated cathode operated at 1.7 amps of heater current (Heatwave, Part No. 101105, Crescent Valley, BC). The indirectly heated cathode was spot-welded in-house, directly on the standard Bruker Daltonics mounting flange for the internal filament, with no additional grids. Mass spectra were obtained with the QFTMS by accumulating ions in the ESI source hexapole and running the Q mass filter in non mass-selective RF-only mode so that ions of a broad *m/z* range (300-2000) were passed to the FTMS analyzer cell. The species of interest were isolated in the gas-phase by setting the Q mass filter to pass the *m/z* for ions of interest within a 3.5 *m/z* window. After a clean selection of the desired precursor ion had been confirmed, fragmentation was induced by: a) dropping the potential of the collision cell (CID); b) transferring the intact molecular ions to the analyzer cell and irradiating with a CO₂ laser (IRMPD); or c) transferring the intact molecular ions to the analyzer cell and irradiating with an electron beam (ECD). All MS/MS spectra were acquired in the positive ion mode with 1M data points and 128 time-domain transients. Typical conditions for fragmentation of the pentuply charged cross-linked peptide α -XL- β 1 include: dropping the collision cell potential to -24.5 V for CID, irradiating for 50 msec with a 25 W CO₂ laser run at 35% of full power for IRMPD, and pulsing the electron energy to a value of -5 V with an irradiation time of 3.5 msec for ECD (at all other times the electron emitting surface was biased to +15 V to prevent electrons from being emitted toward the analyzer cell). Automated data-reduction and assignment of the fragmentation spectra were performed by an in-house macro⁶ and software package (MS2Links)^{1,2}, respectively. Peaks in the MS/MS spectra are labeled by the most probable ion assignments; however, inter-residue cleavage sites were deduced solely by unambiguous ion assignments.

Results and Discussion

Proteolysis and MS. Cleavage of membrane proteins by chemical or enzymatic methods can be challenging due to the poor reaction conditions (low pH, detergent, and/or high organic) required for solubility and the poor recovery of hydrophobic peptides in more favorable conditions. In-gel trypsin digestion of Rho cleaved the bond between K₆₆ and K₆₇, but a poor yield and/or recovery of tryptic peptides and missed cleavages were observed (data not shown). These problems have often been reported for tryptic digests; therefore, we investigated CNBr for proteolysis of Rho at conditions optimal for protein solubility (low pH and high organic). Complete CNBr digestion of Rho should yield 9 large peptides (2-6 kDa) and 6 small peptides (0.4-1.4 kDa) following C-terminal cleavage at 16 methionines. In-solution CNBr digestion of purified cross-

linked monomers produced 9 large peptides and 6 small peptides with 100% sequence coverage, few missed cleavages, and good yield as expected (data not shown).

Cross-linking. Nine of 11 lysines (K) and 3 of 6 cysteines (C) are possible amino acid residues for K-C and K-K cross-linking based on the dark-adapted structure for Rho¹⁹ (i.e., unmodified residues on the intra-cellular face of Rho) and residue-specific accessibility studies (data not shown). There are 27 possible K-C cross-links (i.e., $K\Sigma C$ where $K=9$ and $C=3$) and 36 possible K-K cross-links (i.e., $\Sigma K-1$ where $K=9$) for singly cross-linked Rho. Cross-links with only 1 of 2 adjacent lysines (K_{66} or K_{67}) can be expected to be particularly difficult to assign unambiguously.

Following chemical cross-linking of Rho using various K-K or K-C cross-linkers and CNBr proteolysis, new peptides were observed and identified as cross-linked pairs of CNBr peptides. However, several CNBr peptides contained more than one K residue, making unambiguous assignment of the residues involved in the cross-link impossible. One CNBr peptide was particularly problematic. This 37-aa peptide, Rho₅₀₋₈₆, contains 2 adjacent lysines (K_{66} and K_{67}) and was found to form cross-links with two other peptides. Subsequent digestion of these large cross-linked peptides with trypsin, which cleaves at K and R residues, was hindered by poor solubility. Moreover, CNBr digests (and double digests with CNBr + trypsin) failed to cleave the bond between K_{66} and K_{67} . Therefore, large cross-linked peptides containing K_{66} and K_{67} were analyzed by FT-MS/MS.

FT-MS/MS. Cross-linked peptides formed from two large peptides (α at 4240.4185 Da and $\beta 2$ at 3598.9348 Da) containing 2 lysines each, and one small peptide ($\beta 1$ at 991.4663 Da) containing 1 cysteine, were fragmented by CID, IRMPD, and ECD. Annotation of the cross-linked peptides as α - and β - chains was described previously². Figure 1 shows that Rho₅₀₋₈₆ (α) contains 2 adjacent lysines (K_{66} and K_{67}) and 2 possible K-C cross-links when cross-linked with Rho₃₁₀₋₃₁₇ ($\beta 1$) at C_{316} . Likewise, there are 4 possible K-K cross-links when Rho₅₀₋₈₆ (α) is cross-linked to Rho₃₁₈₋₃₄₈ ($\beta 2$) at K_{325} and/or K_{339} . In the latter pair, both chains are large and contain two K residues each, making unambiguous assignment of the points of cross-link attachment especially difficult. In either case, α -XL- $\beta 1$ or α -XL- $\beta 2$, where XL is the cross-linker, cleavage of the bond between K_{66} and K_{67} in α is the most critical for unambiguous assignment of intra-molecular cross-links. Note that the methionines at sequence positions 86 and 317 are converted to homo-serine lactones, because of the cyanogen bromide digestion used in this work.

α -XL- $\beta 1$. CID and IRMPD of the sextuply charged cation $[M+EMCS+6H]^{6+}$ at 905.1676 m/z ($\Delta m = 0.6$ ppm) from the cross-linked peptide α -EMCS- $\beta 1$ (5424.9587 Da) yielded abundant b- and y-type ions (Figures 2A and 2B), respectively. Sequence-specific identification of cross-linked peptides such as this improves confidence in cross-link assignments. However, the product ion spectra were complicated by numerous internal fragment ions and cleavage between K_{66} and K_{67} was never observed. Based on our CID and IRMPD results, we are unable to determine whether K_{66} or K_{67} is cross-linked to C_{316} .

ECD of the $[M+EMCS+6H]^{6+}$ ion yielded abundant c- and z-type ions with few internal fragment ions (Figure 3A). Likewise, ECD of the sextuply charged cation $[M+SIA+6H]^{6+}$ at 879.6533 m/z ($\Delta m = 0.7$ ppm) from the cross-linked peptide α -SIA- $\beta 1$ (5271.8797 Da) provided extensive backbone fragmentation (Figure 3B). ECD of the

pentuply charged cation $[M+5H]^{5+}$ for the control (unmodified) peptide α at 849.0904 m/z ($\Delta m = 0.7$ ppm, 4240.4185 Da) is shown in Figure 3C. Cleavage of the bond between K_{66} and K_{67} was observed in every case. Thus, ECD revealed that K_{67} (and not K_{66}) is cross-linked to C_{316} . This was clearly demonstrated by comparing the ECD spectrum for control peptide α (Figure 3C) with the ECD spectra for cross-linked peptides (Figures 3A-B). Figure 3C shows the presence of $c_{17\alpha}$ and $z_{21\alpha}$ ions for K_{66} and $c_{18\alpha}$ and $z_{20\alpha}$ ions for K_{67} . However, the $c_{18\alpha}$ and $z_{20\alpha}$ ions are replaced by $c_{18\alpha+M\beta}$ and $z_{20\alpha+M\beta}$ ions in Figures 3A-B (where $M\beta = XL+\beta 1$, $\beta 1 = 991.4663$ Da, and $XL = 193.0739$ Da for EMCS or 39.9949 Da for SIA). N-terminal cleavage at prolines was never observed by ECD, as expected^{14, 16}. While complementary sequence-specific information was obtained by combining cross-link assignments for α -XL- $\beta 1$ from CID, IRMPD, and ECD, only ECD spectra contained crucial c- and z-ions resulting from cleavage of the bond between K_{66} and K_{67} .

α -XL- $\beta 2$. Rho, like most mammalian membrane proteins, contains several PTMs that further compound the problem of unambiguously assigning intra-molecular chemical cross-links to specific amino acid residue(s). However, ECD of the septuply charged cation $[M+DSS+7H]^{7+}$ at 1140.6378 m/z ($\Delta m = 1.0$ ppm) from the cross-linked peptide α -DSS- $\beta 2$ (7977.4214 Da) revealed full palmitoylation of adjacent cysteines (C_{322} and C_{323}) and cross-linking of K_{67} (and not K_{66}) with K_{325} and K_{339} (Figure 4). Consistent with the nonergodic ECD mechanism and a previous study on grehlin with an unusual fatty acid modification²², the palmitoyl groups were retained on all c- and z-ions that contained C_{322} and C_{323} . The ECD spectrum in Figure 4 shows the presence of the unmodified $c_{17\alpha}$ ion for K_{66} , the absence of the unmodified $c_{18\alpha}$ ion for K_{67} , and the presence of the cross-linked and doubly-palmitoylated $c_{20+M\beta}$ ion containing both K_{66} and K_{67} (where $M\beta = XL+\beta 2$, $\beta 2 = 3598.9348$ Da, and $XL = 138.0681$ Da for DSS). No unmodified $z_{20\alpha}$ and $z_{21\alpha}$ ions or cross-linked $z_{20+M\beta}$ and $z_{21+M\beta}$ ions were observed. Based on this evidence, K_{67} (and not K_{66}) is involved in cross-linking with K_{325} and/or K_{339} . No fragments were generated from cleavage between K_{325} and K_{339} on the $\beta 2$ chain either; however, the presence of doubly-palmitoylated (and not cross-linked) c- and z-ions suggests that both K_{325} and K_{339} were cross-linked. While alternative explanations for these observations exist (i.e., doubly-palmitoylated fragments could possibly be formed from 2 bond cleavages of the corresponding cross-linked peptides), it is highly unlikely that these cross-link assignments are incorrect.

Comparison with Known Structure. The previous discussion shows that cross-links with only 1 of 2 adjacent lysines (K_{67} and not K_{66}) will form with C_{316} , K_{325} , and K_{339} . For example, the cross-linking reaction between K_{67} to C_{316} could begin with the attachment of the maleimide (EMCS) or iodo (SIA) group of the cross-linker to the thiol group of C_{316} followed by attachment of the succinimidyl ester end of the cross-linker to the ϵ -amino group of K_{67} . Studies of cross-link formation as a function of distance, cross-linker flexibility, protein conformation, and reactivity are currently underway. However, based on a 40 ns molecular dynamics trajectory of Rho with explicit lipid bilayer in solvent, the calculated average solvent accessibility of K_{67} is ~ 3 -fold greater than K_{66} ²³. Furthermore, alkylation experiments of Rho in lipid bilayers show that only one lysine in peptide α is reactive with 50X NHS-acetate and only $\sim 75\%$ is fully alkylated (both lysines) even in the presence of 1000X NHS-acetate (data not shown). Finally, K_{66} vs.

K₆₇ distances to C₃₁₆ are 13.1 vs. 9.2 Å and 11.0 vs. 9.9 Å for the X-ray structure¹⁹ and molecular dynamics trajectories²³, respectively. While we cannot know the relative contribution of reactivity vs. distance from this information, and other factors may contribute to the rates of cross-link formation at K66 and K67, the solvent accessibility information and distances from the known structure are consistent with our results that show only K66 participates in cross-linking reactions.

Fragmentation of Cross-Linked Peptides. Different fragmentation mechanisms for bonds involving the cross-linker for α -SIA- β 1 and α -EMCS- β 1 are suggested by their ECD spectra (Figures 3A-B). Bond cleavages observed by ECD are shown in Figure 5 where the frequency of cleavage at sites I-III, deduced from the relative abundance of ions formed from cleavage at each site, is indicated by the size of the lightning bolt (not to scale). Major fragmentation of the cross-links occurs at cleavage sites II for EMCS and II and III for SIA. This corresponds to cleavage on either side of sulfur, suggesting that sulfur could be a hydrogen acceptor that directs fragmentation. Evidence for this is found by the observation that cleavage adjacent to the electron withdrawing succinimide group at site II in EMCS (Figure 5A) is more pronounced than cleavage at site II in SIA (Figure 5B). Minor peaks in the ECD spectra, products of selective accumulation in the collision cell, appear as b- and y-type ions from cleavage at site I. While these observations may prove useful for unambiguous assignment of chemical cross-links, it is clear that the fragmentation mechanisms of cross-linked peptides merit further study.

Conclusions

This work effectively demonstrates unambiguous assignment of intra-molecular chemical cross-links to specific amino acid residue(s), when two or more possible cross-linking sites are adjacent in the cross-linked protein, by FT-MS/MS. To our knowledge this work also presents the first demonstration that ECD can be used to characterize S-linked fatty acid acylation on cysteines. The impact of incorrect (or ambiguous) intra-molecular cross-link assignments on models of the tertiary structure of a cross-linked protein is expected to be significant. Likewise, the impact of incorrect (or ambiguous) inter-molecular cross-link assignments on protein complex topology and protein-protein interactions may also be significant. Therefore, the comprehensive fragmentation of large, multiply-charged, peptides by CID, IRMPD, and particularly ECD, in conjunction with the high resolution and mass accuracy of FT-MS/MS, is expected to be a valuable means of characterizing even difficult proteins such as mammalian membrane proteins with both chemical- and post-translational modifications.

Acknowledgments

The authors are grateful to the DOE Laboratory Directed Research and Development program at Sandia National Laboratories (a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000) for supporting this work.

References

- (1) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc Natl Acad Sci U S A* **2000**, *97*, 5802-5806.
- (2) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J Am Soc Mass Spectrom* **2003**, *14*, 834-850.
- (3) Dihazi, G.; Sinz, A. *Rapid Commun Mass Spectrom* **2003**, *17*, 2005-2014.
- (4) Collins, C. J.; Schilling, B.; Young, M.; Dollinger, G.; Guy, R. K. *Bioorg Med Chem Lett* **2003**, *13*, 4023-4026.
- (5) Pearson, K. M.; Pannell, L. K.; Fales, H. M. *Rapid Commun Mass Spectrom* **2002**, *16*, 149-159.
- (6) Kruppa, G. H.; Schoeniger, J.; Young, M. M. *Rapid Commun Mass Spectrom* **2003**, *17*, 155-162.
- (7) Novak, P.; Young, M. M.; Schoeniger, J. S.; Kruppa, G. H. *Eur J Mass Spectrom (Chichester, Eng)* **2003**, *9*, 623-631.
- (8) Back, J.; de Jong, L.; Muijsers, A.; de Koster, C. *J Mol Biol* **2003**, *331*, 303-313.
- (9) Sinz, A. *J Mass Spectrom* **2003**, *38*, 1225-1237.
- (10) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom Rev* **1998**, *17*, 1-35.
- (11) Hakansson, K.; Cooper, H.; Hudgins, R.; Nilsson, C. *Curr Org Chem* **2003**, *7*, 1503-1525.
- (12) Sinz, A.; Wang, K. *Biochemistry* **2001**, *40*, 7903-7913.
- (13) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal Chem* **1994**, *66*, 2809-2815.
- (14) Zubarev, R.; Kelleher, N.; McLafferty, F. *J.A.C.S.* **1998**, *120*, 3265-3266.
- (15) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal Chem* **2000**, *72*, 563-573.
- (16) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. *J Am Soc Mass Spectrom* **2001**, *12*, 245-249.
- (17) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Sorensen, E. S.; Zubarev, R. A. *Anal Chem* **2003**, *75*, 2355-2361.
- (18) Emmett, M. R. *J Chromatogr A* **2003**, *1013*, 203-213.
- (19) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739-745.
- (20) White, S. http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html **2003**.
- (21) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. *J Mass Spectrom* **2004**, *39*, 322-328.
- (22) Guan, Z. *J Am Soc Mass Spectrom* **2002**, *13*, 1443-1447.
- (23) Crozier, P. S.; Stevens, M. J.; Forrest, L. R.; Woolf, T. B. A.-S. N. L., POB 5800, MS 1411, Albuquerque, NM 87185 USA AD - Sandia Natl Labs, Albuquerque, NM 87185 USA AD - Johns Hopkins Univ, Sch Med, Dept Physiol, Baltimore, MD 21205 USA *J Mol Biol* **2003**, *333*, 493-514.

Figure 2. CID and IRMPD of the sextuply charged cation $[M+EMCS+6H]^{6+}$ at $905.1676\ m/z$ ($\Delta m = 0.6\ ppm$) from the cross-linked peptide α -EMCS- β 1 (5424.9587 Da) yielded abundant b- and y-type ions (Figures 2A and 2B), respectively. However, the product ion spectra were complicated by numerous internal ions and cleavage between K_{67} and K_{68} was never observed.

Fig 2A: CID of α -EMCS- β 1

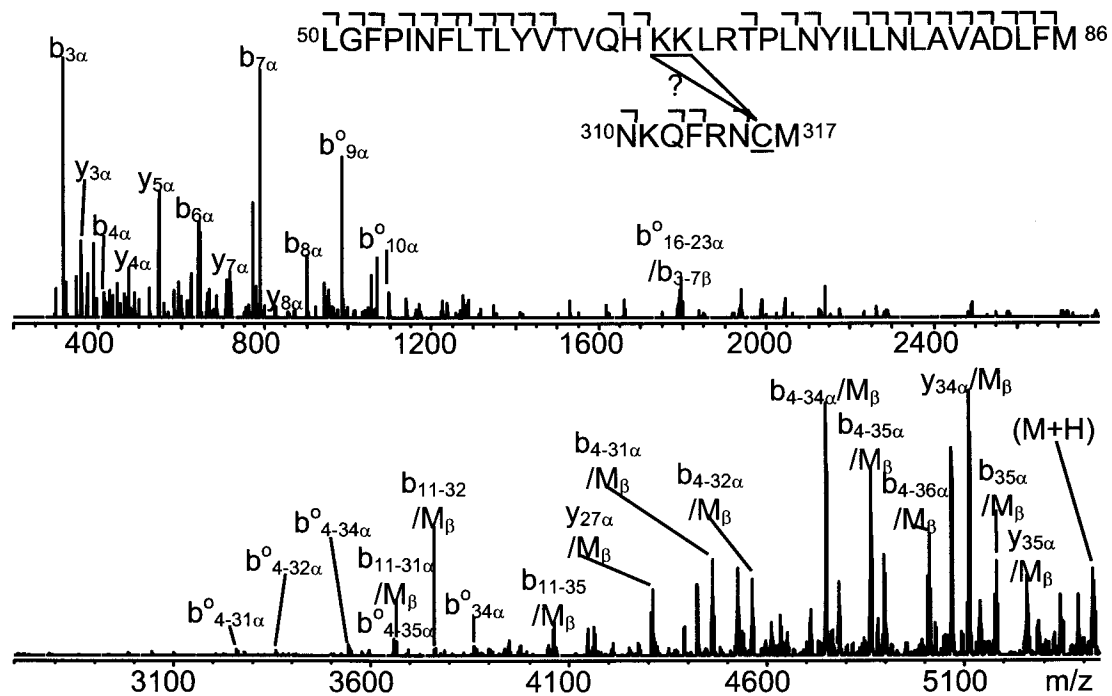


Fig 2B: IRMPD of α -EMCS- β 1

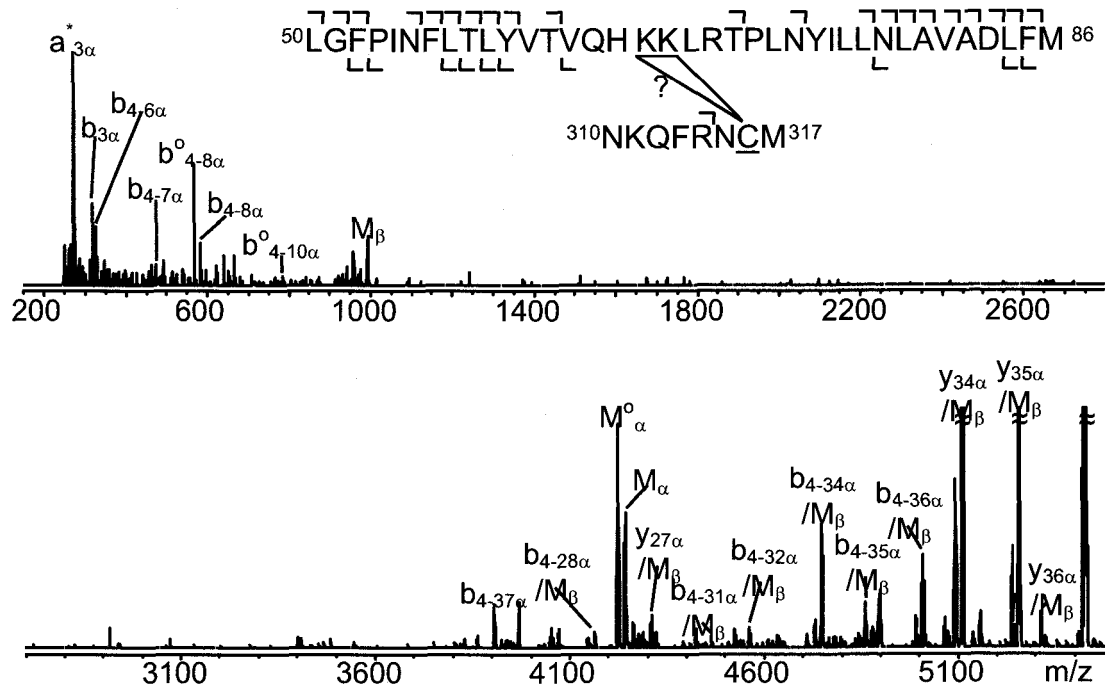


Figure 3. ECD of the $[M+EMCS+6H]6+$ ion yielded abundant c- and z-type ions with few internal fragment ions (Figure 3A). Likewise, ECD of the sextuply charged cation $[M+SIA+6H]6+$ at 879.6533 m/z ($\Delta m = 0.7$ ppm) from the cross-linked peptide α -SIA- β 1 (5271.8797 Da) provided extensive backbone fragmentation (Figure 3B). ECD of the pentuply charged cation $[M+5H]5+$ for the control peptide α at 849.0904 m/z ($\Delta m = 0.7$ ppm, 4240.4185 Da) is shown in Figure 3C. Cleavage of the bond between K66 and K67 was observed in every case.

Fig 3A: ECD of α -EMCS- β 1

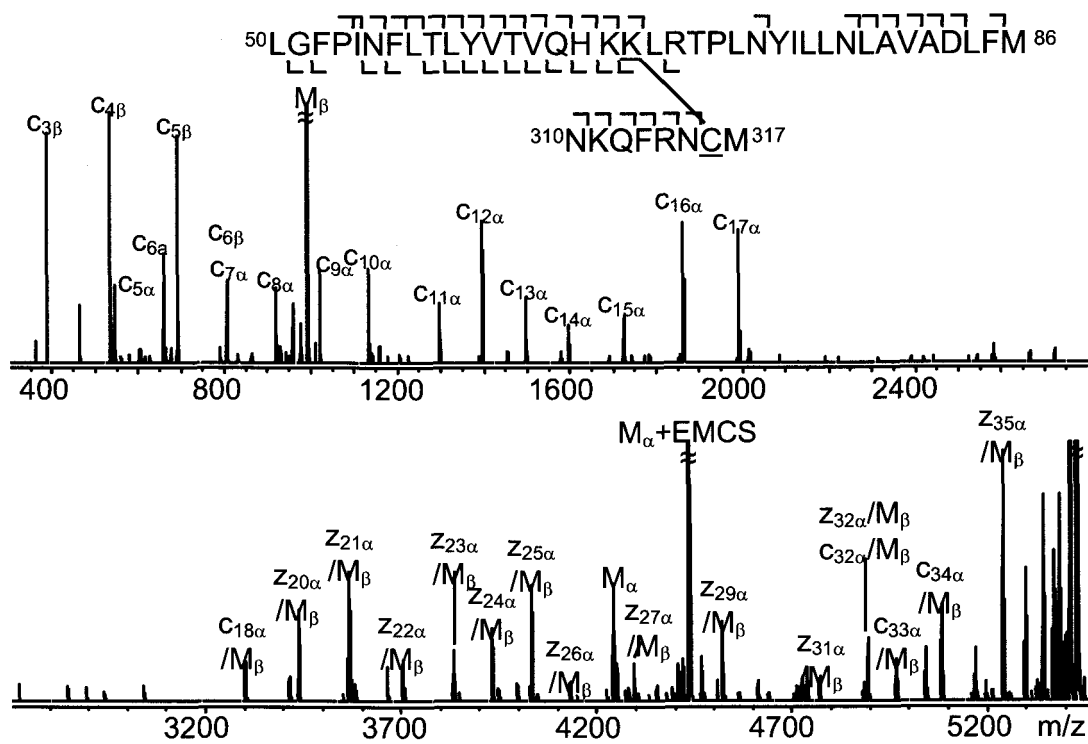


Fig 3B: ECD of α -SIA- β 1

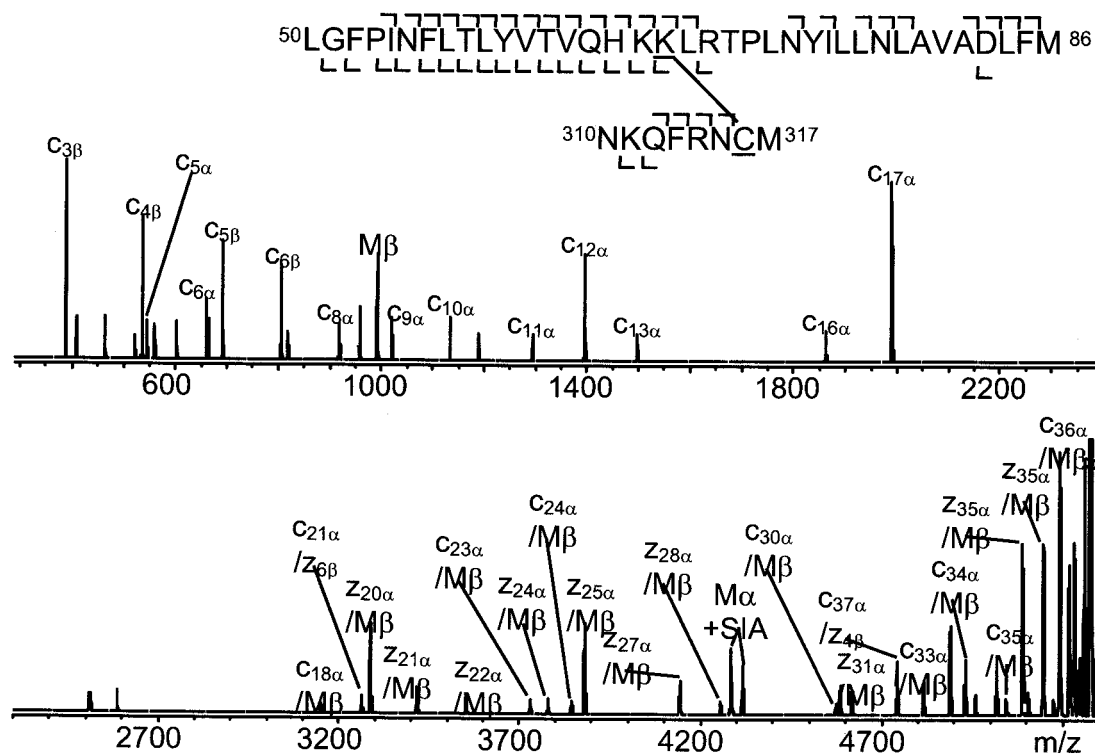


Fig 3C: ECD of control α

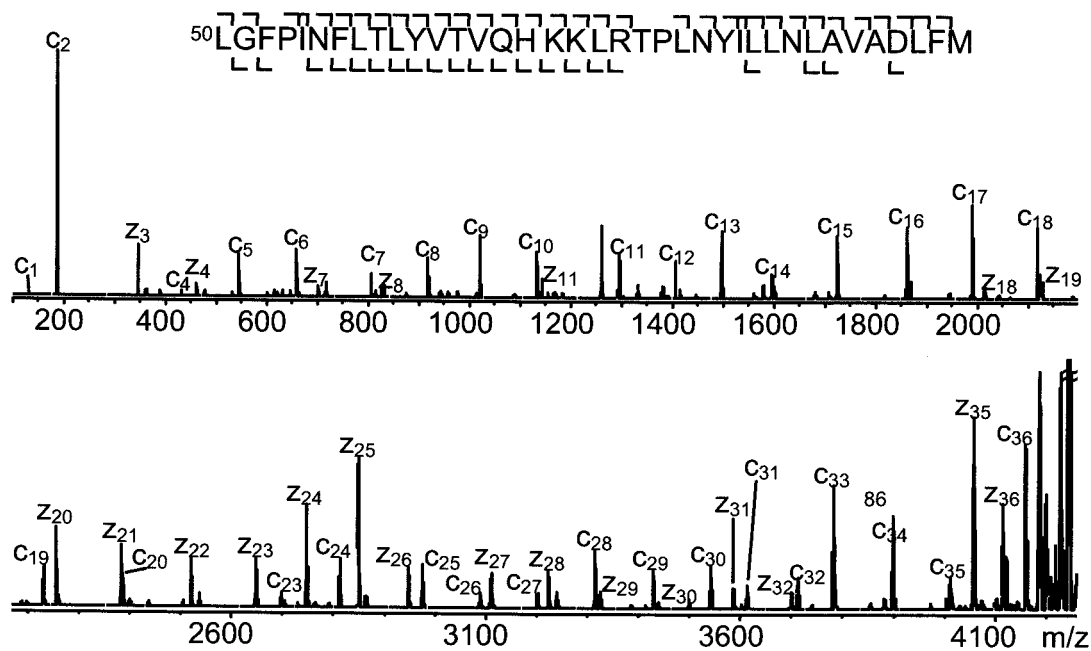
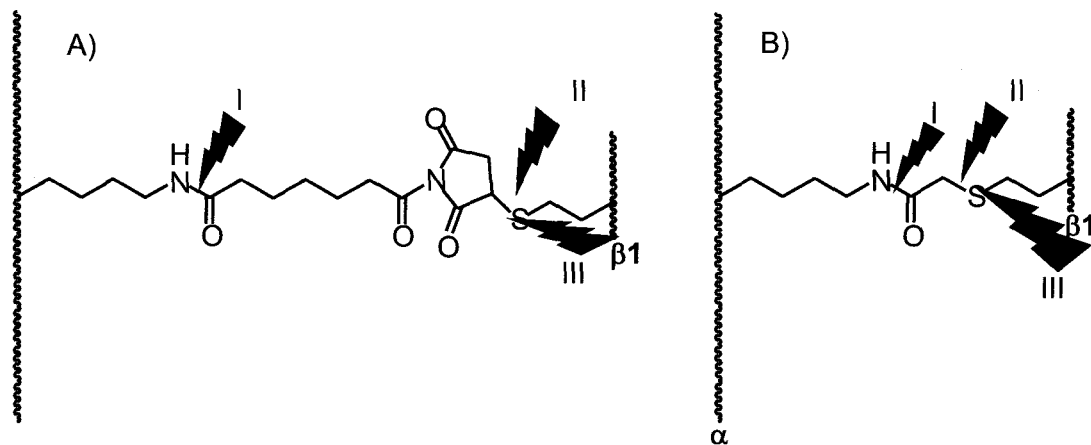


Figure 5. Bond cleavages observed by ECD for (A) α -EMCS- β 1 and (B) α -SIA- β 1. The frequency of cleavage at sites I-III, deduced from the relative abundance of ions formed from cleavage at each site, is indicated by the size of the lightning bolt (not to scale).



Distribution

1	MS9671	M. J. Ayson, 08321
1	MS9292	R. B. Jacobsen, 08321
1	MS9671	P. Lane, 08321
1	MS9292	K. L. Sale, 08321
1	MS9292	M. M. Young, 08321
1	MS9292	J. S. Schoeniger
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	MS0323	D. Chavez, LDRD Office, 1011