

SANDIA REPORT

SAND2004-8079

Unlimited Release

Printed March 2004

The Interfacial Bioscience Grand Challenge

M.J. Ayson, A. Burns, P.S. Crozier, K. Evans, J. Faulon, J. Hong, R.B. Jacobsen, G.H. Kruppa, P. Lane, P. Novak, K.L. Sale, M.J. Stevens, 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.



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>



SAND2004-8079
Unlimited Release

The Interfacial Bioscience Grand Challenge

Marites J. Ayson, Alan Burns, Paul S. Crozier, Kervin Evans, Jean-Loup Faulon, Joohee Hong, Richard B. Jacobsen, Gary H. Kruppa, Pamela Lane, Petr Novak, Kenneth L. Sale, Mark J. Stevens, Nichole L. Wood, Malin M. Young, and Joseph S. Schoeniger

Sandia National Laboratories
Livermore, California

This page intentionally left blank.

Table of Contents

SECTION I: CHEMICAL CROSS-LINKING AND MASS SPECTROMETRY APPLIED TO DETERMINATION OF PROTEIN STRUCTURE AND DYNAMICS	7
Chapter 1: Three-dimensional distance constraints for native bovine rhodopsin determined by chemical cross-linking and high-resolution mass spectrometry (MS3-D).....	9
<i>Introduction</i>	10
<i>Results</i>	11
<i>Acknowledgements</i>	21
<i>References</i>	22
<i>Tables</i>	26
<i>Figures</i>	29
Chapter 2: A Top down Method for the Determination of Residue Specific Solvent Accessibility in Proteins	34
<i>Introduction</i>	35
<i>Experimental</i>	35
<i>Results and discussion</i>	36
<i>Conclusions</i>	37
<i>Acknowledgements</i>	38
<i>References</i>	39
<i>Figures</i>	41
<i>Tables</i>	44
Chapter 3: A Top-down Approach to Protein Structure Studies using Chemical Crosslinking and Fourier Transform Mass Spectrometry.....	46
<i>Introduction</i>	47
<i>Experimental</i>	48
<i>Results</i>	49
<i>Discussion</i>	50
<i>Conclusions</i>	51
<i>Acknowledgements</i>	52
<i>References</i>	53
<i>Tables</i>	55
<i>Figures</i>	57
SECTION II: COMPUTATIONAL MODELING OF MEMBRANE PROTEIN STRUCTURE AND DYNAMICS. 64	
Chapter 4: Exploring the conformational space of membrane protein folds matching distance constraints	66
<i>Introduction</i>	67
<i>Results and discussion</i>	68
<i>Conclusions</i>	72
<i>Methods</i>	73
<i>References</i>	75
<i>Tables</i>	77
<i>Figures</i>	80
Chapter 5: Comparison between isolated rhodopsin helices simulations and intact rhodopsin simulation	86
<i>Introduction</i>	87
<i>Methods</i>	87
<i>Results</i>	88
<i>Conclusion</i>	92
<i>References</i>	93
<i>Tables</i>	95
<i>Figures</i>	96

SECTION III: STUDIES OF TOXIN-MEMBRANE INTERACTIONS USING SINGLE MOLECULE BIOPHYSICAL METHODS.....	106
Chapter 6: Biophysical Studies of Single Nanopores Formed by Protein Toxins Bound to Membranes.....	108
<i>Introduction</i>	110
<i>Materials and Methods</i>	110
<i>Results</i>	114
<i>Summary</i>	118
<i>References</i>	119
Chapter 7: Domain structure in model membrane bilayers investigated by simultaneous atomic force microscopy and fluorescence imaging	120
<i>Introduction</i>	121
<i>Materials and methods</i>	122
<i>Results and discussion</i>	123
<i>Conclusions</i>	125
<i>References</i>	126
<i>Figures</i>	128
<i>Acknowledgments</i>	133
DISTRIBUTION	134

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

This page intentionally left blank.

Chapter 1: Three-dimensional distance constraints for native bovine rhodopsin determined by chemical cross-linking and high-resolution mass spectrometry (MS3-D)

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

Abstract

Membrane proteins make up a diverse and important subset of proteins for which structural information is limited, despite recent progress in X-ray crystallography and NMR spectroscopy techniques. Alternative strategies such as disulfide mapping, spin labeling, and solvent accessibility studies have been used to deduce structural information about these proteins. Recent a method combining the use of mass spectrometry to identify inter-residue cross-links formed by covalent cross-linking reagents, and the use of these cross-links to constrain structural models of proteins, referred to as “MS3-D” or mass spectrometry in three dimensions, has been reported. Subsequent reports indicated that this approach appears to have promise for defining large sets of inter-atomic distance constraints in proteins. We describe general methods that expand the applicability of MS3-D to membrane proteins and membrane protein complexes, and address issues regarding the interpretation of chemical cross-links as quantitative inter-atomic and inter-residue distance constraints. In addition to multiple experimentally measured constraints, we also present a computational framework for relating the observed cross-linked peptides to the structure and dynamics of the model membrane protein bovine rhodopsin. We compared experimental cross-linked distances to side-chain distances determined from a) the energy minimized X-ray crystal structure of bovine rhodopsin, b) a 40 ns molecular dynamics simulation of dark-adapted rhodopsin in an explicit lipid bilayer, and c) Monte Carlo conformational search procedures. Modeling results indicate that the observed cross-linked peptides are consistent with known structure and dynamics of bovine rhodopsin and that the use of cross-linkers with nested lengths can allow the determination of inter-atomic distances in native membrane proteins.

Introduction

Recent advances in genomic and proteomic approaches for the identification of new proteins have led to the need for rapid analysis of 3-D protein structure. Membrane proteins, encoded by ~20% of the genes in almost all organisms, play a central role in intracellular signaling, energy and material transport, cell intoxication and pathogenesis, and cell recognition and motility. However, to date less than 70 integral membrane protein structures have been resolved by X-ray crystallography and/or NMR spectroscopy.¹ Progress remains slow, particularly in the structural analysis of eukaryotic membrane proteins, due to the instability of these proteins in environments lacking phospholipids, their tendency toward aggregation and precipitation, low protein abundance and expression, and sample purity issues which have hindered the application of standard structural determination methods.

Such technical challenges necessitate the use of a variety of techniques to discern the three dimensional structure 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 previously been used to obtain low to moderate resolution distance constraints in membrane proteins²⁻¹². These methods, however, typically require the construction and expression of site-specific histidine or cysteine mutants of the native protein. Young and co-workers recently reported an approach to protein structure analysis, referred to as MS3-D or mass spectrometry in three dimensions, which utilizes chemical cross-linking, mass spectrometry, and computer modeling to produce moderate-resolution structures (Young et al. 2000; Schilling et al. 2003). Compared to methods that use chemical cross-linking and mass spectrometry to identify protein partners and analyze the topology of protein complexes,¹³⁻¹⁵ the goal of MS3-D is to produce a substantial number of intramolecular distance constraints by determining which residues form cross-links when the protein is exposed to a reagent of a given length. In MS3-D, a protein in its native state is reacted with a bifunctional cross-linking reagent with known reactivity, and the internally cross-linked monomeric protein is purified and chemically or enzymatically digested. Mass spectrometric analysis combined with automated prediction of the cross-linked peptide masses allows for the identification of cross-links and the determination of inter-residue cross-linked distances for use in structural modeling. Although this approach has recently been applied to fold recognition, solvent accessibility, and other structural studies of soluble proteins,¹⁶⁻²¹ the application of MS3-D to the study of membrane protein structure has not previously been reported.

We describe here a version of the MS3-D approach for the determination of three-dimensional distance constraints in membrane proteins. The application of MS3-D to the study of membrane proteins required significant optimization of the a) cross-linking conditions, b) proteolysis conditions to ensure complete fragmentation, c) LC-MS methods to obtain full sequence coverage, and d) data analysis software for the identification of cross-linked peptides. We chose the intensively studied G-protein-coupled photoreceptor rhodopsin as a model protein, because large quantities of it can be purified in its native rod outer segment (ROS) membrane and it has a known structure against which we can validate our results.²² Using dark-adapted bovine rhodopsin in ROS, we show that a significant number of cross-linked peptides can be reliably identified that carry information about the inter-atomic distances within the rhodopsin monomer.

We also considered the best way to relate the formation (or failure to form) 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 soluble²⁰ or a membrane protein.²³ 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.¹⁶ 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.²⁴ 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.²² 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 simulation²⁵ were also used to identify side-chain and/or backbone dynamics that could lead to the observed cross-links. Using the cross-linking data, we were able to determine whether the distance constraints implied by the observed cross-links were consistent with the known structure and dynamics of rhodopsin, and, in select cases, whether nested cross-linking experiments result in valid high-resolution inter-atomic DCAs.

Results

Sample Preparation and Analysis by LC-MS

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, several groups have recently made progress in the preparation of integral membrane proteins, including rhodopsin, for mass spectrometry²⁶⁻³⁰. In our work, selected methods were optimized in order to maximize the recovery of cross-linked rhodopsin for subsequent LC-MS and tandem MS analysis.

Formation and identification of inter-residue cross-links was carried out following the MS3-D approach described previously²⁰ with adaptation for membrane protein samples. 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 heterobifunctional C-K reagents used (see Table 1.1) have a primary-amine reactive N-hydroxysuccinimide (NHS) ester group on one end and a sulfhydryl-reactive group (iodoalkyl or maleimide) on the other, with sequentially increasing linker arm lengths (refer to the Materials section for the definition of cross-linker abbreviations). The homobifunctional K-K reagents used have NHS ester groups at both ends, and also have sequentially increasing linker lengths (Table 1.2). During this step, the cross-linker can react to form a cross-link between amino acid residues within a single protein, or it can react with residues in different proteins to form a covalent multimer. Without separation of protein monomers, proteolytic cleavage of the reaction mixture would result in cross-linked peptides containing both intra- and intermolecular cross-links. Since intra-molecular distance constraints were desired, the molar excess of cross-linker to protein was optimized (typically 50:1) in order to achieve a high yield of monomeric protein following the cross-linking reaction (greater than two-thirds of the crude mixture as visualized by Coomassie stained SDS-PAGE, Figure 1.1a).

To rigorously isolate monomeric rhodopsin, dimers and higher order multimers were removed following the cross-linking reaction by tube gel electrophoresis and fractionation. The detergent, ammonium lauryl sulfate (ALS) was used in the electrophoretic separation instead of sodium dodecyl sulfate in order to minimize sodiation of rhodopsin peptides during mass spectral analysis. Fractions containing monomeric rhodopsin were pooled and concentrated (Figure 1.1b), and the residual detergent and lipids were removed from the protein by chloroform/methanol/water (CMW) phase separation.³¹ The pellet was then repeatedly washed with acetonitrile, 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 1.2. 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 1.2a). 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^{26,29}. 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 1.2b 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 1.2d, 50-86x310-317) are typically of much lower intensity than the SICs of precursor peptides (e.g. Figure 1.2c, 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^{32,33}. 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 a putative cross-linked species only in the LC-MS spectrum of cross-linked sample and not in 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 an MS spectrum in the LC-MS run which contained relatively intense peaks of readily identifiable unmodified rhodopsin peptides known to be present in 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 1.1 and 1.2). 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.²⁰ For each monoisotopic mass in the experimental results, ASAP searched the library for masses within \pm 5-10 ppm and reported all matching assignments.

Identified cross-links. Peaks assigned by ASAP to cross-linked peptides 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 Results, Identification of Cross-Linked Products, 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 1.1 and 1.2). 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 1.3) 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 1.1). Cross-linking results with the three homobifunctional NHS ester cross-linking reagents, DST, DSG, and DSS are also reported (Table 1.2). 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 as calibration standards. This high mass accuracy is sufficiently high to provide unambiguous assignment of each cross-linked species without the use of isotopically labeled reagents or MS/MS. The identification of peptides based on the assignment of more than one of its charge states observed by ESI-MS, along with the unique mass shift observed with each cross-linker tested, provided a degree of redundancy to confirm the identity of the cross-linked products. An example of such a mass shift is shown in the spectra of the 50-86x310-317 peptides cross-linked with SIA, GMBS, EMCS, and LC-SMCC (Figure 1.4). In each mass spectrum, 50-86x310-317 displays a unique m/z at each charge state observed, corresponding to the difference in mass of each cross-linker. In this way, multiple cross-linkers of varying spacer arm length can be used to provide further confirmation of the observed cross-linking pattern.

MS/MS. The identity of certain cross-linked peptides was also confirmed using MS/MS fragmentation. Various cross-linked products were purified by semi-preparative HPLC and fractions containing specific cross-linked peptides were analyzed by direct infusion/Q-CID-FTMS (see methods). Results for the EMCS cross-linked species of 50-86x310-317 are shown in Figure 1.5. Assignment of the fragmentation spectra was performed by the software program MS2LINKS (see Methods, Data Analysis below), developed in our laboratory, which compares fragment ions to a theoretical library of all cross-linking possibilities based on common peptide fragmentation pathways. Using a 3 ppm error threshold, a series of b and y ions were successfully assigned for each peptide present, 50-86 and 310-317. Moreover, a series of b and y ions were also found corresponding to the cross-linked peptide pair, unambiguously identifying the mass addition corresponding to the reaction of EMCS with peptides 50-86 and 310-317.

Since two lysine residues are present in both peptides 50-86 (K66/K67) and 318-348 (K325/K339), LC-MS alone is not sufficient to identify the location of the cross-linked residues within these peptides. Consequently, cross-linked products containing these peptides result in several ambiguous cross-links (Tables 1.1 and 1.2). Unfortunately, due to the size of peptides 50-86 and 318-348 (> 3 KDa) and the location of the two lysine residues situated in the internal region of the peptide fragment, fragmentation was not observed between these lysine residues using CID, or Infra-red Multiphoton Dissociation techniques. Due to the poor solubility of the peptides, tryptic digest has also proven problematic. Preliminary results with Q-FTICR MS/MS fragmentation using electron capture dissociation (ECD) of the 50-86x310-317 EMCS cross-linked peptide (data not shown) indicate the presence of a K67xC316 cross-link. Further work is underway on peptides containing ambiguous cross-links to determine the utility of this advanced MS technique for resolving the exact identities of the residues involved in cross-link formation across the two series of nested-length cross-linkers. .

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,³⁴ 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 1.1 and 1.2 report these distances, which were determined using a 3-D structure of the cross-linker constructed in Chem3D³⁵ and optimized using MM3³⁶⁻³⁸ (refer to Methods section). Calculated maximum lengths are in good agreement with the upper range of N-N distances reported by Houk and co-workers³⁴ 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 1.1 and 1.3). 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), 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 structure²². 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 bilayer²⁵ 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 1.3 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 expand 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 chromatography²⁰) and sample recovery (compared to slab PAGE and electroelution). Complete reduction and alkylation of even the most buried cysteines was routinely 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, as opposed to CNBr digestion in 70% TFA, which provided nearly complete cleavage and sequence coverage, along with excellent reproducibility. We found in-gel digests, while effective for proteomics applications, were of limited use in our study, since larger hydrophobic peptides elute poorly from the gel, and reproducibly complete sequence coverage is important when addressing structural questions and eliminating doubt about the identity of cross-linked peptides.

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^{16,19-21}. 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.^{16,24} We believe that this approach is superior to using mixed isotopically-labeled reagents and pattern recognition for data analysis and cross-linked peptide identification^{39,40} 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 ¹²C or single ¹³C 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 the absence of inter-peptide cross-links. 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.⁴¹ 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:

- Are the observed cross-linked peptides compatible with the crystal structure backbone if side-chain rearrangements are allowed?
- Are the observed cross-links compatible with realistic side-chain and backbone dynamics?
- 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 we use 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 1.3). 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 disulphide 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 disulphide bond formation or failure of bond formation during or after folding. In MS-3D, the cross-linking reagent attaches first at 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 chance 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 computer²⁵. 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 are 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 1.3). 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 points out that cross-linking is 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 maleimide 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” (Blaustein 2000) 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.²² Khorana and co-workers^{42,43} observed the formation of cysteine-cysteine disulfide bridges between residues on the cytoplasmic face of rhodopsin that are greater than 12 Å apart ($C\alpha-C\alpha$) in the X-ray crystal structure (see review by Meng et al.⁴⁴). 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^{20,45}. 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.^{46,47} 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.¹⁵ 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.²³ 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.

Experimental

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).

Methods

ROS membrane purification. Rod outer segments (ROS) from bovine retinas were prepared under dim red light (> 650 nm) as previously described.⁴⁸ Briefly, ROS fragments were isolated using the sucrose flotation method.⁴⁹ and purified via 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 (16,000xg, 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.⁵⁰ Samples were incubated at room temperature for 30 min. to complete solubilization and the concentration of rhodopsin was determined based on the absorbance difference at 500 nm before and after illumination ($\epsilon = 42,700 \text{ M}^{-1} \text{ cm}^{-1}$).⁵¹

Cross-linking and labeling experiments. All reactions were performed under dim red light unless otherwise indicated. Lysine-lysine cross-linking was achieved using DST, DSG, DSS, and cysteine-lysine cross-links were obtained following reaction with SIA, GMBS, EMCS, and LC-SMCC (refer to Figures 1.6 - 1.8 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 5-10 mM Tris-HCl, pH 7.5, for 15 min. at room temperature. Cysteine-lysine cross-linking was achieved using the same method except with the following modification. Samples were first reacted in 50 mM HEPES buffer, 100 mM NaCl, to facilitate the reaction of lysine residues at pH 7.5. Following incubation at 37°C (15 min.), samples were centrifuged to

remove the excess cross-linker, resuspended in 50 mM pyridine buffer (pH 6-6.8), and incubated at 37°C (30 min.) to facilitate reaction of the cysteine residues. Samples were subsequently quenched with 5-10 mM Tris-HCl, pH 7.5, and 0.01% β -mercaptoethanol for 15 min. at room temperature. Labeling studies were performed using NHS-acetate. Native rhodopsin (10 μ M) in ROS was incubated with 50 to 1000 fold molar excess of NHS-acetate for 30 min. at 37°C, in 50 mM HEPES buffer, 100 mM NaCl, pH 7.4, and the reaction was quenched with 5-10 mM Tris-HCl, pH 7.5.

Following the cross-linking and labeling reactions, disulfide bonds were reduced with 50 mM tris(2-carboxyethyl)phosphine^{32,33} in 50 mM HEPES buffer, 100 mM NaCl, pH 7.5, for 30 min. at 37°C. Samples were alkylated with 150 mM 4-vinylpyridine⁵² at room temperature for 30 min, and washed several times using the same buffer to remove the 4-vinylpyridine. Both reduction and alkylation were performed prior to removal of the protein from the membrane.

Monomer purification and cleavage with CNBr. Monomeric rhodopsin were separated from intermolecular cross-linked species and contaminants by preparative ALS-PAGE using a miniprep cell (Bio-Rad) containing a column gel (5 cm length, 1 cm diameter, 11 % acrylamide resolving gel; 2 cm length, 4 % acrylamide stacking gel, 1X Tris-Glycine buffer with 0.1% ALS) equipped with a peristaltic pump and a fraction collector. Eluted protein fractions were collected and analyzed by SDS-PAGE and Coomassie staining, and pooled fractions were concentrated using centrifugal filters (5-10K MWCO, Millipore). Purified rhodopsin was subsequently delipidated by chloroform/methanol/H₂O extraction³¹, and the resulting pellets were washed with acetone and allowed to air dry. A stock solution was prepared of high purity (>99%) CNBr in acetonitrile (4M). Rhodopsin samples were first dissolved in 100% TFA, diluted to 70% TFA with H₂O, and then chemically digested with 100-200 mM CNBr. Control experiments indicated that equivalent results were obtained following a 6 hour and 16 hour CNBr digest. Generally, sample vials were flushed with nitrogen and vortexed overnight at room temperature (in the dark). Following CNBr digestion, samples were dried under vacuum (Savant speed-vac), washed with acetonitrile, and dried again to remove the residual TFA. Dried samples were stored at -80°C until further analysis. Failure to scrupulously remove methanol from the CMW pellet results was found to result in substantial methylation of the resulting peptides as observed by ESI-MS.

LC-MS. CNBr digested rhodopsin fragments were separated by reverse-phase HPLC.²⁶ Peptide fragments were dissolved in formic acid, acetonitrile and water (10:10:80) to a final concentration of 1 mg/ml. The sample (8 μ l) was loaded onto a polystyrene/divinylbenzene 0.2x150mm column (PLRP/S, Michrom Bioresources, Auburn, CA) and chromatographic separation was performed using a gradient elution with 5% acetic acid, 2.5% acetonitrile and 2.5% isopropanol as solvent A and 4% acetic acid, 40% acetonitrile and 50% isopropanol as solvent B. Starting at 0% B for the first minute, the gradient increased to 15% B after 5 minutes, then to 60% B after 35 minutes and finally it increased to 100% B after 45 minutes. The flow rate was 5 μ l/min. All chromatographic separation were performed at 25°C using an Agilent 1100 capillary HPLC system (Agilent Technologies, Palo Alto, CA) and the column was connected directly to the mass spectrometer with no split.

Mass spectrometry analysis was performed using an APEX II FTMS instrument 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 preparative column (5 μ particle size, 300 Å 2.0x150mm), thermostated at 30°C. The LC elution conditions were as follows (% v/v): Buffer A = 2.5: 2.5: 5: 90 isopropanol: acetonitrile: acetic acid: water; buffer B = 4: 6: 40: 50 acetic acid: water: acetonitrile: isopropanol. 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. The samples were prepared by re-suspending the dried CNBr digested pellet with 10% trifluoroacetic

acid or formic acid in buffer A. Final concentration of the sample was ca. 1 mg/mL and 200-300 μ g was loaded for each separation. The eluent fractions were collected using a Bio-Rad 2110 fraction collector and their mass spectra were analyzed by direction infusion on an ESI-QTOF mass spectrometer (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 using an ESI-FTICR mass spectrometer (Apex II, Bruker Daltonics). 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.

Data analysis. Reduction of the MS spectra is automated by a macro developed in our laboratory and implemented in the Xmass software package (Bruker Daltonics) as described previously.²¹ The data reduction and associated software written in ‘C’ reduces the peak list for each LC scan to a set of unique monoisotopic masses, assuming that the individual isotopic peaks are resolved for all charge states. The final output table produced by the program contains the monoisotopic protonated molecular weight for each fragment, the charge states that were observed for that fragment, the relative abundance of the most abundant peak observed for that fragment, and finally whether the monoisotopic peak was observed in the spectrum. The list of monoisotopic masses were matched to a theoretical library of unmodified, labeled, and/or cross-linked peptides generated by ASAP (Automated Spectrum Assignment Program) described previously²⁰. ASAP accepts user input on the protein sequence, cross-linker mass modification and amino acid specificity, a mass error limit, the proteolytic enzyme or chemical, the maximum charge state, and a peak abundance threshold. ASAP constructs a proteolytic peptide virtual library, including all possible cross-linked peptides. For each monoisotopic mass in the experimental results, ASAP searches the library for masses within the error limit, and reports all matching assignments for each mass peak.

In some cases, the identity of cross-linked peptides was confirmed using MS/MS fragmentation, which was analyzed using MS2Links, an updated software version of MS2Assign described previously.⁵³ MS2Assign and MS2Links were developed for the assignment of tandem mass spectra of unmodified, labeled, and/or cross-linked peptides. The program generates a theoretical library containing all cross-linking possibilities based on common peptide fragmentation that result in a,b,c-type, x,y,z-type, and internal and immonium ions with associated common losses of H₂O, NH₃, CO, and CO₂. Given an input m/z fragment list, the program will assign the monoisotopic masses that are within a defined ppm error to species in the theoretical library. 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.³⁵ 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 MM3³⁶⁻³⁸ 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.^{54,55} Both the CHARMM⁵⁴ and NAMD⁵⁶ 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.²⁵ 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 SURFACE command in CHARMM computes the SASA using the Lee and Richards methods.⁵⁷ A probe radius of 1.4 \AA was used for all SASA calculations.

Acknowledgements

We would like to thank Helgi Adalsteinsson for helpful discussions and assistance in calculating the maximum cross-linker lengths. 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. (2003). *Membranes Proteins of Known 3D Structure* at http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html.
2. Nakanishi, K., Zhang, H., Lerro, K. A., Takekuma, S., Yamamoto, T., Lien, T. H., Sastry, L., Baek, D. J., Moquin-Pathey, C., Boehm, M. F. & et al. (1995). Photoaffinity labeling of rhodopsin and bacteriorhodopsin. *Biophys Chem* 56, 13-22.
3. Nakayama, T. A. & 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, 15762-9.
4. Yu, H., Kono, M., McKee, T. & 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., Khorana, H. G. (1990). Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *J Biol Chem* 265, 17520-4.
6. Resek, J. F., Farahbakhsh, Z. T., Hubbell, W. L. & Khorana, H. G. (1993). Formation of the meta II photointermediate is accompanied by conformational changes in the cytoplasmic surface of rhodopsin. *Biochemistry* 32, 12025-32.
7. Sheikh, S. P., Zvyaga, T. A., Lichtarge, O., Sakmar, T. P. & Bourne, H. R. (1996). Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature* 383, 347-50.
8. Wu, J. & 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, 14498-502.
9. Altenbach, C., Cai, K., Klein-Seetharaman, J., Khorana, H. & Hubbell, W. (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.
10. Altenbach, C., Klein-Seetharaman, J., Cai, K., Khorana, H. & Hubbell, W. (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.
11. Farrens, D., Altenbach, C., Yang, K., Hubbell, W. & Khorana, H. (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. & 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, 243-90.
13. Rappsilber, J., Siniosoglou, S., Hurt, E. C. & Mann, M. (2000). A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* 72, 267-75.
14. Bennett, K. L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P. & Roepstorff, P. (2000). Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein Sci* 9, 1503-18.

15. 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. & Muijsers, A. O. (2002). A structure for the yeast prohibitin complex: Structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Sci* 11, 2471-8.
16. Novak, P., Young, M. M., Schoeniger, J., Kruppa, G. H. (2003). A top down approach to protein structure studies using chemical crosslinking and fourier transform mass spectrometry. *European Journal of Mass Spectrometry*. In Press.
17. Taverner, T., Hall, N. E., O'Hair, R. A. & Simpson, R. J. (2002). Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem* 277, 46487-92.
18. Dihazi, G. H. & Sinz, A. (2003). Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 17, 2005-14.
19. Novak, P., Kruppa, G. H., Young, M. M., Schoeniger, J. A top down method for the determination of residue specific accessibility in proteins. *Journal of Mass Spectrometry*. Accepted.
20. Young, M., Tang, N., Hempel, J., Oshiro, C., Taylor, E., Kuntz, I., Gibson, B. & 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.
21. Kruppa, G. H., Schoeniger, J. & 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, 155-62.
22. 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. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739-45.
23. Faulon, J. L., Sale, K. & Young, M. (2003). Exploring the conformational space of membrane protein folds matching distance constraints. *Protein Sci* 12, 1750-61.
24. Blaustein, R. O., Cole, P. A., Williams, C. & Miller, C. (2000). Tethered blockers as molecular 'tape measures' for a voltage-gated K⁺ channel. *Nat Struct Biol* 7, 309-11.
25. Crozier, P. S., Stevens, M. J., Forrest, L. R. & 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.
26. Ball, L. E., Oatis, J. E., Jr., Dharmasiri, K., Busman, M., Wang, J., Cowden, L. B., Galijatovic, A., Chen, N., Crouch, R. K. & Knapp, D. R. (1998). Mass spectrometric analysis of integral membrane proteins: application to complete mapping of bacteriorhodopsins and rhodopsin. *Protein Sci* 7, 758-64.
27. Kraft, P., Mills, J. & 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.
28. Ablonczy, Z., Kono, M., Crouch, R. & Knapp, D. (2001). Mass spectrometric analysis of integral membrane proteins at the subnanomolar level: Application to recombinant photopigments. *ANALYTICAL CHEMISTRY* 73, 4774-4779.
29. Barnidge, D. R., Dratz, E. A., Sunner, J. & 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, 816-24.

30. Ablonczy, Z., Goletz, P., Knapp, D. & Crouch, R. (2002). Mass spectrometric analysis of porcine rhodopsin. *Photochemistry and Photobiology* 75, 316-321.
31. Wessel, D. & 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, 141-3.
32. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & 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.
33. 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.
34. Green, N. S., Reisler, E. & Houk, K. N. (2001). Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. *Protein Sci* 10, 1293-304.
35. CambridgeSoft Corporation, C. D., Cambridge, MA, 02140 USA. Chem3D.
36. Allinger, N. Y., YH; and Lii, JH. (1989). Molecular Mechanics- The MM3 Force-Field for Hydrocarbons. 1. *J Am Chem. Soc.* 111, 8551-8566.
37. 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.
38. 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.
39. Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S. & Steinmetz, M. O. (2001). Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem* 73, 1927-34.
40. Collins, C. J., Schilling, B., Young, M., Dollinger, G. & Guy, R. K. (2003). Isotopically labeled crosslinking reagents: resolution of mass degeneracy in the identification of crosslinked peptides. *Bioorg Med Chem Lett* 13, 4023-6.
41. Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A. & Palczewski, K. (2003). Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* 421, 127-8.
42. Cai, K., Klein-Seetharaman, J., Hwa, J., Hubbell, W. & Khorana, H. (1999). Structure and function in rhodopsin: Effects of disulfide cross-links in the cytoplasmic face of rhodopsin on transducin activation and phosphorylation by rhodopsin kinase. *BIOCHEMISTRY* 38, 12893-12898.
43. Cai, K., Langen, R., Hubbell, W. & Khorana, H. (1997). Structure and function in rhodopsin: Topology of the C-terminal polypeptide chain in relation to the cytoplasmic loops. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 94, 14267-14272.
44. Meng, E. C. B., H. R. (2001). Receptor activation: what does the rhodopsin structure tell us? *Trends Pharmacol Sci* 22, 587-93.
45. Reva, B., Finkelstein, A. & Topiol, S. (2002). Threading with chemostructural restrictions method for predicting fold and functionally significant residues: application to dipeptidylpeptidase IV (DPP-IV). *Proteins* 47, 180-93.

46. Albrecht, M., Hanisch, D., Zimmer, R. & Lengauer, T. (2002). Improving fold recognition of protein threading by experimental distance constraints. *In Silico Biol* 2, 325-37.
47. Sommer, I., Zien, A., von Ohsen, N., Zimmer, R. & Lengauer, T. (2002). Confidence measures for protein fold recognition. *Bioinformatics* 18, 802-12.
48. Okada, T., Matsuda, T., Kandori, H., Fukada, Y., Yoshizawa, T. & 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, 4940-6.
49. Papermaster, D. S. (1982). Preparation of retinal rod outer segments. *Methods Enzymol* 81, 48-52.
50. Okada, T., Takeda, K. & 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, 495-9.
51. Hong, K. & Hubbell, W. L. (1973). Lipid requirements for Rhodopsin regenerability. *Biochemistry* 12, 4517-23.
52. Friedman, M. (2001). Application of the S-pyridylethylation reaction to the elucidation of the structures and functions of proteins. *J Protein Chem* 20, 431-53.
53. Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. (2003). MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom* 14, 834-50.
54. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4, 187-217.
55. MacKerell, A. D. J., Brooks, B. R., Brooks, C. L. I., Nilsson, L., Roux, B., Won, Y. & 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.), Vol. 1, pp. 271-277. John Wiley & Sons, Chichester.
56. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. & Schulten, K. (1999). NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* 151, 283-312.
57. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55, 379-400.

Tables

Table 1.1. Observed M + H⁺ (Da) and experimental error (ppm) of cysteine-lysine cross-linked rhodopsin peptides. Protonated masses correspond to the monoisotopic peak (or the ¹³C peak where indicated).

Cross-links	Cross-linked peptides	Cysteine-Lysine Cross-linkers			
		SIA 4.0 Å ¹	GMBS 9.7 Å	EMCS 11.1 Å	LC-SMCC 15.0 Å
C316xK66/67	310-317, 50-86	5272.9097 3.7 ppm	5397.9458 7.2 ppm	5425.9624 0.8 ppm	5565.0869 4 ppm
C316xK325/339	310-317, 318-348 + 2 Palm ²	4632.4387 ppm	4756.4448 1.1 ppm	4784.46683 ppm	4923.5869 1.2 ppm
C140xK141	87-143 internal + 1 PE	n.d.	6414.2285 (¹³ C) 4.3 ppm	6442.227 (¹³ C) 9.4 ppm	6581.4228 (¹³ C) 5.4 ppm
C316xK311	310-317 internal	n.d.	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 Palm = palmitoyl groups attached to C322 and C323.

3 PE = pyridyl ethylation of a cysteine residue.

Table 1.2. Observed $M + H^+$ (Da) and experimental error (ppm) of lysine-lysine cross-linked rhodopsin peptides. Protonated masses correspond to the monoisotopic peak (or the ^{13}C peak where indicated).

Cross-links	Cross-linked peptides	Lysine-Lysine Cross-linkers		
		DST 6.3 Å ¹	DSG 7.4 Å	DSS 11.3 Å
K66/67xK325/339	50-86, 318-348 + 2 Palm ²	7955.3447 (^{13}C) 1.6 ppm	7937.4268 (^{13}C) 5.5 ppm	7979.4067 (^{13}C) 3 ppm
K311xK66/67	50-86, 310- 317+1PE3	5451.9126 5.8 ppm	5434.9717 (^{13}C) 0.3 ppm	5475.9912 4.7 ppm
K311xK325/339	310-317, 318- 348 + 2 Palm + 1 PE	4811.4688 (^{13}C) 1 ppm	4793.5093 (^{13}C) 4.1 ppm	4835.541 (^{13}C) 0.9 ppm
K66xK67	50-86 internal	4355.3784 9.6 ppm	4338.4404 (^{13}C) 2.1 ppm	4379.4756 4 ppm
K325xK339	318-348 internal	n.d.	3696.9828 (^{13}C) 4.6 ppm	3737.9888 5.6 ppm

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

2 Palm = palmitoyl groups attached to C322 and C323.

3 PE = pyridyl ethylation of a cysteine residue.

Table 1.3. Structural analysis using cross-linking distance constraints`

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

1 Length of the most extended structure of the shortest cross-linker for which a cross-link was detected

2 DCA refers to the distance of closest approach of the two cross-linked atoms determined using the following methods: X-ray refers to energy (CHARMM22) minimized X-ray structure (1F88) in which the 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

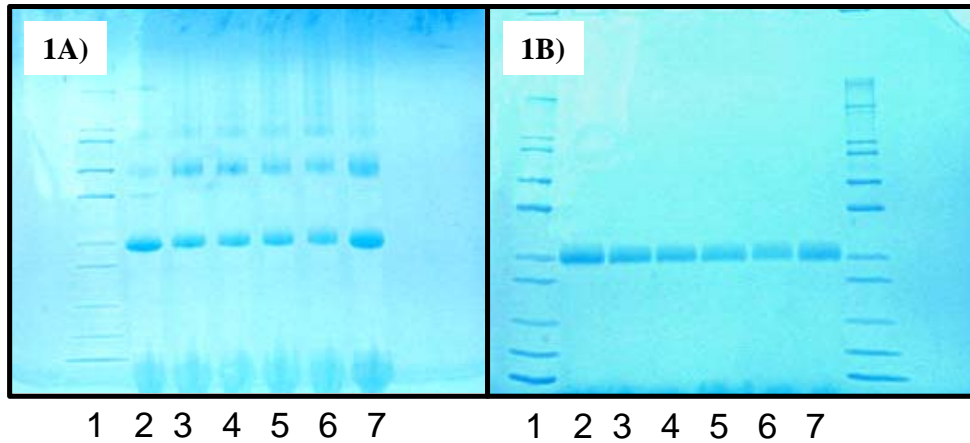


Figure 1.1. Cross-linking and purification of monomeric rhodopsin as visualized by Coomassie stained 4-20% Tris-Glycine SDS-PAGE. Gel (a) Lane 1: Mark 12™ Protein ladder (Invitrogen), 2: Dark adapted rhodopsin in ROS following reduction and alkylation (control), 3-7: Dark adapted rhodopsin in ROS, cross-linked, reduced and alkylated; (b) The same protein samples following tube gel electrophoresis, fractionation, and concentration to purify the monomeric rhodopsin.

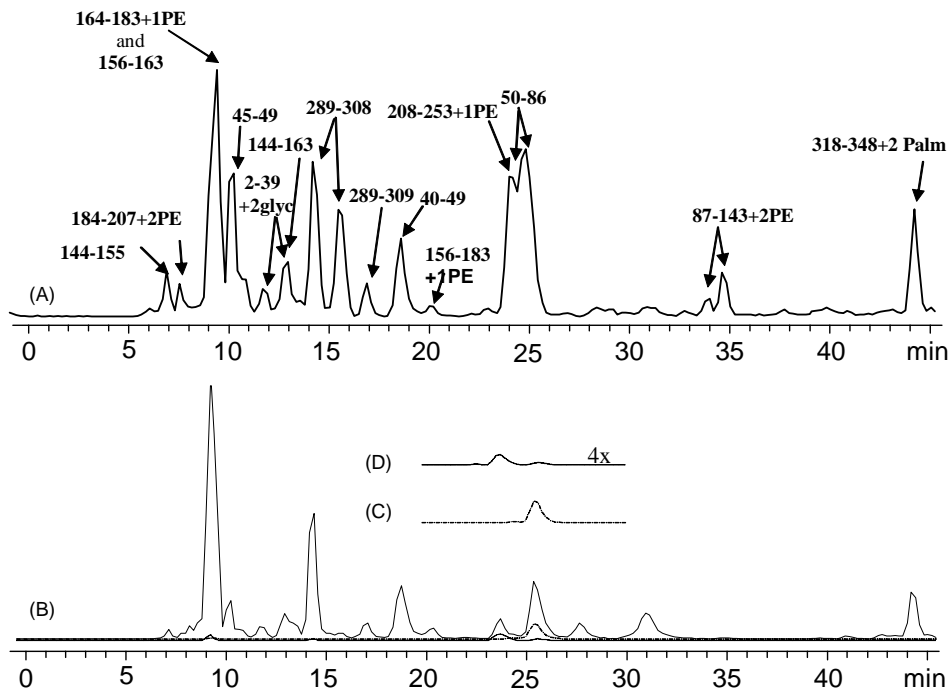


Figure 1.2. Total ion chromatograms (TIC) of the reverse phase HPLC separation on a PRLP-S column and ESI-FTMS of CNBr digested (a) rhodopsin and (b) EMCS cross-linked rhodopsin. Selected ion chromatogram of (c) the unmodified 50-86 peptide and (d) the 50-86x310-317 peptide cross-linked with EMCS. Figure 1.2d is scaled 4x.

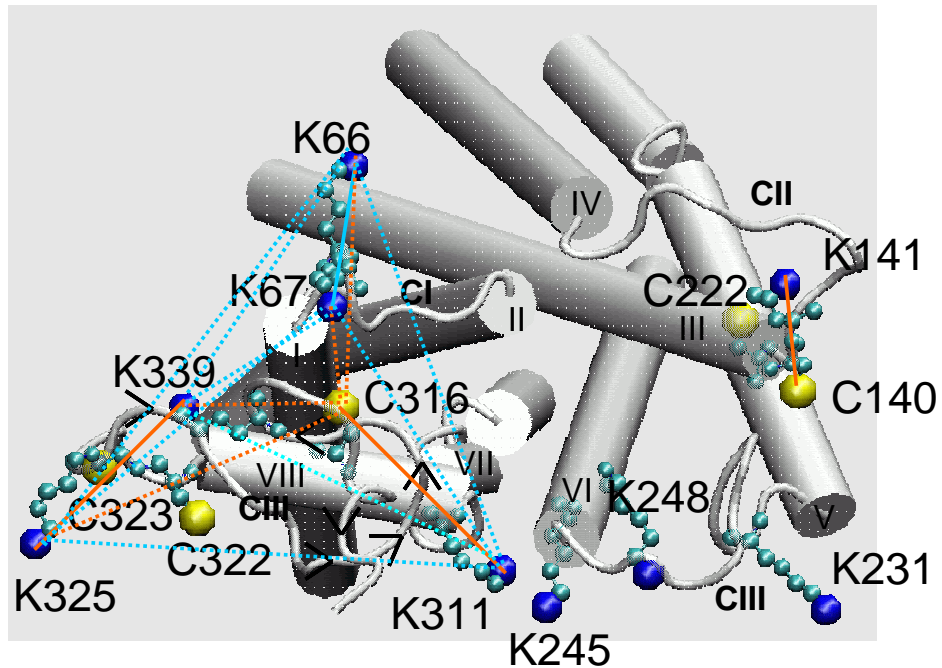


Figure 1.3. Cytoplasmic face of rhodopsin displaying chemically reactive Cys and Lys residues. Solid lines indicate un-ambiguous cross-links, including K141x C140, C316xK311, K66x67, and K325xK339. Dashed lines represent potential cross-links, including C316xK66/67, C316xK325/339, K311xK66/67, and K311xK325/339.

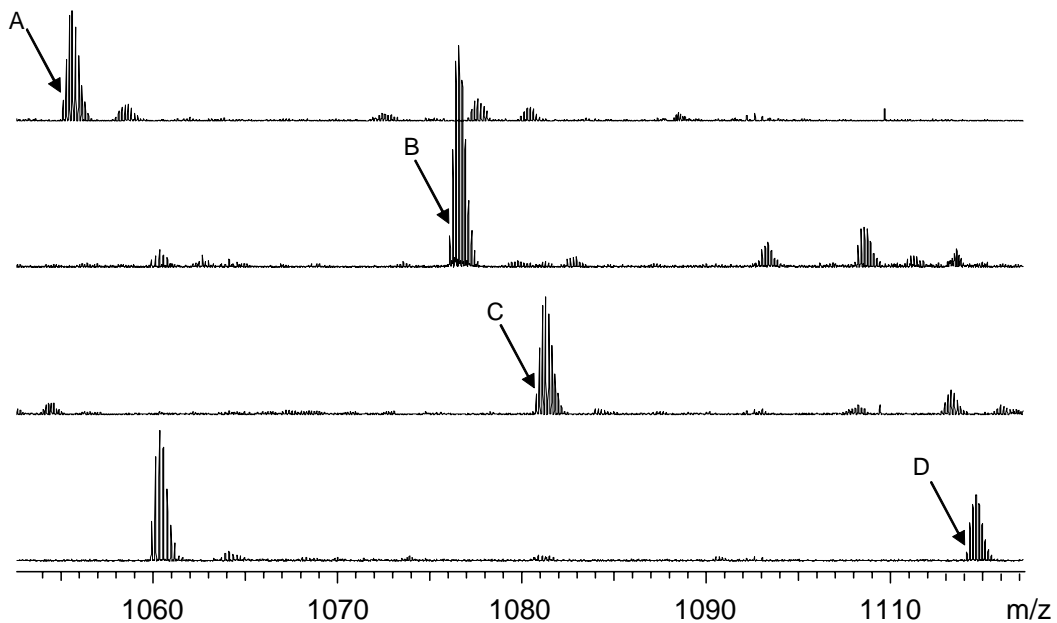


Figure 1.4. Mass spectra of peptides 50-86x310-317, cross-linked with SIA (a) m/z 1055.3881⁺⁵, signal/noise (S/N) = 241.4), GMBS (b) m/z 1080.38319⁺⁵, S/N = 717.7), EMCS (c) m/z 1085.9978⁺⁵, S/N = 462.7), and LC-SMCC (d) m/z 1113.8155⁺⁵, S/N = 344.6). S/N based on most abundant isotope peak. See Table 1.1 for observed $M + H^+$ (Da) and experimental error (Δ ppm).

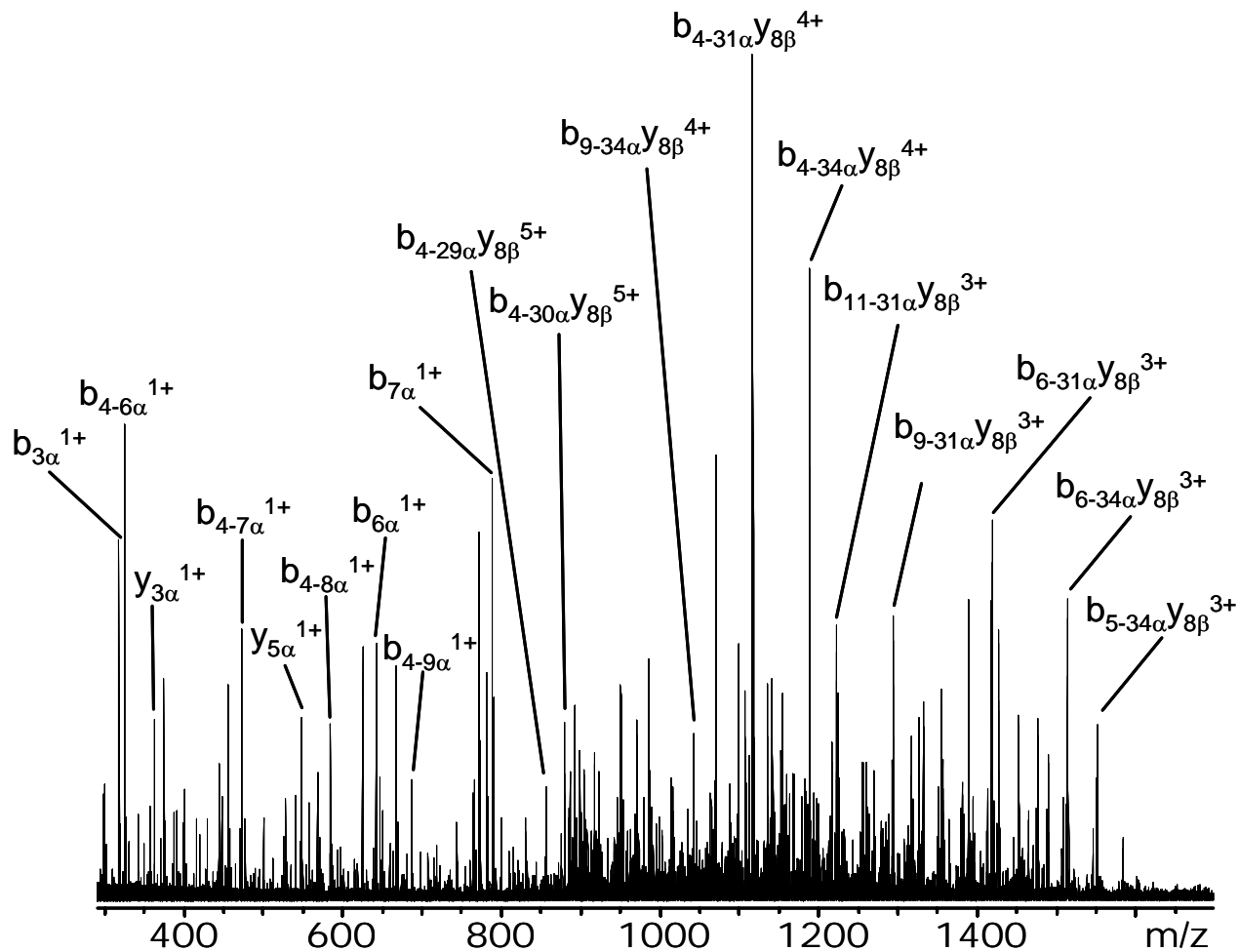


Figure 1.5. Tandem mass spectrum of EMCS cross-linked rhodopsin peptides 50-86x310-317. Only the most abundant ions are labeled due to the complexity of the spectrum. Ion classification is based on the previously established nomenclature.⁵³

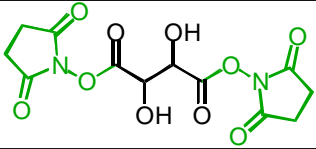
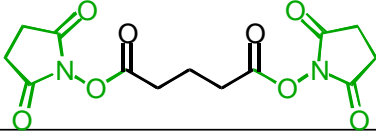
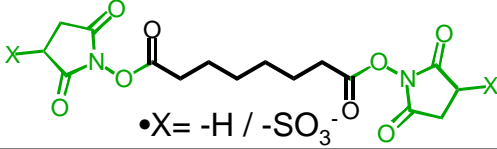
Cross-linker	Structures	N-N distances (Å)
DST		6.32
DSG		7.40
DSS / BS ³	 •X= -H / -SO ₃ ⁻	11.25

Figure 1.6. Lysine-Lysine Cross-linkers

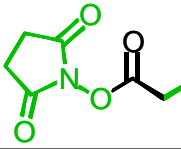
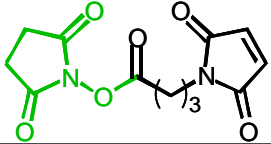
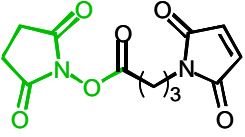
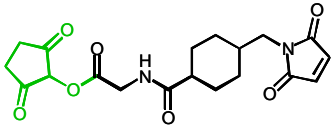
Cross-linker	Structures	S-N distances (Å)
SIA		4.0
GMBS		9.7
EMCS		11.1
LC-SMCC		15.0

Figure 1.7. Cysteine-Lysine Cross-linkers

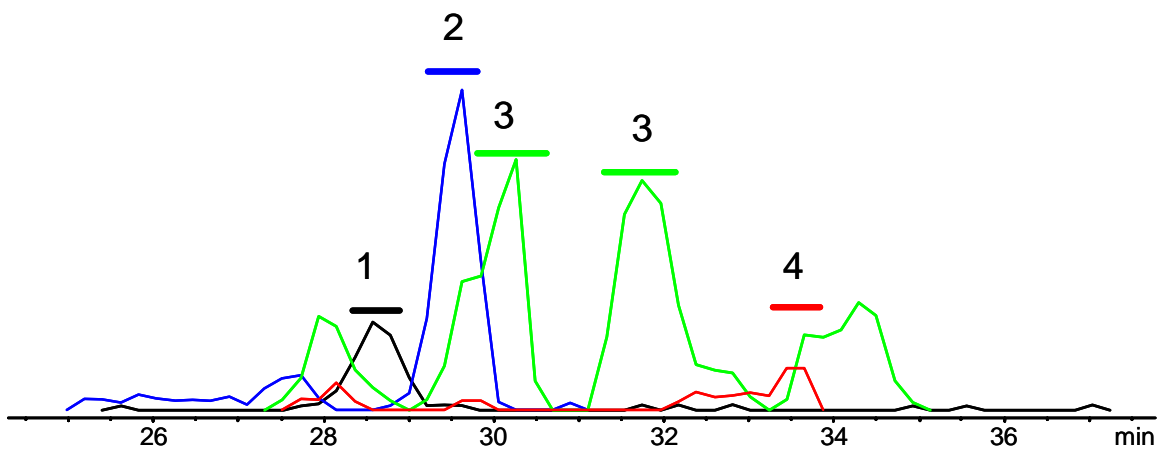


Figure 1.8. Selected ion chromatograms of peptide 208-253 + 1 PE (pyridoethyl group) following acetylation with NHS-acetate. 1) M/z 887, un-modified, 2) M/z 894, 1 Lys-Ac, 3) M/z 901, 2 Lys-Ac, and 4) M/z 908, 3 Lys-Ac.

Chapter 2: A Top down Method for the Determination of Residue Specific Solvent Accessibility in Proteins

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

Abstract

We present a method employing top-down FTMS for the rapid profiling of amino acid side-chain reactivity. The reactivity of side-chain groups can be used to infer residue-specific solvent accessibility and can also be used in the same way as H/D exchange reactions to probe protein structure and interactions. We probed the reactivity of the N-terminal and ϵ -lysine amino groups of ubiquitin by reaction with N-hydroxysuccinimidyl acetate (NHSAc), which specifically acetylates primary amines. Using a hybrid Q-FTMS instrument, we observed several series of multiply acetylated ubiquitin ions that varied with the NHSAc:protein stoichiometry. We isolated and fragmented each member of the series of acetylated ubiquitin ions in the front end of the instrument, and measured the fragment ion masses in the FTMS analyzer cell to determine which residue positions were modified. As we increased the stoichiometric ratio of NHSAc:protein, identification of the fragments from native protein and protein with successively increasing modification allowed the assignment of the complete order of reactivity of the primary amino groups in ubiquitin (Met1 \approx Lys 6 \approx Lys 48 \approx Lys 63 > Lys 33 > Lys 11 > Lys 27, Lys 29). These results are in excellent agreement with the reactivity expected from other studies and predicted from the known crystal structure of ubiquitin. The top-down approach eliminates the need for proteolytic digestion, HPLC separations, and all other chemical steps except the labeling reaction, making it rapid and amenable to automation using small quantities of protein.

Introduction

In the course of optimizing conditions for the chemical cross-linking of proteins¹⁻³ our group has become interested in how the intrinsic reactivity of lysines in the native state structure of a protein influences the outcome of cross-linking reactions employing lysine-reactive reagents. Residue specific reactivity is also of great current interest in structural biology as it provides information about the solvent accessibility and/or reactivity of a residue⁴⁻¹⁷ that can be used in the same manner as H/D exchange reactions to probe protein structure and interactions.¹⁸

Many different techniques have been used to monitor protein solvent accessibility such as Edman degradation⁹, NMR studies^{13, 15}, radioactive and/or fluorescent probes^{10, 16}, and mass spectrometric peptide mapping.^{7, 11, 12, 14, 17} While NMR can give very detailed information about the environment of a specific residue, milligrams of wild-type and genetically modified protein is required. The other techniques all involve proteolytic digestion of the protein followed by identification of the modified residue by mass spectrometry, or sequencing methods.

In work presented here, the top-down mass spectrometry technique¹⁹ was combined with primary amino group labeling,²⁰ and automated software for data reduction and analysis to study the relative solvent accessibility of lysines and the amino-terminus in ubiquitin. The top-down method requires only microgram quantities of wild-type protein, and there is also no need to use time-consuming steps such as proteolytic digestion, liquid chromatography, electrophoretic separation etc. Moreover, the required mass spectrometric analysis can be executed rapidly. For example, a series of MS/MS spectra could be obtained in less than one hour (even without automated data acquisition), which allow assignment of the complete reactivity order of the primary amines within ubiquitin.

Experimental

Materials and protein acetylation

Ubiquitin (Sigma, St. Louis, MO) was dissolved in a 50mM triethylamine (Sigma, St. Louis, MO)/bicarbonate (pH 7.6) buffer (native condition) and/or in a 50mM triethylamine/bicarbonate (pH 7.6) buffer containing 6M guanidine hydrochloride (Sigma, St. Louis, MO) (denaturing condition) at a concentration of 0.5mg/ml. Protein solutions (native and denatured) were split into the twelve independent reactions, six native and six denaturing. Each reaction contained 50ug of protein. Ubiquitin was modified by various concentrations (1:1, 1:3, 1:9, 1:27 and 1:71, ubiquitin:NHSAc) of NHSAc under native and denaturing condition at room temperature for one hour. Reaction in the absence of acetylating agent was performed as a control. All samples were purified using a one-step protein trap cartridge cleanup (Macro Traps, Michrom BioResources, Auburn, CA) and were diluted to a 2 μ M final protein concentration in aqueous solution of 6% acetic acid and 50% methanol (EM Science, Gibbstown, NJ) for subsequent mass spectrometric analysis.

Mass spectrometry

The recent development of mass-selective front-end interfaces for FTMS allowing selective accumulation and dissociation of ions prior to injection into the FTMS analyzer cell has greatly enhanced the capabilities of FTMS instruments.^{21, 22} Experiments in this study were performed on a commercial prototype APEX Q-FTMS instrument equipped with a 7.0 T superconducting magnet and an Apollo ESI ion source (Bruker Daltonics, Billerica MA). The Q front-end consists of a quadrupole mass filter followed by a hexapole collision cell as shown in Figure 2.1. By switching the potentials on the exit lenses indicated in Figure 2.1 appropriately under the control of the data acquisition computer, ions could be accumulated either in the hexapole of the Apollo ESI source, or in the hexapole collision cell of the Q front end, prior to transfer to the FTMS analyzer cell. Mass spectra of the reaction product mixtures were obtained by accumulating ions in the ESI source hexapole and running the quadrupole mass filter in non mass-selective (Rf-only) mode so that ions of a broad m/z range (200-

2000) were passed to the FTMS analyzer cell. Ubiquitin with the desired number of acetylations was isolated from the mixture by setting the quadrupole mass filter to pass the m/z of interest, with a resolution of approximately ± 1 m/z . The quadrupole mass filter had sufficient mass resolution to selectively pass only the species of interest in each case, as determined by accumulating the selected ions in the collision cell at low collision energy and then transferring those ions to the FTMS analyzer cell to obtain a precursor ion mass spectrum. After the clean selection of the desired precursor ion was confirmed, the DC offset on the collision cell was dropped, providing approximately 25 V of collision energy to the ions, which induced fragmentation. The collision energy was adjusted for each species to give extensive fragmentation, and the fragments were simultaneously produced and accumulated in the collision cell. Acetylation of successive primary amino groups in ubiquitin resulted in a shift of the charge state distribution observed by ESI to lower charge states, as acetylation turns a very basic, positive charge-carrying amino group into a much less polar amide. However, for ubiquitin with up to 5 acetylations, the +10 charge state was observed in the ESI spectrum and could be selected by the quadrupole mass filter and fragmented to give excellent MS/MS spectra. For ubiquitin with 6 and 7 acetylations, the +10 charge state was not observed, and the +7 charge state was selected by the quadrupole mass filter and fragmented. While excellent S/N could be obtained for fragmentation of the +7 charge state, the number of fragments observed was significantly lower for this charge state than for the +10.

Results and discussion

Ubiquitin was acetylated using a range of NHSAc concentrations under native and denaturing conditions. All samples were subjected to high-resolution FT-ICR MS. Ubiquitin contains eight primary amino groups (Figure 2.2) that can be acetylated. The data in Figure 2.3 show that ubiquitin could be sequentially acetylated with up to 6 modifications. Surprisingly, under native conditions it was not possible to modify all of the primary amino groups even with a 71 times molar excess of NHSAc. Two sites proved to be highly resistant to labeling as there was only a weak peak indicating addition of a seventh acetyl group and no evidence for the addition of an eighth. Under denaturing conditions, ubiquitin was completely acetylated with a 27-fold excess of acetylating agent (data not shown), which shows that the resistance to labeling was due to the native state structure of the protein. Based on these results, native protein (as a control), and protein with 1 through 7 acetylations were isolated and subjected to MS/MS experiments.

Fragment ion spectra were internally calibrated using several unmodified ubiquitin fragment peaks, resulting in low ppm mass accuracies as shown in Table 2.1. All product ion spectra were then transformed to a list of singly charged fragment ions by a software package developed in-house that incorporates elements of the THRASH²³ and Zscore²⁴ algorithms. The assignment of the monoisotopic masses was performed by MS2PRO, a software package developed at Sandia National Laboratories that can predict and assign all fragments from a protein sequence containing modifications and cross-links. An input file to MS2PRO specifies the molecular weight and residue specificity of the cross-linking reagent. The mass error threshold was kept below 5 ppm for all assignments and the final software output was checked by hand for each collision spectrum. Specifically, all assigned fragments that contained modifications that were useful in establishing the reactivity order were manually checked. These assignments were accepted only if they were absent in the control MS/MS spectrum of the native protein, had reasonable peak shapes, isotopic distributions, and S/N. The nomenclature used for the ions was chosen to be consistent with a recent paper proposing new nomenclature for MS/MS fragment ions from modified and cross-linked proteins and peptides, and references contained therein.²⁵

To determine the relative reactivity of the primary amine-containing residues, MS/MS spectra of the unmodified protein were compared to MS/MS spectra of the protein with increasing numbers of acetylations. Key results are shown in Table 2.1 and Figures 2.4 and 2.5. Detailed analysis of the MS/MS spectrum of the singly acetylated protein revealed the most reactive residues to be Met 1, Lys 6, Lys 48 and Lys 63 as indicated by the sequence ions in Table 2.1 and Figure 2.4. The y40 sequence ion contains two lysine residues (Lys 48, Lys 63) that are modified rapidly (Figure 2.4a). The internal ion b(37-58) contains only Lys 48, while the y14 sequence ion contains only Lys 63, and the spectra of these two ions in Figures 2.4b and 2.4c show that these two lysines are approximately equally reactive. The singly acetylated b5, b6, b8, b11 and b12 sequence ions (Table 2.1), were all

observed in the MS/MS spectrum of singly acetylated ubiquitin. There are three primary amino groups in the sequence covered by these ions. The b6, b11 and b12 ions are observed doubly acetylated in the MS/MS spectra of both doubly and triply acetylated ubiquitin. However, in the MS/MS spectrum of triply acetylated ubiquitin, there was no evidence for triply acetylated b11 or b12 sequence ions, indicating that Met 1 and Lys 6 react in the first stages of the acetylation reaction, but not Lys 11. The relative order of reactivity of Met 1 and Lys 6 are difficult to assign, as no internal ions containing only the Lys 6 residue without other lysines were observed. However, a comparison of the relative ratios of the modified to unmodified fragment ion intensities in the MS/MS spectra of the b5 and b6 ions provides some information regarding this reactivity order. The ratio of unmodified b5 ion to the singly acetylated b5 ion was 3:1 and the relative ratio of unmodified b8 to singly acetylated b8 was 1:1 in the MS/MS spectrum of singly acetylated ubiquitin, indicating the first acetylation occurs with approximately equal probability at these two sites.

The next residue in decreasing order of reactivity is Lys 33. This finding is based on the interpretation of the MS/MS spectra of the doubly acetylated protein fragments shown in Figure 2.5. The y58 ion from the doubly acetylated protein MS/MS experiment appears unmodified in Figure 2.5a, and with 1 and 2 acetyl modifications. As shown in Figure 2.5b, the internal ion b(19-32) does not appear modified in any of the MS/MS spectra, indicating that Lys 27 and Lys 29 are among the least reactive residues in the protein. The internal b(19-36) ion, in contrast, is singly acetylated in the MS/MS spectrum of the doubly acetylated ubiquitin (Figure 2.5c). b(19-36) contains Lys 27, Lys 29, and Lys 33, therefore Lys 33 must be the reactive residue, since Lys 27 and Lys 29 were shown to be unmodified in the MS/MS spectrum of doubly acetylated ubiquitin in Figure 2.5b.

Lys 11 follows Lys 33 in decreasing order of reactivity. The first evidence of its modification was found in MS/MS spectra of the five-fold acetylated ubiquitin. Triply acetylated ions b11 and b12 confirmed its modification in this spectrum (Table 2.1). Furthermore, there was no evidence for modification at Lys 27 and Lys 29 in the MS/MS of spectra up to quintuply acetylated ubiquitin, because an acetylated b(19-32) ion was not observed in (Figure 2.5b).

This leaves Lys 27 and Lys 29 as the least reactive residues in the protein. One of these lysine residues was found to be reactive as shown by the presence of a y58 ion with four acetylations in the MS/MS spectrum of 6- and 7-fold acetylated ubiquitin. Since the y58 ion contains only 5 lysine residues, one of the four must be Lys 27 or Lys 29. As noted in the experimental section, the +10 charge state of the 6- and 7-fold acetylated ubiquitin was not observed in the whole protein ESI spectra, so the +7 charge state was chosen as the precursor ion for fragmentation of 6- and 7-fold acetylated ubiquitin. The +7 charge state did not fragment as extensively and yielded few internal fragment ions, as might be expected on the basis of the reduced coulombic repulsion in the lower charge state. The lower extent of fragmentation meant that it was not possible to find fragments that could be used to differentiate the reactivity of Lys 27 and Lys 29.

Conclusions

While fragments that contained only a single lysine were obtained only for Lys 48 and Lys 63, a careful analysis of the extent of acetylation in other sets of related fragments allowed a qualitative order of reactivity to be assigned for all of the primary amino groups in ubiquitin. Based on these analyses, the complete order of reactivity of the primary amino group in ubiquitin was determined to be Met 1 \approx Lys 6 \approx Lys 48 \approx Lys 63 > Lys 33 > Lys 11 > Lys 27, Lys 29. These results are in excellent agreement with crystal structure data,⁸ which indicate that Lys 11, Lys 27 and Lys 29 are all involved in strong hydrogen bonds to carboxylic acid moieties on other side chains in the protein. Strong H-bonding interactions of Lys 11 with Glu 34 and Lys 29 with Asp 21 have also been proven by single site mutation combined with NMR¹⁵. The crystal structure indicates that the more reactive lysines are involved in much weaker hydrogen bonds to backbone carbonyl groups (Lys 48, Lys 33) or not involved in H-bonding at all (Lys 63). These results also agree with the recent observation that the formation of polyubiquitin involves one of the most reactive lysine residues found in this study, Lys 48.²⁶ In addition, another biologically important form of polyubiquitin that serves as a non-proteolytic marker of protein targets has been identified, and this variant of polyubiquitin is linked via Lys 63.^{27, 28} Solvent accessibility calculations

using the GRASP package show that all of the lysine sidechains in ubiquitin are solvent accessible to an approximately equal extent (using the structure 1UBQ in the PDB databank, data not shown). Using Grasp, we also performed electrostatic point potential calculations that correlate strongly with the lysine reactivity order in ubiquitin: Met 1 (-2.096) \approx Lys 6 (29.2) \approx Lys 48 (17.1) \approx Lys 63 (14.47) > Lys 33 (5.5) > Lys 11 (4.8) > Lys 27 (3.7) , Lys 29 (3.1), where the values in parenthesis are the electrostatic point potentials at the primary amino nitrogen in kcal/mol. . The pKa of an N-terminal methionine amine is \sim 8.0 versus 10.53 for the lysine sidechain, which translates to a reactivity difference of greater than 100-fold. Comparing the electrostatic point potentials for the N-terminus to the lysine sidechains in this case is therefore not appropriate, as local electrostatic effects may not offset the inherent reactivity difference between the two types of reactive amines. These calculations indicate that the local electrostatic environment can also play a large role in influencing the effective pKa, and thus the reactivity, of a lysine side-chain. Thus it is important to note that in studies using reactive side-chain labeling, that a reactive side-chain must be solvent accessible, but lack of reactivity may be due to effects besides solvent accessibility, such as hydrogen bonding.

The reactivity order we observed in this study also correlates well with the cross-links we previously observed in ubiquitin using dissuccinimidyl suberate.¹ The cross-links observed in the previous study were between only the most reactive lysines. These results emphasize the fact that a cross-link can serve as a distance constraint if it is observed, but the absence of a cross-link formation cannot be used as a constraint since lack of reactivity rather than distance may inhibit the formation of cross-links.

The top-down approach using fragmentation in the front end of a hybrid Q-FTMS, combined with the mass accuracy, and resolution of the FTMS mass analyzer has been shown to be useful for rapid solvent accessibility studies on reactive residues. Future work will involve extending the method to larger proteins and a variety of protein labeling reagents that can specifically react at residues other than lysine. We also plan to develop automated labeling and data acquisition schemes to obtain more information from smaller quantities of protein.

Acknowledgements

We would like to acknowledge Marti Head for helpful discussions regarding ubiquitin polymerization. We also very gratefully acknowledge the instrument time on the prototype Q-FTMS instrument at Bruker Daltonics and the support provided by Paul Speir and Ryan Danell of Bruker Daltonics during the experimental work. This work was supported by 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. Kruppa GH, Schoeniger J, Young MM. A top down approach to protein structural studies using chemical cross-linking and fourier transform mass spectrometry. *Rapid Commun Mass Spectrom.* 2003; 17: 155-62.
2. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A.* 2000; 97: 5802-6.
3. Gibson BW. A similar study has recently been undertaken by a group at the Buck Institute for Age Research, in Novato, CA. Their work focused on elucidating the relative reactivities of the primary amino groups in cytochrome c, using a bottom up approach, results to be submitted for publication. Private communication from prof. B. W. Gibson, Buck Institute for Age Research.
4. Batra PP, Roebuck MA, Uetrecht D. Effect of lysine modification on the secondary structure of ovalbumin. *J Protein Chem.* 1990; 9: 37-44.
5. Bazaes S, Silva R, Goldie H, Cardemil E, Jabalquinto AM. Reactivity of cysteinyl, arginyl, and lysyl residues of escherichia coli phosphoenolpyruvate carboxykinase against group-specific chemical reagents. *J Protein Chem.* 1993; 12: 571-7.
6. Chen JG, Liu-Chen S, Rudnick G. Determination of external loop topology in the serotonin transporter by site-directed chemical labeling. *J Biol Chem.* 1998; 273: 12675-81.
7. Fiedler W, Borchers C, Macht M, Deininger SO, Przybylski M. Molecular characterization of a conformational epitope of hen egg white lysozyme by differential chemical modification of immune complexes and mass spectrometric peptide mapping. *Bioconjug Chem.* 1998; 9: 236-41.
8. Jabusch JR, Deutsch HF. Isolation and crystallization of ubiquitin from mature erythrocytes. *Prep Biochem.* 1983; 13: 261-73.
9. Jabusch JR, Deutsch HF. Localization of lysines acetylated in ubiquitin reacted with p-nitrophenyl acetate. *Arch Biochem Biophys.* 1985; 238: 170-7.
10. Jackson GE, Young NM. Determination of chemical properties of individual histidine and tyrosine residues of concanavalin a by competitive labeling with 1-fluoro-2,4-dinitrobenzene. *Biochemistry.* 1986; 25: 1657-62.
11. Kalkum M, Przybylski M, Glocker MO. Structure characterization of functional histidine residues and carbethoxylated derivatives in peptides and proteins by mass spectrometry. *Bioconjug Chem.* 1998; 9: 226-35.
12. Leite JF, Cascio M. Probing the topology of the glycine receptor by chemical modification coupled to mass spectrometry. *Biochemistry.* 2002; 41: 6140-8.
13. Macdonald JM, Haas AL, London RE. Novel mechanism of surface catalysis of protein adduct formation. Nmr studies of the acetylation of ubiquitin. *J Biol Chem.* 2000; 275: 31908-13.
14. Suckau D, Mak M, Przybylski M. Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc Natl Acad Sci U S A.* 1992; 89: 5630-4.
15. Sundd M, Iverson N, Ibarra-Molero B, Sanchez-Ruiz JM, Robertson AD. Electrostatic interactions in ubiquitin: Stabilization of carboxylates by lysine amino groups. *Biochemistry.* 2002; 41: 7586-96.

16. Winkler MA, Fried VA, Merat DL, Cheung WY. Differential reactivities of lysines in calmodulin complexed to phosphatase. *J Biol Chem.* 1987; 262: 15466-71.
17. Zappacosta F, Ingallinella P, Scaloni A, Pessi A, Bianchi E, Sollazzo M, Tramontano A, Marino G, Pucci P. Surface topology of minibody by selective chemical modifications and mass spectrometry. *Protein Sci.* 1997; 6: 1901-9.
18. Engen JR, Smith DL. Investigating protein structure and dynamics by hydrogen exchange ms. *Anal Chem.* 2001; 73: 256A-265A.
19. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* 1999; 121: 806 -812.
20. Hermanson GT. Bioconjugate techniques. 1996; 127.
21. Wang YS, SDH; Hendrickson, CL; Marshall, AG. Mass-selective ion accumulation and fragmentation in a linear octopole ion trap external to a fourier transform ion cyclotron resonance mass spectrometer. *International Journal of Mass Spectrometry.* 2000; 198: 113-120.
22. Belov ME, Nikolaev EN, Anderson GA, Udseth HR, Conrads TP, Veenstra TD, Masselon CD, Gorshkov MV, Smith RD. Design and performance of an esi interface for selective external ion accumulation coupled to a fourier transform ion cyclotron mass spectrometer. *Anal Chem.* 2001; 73: 253-61.
23. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of the American Society for Mass Spectrometry.* 2000; 11: 320-332.
24. Zhang ZQ, Marshall AG. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *Journal of the American Society for Mass Spectrometry.* 1998; 9: 225-233.
25. Schilling B, Row RH, Gibson BW. Ms2assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *Journal of the American Society for Mass Spectrometry.* In press;
26. Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol Rev.* 2002; 82: 373-428.
27. Spence J, Sadis S, Haas AL, Finley D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol.* 1995; 15: 1265-73.
28. McKenna S, Moraes T, Pastushok L, Ptak C, Xiao W, Spyropoulos L, Ellison MJ. An nmr-based model of the ubiquitin-bound human ubiquitin conjugation complex mms2{middle dot}ubc13. The structural basis for lysine 63 chain catalysis. *J Biol Chem.* 2003; 278: 13151-13158.

Figures

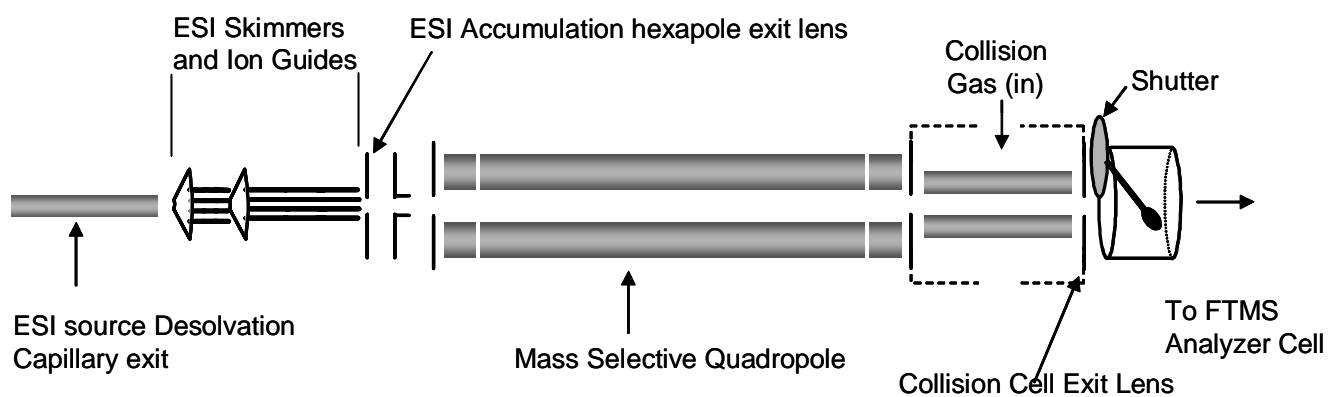


Figure 2.1. Diagram of the mass-selective front end accumulation quadrupole FTMS used in these studies.

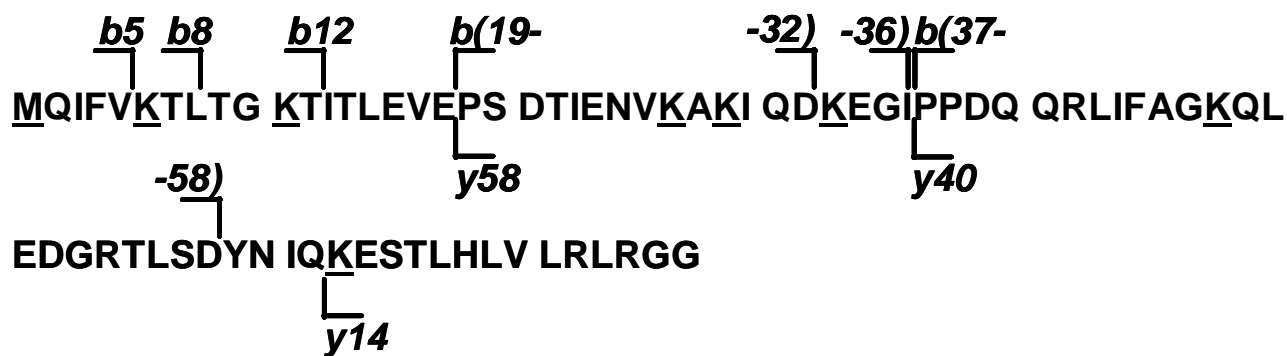


Figure 2.2. Ubiquitin sequence and relevant sequence ions. The reacting residues are underlined.

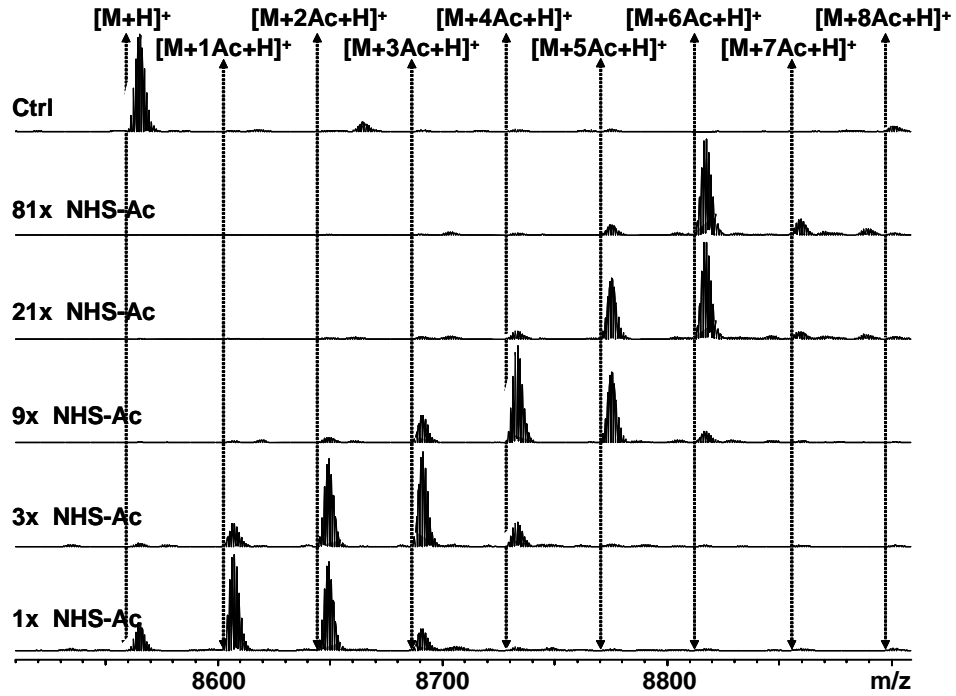


Figure 2.3. FTMS spectra of native protein and Ubiquitin modified by different molar ratios of NHS-Ac.

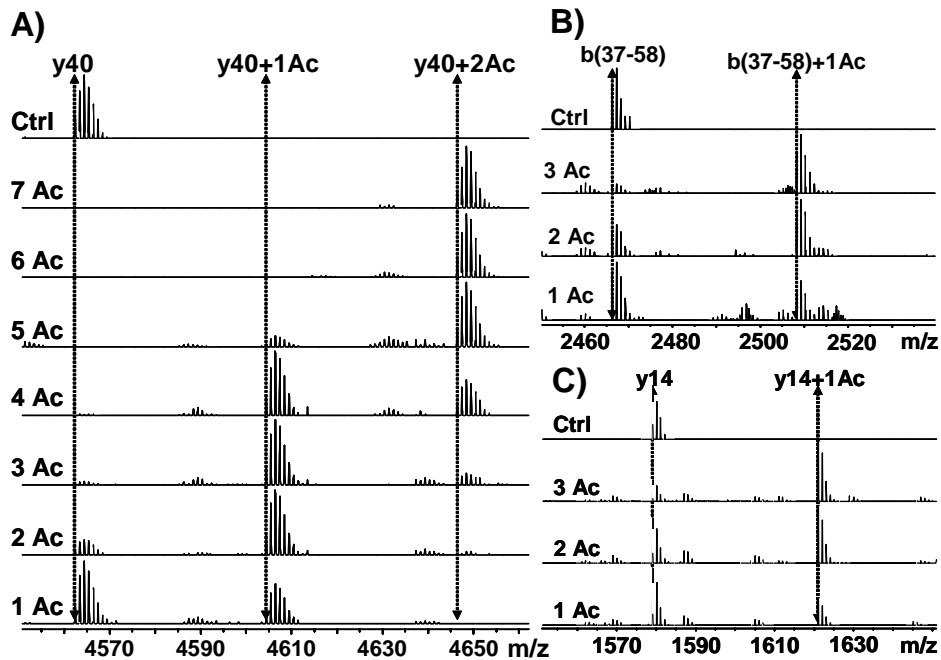


Figure 2.4. Q-CID MS/FTMS spectra of native and sequentially acetylated ubiquitin. The Y-axis labels indicate the number of acetylations present on the ion isolated for MS/MS, "CTRL" indicates a control MS/MS spectrum of the unmodified protein. The detailed analysis (see text) of y40 (A) with b(37-58) (B) and y14 (C) reveals an equal reactivity of K48 and K63. They are both present with one acetyl group in the MS/MS spectrum of singly modified ubiquitin.

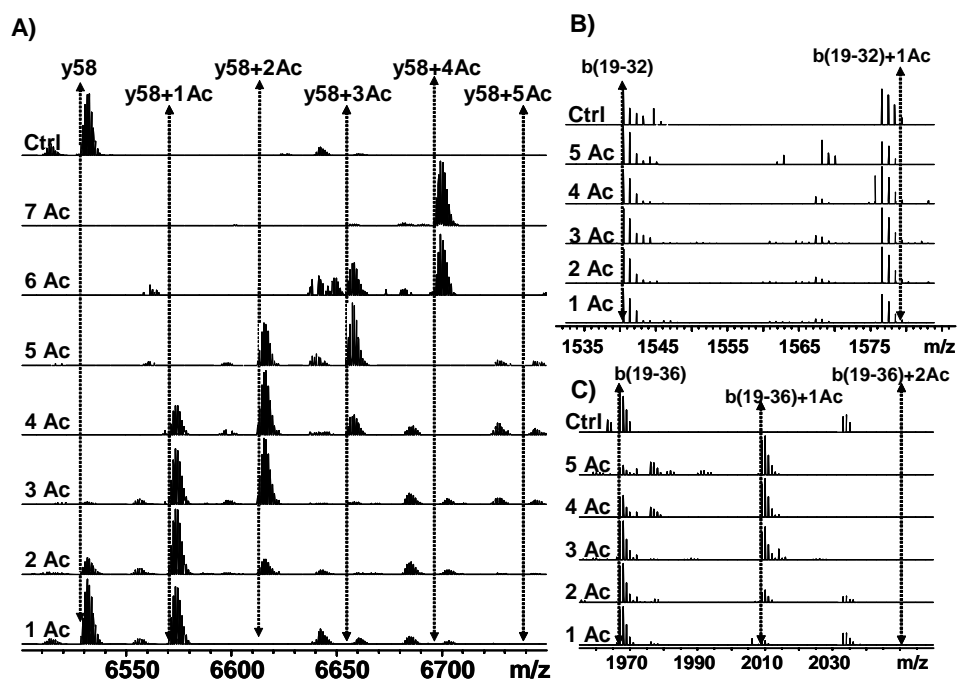


Figure 2.5. Q-CID MS/FTMS spectra of native and sequentially acetylated ubiquitin. The Y-axis labels indicate the number of acetylations present on the ion isolated for MS/MS, "CTRL" indicates a control MS/MS spectrum of the unmodified protein. The detailed analysis of y58 (A) with b(19-32) (B) and b(19-36) (C) reveals K33 as the second most reactive residue and K27, 29 as less reactive. Modified K33 is present in the MS/MS spectrum of doubly acetylated Ubiquitin while K27, 29 are not modified in the MS/MS spectrum of five-fold acetylated protein.

Tables

Table 2.1. Series of b ions from the N-terminus that allows assignment of the reactivity of M1, K6, K11.

Fragment	Precursor	Ubiquitin + 1 acetate	Ubiquitin + 2 acetates	Ubiquitin + 3 acetates	Ubiquitin + 4 acetates	Ubiquitin + 5 acetates	Ubiquitin + 6 acetates	Ubiquitin + 7 acetates
B5	(M+H) ⁺ _{obs.}	619.3266	619.3274	619.3277	619.3273	619.3273	n.o.	n.o.
	Δ (ppm)	2.0	0.7	0.2	0.9	0.9	-	-
B5+ 1 Ac	(M+H) ⁺ _{obs.}	661.3382	661.3387	661.3381	661.3383	661.3384	661.3383	661.3386
	Δ (ppm)	0.3	0.5	0.5	0.1	0.0	0.1	0.4
B6	(M+H) ⁺ _{obs.}	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.
	Δ (ppm)	-	-	-	-	-	-	-
B6+ 1 Ac	(M+H) ⁺ _{obs.}	789.4337	789.4338	789.433	789.4335	789.4336	789.4332	n.o.
	Δ (ppm)	0.5	0.5	0.5	0.2	0.3	0.2	-
B6+2 Ac	(M+H) ⁺ _{obs.}	n.o.	831.4446	831.4433	831.4448	831.4449	831.4437	831.4445
	Δ (ppm)	-	0.8	0.7	1.1	1.2	0.2	0.7
B8	(M+H) ⁺ _{obs.}	961.5538	961.5532	961.5536	n.o.	n.o.	n.o.	n.o.
	Δ (ppm)	0.8	1.3	1	-	-	-	-
B8+1 Ac	(M+H) ⁺ _{obs.}	1003.5662	1003.5652	1003.5646	1003.5652	1003.5648	1003.5653	n.o.
	Δ (ppm)	1.2	0.1	0.5	0.1	0.2	0.2	-
B8+2 Ac	(M+H) ⁺ _{obs.}	n.o.	n.o.	n.o.	1045.5753	1045.5756	1045.5760	1045.5747
	Δ (ppm)	-	-	-	0.2	0.0	0.5	0.8
B11	(M+H) ⁺ _{obs.}	1247.7188	1247.7167	n.o.	n.o.	n.o.	n.o.	n.o.
	Δ (ppm)	0.1	1.6	-	-	-	-	-
B11+1 Ac	(M+H) ⁺ _{obs.}	1289.7314	1289.7285	n.o.	1289.7292	n.o.	n.o.	n.o.
	Δ (ppm)	1.8	0.5	-	0.1	-	-	-
B11+2 Ac	(M+H) ⁺ _{obs.}	n.o.	1331.7426	1331.7401	1331.7408	1331.7405	n.o.	n.o.
	Δ (ppm)	-	2.1	0.3	0.8	0.5	-	-
B11+3 Ac	(M+H) ⁺ _{obs.}	n.o.	n.o.	n.o.	n.o.	1373.7476	n.o.	1373.7495
	Δ (ppm)	-	-	-	-	2.0	-	0.5
B12	(M+H) ⁺ _{obs.}	1348.7649	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.
	Δ (ppm)	1.1	-	-	-	-	-	-
B12+1 Ac	(M+H) ⁺ _{obs.}	1390.7758	1390.7771	1390.7769	1390.7787	n.o.	n.o.	n.o.
	Δ (ppm)	0.8	0.2	0.0	1.3	-	-	-
B12+2 Ac	(M+H) ⁺ _{obs.}	n.o.	1432.7877	1432.7874	1432.7885	1432.7887	n.o.	n.o.
	Δ (ppm)	-	0.2	0.1	0.7	0.9	-	-
B12+ 3Ac	(M+H) ⁺ _{obs.}	n.o.	n.o.	n.o.	n.o.	1474.7971	1474.7982	1474.7972
	Δ (ppm)	-	-	-	-	0.6	0.2	0.5

This page intentionally left blank.

Chapter 3: A Top-down Approach to Protein Structure Studies using Chemical Crosslinking and Fourier Transform Mass Spectrometry

Petr Novak, Malin M. Young, Joseph S. Schoeniger, Gary H. Kruppa

Abstract

In a preliminary communication we described a top down approach to the determination of chemical crosslink location in proteins using FTMS. We have since extended the approach to use a series of homobifunctional crosslinkers with the same reactive functional groups, but different crosslinker arm lengths. Correlating crosslinking data across a series of related linkers allows the distance constraint derived from a crosslink between two reactive side chains to be determined more accurately, and increases the confidence in the assignment of the crosslinks. In ubiquitin there are 7 lysines with primary amines and the amino terminus. Dissuccinimidyl suberate (DSS, crosslinker arm length = 11.4 Å) dissuccinimidyl glutarate (DSG, crosslinker arm length = 7.5 Å), and dissuccinimidyl tartrate (DST, crosslinker arm length = 5.8 Å) are homobifunctional crosslinking reagents that react specifically with primary amines. Using MS/MS on the singly internally crosslinked precursor ion of ubiquitin, we found crosslinks with DSS and DSG between the amino terminus and Lys 6, between Lys 6 and Lys 11, and between Lys 63 and Lys 48. Using dissuccinimidyl tartrate DST, the shortest crosslinker in the series, only the crosslinks between the amino terminus and Lys 6, and between Lys 6 and Lys 11 were observed. The observed crosslinks are consistent with the crystal structure of ubiquitin, if the lysine side-chains and the amino terminus are assumed to have considerable flexibility. In a separate study we probed the reactivity of the primary amino groups in ubiquitin using the amino acetylating reagent N-hydroxy succinimidyl acetate (NHSAc), and a top down approach to localize the acetylated lysine residues. The reactivity order obtained in that study (M1~ K6~ K 48~ K63) > K33 > K11 > (K27, K29), shows that the crosslink first formed in ubiquitin by reaction with DSS and DSG occur between the most reactive residues.

Introduction

A new method has recently been proposed for using chemical crosslinking with Mass Spectrometry to study the 3-Dimensional structure of proteins, and named MS3D^{1,2}. As originally demonstrated, the procedure for carrying out MS3D studies started with reacting the protein with a crosslinking reagent in its native state in solution. After purification, the monomeric protein, including the protein molecules containing crosslinker related modifications, was digested proteolytically. Mass spectrometric analysis combined with computer software prediction of the masses expected for crosslinked proteolytic peptides allowed the crosslinked amino acid residues to be identified. The length of the cross-linking reagent linker arm then provided a constraint for the maximum distance between the identified residues.

In a preliminary communication we described a top down approach to the determination of chemical crosslink positions in a protein using FTMS.³ In the top down approach, the crude crosslinked protein mixture is injected into an ESI-FTMS instrument, and the crosslink positions localized by multiple stages of fragmentation and mass spectrometry, rather than proteolytic digestion followed by LC/MS. This approach is analogous to the top down approach to protein sequence and post-translational modification analysis proposed by McLafferty et al. and has all the advantages described in the original papers⁴⁻⁸. We have demonstrated certain advantages to our top down approach unique to the problem of localizing crosslinks in proteins. The top down approach eliminates the need for separation of the monomeric protein species; this is accomplished via a 'gas-phase purification' in the FTMS analyzer cell. In addition, this gas-phase purification is significantly better than gel or size-exclusion chromatography separation, because the low resolution of these methods results in the isolation of a complex mixture of monomeric species. This mixture may include unmodified protein, protein with deadend crosslinks, and singly or multiply crosslinked species. Finding mass spectrometric peaks from crosslinked proteolytic fragments which may be of low abundance compared to other fragments in the digest of the monomer mixture is a problem that has been discussed previously both for the study of protein complexes via intermolecular crosslinks^{9,10} and protein tertiary structure via intra-molecular crosslinks^{11,12}. Previous studies have suggested the use of cross-linkers with unusual isotope patterns or labels to simplify the identification of crosslinked proteolytic fragments in the mass spectra^{10,12}. In the FTMS analyzer cell it is possible to isolate only the singly internally crosslinked species, which greatly enhances the probability of observing and identifying crosslinks in the fragment spectrum.

The top down approach also eliminates the need for proteolytic digestion; this is replaced by a gas-phase fragmentation and MS/MS analysis of the fragments to localize the crosslinks. Furthermore, this approach eliminates the need for HPLC-MS of the proteolytic peptides, which is necessary for the analysis of complex proteolytic peptide mixtures due to the lower resolution of other mass analyzers typically used, and to minimize the chance that low abundance proteolytic peptides will not be observed due to ionization suppression effects. Another benefit of eliminating proteolysis and LC/MS is that common peptide modifications due to the reaction and solvent conditions of proteolysis and HPLC (e.g., formylation, oxidation, sodiation) are eliminated. In our experience, these effects can considerably complicate the interpretation of bottom up MS3D experiments. The resolution and mass accuracy of FTMS make the identification of cross-linked fragments unambiguous, and while the MS/MS spectra are complex, the spectral data reduction and interpretation have been automated.

We have extended our initial work on ubiquitin to include results from a series of primary amine-specific, homobifunctional crosslinkers differing only in the crosslinker arm length. Using a single crosslinker, only an upper bound to the distance between two reactive residues can be obtained, which is less than or equal to the maximum distance that the crosslinker arm can span. If a series of crosslinkers with decreasing arm lengths and the same reactive groups are used, it is possible to reduce this upper bound distance, or even derive a lower bound if the crosslink between two specific positions disappears with shorter crosslinker arm length. For these studies we have implemented a new Q-FTMS configuration that enables mass selective accumulation and fragmentation of ions in the front-end of the instrument prior to FTMS analysis. The new instrumental configuration eliminates the need for time-consuming MSⁿ analyses in the FTMS analyzer cell, and yields extensive fragmentation that facilitates the localization of the crosslinks.

Our results show that when ubiquitin is crosslinked, the singly internally crosslinked species has only three crosslinks, K1-K6, K6-K11, and K48-K63. Ubiquitin has 7 lysine residues and the amino terminus that are potential reactive sites for disuccinimidyl ester crosslinking reagents. We will discuss the implications of side chain reactivity for the formation of intra-molecular crosslinks in proteins and show that the crosslinks we are able to localize from singly crosslinked ubiquitin precursor ions are formed between the most reactive lysine residues.

Experimental

Materials and crosslinking

Ubiquitin (Sigma, St. Louis, MO) was dissolved in a 50mM triethylamine/bicarbonate (pH 7.6) buffer at a concentration of 0.25 to 0.5 mg/ml, or 25-50 micromolar. Typically each reaction contained 50 ug of protein. The homobifunctional crosslinking reagents disuccinimidyl suberate (DSS), disuccinimidyl glutarate (DSG) and disuccinimidyl tartrate (DST) (Pierce, Rockford, IL) were added to the buffer as concentrated solutions in DMSO, 0.5-1.0 uL of stock crosslinker solution was added to 100 microliters of the protein and mixed well. Thus the DMSO concentration was kept below 1% v/v, and DMSO is not expected to destabilize the native state protein structure at this concentration. Mass spectrometric measurements indicated that the reaction was complete in 30 minutes to one hour, but we typically let the reaction mixture sit at room temperature for several hours or overnight. The crude reaction mixture was subjected to a simple cleanup using a one-step protein trap cartridge (Macro Traps, Michrom BioResources, Auburn, CA) to remove unwanted buffer and small molecule byproducts and contaminants from the crosslinking reagent. The cleaned samples in the eluent solvent from the macrotrap were diluted to a 2 μ M final protein concentration in aqueous solution of 6% acetic acid and 50% methanol for subsequent mass spectrometric analysis.

Mass spectrometry

The recent development of mass-selective front-end interfaces for FTMS allowing selective accumulation and dissociation of ions prior to injection into the FTMS analyzer cell has greatly enhanced the capabilities of FTMS instruments.^{13,14} The mass spectrometry in this study was performed on a Bruker Daltonics APEX Q-FTMS instrument equipped with a 7.0 Tesla superconducting magnet, and an Apollo ESI source, at the manufacturer's facility in Billerica, MA. Additional experiments were subsequently performed on the APEX II FTMS in our laboratories following an upgrade to the APEX Q configuration, which consisted of installing the quadrupole and collision cell unit between the Apollo ESI source and the rest of the instrument. Mass spectra of the reaction product mixtures were obtained by accumulating ions in the ESI source hexapole and running the quadrupole mass filter in non mass-selective (Rf-only) mode so that ions of a broad m/z range (200-2000) were passed to the FTMS analyzer cell. Ubiquitin with a single internal crosslink was isolated from the mixture by setting the quadrupole mass filter to pass the m/z of interest, with a resolution of approximately ± 1 m/z . The quadrupole mass filter had sufficient mass resolution to selectively pass only the species of interest in each case, as determined by accumulating the selected ions in the collision cell at low collision energy and then transferring those ions to the FTMS analyzer cell to obtain a precursor ion mass spectrum. After the clean selection of the desired precursor ion was confirmed, the DC offset on the collision cell was dropped, providing approximately 25 V of collision energy to the ions, which induced fragmentation. The collision energy was adjusted for each species to give the desired degree of fragmentation, and the fragments were simultaneously produced and accumulated in the collision cell. The accumulation time also affected the degree of fragmentation and was adjusted to give the maximum possible FTMS signal while retaining the desired degree of fragmentation. All MS/MS spectra presented in this work were acquired in this way, and we refer to this mode of operation as Q-CID FTMS. Accumulation times for the spectra shown here were typically 3-5 seconds, and 64 scans were averaged, with each scan consisting of 3-5 seconds ion accumulation time and 1 second to acquire a 1024k transient. Unmodified ubiquitin parent ions of the same charge state were also fragmented under the same conditions as a control.

Data Reduction and Analysis

The transient signals were processed by Gaussian apodization, Fourier transformation and magnitude calculation. Low ppm mass accuracies were obtained by using several abundant, unmodified b- and y-ions from ubiquitin as internal calibrants. Reduction of the spectra is automated by a previously described macro that reduces the peak list to a set of unique monoisotopic masses.³ Deconvoluted spectra were calculated in XMASS using the Zscore algorithm¹⁵, which is implemented in the XMASS software package.

Assignment of the fragmentation spectra is performed by MS2LINKS, developed in our laboratory. MS2LINKS accepts as user input the protein sequence, the mass modification and amino acid specificity of the cross-linking reagent and calculates a complete theoretical fragmentation library. The fragmentation library contains all possible cross-linking possibilities for the b-type, y-type and internal fragment ions (c-,z- and x-,y- type ions can also be calculated and assigned if desired). Given an input m/z fragment list, MS2PRO assigns the m/z of peaks that are within a defined ppm error to fragments in the theoretical library. When the input list indicates that the monoisotopic peak may have been or definitely was not observed, MS2LINKS also checks for matches to the first ¹³C isotope peak of the species in the theoretical library.

Results

The deconvoluted spectra of ubiquitin after reaction with the three crosslinkers DST, DSG and DSS are shown in Figure 3.1. Also shown in Figure 3.1 are the structures and maximum crosslinker arm lengths for each of the crosslinkers.¹⁶ The mass accuracy and resolution of FTMS make the assignment of the various crosslinker related modification unambiguous, and allows the stoichiometry of the crosslinking reaction to be optimized. The ubiquitin concentration which gave the best relative yield of the singly internally crosslinked species while minimizing intermolecular crosslinking was 25 micromolar. For DSS and DSG the optimum concentration was equimolar to the ubiquitin concentration, while for DST an eight times molar excess of the crosslinker yielded the best results. The spectra were internally calibrated using the most abundant isotopes of the unmodified ubiquitin charge states, allowing the exact mass measurement of the singly internally crosslinked species from the +10 charge state: ubiquitin + DST, experimental $[M+H]^+ = 8674.601$, theoretical = 8674.606, error = 0.6 ppm; ubiquitin + DSG, experimental $[M+H]^+ = 8656.654$, theoretical = 8656.645, error = 1.0 ppm; ubiquitin + DSS, experimental $[M+H]^+ = 8698.685$, theoretical = 8698.692, error = 0.8 ppm.

Figures 3.2-3.4 show comparisons of the Q-CID FTMS spectra obtained by selecting and fragmenting the singly internally crosslinked species in the +10 charge state. In all cases, the MS/MS spectrum of ubiquitin is shown at the bottom as a control. An overview of the deconvoluted MS/MS spectra for each singly internally crosslinked species is shown in Figure 3.2. While a few clear differences can be seen in the overview, it is necessary to examine the spectra in some detail, guided by the results of the automated data reduction and analysis software, to derive information to localize the crosslink positions. Figure 3.3 shows expanded views of the MS/MS spectra taken under CID conditions that were adjusted to optimize the yield of the large b52 and y58 fragments. Figures 3.3a-3.3c are deconvoluted spectra, and the fragments which are shifted by the exact mass corresponding to the crosslinker are labeled. Figure 3.3d shows the raw, multiply charged spectrum comparing the 19-36 internal fragment ion region. In order to more accurately determine the crosslink locations, MS/MS spectra were measured with the collision and accumulation conditions optimized to give extensive fragmentation. The spectra are complex and contain many small singly and doubly charged fragments, which the automated data reduction software correctly identifies, but which are not correctly assigned by the available deconvolution software. Thus in Figure 3.3d and 4 expansions of the raw, multiply charged MS/MS spectra are shown in the region of ions that were important for the identification of crosslink positions. Some of the fragments that were most important for determining crosslink locations are listed in Table 2.1, together with mass measurement accuracies.

Discussion

Analysis of crosslink positions. The spectra in Figure 3.3 allow the assignment of most of the crosslinking to near either the carboxy terminus or the amino terminus. An annotated sequence of ubiquitin is shown in Figure 3.5, to aid in the discussion of the MS/MS spectra. The y58 and b52 ions both contain a large part of the sequence, including lysines 27, 29 and 33 in the center of the sequence. However, Figure 3.3c shows that the y40 is present both unmodified and with a crosslink when starting with DSS or DSG crosslinked ubiquitin (but not DST). This immediately allows an assignment of a crosslink between K48 and K63, as these are the only two lysines in the sequence covered by y40. This cannot be the only crosslink as demonstrated by the b52 ion, which is seen only with a crosslink, when starting from all three singly crosslinked precursor ions. For DSS and DSG the presence of a b52 ion with and without a modification would be expected, since from Figure 3.3c we know that some of the crosslink must be between K48 and K63. However, examination of the sequence shows that if the crosslink is not in the part of the sequence covered by the b52 ion, then the only other possibility is the K48-K63 crosslink, and K48 is in the part of the sequence covered by the b52 ion. Thus if the crosslink is not entirely in the part of the sequence covered by the b52 ion, then the formation of the b52 ion is inhibited by the K48-K63 crosslink, as indicated on Figure 3.5. Figure 3.3c shows that DST does not crosslink K48 to K63, and thus all of the crosslinking in this case does occur in the part of the sequence covered by the b52 ion. Figure 3.3d shows the multiply charged spectrum in the region of the 19-36 internal ion. Internal ions in peptide mass spectrometry are normally designated by their one-letter amino acid sequence, but for whole protein MS/MS this can be quite long. We will designate internal ions by the sequence range covered using the normal N-terminus to C-terminus numbering. The 19-36 internal ion is abundant as it is formed by cleavages next to proline on both ends. The arrows on the spectrum indicate the positions where peaks would be observed corresponding to this ion containing a DST, DSG or DSS crosslink. In addition to this particular ion, an extensive sequence of 19-X internal ions is observed, two are shown in Table 3.1, and in none of these ions is there evidence for crosslink formation with any of the crosslinkers employed. While it is difficult to entirely rule out crosslink formation at these positions based on this evidence, since the crosslinker may also cleave or be present in low abundance, it does focus the search for the major crosslink containing peaks to other regions of the sequence.

As shown in Figure 3.4a, the b12 ion is clearly observed to be present with masses corresponding to the fragment with a DST, DSG, and DSS crosslink. The assignments from the automated reduction and analysis software shown in Table 3.1 were all confirmed by examination of the spectra and comparison with the control spectra as illustrated in Figure 3.4. Each crosslink had to be observed in a series of related fragments, and with more than one crosslinking reagent, and all fragments related to the crosslink had to be absent in the control spectrum to confirm an assignment. There are three primary amines in the sequence covered by the b12 ion, the amino terminal methionine (M1), K6, and K11. Thus there are three possible crosslinks, and two are confirmed to be present by the spectra shown in Figures 3.4b and 3.4c. The b7 ion shown in Figure 3.4b is observed unmodified in the MS/MS spectra of all three crosslinked samples, and peaks corresponding to crosslink containing b7 fragments are also observed, showing unambiguously that M1 must crosslink to K6, since M1 and K6 are the only residues with primary amines in the sequence covered by the b7 ion. Likewise, Figure 3.4c shows that the 4-13 internal fragment is present with a crosslink for all three crosslinkers, showing that K6 must crosslink to K13, the only two lysines in the part of the sequence covered by this internal fragment. A crosslink from M1-K11 cannot be assigned or ruled out based on the fragmentation patterns observed.

Correlation with protein structure. The crystal structure of ubiquitin, with the observed crosslinks indicated is shown in Figure 3.6. Table 3.2 lists the crosslinks observed with the $N\epsilon$ - $N\epsilon$ primary amine and $C\alpha$ - $C\alpha$ interatomic distance from the crystal structure.¹⁷ Also shown in Table 3.2 are the distance constraints between the primary amines (the $N\epsilon$ - $N\epsilon$ distances) that can be derived from the maximal crosslinker arm length. Note that the $N\epsilon$ - $N\epsilon$ distance in the crystal structure is greater than the maximum distance implied by the observation of a crosslink for all of the observed crosslinks. The crystal structure is a static snapshot of a dynamic system, and lysine side-chains are expected to have considerable flexibility. A $C\alpha$ - $C\alpha$ upper limit for the crosslinking distance can be derived by adding twice the length of a fully extended lysine sidechain (6.3 Å) to the length of the crosslinker arm. These $C\alpha$ - $C\alpha$ constraints are shown in Table 3.2. (For an amino terminus to lysine crosslink,

6.3 Å is added for one lysine sidechain and 1.7 Å is added for the terminal N-C bond distance). For the M1-K6 crosslink, the crystal structure $C\alpha-C\alpha$ distance is still greater than that implied by the observation of a crosslink with DST assuming full lysine side-chain flexibility. However it is greater by only a few Å, and it is likely that the amino terminus is sufficiently flexible to allow the terminal amino group to move close enough to the K6 amino group to allow formation of the crosslink. For the K6-K11 crosslink, the $N\epsilon-N\epsilon$ is 14 Å in the crystal structure. However, we examined the structure and found that it was relatively easy to move the primary amines to within a few Å of each other through strain-free rotations about the C-C bonds in the lysine chain. For the K48-K63 crosslink we obtained both an upper limit to the distance between these residues from the observation of a crosslink with DSS and DSG, and a lower limit on the distance from the absence of a crosslink with DSG. In the crystal structure the $C\alpha-C\alpha$ distance is actually smaller than the lower limit predicted by our analysis. This is probably due to the assumptions made in our derivation of the maximum possible $C\alpha-C\alpha$ distance that a crosslinker can span, namely we have used the maximal crosslinker arm length, the fully extended lysine side chain length, and assumed full flexibility. Molecular dynamics simulations would be required to determine the probability of the side chains and the crosslinker arm achieving this fully extended configuration. Such calculations and correlation with a wider variety of protein structures are needed to define an accurate upper limit to the $C\alpha-C\alpha$ that can be spanned by primary amine specific labeling reagents.

Correlation of crosslink formation with primary amino group reactivity. There are eight primary amino groups in ubiquitin; seven on lysines and the amino terminus. Many crosslinks that are possible based on the distance between the primary amines were not detected by the Q-CID FTMS experiments in this study. In particular the middle lysines in the sequence, K27, K29, and K33 were not observed to form crosslinks. For the experiments presented here, the crosslinking reaction product containing only a single internal crosslink was specifically isolated and probed by MS/MS. The crosslinks that were not observed could be missing either due to distance or low reactivity with the N-hydroxy succinimidyl esters. To probe the issue of reactivity, we used N-hydroxy succinimidyl acetate (NHSAc) as a reagent to specifically label the primary amines in ubiquitin using the same chemistry as the crosslinkers. By isolating precursor ions of ubiquitin with successively increasing numbers of acetylations, we were able to determine the complete order of reactivity of the primary amino groups: (M1 \approx K6 \approx K48 \approx K63 > K33 > K11 > K27, K29).¹⁸ K27 and K29 were shown to be very unreactive toward NHSAc, so it is not surprising that the first crosslink formed in ubiquitin does not involve these two lysines. The crosslinks observed all involve the most reactive lysines, except for K11, which crosslinks with other nearby very reactive lysines. Thus the crosslinking results are entirely consistent with the reactivity results, the only anomaly being that no crosslinks were observed involving the relatively reactive K33, even though the $C\alpha-C\alpha$ distances between K33 and several other reactive lysines are within the crosslinking distances for DSG and DSS.

Conclusions

Previously we demonstrated that top down SORI-CID FTMS/MS experiments on isolated ubiquitin precursor ions containing a crosslink could be used to determine the position of the crosslink.³ In this work we have extended this to ubiquitin crosslinked with a series of homobifunctional disuccinimidyl ester crosslinking reagents with different crosslinker arm lengths. Analysis of the MS/MS spectra confirmed the presence of crosslinks K1-K6 and K6-K11 for DSS, DSG and DST. The K48-K63 crosslink was observed only for DSS and DSG, but not for the shortest crosslinker arm reagent, DST. The observation of crosslink related peaks for a series of reagents increases the confidence in the assignments, and provides improved distance constraints. In principle the inability of DST to crosslink K48 to K63 could provide a lower bound to the distance between these residues. However, molecular dynamics simulations and further experiments are required to determine reliable estimates for lower bounds to distances when using a series of crosslinkers. We also showed that front-end MS/MS in a hybrid Q-CID FTMS instrument was extremely useful in obtaining extensive fragmentation to allow assignment of the crosslink positions.

An important caveat is that the absence of a crosslink can only be used if the crosslink was observed with other related crosslinking reagents with longer arm lengths. If no crosslink is observed between two reactive side-

chains, this does not necessarily mean they are too far apart to crosslink. They may be on opposite sides of the protein three-dimensional structure so that the crosslinker cannot reach both sidechains without excessive steric hindrance. Even though lysines are generally considered to be polar, solvent accessible, and reactive side-chains, in other work we showed that there is a substantial range in the reactivities of the primary amino groups in ubiquitin.¹⁸ In this work, we showed that the first crosslink in ubiquitin forms between the most reactive primary amino acid residues. We plan to isolate and fragment ubiquitin ions containing more than a single internal crosslink to determine if crosslinks to less reactive residues occur after the most reactive residues have been crosslinked.

The flexibility of the Q-CID FTMS instrumentation opens many experimental pathways that we plan to investigate. One example would be using in-source CID to produce an abundant crosslink containing fragment, and then isolating that ion in the quadrupole and fragmenting it in the front end collision cell. The fragments could then be further probed in the FTMS cell using high-resolution isolation and fragmentation with collisions, IRMPD or ECD. Such CID-Q-CID-FTMS and CID-Q-CID-FTMS/MS experiments should greatly enhance our ability to unambiguously assign crosslink positions. We are currently working on combining these instrumental developments with new crosslinking chemistry to obtain a larger number of crosslinks in ubiquitin, and to extend the top down method for crosslink determination to larger proteins.

Acknowledgements

We gratefully acknowledge the instrument time provided to us by Paul Speir at Bruker Daltonics.

References

1. M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson G. Dollinger, "High Throughput Protein Fold Identification by Using Experimental Constraints Derived from Intramolecular Cross-Links and Mass Spectrometry", *Proc. Natl. Acad. Sci.*, 97, 5802, (2000).
2. B. Schilling, Row, R. H., Gibson, B. W., Guo, X., Young, M. M., "Ms2assign, Automated Assignment and Nomenclature of Tandem Mass Spectra of Chemically Crosslinked Peptides." *Journal of the American Society for Mass Spectrometry*, In Press, Available on-line, (2003).
3. G. H. Kruppa, J. Schoeniger M. M. Young, "A Top Down Approach to Protein Structural Studies Using Chemical Cross-Linking and Fourier Transform Mass Spectrometry", *Rapid Comm. Mass Spectrom.*, 17, 155, (2003).
4. N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson F. W. McLafferty, "Top Down Versus Bottom up Protein Characterization by Tandem High-Resolution Mass Spectrometry", *J. Am. Chem. Soc.*, 121, 806, (1999).
5. E. K. Fridriksson, A. Beavil, D. Holowka, H. J. Gould, B. Baird F. W. McLafferty, "Heterogeneous Glycosylation of Immunoglobulin E Constructs Characterized by Top-Down High-Resolution 2-D Mass Spectrometry", *Biochemistry*, 39, 3369, (2000).
6. F. W. McLafferty, "Tandem Mass Spectrometric Analysis of Complex Biological Mixtures", *Int. J. Mass Spectrom.*, 212, 81, (2001).
7. S. K. Sze, Y. Ge, H. Oh F. W. McLafferty, "Top-Down Mass Spectrometry of a 29-Kda Protein for Characterization of Any Posttranslational Modification to within One Residue", *Proc. Natl. Acad. Sci.*, 99, 1774, (2002).
8. Y. Ge, B. G. Lawhorn, M. ElNaggar, E. Strauss, J. H. Park, T. P. Begley F. W. McLafferty, "Top Down Characterization of Larger Proteins (45 Kda) by Electron Capture Dissociation Mass Spectrometry", *J. Am. Chem. Soc.*, 124, 672, (2002).
9. J. Rappsilber, S. Siniossoglou, E. C. Hurt M. Mann, "A Generic Strategy to Analyze the Spatial Organization of Multi-Protein Complexes by Cross-Linking and Mass Spectrometry", *Anal. Chem.*, 72, 267, (2000).
10. D. R. Muller, P. Schindler, H. Towbin, U. Wirth, H. Voshol, S. Hoving M. O. Steinmetz, "Isotope Tagged Cross Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis", *Anal. Chem.*, 73, 1927, (2001).
11. X. H. Chen, Y. H. Chen V. E. Anderson, "Protein Cross-Links: Universal Isolation and Characterization by Isotopic Derivatization and Electrospray Ionization Mass Spectrometry", *Anal. Biochem.*, 273, 192, (1999).
12. K. M. Pearson, L. K. Pannell H. M. Fales, "Intramolecular Cross-Linking Experiments on Cytochrome C and Ribonuclease a Using an Isotope Multiplet Method", *Rapid Commun. Mass Spectrom.*, 16, 149, (2002).
13. Y. S. Wang, SDH; Hendrickson, CL; Marshall, AG, "Mass-Selective Ion Accumulation and Fragmentation in a Linear Octopole Ion Trap External to a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer", *International Journal of Mass Spectrometry*, 198, 113, (2000).
14. M. E. Belov, E. N. Nikolaev, G. A. Anderson, H. R. Udseth, T. P. Conrads, T. D. Veenstra, C. D. Masselon, M. V. Gorshkov R. D. Smith, "Design and Performance of an Esi Interface for Selective

External Ion Accumulation Coupled to a Fourier Transform Ion Cyclotron Mass Spectrometer", *Anal Chem*, 73, 253, (2001).

15. Z. Q. Zhang A. G. Marshall, "A Universal Algorithm for Fast and Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra", *J. Am. Soc. Mass Spectrom.*, 9, 225, (1998).
16. N. S. Green, E. Reisler K. N. Houk, "Quantitative Evaluation of the Lengths of Homobifunctional Protein Cross-Linking Reagents Used as Molecular Rulers", *Protein Sci.*, 10, 1293, (2001).
17. S. Vijay-Kumar, C. E. Bugg W. J. Cook, "Structure of Ubiquitin Refined at 1.8 a Resolution", *J Mol Biol*, 194, 531, (1987).
18. P. Novak, G. H. Kruppa, M. M. Young, J. Schoeniger, "A Top Down Method for the Determination of Residue Specific Accessibility in Proteins", *Journal of Mass Spectrometry*, Accepted.

Tables

Table 3.1. Selected List of Fragment Ions Used to Assign Crosslink Position.

Fragment	Precursor	Ubiquitin	Ubiquitin + DST	Ubiquitin+DSG	Ubiquitin+ DSS
b7-H ₂ O	(M+H) ⁺ _{obs.}	830.4611	944.4555	926.4814	N.O.
	Δ (ppm)	1.4	0.3	0.3	
b7	(M+H) ⁺ _{obs.}	848.4706	962.4658	944.4915	986.5393
	Δ (ppm)	0.1	0.3	0.1	0.7
b9-H ₂ O	(M+H) ⁺ _{obs.}	1044.5907	1158.5858	N.O.	N.O.
	Δ (ppm)	0.9	1.1	-	-
b9	(M+H) ⁺ _{obs.}	1062.6023	1176.5983	N.O.	N.O.
	Δ (ppm)	0.1	0.6		
b12	(M+H) ⁺ _{obs.}	1348.7683	1462.7547	1444.7855	1486.8351
	Δ (ppm)	1.4	4.8	1.4	0.4
4-13	(M+H) ⁺ _{obs.}	1089.6667	1203.6652	1185.6884	1227.7372
	Δ (ppm)	0.5	2.2	0.1	1.6
6-12	(M+H) ⁺ _{obs.}	730.4454	844.4407	826.4669	868.5146
	Δ (ppm)	1.4	1.2	0.8	0.1
6-16	(M+H) ⁺ _{obs.}	1186.7031	1300.7034	1282.7267	N.O.
	Δ (ppm)	1.3	2.6	0.7	-
19-30	(M+H) ⁺ _{obs.}	1278.7050	N.O.	N.O.	N.O.
	Δ (ppm)	0.7	-	-	-
19-36	(M+H) ⁺ _{obs.}	1967.0451	N.O.	N.O.	N.O.
	Δ (ppm)	1.3	-	-	-
y37	(M+H) ⁺ _{obs.}	4253.3271	N.O.	4349.3457	4391.3930
	Δ (ppm)	4.1	-	0.6	0.5
y40	(M+H) ⁺ _{obs.}	4562.4614	N.O.	4658.4814	4700.5322
	Δ (ppm)	0.4	-	0.1	1.0
y58	(M+H) ⁺ _{obs.}	6528.5117	N.O.	6624.5397	6656.5593
	Δ (ppm)	2.2	-	3.2	0.9
b52	(M+H) ⁺ _{obs.}	5835.1396	5949.1142	5931.1333	5973.2239
	Δ (ppm)	2.3	1.8	2.2	5.0

Table 3.2. Summary of the Observed Crosslinks and Correlation with the Crystal Structure.

Crosslink	DSS	DSG	DST	N ϵ -N ϵ Constraint ^a	C α -C α Constraint ^b	N ϵ -N ϵ Crystalc	C α -C α Crystal ^c
M1-K6	YES	YES	YES	d < 5.8 Å	d < 13.8 Å	21.0 Å	16.5 Å
K6-K11	YES	YES	YES	d < 5.8	d < 18.4 Å	14 Å	6.0 Å
K48-K63	YES	YES	NO	5.8 Å < d < 7.5 Å	18.4 Å < d < 20.1 Å	19.9 Å	17.9 Å

- a) Ref. 15.
- b) See text for a discussion of how the C α -C α constraints were calculated
- c) The distances were measured using the coordinates in the Protein Data Bank, from the original data deposited from Ref. 15.

Figures

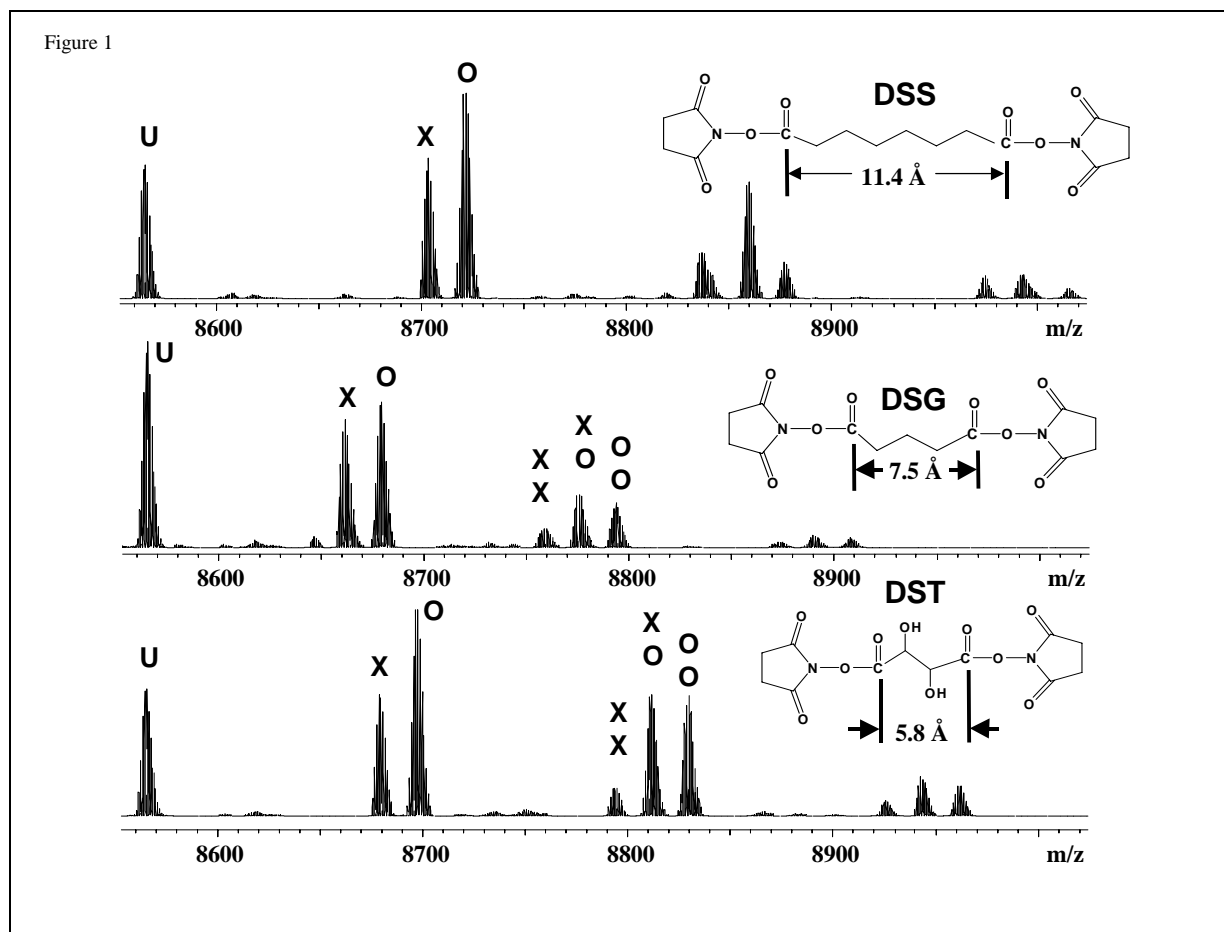


Figure 3.1: The deconvoluted ESI FTMS spectrum of ubiquitin after reaction with DSS, DSG and DST. X indicates a peak corresponding to ubiquitin with an internal crosslink, O indicates a peak with a deadend crosslinker modification, and multiple modifications are indicated by X's and O's stacked above each other.

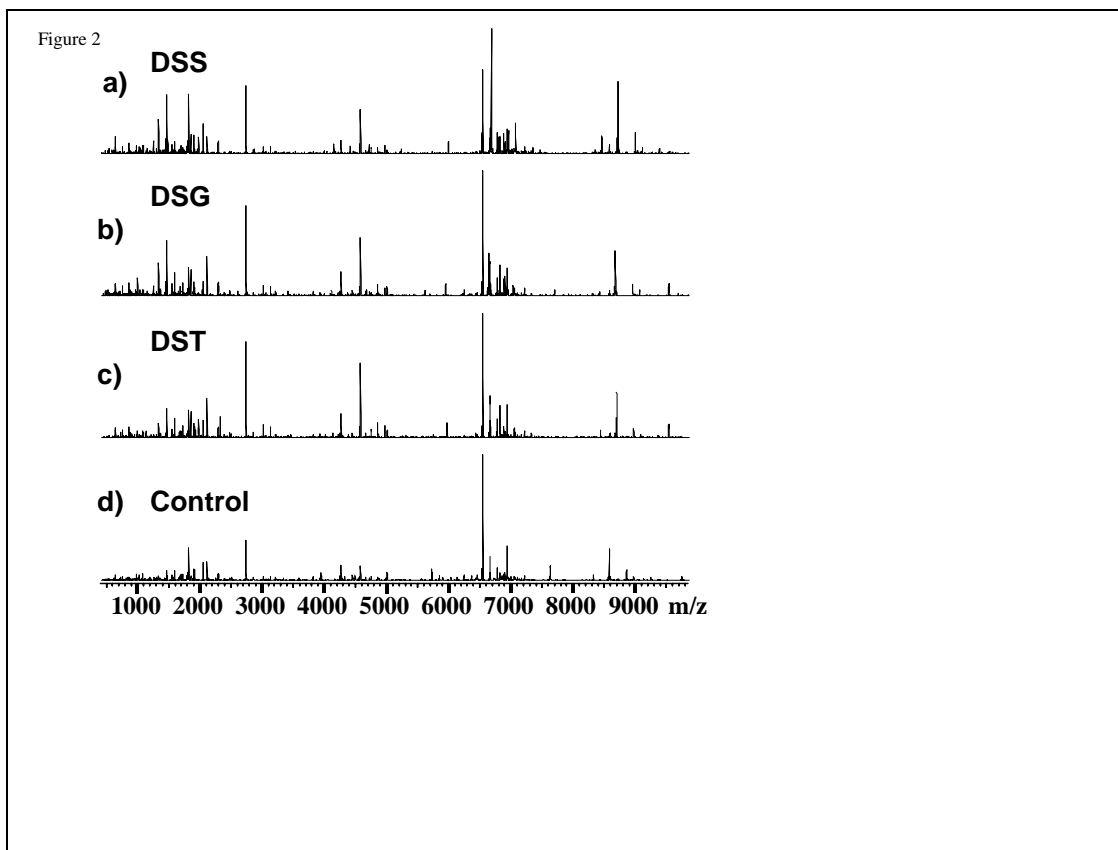


Figure 3.2: An overview of the Q-CID FTMS MS/MS spectra of +10 charge states of ubiquitin with a single internal DSS (a), DSG (b) and DST (c) crosslink, and the +10 charge state of unmodified ubiquitin (d).

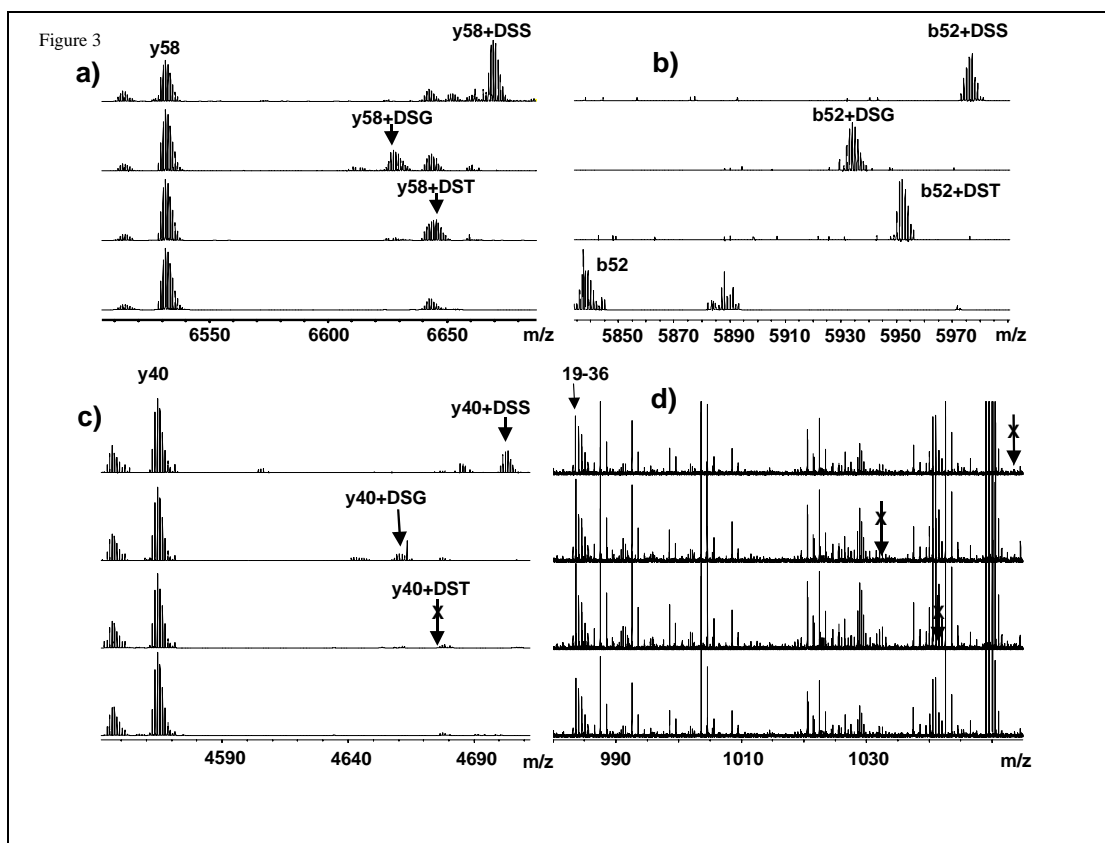


Figure 3.3: Expansions of the MS/MS spectra shown in Figure 3.2. Note that 19-36 denotes the internal sequence ion starting at position 19 and ending at position 36. In 3a-3d, the MS/MS spectra from top to bottom are the ubiquitin ion modified with a single internal DSS crosslink, a single DSG crosslink, a single DST crosslink, and unmodified ubiquitin, all isolated for MS/MS in the +10 charge state.

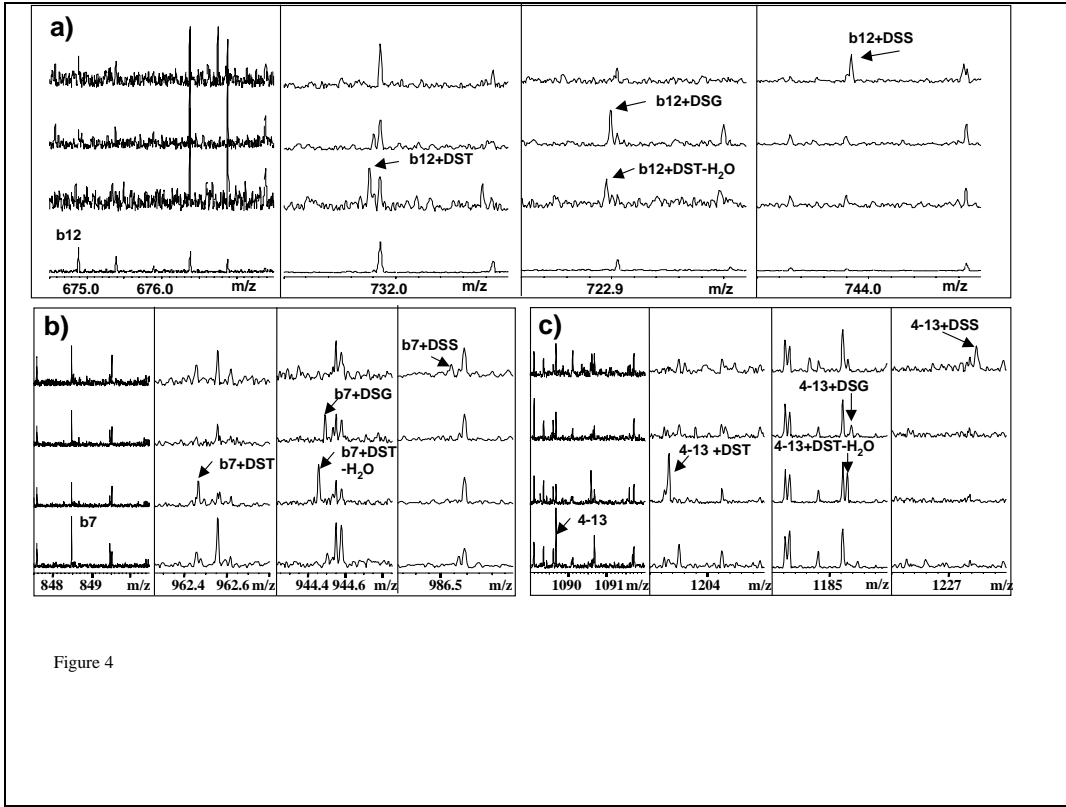


Figure 3.4: Expansions showing an example of peaks used to localize crosslinks near the amino terminus. In 4a-4d, the MS/MS spectra from top to bottom are from the ubiquitin ion modified with a single internal DSS crosslink, a single DSG crosslink, a single DST crosslink, and unmodified ubiquitin, all isolated for MS/MS in the +10 charge state.

Figure 5

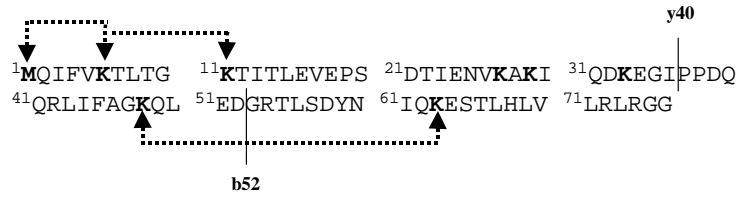


Figure 3.5: The annotated sequence of ubiquitin. Crosslinks are noted by dotted lines with arrows connecting the crosslinked positions on the sequence.

Figure 6

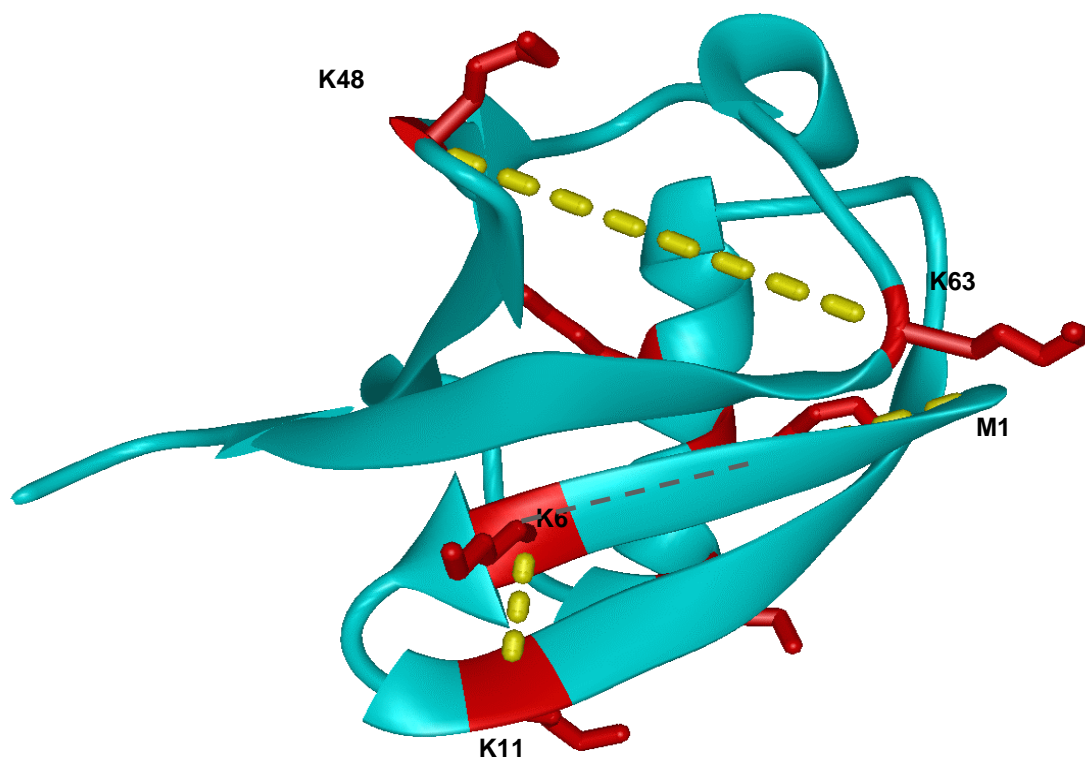


Figure 3.6: The structure of ubiquitin with the observed crosslinks indicated. Note that the M1-K6 crosslink is partially obscured by the structure.

This page intentionally left blank.

Section II: Computational Modeling of Membrane Protein Structure and Dynamics

This page intentionally left blank.

Chapter 4: Exploring the conformational space of membrane protein folds matching distance constraints

Jean-Loup Faulon, Ken Sale, Malin Young

Abstract

Herein we present a computational technique for generating helix-membrane protein folds matching a predefined set of distance constraints, such as those obtained from NMR NOE, chemical cross-linking, dipolar EPR and FRET experiments. The purpose of the technique is to provide initial structures for local conformational searches based on either energetic considerations or ad-hoc scoring criteria. In order to properly screen the conformational space, the technique generates an exhaustive list of conformations within a specified root-mean-square deviation (RMSD) where the helices are positioned in order to match the provided distances. Our results indicate that the number of structures decreases exponentially as the number of distances increases, and increases exponentially as the errors associated with the distances increases. We also find the number of solutions to be smaller when all the distances share one helix in common compared to the case where the distances connect helices in a daisy chain manner. We find that for 7 helices, at least 15 distances with errors up to 8 Å are needed to produce a number of solutions that is not too large to be processed by local search refinement procedures. Finally, without energetic considerations, our enumeration technique retrieved the transmembrane domains of Bacteriorhodopsin (PDB entry 1c3w), Halorhodopsin (1e12), Rhodopsin (1f88), Aquaporin-1 (1fqy), Glycerol uptake facilitator protein (1fx8), Sensory Rhodopsin (1jgj), and a subunit of Fumarate reductase flavoprotein (1qlaC) with $C\alpha$ level RMSDs of 3.0 Å, 2.3 Å, 3.2 Å, 4.6 Å, 6.0 Å, 3.7 Å and 4.4 Å, respectively.

Introduction

Unlike soluble proteins, only a few membrane protein structures have been solved using conventional methods such as NMR and crystallography. Considering the pace at which membrane protein structures are elucidated experimentally, membrane protein fold recognition based on structural homology does not appear to be a practical option for the near future.

Alternative membrane protein structural modeling approaches take advantage of the fact that many membrane proteins contain regions of highly hydrophobic transmembrane helical fragments (e.g., helix bundle). The structural constraints placed on transmembrane helices by the lipid bilayer limits the number of possible membrane protein folds to a point that several ab-initio computational techniques have been considered (Bowie 1999; Nikiforovich et al. 2001; Vaidehi et al. 2002). These techniques decompose the membrane protein folding problem into the following steps: (1) prediction of transmembrane regions, (2) construction and optimization of individual helices, (3) assembly of the helix bundle, and (4) addition of the interhelical loops.

Several codes have been developed to predict the transmembrane regions of step (1) (Hirokawa et al. 1998; Gromiha 1999; Nikiforovich et al. 2001; Vaidehi et al. 2002). The regions are generally determined using hydrophobicity analysis. Step (2) requires energy minimization or molecular dynamics simulations (Vaidehi et al. 2002) to predict sequence-specific distortions in the helices, such as kinks induced by proline. While step (3) is the main subject of this paper, the approaches that have so far been taken to attack the problem of assembling helix bundles are different in nature. Bowie (Bowie 1999) argues that the conformational space of a membrane protein can effectively be sampled and gives a truly ab-initio technique where all possible helix bundles are enumerated. However, in Bowie's calculations the orientations of the individual helices around their respective axes are not taken into account. Nikiforovich et al. (Nikiforovich et al. 2001) use the similarity between the X-ray structures of bacteriorhodopsin and rhodopsin to find the helix packing in the membrane plane. Specifically, the intersections between the helical axes and the plane crossing the membrane are fixed at values derived from the two X-ray structures. Vaidehi et al. (Vaidehi et al. 2002) orient each helical axis of the helix bundle according to the 7.5 Å electron density map of rhodopsin. Although the two previous methods use energetic calculations and molecular simulations to further refine the helical arrangements, both techniques are potentially biased toward the structures of bacteriorhodopsin and rhodopsin and have yet to be validated for other membrane proteins. The addition of loops in step (4) can be performed using commercially available software such as WHATIF (Vriend 1990) and SCWRL (Bowers et al. 1997) and Jackal (Xiang et al. 2002), while other solutions that segregate between small and medium loops have also been proposed (Nikiforovich et al. 2001).

The technique presented in this paper is concerned with helix bundle assembly (step 3 above). We assume that transmembrane regions have been predicted and that individual helices have been built and optimized. The method takes as input a set of helices in PDB format and a set of distances between pairs of atoms on these helices. The technique outputs all possible helix arrangements matching the input data, each solution differing from another by a predefined RMSD. The arrangements that are produced are constructed at an atomistic level and can thus be further refined using local conformational search.

Distance constraints are needed since, as we will see in the next section, the conformation space for membrane proteins is too large to perform energetic calculations on each conformation. These distances can come from a variety of experiments including chemical cross-linking, dipolar electron paramagnetic resonance (dipolar EPR), fluorescence resonance energy transfer (FRET), and NMR to name a few. While each of these methods has advantages and disadvantages in terms of high throughput capability, distance measurement accuracy and structural model building, the purpose here is to show the utility of a set of distance constraints for membrane protein structural modeling. For validation purposes on a set of membrane proteins for which crystal structures are available, the distances we consider in this paper correspond to pairs of amino acids (K-K, K-D, K-E, K-C and C-C) that could potentially be cross-linked using chemical cross-linkers. However, we stress that the method is not limited to the consideration of distances derived from cross-linking experiments and also demonstrate the

method on the dark-adapted rhodopsin structure (1f88) using a set of disulfide mapping, cross-linking and dipolar EPR distances gathered from the literature.

Essential to the understanding of our method is the concept of a distance graph, which is used on several occasions in the paper and is illustrated in Figure 4.1. The distance graph for a given set of helices and inter-helical distances is derived by representing each helix as a vertex and each distance as an edge connecting the two vertices corresponding to the helices between which the distance was measured: for each distance between two helices an edge is added to the graph between the corresponding vertices. Associated with a distance graph is its radius. The radius of the distance graph is the minimum number of consecutive edges one has to follow to reach all the vertices of the graph.

Throughout the paper we also use a specific terminology to characterize increasing levels of membrane protein fold complexity. A *template* is a set of helices fully positioned in the membrane. Precisely, the helices are positioned in a reference system with origin in the bilayer central plane, and with *z-axis* pointing toward the extracellular side of the membrane and parallel to the bilayer normal. Since we do not attempt to position our helices toward other components of the membrane such as lipids, the origin is chosen arbitrarily in the bilayer central plane, and the *x-* and *y-axes* are any pair of orthogonal vectors in that plane. We refer to an *unlabeled* template when the sequence order of the helices is not known. In the unlabeled case, the coordinates of the helix center of mass in the bilayer reference system and the coordinates of the helix axis unit vector are used to represent each helix. A template is said to be labeled when the order of the helices is known, i.e, the helix numbered *i* in the template corresponds to the *ith* membrane helix in the protein sequence. Each helix in a *labeled* template is represented by its sequence number and the coordinates of both its center of mass and its unit axial vector. An *oriented template* is a labeled template where the helices have been oriented around their helical axes. A helix in an oriented template is represented as a labeled template to which the coordinates of any backbone atom have been added. Finally, an *atomistic* template is a template composed of the coordinates in the bilayer reference system of all atoms of the helices. If one assumes that backbone atoms in helices are fixed, then there is only one atomistic template per oriented template, since if one knows the position of the center of mass, the axis vector, and the position of any backbone atom of a helix, then one knows the position of all backbone atoms of that helix. Consequently we will not make a distinction between oriented templates and atomistic templates.

Results and discussion

In the following three subsections we probe the number of membrane protein templates generated using our technique with various constraints. In the fourth subsection we examine the ability of our technique to retrieve known membrane protein structures.

Conformational space size without distance constraints

Bowie (Bowie 1999) argues that, for membrane proteins of up to seven helices, unlabeled templates can be generated to cover the entire conformational space of the membrane protein folds. Precisely, he observes that conformational space is at least 80% covered by no more than 10 templates for 3 helices, 250 templates for 4 helices, 2,500 templates for 5 helices, 25,000 templates for 6 helices, and 150,000 templates for seven helices. Now, for *n* helices there are $2n!$ ways to label the helices, which gives 10,080 labelings per unlabeled template for seven helices; however, this number can drastically be reduced by adding inter-helix connection constraints (Bowie 1999). These constraints specify that (1) consecutive helices must be in contact, (2) the distance between the end points of consecutive helices must be in the range observed in known membrane protein structures, and (3) loops connecting helix end points cannot cross each other. Figure 4.2 gives the distribution of the number of labelings over a set of 50,000 unlabeled templates composed of seven helices. This distribution was obtained by running an algorithm described in the Method section. The average number of labelings per template is 27.25 with a large standard deviation of 26.07. Constraint number (3) is not valid for all known membrane proteins. In fact, the structure of aquaporin-1 (1fqy) provides an example in which the loop between helices 3 and 5 and the loop between helices 5 and 6 may cross. Removing this constraint only slightly increases the number of labelings per template to 33.67. These numbers are comparable to the number of labelings, 30, reported by Bowie (Bowie

1999), and a library of 150,000 unlabeled structures gives about 4.5×10^6 labeled templates. Recall that a labeled template is oriented once an arbitrary atom has been positioned for each helix in the bilayer reference system. If one imposes two orientations separated by at least $\Delta\alpha$, then for n helices there are $(360/\Delta\alpha)^n$ possible orientations per labeled template. For ideal helices, e.g. poly-alanine helices, where the alpha-carbons are at a distance less than 2.5 Å from their helix axis, $\Delta\alpha = 60^\circ$ leads to a RMSD no greater than 2.5 Å between two consecutive orientations. Thus, using $\Delta\alpha = 60^\circ$, a labeled library of 4.5×10^6 labeled templates with seven helices will give 78,125 possible orientations per template, and a total of 351.5×10^9 oriented or atomistic templates. Even if these templates are clustered with an RMSD of 2.5 Å, which according to Bowie leads to three times fewer structures (Bowie 1999), the number is still too large to perform energetic calculations on each template.

Conformational space size with exact distance constraints

In this subsection, we probe the number of oriented templates when exact inter-helix distances are provided. We assume we have generated all possible labeled templates and we are given a set of exact distances between backbone atoms of the helices. All backbone atoms are fully positioned relative to their respective helices and all helices belong to a labeled template. Since each helix can freely rotate around its axis, each atom lies on a circle perpendicular to its helix axis (cf. Figure 4.3). Now let us assume some helix H_i has already been positioned in the bilayer reference system and also assume that a distance d_{ij} is given between an atom of H_i and an atom of a second helix H_j . Given these assumptions, there are no more than two possible positions for H_j . In fact, as illustrated in Figure 4.3, finding the position(s) of helix H_j using distance d_{ij} from helix H_i that has already been positioned, leads to either no position (A), one position (B), or two positions (C), depending on the number of points obtained when intersecting a circle with a sphere. Now that helix H_j has been oriented, positioning helix H_k from H_j using distance d_{jk} should lead to no more than two positions for each H_j position, and consequently to no more than four positions from H_i . While according to the previous subsection we have $360/\Delta\alpha - 1$ possible positions for H_i , if all $n-1$ helices are positioned from H_i , the final number of oriented templates is bounded by $2^{n-1}(360/\Delta\alpha - 1)$. This result is valid as long as there is no cycle in the distance graph. Indeed, as illustrated in Figure 4.4, when a cycle is introduced only one position remains for all helices belonging to the cycle. Although we do not rigorously prove the above statement, in all the tests carried out for this paper we observed that only one position per helix permits formation of a cycle. Let n_0 be the number of helices that are not linked to any other helix through a distance and let n_2 be the number of helices that are in at least one cycle in the distance graph. According to the above discussion the total number of orientations of a given labeled template is at most $2^{n-n_0-n_2}(360/\Delta\alpha - 1)^{n_0}$. Note that this number ranges from one when $n_2 = n$, i.e., when the distance graph is a cyclic graph, to $2^6(360/\Delta\alpha - 1) = 320$ for seven helices and $\Delta\alpha = 60^\circ$ when the distance graph is connected but contains no cycles.

Conformational space size with experimental distance constraints

Experimental distances are never exact. For instance NOE NMR distances are generally reported in bins of 2 Å and thus have an associated error of at least 1 Å. Cross-link distances have errors depending on the length and the flexibility of the cross-linker and the crosslinked amino acid sidechains. For instance, the homobifunctional lysine-specific crosslinker BS³ (Pierce Biotechnology, Inc.) yields C α -C α distance information ranging from 5-24 Å (Young et al. 2000). Dipolar EPR distance measurements, ignoring interspin orientation, have errors of approximately 5 Å (Rabenstein and Shin 1995). Intuitively, for a given set of distances, as error increases so should the number of oriented templates matching the provided data. Less intuitive is the fact that the number of oriented templates also increases with the radius of the distance graph. In order to simplify our discussion, let us assume as illustrated in Figure 4.5 that we have 4 points P_1 through P_4 embedded in a one-dimensional space. In the first case (A), distances d_{12} , d_{13} , and d_{14} are provided, each having an error of one arbitrary unit. Note that since each point is linked to P_1 , the radius of the distance graph is one. We are interested in counting the number of possible configurations that are separated by one unit. Clearly, each point P_i , $i=2,3,4$ has three possible positions for which each configuration is separated by one unit: P_{i-1} , P_i , and P_{i+1} , where $P_i = P_1 + d_{1i}$. The total number of configurations is thus 27. In case (B), the provided distances are d_{12} , d_{23} , and d_{14} and the radius of the distance graph is two. For the reason mentioned above, points P_2 and P_4 have three possible configurations, and

each P_2 position leads to three possible positions for P_3 : $P_2+d_{23}-1$, P_2+d_{23} and $P_2+d_{23}+1$. Now, the three positions for P_2 are $P_1+d_{12}-1$, P_1+d_{12} , and $P_1+d_{12}+1$, and these lead to five final positions for P_3 relative to P_1 : $P_1+d_{12}+d_{23}-2$, $P_1+d_{12}+d_{23}-1$, $P_1+d_{12}+d_{23}$, $P_1+d_{12}+d_{23}+1$, and $P_1+d_{12}+d_{23}+2$. Consequently the total number of configurations for case (B) is 45.

The algorithm used for enumerating oriented templates matching a set of distances with errors is detailed in the Method section. This algorithm was used to probe changes in the numbers of unlabeled, labeled, and oriented templates with increasing number of distances, increasing error, and increasing radius of the distance graph. All runs were performed on a library of 150,000 unlabeled templates composed of seven helices generated using Bowie's code, which is also briefly described in the Method section. The distances were compiled for rhodopsin (PDB entry 1f88) and are listed in Figure 4.1. All labeled templates matching the distances were clustered with a clustering algorithm given in the Methods section using clustering RMSD values of 4 Å and 6 Å. Oriented templates were not clustered; however, the number of oriented templates matching the provided distances was computed on the clustered labeled templates.

Table 4.1 and Figure 4.6 report the number of templates with number of distances ranging from 7 to 38. The 38 distances represent potential chemically cross-linkable amino acid pairs (K-K, K-D, K-E, K-C, C-C) and are listed in Figure 4.1. From this set, distances were removed at random and an error of ± 4 Å was attributed to each residual distance. It is clear from Figure 4.6 that the number of unlabeled, labeled, and oriented templates decreases exponentially as the number of distances increases. Specifically, the slopes in Figure 4.6 are -0.10 for unlabeled templates, -0.13 for labeled templates, and -0.22 for oriented templates. For Nd distances, the number of unlabeled templates scales as $150,000 \times 0.78^{Nd}$, the number of labeled templates scales as $4.5 \times 10^6 \times 0.73^{Nd}$, and the number of oriented template scales as $351.5 \times 10^9 \times 0.59^{Nd}$. Hence, the number of oriented templates decreases slightly faster than the other template types as the number of distances increases.

As mentioned in the introduction, the purpose of our enumeration algorithm is to provide starting structures for further refinement with energetic calculations and molecular simulations. Local conformational search is a well-established technique in computational chemistry and has long been used for soluble protein fold recognition (Godzik et al. 1992). More recently, local search techniques have been proposed that are specific for membrane proteins. Nikiforovich *et al.* (Nikiforovich et al. 2001) propose a technique for searching the energetically favored orientation of pair of helices. Provided the two helices have been positioned in the bilayer reference system, the method exhaustively enumerates all possible helix rotations with an angular increment of 30° . Energy calculations are performed using the ECEPPP/2 force fields (Dunfield et al. 1978) and low energy conformations are retained. Using this method Nikiforovich *et al.* generated low energy conformations for bacteriorhodopsin that are 3.13 Å RMSD from the crystal structure. Vaidehi et al. (Vaidehi et al. 2002) report a molecular dynamics code, COARSEROT, which performs coarse-grain rotations of the helical orientations around their axis through a grid of rotation angles. Using this technique the authors retrieved the seven-helix bundles of bacteriorhodopsin and bovine rhodopsin with RMSDs of 2.9 Å and 3.1 Å, respectively.

Now, for each labeled template, in order to fully explore the conformational space of that template, one has to rotate all the helices around their axes and compute energetic parameters for each orientation. As detailed in the Method section, helix orientations in an oriented template are enumerated using an angular increment of $\Delta\alpha = 60^\circ$, thus, the local conformational space of each oriented template can fully be explored by rotating each helix with an angle no greater than 60° . Consequently, the refinement procedures mentioned above should be able to process $(360/60)^7 = 279,936$ oriented templates of seven helices in the same computational time it takes to run one labeled template. Using the latter number as a threshold (Table 4.1) for systems of seven helices at least 15 distances with an error of 4 Å are needed to be able to explore the conformational space in a reasonable computational time.

Table 4.2 and Figure 4.7 give the number of templates for the 38 distances listed in Figure 4.1 with increasing distance error. All template numbers increase exponentially with the distance error. The slopes computed from Table 4.2 are 0.62 for unlabeled and labeled templates and 0.75 for oriented templates. The scaling behavior for unlabeled and labeled templates is $0.25 \times 4.2^\epsilon$, where ϵ is the error, and $0.35 \times 5.6^\epsilon$ for oriented templates.

Interestingly, while the slopes are more pronounced for decreasing error than increasing number of distances (-0.62 vs. -0.10 for unlabeled templates) the number of templates is not as large as one might expect, and up to a fairly big error of ± 8 Å, the number of oriented templates is small enough to be manageable by local conformational refinements.

The number of templates with increasing distance graph radius is given in Table 4.3. Although not as pronounced as with the distance error, the number of templates increases exponentially with the radius. The scaling computed on Table 4.3 is 500×2.9^r , where r is the radius, for unlabeled and labeled templates, and 2500×4.5^r for oriented templates. As a consequence of this finding, it is preferable to design experiments that will compute distances from the same reference helix rather than generating distances linking helices in a daisy chain manner.

Specific membrane proteins

In order to validate the value of the proposed method for generating structures suitable for energetic refinement, we considered seven known crystal structures with varying number of transmembrane helices and attempted to predict the positions of the transmembrane helices. Specifically, only those portions of the transmembrane helices completely embedded in the membrane were considered (e.g., the residue 109 to residue 120 helix of 1fx8 was not considered). For each structure we derived a set of distance constraints corresponding to pairs of amino acids (K-K, K-D, K-E, K-C and C-C) that could potentially be cross-linked using chemical cross-linkers. Specifically, we considered the following cases (PDB identifier and number of distance constraints in parentheses): Bacteriorhodopsin (1c3w, 60), Halorhodopsin (1e12, 9), Rhodopsin (1f88, 38), Aquaporin-1 (1fqy, 17), Glycerol uptake facilitator protein (1fx8, 43), Sensory Rhodopsin (1jgj, 18), and a subunit of Fumarate reductase flavoprotein (1qlaC, 58). For all cases the error assigned to each distance constraint was ± 4 Å and library sizes of 300,000 unlabeled templates for 7 helices (twice the size used earlier), 150,000 for 6 helices, and 50,000 for 5 helices were searched.

Table 4.4 reports the results for these seven cases in terms of number of distance constraints, number of solutions and minimum RMSD among the computed structures and the structure from the PDB across all C α atoms (C α – RMSD) for each case (RMSD calculated using MMTSB Tool Set (Feig et al. 2001)). In general, the predicted positions of the helices in the helix assemblies were in good agreement with the PDB structures, having C α – RMSD values ranging from 2.3 Å to 6.0 Å for the seven test cases. As discussed above, the number of solutions declines exponentially with the number of distances. For our test cases the number of solutions is heavily dependent on the number of distance constraints ranging from 8,274,078 for the structure having only 9 distances to a minimum of 26 for the case with 58 constraints. Since our method involves searching a dataset of potential structures for structures that satisfy a set of distance constraints, we did not expect the quality of the best structures to depend on the number of distance constraints. Table 4.4 supports this expectation with very little correlation between RMSD and number of constraints. This lack of correlation is also due in part to the fact that the quality of the computed structure is a function of the number of distance constraints, the error on each distance and the combination of sites among which the distances were measured. Our stated goal is to provide starting structures for further refinement using either energetic constraints (force field methods) or ad-hoc scoring functions based on *Bioinformatics* data. To this end these results clearly show that, given a set of experimental distance constraints, our method generates sets of helical positions close to those of the known structure, providing a set of candidate structures from which further structural refinements can be performed.

As a second validation we then predicted the positions of the seven transmembrane helices of dark-adapted Rhodopsin (1f88) using a set of distance constraints compiled from various experiments reported in the literature (Yeagle et al. 2001). These included dipolar EPR distances (Farrens et al. 1996; Yang et al. 1996; Albert et al. 1997; Galasco et al. 2000), disulfide mapping distances (Yu et al. 1995; Sheikh et al. 1996; Cai et al. 1997; Cai et al. 1999; Yu et al. 1999) and distances from electron cryo-microscopy (Unger and Schertler 1995). The distance constraints are shown in Table 4.5. The average distance error is ± 3.75 Å.

Since the published EPR dipolar distances are between nitroxide spin labels, they do not directly correspond to distances between helical axes. To better represent the dipolar EPR distances, we used a Monte-Carlo

minimization based method (Sale et al. 2002) to uncover the most probable orientations of the spin label at each of the labeled sites V139, K248, V249, V250, T251 and R252. From these orientations the mean length of the $C\alpha$ to nitroxide N vector for the six sites was approximately 9 Å. This length was used to adjust the lower and upper limits of the reported distances in order to better represent the inter-nitroxide distances as helix backbone distances. For the disulfide mapping distances we used a $C\alpha$ to $C\alpha$ distance of 5.68 Å, which corresponds to two $C\beta$ to $S\gamma$ bonds (1.82 Å) and one $S\gamma$ to $S\gamma$ bond (2.04 Å), plus and minus the reported error.

The results are shown in the last row of Table 4.4. Using a set of 27 distance constraints with differing amounts of error, only 87 helical bundles of the possible $300,000 \times 30 \times 78,125 = 0.7 \times 10^{12}$ seven-helix configurations were found that matched the set of distances. This yielded structures with $C\alpha$ - RMSD values ranging from 4.3 Å to 9.5 Å. These results show that even in the case of using only 27 distance constraints taken from a variety of experimental methods with differing levels of error that a reasonable number of structures suitable for further refinement can be extracted from a large dataset of possible helix bundles.

Conclusions

We have demonstrated a method for generating helical bundles (oriented templates) satisfying a given set of inter-helical distance constraints. Using the structure of rhodopsin (1f88) as a model, we showed (1) the number of templates decreases exponentially as the number of distance constraints increases, (2) the number of templates increases exponentially with increasing error on the distances, (3) the number of templates increase exponentially with increasing radius of the distance graph and (4) experiments designed to measure distances from the same reference helix are preferable to those designed generate distances linking helices in a daisy chain manner. We also note that the slopes derived from Tables 4.1 and 4.2 indicate that the number of solution templates decreases faster with decreasing error on the distance constraints than with increasing number of distances increases. This indicates that in order to cut down the number of solutions one may want to lower the error of a given experiment rather than adding new data points.

Conclusion (4) is particularly important to the design of experiments in which distances are derived from the interaction of external probe molecules incorporated at specific sites such as in FRET and dipolar EPR experiments. Our findings suggest that fewer measurements from a common helix to the other helices result in a reduced set of possible structures compared to many measurements between distinct pairs of amino acids. The obvious consequence of this finding is the need for fewer experiments involving amino acid mutations and labeling. For potentially higher throughput methods such as chemical cross-linking, this result may not assist in the initial experimental design as cross-linking does not require a directed mutagenesis and labeling strategy. Conclusion (4) may, however, provide a guide for the selection of cross-linking reagents or sites for mutagenesis once a preliminary structural model has been constructed.

Using a set of seven helical bundles with from 5 to 7 helices and varying number of distance constraints as test cases we found that a reasonable number (in terms of starting points for further refinement) of helical bundles with helix positions near the native structure can be generated. The same conclusion can also be drawn for rhodopsin (1f88) using 27 experimental distances taken from the literature. This finding has important consequences in that further refinement using either force field or ad-hoc scoring methods with local conformational search should produce reasonably accurate model structures. Furthermore, given a set of model structures, experiments can then be designed to differentiate among the models and provide data for further structural refinement.

While the technique developed in this paper has been demonstrated on transmembrane proteins with only up to seven helices, there are no limitations on the number of helices that can be processed. Note however that according to Bowie (Bowie 1999) the initial library of unlabeled templates has a size that increases exponentially with the number of helices. Consequently the method presented here is somewhat limited by the ability to generate a library of unlabeled templates comprising more than 7 helices that is representative of the entire

conformational space. One possible extension of this work is to modify Bowie's code and incorporate distance information when generating unlabeled templates.

Another possible extension of the code is to handle not only helices but also beta strands and beta sheets. This extension of our technique is straightforward as beta strands and sheets can be represented as special helices; that is, they too can be modeled with a center of mass and a unit vector. Furthermore, as with helices if one knows the position of one atom of the beta strand or sheet one knows the position of all atoms. An important difference in the beta strand case is the additional constraints imposed on the structure by the inter-strand hydrogen-bonding network. If pairs of adjacent strands can be determined with their respective orientations, the solution space of the corresponding template library is expected to be quite small relative to the library of an all-helical system of equivalent size.

Methods

Generating unlabeled templates

Unlabeled templates were generated using Bowie's software, which is described in detail in a previous paper (Bowie 1999). This software generates random unlabeled templates matching specific criteria determined from known membrane protein structures (Bowie 1997). Briefly, the angle between each helix axis and the bilayer normal is random but smaller than 40° . The distance of closest approach between two neighboring helices is no greater than 13.4 \AA and no smaller than 6 \AA . The angle between the axes of two neighboring helices follows a distribution computed on known membrane proteins and reported by Bowie (Bowie 1997). The output is a library of unlabeled templates each satisfying the above criteria. The user specifies the number of helices and the desired number of templates.

Generating labeled templates

Labeled templates were enumerated for each element of the library of unlabeled templates. Theoretically, for n helices there are $2n!$ labelings of the helices. However, this number can drastically be reduced if one adds the three following interhelix connection constraints: (1) the distance of closest approach between consecutive helices is in the $6\text{-}13.4 \text{ \AA}$ range, (2) the distance between the end points of consecutive helices is in the $7\text{-}22 \text{ \AA}$ range, and (3) there are no cross-over connections. The justifications for the above constraints can be found in (Bowie 1999). Our enumeration algorithm finds all possible labels for each helix sequentially. The above constraints are applied on-the-fly by the algorithm even if the template is not yet fully labeled. Running this algorithm, we observed that most labelings are rejected before completion, and we were able to process libraries comprising several thousand templates quite efficiently in a couple hours of CPU time on an SGI/O2 workstation.

Generating oriented or atomistic templates with distance constraints

Recall that a labeled template is oriented once an arbitrary backbone atom has been positioned in the bilayer reference system for each helix. As discussed, for n helices there are $n^{360/a-1}$ possible orientations per labeled template, each being separated by at least one $\Delta\alpha$ angle. This number can be reduced by introducing distance constraints using the algorithm described next and illustrated in Figures 4.3 and 4.4. In order to minimize the effect of the distance graph radius (cf. Table 4.3 and Results and discussion section), the first step of the algorithm consist of finding the helix that gives the smallest radius in the distance graph. If the graph is not connected (radius is infinite) then the procedure halts, as the helices cannot be oriented due to a lack of distances. Otherwise, the algorithm then tests all possible orientations of the initial helix. Each orientation is incremented by a 10° angle until a solution is found and by a 60° angle if the previous angle led to at least one solution. The 10° angular increment is small enough to avoid missing solutions, while the 60° increment was chosen to provide a 2.5 \AA RMSD between consecutive solutions. For each tested orientation of the initial helix the following steps are applied until there are no more distances to be considered. (1) A helix, H_i , also named the current helix, is chosen such that this helix has already been positioned and is attached through distances to helices that have not yet been

positioned. Note that if no such helix can be found then all helices have been positioned, and the algorithm then verifies that all distance constraints are satisfied and rejects the solution if there are unsatisfied distance constraints. Otherwise, (2) one chooses an unused distance d_{ij} with associated error ε_{ij} between two atoms P_i belonging to the current helix H_i , and P_j belonging to a helix H_j . Since H_i is the current helix, P_i is fully positioned in the bilayer reference system, while as illustrated in Figure 4.3, P_j lies on a circle orthogonal to V_j , the axis of H_j . This circle is centered on $C_j + (C_j P_j \cdot V_j) V_j$, where C_j is the center of mass of H_j , and has a radius $|C_j P_j \times V_j|$ (norm of the product between vectors $C_j P_j$ and V_j). (3) The position of P_j is found by intersecting the circle on which P_j lies and the sphere centered on P_i of radius $d_{ij} \pm \varepsilon_{ij}$. The intersection of a sphere and a circle leads to no solution, one position, or two positions. If no solution can be found the procedure halts and a new position of the initial helix is tested. Otherwise for each P_j position found, helix H_j is rotated clockwise and counterclockwise in 60° angular increments and all new P_j positions that are at distances $d_{ij} \pm \varepsilon_{ij}$ from P_i are retained. Again, 60° was chosen to provide a 2.5 \AA RMSD between consecutive solutions. Finally, for each P_j positions stored, steps (1) through (3) are applied until all distances have been used. Each time all distances have been processed successfully the resulting template is a solution and is added to the library of oriented templates.

Clustering labeled templates

Labeled templates are represented by numbered helices located in the bilayer reference system through the coordinates of their center of mass and unit axial vector. Templates are clustered using a predefined RMSD^{Hel} value computed at the helix level. Precisely, RMSD^{Hel} between two templates is calculated by averaging the RMSD between the centers of mass and the two end points of the helices of the templates. Prior to performing the calculation, the templates are rotated in the bilayer plane in order to align the vectors joining the first two helices. Bowie compared RMSD values for a full atom model with RMSD^{Hel} and found that for 100 structures with RMSD^{Hel} in the range $3.9\text{-}4.1 \text{ \AA}$, the average RMSD of the full atom models was 2.6 \AA for 86% of the atoms. In the present paper, we clustered labeled templates using RMSD^{Hel} of 4 \AA and 6 \AA .

References

1. Albert, A.D., Watts, A., Spooner, P., Grobner, G., Young, J., and Yeagle, P.L. 1997. A solid state NMR characterization of the substrate binding specificity and dynamics for the L-fucose-H⁺ membrane transport protein of E. coli. *Biochim. Biophys. Acta* 1328: 74 - 82.
2. Bowers, M., Cohen, F.E., and Dunbrack, R.L., Jr. 1997. Sidechain prediction from a backbone-dependent rotamer library: A new tool for homology modeling. *Journal of Molecular Biology* 267: 1268-1282.
3. Bowie, J.U. 1997. Helix packing in membrane proteins. *Journal of Molecular Biology* 272: 780-789.
4. Bowie, J.U. 1999. Helix-bundle membrane protein fold templates. *Protein Science* 8: 2711-2719.
5. Cai, K., Klein-Seetharaman, J., Hwa, J., Hubbell, W.L., and Khorana, H.G. 1999. Structure and Function in Rhodopsin. Effects of Disulfide Cross-Links in the Cytoplasmic Face of Rhodopsin on Transducin Activation and Phosphorylation by Rhodopsin Kinase. *Biochemistry* 38: 12893 - 11898.
6. Cai, K., Langen, R., Hubbell, W.L., and Khorana, H.G. 1997. Structure and function in rhodopsin: topology of the C-terminal
7. RT polypeptide chain in relation to the cytoplasmic loops. *Proc. Natl. Acad. Sci. USA* 94: 14267 - 14272.
8. Dunfield, L.G., Burgess, A.W., and Scheraga, H.A. 1978. Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. *J. Phys. Chem.* 82: 2609-2616.
9. 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.
10. Feig, M., Karanicolas, J., and Brooks, C.L.I. 2001. MMTSB Tool Set. MMTSB NIH Research Resource, The Scripps Research Institute.
11. Galasco, A., Crouch, R.K., and Knapp, D.R. 2000. Intrahelic arrangement in the integral membrane protein rhodopsin investigated by site-specific chemical cleavage and mass spectrometry. *Biochemistry* 39: 4907 - 4914.
12. Godzik, A., Kolinski, A., and Skolnick, J. 1992. Topology fingerprint approach to the inverse folding problem. *Journal of Molecular Biology* 227: 227-238.
13. Gromiha, M.M. 1999. A simple method for predicting transmembrane alpha-helices with better accuracy. *Protein Eng.* 12: 557-561.
14. Hirokawa, T., Boon-Chieng, S., and Mitaku, S. 1998. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14: 378-379.
15. Nikiforovich, G.V., Galaktionov, S., Balodis, J., and Marshall, G.R. 2001. Novel approach to computer modeling of seven-helical transmembrane proteins: Current progress in the test case of bacteriorhodopsin. *Acta Biochimica Polonica* 48: 53-64.
16. Rabenstein, M.D., and Shin, Y.K. 1995. Determination of the distance between two spin labels attached to a macromolecules. *Proc. Natl. Acad. Sci. USA* 92: 8239-8243.

17. Sale, K.L., Sar, C., Sharp, K.A., Hideg, K., and Fajer, P. 2002. Structural determination of spin label immobilization and orientation: A Monte Carlo minimization approach. *Journal Magnetic Resonance* 156: 104-112.
18. 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.
19. Unger, V.M., and Schertler, G.F. 1995. Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. *Biophysical Journal* 68: 1776-1786.
20. Vaidehi, N., Floriano, W.B., Trabanino, R., Hall, S.E., Freddolino, P., Choi, E.J., Zamanakos, G., and Goddard III, W.A. 2002. Prediction of structure and function of G protein-coupled receptors. *PNAS* 99: 12622-12627.
21. Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graphics* 8: 52-56.
22. Xiang, Z., Soto, C., and Honig, B. 2002. Evaluating Conformational Free Energies: The Colony Energy and its Application to the Problem of Loop Prediction. *Proc. Natl. Acad. Sci. USA* 99: 7432-7437.
23. Yang, K., Farrens, D.L., Hubbell, W.L., and Khorana, H.G. 1996. Structure and function in rhodopsin. Single cysteine substitution mutants in the cytoplasmic interhelical E-F loop region show position-specific effects in transducin activation. *Biochemistry* 35: 14040 - 14046.
24. Yeagle, P.L., Choi, G., and Albert, A.D. 2001. Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms. *Biochemistry* 40: 11932-11937.
25. 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. *Proceedings of the National Academy of Sciences USA* 97: 5802-5806.
26. Yu, H., Kono, M., McKee, T.D., and Oprian, D.D. 1995. A general method for mapping tertiary contacts between amino acid residues in membrane embedded proteins. *Biochemistry* 34: 14963 - 14969.
27. Yu, H., Kono, M., and Oprian, D.D. 1999. State-dependent Disulfide Cross-linking in Rhodopsin. *Biochemistry* 38: 12028 - 12032.

Tables

Table 4.1. Number of seven-helix membrane protein templates vs. number of distances. Results obtained using a distance error of $\pm 4 \text{ \AA}$ and clustered with RMSD^{Hel} values of 4 \AA AND 6 \AA (see methods for definition of RMSD^{Hel}).

nbr distances	unlabeled templates	labeled templates	oriented templates	labeled clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	labeled clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$
7	39606	84289	2671680	50189	8192	1590824	259659
10	32298	66295	1926966	36480	5372	1060347	156145
15	12826	18212	394705	10738	1954	232722	42349
20	1274	1369	14656	573	127	6134	1360
30	71	73	518	65	32	461	227
38	33	35	44	31	16	39	20

Table 4.2. Number of seven-helix membrane protein templates vs. distance error. Results obtained using the 38 distances listed in Figure 4.1 and clustered with RMSD^{Hel} values of 4 \AA AND 6 \AA (see methods for definition of RMSD^{Hel}).

error (\AA)	unlabeled templates	labeled templates	oriented templates	labeled clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	labeled clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$
± 4	33	35	44	31	16	39	20
± 5	408	417	2314	318	126	1765	699
± 6	2259	2339	16575	1578	438	11182	3104
± 7	6468	6892	65117	4551	1049	42999	9911
± 8	12851	14162	159832	9520	2073	107442	23396

Table 4.3. Number of seven-helix membrane protein templates vs. distance graph radius. Results obtained using 17 distances extracted from Figure 4.1 and a distance error of $\pm 4 \text{ \AA}$ and clustered with RMSD^{Hel} values of 4 \AA AND 6 \AA (see methods for definition of RMSD^{Hel}). The distance graph corresponding to radius 3 is given in Figure 4.1.

radius	unlabeled templates	labeled templates	oriented templates	labeled clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	labeled clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 4 \text{ \AA}$	oriented clustered $\text{RMSD}^{\text{Hel}} = 6 \text{ \AA}$
1	1354	1487	11510	1104	362	8545	2802
2	4873	5854	73474	4591	1387	57622	17408
3	10056	13491	229830	10293	2671	175350	45503

Table 4.4. Results for seven membrane protein structures using distance constraints between potential cross-linking amino acid pairs (K-K, K-D, K-E, K-C, C-C) and for the structure of rhodopsin using a set of experimental distances. The number of solutions reported for ± 4 Å error are unclustered oriented or atomistic templates. Note that the number of solution obtained for rhodopsin (108) is different from the one found in the first row of Table 4.2 (44), because we doubled the size of the unlabeled templates library. The average distance error for rhodopsin in the last row is ± 3.75 Å.

Protein	PDB entry	Number of helices ¹	Number of distances	Number of solutions for $\varepsilon = 4$ Å	$C\alpha$ RMSD ² (Å)
Fumarate reductase flavoprotein	1qlaC	5	58	26	4.4
Aquaporin-1	1fqy	6	17	1,154,191	4.6
Glycerol uptake facilitator protein	1fx8	6	43	208	6.0
Bacteriorhodopsin	1c3w	7	60	330	3.0
Halorhodopsin	1e12	7	9	8,274,078	2.3
Sensory Rhodopsin	1jgj	7	18	329,502	3.7
Rhodopsin	1f88	7	38	108	3.2
Rhodopsin (Experimental Constraints given in Table 4.5)	1f88	7	27	87	4.3

1 Only those helices completely within the membrane were included.

2 Minimum $C\alpha$ - RMSD among all solutions (oriented templates) and the PDB entry shown in column 2.

Table 4.5. Experimental distances used for the Rhodopsin structure.

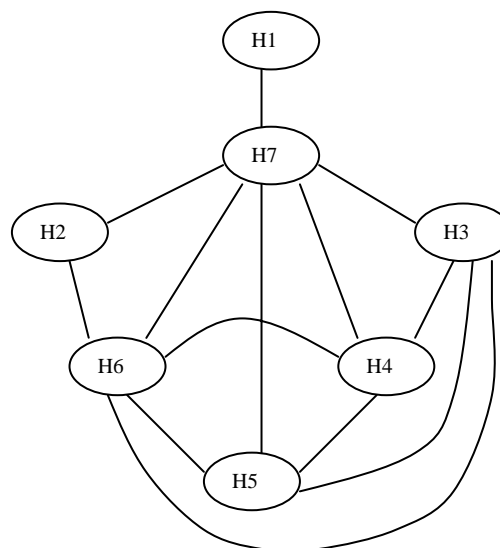
Helix1	Helix2	Residue1	Residue2	Minimum Distance	Maximum Distance
H3	H6	VAL_139	LYS_248	3.80	22.20
H3	H6	VAL_139	GLU_249	8.30	26.70
H3	H6	VAL_139	VAL_250	8.30	26.70
H3	H6	VAL_139	THR_251	3.80	22.20
H3	H6	VAL_139	ARG_252	8.30	26.70
H5	H6	VAL_204	PHE_276	3.18	8.18
H3	H5	CYS_140	CYS_222	1.18	10.18
H3	H5	CYS_140	GLN_225	4.18	7.18
H3	H6	ARG_135	VAL_250	2.18	9.18
H3	H5	TYR_136	CYS_222	4.18	7.18
H3	H5	TYR_136	GLN_225	3.18	8.18
H2	H3	PRO_71	GLU_134	7.00	15.00
H2	H3	GLY_90	PHE_116	5.00	10.00
H2	H4	PRO_71	ALA_153	4.00	11.00
H2	H4	MET_86	LEU_172	15.00	20.00
H3	H5	TYR_136	LEU_226	5.00	10.00
H3	H5	LEU_125	PRO_215	5.00	10.00
H4	H5	HIS_152	GLN_225	18.00	22.00
H5	H6	LEU_216	LEU_258	9.00	13.00
H6	H7	MET_253	MET_305	5.00	9.00
H6	H7	CYS_264	SER_298	5.00	9.00
H1	H7	MET_39	ILE_286	9.00	14.00
H3	H6	GLY_114	TYR_268	14.00	18.00
H4	H6	PRO_171	TYR_268	16.00	21.00
H2	H6	ASN_73	VAL_250	10.00	15.00
H1	H6	THR_62	VAL_250	16.00	20.00
H1	H6	LEU_47	CYS_264	15.00	20.00

Figures

(A)

Helix1	Helix2	Residue1	Residue2	Distance
H1	H7	GLUN_33	LYS_296	21.8
H2	H6	ASP_83	LYSN_245	29.7
H2	H6	ASP_83	LYS_248	26.4
H2	H7	ASP_83	LYS_296	9.8
H3	H4	CYSH_110	CYSH_167	16.5
H3	H4	CYS_140	CYSH_167	30.5
H3	H5	CYSH_110	CYSH_222	36.3
H3	H5	CYS_140	CYSH_222	9.2
H3	H6	CYSH_110	CYSH_264	21.2
H3	H6	GLU_122	LYSN_245	30.0
H3	H6	GLU_122	LYS_248	25.7
H3	H6	GLU_134	LYSN_245	16.3
H3	H6	GLU_134	LYS_248	13.0
H3	H6	CYS_140	LYSN_245	14.2
H3	H6	CYS_140	LYS_248	11.6
H3	H6	CYS_140	CYSH_264	28.0
H3	H7	CYSH_110	LYS_296	15.0
H3	H7	GLU_113	LYS_296	10.8
H3	H7	GLU_122	LYS_296	12.6
H3	H7	GLU_134	LYS_296	26.6
H3	H7	CYS_140	LYS_296	33.7
H4	H5	CYSH_167	CYSH_222	22.8
H4	H6	GLU_150	LYSN_245	22.1
H4	H6	GLU_150	LYS_248	20.8
H4	H6	CYSH_167	LYS_248	32.2
H4	H6	CYSH_167	CYSH_264	18.3
H4	H7	GLU_150	LYS_296	30.1
H4	H7	CYSH_167	LYS_296	17.6
H5	H6	CYSH_222	LYSN_245	19.5
H5	H6	CYSH_222	LYS_248	15.1
H5	H6	CYSH_222	CYSH_264	20.9
H5	H7	GLU_201	LYS_296	25.2
H5	H7	CYSH_222	LYS_296	27.0
H6	H7	LYSN_245	LYS_296	33.2
H6	H7	GLU_247	LYS_296	29.9
H6	H7	LYS_248	LYS_296	29.3
H6	H7	GLU_249	LYS_296	26.9
H6	H7	CYSH_264	LYS_296	8.1

(B)



(C)

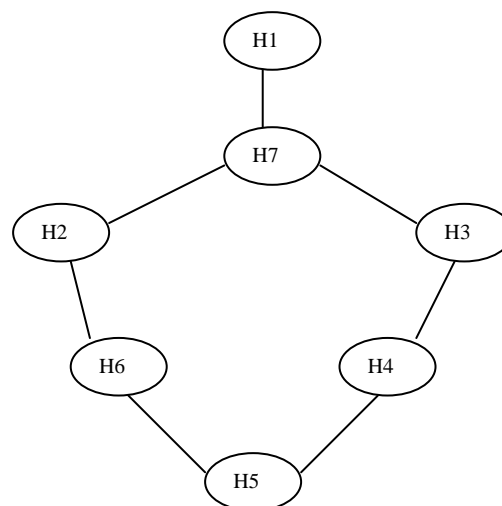


Figure 4.1. Distance constraints and distance graph. (A) Distance list computed for rhodopsin (1f88). (B) Corresponding distance graph. An edge is drawn between two helices if a distance exists between these helices. All helices are linked to helix H7, thus the radius of the distance graph is 1. (C) Distance graph corresponding to the distances in italic in the distance list (A). The radius of this distance graph is 3 since every helix can be reached from H7 using no more than 3 links.

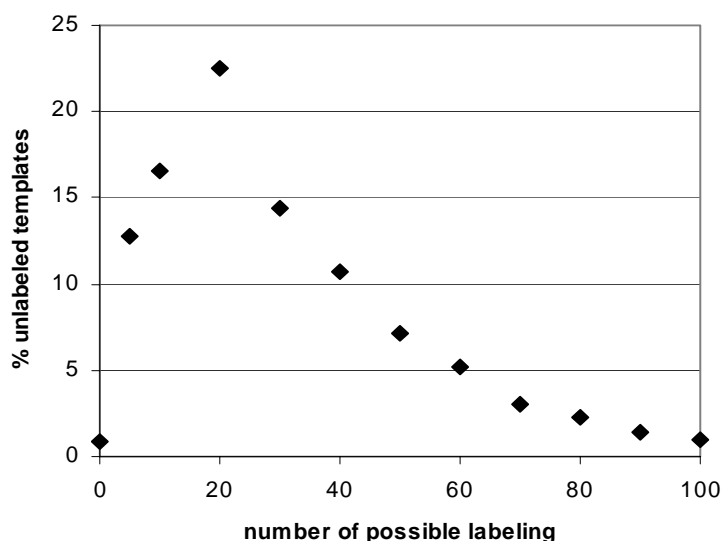


Figure 4.2. Labeling distribution for seven-helix templates

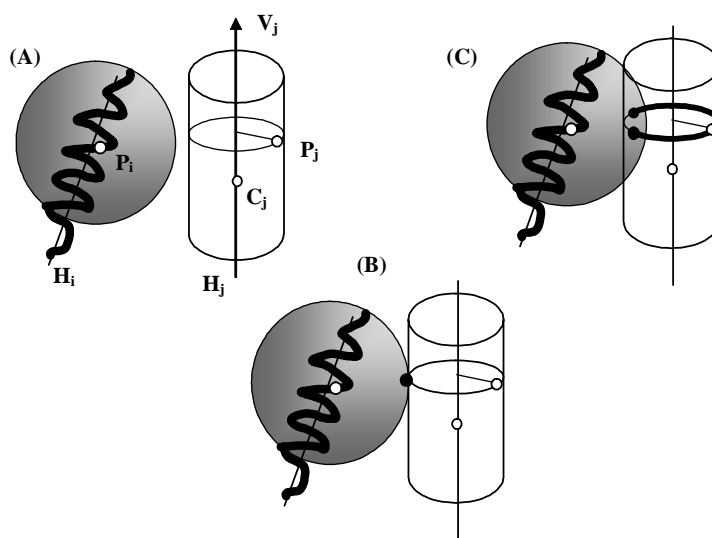


Figure 4.3. Helix positioning using distance constraints.

We assume helix H_i has already been positioned and we are provided with the distance d_{ij} between points P_i and P_j . Helix H_j is defined by its center of mass C_j and its unit vector V_j . P_j is an atom of H_j , which without distance information can be at any position on a circle induced by the possible orientations of H_j around its axis. Positioning helix H_j consists of finding a position for P_j such that P_j lies on its circle and is at distance d_{ij} from P_i , or in other words is on the sphere centered on P_i and of radius d_{ij} . (A) Circle and sphere do not intersect; all possible P_j positions are too far from P_i . The other situation without solution (not represented here) is when all possible P_j positions are too close to P_i , which case the circle is inside the sphere. (B) The circle is tangent to the sphere; only one P_j position is at distance d_{ij} from P_i . (C) The circle crosses the sphere and the two intersection points are the two possible positions for P_j .

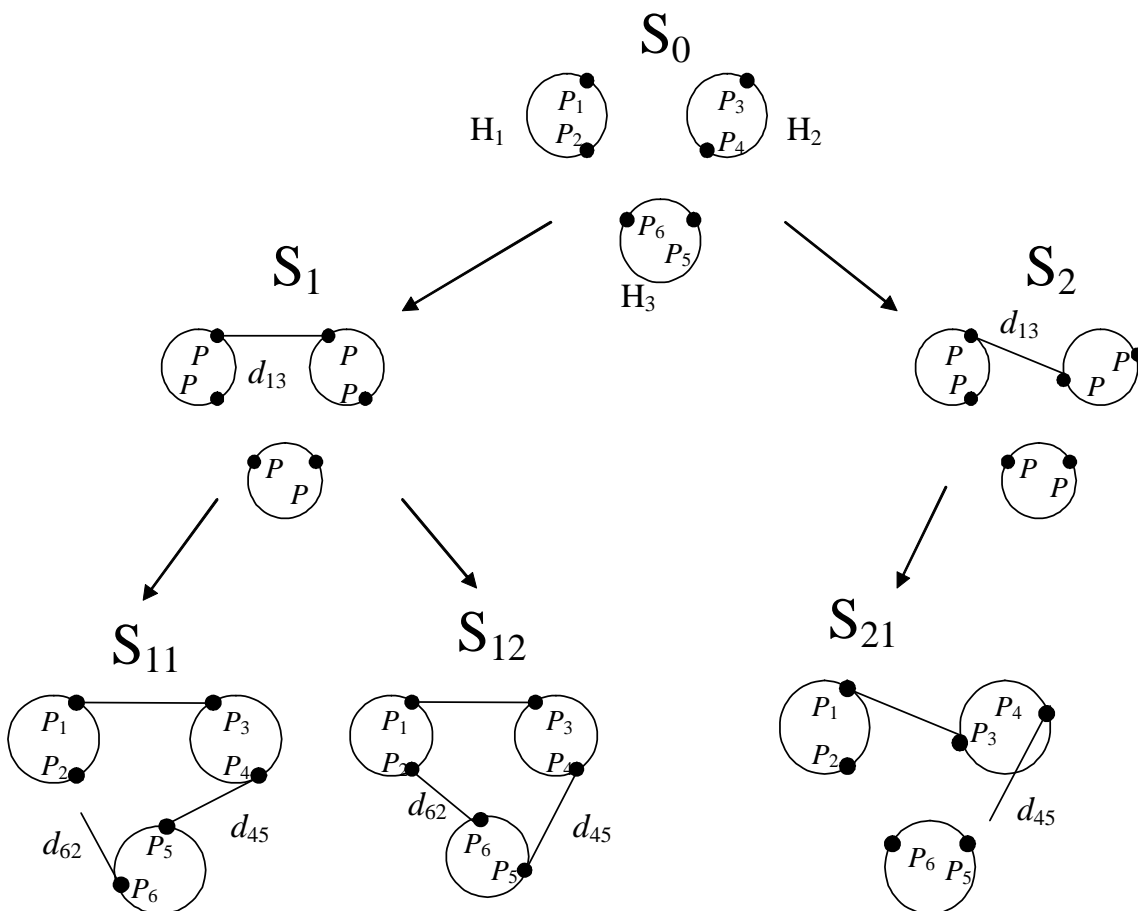


Figure 4.4. Enumerating oriented templates. S_0 is the initial labeled template composed on three helices viewed from the top. We assume helix H_1 has already been positioned and distances $d_{13} = D(P_1, P_3)$, $d_{45} = D(P_4, P_5)$, and $d_{62} = D(P_6, P_2)$ are provided. Helices H_2 and H_3 are arbitrarily positioned in S_1 . As explained in Figure 4.3, there are two positions for P_3 both being at distance d_{13} from P_1 ; these are represented by templates S_1 and S_2 . Similarly there are two positions for P_5 at distance d_{45} from P_4 (templates S_{11} and S_{12}). Template S_{11} is rejected because distance d_{62} is not matched, template S_{21} is also rejected because distance d_{45} is too short. Template S_{12} is the only one passing all distance tests.

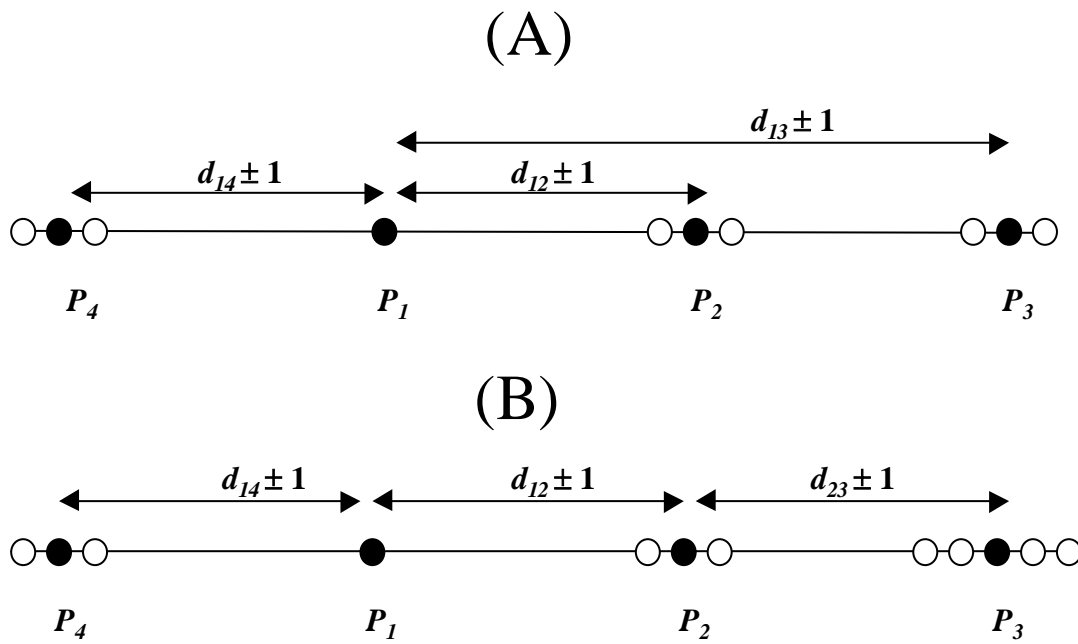


Figure 4.5. (A) The radius of the distance graph is 1 and the number of possible positions is 3 for points P_2 , P_3 and P_4 . The total number of configurations is 27. (B) The radius of the distance graph is 2, points P_1 and P_4 have 3 possible positions and point P_3 has 5 positions, thus the total number of configurations is 45.

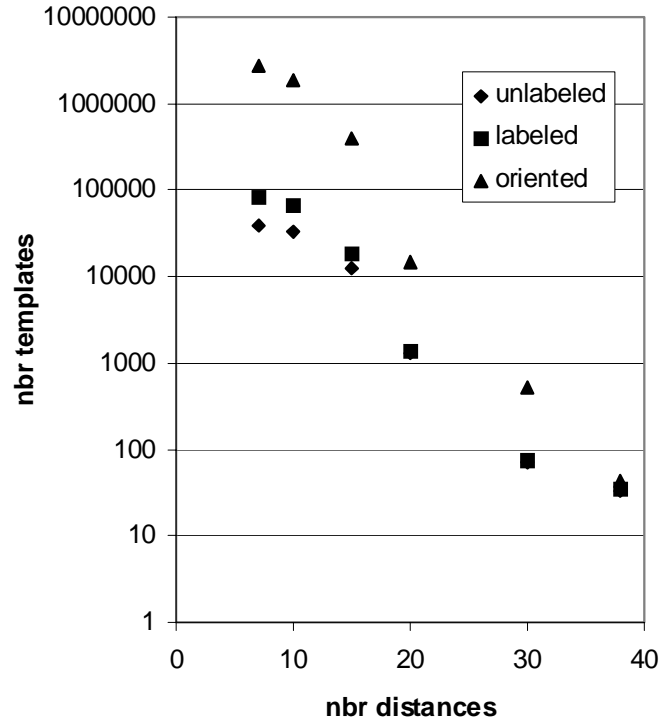


Figure 4.6. Number of seven-helix membrane protein templates vs. number of distances (data listed in Table 4.1).

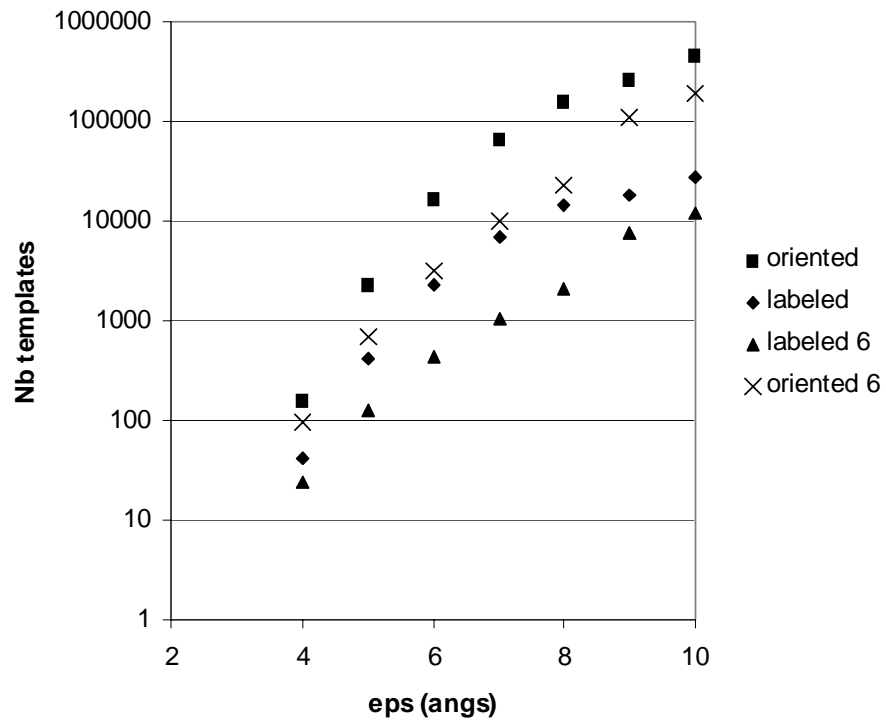


Figure 4.7. Number of seven-helix membrane protein templates vs. distance error (data listed in Table 4.2).

This page intentionally left blank.

Chapter 5: Comparison between isolated rhodopsin helices simulations and intact rhodopsin simulation

Paul S. Crozier, Mark J. Stevens and Joseph S. Schoeniger

Abstract

The seven transmembrane helices of bovine rhodopsin are examined via molecular dynamics simulation, with comparison of helix characteristics as simulated in intact rhodopsin versus simulation of the isolated helices. The all-atom molecular dynamics simulations included explicit lipid bilayers and explicit solvent molecules in all cases, and were 20 nanoseconds each for the isolated helix simulations and 40 nanoseconds for the intact rhodopsin simulation. The simulations show that helix behavior is strongly dependent on the rest of the protein to determine the helix tilt, kink, and secondary structure characteristics. Key elements of secondary structure that are present in native rhodopsin were maintained in the intact rhodopsin simulation, but were absent in the isolated helix simulations due to the lack of the inter-helical interactions and the loop:helix interactions that are needed in order to form and maintain the native secondary structure.

Introduction

Even though membrane proteins are extremely important to cell function and make up 20-30% of the genome,¹ few well-defined structures of membrane proteins have been published owing partly to the difficulty of membrane protein crystallization. The dark-adapted form of bovine rhodopsin is a notable exception²⁻³ and provides considerable insight into the structure and function of a large class of membrane proteins. The protein folding problem is especially difficult in the case of membrane proteins since the mechanism of assembly is not entirely clear and involves the membrane. It has been postulated that individual transmembrane helices form first, followed by formation of the native transmembrane helix bundle conformation⁴. This presupposes that transmembrane helices are independently stable units of secondary structure, and raises the question: if isolated, do individual transmembrane helices maintain native conformation, or are their structures highly dependent on the remainder of the protein? This article addresses this and other relevant questions via molecular dynamics (MD) simulation of isolated transmembrane helices from rhodopsin as well as MD simulation of the known structure of intact dark-adapted rhodopsin.

Goddard's group built models of membrane proteins using a computational approach, where transmembrane regions were predicted, and idealized α -helices built and optimized using torsional MD⁵. Further refinement of their model included all-atom MD simulation in an explicit lipid bilayer using the MPSim software.⁶ Their successes have provided motivation for our exploration of the hypothesis that simulation of isolated alpha helices can be used to predict helix irregularities. However, we show that isolated helix simulation cannot be used for accurate prediction of helix tilts, kinks, and secondary structure characteristics. The other helices and loops of the given transmembrane protein strongly affect helix structure.

Biggin and Sansom give an excellent summary of α -helix simulation studies in their 1999 review article.⁷ Here, we give a brief overview of recent work specific to all-atom simulation. Previous MD studies of isolated helices have, to a limited degree, demonstrated the ability to predict membrane protein secondary structure, including kinks, bends, and tilts. Bright *et al.*,⁸ Cordes *et al.*,⁹ and Bright *et al.*¹⁰ have shown that helix distortions induced by proline and glycine motifs can be studied by MD simulation of isolated helices in a membrane-mimetic biphasic octane/water system. Woolf¹¹⁻¹³ performed all-atom MD simulation of isolated bacteriorhodopsin helices in an explicit dimyristoyl phosphatidylcholine (DMPC) lipid environment. Due to the CPU-intensive nature of all of the calculations mentioned above, each was limited in terms of simulation time (up to a 1.0 ns each), lipid to helix ratio (12:1), and long-range electrostatics (12 Å cutoff). In contrast, the isolated helix simulations of the present study are 20 ns in duration, had a lipid-to-helix ratio of 30:1, and use the P³M long-range electrostatics method.

Methods

Rather than perform simulations of isolated helices in a vacuum, a mean field environment, or a membrane-mimetic environment,⁸ we have performed these simulations in an explicit DOPC lipid environment immersed in an aqueous (TIP3P) ionic (NaCl) environment. This has the advantage of a much more accurate description of the interfacial region between the aqueous phase and the hydrophobic core of the lipid bilayer. No approximation to hydrophobic matching is needed and fluctuations of the hydrophobic region are allowed. Also avoided are problems associated with inadequate modeling of the environmental viscosity that likely influences the bending dynamics of the transmembrane helices. The disadvantage of including the lipid and solvent environments in atomistic detail is that these simulations are very computationally demanding.

Long-range electrostatics calculations can also be computationally demanding. But inadequate representation of the long-range electrostatics can introduce unwanted artifacts, so we have used full particle-particle/particle-mesh (P³M) Ewald electrostatics to avoid errors associated with cutoff electrostatics.

Good lipid mixing and system equilibration are difficult to achieve on the currently feasible nanoseconds time scale of all-atom molecular simulation. In an effort to reduce time scale concerns, we have built ensembles from

pre-equilibrated lipids according to the protocol established by Woolf and Roux¹⁴ to minimize the needed equilibration time, and performed relatively long time scale simulations of 20 nanoseconds for each isolated transmembrane helix. Initialization and equilibration was done using the CHARMM molecular modeling package, and subsequently, each simulation required approximately 20 CPU days on 12 processors of CPlant¹⁵ using the LAMMPS¹⁴ parallel MD software package.

The seven transmembrane helices of rhodopsin were defined as the residues between and including the first and last residues as given in Table 5.1. In order to set up the membrane protein ensembles in an explicit lipid bilayer, we followed the procedure of Woolf and Roux.¹⁴ A brief description of the procedure follows. Each isolated helix simulation was initialized by placement of the idealized helix perpendicular to the membrane surface and centered in the membrane. DOPC lipid molecules were then chosen at random from a pre-equilibrated library of lipids created from a pure lipid bilayer simulation. Lipids were fitted around the protein and rotation and translation moves were performed to minimize atom overlaps, which were then followed by energy minimization. Following construction of the lipid bilayer around the isolated helix, each system was immersed in TIP3P water and NaCl ions were added to neutralize the net system charge and achieve ~0.1 M salt bath solution. After relaxation by further energy minimization, equilibration dynamics were performed for 25 ps.

All simulations were performed at a constant temperature of 307 K and at a constant pressure of 1 atm using a Nosé-Hoover thermostat/barostat. In order to maintain a constant membrane surface area, the simulation box was allowed to fluctuate only in the direction perpendicular to the membrane, with the box dimensions in the other two directions held constant. After initial equilibration, production runs of 20 ns for each of the seven isolated transmembrane helices were performed for comparison with a 40 ns intact rhodopsin simulation performed under similar conditions and reported elsewhere.¹⁷

In order to allow motion free of artificially imposed restraints, none of the isolated helix ends were tethered, nor were the backbone atoms fixed in place after the initial equilibration. With no restraints imposed, it will be shown that the isolated helices have a tendency to unravel near the ends and to degenerate away from the initial α -helix conformation.

Results

The simulation time frame of 20-40 ns is adequate to observe a great deal of helix motion, especially in the case of the isolated helix simulations. Substantial changes in helix tilt, kink, conformation, and energetics were observed, with much larger changes seen in the isolated helices than in the intact rhodopsin helices. Although the isolated helix simulations were initialized with each helix placed perpendicular to the membrane in the idealized α -helix conformation, the helices quickly devolved into less ordered structures with large tilt and kink angles. Large changes in helix conformation continued to occur throughout the entire 20 ns of each isolated helix simulation, signaling that, in some cases, it is possible that the equilibrium ensemble has yet to be reached, since the simulation moves far from the starting structure. This behavior is not unlike that observed in other isolated helix simulations,^{7-14,18} although the present simulations give a longer and more complete picture.

Helix Tilt and Kink Angles

Figures 5.1 and 5.2 display tilt and kink angles of the transmembrane helices from the intact rhodopsin simulation on the left and the isolated helix simulations on the right. The tilt angle, defined as the angle between the z -axis, perpendicular to the membrane surface, and the major axis for the cylinder that best fits the helix, were then computed for each snapshot and plotted as a function of time in Figure 5.1. For the kink angles, sections of each helix above and below each kink point were fit to cylinders and the kink angle was defined as the angle between the axes of the contiguous cylinders on the same α -helix. Kink angles are plotted as a function of time in Figure 5.2, with the hinge point residue for each kink listed in the caption. Experimentally measured tilt and kink angles are marked with an arrow and a straight line, representative of the angles computed from the static x-ray structure of rhodopsin.

Tilt angle fluctuations on the picosecond time scale are on the order of a few degrees in both the intact rhodopsin simulation and in the isolated helix simulations. But over the course of the simulation, the range between the minimum and maximum tilt angle in the intact rhodopsin simulation is 15-17° for each helix, while the range between the minimum and maximum tilt angle in the corresponding isolated helix simulations is 39-52° and would likely be larger if the simulations were continued. The large range for the isolated helix simulations can partly be blamed on the fact that the isolated helix simulations were started with the helices perpendicular to the membrane, whereas in the case of the intact rhodopsin simulation, the helices were initialized in their known x-ray structure positions. In particular, isolated helices 3 and 5 achieve tilt angles close to the crystal values (Figure 5.1). However, for helices 4 and 7, the trend was almost entirely away from the experimentally measured tilt angle. The tilt angle of isolated helix 1 is steadily increasing and deviates from the crystal value by 7° at the end of the run.

Comparison with the experimentally measured tilt angles shows that deviations were much larger for the isolated helices than for the helices in the intact rhodopsin. The absolute value of the difference between the simulated tilt angle and the experimentally measured tilt angle, as averaged over all 7 transmembrane helices for each snapshot, was 5° for the intact rhodopsin simulation, while it was 17° for the isolated helix simulations. Likewise, examining just the endpoint snapshot reveals that the intact rhodopsin simulation helix tilt angles varied from the experimentally measured tilt angles by an average of only 5°, whereas they varied by an average of 19° in the isolated helix simulations.

Similar to the tilt angles, fast fluctuations in kink angles are on the order of 10° in both the intact rhodopsin simulation and in the isolated helix simulations (Figure 5.2). But, again, the ranges between the minimum and maximum kink angles in the isolated helix simulations are much larger than in the intact rhodopsin simulation, indicating much more freedom of motion around kinking residues in the case of the isolated helix simulations. Isolated helices 1 and 3(b) have kink angles similar to the corresponding intact rhodopsin helices, albeit with larger fluctuations. By the end of the simulation, isolated helix 3(a) has relaxed to the same value as the intact rhodopsin helix 3(a), although it is not clear that it would have remained at that value if the run had been continued even longer. Isolated helices 2, 6, and 7 differ significantly from the corresponding intact rhodopsin helices.

Very large kink angles occurred in some of the isolated helices, but almost never exceeded 45° in the intact rhodopsin. We show an example of a very large kink angle in Figure 5.3, where the simulation of isolated helix 2 produced a kink centered on Ala 80. Interestingly, no kink formed around Gly 89 and Gly 90 as in the native rhodopsin structure. Without the constraints at either end of the helix and without the hydrogen bonding networks of the neighboring helices that are present in native rhodopsin, the isolated helix is free to kink in a different location and to solvent match, exhibiting exaggerated kinking in a non-native location in order to minimize contact of hydrophobic residues with water and contact of hydrophilic residues with the lipid tails.

HELANAL Classifications

Kumar and Bansal¹⁹⁻²⁰ have devised a system for dividing α -helices into kinked, curved, or linear classifications, and have shown that almost all α -helices fall into these categories. We have used HELANAL, their analysis program, to classify each helix snapshot from the simulations as kinked, curved, linear, or ambiguous. The present simulations show that the individual helices of rhodopsin are primarily classified as kinked, but for a small fraction of snapshots during the course of the simulations are better classified as curved or linear.

According to the HELANAL classification system, kinked helices are characterized by local bending angles greater than 20°. Linear and curved helices have no local bending angles greater than 20°, and the average of their local bending angles is less than 10°. A helix is considered linear rather than curved if a fit to a straight line is better than a similar rmsd fit to an arc. If the fit to an arc is better, the helix is considered curved rather than linear.

Due to the fact that each of the transmembrane helices in rhodopsin has at least one kink-inducing residue, rhodopsin's helices predominately exist in a kinked state. It has been speculated that kinks act as molecular hinges that allow membrane proteins to undergo conformational changes necessary for signaling.⁸⁻⁹ Kinks in rhodopsin's helices are likely important for conformational changes during photoisomerization¹⁷. As evidenced by the HELANAL analysis data presented in Table 5.2, all of the transmembrane in both the intact rhodopsin and the isolated helices simulations are predominately categorized as kinked during the vast majority of the time. Non-kinked behavior was distributed fairly evenly across the time frame of each of the simulations rather than distributed as blocks of non-kinked time. In the intact rhodopsin simulation, only helices 1, 3, and 5 exhibited substantial curved helix behavior, with helix 3 being the only helix to exhibit some linear behavior. In contrast, the isolated helix simulations showed an even greater majority of kinked helix behavior, with only helices 1 and 4 showing some curved helix behavior, and only helix 4 showing a small fraction of linear helix behavior.

Overall, the HELANAL analysis points to the fact that the presence of the entire intact membrane protein is needed in order to maintain the native helix geometric configurations. This conclusion will be further reinforced and explored in greater detail at the residue level in what follows.

DSSP Comparisons

In addition to examining each of the helices in its entirety in terms of its overall shape, we have examined the individual residues of each helix in terms of its predominate secondary structure characteristics using the DSSP (Define Secondary Structure of Proteins) program of Kabsh and Sander.²¹ Secondary structure for each of the simulated helices was computed as a function of time, giving a dynamic picture of the secondary structure of each residue of each helix, as shown in Figures 5.4-5.17. In general, the isolated helices exhibited much less order in their structure than did the helices in intact rhodopsin (especially helices 3 and 6). Helix 7 is the notable exception.

The isolated helices were initiated as completely idealized α -helices, but they quickly departed from that state, showing kinking, turns, and unraveling, particularly at the ends. Even though unraveling progressed somewhat over the course of the simulations, in general, it appears that a reasonably stable steady state has been achieved. Irregular helix formation, involving 3_{10} helices and π -helices, occurs much more frequently in the isolated helix simulations than in the intact rhodopsin simulations. While some of this behavior can be blamed on force field inadequacies,²² it is not uncommon to observe overcoiling or undercoiling near the ends of α -helices in their native structures.

A simulated isolated rhodopsin helix can have kinks present that are not observed in the corresponding intact rhodopsin helix simulation. Likewise, a kink present in the intact rhodopsin will not necessarily be present in the corresponding isolated helix simulation. A great deal of helix distortion and unraveling near the ends was observed in the isolated helix simulations and not in the intact rhodopsin helices. The following paragraphs discuss in more detail specific differences observed between each helix in the isolated and intact simulations.

Helix 1 is nearly ideal and has very little kinking near Pro 53 in both the intact rhodopsin simulation and in the isolated helix simulation. In general, the presence of proline suggests that there will be α -helix distortion, but very little distortion occurs in this case. Even less distortion occurred around Pro 53 in the isolated helix simulation than in the intact rhodopsin simulation. This is in contrast to the large and growing α -helix distortion at either end of the helix in the isolated helix simulation, where it tends to degenerate into π -helix and hydrogen bonded turn regions. The isolated helix 1 is very different from the intact rhodopsin helix 1, primarily at the ends. The intact rhodopsin helix does not become non-helical until Glu 33 on the extracellular side and His 65 on the cytoplasmic side. In contrast, the isolated helix becomes non-helical from Phe 37 to the extracellular compartment and from Leu 59 to the cytoplasmic compartment, even unraveling down to Pro 53 in some instances. The intact rhodopsin helix exhibits a kink at Pro 53, which shows up as "H-bonded turn" at Phe 52 in Figure 5.4. This kink is also hinted at in the isolated helix 1 simulation, but is less distinct and shifted slightly, as seen by the faint "H-bonded turn" between residues 50 and 52 in Figure 5.5.

The only proline residue in helix 2 is Pro 71, found near the cytoplasmic end of the helix. Pro 71 causes similar kinking distortion in the intact rhodopsin simulation and in the isolated helix simulation. But kinking near Pro 71 is mild compared to the major kinking center found in native rhodopsin helix 2 centered on Gly 89 and Gly 90. Kinking around these two glycines is glaringly absent in the isolated helix simulation. The major kink found in the intact rhodopsin simulation centered on Gly 89 and Gly 90 places Gly 90 close to the Schiff base, Glu 113 of helix 3. In the absence of helix 3 and the rest of rhodopsin, no kink forms around Gly 89 and Gly 90, indicating that the tertiary structure holds the secondary structure in place. In native rhodopsin, the structure of helix 2 is influenced by interhelical H-bonding networks involving Asp 83 and Asn 78.² In contrast to the intact rhodopsin simulation of helix 2, the isolated helix 2 simulation exhibits kinking in the region near Ala 80 rather than in the region near Gly 90. The DSSP analysis of the intact rhodopsin helix implies that the helix consists of residues 73 to 100 with 88-91 being non-helical, but maintaining the H-bonded turn structure, which is fairly constant throughout the run. The isolated helix is different on the extracellular end, where for $t > 8$ ns, residues 96-99 tend to be either an H-bonded turn or a π -helix. The cytoplasmic end of the isolated helix is only slightly different from the intact helix of the same residue range. The kink that is evidenced by a few residues of non-helical H-bonded turn moves to residues 77-81 in the isolated helix.

Helix 3 in the intact rhodopsin simulation is almost entirely α -helical, while the simulation of isolated helix 3 shows a great deal of kinking and non-ideality according to the DSSP analysis. Gly 120 and Gly 121 do not distort helix 3 in the intact rhodopsin simulation, but cause a distinct kink in the simulation of isolated helix 3. Likewise, a turn occurs near Glu 113 and Gly 114 in the isolated helix but is entirely absent in the intact rhodopsin simulation. In the case of helix 2, there was kinking around glycines in the intact rhodopsin simulations without kinking around glycines in the isolated helix simulations, whereas the opposite is true in the case of helix 3.

The simulations of helix 4 are much more similar to each other. The two proline residues in a row, Pro 170 and Pro 171 induce a non-helical region with no inter-residue hydrogen bonds (labeled "bend" by DSSP) in helix 4, which appears in the DSSP analyses as a bend at Ala 169. The bend is seen in the intact rhodopsin simulation as well as in the isolated helix simulation, albeit more pronounced and focused on Ala 169 in the intact rhodopsin simulation. But residues 169-173 are more floppy in the isolated helix simulation. The cytoplasmic side of isolated helix 4 shows H-bonded turns from residue 150 to 152 for $t = 1-3$ ns, and for $t = 17-20$ ns residues 150 to 157 are H-bonded turns or π -helix. This behavior is not seen at all in the intact rhodopsin simulation of helix 4.

Kinking in helix 5, induced in part by Pro 215, is evidenced by a H-bonded turn region seen near Phe 212 in both the isolated helix simulation and in the intact rhodopsin simulation, but is more pronounced in the intact rhodopsin simulation. (See Figures 5.12 and 5.13) In the isolated helix simulation, the extracellular end of helix 5 near Glu 201 and Ser 202 begins as part of the α -helical, but devolves into an H-bonded turn region after about 6.5 ns. Similar behavior occurs on the cytoplasmic end of helix 5. After $t = 15$ ns, undercoiling of helix 5 transforms the region from residues 221 to 225 into a π -helix. This may be at least partially due to inadequacies in the CHARMM 22 force field, which has a tendency to favor π -helices in structures that should be α -helices.²²

From the rhodopsin x-ray structure (Ref. 2, p. 741), it can be observed that Phe 293 and Phe 294 of helix 7 interact with Leu 40 of helix 1 and Cys 264 of helix 6, and that this interaction is likely important to rhodopsin function since it is facilitated by the distortion of helix 6 in the region around Ile 263. Both the intact rhodopsin simulation and the isolated helix simulation show a large kink with a highly variable kink angle near Ile 263. (See Figure 5.2.) Interestingly, the DSSP analysis shows considerable helix distortion around Ile 263 in the isolated helix simulation, but the intact rhodopsin simulation shows only very faint distortion of the α -helical structure of helix 6, even with the kink present. The DSSP analysis shows that the intact rhodopsin helix is characterized as almost entirely α -helical, while the isolated helix 6 shows π helix formation near the cytoplasmic end, as seen in helix 5. Interestingly, this time the π helix transforms back into a α -helix before the end of the simulation.

With a considerably elongated mid-section, the intact rhodopsin helix 7 has a much richer structure than the other intact rhodopsin helices. The center section of helix 7 is primarily non-helical, with Thr 297 and Ser 298 in an

unstructured region for almost all times. Residues 299-301 are classified as an H-bonded turn region. Ala 295 and Lys 296 are primarily designated H-bonded turn residues, except for between $t = 11$ ns and $t = 24$ ns, when they usually form part of a tightly coiled 310-helix. The contrast with the isolated helix is considerable. None of the middle structure is present; it is characterized as helical for the entire simulation. After $t = 12$ ns, the α -helical region near the extracellular end actually grows, while the opposite occurs at the cytoplasmic end of the helix. Again, we see frequent π -helix formation in the isolated helix simulation, this time including the region between Asn 302 and Met 308.

In all seven rhodopsin transmembrane helices, very substantial structural differences were observed between the isolated helix simulations and the intact rhodopsin simulation. The observed differences are due to the unraveling of the untethered ends in the isolated helix simulations and the effects of the rest of the rhodopsin molecule beyond the helix in question, which plays a key role in enforcing helix secondary structure through hydrogen bonding networks and other tertiary structure interactions.

Conclusion

Simulations of isolated helices have been compared against the corresponding intact rhodopsin simulation from the level of the entire helix down to the structural details of the individual residues that make up the helices. These relatively long time scale simulations have shown that helix characteristics slowly evolve and require a long time to equilibrate. Use of all-atom MD simulations as a screening tool for evaluating individual transmembrane helix characteristics is both expensive and impractical.

It is not possible to use isolated helix simulations to predict membrane protein helix structures, even if simulated in a native-like lipid environment. The rest of the protein, with its tertiary structure, is required to enforce the helix secondary structure. Solvent matching alone is a necessary, yet insufficient, condition for tilt and kink angle determination or for proper folding.

Although all-atom simulation of isolated helices has been shown to be an inadequate method for helix characteristics prediction, good predictions were achieved in some cases. Reasonably favorable comparisons between the isolated helix simulations and the intact rhodopsin simulation were shown for the tilt angles of helices 3 and 5, the kink angles of helices 1 and 3, the HELANAL comparison of helices 2, 5, 6, and 7, and the DSSP comparison of helices 1, 4, and 5.

References

1. Erik Wallin and Gunnar Von-Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, Archaeal, and eukaryotic organisms, *Protein Science*, 7 (1998) 1029-1038.
2. Krzysztof Palczewski, Takashi Kumasaka, Tetsuya Hori, Craig A. Behnke, Hiroyuki Motoshima, Brian A. Fox, Isolde Le Trong, David C. Teller, Tetsuji Okada, Ronald E. Stenkamp, Masaki Yamamoto, and Masashi Miyano, Crystal structure of rhodopsin: a G protein-coupled receptor, *Science*, 289 (2000) 739-745.
3. David C. Teller, Tetsuji Okada, Craig A. Behnke, Krzysztof Palczewski, and Ronald E. Stenkamp, Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs), *Biochemistry*, 40 (2001) 7761-7772.
4. Jean-Luc Popot and Donald M. Engelman, Membrane protein folding and oligomerization: the two-stage model, *Biochemistry*, 29 (1990) 4031-4037.
5. Nagarajan Vaidehi, Wely B. Floriano, Rene Trabanino, Spencer E. Hall, Peter Freddolino, Eun Jung Choi, Georgios Zamanakos, and William A. Goddard III, Prediction of structure and function of G protein-coupled receptors, *Proceedings of the National Academy of Sciences of the United States of America*, 99 (2002) 12622-12627.
6. Kian-Tat Lim, Sharon Brunett, Mihail Iotov, Richard B. McClurg, Nagarajan Vaidehi, Siddharth Dasgupta, Stephen Taylor, and William A. Goddard III, Molecular dynamics for very large systems on massively parallel computers: the MPSim program, *Journal of Computational Chemistry*, 18 (1997) 501-521.
7. Phil C. Biggin and Mark S. P. Sansom, Interactions of α -helices with lipid bilayers: a review of simulation studies, *Biophysical Chemistry*, 76 (1999) 161-163.
8. Joanne N. Bright, Indira H. Shrivastava, Frank S. Cordes, and Mark S. P. Sansom, Conformational dynamics of helix S6 from Shaker potassium channel: simulation studies, *Biopolymers*, 64 (2002) 303-313.
9. Frank S. Cordes, Joanne N. Bright, and Mark S. P. Sansom, Proline-induced distortions of transmembrane helices, *Journal of Molecular Biology*, 323 (2002) 951-960.
10. Joanne N. Bright and Mark S. P. Sansom, The flexing/twirling helix: exploring the flexibility about molecular hinges formed by proline and glycine motifs in transmembrane helices, *The Journal of Physical Chemistry B*, 107 (2003) 627-636.
11. Thomas B. Woolf, Molecular dynamics of individual α -helices of bacteriorhodopsin in dimyristoyl phosphatidylcholine. I. Structure and dynamics, *Biophysical Journal*, 73 (1997) 2376-2392.
12. Thomas B. Woolf, Molecular dynamics simulations of individual α -helices of bacteriorhodopsin in dimyristoylphosphatidylcholine. II. Interaction energy analysis, *Biophysical Journal*, 74 (1998) 115-131.
13. Thomas B. Woolf, Bacteriorhodopsin α -helices in lipid settings: insights for structure prediction, *International Journal of Quantum Chemistry*, 69 (1998) 105-116.
14. Thomas B. Woolf and Benoit Roux, Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer, *Proceedings of the National Academy of Sciences of the United States of America*, 91 (1994) 11631-11635.

15. Ron Brightwell, Lee Ann Fisk, David S. Greenberg, Tramm Hudson, Mike Levenhagen, Arthur B. Maccabe, Rolf Riesen, Massively parallel computing using commodity components, *Parallel Computing*, 26 (2000) 243-266.
16. Steven J. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *Journal of Computational Physics*, 117, (1995) 1-19.
17. Paul S. Crozier, Mark J. Stevens, Lucy R. Forrest, and Thomas B. Woolf, Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change, *Journal of Molecular Biology*, 333 (2003) 493-514.
18. Lucy R. Forrest, D. Peter Tieleman, and Mark S. P. Sansom, Defining the transmembrane helix of M2 protein for influenza A by molecular dynamics simulations in a lipid bilayer, *Biophysical Journal*, 76 (1999) 1886-1896.
19. Sandeep Kumar and Manju Bansal, Structural and sequence characteristics of long α -helices in globular proteins, *Biophysical Journal*, 71 (1996) 1574-1586.
20. Sandeep Kumar and Manju Bansal, Geometrical and sequence characteristics of α -helices in globular proteins, *Biophysical Journal*, 75 (1998) 1935-1944.
21. Wolfgang Kabsh and Christian Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 22 (1983) 2577-2637.
22. Michael Feig, Alexander D. MacKerell, Jr., and Charles L. Brooks, III, Force field influence on the observation of π -helical protein structures in molecular dynamics simulations, *The Journal of Physical Chemistry B*, 107 (2003) 2831-2836.

Tables

Table 5.1. Rhodopsin residues included in each of the seven transmembrane isolated helix simulations.

Helix	First residue	Last residue
1	33	65
2	70	101
3	105	140
4	149	173
5	199	226
6	245	278
7	284	309

Table 5.2. HELANAL analysis results for the seven transmembrane helices of rhodopsin. The percentage of time that each helix spent in each HELANAL classification is shown.

Intact rhodopsin	helix 1	helix 2	helix 3	helix 4	helix 5	helix 6	helix 7
kinked	85%	100%	73%	100%	99%	100%	100%
curved	15%	0%	18%	0%	1%	0%	0%
linear	0%	0%	7%	0%	0%	0%	0%
ambiguous	0%	0%	1%	0%	0%	0%	0%
Isolated helices	helix 1	helix 2	helix 3	helix 4	helix 5	helix 6	helix 7
kinked	98%	100%	100%	89%	100%	100%	100%
curved	2%	0%	0%	9%	0%	0%	0%
linear	0%	0%	0%	1%	0%	0%	0%
ambiguous	0%	0%	0%	0%	0%	0%	0%

Figures

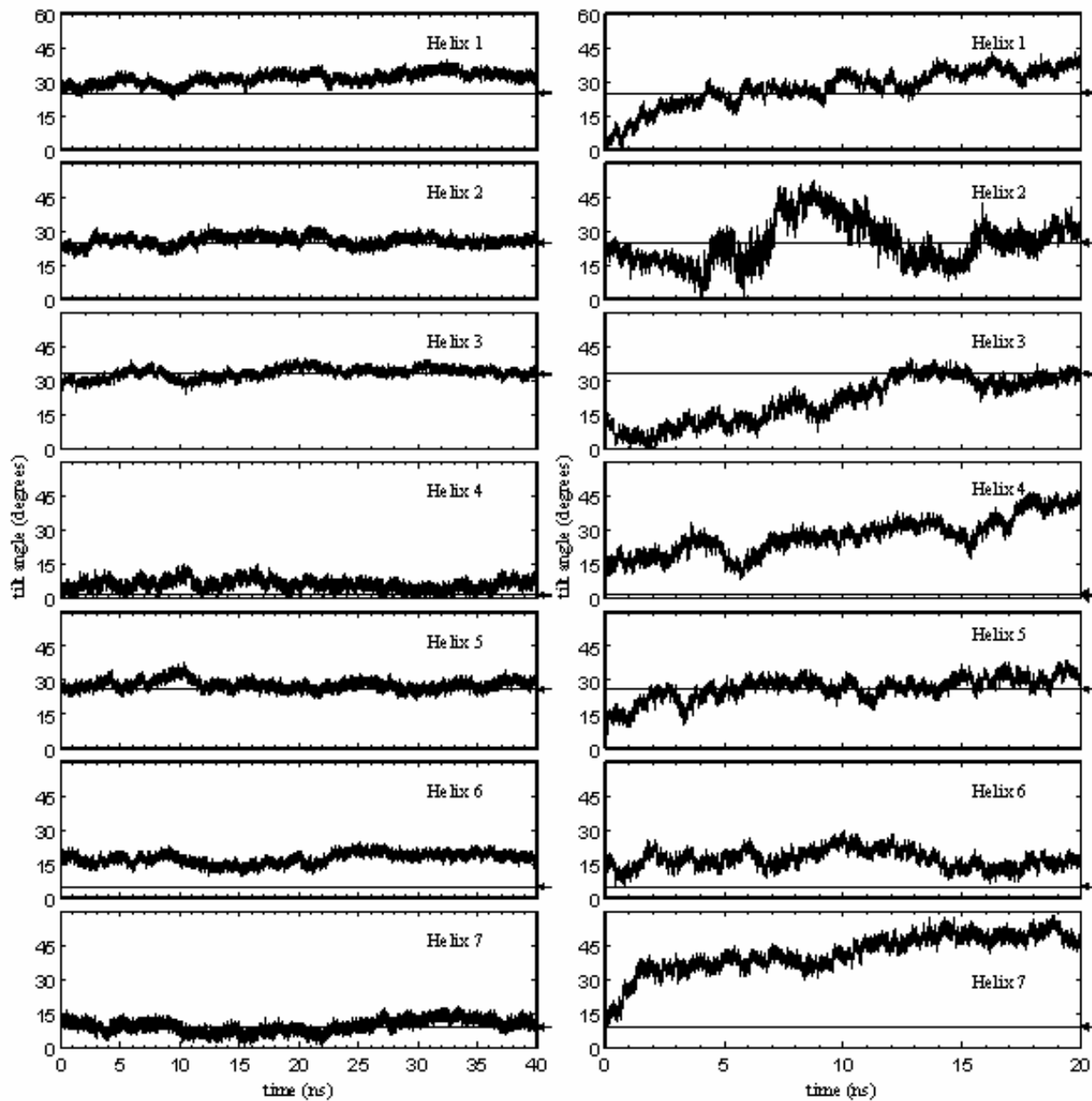


Figure 5.1. Helix tilt angles as a function of time are shown with the intact rhodopsin helices on the left and the isolated helices on the right. Static experimentally-measured tilt angles are marked by the flat lines and by the arrows on the right.

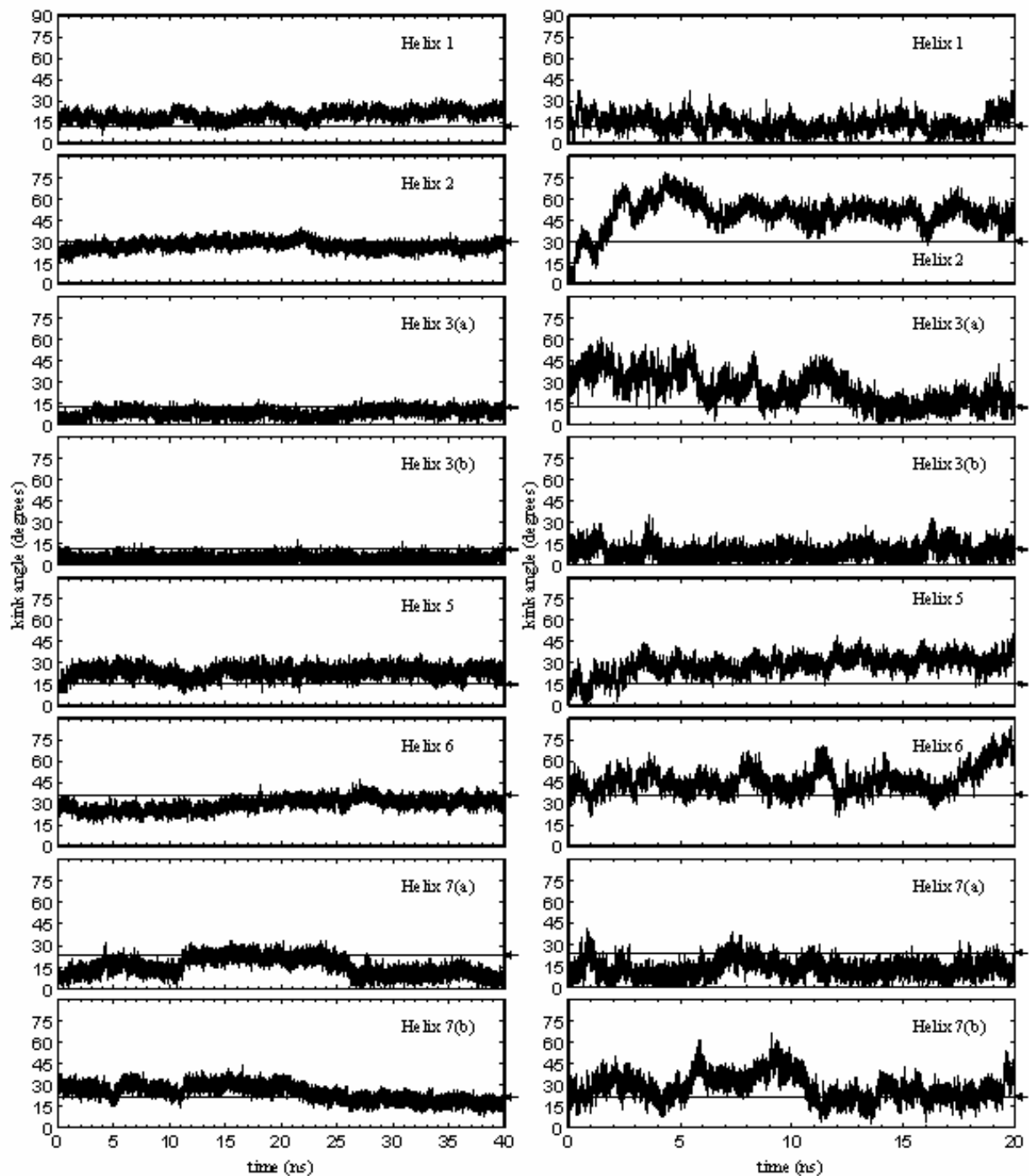


Figure 5.2. Helix kink angles as a function of time are shown with the intact rhodopsin helices on the left and the isolated helices on the right. Static experimentally-measured kink angles are marked by the flat lines and by the arrows on the right. Residues around which kink angles were measured are as follows: Pro 53 for Helix 1, Gly 89 for Helix 2, Gly 120 for Helix 3(a), Ser 127 for Helix 3(b), His 211 for Helix 5, Pro 267 for Helix 6, Pro 291 for Helix 7(a), and Pro 303 for Helix 7(b).

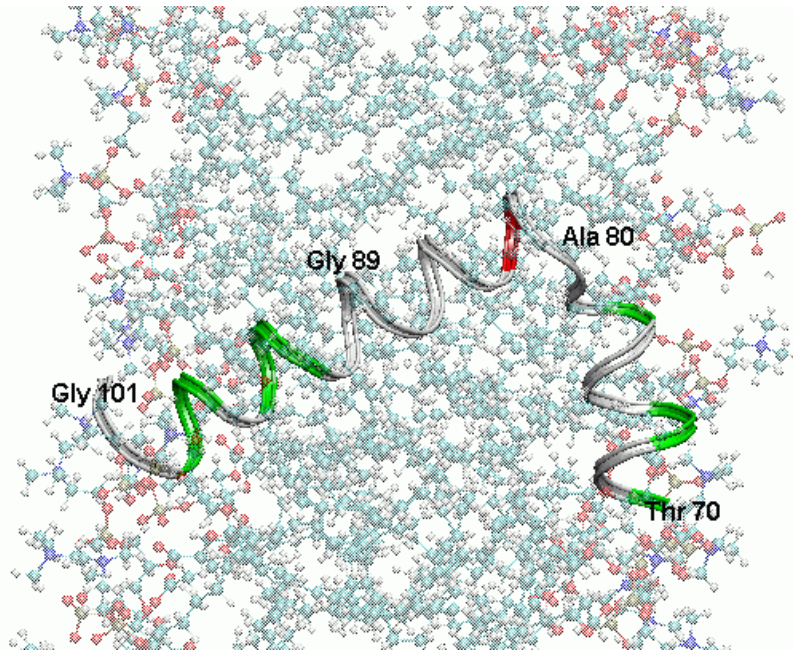


Figure 5.3. Snapshot from the simulation of isolated helix 2 taken at $t = 4.5$ ns showing the large kink angle centered on Ala 80. No kink forms around Gly 89 and Gly 90 as in native rhodopsin. Acidic residue Asp 83 is shown in red. Polar residues are shown in green and nonpolar residues are shown in white. The water has been omitted for clarity. The helix starts in the bottom right corner with cytoplasmic residue Thr 70 and ends on the left with extracellular residue Gly 101.

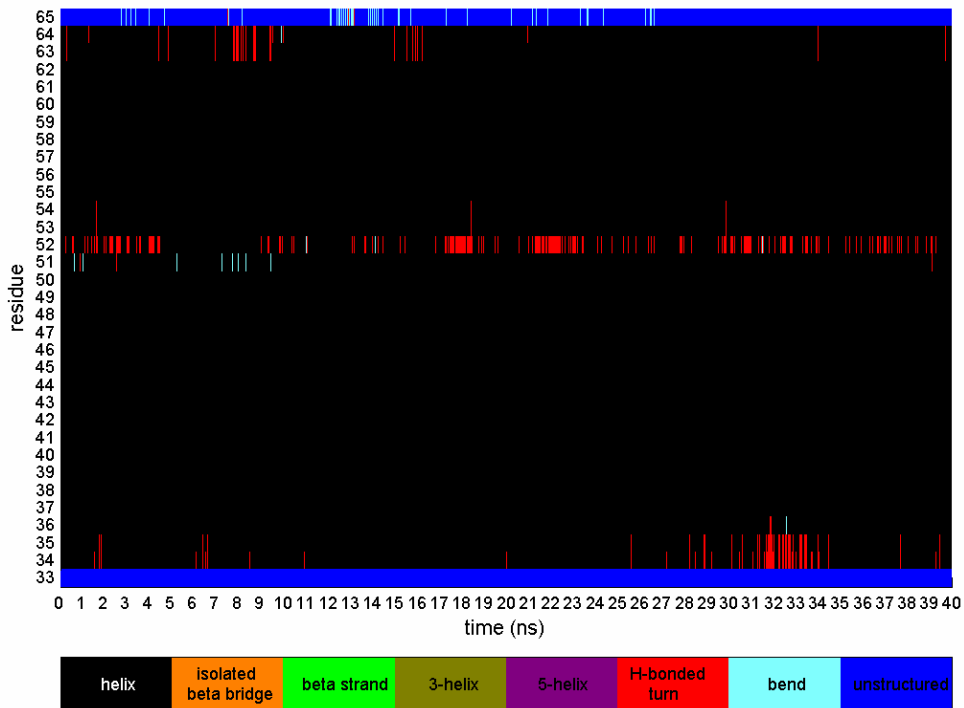


Figure 5.4. Secondary structure time evolution of intact rhodopsin helix 1.

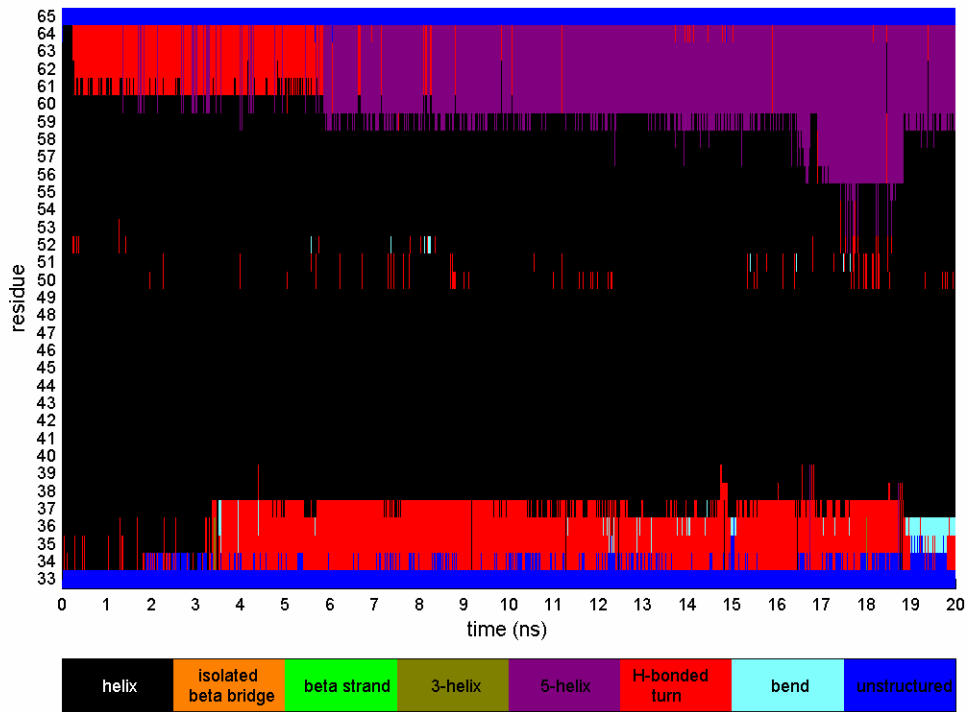


Figure 5.5. Secondary structure time evolution of isolated helix 1.

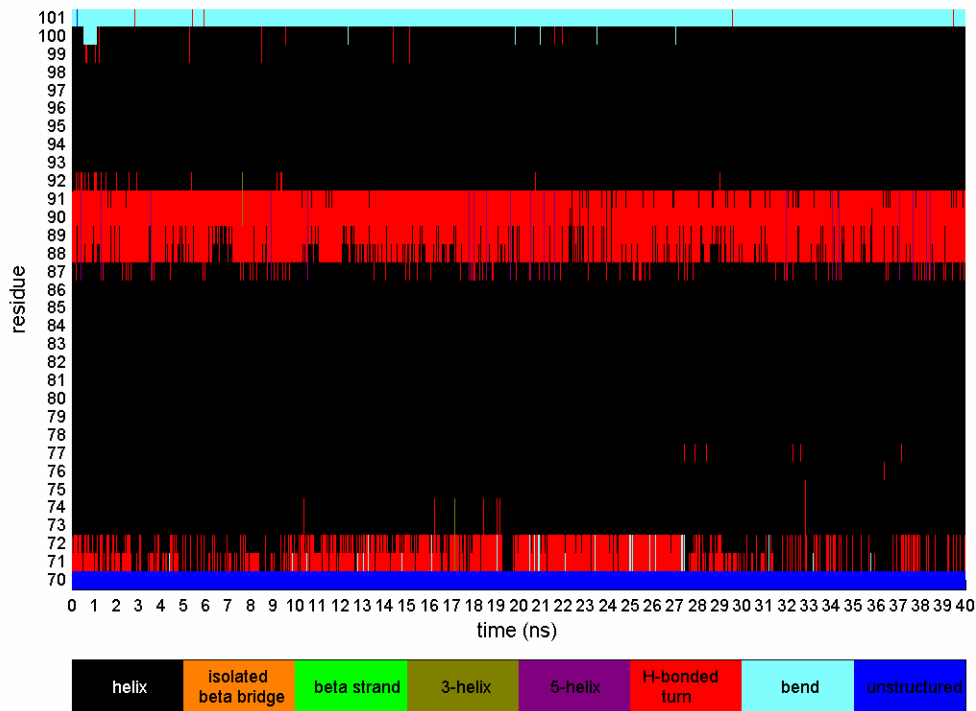


Figure 5.6. Secondary structure time evolution of intact rhodopsin helix 2.

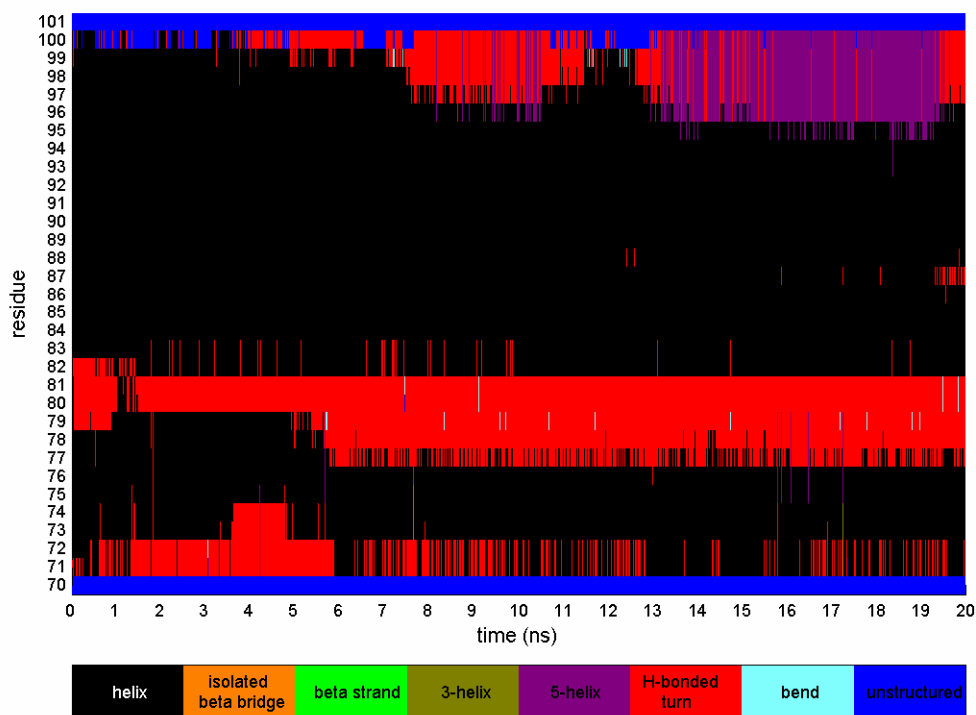


Figure 5.7. Secondary structure time evolution of isolated helix 2.

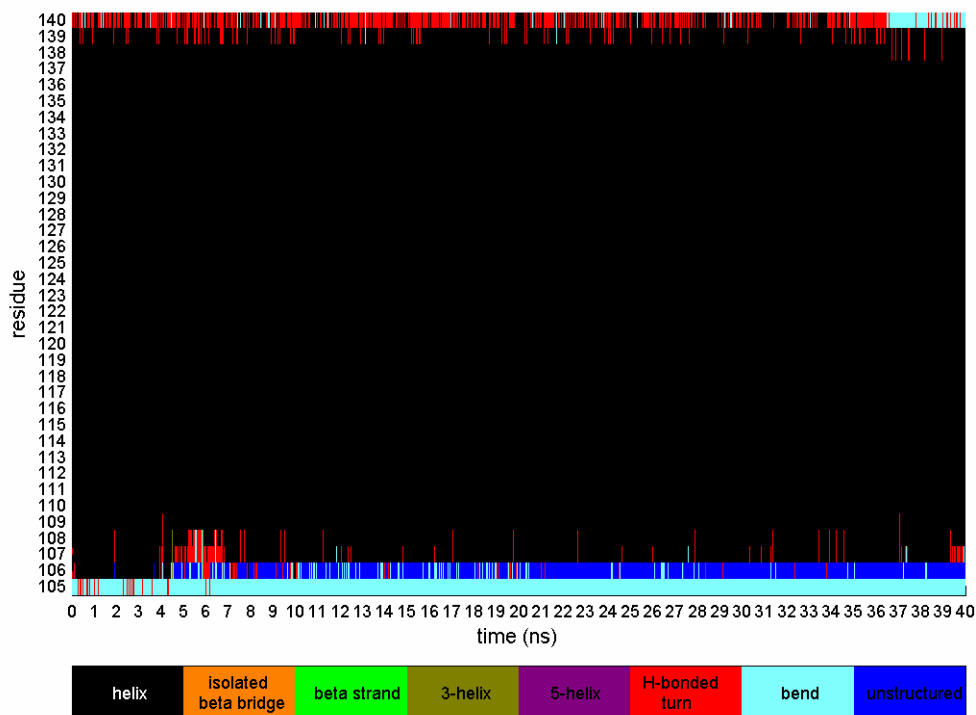


Figure 5.8. Secondary structure time evolution of intact rhodopsin helix 3.

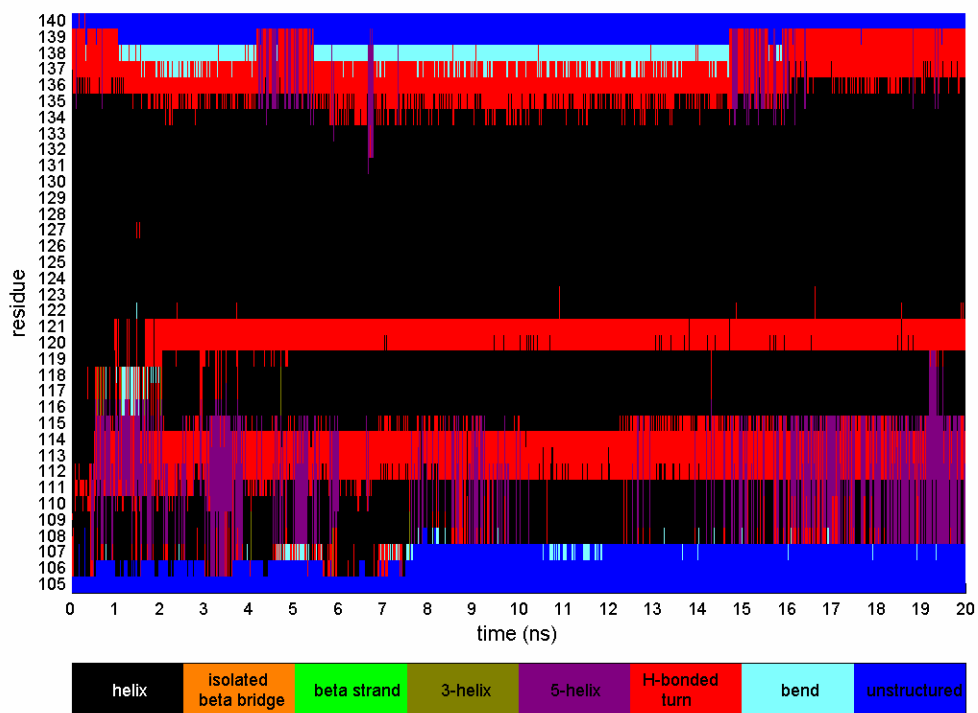


Figure 5.9. Secondary structure time evolution of isolated helix 3.

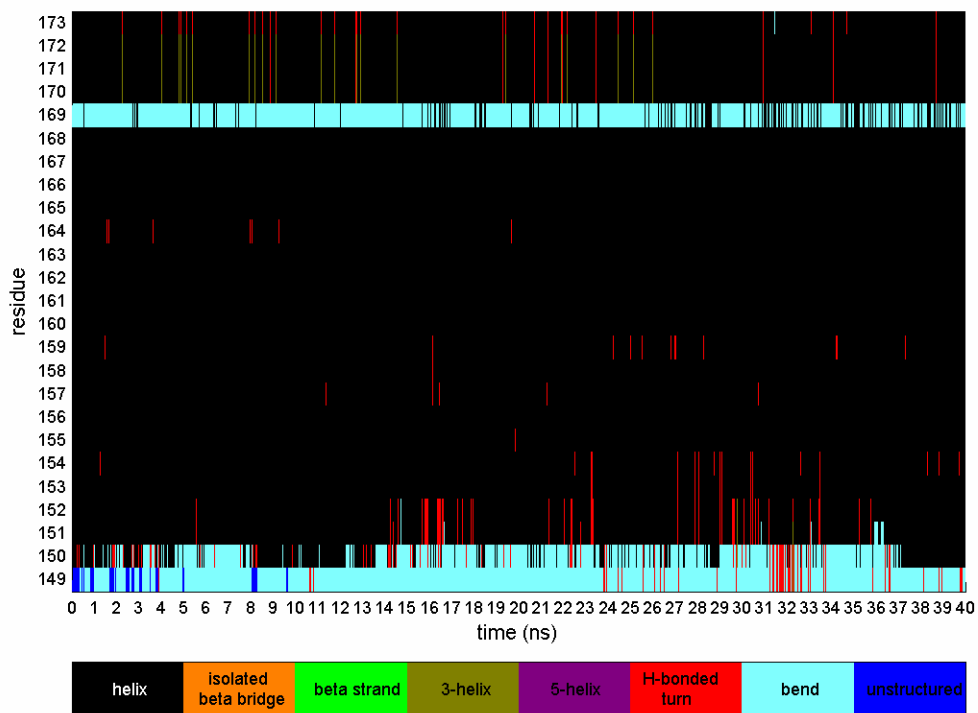


Figure 5.10. Secondary structure time evolution of intact rhodopsin helix 4.

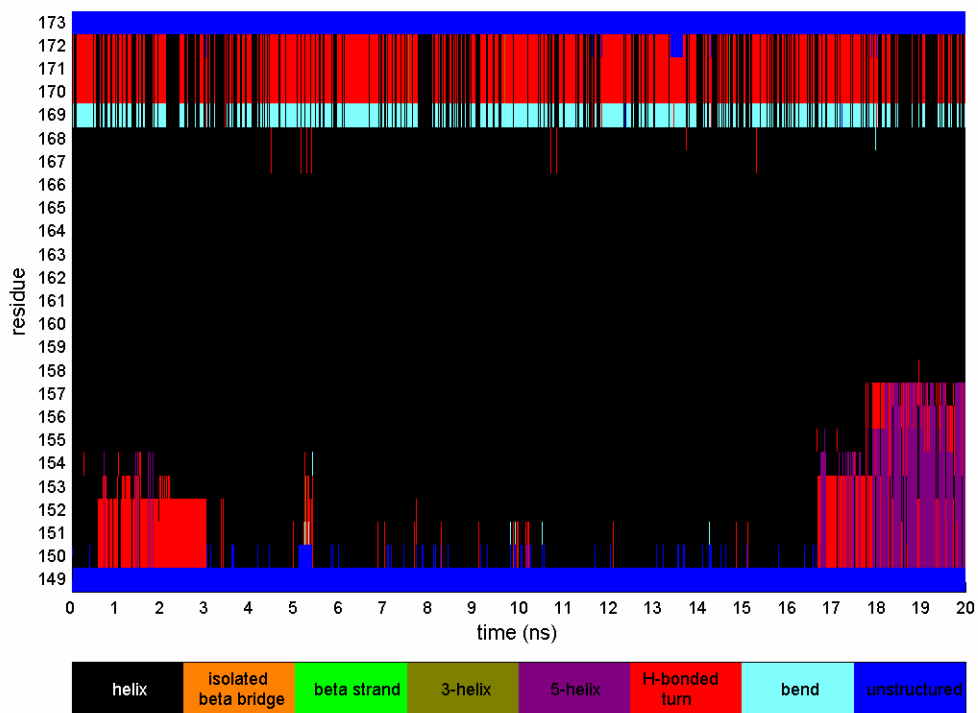


Figure 5.11. Secondary structure time evolution of isolated helix 4.

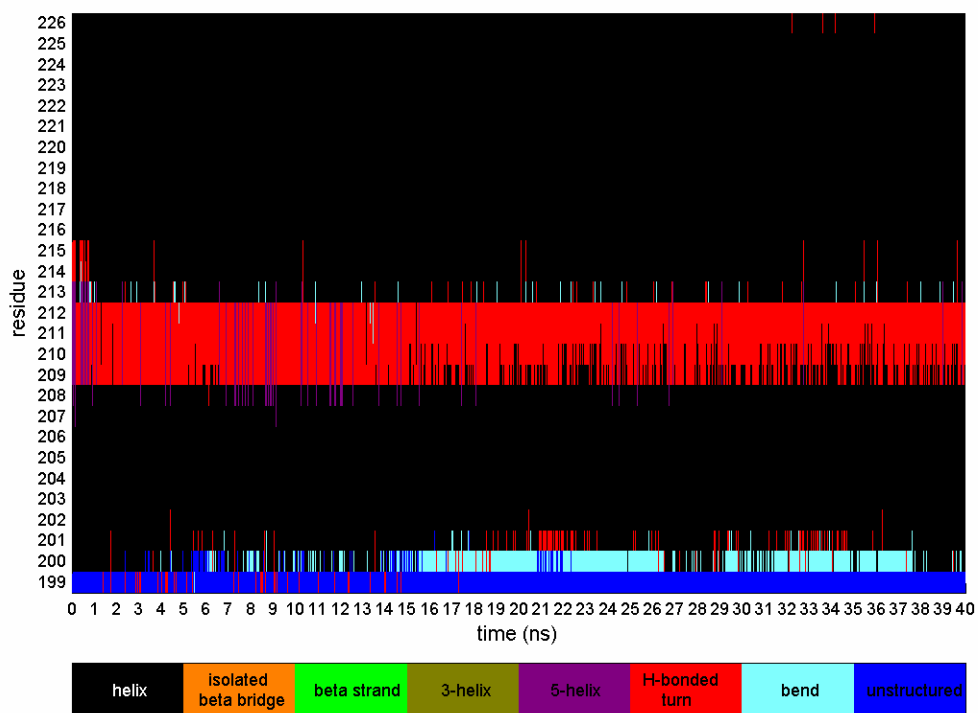


Figure 5.12. Secondary structure time evolution of intact rhodopsin helix 5.

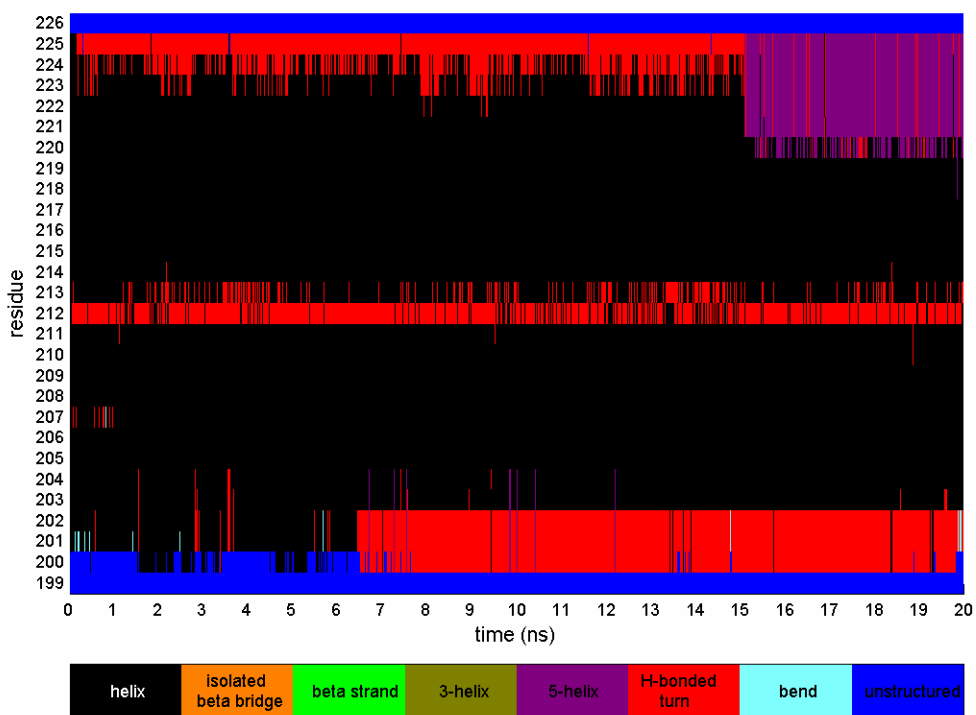


Figure 5.13. Secondary structure time evolution of isolated helix 5.

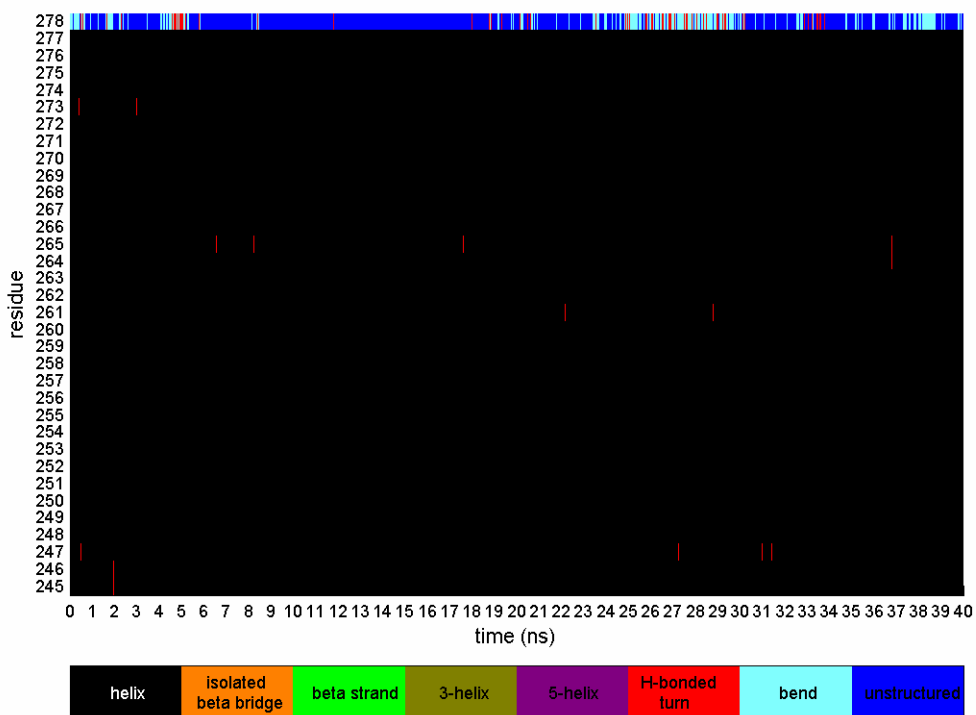


Figure 5.14. Secondary structure time evolution of intact rhodopsin helix 6.

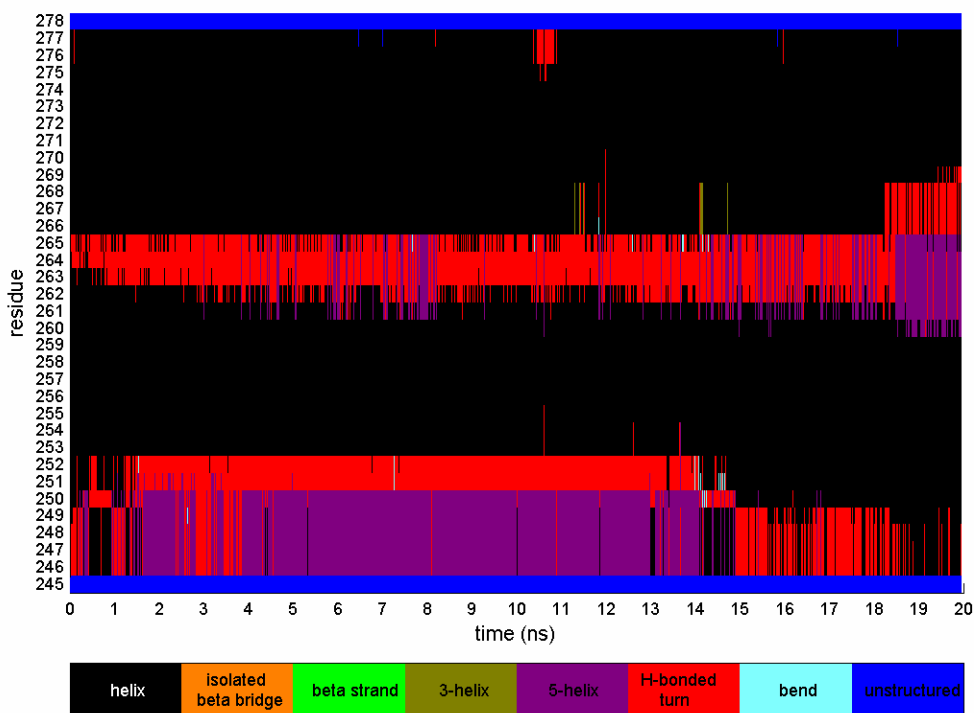


Figure 5.15. Secondary structure time evolution of isolated helix 6.

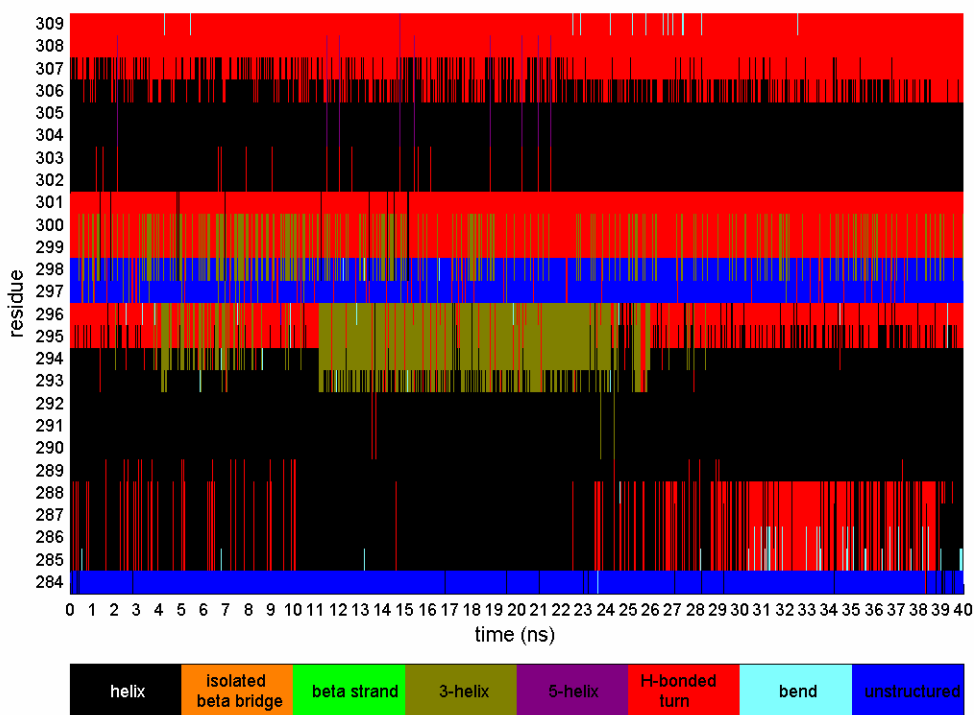


Figure 5.16. Secondary structure time evolution of intact rhodopsin helix 7.

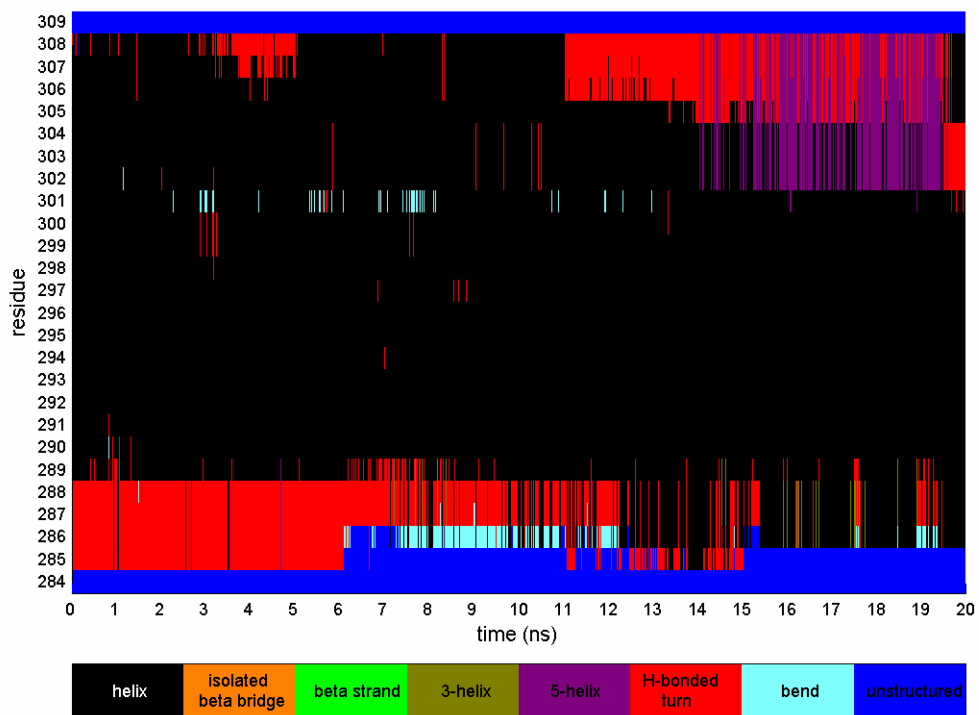


Figure 5.17. Secondary structure time evolution of isolated helix 7.

Section III: Studies of Toxin-Membrane Interactions using Single Molecule Biophysical Methods

This page intentionally left blank.

Chapter 6: Biophysical Studies of Single Nanopores Formed by Protein Toxins Bound to Membranes

Kervin Evans and Alan Burns

This page intentionally left blank.

Introduction

The work presented here describes the electrochemical detection of pores formed in membranes by tetanus neurotoxin. Tetanus toxin is a protein toxin produced by *Clostridium tetani* composed of two polypeptide chains (light chain and heavy chain). It is very structurally similar to other clostridial exotoxins (e.g., botulinum toxin A and B) (Lacy and Stevens 1999). Tetanus toxin is known to bind to cells and lipid bilayers containing ganglioside lipids via the C-terminal fragments of its heavy chain (Singh et. al., 2000), and the heavy chain of clostridial toxins appears to catalyze the leakage of dye from lipid vesicles (Fu and Singh, 1999). It is believed that tetanus toxin acts by binding to GT1b ganglioside molecules on neurons via its heavy chain (Kozaki et. al., 1998), becoming internalized into a cell vesicle, and then inserting the light chain through the cell wall into the cytoplasm when the vesicle becomes acidified (Kalandakanond and Coffield 2001). Electrochemical experiments done on lipid membranes exposed to the pore-forming toxin α -hemolysin have demonstrated that formation of pores single molecules in a membrane could be detected, and that large, single-stranded biological molecules (i.e. RNA, DNA, or proteins) passing through single ion channels would displace the ions flowing through the channels (Kasianowicz et. al., 1996, Akeson, Branton et al. 1999; Meller, Nivon et al. 2000; Movileanu, Howorka et al. 2000). This ionic displacement produces a transient blockage of the measure current through the nanopore and creates a detectable signal. The goal of this study was to establish conditions under which single pores formed by tetanus toxin could be reliably formed and detected (Lebeda and Singh, 1999), in order to be able to study the phenomenon of pore formation and protein translocation through the membrane.

Materials and Methods

Materials

Lipids (1,2-diphytanoyl- *sn*-Glycero-3-phosphatidylcholine [DPh-PC], 1,2-distearoyl-*sn*-Glycero-3-phosphatidylcholine [DSPC],) and cholesterol (chol) were purchased as dried powder from Avanti Polar Lipids, Inc. without further purification. Lipids were generally stored in chloroform at -20°C . Receptor lipids (monosialoganglioside-GT1b [GT1b]) were purchased from Sigma Chemical Company (St. Louis, MO), without further purification. Receptor lipids were stored in a chloroform/methanol mixture at 2:1 ratio.

Water used in all experiments was purified using a Barnstead ultrapure reversed osmosis system to give high resistivity $17.8\text{ M}\Omega\text{-cm}$. All organic solvents (ethanol, hexane, hexadecane, hexane, chloroform, methanol) were purchased from Fisher Scientific Company, ACS grade or better. Also, sodium chloride, potassium chloride, sodium hydroxide, potassium hydroxide, glycerol, cobalt chloride, ethylenediaminetetraacetic acid (EDTA), glycerol, and tris(hydroxymethyl)aminemethane [Tris-HCl] were purchased from Fisher Scientific Company at ACS grade or better. 2-[4-(2-Hydroxyethyl)-1-piperazine]ethanesulfonic acid [HEPES] was purchased from Acros Organics. Sequences of DNA were purchased from Sigma-Genosys (Woodlands, TX) as desalted powders at 0.05- or 0.2- μM concentration. The fluorescent probe calcein was purchased from Molecular Probes, Inc. (Eugene, OR). All chemicals were used without further purification.

Alpha-hemolysin toxin was purchased from Calbiochem (now EMD Sciences, San Diego, CA) or List Laboratories, Inc. (Campbell, CA). The tetanus toxin and its C-fragment were purchased from Calbiochem. No further purification was done.

DNA was stored dry in the freezer at -20°C until needed. For experiments, DNA was rehydrated in 400 μl of buffer (1 mM EDTA, 10 mM Tris-HCl, pH 8 ± 0.05) and kept on ice. Otherwise, rehydrated DNA was kept in the refrigerator at 4°C .

Methods

Single-Toxin-Ion-Channel Chamber & Bilayer Set-Up

The single-ion-channel chamber for toxin nanopores was a Teflon shell with two wells connected by Teflon tubing. The connecting Teflon tubing was a piece of tubing (~ 1 – 1.25 inches long) with shrink-wrap Teflon melted to form a cap (Figure 6.1). The orifice in the shrink-wrap cap (where bilayers were formed) was created by gently pushing a steel stylus (tip was fabricated to 25 μm diameter) through the cap as the shrink-wrap was reheated.

The single ion channel chamber for toxin nanopores was cleaned in boiling nitric acid (~15 - 25% v/v) for approximately 20 minutes prior to each experiment. Once the chamber cooled, it was rinsed three times in nanopure water. The connecting tube was then flushed with 150 μl of water, ethanol, and hexane each. After each solvent flush, air was then drawn through the connecting tube to dry it.

It was found that a bilayer was stable for longer periods of time if a thin film of lipids were dried inside the tubing to “pre-coat” the orifice. Therefore, lipids for coating the tube were prepared by placing 50 μl of stock DPh-PC (10- $\mu\text{g}/\mu\text{l}$) in a culture test tube and gently drying with a stream of nitrogen. These lipids were rehydrated in 250 μl of hexane and capped for later use. Lipids for forming the bilayer were prepared by placing 15 μl onto a glass slide and allowed to air-dry. The connecting tube was coated with lipids by allowing 5 μl of the coating lipids to wick into the tubing. Air was then forced through the tubing to dry the hexane, leaving a thin film of lipids inside the tubing and possibly outside of the cap. After 5 – 15 minutes of further drying, buffer solution was slowly forced through the tubing to fill it and the two wells. A bubble-free tube was confirmed by a square-wave signal when the seal test on the amplifier was turned on (Figure 6.2). Otherwise, the electrical signal due to a bubble in the tubing or across the 25- μm orifice resembled that of a bilayer, transient spikes ~ 600 – 1000 pA in amplitude.

Once the wells on the chamber were filled with buffer, a bilayer was formed across the orifice. To form a thin bilayer, a pipette with a disposable tip attached was dipped into hexadecane just so that the solvent wicked into the tip (1 μl). This aliquot of solvent was placed onto the glass slide away from the dried lipid spot. After wiping the pipette tip to remove excess hexadecane, the tip was dipped into the solvent aliquot to remove a smaller amount and thoroughly mixed with an edge of the dried lipid spot. This was continued until a small amount of lipid and solvent formed a goeey mix and a small amount was transferred onto the tip. Finally, “lipid bubbles” were blown across the orifice until the above square wave became transient spikes (The transient spikes were the discharging signal of the bilayer, which acted as a capacitor. Therefore, the thinner the bilayer became the larger the capacitance became.). Typically, thin bilayers had transient spikes that peaked 2000 ± 200 pA (anything less was usually an indication that the organic solvent used to “paint” the bilayer did not get compress out of the middle of the bilayer, therefore leaving a “thick” bilayer) (Figure 6.3). Bilayers were allowed to set 5 – 15 minutes before adding toxin to form a single nanopore.

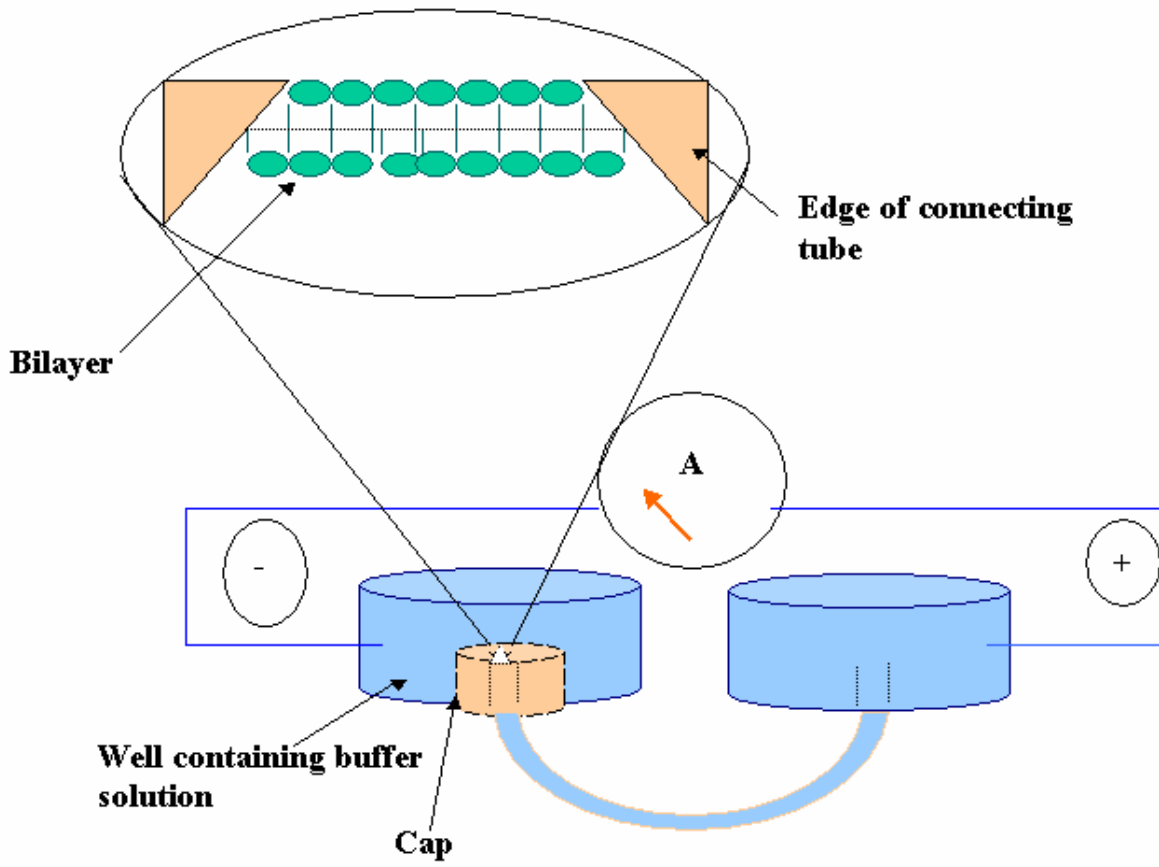


Figure 6.1. General set up of single-ion-channel chamber. Teflon tubing with a shrink-wrapped cap connects two separate reservoirs. At the end of the cap is the orifice where bilayers are formed. The cis reservoir is on the left; the *trans* one is on the right.

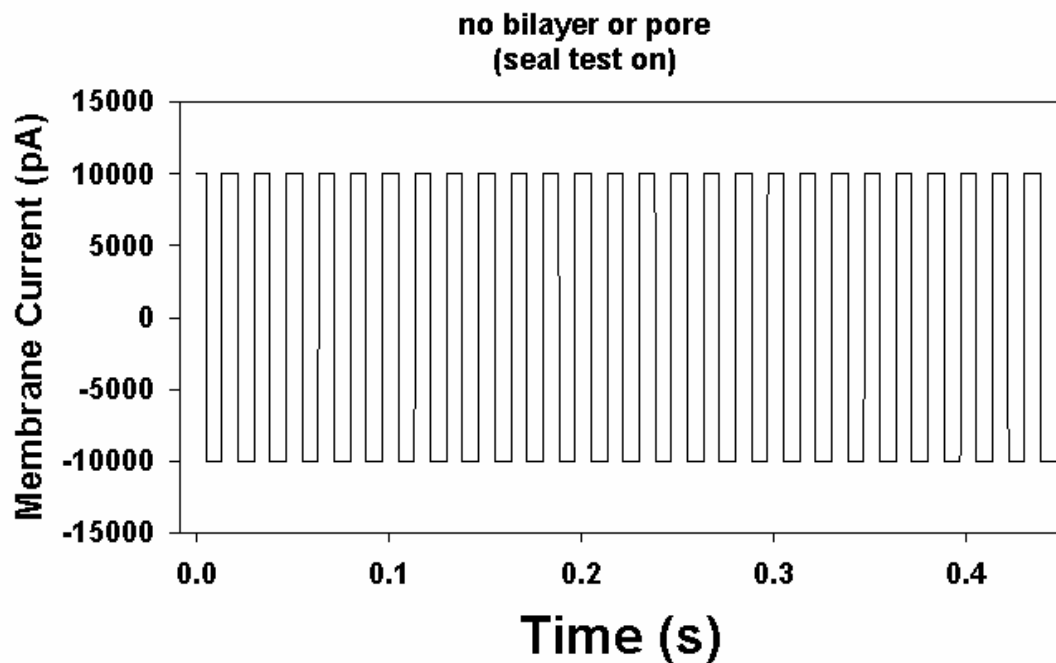


Figure 6.2. This is the electrical signal during the seal test when no bilayer or air bubble was present. The seal test was run at 5 mV / 8 ms.

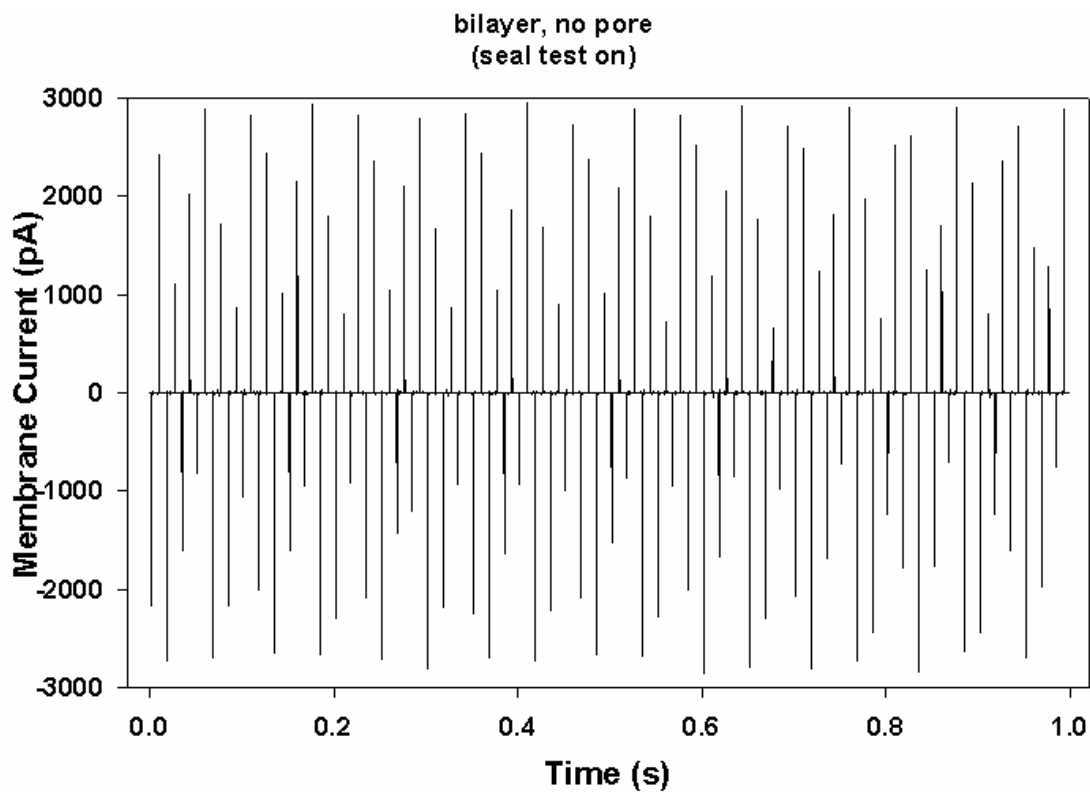


Figure 6.3. Electrical signal of a thin DPh-PC bilayer at +120 mV. The square wave from the seal test becomes transient spikes $\geq 2000 \pm 200$ pA when a bilayer thins over the orifice of the single-ion-channel chamber.

Alpha-hemolysin was stored in 1:1 glycerol/water mix at 0.5 $\mu\text{g}/\mu\text{l}$ prior to use. For each experiment, 10 μl of toxin was aliquoted to an eppendorf tube, mixed with 90 μl of 1 M KCl, 10 mM HEPES at $\text{pH } 8 \pm 0.05$ to give an experimental concentration of 0.05 $\mu\text{g}/\mu\text{l}$, and stored on ice during the experiment. The buffer was the same for all DNA translocation experiments (1 M KCl, 10 mM HEPES, $\text{pH } 8$). Toxin was added typically 1 - 2 microliters at a time to the same well where the bilayer was formed (cis well). The current was monitored in DC mode. Typical currents before nanopore formation were < 3 pA. Once a single nanopore of α -hemolysin toxin was established (indicated by a sudden increase in current to ~ 120 pA) in a bilayer of diphytanoyl phosphatidylcholine (DPh-PC), the excess toxin was flushed out with 3 – 5 ml of buffer. Established single nanopores were allowed to settle for 5 minutes before single stranded DNA was introduced into the cis-well (negative electrode) for translocation studies. In a typical experiment, DNA was added at a final concentration of 2.5- μM .

For tetanus toxin experiments, the bilayer was either DPhPC/GT1b @ 95:1, or DSPC/chol/GT1b @ 47.5:47.5:5 mole percents. The buffer contained 150 mM NaCl, 5 mM citrate-HEPES-MES, at $\text{pH } 7.4$ (or 4) ± 0.05 .

With the cis well negative, typically 120 mV was applied such that a positive current was detected when a nanopore opened. Current measurements were measured using a patch-clamp amplifier and head-stage (Axopatch 200B and CV203BU, Axon Instruments, Foster City, CA). The amplified signals were low-pass filtered at 5 kHz and digitized (at 250 kHz for α -hemolysin, 20 kHz for tetanus) with a Digidata 1320A from Axon Instruments.

Results

Tetanus Toxin Nanopore Activity

As we explored the pore formation of tetanus toxin, we wanted to know whether a lipid system containing the ganglioside GT1b would readily form a stable bilayer. Our first attempts were done with the lipid components of 95-mol% DPh-PC and 5-mol% GT1b (DPh-PC was a lipid system that we have made thin bilayer with great success and ease). We then tried DSPC/chol/GT1b at 47.5/47.5/5 mol% (lipid system chosen to keep conditions the same as those in AFM experiments). Both bilayer systems formed stable, thin (transient spikes ~ 2000 pA) bilayers for 30 – 60 minutes in 150 mM NaCl, at $\text{pH } 7.4$ (and $\text{pH } 4$) and 120 mV. The bilayer system containing DPh-PC formed a thin bilayer within 15 minutes, using the standard solvent hexadecane. However, bilayers of DSPC/cholesterol/GT1b took longer (30 – 60 minutes) to form, and required a solvent mixture of decane and hexadecane.

Early atomic force microscopy images (Andrea Slade, unpublished) indicated that tetanus C-fragment (Tet-C, the portion of the heavy chain that contains the binding site for GT1b, ~ 50 kD in size) bound to the bilayer of dipalmitoyl phosphatidylcholine/dipalmitoleoyl phosphatidylcholine/GT1b @ 45/45/10 mol% at $\text{pH } 7.4$ after ~ 2 hours of incubation. And defects (~ 40 - 80 nm in diameter) formed in the bilayer after ~ 12 hours. Therefore, our next step was to evaluate the pore formation activities of Tet-C in DPh-PC/GT1b (95:5 mol%) bilayer. Initial current measurements showed that at least one pore (most likely several pores) formed about one hour and 45 minutes after adding ~ 0.65 μg Tet-C (4.06 $\mu\text{g}/\text{ml}$; we used 13-times the amount required for the α -hemolysin experiments because a single tetanus pore is much smaller, $\sim 1 - 8$ pA/tetanus pore vs 120 pA/ α -hemolysin pore) to the cis-well. This was unexpected since all of the literature on pore formation of Tet-C suggested that Tet-C did not form any nanopore structures. Therefore, we analyzed our samples of Tet-C for purity by gel analysis. Size exclusion gel analysis showed that our stock Tet-C contained light bands of contaminants approximately in the size range of the heavy chain of tetanus (~ 100 kD) and the full tetanus toxin (~ 150 kD) (Figure 6.4). This very well would explain the unexpected results of pore formation with the Tet-C sample. Small amounts of tetanus heavy chain and/or full toxin in a 20- μg stock solution of Tet-C are more than enough contaminants to form pores (it takes 20 – 50 ng of α -hemolysin to form a pore).

We thus evaluated the pore forming properties of the full tetanus toxin in the DSPC/cholesterol/GT1b and DPh-PC/GT1b bilayers as a function of pH. For the DSPC/cholesterol/GT1b bilayer we found that at 125 ng/ml and +120 mV bursts (not continuously open) of pores formed with a maximum current around 50 pA (Figure 6.5). As the concentration was increased to 250 and 500 ng/ml the frequency of pore formation (bursts) reduced over a 30-minute period. This implied that tetanus might form aggregates in solution. Dynamic light scattering data showed that tetanus particle size increased with increased concentration, further suggesting that tetanus aggregated with increased concentration. The particle size of tetanus was also pH-dependent where at or above 1.67 $\mu\text{g/ml}$ the aggregate size at pH 4 was nearly twice that of aggregates at pH 7.4 (Figure 6.6). However, no pores formed at pH 4.

Upon adding tetanus at $\sim 5 \mu\text{g/ml}$ to DPh-PC/GT1b (95/5 mol%) at pH 7.4, pores readily formed. The pores were opening and closing more frequently than at the lower concentration in the DSPC/cholesterol/GT1b bilayer system. At 120 mV and 150 mM NaCl, the pores had a maximum current peak around 120 pA. As in the other bilayer system, there were no pore formations at pH 4 in the DPh-PC/GT1b system (Figure 6.7).

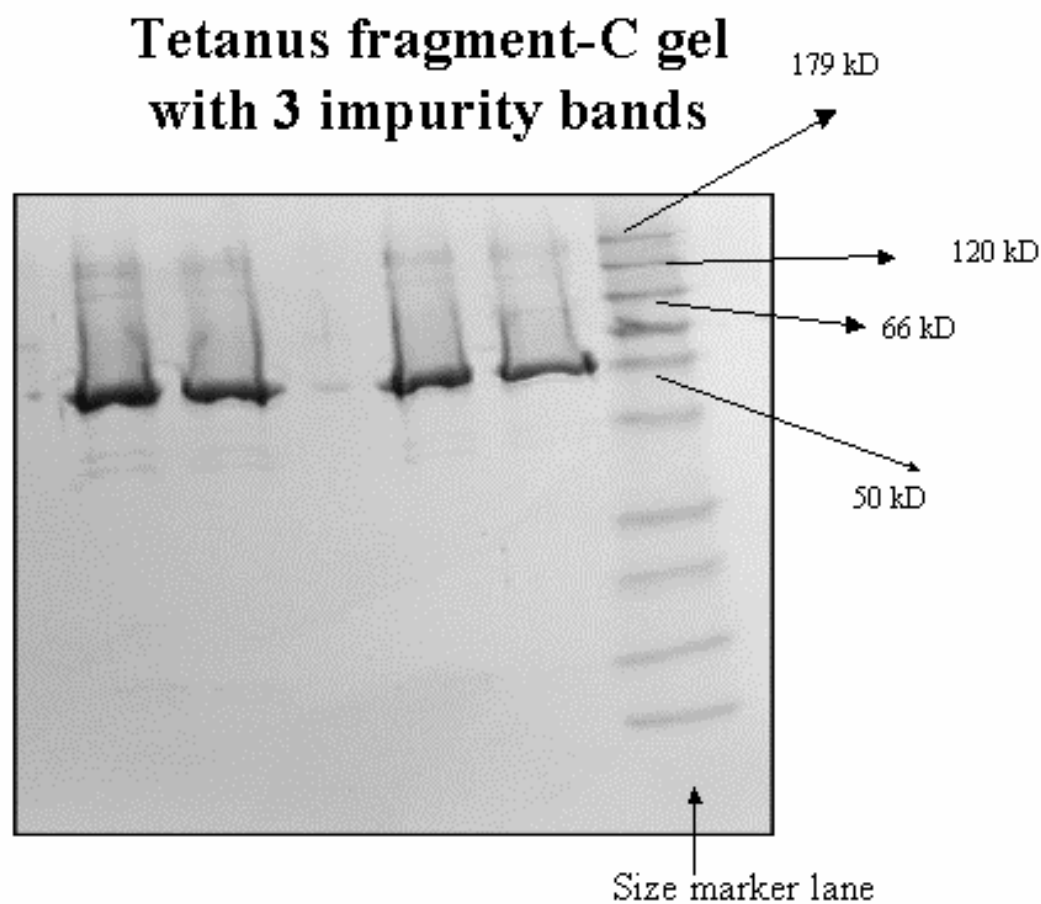


Figure 6.4. SDS-Gel of the tetanus C-fragment showed contamination of the stock solution. Tetanus C-fragment is the dark band around 50 kD. The other two bands are near the size of the heavy chains (~ 100 kD) and full toxin (~ 150 kD).

Tetanus Toxin Pores in DSPC/cholesterol/GT1b
(47.5/47.5/5 mol%) Bilayer @ pH 7.4

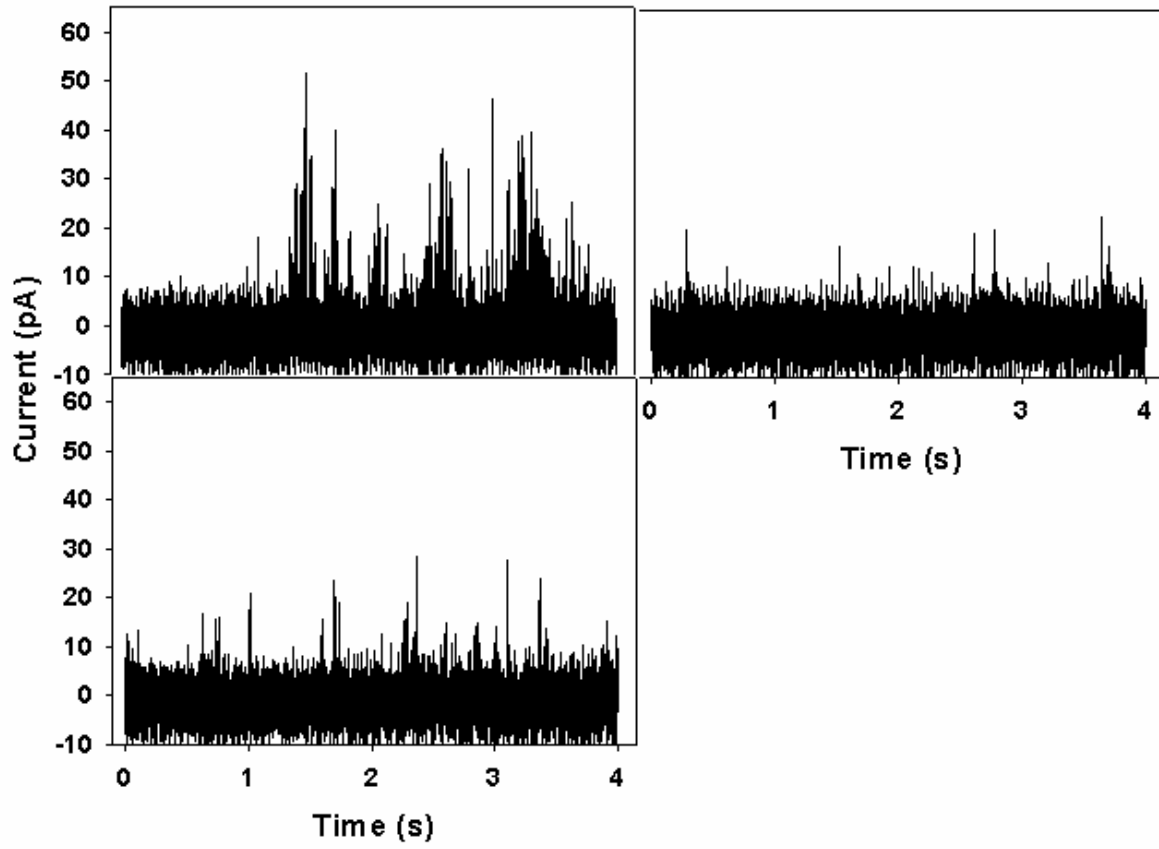


Figure 6.5. Tetanus toxin formed pores in DSPC/cholesterol/GT1b (47.5/47.5/5 mol%) bilayer. Pore activity appeared to decrease with added toxin concentration.

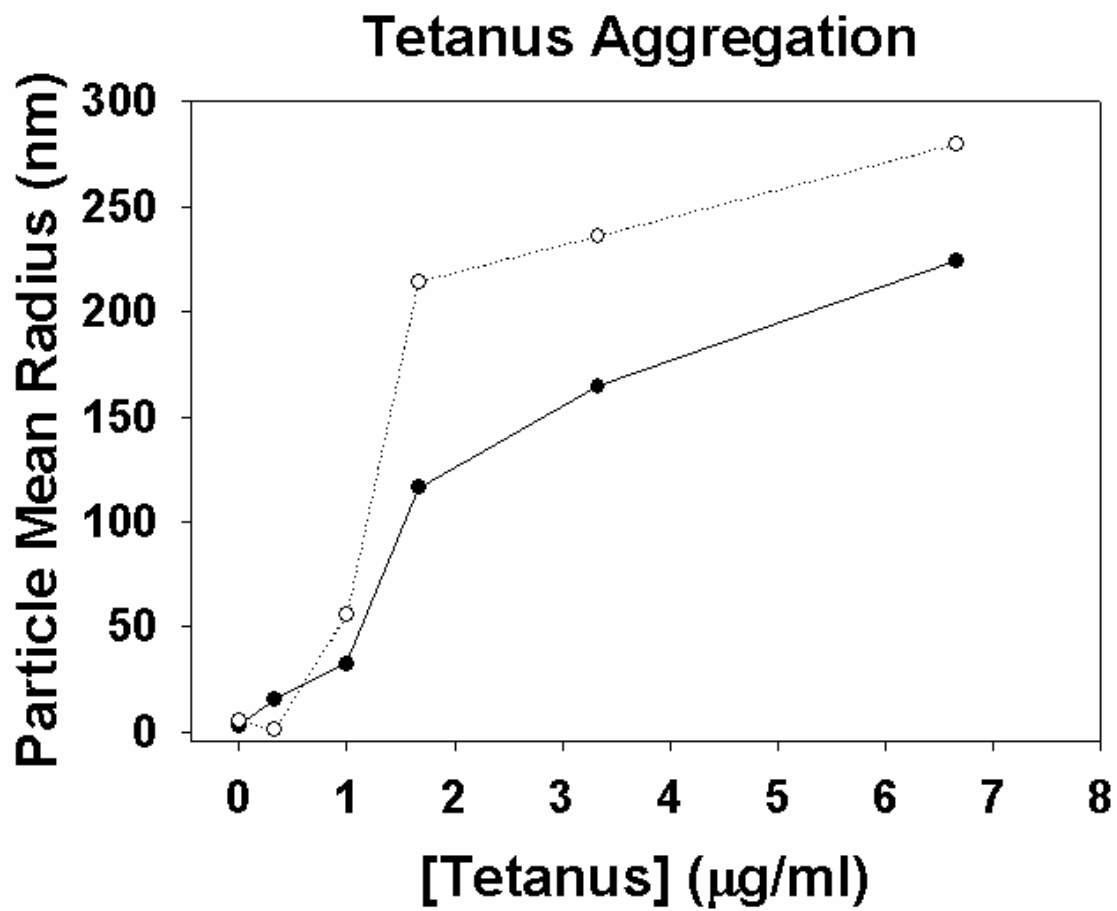


Figure 6.6. Tetanus aggregation is pH-dependent. Tetanus aggregates in solution at 1.67 $\mu\text{g/ml}$ were twice as big at pH 4 than at pH 7.4.

Tetanus Toxin in DPh-PC/GT1b (95/5 mol%) Bilayer @ pH 7.4

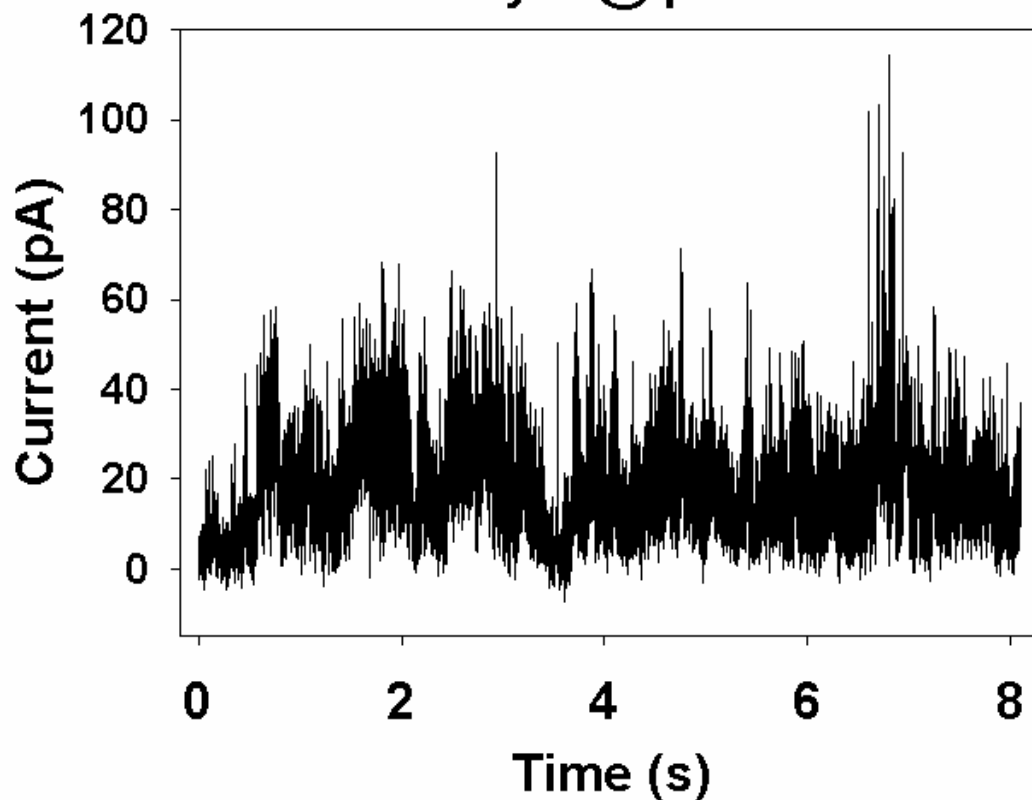


Figure 6.7. Tetanus toxin formed pores more readily at 5 $\mu\text{g/ml}$ in DPh-PC/GT1b (95/5 mol%) bilayer at pH 7.4. Current peaked around 120 pA at +120 mV and 150 mM NaCl.

Summary

We have the capabilities to study tetanus in the presence of leak-free bilayers and as a function of physiological conditions. We have demonstrated that tetanus does form pores in two different bilayer systems at pH 7.4, but not pH 4.

References

1. Akesson, M., D. Branton, et al. (1999). Microsecond Time-Scale Discrimination among Polycytidylic Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments within Single RNA Molecules. *Biophys. J.* 77(6): 3227-3233.
2. Fu, F. N. and B. R. Singh (1999). Calcein permeability of liposomes mediated by type A botulinum neurotoxin and its light and heavy chains. *Journal of Protein Chemistry*; 18(6): 701-707.
3. Kalandakanond, S. and J. A. Coffield (2001). Cleavage of SNAP-25 by botulinum toxin type A requires receptor-mediated endocytosis; pH-dependent translocation; and zinc. *Journal of Pharmacology and Experimental Therapeutics*; 296(3): 980-986.
4. Kasianowicz, J. J., E. Brandin, et al. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America*: 13770-13773.
5. Kozaki, S., Y. Kamata, et al. (1998). Ganglioside GT1b as a complementary receptor component for Clostridium botulinum neurotoxins. *Microbial Pathogenesis*; 25(2): 91-99.
6. Lacy, D. B. and R. C. Stevens (1999). Sequence homology and structural analysis of the clostridial neurotoxins. *Journal of Molecular Biology*; 291(5): 1091-1104.
7. Lebeda, F. J. and B. R. Singh (1999). Membrane channel activity and translocation of tetanus and botulinum neurotoxins. *Journal of Toxicology-Toxin Reviews*; 18(1): 45-76.
8. Meller, A., L. Nivon, et al. (2000). Rapid nanopore discrimination between single polynucleotide molecules. *PNAS* 97(3): 1079-1084.
9. Movileanu, L., S. Howorka, et al. (2000). Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nature Biotechnology*: 1091-1095.
10. Singh, A. K., S. H. Harrison, et al. (2000). Gangliosides as receptors for biological toxins: Development of sensitive fluoroimmunoassays using ganglioside-bearing liposomes. *Analytical Chemistry*; 72(24): 6019-6024

Chapter 7: Domain structure in model membrane bilayers investigated by simultaneous atomic force microscopy and fluorescence imaging

A. R. Burns

Abstract

Simultaneous atomic force microscope (AFM) and submicron confocal fluorescence imaging of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) lipid domain structures in 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) is presented. Lipids labeled by fluorescent probes either at the headgroups or tailgroups enable domain contrast in fluorescence imaging on the basis of partitioning between the gel (DPPC) and disordered liquid (DOPC) phases. However, correlation with AFM topographic information reveals that they do not always faithfully report exact gel domain size or shape. Furthermore, we find that the fluorescence contrast decreases significantly with domain size, such that small domains observed with AFM are not observed in fluorescence images despite adequate optical resolution. We attribute these effects in part to broadened partitioning of the probe lipids across the domain boundaries. Binding of fluorescent Alexa 488-conjugated cholera toxin B subunits to GM1 gangliosides in DPPC domains correlates well with AFM topographic information to the limit of optical resolution. However, it also may reveal the presence of dilute GM1 components in the fluid phase that have no topographic contrast. In all cases, the complete correlation of topographic and fluorescence images provides evidence that gel-phase domains occur across both leaflets of the bilayer.

Introduction

Supported lipid bilayers have been studied extensively as models for biological membranes¹⁻³ because of the ease by which they can be characterized by numerous techniques, particularly imaging microscopies. A 1-2 nm layer of water¹ between the solid support (usually glass or mica) and the lipids allows freedom of movement of both leaflets, thus preserving integrity of the bilayer as a membrane to a good approximation. Recently, there has been considerable activity in the study of distinct phase domains in multicomponent lipid bilayers and monolayers, both supported⁴⁻¹² and unsupported¹³⁻¹⁸, as models for “lipid rafts,” believed to be important in cellular signaling processes¹⁹⁻²⁴. Briefly, trans-membrane and membrane-associated proteins involved in cellular signaling often have been shown to be components of densely-packed, detergent-resistant gel-like domains of saturated lipids within the membrane^{19,21,22}. The role of these “rafts” is thought to be one of binding or sequestering specific proteins for specialized functions within the membrane¹⁹⁻²⁴. This association of proteins with lipid domains may in turn be mediated by much smaller “shells” of lipids surrounding the protein²⁵. Thus many studies of lipid domains in model systems have focused on phase separation within multicomponent lipid mixtures, on the basis of head group interactions and/or acyl chain structure, and the subsequent partitioning of protein-binding glycosphingolipids between the phases.

Domain structure in bilayer and monolayer assemblies has been characterized by numerous fluorescence microscopy methods including far-field^{14,15,17,26} and near-field imaging²⁷⁻²⁹, fluorescence recovery after photobleaching^{10,13}, and fluorescence quenching^{16,18}. Indeed, these techniques have been useful for domain studies on cellular membranes as well^{19,30-34}. Fluorescent labels are usually placed on specific lipids that subsequently partition between the gel-phase (or liquid-ordered) raft domains and the liquid-disordered domains, although labeling of interacting proteins is also used. It is difficult to minimize the influence of the label on the lipid interaction with its neighbors; thus the location of the label, on the head group or acyl chains, as well as acyl chain length¹⁸, can lead to significant differences in partitioning.

Atomic force microscopy (AFM) has been used to image organizational structure in supported lipid monolayers and the top leaflet of bilayers on the basis of topographic height differences between gel-phase and liquid-disordered domains^{4,5,7,8,12}. The origin of this height difference most likely arises because the “taller” gel-phase domains of saturated acyl chains are more tightly packed relative to liquid disordered domains (usually lipids with unsaturated acyl chains), thus they occupy a smaller area per molecule and are oriented more normal to the substrate⁴. Microdomains of glycosphingolipids, such as GM1, have been resolved on the basis of their large head groups^{5,7,8,12}. Like fluorescence imaging, AFM can be routinely performed with the model membranes in fluid environments, thus allowing lipid mobility and full interaction with solutes such as proteins. It has also been well established that AFM can provide undistorted molecular-scale lateral (< 2 nm) and vertical (<0.1 nm) resolution of compliant membrane structures, due in large part to sensitive force feedback detection (< 1 nN) and sharp probes (<20 nm tip radius)³⁵⁻³⁷. As with all scanning probes, AFM is primarily useful for imaging relatively “static” features that are not rapidly diffusing or changing shape.

The ability to correlate information gathered from fluorescence imaging on supported membranes in a fluid environment with detailed lateral structures mapped out with AFM would be highly advantageous. In a step towards this goal, the utility of performing simultaneous AFM and submicron confocal fluorescence imaging is demonstrated on lipid bilayers consisting of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) gel-phase domains in a 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) liquid phase. It is shown that since the fluorescent labels are subject to partitioning between the gel (DPPC) and disordered (DOPC) lipid phases, they do not always faithfully report exact gel domain size or shape. Furthermore, the fluorescence contrast decreases significantly with domain size, such that small domains observed with AFM are not observed in fluorescence images despite adequate optical resolution. Fluorescent markers do, however, reveal the presence of dilute components in the fluid phase that have no topographic contrast.

Materials and methods

Lipids and protein

1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), and GM1 ovine brain ganglioside (GM1) were purchased from Avanti Polar Lipids and used without further purification. N-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionyl)-1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (BODIPY-DPPE, also known as BODIPY-DHPE), N-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionyl)-C5-ganglioside GM1 (BODIPY-C5-GM1), and Alexa 488-conjugated cholera toxin B subunit (Alexa 488-CTX-B) were purchased from Molecular Probes and used without purification unless otherwise indicated.

Supported bilayers

Lipid bilayers are formed on glass coverslips by the method of vesicle fusion^{2,3}. Single unilamellar vesicles are prepared by first dissolving the lipids in chloroform, followed by rotary evaporation of the solvent and drying for over 12 hours under high vacuum. The lipids are re-suspended by adding phosphate-buffered saline (PBS) buffer (100 mM NaCl, 40 mM sodium phosphate, pH 7.4) and vortexing, followed by de-gassing with nitrogen. The mixtures are tip-sonicated for 4-minute cycles over 20 minutes while being cooled with an ice-bath. The resulting vesicle suspensions are centrifuged at 15,000 rpm, with the supernatant filtered through a 0.2 μm syringe filter. Dynamic light scattering characterization (Protein Solutions) usually indicates vesicles diameters are in the range of 50-100 nm.

Glass coverslips (0.13-0.17 mm thickness) are cleaned in (7:3) $\text{H}_2\text{SO}_4/\text{H}_2\text{O}$, (*caution: this is potentially explosive when reacting with organics*), rinsed thoroughly in distilled and ultra pure water (Barnstead Nanopure), and stored under ultra pure water (18M Ω -cm). Just prior to use, the coverslips are dried under a stream of pure, dry nitrogen and mounted in a Leiden coverslip dish (Harvard Apparatus). The ~ 3 mM vesicle solution is pipetted onto the coverslip and diluted (1:5) with imaging PBS buffer (150 mM NaCl, 50 mM sodium phosphate, 1.5 mM NaN_3 , pH 7.4). After two hours of incubation at 60°C, followed by cooling to room temperature, the lipid bilayer is rinsed thoroughly with imaging PBS buffer and mounted on the microscope. When required, incubation of the lipid bilayers with Alexa 488-CTX-B stock solution (10 $\mu\text{g}/\text{ml}$, in imaging PBS buffer) is performed for 1-2 minutes, followed by thorough rinsing with imaging PBS buffer. Prior to use, the stock solution was passed through chromatography columns (Sephadex G25 or Bio Rad Micro Bio Spin 6) to reduce the amount of unconjugated Alexa 488.

AFM and fluorescence microscopy

The experimental apparatus for obtaining simultaneous AFM and fluorescence images of lipid bilayers is shown schematically (not to scale) in Figure 7.1. The instrument is based on an inverted microscope (Olympus IX-70) platform modified to accommodate an AFM scan head (Digital Instruments Bioscope G scanner). The microscope is mounted on a vibration isolation air table that is also acoustically- and light-isolated. The liquid cell (coverslip dish) is mounted on a flat-plate XY scanner (Nanonics) that allows full optical access below and AFM access above the sample. Excitation light from a continuous 488 nm Ar^+ laser (on a separate table) is coupled into a single-mode optical fiber that ports the light to the microscope table and also serves as a spatial filter. A fiber coupler (FC1 in Figure 7.1) then forms a collimated Gaussian beam that reflects off a dichroic mirror (Chroma Q495LP) and fills the back plane of a 100X (1.3 NA) oil immersion objective. The objective focuses the light ($< 2 \mu\text{W}$) to a 250-300 nm spot size spatially coincident on the sample with the pyramidal tip on the end of the AFM cantilever, as can be seen through the microscope eyepiece. This alignment remains fixed as the sample is scanned. Epi-fluorescence emission is collected by the same objective, passes through the dichroic mirror, through two notch filters (488 and 670 nm, Kaiser), and through a 500-580 nm band pass filter (Chroma HQ 540/80). The

fluorescence is then spatially filtered in a confocal manner by coupling (FC2 in Figure 7.1) into a 50 μm diameter core multimode fiber connected to an avalanche photodiode detector (EG&G SPCM-AQR).

A single controller (RHK Technology) is used for sample scanning, AFM feedback, and photon counting of the fluorescence, thus providing simultaneous topographic and fluorescence images. Fluorescence background count rates for blank substrates were less than 1 KHz. A slight nm-scale offset between the two images is possible due to the alignment of the laser focus on the AFM lever. All the AFM data presented here were acquired with levers (Olympus TR400-PSA, nominal force constant of 0.08 N/M) in contact mode under a feedback load 0.2 nN. Force calibration was checked with the thermal power spectrum method³⁸. The XY scanner plate was calibrated with a 463 nm square grid grating (Ted Pella, Inc.), and the AFM head was calibrated in Z with known 25.5 nm steps (TGZ01, NT-MDT Mikromasch).

Results and discussion

Lipid bilayers were created from a 3:1 molar mixture of DOPC/DPPC and various fluorescent probes. In all cases, the vesicle fusion process, which occurred above the gel-fluid transition temperature of DPPC (41° C), produced bilayers without apparent defects. Due to the allotment of at least one hour for temperature equilibration of the cantilevers in the coverslip dish, there was ample time for the stabilization of the gel/fluid phases¹¹. Thus the images shown below were static for the period of several hours.

In Figure 7.2 are shown 3.0 μm AFM and fluorescence images acquired simultaneously of a DOPC/DPPC (3:1) bilayer with 0.5% BODIPY-DPPE. The gel domains of DPPC are readily observed in the AFM images (Figure 7.2a) on the basis of the 1.1 \pm 0.2 nm height above the DOPC fluid phase, consistent with previous results^{4,5,7,8,12}. The DPPC domains have irregular shapes, and a size distribution ranging from 10 nm to >1 μm . In the corresponding fluorescence image, (Figure 7.2b) the domains are darker than the surrounding DOPC, thus revealing that the BODIPY-DPPE is excluded from the DPPC domains. Due in part to the 300 nm limit in confocal optical resolution, the domain boundaries are blurred in the fluorescence image and thus lack the detail observed in the topographic image. Also, by having both AFM and fluorescence images one can readily observe that in the fluorescence image larger domains are darker than the smaller ones.

The BODIPY label is situated on the DPPE headgroup, thus leaving the saturated acyl chains unperturbed relative to the DOPC/DPPC matrix. Since BODIPY-DPPE and DPPC have identical saturated acyl chains, it is surprising that the probe lipid tends to be excluded from the tightly-packed DPPC domains. It is thus perhaps due to headgroup interactions that it goes into the DOPC fluid phase. Similar behavior was observed for headgroup-labeled DPPE probes in *Langmuir*-Blodgett monolayers^{9,10,26}. The BODIPY-DPPE is not completely excluded from the gel domains, however, as can be seen from the fluorescence intensity profile in Figure 7.2c. The lowest count rates, corresponding to the darkest regions, still exceed the background count rate of 1 KHz by several orders of magnitude. For the most part, the gradual slope of the fluorescence line profile relative to the sharp domain boundaries delineated in the topograph line profile (also shown in Fig 2C), is due to the limited (300 nm) optical resolution. However, there is a small, additional broadening that indicates that the BODIPY-DPPE partitioning is not sharp across the boundary. This gradient of the BODIPY-DPPE concentration associated with the domain boundary is more obvious in Figure 7.3, where AFM and fluorescence images are shown for a very large DPPC domain in the same sample.

The fluorescent line profile in Figure 7.3c clearly indicates that the relative exclusion of the BODIPY-DPPE from the large DPPC domain increases with distance from the domain boundary to the center of the domain. For the sake of comparison with previous results, one can estimate a maximum relative partition coefficient $R = (IG/IG+IL)$, where IG is the intensity at the center of the domain and IL is intensity well outside the domain. For Figure 7.3c, $R \cong 0.25$; whereas for Figure 7.2c, $R \cong 0.37$. For fluorescein headgroup-labeled DPPE, a coefficient $R=0.14$ has been measured for partitioning into large (>1.0 μm) sphingomyelin-chloesterol domains in DOPC¹⁰. Thus the R value for Figure 7.3c is comparable, despite the change in lipids and probe. The larger R value for the smaller domain in Figure 7.2c points to an important aspect of the simultaneous AFM and fluorescence images

presented in Figure 7.2 and Figure 7.3: the smallest domains exhibit the weakest fluorescence contrast not only because they are more difficult to optically resolve, but also because the relative partition coefficient of the probe lipid is approaching 0.5 (i.e., uniform coverage). The apparent gradient in BODIPY-DPPE probe concentration across the domain boundaries also has the potential to mask irregular domain shapes (possibly indicating solid-like domains³⁹) and make them appear more circular (indicating more fluid-like domains). Furthermore, a very important question arises from these observations: is the boundary between the DOPC fluid-like region and DPPC gel-like region really as sharp as that indicated by the AFM topographic data? A slight mixing of the two components across the domain boundary would help to explain the partitioning results and also make possible an initial decrease of the BODIPY-DPPE concentration on the “DOPC side” of the topographically-defined boundary.

The glycosphingolipid GM1 has been shown in several fluorescence^{9,10,17,18,26} and AFM studies^{5,7,8,12} to partition into the gel-phase domains. It is even thought to be immiscible in DOPC²⁶. However, when the acyl chains are modified with a fluorescent probe, as is the case with BODIPY-C5-GM1, one can see from the 6.0 μm images presented in Figure 7.4 that the tailgroup-labeled GM1 is predominantly excluded from the DPPC gel-phase and resides in the DOPC matrix. This observation is consistent with previous bilayer domain studies using tailgroup-labeled sphingolipids^{16,17}. The exclusion of the tail-group labeled BODIPY-C5-GM1 from the gel phase is most likely due to the larger molecular area induced by steric interference between the probe and the longer acyl chains¹⁸. From the fluorescence line profile (Figure 7.4c), the BODIPY-C5-GM1 partition coefficient $R \cong 0.25$ at the center of the domain, where it is fairly flat over at least 1.5 μm . Just as discussed above with the headgroup-labeled DPPE, we see in Figures 7.4b and 7.4c, a more gradual partitioning of the tail-group labeled GM1 between the two phases than is dictated by the phase boundary or the 300 nm optical resolution. Finally, as expected for AFM imaging in a fluid phase, the large GM1 head groups are completely unresolved and no aggregation of the GM1 is observed in the DOPC regions. However, the DPPC domains in Figure 7.4a are now 0.8 ± 0.1 nm higher than the surrounding DOPC, a slightly smaller difference than that (1.1 ± 0.2 nm) observed in Figures 7.2 and 7.3 in the absence of GM1, and which may reflect the presence of GM1.

An important issue that is mentioned in numerous studies of lateral structure in model lipid bilayers is the superposition of domains between the bilayer leaflets; i.e., do the domains span both leaflets of the bilayer? Much evidence has been presented, particularly via fluorescence techniques^{9,13,14,17}, that this is indeed the case. The evidence resides in that no overlapping of different-sized domains is observed, nor are there “gray” features that derive from emission from only one leaflet and thus are half the intensity of emission from bilayer domains. These observations are also supported by the fluorescence images obtained in the present work. However, as discussed above, one notes that domains in Figures 7.2 and 7.3 become less and less dark with shrinking size due in part to the diminished exclusion of the probe lipid. Thus the “gray” criteria would lead to erroneous conclusions for small domains (<1 μm) relative to large domains (>1 μm). The question of leaflet superposition of domains is more easily verified here than in previous studies because there is always a correlation between the AFM and fluorescence images. For each domain in the fluorescence there is a corresponding feature in the AFM topography. If there was poor registry between leaflets, then one would expect to see light or dark regions in the fluorescence images without topographic correlation.

Alexa 488-CTX-B binds to GM1 receptors and thus is an effective way detect or “stain” unlabeled GM1 in the upper leaflet^{9,10,26}. In Figure 7.5, the simultaneous AFM and fluorescence images are shown for a (3:1) DOPC/DPPC bilayer containing 1% GM1 that is subsequently incubated with Alexa 488-CTX-B. As expected from previous results^{9,10,26}, the unlabeled GM1 is now readily incorporated into the DPPC domains seen in the AFM image (Figure 7.5a), and which appear bright in the fluorescence image (Figure 7.5b) due to the presence of GM1-bound Alexa 488-CTX-B. Due to the toxin binding to GM1, the DPPC domains in the AFM image now have a height of 3.2 ± 0.3 nm relative to the DOPC matrix. High resolution images (not shown) of the DPPC domains reveal a uniform coverage of close-packed 7 nm features that correspond to CTX-B pentamers³⁷. The fluorescence image shown in Figure 7.5b has distinct features not seen in the other data (Figures 7.2-7.4). The domains are uniformly bright and do not exhibit a fading towards the domain boundary. Thus the small domains (<1 μm) are as bright as the larger domains and the sharpness of the fluorescence features as seen in the line

profile (Figure 7.5c) is limited only by the 300 nm optical resolution of the domain boundaries and not by partitioning of the GM1 to the DOPC fluid phase. Thus it appears that the unlabeled GM1 is predominantly restricted to the DPPC gel phase and is uniformly distributed in it. However, there is a much larger than expected fluorescence background in the liquid DOPC phase. It may be attributed in part to unconjugated Alexa 488 probe, although the Alexa 488-CTX-B stock solution was passed through separation columns in an attempt to remove unconjugated probe. The background may also be attributed to a small component of the GM1 lipids (purified from ovine brain) that have acyl chains that could be excluded from the gel phase. Further studies on other lipid mixtures, such as DOPC/sphingomyelin/cholesterol¹⁰, where GM1 was shown to have a partition factor $R \cong 0.95$ for the gel-phase, are required before a conclusion can be drawn. In any case, an important point to note here is that fluorescence data has the potential to reveal the presence of GM1 in the fluid phase that would not be ordinarily detected by AFM.

Conclusions

Independent fluorescence and AFM imaging methods have been successfully used in the past to investigate domain formation in multicomponent lipid bilayers. It has been demonstrated here, for 3:1 DOPC/DPPC bilayers supported on glass, that the simultaneous acquisition of fluorescence and AFM topographic information provides insights on lateral organization that either technique alone would not necessarily be able to provide. Well-defined gel-phase DPPC domains with irregular shapes and sizes ranging from 10 nm up to several microns were observed in AFM images on the basis of the ~ 1 nm height difference above the surrounding DOPC fluid-phase. In the fluorescence images, those same (resolvable) domains were observed by the selective partitioning of probe-labeled lipids for the liquid-phase.

Correlation between the two sets of images revealed that the fluorescent probe lipid partitioning was not uniform across the domains, thus potentially distorting domain size and shape beyond that imposed by optical resolution. Factors that govern probe lipid partitioning across domain boundaries need to be studied in more detail. Models that incorporate both chemical and physical factors that not only lead to domain formation but also relative partitioning of tail-group and head-group labeled probes are required. This is complicated by the unknown lipid composition at these boundaries, despite the abrupt height difference in the bilayer topography. A more precise understanding of the topographical height difference between gel and fluid domains is also necessary in order to understand the multicomponent composition at the domain boundaries.

The results presented here also possibly indicate that the glycosphingolipid GM1 may not reside exclusively in the densely-packed gel-phase domains as could be construed from AFM, but may have some finite concentration in the fluid phase as detected by fluorescence from bound, labeled protein. Future studies of lateral structure and signaling processes in model membranes and harvested biological membranes will surely benefit from combined techniques such as that presented here. Moreover, this work provides a basis and motivation for improvements in optical resolution provided by near-field enhanced fluorescence excitation at metallic AFM probe tips^{40, 41}.

References

1. Johnson, S. J.; Bayerl, T. M.; McDermott, D. C.; Adam, G. W.; Rennie, A. R.; Thomas, R. K.; Sackmann, E. *Biophys. J.* 1991, 59, 289.
2. Sackmann, E. *Science* 1996, 271, 43.
3. Boxer, S. G. *Curr. Opin. Chem. Biol.* 2000, 4, 704.
4. Dufrêne, Y.; Barger, W.; Green, J.-B.; Lee, G. *Langmuir* 1997, 13, 4779.
5. Vié, V.; Mau, N.; Lesniewska, E.; Goudonnet, J.; Heitz, F.; Le Grimellec, C. *Langmuir* 1998, 14, 4574.
6. Kramer, A.; Wintergalen, A.; Sieber, M.; Galla, H. J.; Amrein, M.; Guckenberger, R. *Biophys. J.* 2000, 78, 458.
7. Yuan, C.; Johnston, L. J. *Biophys. J.* 2001, 81, 1059.
8. Milhiet, P.-E.; Vié, V.; Giocondi, M.-C.; Le Grimellec, C. *Single Mol.* 2001, 2, 109.
9. Dietrich, C.; Bagatolli, L. A.; Volovyk, Z. N.; Thompson, N. L.; Levi, M.; Jacobson, K.; Gratton, E. *Biophys. J.* 2001, 80, 1417.
10. Dietrich, C.; Volovyk, Z. N.; Levi, M.; Thompson, N. L.; Jacobson, K. *Proc. Natl. Acad. Sci. USA* 2001, 98, 10642.
11. Giocondi, M.-C.; Vié, V.; Lesniewska, E.; Milhiet, P.-E.; Zinke-Allmang, M.; Le Grimellec, C. *Langmuir* 2001, 17, 1653.
12. Yuan, C.; Furlong, J.; Burgos, P.; Johnston, L. J. *Biophys. J.* 2002, 82, 2526.
13. Almeida, P. F. F.; Vaz, W. L. C.; Thompson, T. E. *Biochemistry* 1992, 31, 7198.
14. Korlach, J.; Schwille, P.; Webb, W. W.; Feigensohn, G. W. *Proc. Natl. Acad. Sci. USA* 1999, 96, 8461.
15. Bagatolli, L. A.; Gratton, E. *Biophys. J.* 2000, 79, 434.
16. Wang, T.-Y.; Silviu, J. R. *Biophys. J.* 2000, 79, 1478.
17. Samsonov, A. V.; Mihalyov, I.; Cohen, F. S. *Biophys. J.* 2001, 81, 1486.
18. Wang, T.-Y.; Silviu, J. R. *Biophys. J.* 2003, 84, 367.
19. Simons, K.; Ikonen, E. *Nature* 1997, 387, 569.
20. Brown, R. E. *J. Cell Sci.* 1998, 111, 1.
21. Harder, T.; Scheiffele, P.; Verkade, P.; Simons, K. *J. Cell. Biol.* 1998, 141, 929.
22. Brown, D. A.; London, E. *Ann. Rev. Cell Dev. Biol.* 1998, 14, 111.
23. Jacobson, K.; Dietrich, C. *Trends Cell Biol.* 1999, 9, 87.
24. Simons, K.; Ikonen, E. *Science* 2000, 290, 1721.
25. Anderson, R.; Jacobson, K. *Science* 2002, 296, 1821.

26. Radhakrishnan, A.; Anderson, T. G.; McConnell, H. M. *Proc. Natl. Acad. Sci. USA* 2000, 97, 12422.
27. Hollars, C. W.; Dunn, R. C. *Biophys. J.* 1998, 75, 342.
28. Vaccaro, L.; Schmid, E. L.; Ulrich, W.-P.; Vogel, H.; Duschl, C.; Marquis-Weible, F. *Langmuir* 2000, 16, 3427.
29. Yuan, C.; Johnston, L. J. *J. Microsc.* 2002, 205, 136.
30. Enderle, T.; Ha, T.; Ogletree, D. F.; Chemla, D. S.; Magowan, C.; Weiss, S. *Proc. Natl. Acad. Sci. USA* 1997, 94, 520.
31. Hwang, J.; Gheber, L. A.; Margolis, L.; Edidin, M. *Biophys. J.* 1998, 74, 2184.
32. Varma, R.; Mayor, S. *Nature* 1998, 394, 798.
33. Kenworthy, A. K.; Petranova, N.; Edidin, M. *Mol. Biol. Cell* 2000, 11, 1645.
34. Schütz, G. J.; Sonnleitner, M.; Schindler, H. *J. Fluores.* 2001, 11, 177.
35. Müller, D. J.; Fotiadis, D.; Engel, A. *FEBS Letters* 1998, 430, 105.
36. Müller, D. J.; Heymann, J. B.; Oesterhelt, F.; Möller, C.; Gaub, H.; Büldt, G.; Engel, A. *Biochim. Biophys. Acta* 2000, 1460, 27.
37. Czajkowsky, D. M.; Shao, Z. *FEBS Letters* 1998, 430, 51.
38. Hutter, J. L.; Bechhoefer, J. *Rev. Sci. Instrum.* 1993, 64, 1868.
39. Veatch, S. L.; Keller, S. L. *Phys. Rev. Lett.* 2002, 89, 268101.
40. Sánchez, E. J.; Novotny, L.; Xie, X. S. *Phys. Rev. Lett.* 1999, 82, 4014.
41. Yang, T. J.; Lessard, G. A.; Quake, S. R. *Appl. Phys. Lett.* 2000, 76, 378.

Figures

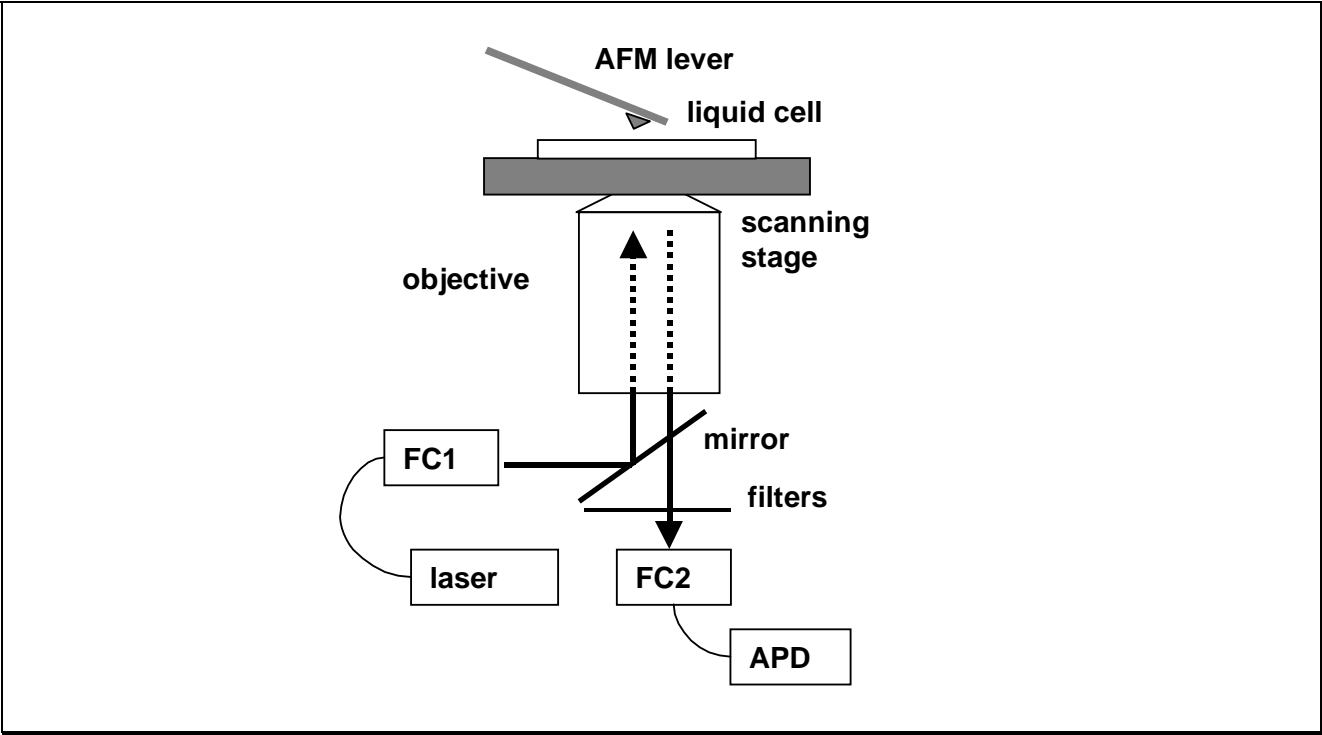


Figure 7.1. Schematic (not to scale) of experimental apparatus for acquiring AFM and fluorescence images. See text for details. FC1 and FC2 are fiber couplers for excitation light and fluorescence emission, respectively. APD is an avalanche photodiode. The scanning stage is mounted on an inverted microscope stage (not shown).

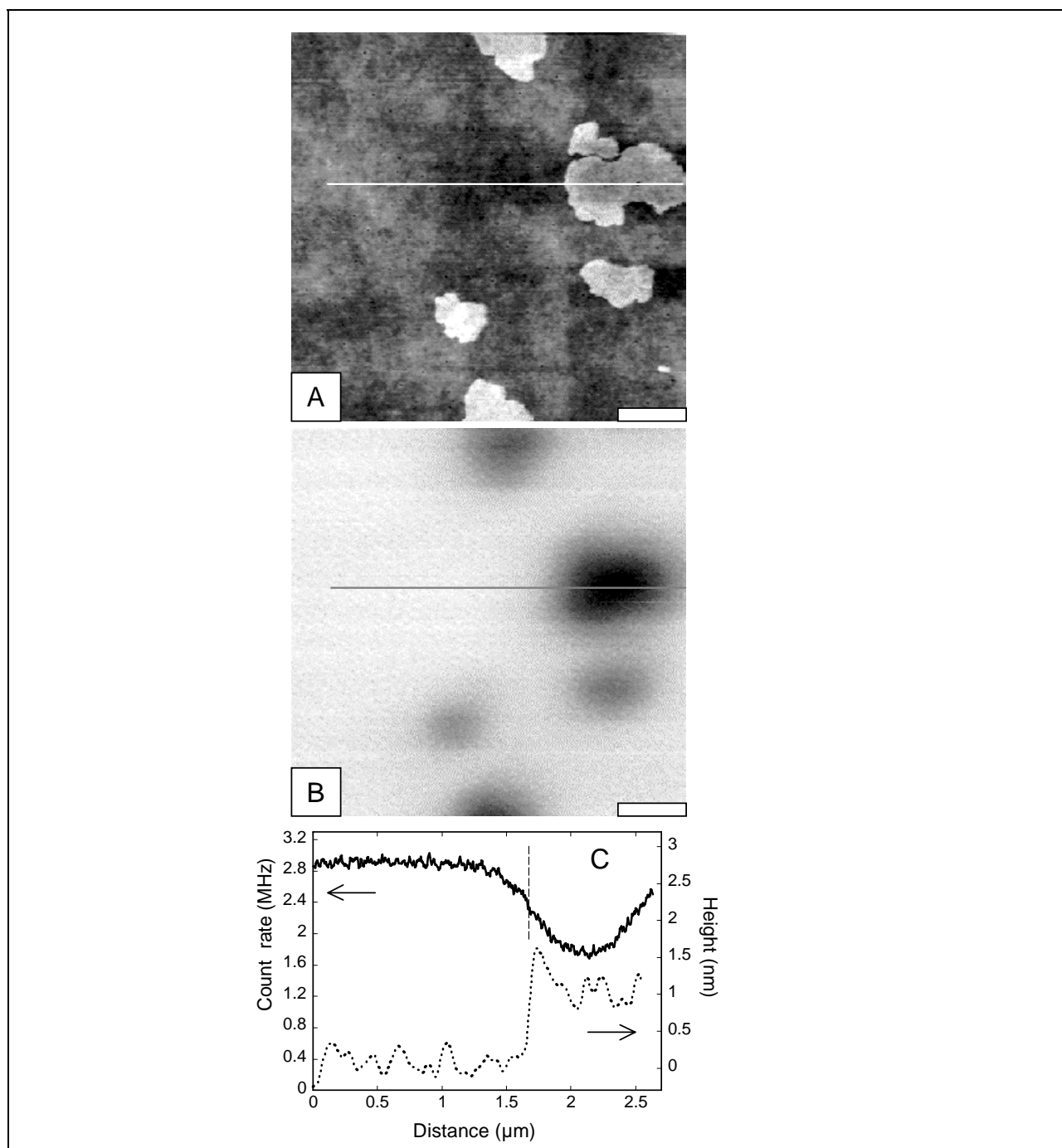


Figure 7.2. 3.0 μm scan featuring topographic (A) and fluorescence (B) images of a 3:1 DOPC/DPPC bilayer, with 0.5% BODIPY-DPPE, on glass (scale bar = 500 nm). The height of the DPPC domains in (A) is 1.1 ± 0.2 nm. In the fluorescence image (B), the dark regions are due to the exclusion of the headgroup-labeled BODIPY-DPPE from the DPPC domains. The smaller domains are not as dark as the largest, indicating less exclusion. In (C), a fluorescence line profile (solid) and corresponding topographic line profile (dotted) are plotted for the largest domain. The left boundary of the domain is marked by the dashed line.

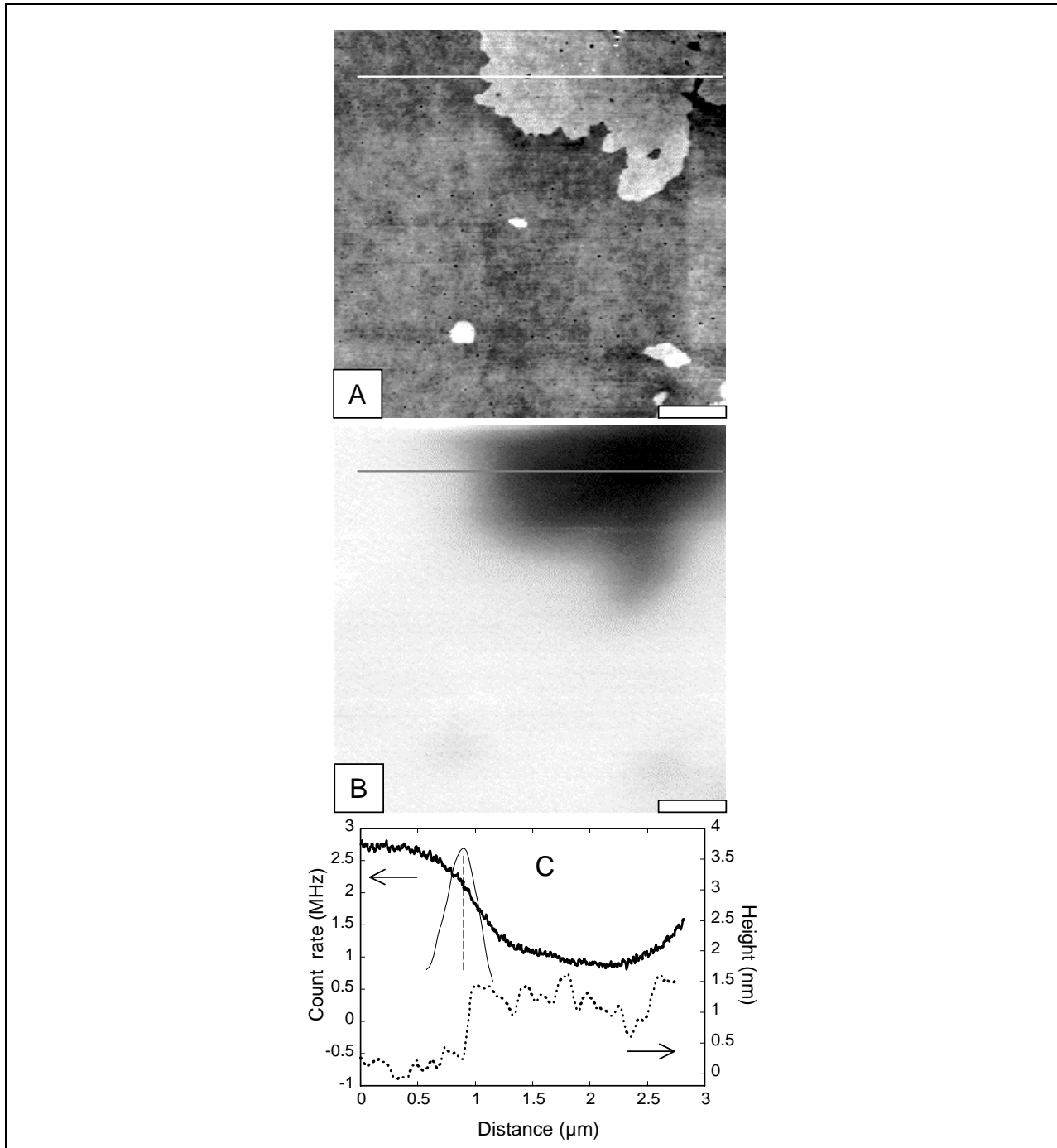


Figure 7.3. 3.0 μm scan featuring topographic (A) and fluorescence (B) images of a 3:1 DOPC/DPPC bilayer, with 0.5% BODIPY-DPPE, on glass (scale bar = 500 nm). The height of the DPPC domains in (A) is 1.1 ± 0.2 nm. The large domain is likely formed by coalescence of smaller domains, as evidenced by residual pockets of DOPC. In the fluorescence image (B), the dark region is due to the exclusion of the headgroup-labeled BODIPY-DPPE from the DPPC domains. It does not appear to be excluded from the small domains, although optical resolution is limiting detection of the smallest. In (C), a fluorescence line profile (solid) and corresponding topographic line profile (dotted) are plotted for the large domain. The left boundary of the domain is marked by the dashed line. Overlaid on this dashed line is a diffraction-limited (300 nm) line profile (thin solid) to show that the fluorescence line profile is considerably broadened beyond the diffraction limit.

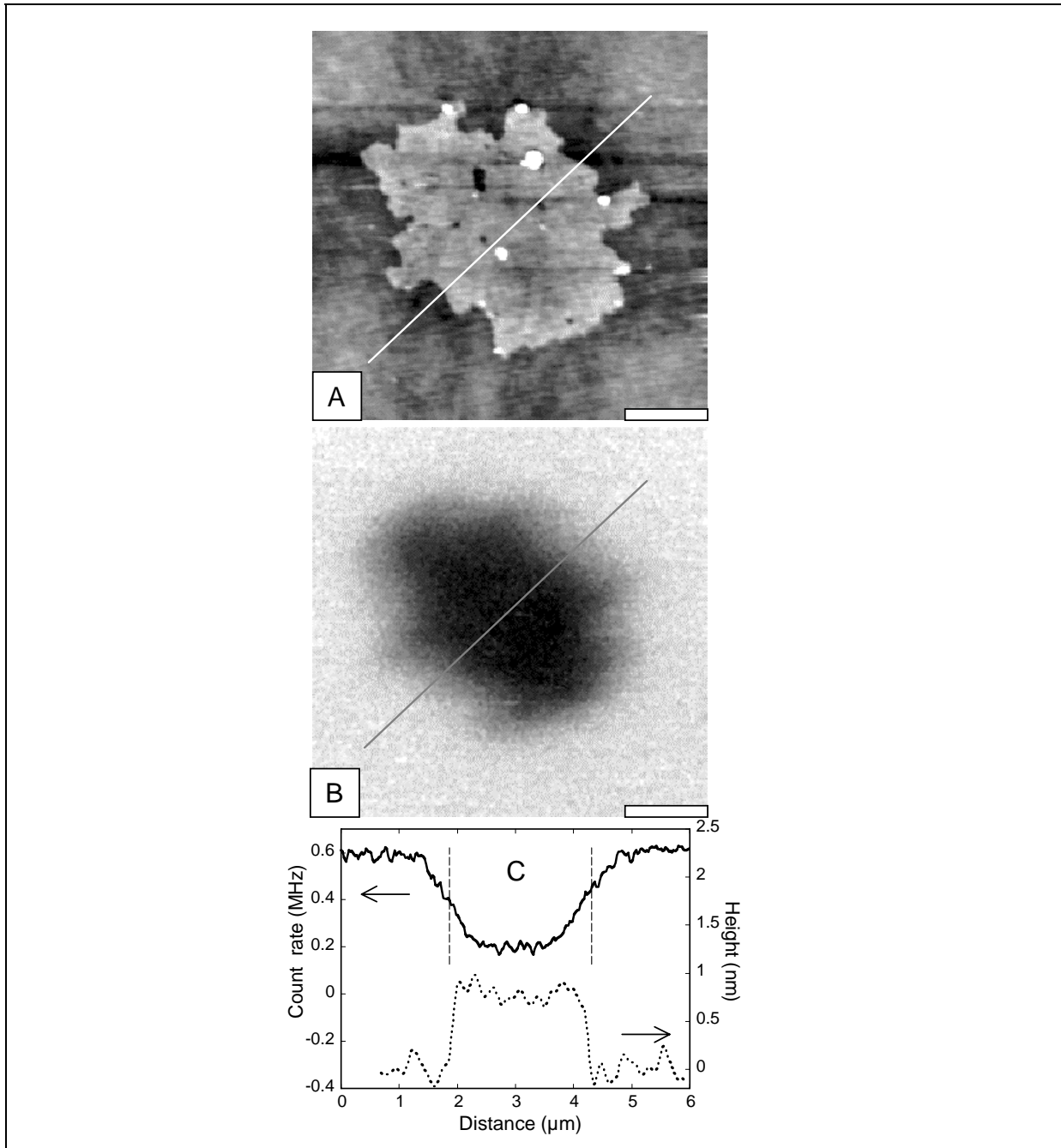


Figure 7.4. 6.0 μm scan featuring topographic (A) and fluorescence (B) images of a 3:1 DOPC/DPPC bilayer, with 0.25% BODIPY-GM1, on glass (scale bar = 1.0 μm). The height of the DPPC domain in (A) is 0.8 ± 0.1 nm. In the fluorescence image (B), the dark region is due to the exclusion of the tailgroup-labeled GM1 from the DPPC domains. In (C), a fluorescence line profile (solid) and corresponding topographic line profile (dotted) are plotted. The left and right boundary of the domain is marked by the dashed lines.

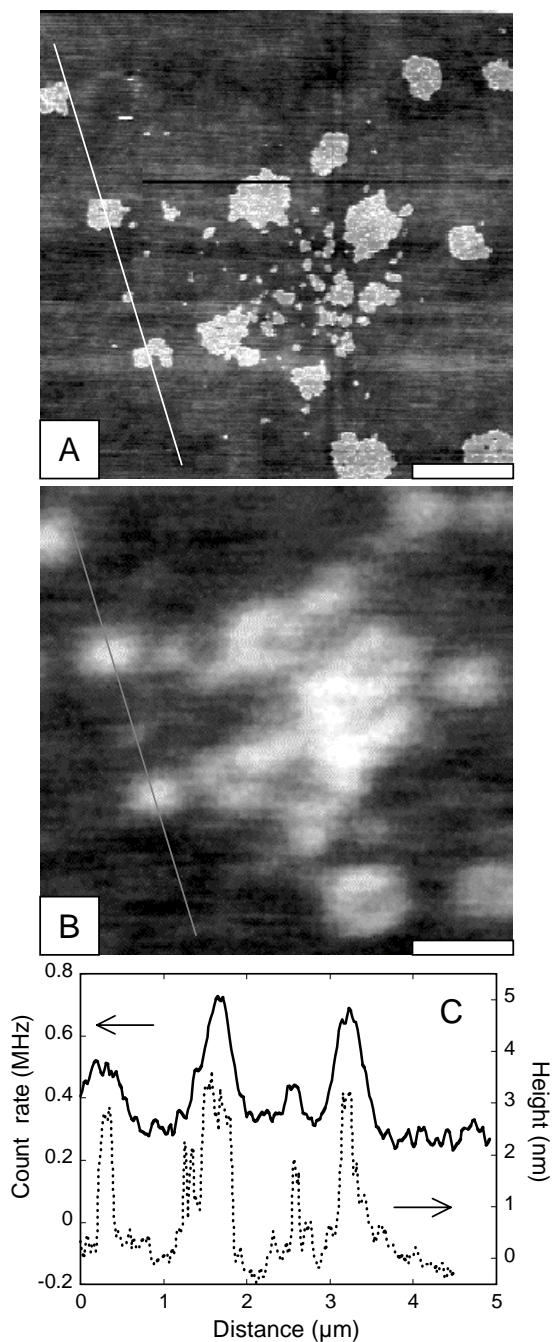


Figure 7.5. 5.0 μm scan featuring topographic (A) and fluorescence (B) images of a 3:1 DOPC/DPPC bilayer, with Alexa 488-CTX-B bound to 1.0 % unlabeled GM1, on glass (scale bar = 1.0 μm). The combined height of the DPPC domains and bound CTX-B in (A) is 3.2 ± 0.3 nm. In the fluorescence image (B), the bright regions are due to the Alexa 488-labeled CTX-B bound to the DPPC domains. In (C), a fluorescence line profile (solid) and corresponding topographic line profile (dotted) are plotted. The fluorescence lineshapes of the two prominent features are diffraction-limited ($\text{fwhm} \cong 300$ nm).

Acknowledgments

The author is indebted to Tina Huber for preparation of lipid vesicles, Jennifer Gaudioso for column chromatography, and Peter Chen for substrate cleaning. Critical reading of the manuscript by Mark Stevens is also gratefully acknowledged.

Distribution

1	MS 9004	Rick Stulen, 8100
1	MS 9951	Len Napolitano, 8140
1	MS 9951	Malin Young, 8141
5	MS 9951	Joe Schoeniger, 8141
1	MS9671	Marites J. Ayson, 8141
1	MS1413	Alan Burns, 1141
1	MS0316	Paul Crozier, 9235
1	MS9951	Jean-Loup Faulon, 9212
1	MS9052	Gary H. Kruppa, 8368
1	MS9671	Pamela Lane, 8141
1	MS9052	Petr Novak, 8368
1	MS9951	Kenneth Sale, 8141
1	MS1411	Mark J. Stevens, 1838
3	MS 9018	Central Technical Files, 8945-1
1	MS 0899	Technical Library, 9616
1	MS 9021	Classification Office, 8511 for Technical Library, MS 0899, 9616 DOE/OSTI via URL
1	MS 0323	D. Chavez, LDRD Office, 1011

This page intentionally left blank.