January 2011

Mass Spectrometry-Based Structural Proteomics: Bottom-Up Protein Footprinting And Top-Down Native Electrospray Of Protein Assemblies

Hao Zhang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:

Michael L. Gross, Chair

Robert E. Blankenship

Ron Bose

Cynthia S. Lo

Liviu Mirica

John-Stephen Taylor

MASS SPECTROMETRY-BASED STRUCTURAL PROTEOMICS: BOTTOM-UP PROTEIN FOOTPRINTING AND TOP-DOWN NATIVE ELECTROSPRAY OF PROTEIN ASSEMBLIES

by

Hao Zhang

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2011

Saint Louis, Missouri

ACKNOWLEDGMENTS

I thank my advisor, Prof. Michael L. Gross, for his guidance and support during my thesis research and graduate study at Washington University. He provided expert direction, a state-of-art research environment and the ability to work with distinguished collaborators.

A special thanks to Henry W. Rohrs, Weidong Cui, Don L. Rempel and Ian Vidavsky for their help in instrumentation, researching design, data analysis and , most importantly, for sharing a friendship during my study and research.

My thanks are extended to those members, both current and past, of NIH NCRR Mass Spectrometry Resource at Washington University in St. Louis: Michael Grayson, Joyce Neff, Daryl Giblin, Ying Li, Bich Vu, Micah Wilcox, Adam Brustkern, Sandy Kerfoot, Manohari Silva, Dian Su, Brian Gau, Jiawei Chen, Richard Huang, Yining Huang, Yuetian Yan and Ying Zhang. Special thanks to Reid Towsend, Richard LeDuc, Cheryl Lichti, Haowei Song from School of Medicine, Washington University in St. Louis.

I also thank my collaborators: Prof. Robert E. Blankenship, Jianzhong Wen from department of chemistry and biochemistry; Prof. Bose Ron, Wei Shen, John Monsey, Michael Boyne from school of medicine, Washington University in St. Louis; Pegah R. Jalili, Janet W. Irungu, Gordon R. Nicol, Kevin B. Ray from Sigma-Aldrich Corporation; James A. Carroll from Pfizer Inc; Prof. Hao Chen, Yun Zhang from department of chemistry, Ohio University.

This thesis research is supported by the National Institutes of Health and National Center for Research Resources (Grant No. 2P41RR000945). Some projects in my research are supported by the Instrument Development for Biological Research Program of the NSF (Grant 0964199) and U.S. Department of Energy (Grant DE-FG02-07ER15902). Part research is from the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (Grant DE-SC 0001035). Additionally, I thank to the graduate school of art and sciences for dean's fellowship.

I thank my Dissertation Advisory and Examination Committee members, Prof. Robert E. Blankenship, Prof. Ron Bose, Prof. Cynthia S. Lo, Prof. Liviu Mirica, Prof. John-Stephen Taylor for their advice and discussions regarding my thesis research.

A special thanks to my parents, grandparents, uncles and friends for their continued support, love and encouragement throughout my graduate studies.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Converting gene-sequence information into functional information about a protein is a major challenge of post-genomic biology(*1*). Proteins have a variety of functions from serving as catalysts to acting as structural components; all these functions are closely related to protein structure. The first step to understand protein function is often a structural study of that protein. Two major approaches, NMR spectroscopy(*2*) and X-ray crystallography(*3*), can provide an atomic-level, 3D structural model of a protein. The applications of these high resolution approaches, however, are limited by protein size, conformational flexibility, and aggregation propensity(*4*). To obtain complementary structural information about proteins, a variety of approaches from traditional structural biology (e.g., circular dichroism, fluorescence spectroscopy) to new advances (e.g., computational prediction, protein footprinting) are required(*1*).

Mass spectrometry (MS) has become an important tool for studying protein structure, dynamics, interactions, and function(*5*). In particular, detailed characterization of protein-ligand interactions is now possible(*6*), a critical step toward understanding biological function.Mass spectrometric analysis of protein structure can take two approaches. First, protein-ligand interactions can be probed by chemical labeling followed by MS analysis to determine the resulting mass shift (extent of labeling) and the location of the labeling. This approach in a titration format gives protein-ligand affinities. The labeling takes place in solution, where biochemistry occurs, and can be under physiological conditions, whereas the mass spectrometer is used for analysis typically by bottom-up proteomic strategies. In the other approach, protein assemblies can also be transferred directly into the gas phase and interrogated by MS to afford structural insights. One can view this is a top-down approach. The measurements refer to a gas-phase

species, and that raises the question of whether the outcomes of the measurements have relevance to the structure and properties of proteins in solution or in a living system.

Although there are differences in experimental format, results, and sensitivity between the two approaches of MS-based protein structural analysis, the similarity of those approaches must not be overlooked(*7*). All MS-based structural analyses rely heavily on the identification of peptides, purified protein species, or protein complexes. This analysis has been accelerated by the developments of MS instrumentation and methodology in protein analysis; the structural information provided by MS-based analysis is greatly facilitated by having a structural model of the protein. The integrated results from MS approaches, traditional structural biology approaches (e.g., NMR and X-ray), and computational modeling give more complete structural information of proteins than that from any one of the approaches alone.

In the first part of thesis, we focus on the development and application of chemical-labeling methods (protein footprinting) in studies of protein conformation. In the second part, a combined top-down approach of native ESI and electron-capture dissociation (ECD) in FTICR MSis presented for structural studies of protein assemblies in the gas phase.

Part 1.The development and application of chemical-labeling methods in studies of protein conformation.

One MS-based approach is protein footprinting(*8*). When a protein binds with a ligand, the solvent accessible surface areas of its residues are affected by the resulting conformational changes and direct ligand interactions. Residues with reduced solvent

accessibility, such as those at the binding site in the holo state, have attenuated labeling or cleavage relative to those of the comparison state, and this signal change is readily resolved by MS. Thus, protein footprinting is well suited to the study of protein-ligand interactions, although its outcome is of lower resolution than that provided by X-ray crystallography and NMR spectroscopy. An advantage of MS-based footprinting over these high resolution methods is that protein-ligand structure can be interrogated in a near physiological setting (i.e., at dilute concentration in solution with appropriate buffer, ionic strength, and bimolecular milieu).

Footprinting was first reported as a cleavage-based strategy to detect protein-DNA binding specificity(*9*). This demonstration inspired similar methods to be developed for the analysis of proteins(*10*); Hanai(*11*) and co-workers first demonstrated that modification reagents could footprint proteins. Protein footprinting as a complementary approachto probing protein conformationhas rapidly developed during the last two decades(*12*).Different footprinting strategies including side-chain derivatizations, hydrogen deuterium amideexchange (H/DX), and hydroxyl-radical modification are now characterized sufficiently to be used (*13-16*). A labeling process usually induces a mass shift in aprotein's molecular weight. The resulting mass shift makes mass spectrometry (MS) a suitable analytical tool for detecting the protein footprinting and locating its site. Coupled withthe labeling strategy, MS has greater sensitivity and speed than many other approaches in protein biochemistry.Several successful MS-based protein footprinting methods have been demonstratedfor monitoring the dynamics of protein folding/unfolding(*4, 17, 18*), measuring protein-ligand binding

affinity(*19*), locating the interface of protein-protein interactions, and probing protein structural information(*20*, *21*).

In MS-based footprinting, protein samples are modified at physiological pH and salt concentration. The modified protein can be directly analyzed by ESI-MS by using a C18 trap column to desalt the sample, or it can be digested by trypsin and analyzed by LC-ESI-MS/MS to quantify the site of modification. With accurate mass measurement, the percent modification (the modification extent) is determined with little ambiguity on the basis of the peak area of the modified peptide divided by the peak areas of all forms of that peptide. The site of modification can be located by MS/MS; for example, carboxyl-group modification is located by interpretingthe product-ion spectrum of the modified peptide(Figure 1).

Chemical-labeling-based protein footprinting can be classified into two groups: site-specific and site-nonspecific footprinting. Site-specific footprinting can provide detailed conformational information about only certain, reactive amino-acid residues. However, the information is limited by a small set of targets (e.g., acetylation modifies only the amino acid lysine). Site-nonspecific footprinting, on the other hand, has a relatively large set of targets and provides more information. The information from a large set of targets, however, increases the complexity in the footprinting data analysis(*7*).

In the site-nonspecific footprinting, there are two subgroups: one is radical-based modifications; the other is hydrogen deuterium amide exchange (H/DX). The most common radical-based labeling approach is hydroxyl-radical protein footprinting(*16*). The hydroxyl radical, as structural probe, has several advantages; it is highly reactive

(hydroxyl radicals can react with the half of amino acids in microseconds of exposure), has a similar size as water and, therefore, goes where water goes, and gives modified products that can be interrogated by tandem MS to provide residue-level information. Hydroxyl radicals can be generated by different reactions: examples are electron-pulse radiolysis, synchrotron radiolysis of water, laser photolysis of hydrogen peroxide, Fenton and Fenton-like reactions, and high-voltage electrical discharges(*12*). In the second chapter of this thesis, we describethe use of fast photochemical oxidation of protein (FPOP), an approach using laser photolysis of hydrogen peroxide to give hydroxyl radicals; this approach was developed in our group(*22*). FPOP modifies proteins on the microsecond time scale. Highly reactive •OH, produced by laser photolysis of hydrogen peroxide, oxidatively modifies the side chains of overone half the common amino acids on this time scale. Owing to the short labeling exposure, only solvent-accessible residues are sampled.We describe anapplication of FPOP in a structural study of various calmodulin-peptide complexes.

The other subgroup in site-nonspecific footprinting is hydrogen deuterium amide exchange (H/DX)(*23*). Deuterium incorporation is used as a probe to measure the solvent accessibility of the protein-backbone amide hydrogens. Although deuterium back exchange and scrambling can limit the application of H/DX, it is still a powerful technique in analyzing protein structure. By digesting the protein and looking at rates at the peptide level, the rate of H/DX can be measured along the entire length of a protein backbone, providing coarse structural information of the whole protein (except for proline). H/DX is unlikely to perturb protein structure during labeling as opposed to other labeling reagents. In the third chapter of this thesis, H/DX was used as protein-structure

check to insure that in the development of carboxyl group modification based protein footprinting, the footprinting itself is not perturbing the protein and does not give misleading labeling results.

In site-specific footprinting, a variety of chemical reagents are used to modify specifically certain amino acids; an example is the 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide(EDC) mediated coupling reaction between glycine ethyl ester (GEE) and the carboxyl group of a protein(*14, 24*). Another example of a reactive amino-acid group is the thiolof cysteine; it can be modified by several reagents (e.g., iodoacetamide, NEM)(*14, 25*). The major drawback of site-specific footprinting is that, because there are so few target residues, it is likely the no conformational information will obtained as there is no target residue in protein region undergoing a change.

In the third chapter of the thesis, the protocol of carboxyl-group-modification-based protein footprinting is developed by using calmodulin as a model protein.The reaction rate of carboxyl-group modification is examined under different conditions.The application of carboxyl-group modification in probing conformational changes in calmodulin induced by calcium binding is demonstrated. In the fourth chapter of this thesis, an application of carboxyl-group-modification-based protein footprinting, mapping the kinase domain dimer interface and phosphorylation-induced conformational changes of a membrane-associated tyrosine kinase (Her4/ErbB-4) is presented. The Her4 kinase domain monomer-dimer equilibrium was measured using a titration format.

Part 2: Top-down fragmentation of protein assemblies: native electrospray and electron capture dissociation in FTICR MS.

Most proteins carry out their functions as a complex or a protein assembly. Those high-order structures varyin molecular mass from kDa to MDa and contain many components(*26, 27*).For example, in the yeast proteome, about 70% of the more than 4000 proteins identified by mass spectrometry (MS) are involved in protein-protein interactions. One estimate is that more than 500 protein complexes are formed, each containing fivesubunits in average (*28*). To understand the function of high order structures, especially those that are hard to study by traditional techniques, complementary methods are needed to provide structural information.

Protein structuresare usually investigated by electron microscopy, X-ray crystallography, NMR spectroscopy, and light microscopy (*26, 29*). One fast growing area in structural biology is the study of protein assemblies, and MS is beginning to play a role(*30*). The major approach for introducing assemblies to the gas phase is native electrospray (ESI)from aqueous solution under conditions that are close to physiological(*31*). Information about stoichiometry, structure, and subunit interactions can result from these MS studies. AlthoughMS cannot provide atomic-level resolution, as can X-ray crystallography and NMR, its advantages for interrogating protein complexes in their near-native states are small sample consumption, high throughput, and unique specificity to make it complementary to traditional structural biology techniques(*32*).

Since Ganem and co-workers(*33*) first demonstrated the application of MS for protein complexes, developments in both methodology and instrumentation of MS have been improved its application in structural biology(*32*); progress has been reviewed periodically(*34-36*). Evidence about keeping complexes near their native conformations in the gas phase was also reported (*37-39*). Technical advances, such as analysis of

membrane protein complexes(*40, 41*) and large protein machine up to MDa(*42*), were also reported.

Two types of mass analyzers, time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FTICR) are appropriate for measuring the high *m/z* protein-assembly ions introduced via native ESI. The principal instrumentation used thus faris thequadrupole/time-of-flight (QToF)(*43*)owing to its high upper-mass limit, simple design, and compatibility forinterface with ion mobility (IM). A small number of groups are using FTICR MScoupled with native ESI (*44, 45*). Although QToF instruments provide good accurate mass, dissociation chemistry, and when coupled with IM, assembly-shape information, identification of thesubunit via sequence information still relies on thestandard bottom-up proteomics(*30*). The disadvantage is that the dynamics, assembly shape and structural layout are lost in any bottom-up analysis.

A top-down approach that integrates those two independent experiments into one would have significant advantages for sample preparation and data interpretation(*46*). Different fragmentation methods, collision-induced dissociation (CID), electron-capture dissociation (ECD), blackbody infrared radiativedissocation (BIRD), infrared multiphoton dissociation (IRMPD) and surface-induced dissocation (SID), are availablefor the top-down approach(*47-53*).

In the top-down approaches, CID and ECD are two effectivefragmentation methods. For CID, precusor ions are accelerated to high kinetic energy and collide with neutral molecules (nitrogen or argon) in the source region or collision cell of mass spectrometers. The kinetic energy of precusor ions is converted into internal energy. With

protonation on an amide nitrogen or a carbonyl oxygen, the C(O)-N cleavage occurs with attack of the adjacent nucleophilic carbonyl oxygen (Figure 3) (*54*). The major sequence-specific ions from CID experiment are y- and b- type ions as in Figure 2. For ECD, low energy electrons are captured by the precursor ions in the FT-ICR trap. After initial electron capture atone of the positively charged sites, transfer of a hydrogen atom to an amide oxygen facilitates β-cleavage of the adjacent N-Cα bond through an aminoketyl radical intermediate. The resulting fragments are c- and z- type ions (Figure 4) (*55*). By comparing the product-ion spectra of protein or peptide ions with calculated mass lists of y-/b- type ions (CID) or c-/z- type ions (ECD), the sequence information of protein/peptide can be elucidated.

For the application of CID, Robinson and co-workers(*56*)useda modified QTOF mass spectrometer and IM to study the unfolding of protein assemblies. Some fragment ions from backbone cleavageswere obtained, suggestingthe potential of tandem MS in structural studies of protein assemblies. A recent study on human transthyretin demonstrated that peptide fragments are generated by a charge-state-dependent decay(*57*). Charge reduction is required for successful CID experiment to break the backbone of protein complexes.

One advantage of hybrid Qq-FTICR MS isits versatile fragmentation for top-down purposes.Using hybrid Qq-FTICR, onecan fragment protein assemblies by ECD or IRMPD in the ICR trap or by CID or ISCID in the front end of the instrument to form highly charged sub-complexes or monomeric subunit. Additionally, a subunit bearing more charges is beneficial for ECD to generate sequence ions while they are stored in the FTICR trap. The high mass resolving power of FTICR also makes feasible the analysis

of some complicated product-ion spectra by separating isotope peaks.Heeren and co-workers(*58*) first reported the application of ECD and FTICR MS in the study of non-covalent protein assemblies. Although they selected each charge state of the assembly in the FTICR trap before ECD, they observed no ECD-induced backbone cleavage productions (i.e., sequence ions).

In the fifth chapter of this thesis, the detail of native ESI of protein assemblies and recent advances in the area of protein assemblies are reviewed more extensively than here. The review suggests that dissociation of protein assemblies in the gas phase by tandem mass spectrometry can be a powerful approach in obtaining structural information. In the sixth chapter of this thesis, an example for top-down sequencing of a protein component in an intact protein assembly (the intact yeast alcohol dehydrogenase tetramer of 147 kDa) by electron-capture dissociation in FTICR MS is presented. To provide structural insights of protein assembly by this top-down approach, the systematic study of electron-capture-dissociation-based top-down of protein assemblies is required. In the seventh chapter, an investigation of several protein assemblies including yeast alcohol dehydrogenase, concanavalin A, and photosynthetic Fenna-Matthews-Olsen protein complex are described in which native ESI and electron-capture dissociation in FTICR MS are used. We found that the relatively free and flexible regions of the subunits in protein assemblies can be sequenced by ECD or activated ECD. This top-down approach provides not only top-down proteomics information of the complex subunits, but also structural insights of protein assemblies.

References

1. Petsko, G. A., and Ringe, D., (Eds.) (2004) *Protein Structure and Function,* New Science Press Ltd, London.
2. Wuthrich, K. (1990) Protein structure determination in solution by NMR spectroscopy, *J Biol Chem 265*, 22059-22062.
3. Scapin, G. (2006) Structural biology and drug discovery, *Curr Pharm Des 12,* 2087-2097.
4. Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *J Mass Spectrom 43,* 1021-1036.
5. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature 422*, 198-207.
6. Schermann, S. M., Simmons, D. A., and Konermann, L. (2005) Mass spectrometry-based approaches to protein-ligand interactions, *Expert Rev Proteomics 2,* 475-485.
7. Chance, M. R. (2008) *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics*, John Wiley & Sons, Inc., Hoboken, NJ.
8. Jay, D. G. (1984) A general procedure for the end labeling of proteins and positioning of amino acids in the sequence, *J Biol Chem 259*, 15572-15578.
9. Mirzabekov, A. D., and Melnikova, A. F. (1974) Localization of chromatin proteins within DNA grooves by methylation of chromatin with dimethyl sulphate, *Mol Biol Rep 1*, 379-384.

10. Sheshberadaran, H., and Payne, L. G. (1988) Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein "footprinting", *Proc Natl Acad Sci U S A 85*, 1-5.

11. Hanai, R., and Wang, J. C. (1994) Protein footprinting by the combined use of reversible and irreversible lysine modifications, *Proc Natl Acad Sci U S A 91*, 11904-11908.

12. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem Rev 107*, 3514-3543.

13. Sperry, J. B., Shi, X., Rempel, D. L., Nishimura, Y., Akashi, S., and Gross, M. L. (2008) A mass spectrometric approach to the study of DNA-binding proteins: interaction of human TRF2 with telomeric DNA, *Biochemistry 47*, 1797-1807.

14. Mendoza, V. L., and Vachet, R. W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry, *Mass Spectrom Rev 28*, 785-815.

15. Zhu, M. M., Rempel, D. L., Du, Z., and Gross, M. L. (2003) Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX, *J Am Chem Soc 125*, 5252-5253.

16. Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes, *Annu Rev Biophys Biomol Struct 35*, 251-276.

17. Boys, B. L., and Konermann, L. (2007) Folding and assembly of hemoglobin monitored by electrospray mass spectrometry using an on-line dialysis system, *J Am Soc Mass Spectrom 18*, 8-16.

18. Stocks, B. B., and Konermann, L. (2009) Structural characterization of short-lived protein unfolding intermediates by laser-induced oxidative labeling and mass spectrometry, *Anal Chem 81*, 20-27.

19. Zhu, M. M., Rempel, D. L., Zhao, J., Giblin, D. E., and Gross, M. L. (2003) Probing Ca2+-induced conformational changes in porcine calmodulin by H/D exchange and ESI-MS: effect of cations and ionic strength, *Biochemistry 42*, 15388-15397.

20. Chitta, R. K., Rempel, D. L., Grayson, M. A., Remsen, E. E., and Gross, M. L. (2006) Application of SIMSTEX to oligomerization of insulin analogs and mutants, *J Am Soc Mass Spectrom 17*, 1526-1534.

21. Xu, G., and Chance, M. R. (2004) Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions, *Anal Chem 76*, 1213-1221.

22. Hambly, D., and Gross, M. L. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin *International Journal of Mass Spectrometry 259*, 124-129.

23. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass Spectrom Rev 25*, 158-170.

24. Hoare, D. G., and Koshland, D. E., Jr. (1967) A method for the quantitative modification and estimation of carboxylic acid groups in proteins, *J Biol Chem 242*, 2447-2453.

25. Lundblad, R. L. (2005) *Chemical Reagents for Protein Modification* Third ed., CRC press.

26. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell, *Nature 450*, 973-982.

27. Ban, N., and Egelman, E. H. Structure and function of large cellular assemblies, *Curr Opin Struct Biol 20*, 207-209.

28. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae, *Nature 440*, 637-643.

29. Hasnain, S. S., and Wakatsuki, S. Advances in biophysical methods: characterisation and visualization of molecules, cells and organism, *Curr Opin Struct Biol 20*, 584-586.

30. Zhou, M., and Robinson, C. V. When proteomics meets structural biology, *Trends Biochem Sci 35*, 522-529.

31. Heck, A. J. (2008) Native mass spectrometry: a bridge between interactomics and structural biology, *Nat Methods 5*, 927-933.

32. Lorenzen, K., and Duijn, E. V. (2010) Current Protocols in Protein Science, In *UNIT 17.12 Native Mass Spectrometry as a Tool in Structural Biology* (Coligan, J. E., Dunn, B. M., Speicher, D. W., and Wingfield, P. T., Eds.), John Wiley & Sons, Inc.

33. Ganem, B., li, Y.-T., and Henion, J. D. (1991) Observation of Noncovalent Enzyme-Substrate and Enzyme-Product Complexes by Ion-Spray Mass Spectrometry, *J Am Chem Soc*, 7818-7819.

34. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom Rev 16*, 1-23.

35. Benesch, J. L., Ruotolo, B. T., Simmons, D. A., and Robinson, C. V. (2007) Protein complexes in the gas phase: technology for structural genomics and proteomics, *Chem Rev 107*, 3544-3567.

36. Heck, A. J., and Van Den Heuvel, R. H. (2004) Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom Rev 23*, 368-389.

37. Leary, J. A., Schenauer, M. R., Stefanescu, R., Andaya, A., Ruotolo, B. T., Robinson, C. V., Thalassinos, K., Scrivens, J. H., Sokabe, M., and Hershey, J. W. (2009) Methodology for measuring conformation of solvent-disrupted protein subunits using T-WAVE ion mobility MS: an investigation into eukaryotic initiation factors, *J Am Soc Mass Spectrom 20*, 1699-1706.

38. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water, *Science 310*, 1658-1661.

39. Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C., and Heck, A. J. (2008) Stability and shape of hepatitis B virus capsids in vacuo, *Angew Chem Int Ed Engl 47*, 6247-6251.

40. Barrera, N. P., Di Bartolo, N., Booth, P. J., and Robinson, C. V. (2008) Micelles protect membrane complexes from solution to vacuum, *Science 321*, 243-246.

41. Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., Seeger, M. A., Miguel, R. N., Korkhov, V. M., van Veen, H. W., Venter, H., Walmsley, A. R., Tate, C. G., and Robinson, C. V. (2009) Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions, *Nat Methods 6*, 585-587.

42. Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C., and Heck, A. J. (2008) High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids, *Proc Natl Acad Sci U S A 105*, 9216-9220.

43. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies, *Anal Chem 74*, 1402-1407.

44. Janis, J., Pasanen, S., Rouvinen, J., and Vainiotalo, P. (2008) Characterization of the pH-dependent dissociation of a multimeric metalloprotein Streptomyces rubiginosus xylose isomerase by ESI FT-ICR mass spectrometry, *J Mass Spectrom 43*, 1376-1380.

45. Kitova, E. N., Kitov, P. I., Bundle, D. R., and Klassen, J. S. (2001) The observation of multivalent complexes of Shiga-like toxin with globotriaoside and

the determination of their stoichiometry by nanoelectrospray Fourier-transform ion cyclotron resonance mass spectrometry, *Glycobiology 11*, 605-611.

46. van Duijn, E. (2010) Current limitations in native mass spectrometry based structural biology, *J Am Soc Mass Spectrom 21*, 971-978.

47. Benesch, J. L. (2009) Collisional activation of protein complexes: picking up the pieces, *J Am Soc Mass Spectrom 20*, 341-348.

48. Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F., and Robinson, C. V. (2006) Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies, *Chem Biol 13*, 597-605.

49. Benesch, J. L., and Robinson, C. V. (2006) Mass spectrometry of macromolecular assemblies: preservation and dissociation, *Curr Opin Struct Biol 16*, 245-251.

50. Jones, C. M., Beardsley, R. L., Galhena, A. S., Dagan, S., Cheng, G., and Wysocki, V. H. (2006) Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions, *J Am Chem Soc 128*, 15044-15045.

51. Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H., and Heck, A. J. R. (2007) Optimizing Macromolecular Tandem Mass Spectrometry of Large Non-Covalent Complexes Using Heavy Collision Gases, *Intl. J. Mass Spectrom. 268*, 198-206.

52. El-Faramawy, A., Guo, Y., Verkerk, U., Thomson, B. A., and Siu, M. (2008) Evaluation of IR Multiphoton Dissociation as a Method for High Mass Protein Clean Up, In *56th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, Co.

53. Felitsyn, N., Kitova, E. N., and Klassen, J. S. (2001) Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior, *Anal Chem 73*, 4647-4661.

54. Wysocki, V. H., Tsaprailis, G., Smith, L. L., and Breci, L. A. (2000) Mobile and localized protons: a framework for understanding peptide dissociation, *J Mass Spectrom 35*, 1399-1406.

55. Sohn, C. H., Chung, C. K., Yin, S., Ramachandran, P., Loo, J. A., and Beauchamp, J. L. (2009) Probing the Mechanism of Electron Capture and Electron Transfer Dissociation Using Tags with Variable Electron Affinity, *J Am Chem Soc 131*, 5444-5459.

56. Benesch, J. L., Ruotolo, B. T., Sobott, F., Wildgoose, J., Gilbert, A., Bateman, R., and Robinson, C. V. (2009) Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments, *Anal Chem 81*, 1270-1274.

57. Pagel, K., Hyung, S. J., Ruotolo, B. T., and Robinson, C. V. Alternate dissociation pathways identified in charge-reduced protein complex ions, *Anal Chem 82*, 5363-5372.

58. Geels, R. B., van der Vies, S. M., Heck, A. J., and Heeren, R. M. (2006) Electron capture dissociation as structural probe for noncovalent gas-phase protein assemblies, *Anal Chem 78*, 7191-7196.

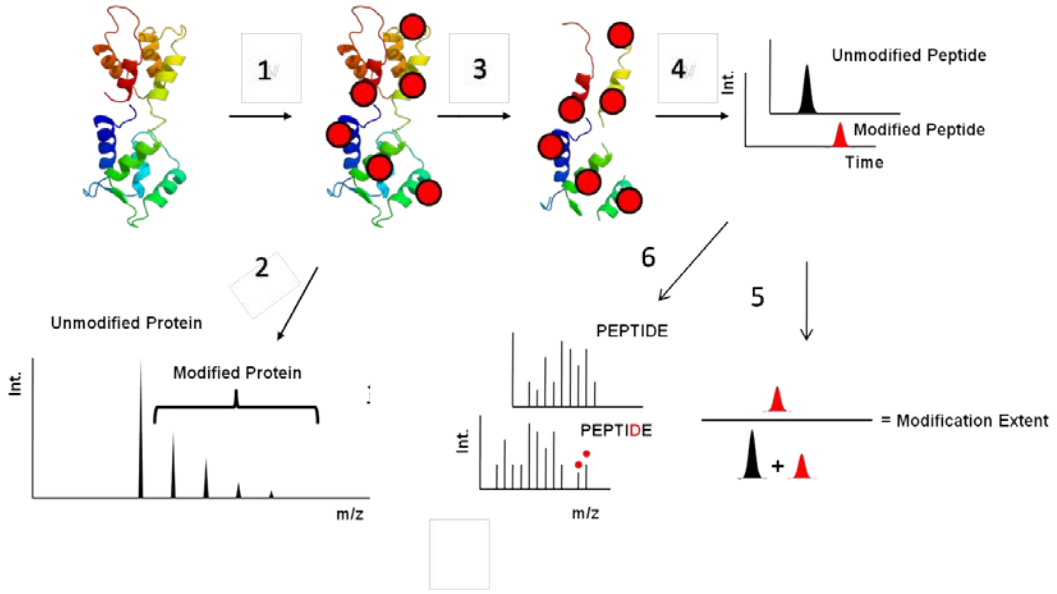


Figure 1.1.Work flow of MS based carboxyl group protein footprint (1) Modification (2) ESI-MS analysis of modified protein (3) proteolytic digestion (4) LC-MS analysis of peptides (5) calculation of the modified peptide level (6) LC-MS/MS analysis of peptides

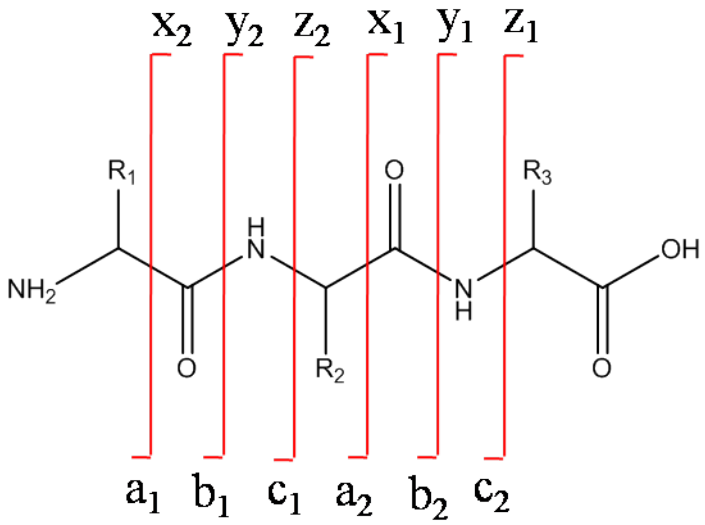


Figure 1.2.Nomenclature for fragment ions of peptide.

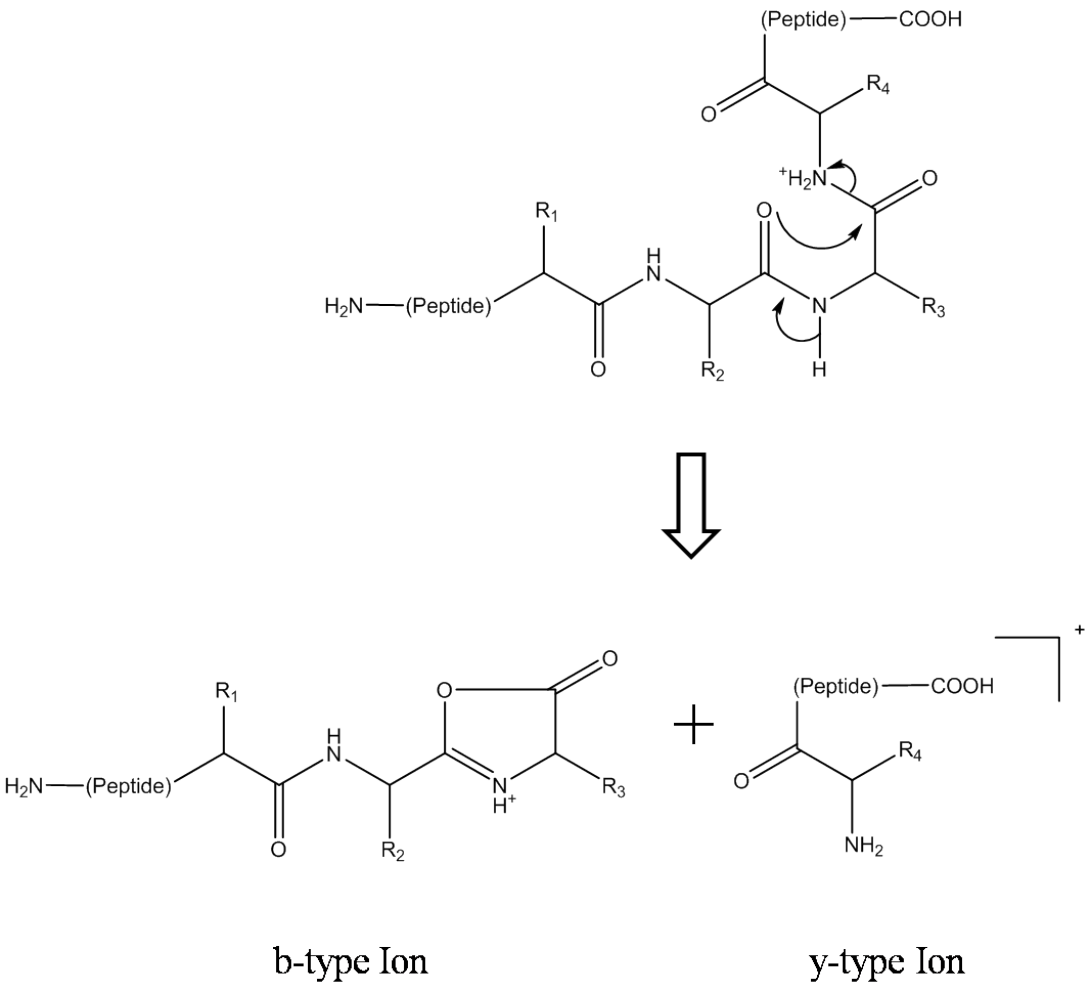


Figure 1.3.A proposed CID fragmentation mechanism.

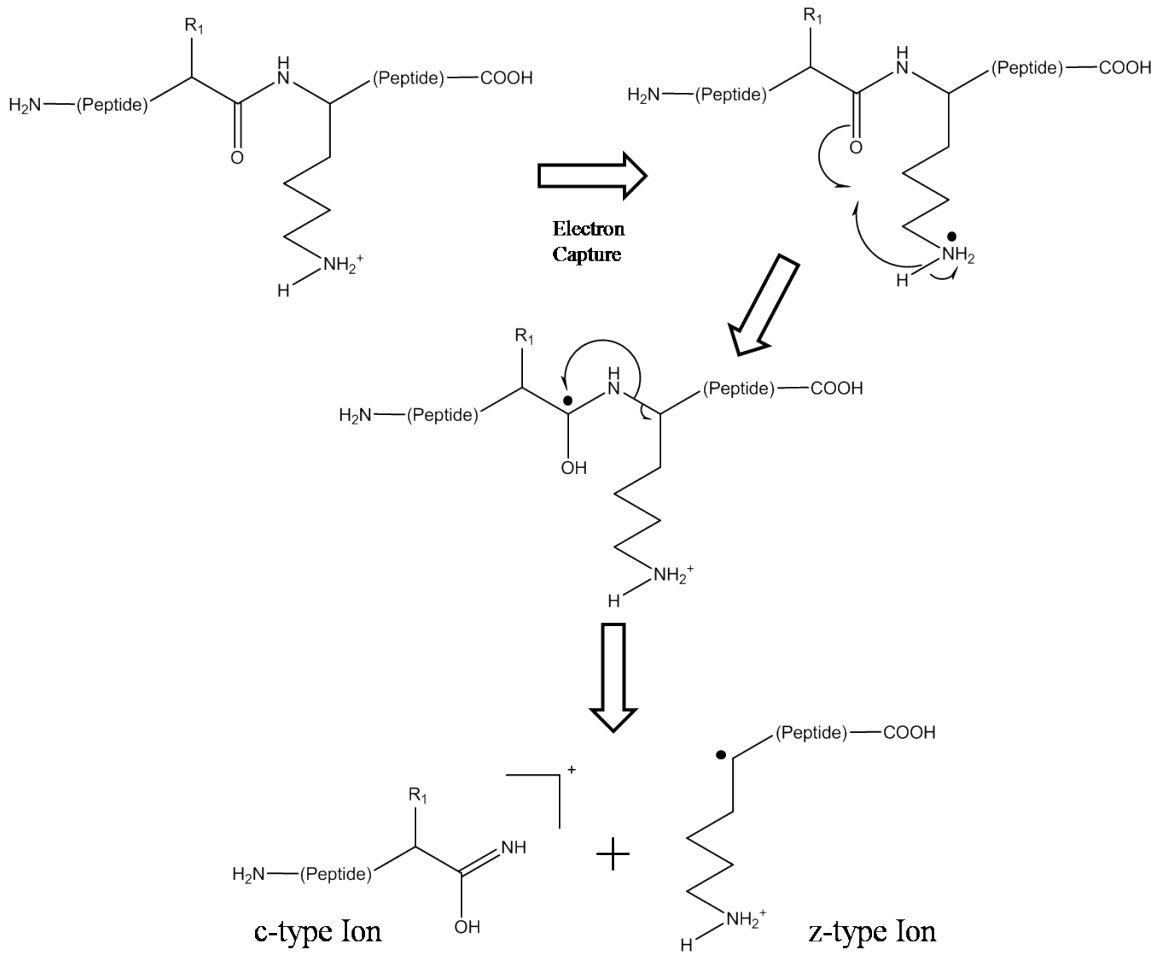


Figure 1.4.A proposed ECD fragmentation mechanism.

Chapter 2

Fast Photochemical Oxidation of Proteins (FPOP) for Comparing Structures of Protein/Ligand Complexes: The Calmodulin-peptide Model System

This chapter is based on recent publication:

H. Zhang, B. C. Gau, L. M. Jones, I. Vidavsky, M. L. Gross, Anal Chem. 2011, 83. 311-8.

Author contributions: H.Z., M.L.G. designed research; H.Z. performed research; H.Z., B.C.G., I.V. contributed analytic tool; H.Z., M.L.G. analyzed data; H.Z., B.C.G., L.M.J., M.L.G. discussed results; and H.Z. wrote the paper.

Abstract

Quantification of the modification extent for the apo and holo states of a protein-ligand complex should provide structurally sensitive information at the amino-acid level to compare the structures of unknown protein complexes with known ones. We report here the use of FPOP to monitor the structural changes of calmodulin in its established binding to M13 of the skeletal muscle myosin light chain kinase. We use the outcome to establish the unknown structures resulting from binding with melittin and mastoparan. The structural comparison follows from a comprehensive examination of the extent of FPOP modifications as measured by proteolysis and LC-MS/MS for each protein-ligand equilibrium. The results not only show that the three calmodulin-peptide complexes have similar structures but also reveal those regions of the protein that became more or less solvent-accessible upon binding. This approach has the potential for relatively high throughput, information-dense characterization of a series of protein-ligand complexes in biochemistry and drug discovery when the structure of one reference complex is known and related complexes are not, as is the case for calmodulin and M13 of the skeletal muscle myosin light chain kinase.

Introduction

In this paper, we use FPOP to show the effect of peptide ligands on calmodulin conformation. Our intention is to set forth a method for comparing structures of protein/ligand complexes when one or more reference structures are available. We chose as a model calmodulin (CaM), a small acidic calcium binding protein(*1*). Increased Ca^{2+}

concentration promotes specific CaM-Ca^{2+} binding; with the uptake of four Ca^{2+} ions, CaM undergoes a conformational changes to a wholly different state capable of ligand binding(*2*). As a Ca^{2+} sensor and signal transducer, Ca^{2+}-bound CaM binds to target proteins to alter their functions, acting as part of a calcium signal transduction pathway(*3*),(*4*). CaM mediates processes such as inflammation, metabolism, muscle contraction, intracellular movement, short-term and long-term memory, nerve growth and the immune response(*5*). One family of CaM targets is Ca^{2+}/Calmodulin–dependent protein kinases such as skeletal muscle myosin light chain kinase (SK-MLCK) (*6*). Some peptides also bind CaM to inhibit the calcium signal transduction pathway. Two such peptides are melittin (Mel) and mastoparan (Mas). Mel, the principal component of honeybee venom, is a 26-residue peptide(*7*), whereas Mas is a 14-residue peptide toxin from wasp(*8*).

To date, no high resolution 3D structures of CaM-Mel or CaM-Mas have been published, although studies by NMR(*9*), H/DX(*10*), fluorescence(*11*), enzyme cleavage(*12*), cross-linking(*13*) and chemical modification were reported(*14*, *15*). Previous studies demonstrated that both Mel and Mas share a target recognition and activation mechanism with the SK-MLCK binding domain (M13 peptide) (Figure 2.1), suggesting their structures are similar (*16*),(*17*, *18*) .

We tested the hypothesis of structural homology among the three CaM-ligand complexes by comparing their quantitative FPOP outcomes using spectra-contrast-angle evaluation (*19*). This method utilizes the CaM-M13 complex as a well-understood template structure for the interpretation of FPOP signals of the CaM-Mel and CaM-Mas complexes. By this example, we demonstrate a method for the detailed characterization

of protein-ligand interactions. Such an approach should be applicable to protein-ligand systems where sub libraries of ligands that are known to bind can be screened for similarity of interaction relative to a standard protein-ligand complex; the approach has higher throughput than NMR and X-ray crystallography. Application of this method for the rapid analysis of protein-ligand complexes may prove useful for studies of disease-related protein complexes, especially in drug discovery.

Material and Methods

Materials

Bovine CaM (calcium free) was obtained from Ocean Biologics Co. (Edmonds, WA). Hydrogen peroxide, water, acetonitrile, formic acid, calcium chloride, phosphate buffered saline powder, EGTA (ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetra-acetic acid), *L*-methionine, *L*-glutamine, melittin from honeybee venom (MW 2846), mastoparan from *Vespula lewisii* (MW 1478), catalase from bovine liver, and trypsin from porcine pancreas at the highest purity available were purchased from Sigma-Aldrich (St. Louis, MO). Skeletal muscle myosin light chain kinase peptide (M13), ^{1}KRRWKKNFIAVSAANRFKKISSSGAL^{26}, was purchased from AnaSpec, Inc. (Fremont, CA).

Protein and peptide stock solution

All proteins and peptides were received as lyophilized powder. A stock solution of CaM was prepared in water, and its concentration determined by UV-absorbance at 280 nm (ε = 2980 M^{-1} cm^{-1}). All FPOP samples contained 10 μM CaM in 10 mM phosphate buffered saline (PBS buffer, 138 mM NaCl, 2.7 mM KCl, pH = 7.4) and 20 mM *L*-glutamine. Peptides, Mel, Mas, M13, were used without further purification.

Formation of Complexes

Eight types of samples were prepared from the CaM stock solution and PBS buffer: Ca^{2+}-free-CaM (100 μM EGTA), Ca^{2+}-free-CaM-Mel (100 μM EGTA, 20 μM Mel), Ca^{2+}-free-CaM-Mas (100 μM EGTA, 20 μM Mas), Ca^{2+}-free-CaM-M13 (100 μM EGTA, 20 μM M13), Ca^{2+}-bound-CaM (100 μM calcium chloride), Ca^{2+}-bound-CaM-Mel complex (100 μM calcium chloride, 20 μM Mel), Ca^{2+}-bound-CaM-Mas complex (100 μM calcium chloride, 20 μM Mas), and Ca^{2+}-bound-CaM-M13 complex (100 μM calcium chloride, 20 μM M13). The protein-peptide ratio in the sample solutions was 1:2. All samples were incubated overnight at 22 °C to allow for equilibration prior FPOP labeling.

FPOP Protocol

The protocol used here was based on a previous report (*20*) with minor changes. A 248 nm KrF excimer laser (GAM Laser Inc, Orlando, FL) tuned to 40 mJ/pulse, was used to irradiate the sample solution. The laser was focused through a 250 mm convex lens (Edmunds Optics, Barrington, NJ) onto 150 μm i.d. fused silica tubing (Polymicro Technologies, Pheonix, AZ) located 125 mm from the lens, giving a 2.5 mm irradiation window; the fused silica polyimide coating in this region was first removed by using a propane torch. The flow rate and pulse frequency were adjusted to guarantee that1.2 x the reaction volume vacated the window between laser shots. The laser pulse frequency was controlled by an external pulse generator (B&K Precision, Yorba Linda, CA).

Hydrogen peroxide was added to each sample to a final concentration of 15 mM just prior to FPOP. Samples of 50 μL were infused through the apparatus and collected in tubes containing 10 μL of 50 nM catalase and 20 mM methionine. Residual hydrogen

peroxide was decomposed by catalase treatment at room temperature for 10 min before storage at 4 °C. FPOP labeling was done in triplicate for each sample type.

ESI MS Analysis with a Q-TOF Mass Spectrometer

All ESI mass spectra were acquired in the positive-ion mode on a Waters (MicroMass) Q-TOF Ultima (Manchester, U.K.) equipped with a Z-spray ESI source. The instrument setup for protein analysis was similar to that reported previously.(*21*)

Trypsin Proteolysis Protocol

Digestion of CaM FPOP sample was conducted according to the previously reported protocol without modification (*22*) .

LC-ESI-MS/MS Analysis of Protein Digest

Digested samples (1 μL) were diluted in 100 μL water with 0.1% formic acid. An aliquot of 5 μL diluted solution was loaded onto a silica capillary column with a PicoFrit™ tip (New Objective, Inc., Woburn, MA) that was custom-packed with C18 reverse phase material (Magic, 0.075 mm × 150 mm, 5 μm, 120 Å, Michrom Bioresources, Inc., Auburn, CA). The gradient was pumped with an Eksigent NanoLC-1D ultra (Eksigent Technologies, Inc. Livermore, CA) at 260 nL/min, from 2% to 60% solvent B (acetonitrile, 0.1% formic acid) over 60 min, then to 80% solvent B for 10 min, followed by a 12 min re-equilibration step by 100% solvent A (water, 0.1% formic acid). The flow was directed by a PicoView Nanospray Source (PV550, New Objective, Inc., Woburn, MA) to an LTQ Orbitrap (Thermo-Scientific, San Jose, CA) with a spray voltage of 1.8-2.0 kV, and a capillary voltage of 27 V. The LTQ Orbitrap was operated in the standard, data-dependent acquisition mode controlled by Xcalibur 2.0.7 software. Peptide mass spectra (*m/z* range: 350-2000) were acquired at high mass resolving power

(60,000 for ions of *m/z* 400) in FT mode. The top six most abundant multiply charged ions with minimal signal intensity at 1000 counts were subjected to CID (collision-induced dissociation) in the linear ion trap. Precursor activation was performed with an isolation width of 2 Da and an activation time of 30 ms. The normalized collision energy was set at 35%. The automatic gain control target value was regulated at 1×10^6 for the FT analyzer and 3×10^4 for ion-trap analyzer with a maximum injection time of 1000 ms for FT and 200 ms for ion trap.

LC-MS Feature Annotation

The LC-MS/MS data were searched for modified and unmodified CaM tryptic peptides by using Mascot 2.2.06 (Matrix Science, London, UK)(*23*) and the NCBI database. All known •OH-side chain reaction products were added into modification database for searching as variable modifications. Modification site assignments were validated by manual inspection of product-ion spectra.

LC-MS/MS .raw files were imported into the Rosetta Elucidator system (v3.3.0.0.220, Rosetta Biosoftware, Seattle, WA) for retention time alignment of shared LC-MS features (*24*). The aligned retention time and peak volume of all high resolution extracted ion chromatogram features across a 5-70 min window were output to a spreadsheet for each sample. Each product-ion spectrum was paired with its LC-MS feature by using its retention time and precursor *m/z*; a table of this pairing was used by the Excel macro written in our lab to annotate all Elucidator-determined LC-MS features with the Mascot assignments linked to their product-ion spectra.

Calculation of Modification Extent

The fraction modified (F_{ox1} in Eq. 1) was calculated for any specific residue as the ratio of signal intensities of each peptide (I_{ox1}) modified at this residue to the total intensities of modified and unmodified peptide signals spanning this residue ($I + I_{ox1} + I_{ox2} + ...+ I_{oxn}$). The changes in modification (R) at a site between Ca^{2+}-free and Ca^{2+}-bound states (Eq. 2) were calculated at the peptide level by using the modification fraction ($F_{\text{Ca free}}$ and $F_{\text{Ca bound}}$) of each peptide, wherein the signal from all modifications on the peptide contribute to the numerator in equation 1.

$$F_{ox\,1} = \frac{\sum I_{ox\,1}}{\sum_{i=1}^{n} I_{ox\,i} + \sum I} \qquad \text{Eq. 1}$$

$$R_{ox} = \frac{F_{Ca\ bound} - F_{Ca\ free}}{F_{Ca\ free}} \qquad \text{Eq. 2}$$

To identify changes induced by complexation, results from CaM (no complex) were used as controls in both comparisons.

Results and discussion

ESI MS of calmodulin submitted to FPOP

Beginning with CaM in the presence of Mel or Mas, we detected by ESI MS multiply charged CaM molecules (Figure 2.2a) that have a Gaussian charge-state distribution centered at the most abundant 15+ ion. We also saw ions for Mel and Mas at lower *m/z*, consistent with dissociation of the complex under ESI conditions. Trace signals for oxidized Mel/Mas were observed as well. The deconvoluted mass spectrum showed that the molecular weight (MW) of CaM is 16790 Da, consistent with CaM

having the expected post-translational modifications of N-terminal acetylation (+42 Da) and trimethylation of K115 (+42 Da) (*22*).

After submitting CaM to FPOP, we found evidence in the mass spectra for significant oxidative modification (Figure 2.2b) as mass shifts +16, 32, 48… Da. The distribution of oxidation products is Poisson-like, serving as an indication that FPOP is fast and probed only a single state of the target protein. This conclusion is consistent with recent results(*25*) showing that the distribution of oxidation products of CaM and two other proteins sensitive to oxidative-induced unfolding are well modeled by a Poisson distribution. The conformation sampled by FPOP for these experiments is singular and invariant during labeling (i.e., the protein did not unfold on the labeling timescale).

The abundances of the +16, +32, and +48 Da modification products, as estimated from their +15 charge state intensities (Figure 2.2b-d), show that slightly more protein is modified in the Ca^{2+}-bound, ligand-free state than the Ca^{2+}-free and Ca^{2+}-bound CaM-ligand states. The Ca^{2+}-free and Ca^{2+}-bound structures are markedly different. Given that the global FPOP product distributions do not substantially distinguish these two, we turned to peptide and residue-level analysis for their and CaM-peptide complexes' structural footprints.

Information Content of LC-ESI-MS/MS proteolyzed FPOPsamples

Mascot processing of the LC-MS/MS data of the trypsin-proteolyzed FPOP samples showed 93% sequence coverage for the three CaM complexes. The aforementioned PTMs were also identified by using Mascot. Of the 12 peptides identified, nine had product-ion spectra for both the oxidatively modified and unmodified

forms for all samples, spanning over 70% of protein sequence. Calmodulin 38-74, a large tryptic peptide, was not detected probably owing to high retention upon sample handling and chromatography. This peptide was not included among the peptides used for quantitative analysis. Small tryptic peptides from ligand peptides (Mel/Mas/M13) were also observed. Owing to the excess ligand present at equilibrium prior to FPOP, labeling occurs for both bound and free ligands; their differential analysis is consequently not possible.

Minimization of post-FPOP oxidative-modification bias

Several factors give rise to post-FPOP oxidation on peptides, including methionine oxidation during proteolytic digestion, sample handling (*26*), and electrospray ionization (ESI) (*27*). Of these modification-biasing signals, ESI-induced oxidation is the easiest to distinguish and thereby to be excluded in our analysis. The reason that we are not misled by ESI-induced oxidation is the hydrophilicity of a peptide is nearly always increased with oxidative-modification so that its reverse-phase LC retention time is earlier than that of its unmodified peptide (Figure 2.3). We routinely observe this in all FPOP labeling experiments (data not shown). As co-elution of an oxidized peptide with its unmodified counterpart is unlikely, we attribute any oxidatively modified peptides co-eluting with the unmodified counterpart to ESI-induced oxidation.

A recent study of post-FPOP sample handling showed that proteins stored in millimolar levels of hydrogen peroxide can oxidize while frozen and at lower temperatures (*28*). As a precaution, we stored FPOP-treated samples at 4 °C after removing hydrogen peroxide with catalase immediately following FPOP labeling. To this mixture was added free methionine to curtail further any post-FPOP oxidation (*29*).

Changes of modification at peptide and residue level

The extents of modification can be compared for CaM in the presence of various peptides with and without bound calcium (Figure 2.4). Comparisons between CaM-Mel/Mas/M13 complexes (Figure 2.4b-d) and CaM (Figure 2.4a) reveal that CaM peptides fall into three groups as discussed in following paragraphs.

Peptides 76-86, 78-86 and 31-37 belong to group I; they have similar trends of modification change for both CaM and CaM-Mel/Mas/M13 complexes. Peptides 76-86 and 78-86 peptides originate from the linker region of CaM and display increased labeling in their Ca^{2+}-bound vs. the Ca^{2+}-free states, whereas peptide 31-37 shows decreased labeling. More importantly, their modification extents are invariant with addition of the peptide ligands (Mel/Mas/M13), indicating that the regions represented by these peptides are not directly involved in any CaM-Mel/Mas/M13 interaction.

Peptides 14-21, 22-30, 107-126, and 127-148 belong to group II; they exhibit different labeling trends for the CaM-Mel/Mas/M13 complexes than for CaM itself. The extents of oxidative modification decrease for peptides 14-21, 107-126 and 127-148 when complexed with peptide ligands (Figure 2.4b,c,d) compared to Ca^{2+}-bound CaM in the absence of peptide ligands, where increases occur (Figure 2.4a). Reverse trends pertain to peptide 22-30. The most significant decrease in oxidative modification occurs for peptide 14-21. The differential labeling between CaM and the CaM-Mel/Mas/M13 complexes indicates that, upon forming the complexes, these regions of CaM are either buried allosterically or directly protected by the ligand.

Group III peptides 1-13 and 95-106 do not share a single labeling trend when CaM binds to peptides. The *N*-terminus peptide, 1-13, from CaM and the CaM-Mel

bound complex (Figure 2.4a, b) show a decrease in oxidative modification whereas that peptide from CaM-Mas and CaM-M13 complexes (Figure 2.4c,d) undergo increased labeling. For peptide 95-106, the trend for CaM, CaM-Mel, and CaM-Mas (Figure 2.4a,b,c) showed decreased labeling whereas the CaM-M13 (Figure 2.4d) showed increased oxidative labeling.

Taken together, we propose that the patterns for all nine peptides compose a unique "fingerprint" of the protein-ligand complex. Although these data are not residue-resolved, they are readily accessible for comparing structures of the complexes because the modification extent of a peptide is a function of the aggregate of its modifiable residues' solvent accessibilities (Figure 2.5a). Examining these fingerprints should establish whether the CaM-Mel/Mas structures are similar to the known structure of CaM/M13 complex.

Turning to product-ion spectra (Figure 2.6), we identified 14 modified residues. Seven are at the N-terminal domain, two from the linker region, and five from the C-terminal domain. Nine have side chains that contain sulfur (M36, M76, M109, M124, M144, and M145) or aromatic rings (F16, F19, Y99), all of which are highly reactive with •OH. The extent of modification is highest for Met (with modification fractions > 0.2). Five aliphatic amino acids (L4, I9, L18, I27, and I85) undergo detectable amounts of modification. The relative modification yields between these residues agree well with those reported in previous studies of the •OH reactivity of amino acids (*30*) . Although we can detect other low-abundance modifications, we did not include them because their assignments based on product-ion spectra were problematic. Instead, we focused on the 14 oxidized residues as a source of comparative structural information of CaM complexes

(Figure 2.7). Peptides containing more than one highly reactive site and undergoing double modifications are considered separately.

Higher resolution provided by residue level FPOP data taken by MS/MS should afford more detailed views of CaM-peptide complexes. Those "zoom-in" views help to elucidate residue-level interactions between the protein and the peptide ligand, a view that is missing in the peptide mass spectra. In the case of peptide 14-21, its modified residues F16, L18, and F19, however, do not share the same trends shown for the full peptide. L18 and F19 show higher modification levels for CaM than for the complexes, which is the same trend seen at the peptide level. F16, however, is modified to nearly the same extent for CaM as for the complexes. Thus, one model of complexation puts L18 and F19 in direct interaction with the peptide ligand and F16 away from this interface. These data for peptide 14-21 show that analysis at the amino-acid level, made efficient because the modifications are irreversible, provides a more resolved picture of protein-ligand interaction than methodologies that only use peptide-level data. Integrating high-resolution views with a reference structure of CaM-M13 can allow the structures of unknown CaM-Mel/Mas complexes to be compared (Figure 2.5b).

Spectral-contrast-angle comparison of the modification patterns of peptides

At this point, one may conclude from analysis at both the peptide level and the amino-acid residue level that the modification extents in the peptide-bound and free states for the three complexes are similar. To test more rigorously the similarity, we employed a spectral-contrast-angle (θ) analysis of the nine peptide modification trends. This spectra-contrast-angle approach can validate the comparison process by providing a confident value (θ) related to similarity. It was used previously to compare the product-

ion spectra of oligodeoxynucleotide isomers (*19*) . The changes of oxidative modification of the various peptides from CaM can be treated in similar way. Although the sequence coverage of the tryptic peptides is ~70%, this statistical approach still is suitable for making comparisons and does not require full coverage. For comparing the CaM-peptide complex with the peptide-free sample, the set of peptide signals can be represented as a basis vector in an N-dimensional space (a_i). In this analysis, each signal is the change in modification level for the calcium-bound vs. calcium-free state. A comparison vector (b_i) is comprised of the same signal data from a different sample replicate or CaM-peptide mixture. The spectral-contrast-angle (Eq. 3) of these vectors provides a single parameter reflecting their similarity; as the similarity increases, $\theta \rightarrow 0$.

The first four entries in Table 2.1 are the average θ for the three pair-wise comparisons of replicates from the independent triplicates of a CaM-peptide or absent-peptide FPOP experiments. These entries provide an expectation level for near identity of $\theta < 35$. The θ determined from the average of each CaM vs CaM-peptide comparison is significantly larger (Table 2.1, 5th entry). This implies a significant change in structure with peptide binding, as is expected, and the small standard deviation conveys the overall similarity of this change among the three peptide complexes. The pair-wise comparisons between each CaM-peptide complex give an average θ, which is similar to the smallest angle from replicate experiments (Table 2.1 last entry). This similarity further supports the conclusion that the two unknown CaM-peptide complexes have similar structures and that they resemble the known CaM-M13 structure. We propose that this statistical approach is a useful means of comparing protein/ligand structures for a series of ligands when one structure is known and can be used as a reference.

$$\cos\theta = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2 \sum_i b_i^2}} \qquad \text{Eq. 3}$$

Comparative structure assignment from FPOP data

The approach of utilizing the "fingerprint" of changes in extent of modification at the peptide and residue levels is more empirical than but complements that of Chance and coworkers (*31*, *32*) who showed that the combination of •OH surface mapping data with computational modeling provides important protein structural data. Our approach not only allows conclusions about the similarities of the protein-ligand structures but also permits establishment of some structural details. Proteins or peptides (including Mel, Mas) that bind to CaM are likely to bind to both N-term and C-term hydrophobic clefts by forming a long alpha helix in a manner similar to that of SK-MLCK M13 peptide (*16*). In addition to the NMR structure (PDB ID: 2BBM) of CaM-M13 complex, the structures of Ca^{2+}-free CaM from NMR (PDB ID: 1CFD) and Ca^{2+}-bound CaM from x-ray crystal structure (PDB ID: 1CLL) are also known (Figure 2.1). The FPOP results for the established structure of the CaM-M13 complex provide a basis for a more detailed structural analysis. For example, we see that L18, F19, M109, M124, M144 and M145 from both N and C terminal domains of the three complexes are protected against •OH reaction upon complex formation with the various ligand peptides. All six are hydrophobic and are located at the interface between CaM and M13, as seen in the NMR structure (Figure 2.5b). M109, M124, M144 and M145 are located inside the hydrophobic cleft of the C-terminus and make contact with the M13 hydrophobic side chains. For the CaM-M13 structure, L18 and F19, which are at the hydrophobic-cleft

center of the N-terminus, become less solvent-accessible when the Ca^{2+}-bound CaM binds with M13. On the other hand, F16 is not part of the hydrophobic cleft in the CaM-M13 complex, but it is hydrophobic and close in proximity to L18 and F19 (Figure 2.5b).

The L4 and I9 side chains have relatively similar solvent accessibilities in CaM and the CaM-M13 complex, suggesting that these residues share similar conformations in CaM and in the peptide complexes. Two other interesting sites are I27 and Y99, which are located in the EF hands. Y99 is in a short anti parallel beta sheet between adjacent calcium binding loops. The aromatic side chain of Y99 points outside the molecule. Owing to its high solvent accessibility and its inherent high •OH reactivity, this tyrosine residue is particularly sensitive to slight changes in structure. This suggests the peptide-binding-induced deprotection with M13 and induced protection with Mel and Mas are small changes although significantly different. I27, however, is on the small loop connecting two alpha helices of the first EF hand motif. Its side chain is inside the hydrophobic cleft in all three of the high resolution CaM structures (Ca^{2+}-free CaM, Ca^{2+}-bound CaM and CaM-M13 complex), consistent with the low reactivity and lack of difference between Ca^{2+}-loaded CaM and the peptide complexes.

Conclusion

Comparison of FPOP data taken for a protein complex of known structure to the FPOP-induced modifications of unknowns by using spectral-contrast-angle provides a way to test whether a protein has similar or different structures when it binds to various ligands. Specifically, a statistical analysis of the modification extent of peptide regions of CaM extricated by tryptic digestion shows that two peptide ligands (i.e., Mel and Mas) bind similarly in the presence of Ca^{2+} as does a third ligand (M13) that forms a known

structure with CaM. Although one could use other proteases in the digestion to obtain more peptides and thereby increase certainty, we did not do this to show that this comparative approach is effective for distinguish the structural differences even when the sequence coverage is not 100%. An advantage of FPOP coupled with MS is that one can examine the modification extents at the amino-acid level to identify those residues that are involved in the conformational changes induced by ligand binding. Those residues that show differential modification levels in the structure of CaM-M13 complex are those that are expected on the basis of the high resolution 3D structure from NMR.

The MS-based FPOP method has the ability to probe structures on a comparative basis for small amounts of protein (pmoles) in relatively short times (fractions of a week) compared to NMR or X-ray crystallography. These simple comparisons may have application in screening many protein/ligand interactions including those in drug development. The FPOP data serve as a "fingerprint" for the ligand interactions. Combining the "fingerprints" with those from H/DX, specific chemical modification (e.g., acetylation of Lys), and cross linking can give detailed information about protein-ligand interactions with reasonable throughput.

Acknowledgements

We thank Dr. Justin Sperry for helpful suggestions and Dr. Henry Rohrs for assistance with the LTQ-Orbitrap studies. This research was supported by the National Centers for Research Resource of the NIH (Grant NO. P41RR000954). Additional support was provided by Merck; MLG is a consultant for Merck.

References

1. Weinstein, H., and Mehler, E. L. (1994) Ca(2+)-binding and structural dynamics in the functions of calmodulin, *Annu Rev Physiol 56*, 213-236.
2. Wriggers, W., Mehler, E., Pitici, F., Weinstein, H., and Schulten, K. (1998) Structure and dynamics of calmodulin in solution, *Biophys J 74*, 1622-1639.
3. Tanaka, T., and Hidaka, H. (1980) Hydrophobic regions function in calmodulin-enzyme(s) interactions, *J Biol Chem 255*, 11078-11080.
4. Degrado, W. F. (1988) Design of peptides and proteins, *Adv Protein Chem 39*, 51-124.
5. Bachs, O., Agell, N., and Carafoli, E. (1992) Calcium and calmodulin function in the cell nucleus, *Biochim Biophys Acta 1113*, 259-270.
6. Blumenthal, D. K., Takio, K., Edelman, A. M., Charbonneau, H., Titani, K., Walsh, K. A., and Krebs, E. G. (1985) Identification of the calmodulin-binding domain of skeletal muscle myosin light chain kinase, *Proc Natl Acad Sci U S A 82*, 3187-3191.
7. Kataoka, M., Head, J. F., Seaton, B. A., and Engelman, D. M. (1989) Melittin binding causes a large calcium-dependent conformational change in calmodulin, *Proc Natl Acad Sci U S A 86*, 6944-6948.
8. Matsushima, N., Izumi, Y., Matsuo, T., Yoshino, H., Ueki, T., and Miyake, Y. (1989) Binding of both Ca2+ and mastoparan to calmodulin induces a large change in the tertiary structure, *J Biochem 105*, 883-887.

9. Seeholzer, S. H., Cohn, M., Putkey, J. A., Means, A. R., and Crespi, H. L. (1986) NMR studies of a complex of deuterated calmodulin with melittin, *Proc Natl Acad Sci U S A 83*, 3634-3638.

10. Zhu, M. M., Rempel, D. L., Zhao, J., Giblin, D. E., and Gross, M. L. (2003) Probing Ca2+-induced conformational changes in porcine calmodulin by H/D exchange and ESI-MS: effect of cations and ionic strength, *Biochemistry 42*, 15388-15397.

11. Chapman, E. R., Alexander, K., Vorherr, T., Carafoli, E., and Storm, D. R. (1992) Fluorescence energy transfer analysis of calmodulin-peptide complexes, *Biochemistry 31*, 12819-12825.

12. Scaloni, A., Miraglia, N., Orru, S., Amodeo, P., Motta, A., Marino, G., and Pucci, P. (1998) Topology of the calmodulin-melittin complex, *J Mol Biol 277*, 945-958.

13. Schulz, D. M., Ihling, C., Clore, G. M., and Sinz, A. (2004) Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes, *Biochemistry 43*, 4703-4715.

14. Steiner, R. F., Albaugh, S., Fenselau, C., Murphy, C., and Vestling, M. (1991) A mass spectrometry method for mapping the interface topography of interacting proteins, illustrated by the melittin-calmodulin system, *Anal Biochem 196*, 120-125.

15. Wong, J. W., Maleknia, S. D., and Downard, K. M. (2005) Hydroxyl radical probe of the calmodulin-melittin complex interface by electrospray ionization mass spectrometry, *J Am Soc Mass Spectrom 16*, 225-233.

16. Yap, K. L., Kim, J., Truong, K., Sherman, M., Yuan, T., and Ikura, M. (2000) Calmodulin target database, *J Struct Funct Genomics 1*, 8-14.

17. Heidorn, D. B., Seeger, P. A., Rokop, S. E., Blumenthal, D. K., Means, A. R., Crespi, H., and Trewhella, J. (1989) Changes in the structure of calmodulin induced by a peptide based on the calmodulin-binding domain of myosin light chain kinase, *Biochemistry 28*, 6757-6764.

18. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR, *Science 256*, 632-638.

19. Wan, K. X., Vidavsky, I., and Gross, M. L. (2002) Comparing similar spectra: from similarity index to spectral contrast angle, *J Am Soc Mass Spectrom 13*, 85-88.

20. Hambly, D., and Gross, M. L. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin *International Journal of Mass Spectrometry 259*, 124-129.

21. Sperry, J. B., Shi, X., Rempel, D. L., Nishimura, Y., Akashi, S., and Gross, M. L. (2008) A mass spectrometric approach to the study of DNA-binding proteins: interaction of human TRF2 with telomeric DNA, *Biochemistry 47*, 1797-1807.

22. Sharp, J. S., and Tomer, K. B. (2007) Analysis of the oxidative damage-induced conformational changes of apo- and holocalmodulin by dose-dependent protein oxidative surface mapping, *Biophys J 92*, 1682-1692.

23. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis 20*, 3551-3567.

24. Neubert, H., Bonnert, T. P., Rumpel, K., Hunt, B. T., Henle, E. S., and James, I. T. (2008) Label-free detection of differential protein expression by LC/MALDI mass spectrometry, *J Proteome Res 7*, 2270-2279.

25. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding, *Analytical Chemistry 81*.

26. Xu, G., Kiselar, J., He, Q., and Chance, M. R. (2005) Secondary reactions and strategies to improve quantitative protein footprinting, *Anal Chem 77*, 3029-3037.

27. Boys, B. L., Kuprowski, M. C., Noel, J. J., and Konermann, L. (2009) Protein oxidative modifications during electrospray ionization: solution phase electrochemistry or corona discharge-induced radical attack?, *Anal Chem 81*, 4027-4034.

28. Hambly, D. M., and Gross, M. L. (2009) Cold chemical oxidation of proteins, *Anal Chem 81*, 7235-7242.

29. Saladino, J., Liu, M., Live, D., and Sharp, J. S. (2009) Aliphatic peptidyl hydroperoxides as a source of secondary oxidation in hydroxyl radical protein footprinting, *J Am Soc Mass Spectrom 20*, 1123-1126.

30. Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes, *Annu Rev Biophys Biomol Struct 35*, 251-276.

31. Zheng, X., Wintrode, P. L., and Chance, M. R. (2008) Complementary structural mass spectrometry techniques reveal local dynamics in functionally important regions of a metastable serpin, *Structure 16*, 38-51.

32. Kamal, J. K., and Chance, M. R. (2008) Modeling of protein binary complexes using structural mass spectrometry data, *Protein Sci 17*, 79-94.

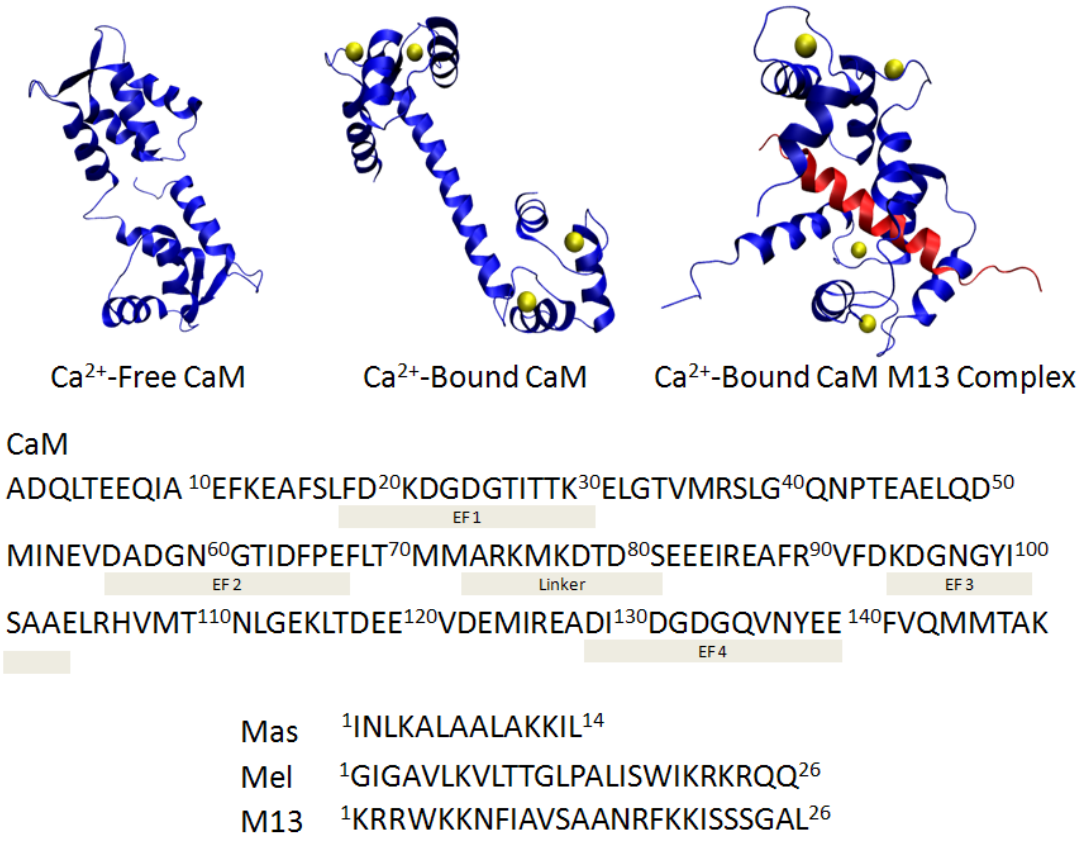


Figure 2.1. 3D structure of CaM and CaM-peptide complex: Ca^{2+}-free CaM structure from the average NMR structure (PDB ID: 1CFD). Ca^{2+}-bound CaM structure from the x-ray crystal structure (PDB ID: 1CLL). Ca^{2+}-bound CaM-M13 complex structure from the average NMR structure (PDB ID: 2BBM). M13 peptide is in pink. Ca^{2+} is displayed as yellow balls.

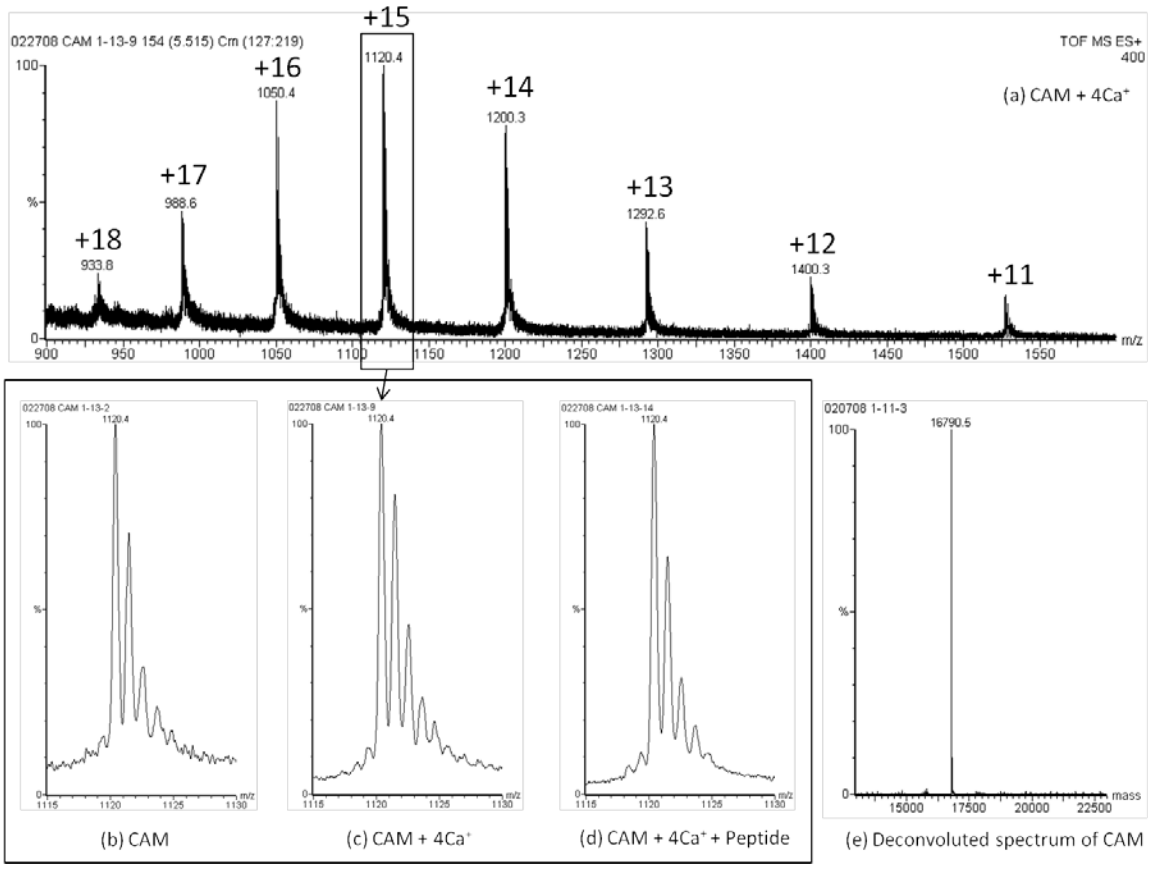


Figure 2.2. ESI Q-TOF mass spectrum: (a) Ca^{2+}-bound CaM (b) Ca^{2+}-free CaM in extended view (c) Ca^{2+}-bound CaM in extended view (d)Ca^{2+}-bound CaM-Mel in extended view (e) deconvoluted spectrum of CaM.

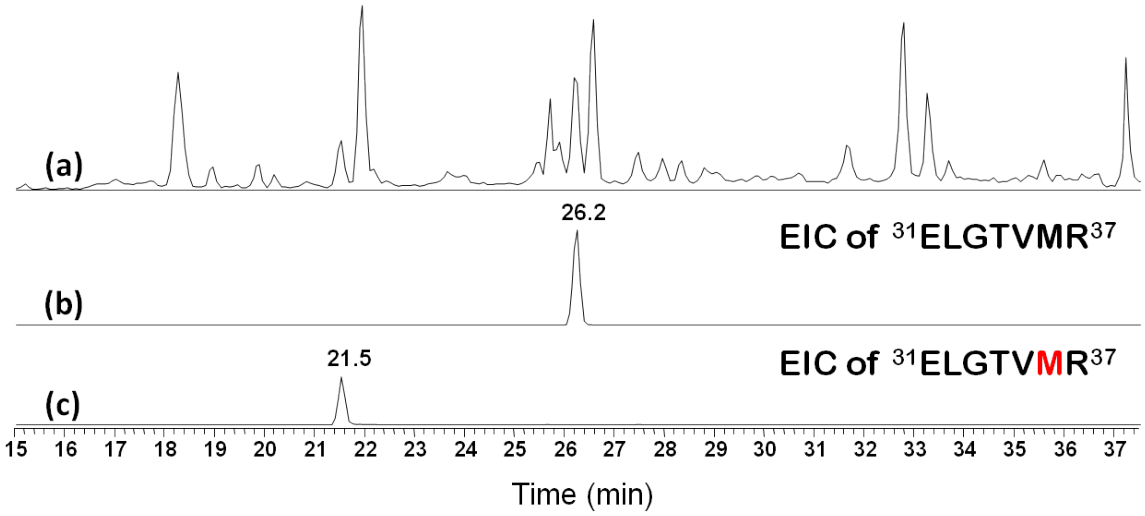


Figure 2.3. LC-MS results of CaM FPOP samples. (a) The Total Ion Chromatogram (TIC) of LC-MS experiment with all peptide peaks were labeled with different retention time. (b) Extracted Ion Chromatogram (EIC) of peptide 31-37 using doubly charged ion. (c) EIC of oxidized peptide 31-37 from doubly charged ion.

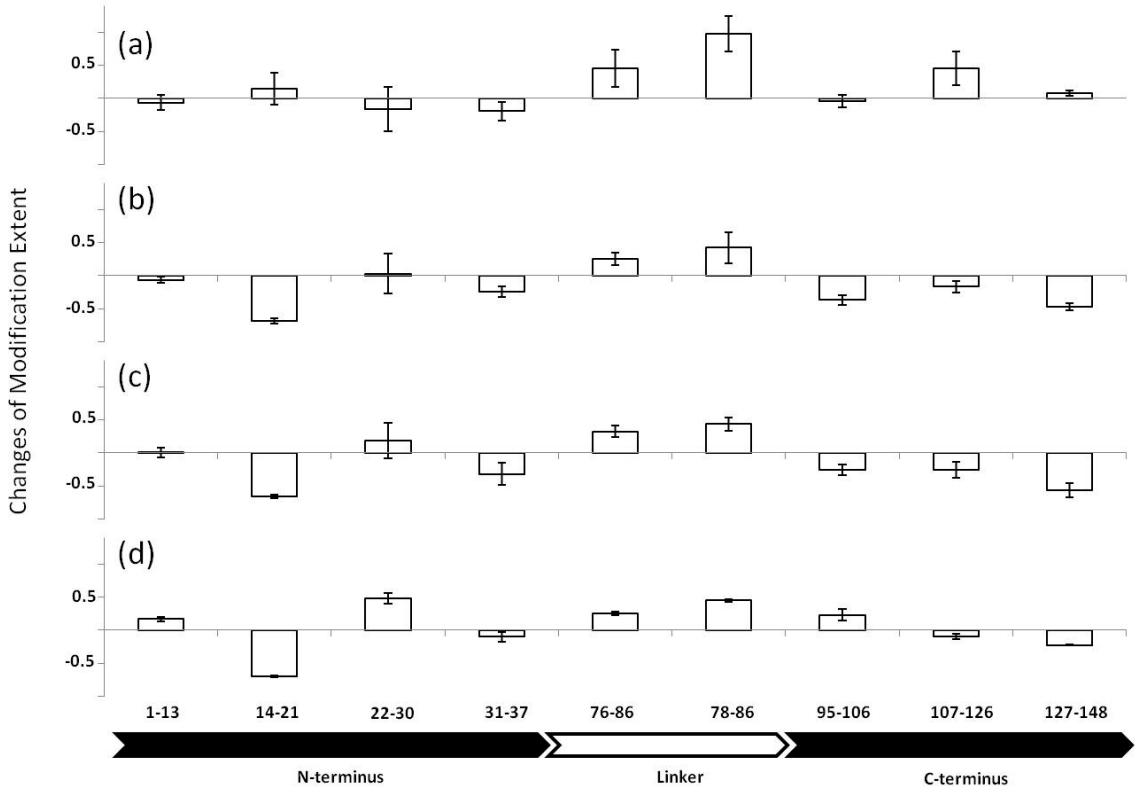


Figure 2.4. Changes of modification extent from Ca2+-free to Ca2+-bound states for various CaM tryptic peptides from different CaM complexes. (a) CaM itself, (b) CaM-Mel Complex, (c) CaM-Mas Complex, (d) CaM-M13 complex.

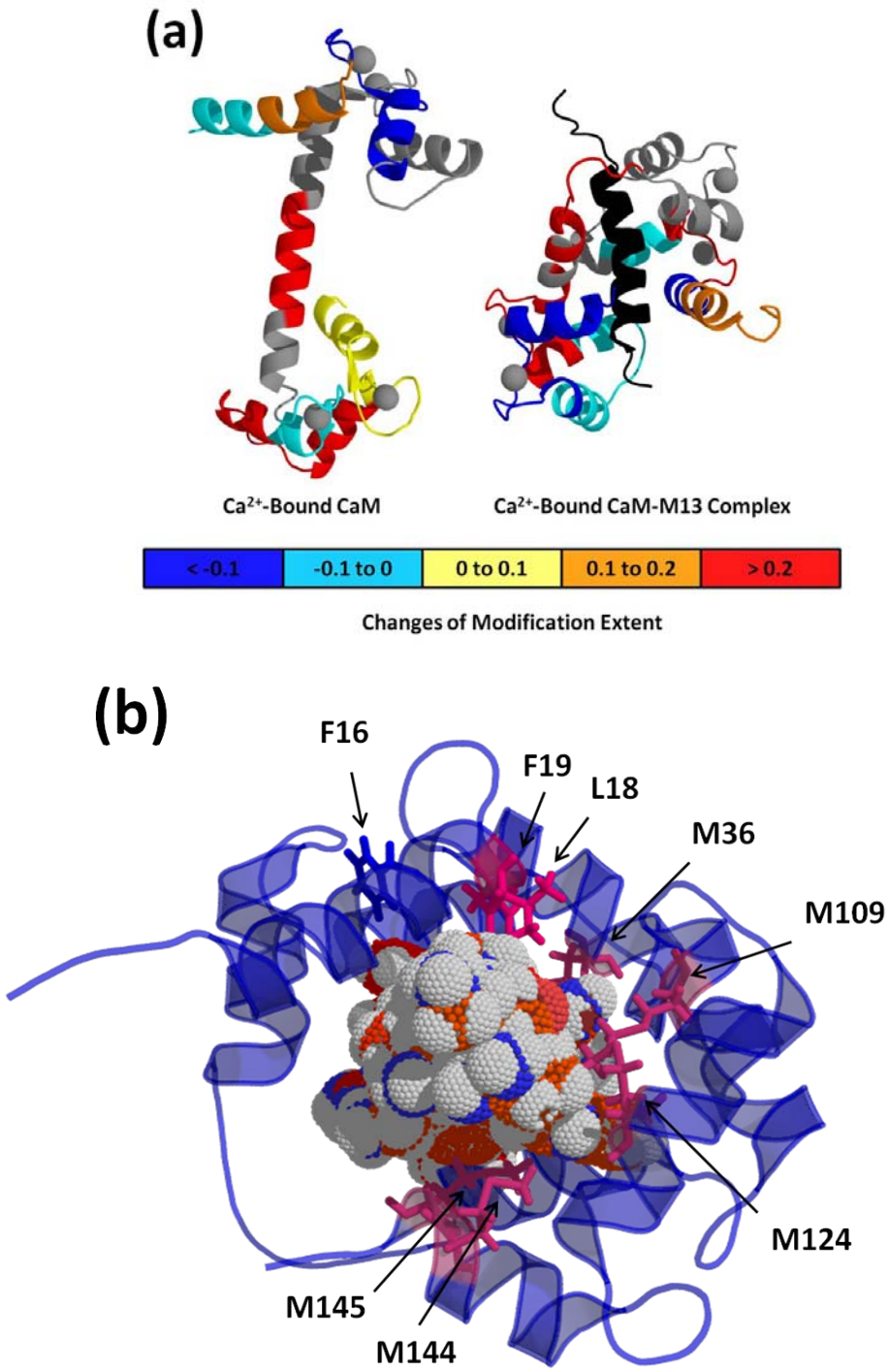


Figure 2.5. FPOP footprinting results on CaM structures. (a) Changes of oxidation are labeled on peptides with different color. M13 is black. (b) Bottom view of CaM-M13 complex. Residues in pink are several modified residues detected by LC-MS experiment. M13 is in the center of structure with ball shape.

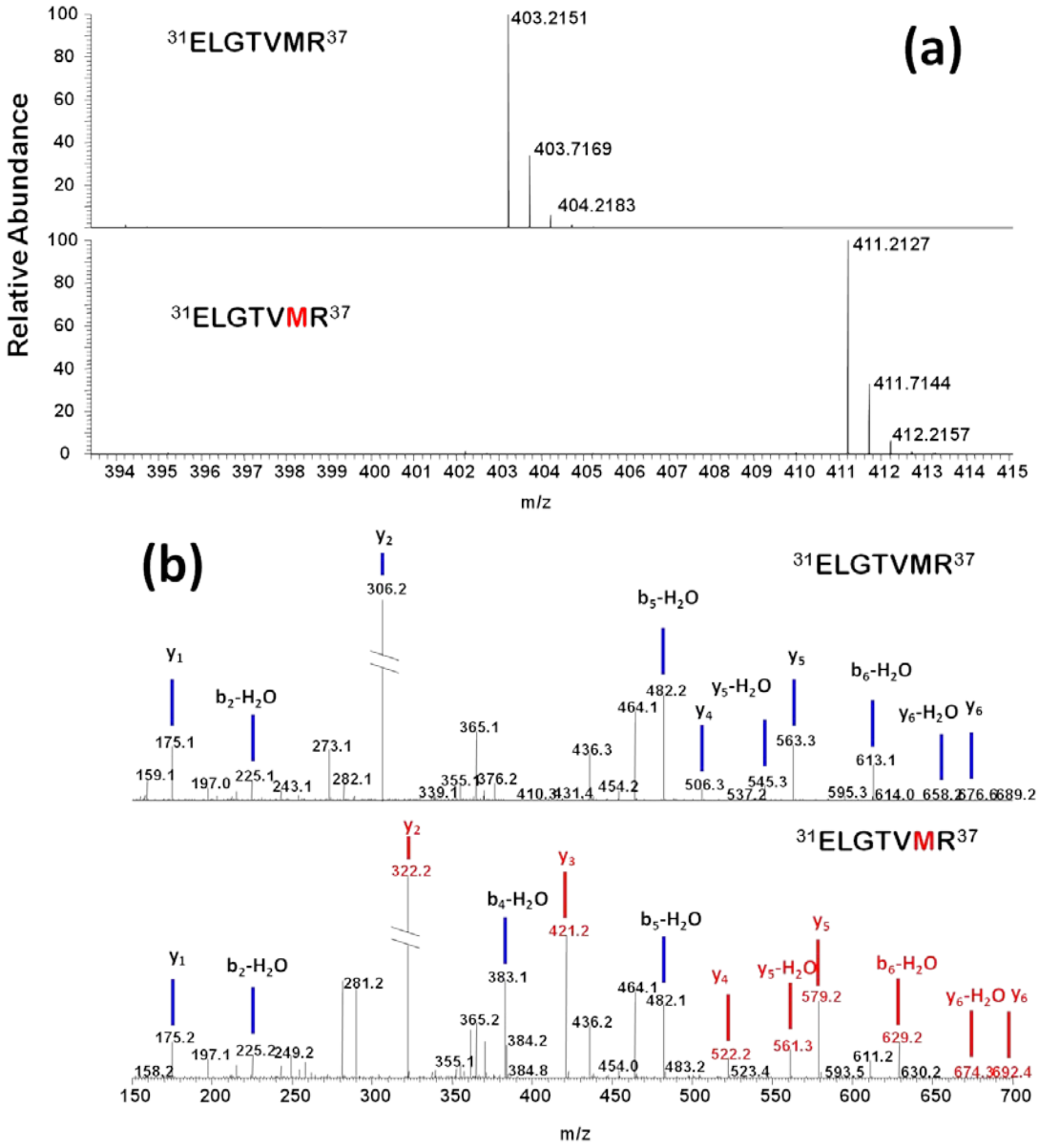


Figure 2.6. Mass spetra of modified peptide: (a). Mass spectra of peptide 31-37 (doubly charged), unmodified peptide (top), and modified peptide (bottom). (b). Product-ion spectra (MS/MS) of unmodified (top) and modified (bottom) peptide 31-37. All ions with +16 mass shift induced by FPOP are in red.

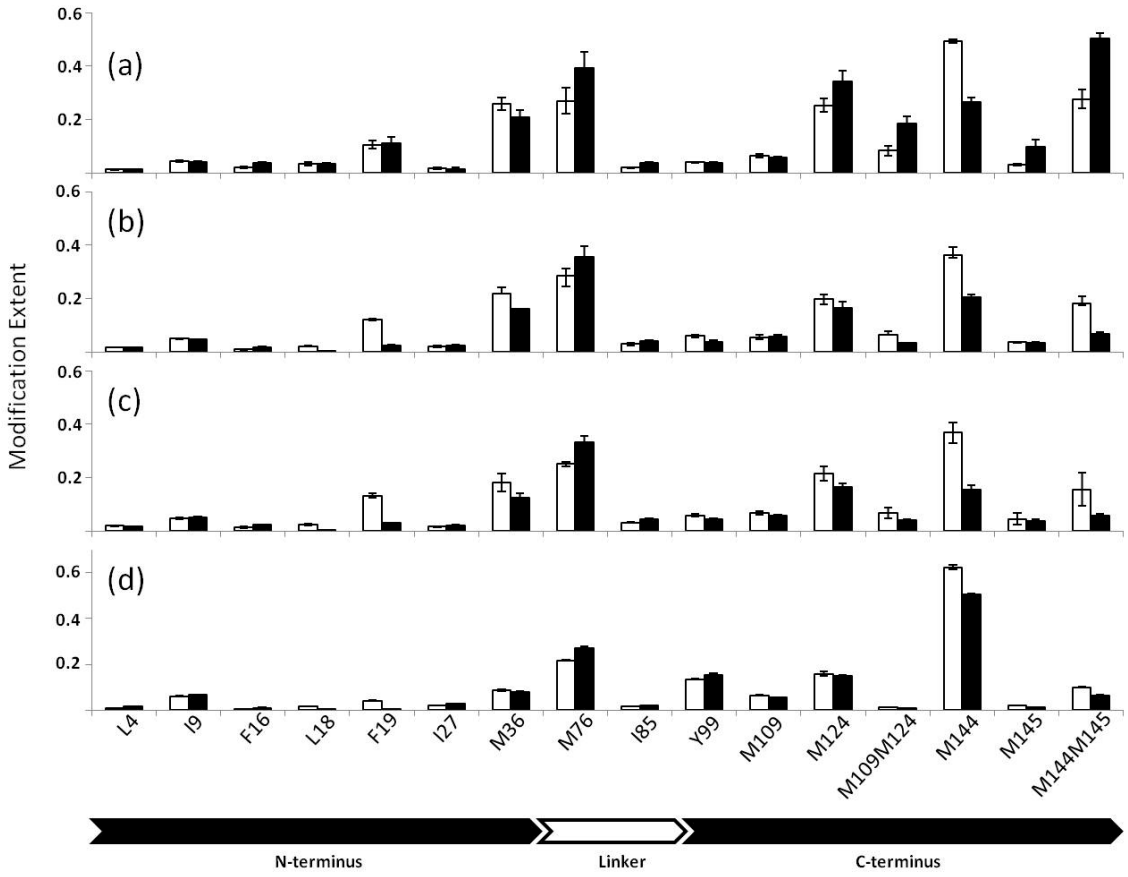


Figure 2.7. Extent of modification for Ca^{2+}-free (white) and Ca^{2+}-bound (black) states of each modified residue for various CaM complexes. (a) CaM itself (b) CaM-Mel Complex (c) CaM-Mas Complex (d) CaM-M13 complex.

Table 2.1. Spectral Contrast Angle Similarity Evaluation on Changes of Oxidation

Items of Comparison	Contrast angle (°)
CaM	34 ± 5
CaM-Mel	32 ± 2
CaM-Mas	31 ± 3
CaM-M13	18 ± 6
CaM vs. CaM-Peptide*	75 ± 1
CaM-Peptide vs. CaM-Peptide	42 ± 5

* Peptide = Mel, Mas, M13

Chapter 3

Mass Spectrometry-Based Protein Carboxyl-Group Footprinting: A Study of the Method

Abstract

Higher order structure is important in determining the function of proteins in biology, and a variety of approaches have been employed to obtain complementary structural information about proteins. Mass spectrometry-based protein footprinting is a fast-growing approach in structural studies of proteins. One labeling-based footprinting approach is the use of a water-soluble carbodiimide, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), in the presence of a glycine ethyl ester (GEE) to modify solvent-accessible carboxyl groups (i.e., on glutamate [E] and aspartate [D]). This chapter describes the development of carboxyl-group modification in protein footprinting. The modification protocol was developed by using the protein calmodulin as a model. Because carboxyl-group modification is a slow reaction relative to protein folding and unfolding, there is an issue that modifications at certain sites may induce protein unfolding and lead to modification at sites that are not solvent-accessible in the wild-type protein. Thus, we checked the protein structural integrity during carboxyl group modification by using hydrogen deuterium amide exchange (H/DX). The study demonstrated that application of carboxyl group modification in probing conformational changes in calmodulin induced by Ca^{2+} binding provides useful information that is not compromised by labeling-induced protein unfolding.

Introduction

To analyze the outcome of protein footprinting by MS, the protein must be separated from its biological matrix, which is essential for its activity and for maintaining its active conformation. Such separation could affect the MS analysis of the footprinting product if the labeling is reversible. Both MS-based H/DX and hydroxyl-radical

footprinting are limited in some applications that require extensive post-labeling separation (*1-3*). For example, MS-based H/DX requires a rapid separation to avoid the back exchange. A protein in a complicated biological system (e.g., a protein embedded in a membrane) is a challenge for rapid separation. Extensive post-labeling separation could induce a false readout in hydroxyl radical footprinting by adding additional oxidations (*4*). Even with the development of efficient separation techniques such as UPLC, studies of complicated membrane proteins by MS-based H/DX and hydroxyl radical footprinting are still difficult. Thus, the general application of protein footprinting in complicated biological systems is still a challenge.

One applicable approach for specific, irreversible modification is the 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) mediated coupling reaction between glycine ethyl ester (GEE) and the carboxyl group of a protein (*5, 6*)(Figure 3.1). The carboxyl groups of protein are activated by EDC, and then the nucleophilic modifying reagent, GEE, attacks the activated carboxyl group to yield the product. This reaction can lead to quantitative modification of solvent-accessible carboxyl groups under mild conditions[20]. The covalently labeled amide product and its hydrolyzed form are relatively stable in solution and can readily remain intact during extensive sample handling and separation.

A footprinting strategy based on modifying carboxyl groups appears appropriate because aspartate and glutamate side chains play important roles in electrostatic interactions and are essential for enzymatic activities in cells. This GEE coupling reaction has been successfully used in probing the enzymatic activity of the mammalian polyamine transport system, pancreatic phospholipase, thymidylate synthase, and cytochrome c oxidase (*7-10*).

Coupled with MS analysis, carboxyl-group modification can probe protein conformation. A previous study from our group demonstrated the application of carboxyl-group modification in determining the membrane orientation of the FMO antenna protein in the green sulfur bacterial photosynthetic system from *Chlorobaculum tepidum* (*11*). After footprinting, the modified FMO protein was separated and analyzed by MS. Information about FMO orientation in the membrane was elucidated by analysis of modification extent at the peptide level. Two advantages of carboxyl group modification—it is specific to carboxyl groups of protein and results in a relatively stable modified product—make this approach attractive for the study of protein conformation.

Here, we describe a systematic development and test of this protein footprinting approach. The reaction rate of carboxyl-group modification was examined under different conditions. The developed protocol was based on a model protein, calmodulin, but now can be used as starting point in designing new footprinting reagents. We used H/DX as a protein structure integrity check method during the carboxyl-group modification. We address the question of whether carboxyl-group modification can provide conformational information for aspartates and glutamates at the residue level in a manner that is complementary to the other protein footprinting methods.

Material and Methods

Chemicals and Proteins

Calcium free calmodulin (CaM) from bovine and troponin C (TnC) from rabbit skeletal muscle were obtained from Ocean Biologics Co. (Edmonds, WA). Water, acetonitrile, calcium chloride, phosphate-buffered saline powder, MES (2-(N-Morpholino)ethanesulfonic acid), HEPES(4-(2-hydroxyethyl)-1-piperazineethanesulfonic

acid), formic acid, glycine ethyl ester, EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, (ethylene glycol-bis(2-aminoethylether)-*N*,*N*,*N*`,*N*`-tetra-acetic acid) (EGTA), ammonium acetate, urea solution (8 M), trifluoroacetic acid, β-lactoglobulin A from bovine milk, trypsin from porcine pancreas were all obtained from Sigma-Aldrich (St. Louis, MO).

Protein Sample preparations

A protein stock solution was prepared as 10 μM protein in 10 mM phosphate buffered saline (PBS buffer, 138 mM NaCl, 2.7 mM KCl, pH = 7.4). The calcium-free CaM samples were prepared by incubating with EGTA as per a previously reported protocol(*12*).

Carboxyl Group Modification

Protein samples (10 μL) were mixed with GEE (2 M in water) 1 μL and EDC (50 mM in water) 1 μL to initiate the reaction at room temperature (21.6 °C). The reaction was quenched by adding 20 μL ammonium acetate (1 M). After modification, all samples were kept at 4 °C. In the pH-dependence experiments, protein samples in a variety of buffer systems, MES/NaOH (pH = 5.5), PBS (pH = 6, 6.5, 7, 7.5), HEPES (pH = 8, 8.5) were prepared and modified in an identical way.

MS of modified protein

ESI mass spectra were acquired in the positive-ion mode on a Waters (MicroMass) Q-TOF Ultima (Manchester, U.K.) and a Bruker MaXis Q-TOF (Bremen, Germany). The instrument setup for Waters Q-TOF for protein analysis was similar to the previous protocol (*12*).

For the Bruker MaXis Q-TOF, the capillary voltage was 3.8-4 kV, nebulizer gas was 0.6 bar, dry gas was 6.0 L/min, and the source temperature (dry temperature) was 180-200 °C. The protein sample (10 pmole) was loaded on a C8 rapid-resolution cartridge (ZORBAX Eclipse XDB-C8, 2.1 x 15 mm, 3.5 μm, Agilent Technologies, Santa Clara, CA). The protein was eluted at 200 uL/min with a LC gradient (Agilent 1200 HPLC, 5 % - 15 % B in 0.3 min, 50 % B in 5.5 min, 100 % B in 6-7.5 min then back to 5% B in 9.5 min, solvent A: water, 0.3 % formic acid; solvent B: 80% acetonitrile, 20% water, 0.3% formic acid).

CD experiment

CaM samples, unmodified native and modified (reaction time 90 s), were desalted to remove extra ammonium acetate. CD spectra were measured at room temperature from 195-300 nm wavelength at 1 nm intervals by a JASCO J815 CD spectrometer (JASCO Analytical Instruments, Tokyo, Japan).

H/DX experiment

Protein samples (2 μL, 10 μM) were mixed with PBS D_2O buffer (18 μL). The hydrogen deuterium exchange was conducted on ice. Exchange was quenched by adding 30 μL 3 M urea solutions with 0.1% trifluoroacetic acid at 0 °C. H/D exchanged samples were directly desalted by LC at 0 °C and analyzed by MS to minimize any back exchange.

Typsin digestion, LC-MS/MS and data analysis

The digestion of CaM samples was conducted according to a previously reported protocol with minor changes (*12*). The labeled protein sample was mixed with trypsin (trypsin: protein = 1:10) without heating or denaturation. LC-MS/MS of the protein

digest and data analysis were similar to the previously reported approach (*12*). Changes were made to search for carboxyl-group modifications instead of FPOP modifications.

Solvent accessible surface area (SASA) calculation

The side-chain solvent accessibility of calcium-free and bound CaM were calculated by using the structures determined by NMR spectroscopy or X-ray crystallography. PDB files of calcium-free CaM (PDB ID: 1CFC) and calcium-bound CaM (PDB ID: 1CLL) were submitted to GETAREA 1.0 (http://curie.utmb.edu/getarea.html) for calculation of individual side chain SASA (*13*).

Results and Discussion

Workflow of Post Labeling MS Analysis

We analyzed the outcomes of protein footprinting MS at both the protein and the amino-acid residue levels. At the protein level, we used desalted protein samples and directly analyzed them by ESI MS. Peaks corresponding to unmodified and modified proteins allowed a comparison of the modification extents for different protein states. In the method development, we optimized the modification conditions based on MS analysis at the protein level even though the results cannot provide information about local conformational changes of the protein. To probe changes in local conformations, we digested the protein and analyzed the resulting peptides by LC-MS/MS, using a typical label-free, quantitative-analysis approach from proteomics. We used both calcium-free and calcium-bound CaM samples to obtain information at the amino-acid residue level.

Modification Conditions for Protein Footprinting

Labeling protein in biological relevant settings should be the first step in protein footprinting. The modification experiment usually generates a mixture of unmodified and modified species. To calculate the modification extent, good MS signals for both the unmodified and modified species are essential. Tuning of the modification condition is required for generating the desired amount of modified species for MS analysis. Ideally, one would aim for a "single hit" in the modification as this would guarantee that the footprinting does not induce unfolding and lead to misleading subsequent modifications. Single-hit modifications are not practical because analysis is difficult.

The coupling reaction between the carboxyl group and GEE is initiated by the attachment of the protein to EDC. The next step involves attach of the carboxyl group to form an O-acylisourea. The nucleophilic GEE attacks the activated carboxyl group to yield the desired amide and urea. The rate of this coupling reaction depends on the reagent ratio, solvent conditions, and the solvent accessibility of the acid residues of the target protein (Figure 3.2).

Carboxyl groups of CaM were modified with excess amounts of GEE and EDC (CaM : GEE:EDC = 1:20000:500) in PBS buffer at room temperature with different reaction times. For each charge state, the peaks representing the modified protein show mass shifts of 85 Da (amide form) and 57 Da (hydrolyzed form of amide product) (Figure 3.3). With increasing reaction time, multiple peaks for the modified protein (+2GEE, +3GEE, +4GEE, +5GEE) became intense in a sequential manner. The results show that the GEE coupling reaction efficiently modified CaM within a timescale of minutes.

The pH of solvent has a major effect on the reaction rate. We examined this effect by modifying CaM at different pHs (from 5.5 to 8.5). The results show that the rate of the coupling reaction can be accelerated by lowering the pH (Figure 3.4). In addition to affecting the coupling reaction, lowering the pH promoted the hydrolysis of the amide product formed in the modification. If the GEE coupling reaction is to be used in probing protein conformational changes that require pH variation, special attention is required to adjust the reaction time. Similarly, lowering the temperature can dramatically reduce the rate of the coupling reaction (Figure 3.5).

The rate of the GEE coupling reaction is also affected by protein structure. We chose three proteins, CaM, BlG, and TnC, as models to be modified by carboxyl-group modification. Although these three proteins have similar sizes, CaM and TnC have more flexible structures (*14-19*), whereas BlG has a more compact and rigid structure owing to two pairs of disulfide linkages(*20*). At the same reaction times, higher levels of multiply modified species were observed for CaM and TnC samples (up to protein + 4 GEE) than for the BlG sample (up to protein + 2 GEE). The modification conditions that we established for the CaM protein can serve as a good starting point for optimization in other applications of carboxyl-group modification in a biologically relevant setting.

Ensuring the Protein's Structural Integrity

Ensuring a protein's structural integrity is a major concern in labeling-based protein footprinting (*6*). Covalent labeling can cause conformational changes by breaking protein noncovalent interactions ("over labeling"). The readouts of over-labeled samples will result in misleading conclusions about protein conformation. Various approaches can be employed to check the structural integrity following a modification:

examples are circular dichroism (CD)(*21*), fluorescence spectroscopy(*22*), and activity assays. Another approach measures the reaction kinetics for individual modification sites (*23*) or the modification pattern at the protein level(*24*). The premise is that the modified protein should show the same chemical reactivity as the unmodified version.

CD spectroscopy is the most common check because it is sensitive to variations in protein secondary structure (*25*). In the development of carboxyl-group footprinting, we analyzed both native and modified CaM (reaction time 90 s) by CD (Figure 3.6). Typical α helix CD curves were observed for both samples. No change occurred in the CD spectrum of modified CaM. The result shows that no significant secondary structural change happened in the time scale of modification. In labeling-based protein footprinting, the protein sample is a mixture; that is, it contains both unmodified and modified species. The majority of the proteins in the mixture is still in an unmodified, native form. Any conformational changes from the small portion of the sample undergoing the modification could be overlooked in the CD spectrum.

Alternatively, MS analysis can identify unmodified and modified species in the mixture of modified and unmodified proteins. Is it possible to employ a second MS-based footprinting method to check for any conformational changes that occur during carboxyl-group modification? H/DX is an ideal protein footprinting method for this purpose because the replacement of H by D induces minimal structural disturbance (*26*). The deuterium uptake of each modified protein state, as well as of the unmodified protein state, can be monitored separately by MS. More important, the H/DX outcome is directly measured as a mass shift of the protein. Any ionization bias induced by the carboxyl-group modification can be ignored in the H/DX experiment.

We analyzed by H/DX three protein samples, CaM, BlG, and TnC, after their carboxylate footprinting (Figure 3.7). The unmodified-protein and each modified-protein state share a similar deuterium uptake curve. The results provide additional and more convincing evidence that no major conformational change occurs during the timescale of carboxyl-group modification. The relatively simple design and high sensitivity to each modified protein state make this H/DX check attractive in the field of protein footprinting. With the fast peptic digestion protocol developed for the H/DX experiment, a more detailed check on the local conformational change than that provided by CD is available, and it should be a powerful control that can be used for the development of other protein footprinting strategies.

GEE labeling probes conformational changes

CaM is a small calcium-binding protein that has two domains; in each domain are two typical EF-hand motifs (helix-loop-helix unit) that bind calcium ions by electrostatic interactions (*27*). Calcium-binding EF hand motifs are rich in negatively charged glutamates, and aspartates. In the typical EF hand motif, the co-ordination of a calcium ion contains seven ligands arranged in a pentagonal bipyramidal fashion (*28*). Among these seven co-ordination sites, five are from the nine-residue loop and two are side chains of amino acids in the helix. In the nine-residue loop, three or four coordination sites are the carboxyl groups of aspartates or glutamates and one is a backbone carbonyl group. CaM undergoes conformational changes upon calcium binding, and these changes activate its binding domain for target proteins in downstream signal transduction.

We used carboxyl-group modification to probe the conformational changes induced by Ca^{2+} binding and found a total of 30 modified aspartates or glutamates in the

CaM sequence. This coverage enables a thorough comparison of the modification extents of calcium-free and calcium-bound CaM. We compare the calculated SASA of calcium-free and bound-CaM in Figure 3.8 to see if they agree with the changes in the extent of modification. The calculated SASA of aspartates or glutamates from those EF hands that coordinate Ca^{2+} show significant decreases upon calcium binding. Similar trends in the modification extents occur for aspartate and glutamate residues from the other EF hands (Figure 3.9). Although the absolute SASA values and the modification extents are different, the decreases induced by calcium binding are clear. For those residues that undergo high modification extents (e.g., EF hand 1), the difference in modification extents between the calcium-free and calcium-bound CaM is statistically significant (99.5% confidence level in T-test). An exception is E104 of EF hand 3 where there is no difference in the modification extent between the calcium-free and calcium-bound CaM. The lack of difference may be masked by the experimental error owing to the low modification extent and the poor S/N. The experimental errors from three biological replicas are from 0.5% to 5% (a total of 30 modified residues have an average error ~2%) in modification extent. This exception emphasizes again that adjusting the modification conditions is important for a successful footprinting experiment. For example, to differentiate the change of modification extent at the sites of low modification extent, adjusting the modification condition is required to increase the modification extent.

Residues E11, D78, D80, E114, E118, E119, and E123 show significant differences between the modification extents and calculated SASA changes. These differences may originate from the dynamics of CaM solution structures. Calcium-free CaM has a flexible structure in solution, and its central linker region can be bent to

accommodate different relative positions of the N- and C-terminal domains (*29*). The calcium-bound CaM, however, has a relatively rigid solution structure (*14*). The central linker forms a seven-turn α helix that connects the two domains. All residues with inconsistent outcomes between modification extents and calculated SASA are from regions of the protein that are close to the central linker region (Figure 3.10). The highly flexible properties of CaM in the central linker region could result in inconsistent readouts from different measurements. The extent for carboxyl-group modification is an average over protein conformational dynamics during the reaction time whereas the calculated SASA results from the different 3D model of protein structures (calcium-free CaM model from NMR spectroscopy (*29*) and calcium-bound CaM model from X-ray crystallography(*14*)). The protein model from X-ray crystallography is a "snap shot" of the protein conformation in crystalline solid state whereas the model from NMR is an integrated picture from solution.

The conformational effects of calcium binding on the EF-hand regions can be detected by the carboxyl-group modification at the residue level for aspartates and glutamates. Those effects were not seen in an FPOP approach to the same system (*12*). H/DX showed similar conformational variations induced by calcium binding at the EF hands (*30, 31*). The results from H/DX, however, are not at the amino-acid residue level. Compared with other footprinting methods, carboxyl-group modification is sensitive to the conformational changes of the protein's carboxyl side chains and capable of providing results at the residue level. More importantly, this approach provides protein conformational information that is complementary to that from other approaches.

Conclusion

We studied carboxyl-group modification by using three model protein systems. A sufficient amount of modified species can be reached in a biologically relevant setting by adjusting the modification conditions. Under these conditions, no major conformational change occurred during the time scale of the carboxyl-group footprinting as determined by H/DX. This carboxyl-group footprinting approach is sensitive to the conformational changes in carboxyl side chains of the protein and can report, for example, the conformational variations induced by calcium binding at the EF hand region of CaM. Carboxyl-group modifications provide structural information that is complementary not only to traditional structural biology methods (e.g., NMR and X-ray) but also to other protein footprinting approaches (e.g., FPOP and H/DX).

References

1. Pan, Y., Stocks, B. B., Brown, L., and Konermann, L. (2009) Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry, *Anal Chem 81*, 28-35.
2. Zhu, Y., Guo, T., Park, J. E., Li, X., Meng, W., Datta, A., Bern, M., Lim, S. K., and Sze, S. K. (2009) Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting, *Mol Cell Proteomics 8*, 1999-2010.
3. Hebling, C. M., Morgan, C. R., Stafford, D. W., Jorgenson, J. W., Rand, K. D., and Engen, J. R. (2010) Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, *Anal Chem 82*, 5415-5419.
4. Perdivara, I., Deterding, L. J., Przybylski, M., and Tomer, K. B. (2010) Mass Spectrometric Identification of Oxidative Modifications of Tryptophan Residues in Proteins: Chemical Artifact or Post-Translational Modification?, *J Am Soc Mass Spectrom 21*, 1114-1117.
5. Hoare, D. G., and Koshland, D. E., Jr. (1967) A method for the quantitative modification and estimation of carboxylic acid groups in proteins, *J Biol Chem 242*, 2447-2453.
6. Mendoza, V. L., and Vachet, R. W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry, *Mass Spectrom Rev 28*, 785-815.

7. Ferreira, J. P., Sasisekharan, R., Louie, O., and Langer, R. (1994) Carbodiimide modification enhances activity of pig pancreatic phospholipase A2, *Eur J Biochem 223*, 611-616.

8. Chen, D. H., Daron, H. H., and Aull, J. L. (1992) 1-Phenyl-3-trimethylaminopropyl carbodiimide: a new inhibitor of thymidylate synthase, *J Enzyme Inhib 5*, 259-268.

9. Torossian, K., Audette, M., and Poulin, R. (1996) Substrate protection against inactivation of the mammalian polyamine-transport system by 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide, *Biochem J 319 (Pt 1)*, 21-26.

10. Taha, T. S., and Ferguson-Miller, S. (1992) Interaction of cytochrome c with cytochrome c oxidase studied by monoclonal antibodies and a protein modifying reagent, *Biochemistry 31*, 9090-9097.

11. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting, *Proc Natl Acad Sci U S A 106*, 6134-6139.

12. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system, *Anal Chem 83*, 311-318.

13. Fraczkiewicz, R., and Braun, W. (1998) Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules, *J. Comp. Chem 19*, 319-333.

14. Kretsinger, R. H., Rudnick, S. E., and Weissman, L. J. (1986) Crystal structure of calmodulin, *J Inorg Biochem 28*, 289-302.

15. Rupp, B., Marshak, D. R., and Parkin, S. (1996) Crystallization and preliminary X-ray analysis of two new crystal forms of calmodulin, *Acta Crystallogr D Biol Crystallogr 52*, 411-413.

16. Han, B. G., Han, M., Sui, H., Yaswen, P., Walian, P. J., and Jap, B. K. (2002) Crystal structure of human calmodulin-like protein: insights into its functional role, *FEBS Lett 521*, 24-30.

17. Kay, L. E., Forman-Kay, J. D., McCubbin, W. D., and Kay, C. M. (1991) Solution structure of a polypeptide dimer comprising the fourth Ca(2+)-binding site of troponin C by nuclear magnetic resonance spectroscopy, *Biochemistry 30*, 4323-4333.

18. Houdusse, A., Love, M. L., Dominguez, R., Grabarek, Z., and Cohen, C. (1997) Structures of four Ca2+-bound troponin C at 2.0 A resolution: further insights into the Ca2+-switch in the calmodulin superfamily, *Structure 5*, 1695-1711.

19. Soman, J., Tao, T., and Phillips, G. N., Jr. (1999) Conformational variation of calcium-bound troponin C, *Proteins 37*, 510-511.

20. Kontopidis, G., Holt, C., and Sawyer, L. (2004) Invited review: beta-lactoglobulin: binding properties, structure, and function, *J Dairy Sci 87*, 785-796.

21. Tong, X., Wren, J. C., and Konermann, L. (2007) Effects of protein concentration on the extent of gamma-ray-mediated oxidative labeling studied by electrospray mass spectrometry, *Anal Chem 79*, 6376-6382.

22. Gupta, S., Cheng, H., Mollah, A. K., Jamison, E., Morris, S., Chance, M. R., Khrapunov, S., and Brenowitz, M. (2007) DNA and protein footprinting analysis of the modulation of DNA binding by the N-terminal domain of the Saccharomyces cerevisiae TATA binding protein, *Biochemistry 46*, 9886-9898.

23. Mendoza, V. L., and Vachet, R. W. (2008) Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection, *Anal Chem 80*, 2895-2904.

24. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Anal Chem 81*, 6563-6571.

25. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc 1*, 2876-2890.

26. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass Spectrom Rev 25*, 158-170.

27. Wilson, M. A., and Brunger, A. T. (2000) The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity, *J Mol Biol 301*, 1237-1256.

28. Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007) Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs, *Biochem J 405*, 199-221.

29. Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B., and Bax, A. (1995) Solution structure of calcium-free calmodulin, *Nat Struct Biol 2*, 768-776.

30. Zhu, M. M., Rempel, D. L., Zhao, J., Giblin, D. E., and Gross, M. L. (2003) Probing Ca2+-induced conformational changes in porcine calmodulin by H/D exchange and ESI-MS: effect of cations and ionic strength, *Biochemistry 42*, 15388-15397.

31. Sperry, J. B., Huang, R. Y.-C., Zhu, M. M., Rempel, D. L., and Gross, M. L. (2010) Hydrophobic Peptides Affect Binding of Calmodulin and Ca^{2+} as Explored by H/D Amide Exchange and Mass Spectrometry, *International Journal of Mass Spectrometry*, doi:10.1016/j.ijms.2010.1008.1013.

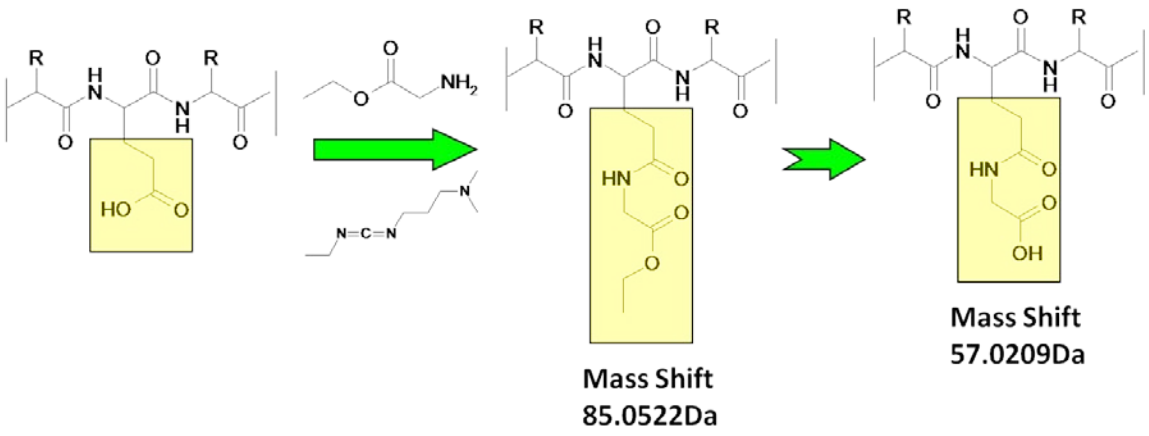


Figure 3.1. Carboxyl group modification. Carbodiimide (EDC) mediates the incorporation of glycine ethyl ester (GEE) to the carboxyl group of aspartate or glutamate side chains.

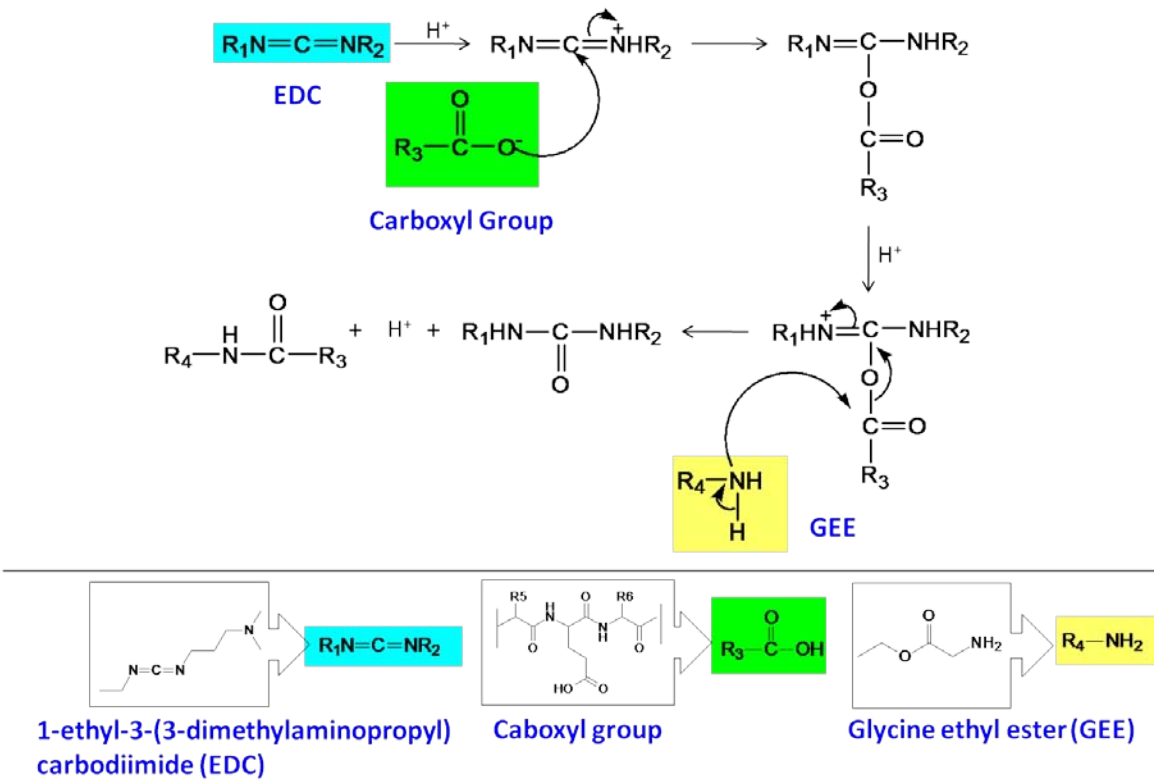


Figure 3.2. Reaction mechanism of EDC mediated GEE coupling.

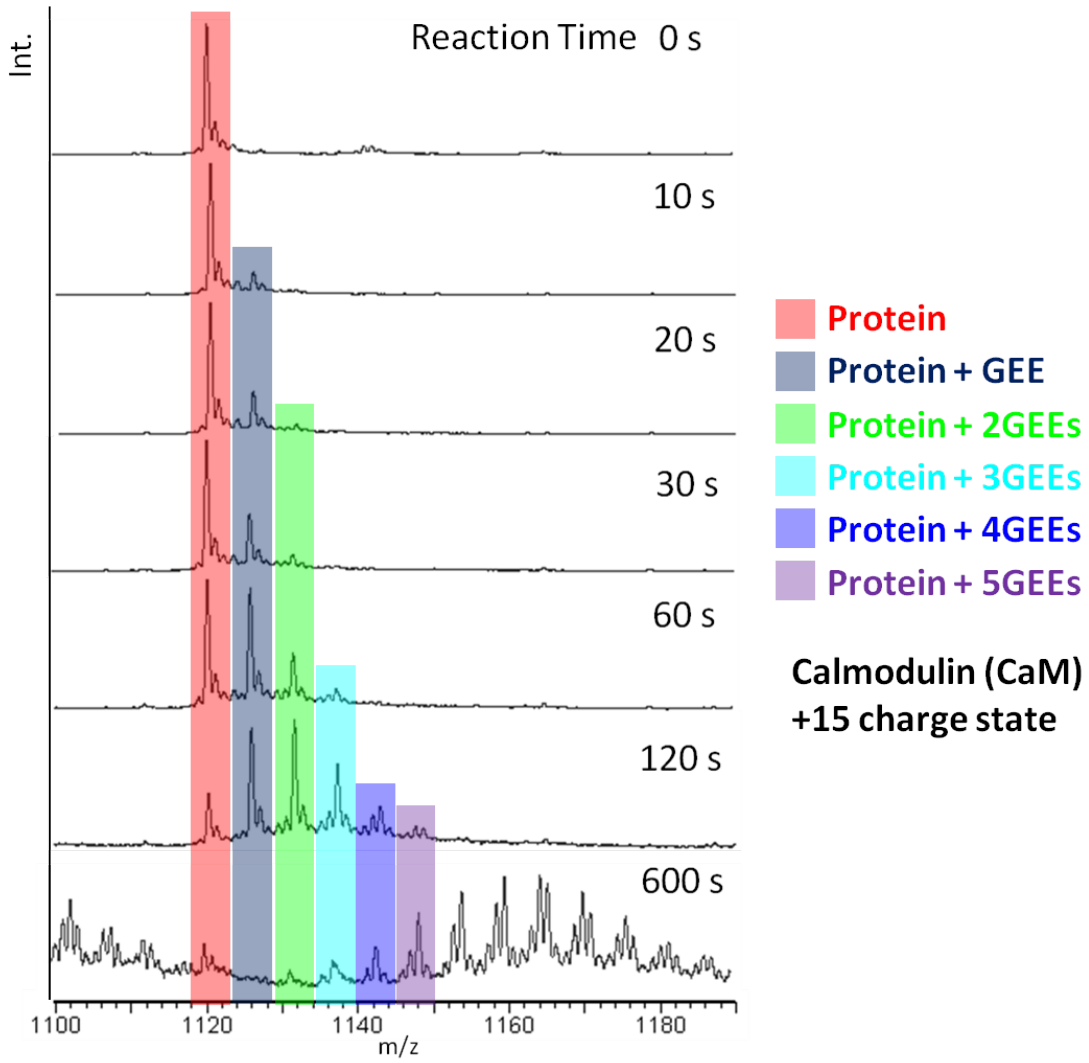


Figure 3.3. Mass spectra of modified CaM samples at different reaction times. CaM was labeled (CaM:GEE:EDC = 1:20000:500) in PBS buffer. Resulting CaM samples were desalted and analyzed by a QTOF mass spectrometer to compare modified CaM samples (+15 charge state) at different reaction times. The unmodified protein state and each modified protein state are highlighted by different colors.

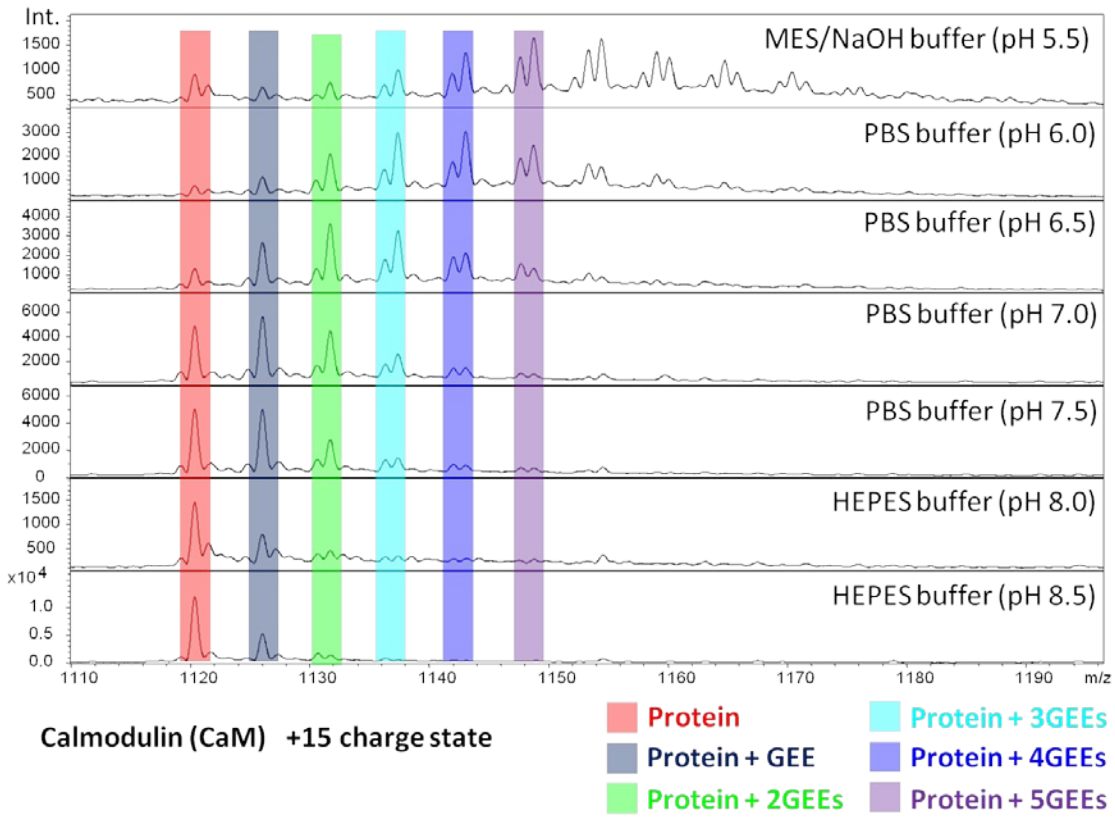


Figure 3.4. Mass spectra of modified CaM in different pH buffer systems. Modified CaM samples (+15 charge state) were compared (the unmodified and each modified protein state are highlighted by different colors).

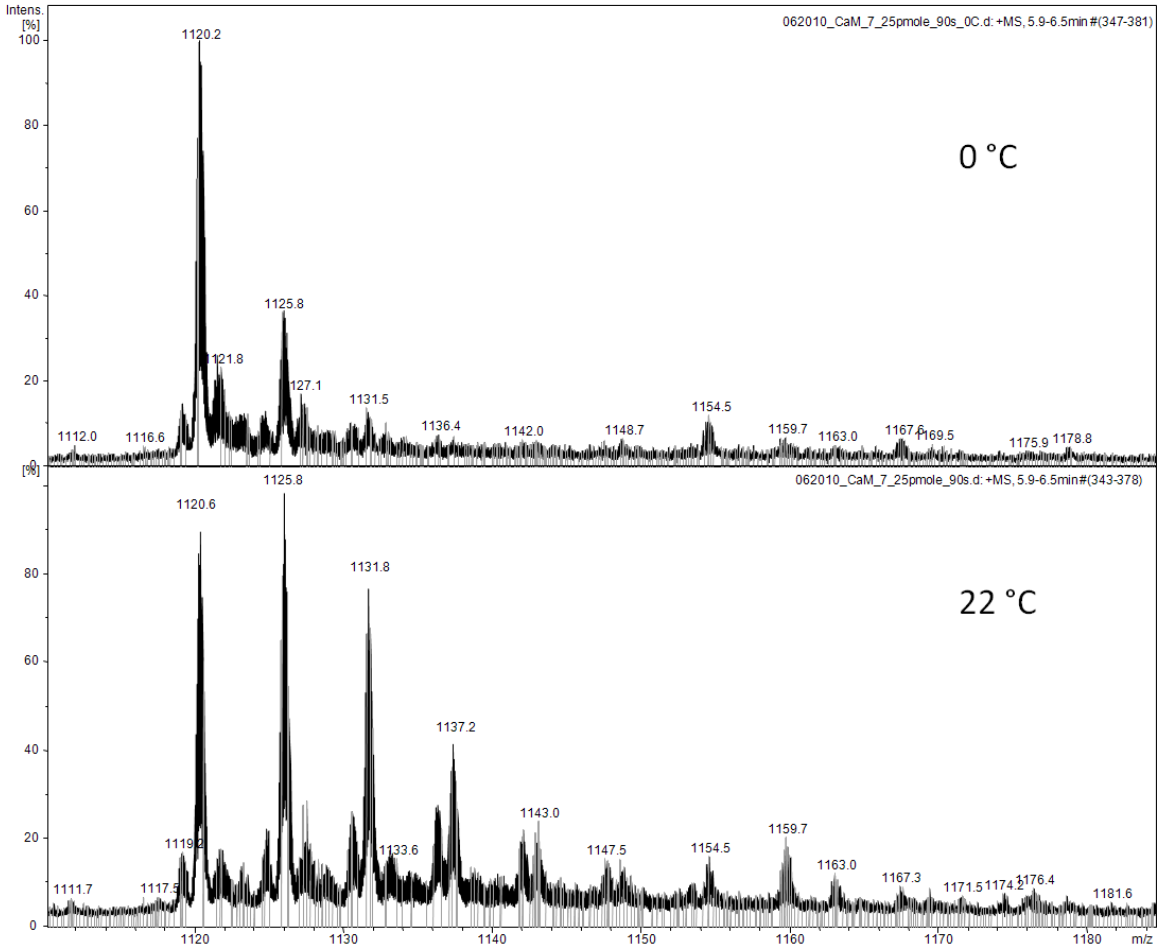


Figure 3.5. Mass spectra of modified CaM samples at different temperatures. CaM was labeled (CaM:GEE:EDC = 1:20000:500) in PBS buffer. Resulting CaM samples were desalted and analyzed by a QTOF mass spectrometer to compare modified CaM samples (+15 charge state).

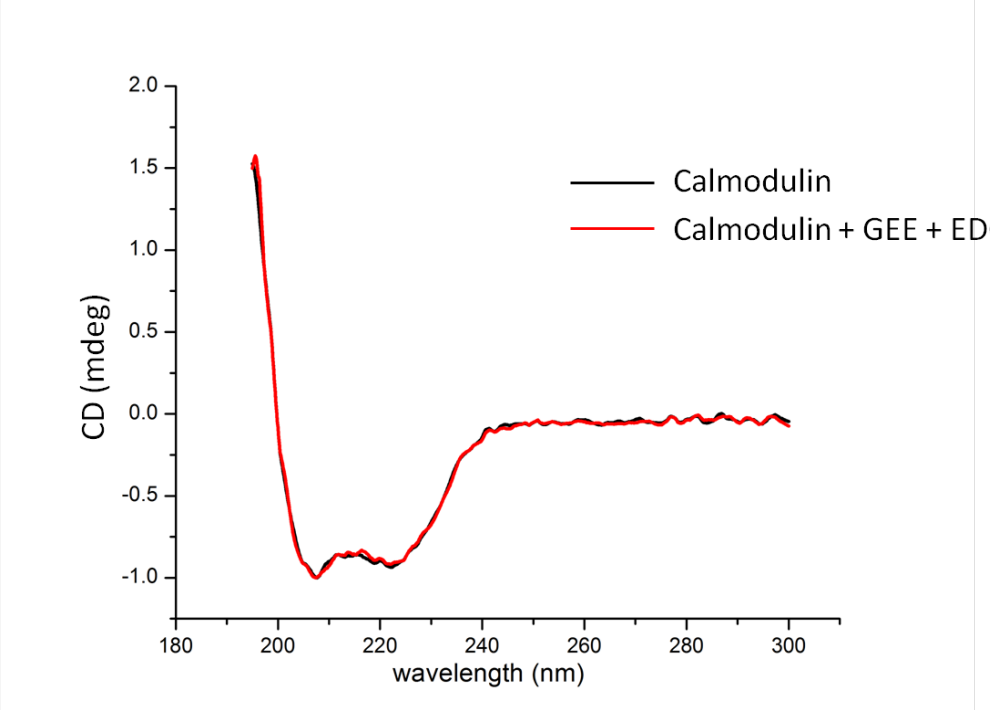


Figure 3.6. CD spectra of unmodified and modified CaM samples. Native CaM and modified CaM (reaction time 90 s) samples were desalted and then analyzed by CD.

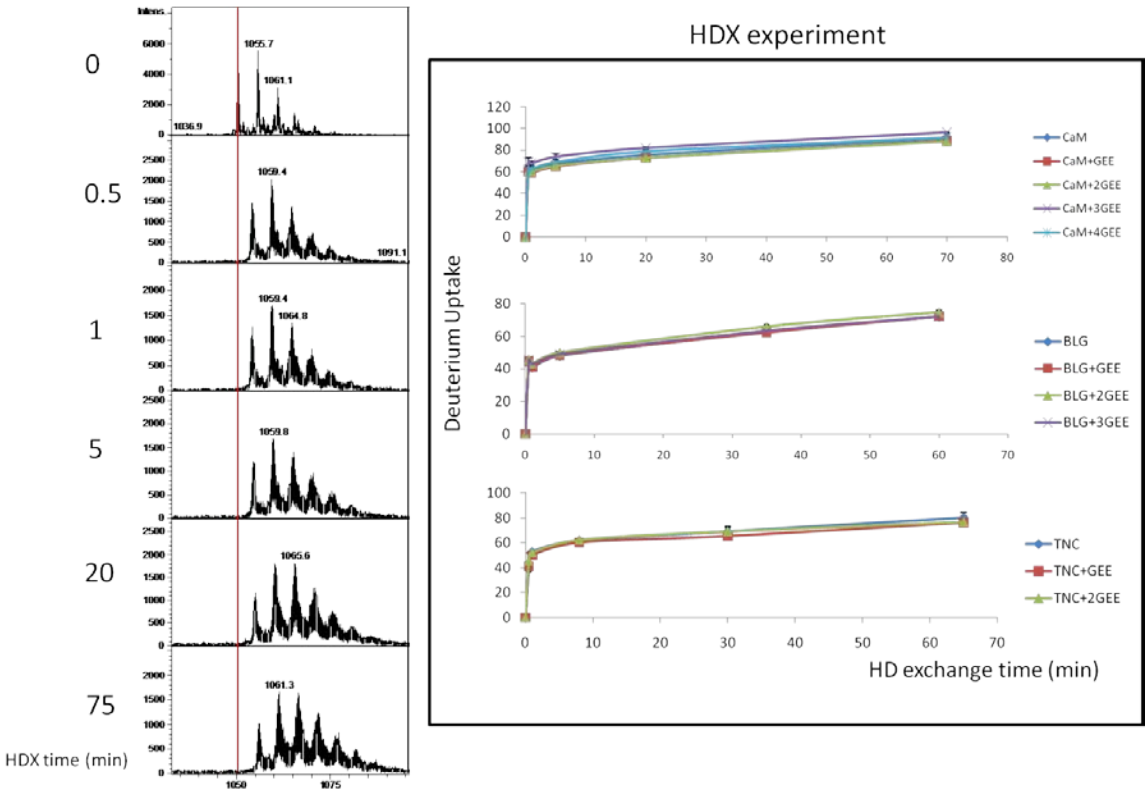


Figure 3.7. HDX results of modified CaM, BlG and TnC, as analyzed by an H/DX experiment. The deuterium uptakes of unmodified protein state as well as each modified protein state were monitored separately. Extended views of the modified-CaM spectra with different exchange times are listed on the left. The deuterium uptake plots of three modified protein are on the right.

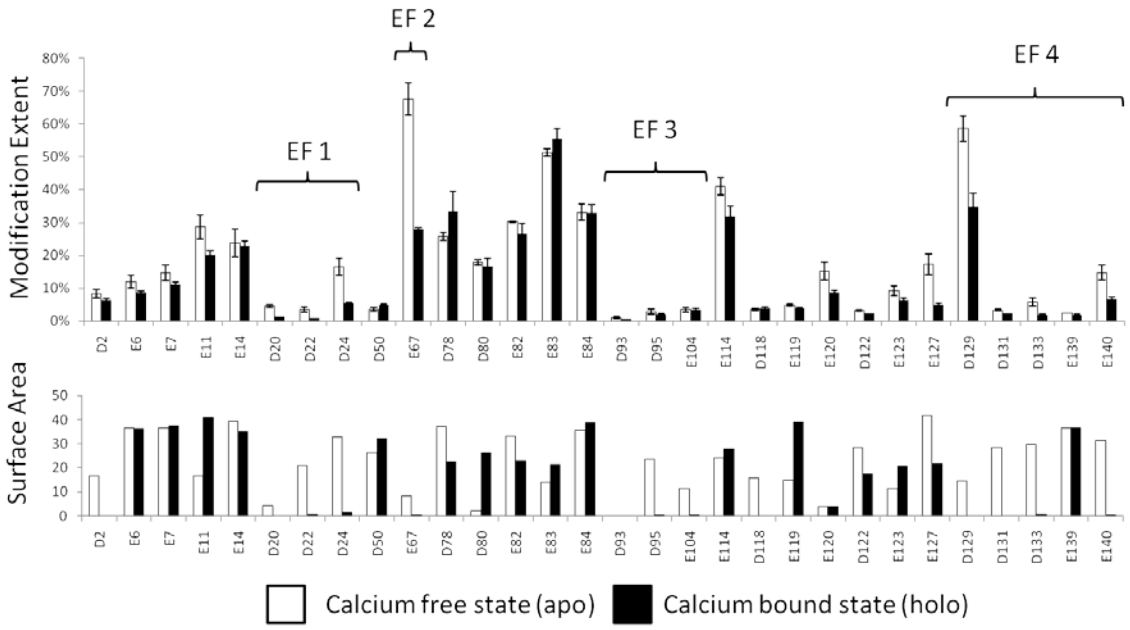


Figure 3.8. The modification extent and calculated SASA of apartate and glutamate residues detected by LC-MS/MS experiment. Two CaM samples, calcium-free and calcium-bound CaM, were analyzed. The modification extent of each detected aspartate and glutamate residue was plotted on the top panel, and calculated SASA values from calcium-free CaM (PDB ID: 1CFC) and calcium-bound CaM (PDB ID: 1CLL) were plotted on the bottom for reference.

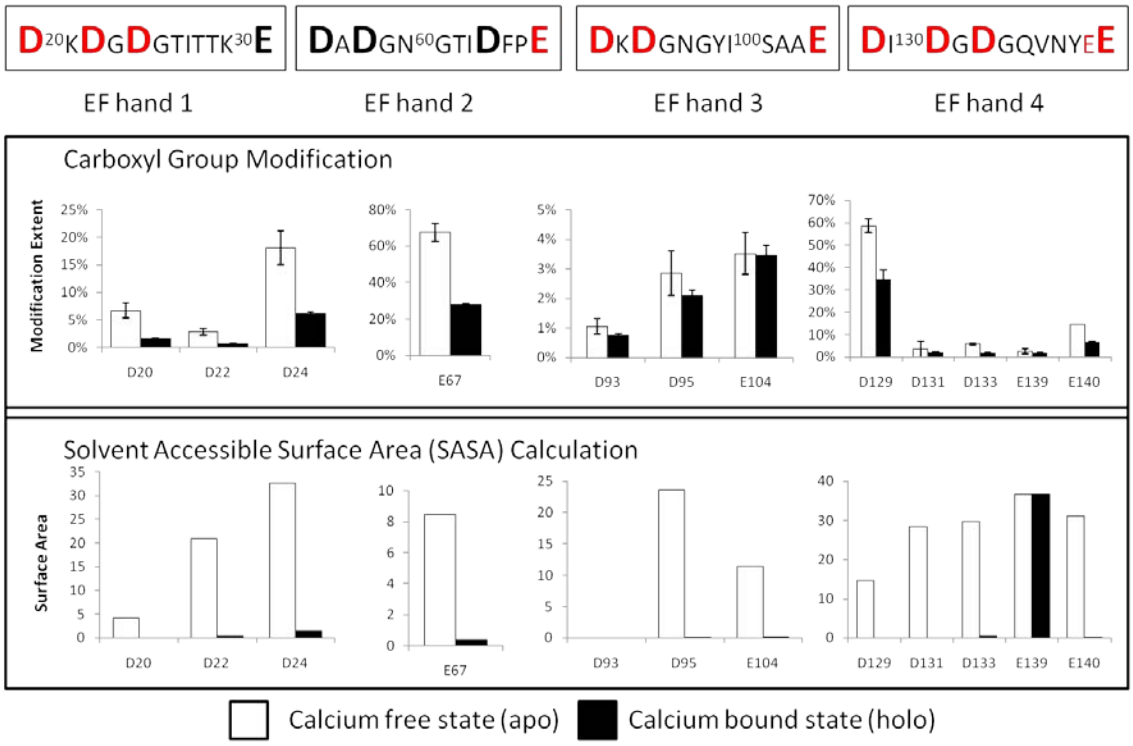


Figure 3.9. The modification extent and calculated SASA of apartate and glutamate residues from CaM EF hand regions. The modification extents of each detected aspartate and glutamate residues from EF hand regions were plotted. Calculated SASA values plotted for reference.

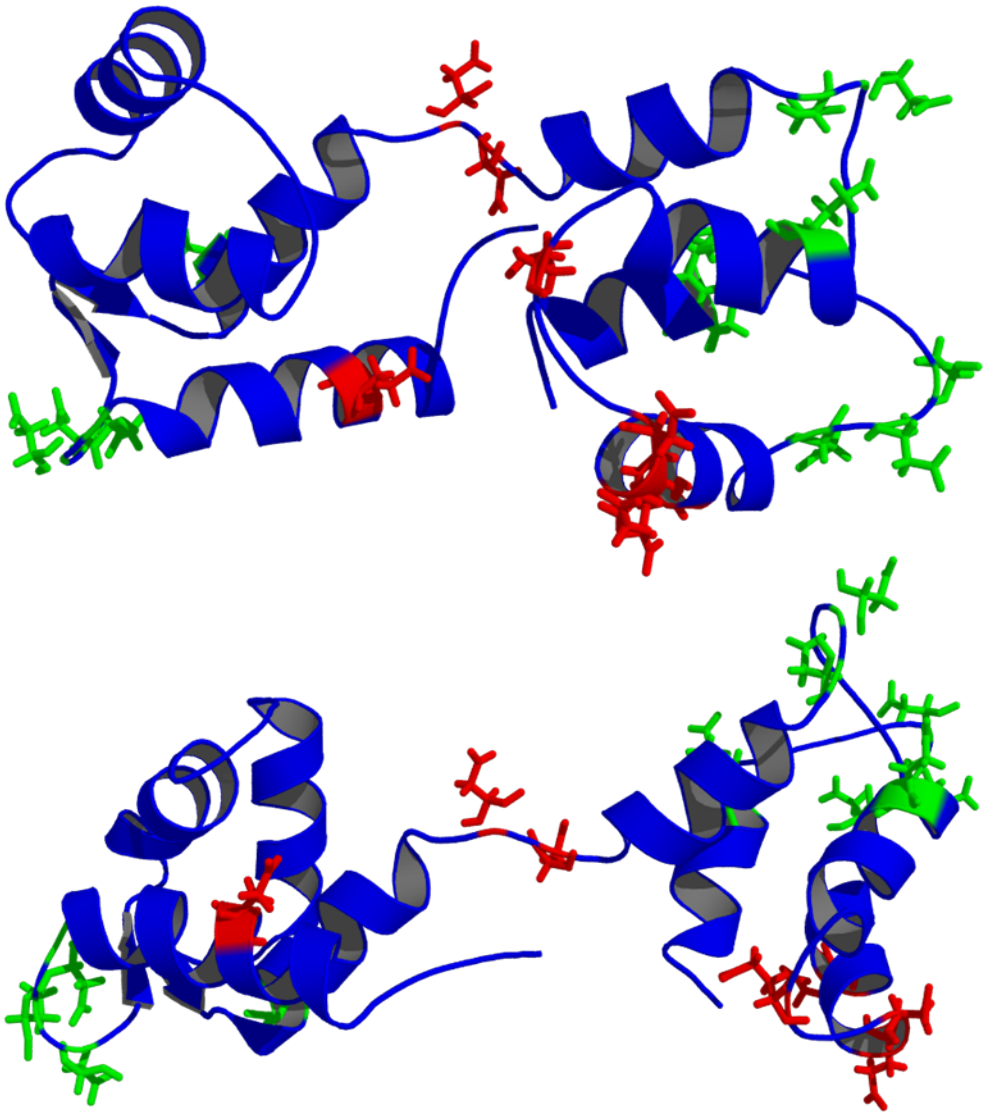


Figure 3.10. The 3D models of calcium free CaM. Two conformations of calcium free CaM from NMR spectroscopy were displayed. The apartate and glutamate residues from EF hand regions were labeled in stick model with green color. The aspartate and glutamate residues with inconsistent outcomes between modification extents and calculated SASA values were labeled in stick model with red color.

Chapter 4

Carboxyl Group Footprinting Maps the Dimerization Interface and Phosphorylation-induced Conformational Changes of a Membrane-associated Tyrosine Kinase

This chapter is based on recent publication:

H. Zhang, W. Shen, D. Rempel, J. Monsey, I. Vidavsky, M. L. Gross, R. Bose, Mol Cell Proteomics. 2011.

Author contributions: H.Z., W.S., D. R.,J. M., I.V., M.L.G., R. B. designed research; H.Z. W.S., J.V. performed research; H.Z., D.R., I.V. contributed analytic tool; H.Z., W.S., D.R., J.M. analyzed data; H.Z., W.S.,D.R.,J.M., I.V., M.L.G., R.B. discussed results; and H.Z., D.R., R.B. wrote the paper.

Abstract

Her4, a receptor tyrosine kinase, is mutated in malignant melanoma and lung cancer. Following binding of its growth-factor ligand, Her4 dimerizes, autophosphorylates, and activates several intracellular signaling pathways. To understand the dimerization and phosphorylation and to provide a fundamental basis for drug design and therapeutic intervention in cancer, we implemented footprinting of Her4 kinase domain.We used a water-soluble carbodiimide, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), in the presence of glycine ethyl ester (GEE), to modify solvent-accessible carboxyl groups from glutamate and aspartate. Digestion and LC-MS/MS of the modified Her4 kinase monomer, homodimer, and the homodimer that was phosphorylated on the activation loop provide information to map the dimerization interface and to determine phosphorylation induced-conformational changes. Thirty seven glutamate/aspartate residues were modified, and four residues showed reduced extents of modification in the Her4 dimer as compared to the monomer. Three of the residues (Glu-690, Glu-692, and Glu-912) are at the dimer interface, on the basis of prior protein crystallography, and one residue (Asp-853) is on the activation loop. The footprinting results confirm the existence of the Her4 kinase domain asymmetric dimer in solution. Further investigation of the Her4 kinase domain conformational changes during autophosphorylation showed that Her4, when phosphorylated at Tyr 850, undergoes less modification on three residues in the activation loop and this suggests that phosphorylation causes this loop to become less flexible. We extended this footprinting approach to measure the dimer association constant for the Her4 kinase domain. Mathematical modeling of titration curves that show modification extent as a function of

protein concentration afforded a Her4 kinase domain dimer association constant in the range 1×10^8 to 4.5×10^8 M^{-1}. The results show that carboxyl group footprinting is an effective means to study membrane-associated Her family homodimers and may be generally useful for other membrane proteins.

Introduction

Her4 is a transmembrane receptor tyrosine kinase that plays a vital role in the cardiovascular system, nervous system, and other tissues(*1*, *2*). Her4 (also called ErbB-4) binds the neuregulin family of growth factors. Genetic knock-out of Her4 in mice results in embryonic lethality due to malformations of the heart and hindbrain(*3*). Her4 functions by binding the neuregulin growth factors, dimerizing, activating its tyrosine kinase domain, and phosphorylating Tyr residues on Her4 itself and on downstream signaling proteins. Protein crystallography showed those regions in both the extracellular domain and kinase domain that contribute to Her4 dimerization(*4*, *5*). The kinase domain dimerization contacts are particularly interesting because they mediate an allosteric activation of the Her4 tyrosine kinase domain(*6*). The Her4 kinase domain dimer is asymmetric where a surface on the N-lobe of one kinase monomer (involving residues 680-691 and 734-745) contacts a surface on the C-lobe of the other kinase monomer (involving residues 910-960). This dimer interface was identified on the basis of crystal packing interactions in the Her4 crystal structure(*5*), but the interface has not yet been verified to exist in solution. Two important questions are: what is the affinity of this dimerization in solution and how does the subsequent phosphorylation step change the conformation of the Her4 kinase domain? We propose that footprinting mass

spectrometry can address these questions, and we report the implementation of a suitable method in this article.

Footprinting to probe protein structure has been rapidly developing over 20 years(*7*). Notable approaches include acetylation (*8*), amide hydrogen/deuterium exchange(H/DX)(*9*), and hydroxyl radical footprinting(*10*). MS-based footprinting methods have been successful for monitoring protein folding/unfolding dynamics(*11*), characterizing protein-ligand interaction(*12, 13*), and protein oligomerization(*14*). For probing complicated systems, such as membrane-associated proteins, H/DX suffers limitations due to significant back exchange accompanying the demanding post-labeling purification. Similarly, for hydroxyl radical based footprinting, an extra purification step (e.g., SDS-PAGE) may induce false readout in oxidations(*15*). Nevertheless, improvements in H/DX(*16*) and hydroxyl radical based footprinting (*17, 18*) methods show promise in probing membrane proteins.

Although less general than H/DX and radical footprinting, labeling of carboxyl groups by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)-mediated incorporation of glycine ethyl ester (GEE) is a possible alternative(*19*). Carboxyl groups of protein are activated by the carbodiimide, and then the nucleophilic reagent, GEE, attacks the activated carboxyl group to form products (Figure 4.1A). The properties of this reaction make it appropriate for probing proteins in complicated systems. The coupling reaction is highly efficient in biologically relevant buffer systems (*20*). Biological mixtures can be directly modified without pre-separation of the proteins, affording an opportunity to probe proteins in their native state. Irreversibly modified products survive any intensive post-labeling purification. An example of this approach is

the determination of the membrane orientation of the FMO antenna protein in photosynthetic *Chlorobaculum tepidum* (*21*).

In this paper, we applied carboxyl group footprinting to study the conformational changes of Her4 kinase domain upon dimerization and phosphorylation when the protein is bound to the surface of a liposome, which serves as a model for the membrane system. To our knowledge, this is the first reported investigation in solution of Her4 dimerization and phosphorylation-induced conformational changes. We expect that differences in the modification extent between Her4 monomer and dimer will show the existence of a Her4 asymmetric dimer in solution and provide information about the dimerization interface. If tyrosine (Tyr-850) phosphorylation indeed induces conformational changes on the activation loop, the carboxyl group footprinting should report this and provide a deeper understanding the Her4 kinase domain dimerization. Moreover, using the carboxyl group modifications in a self titration format, we hope to develop a means to determine the association constant for Her4 kinase domain dimerization. Success of this approach to study the Her (ErbB) family homodimers offers promise for understanding the physiology of cardiovascular and nervous system and for developing new cancer treatments.

Material and Methods

Chemicals and Reagents

Dioleoyl-phosphatidylcholine (DOPC) and nickel-1,2-dioleoyl-sn-glycero-3-([N-(5-amino-1-carboxypentyl)iminodi acetic acid]succinyl)-nickel salt (Ni-NTA-DOGS) were purchased from Avanti Polar Lipids (Alabaster,AL). Water, acetonitrile, formic acid, glycine ethyl ester, EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride,

ammonia acetate, trypsin from porcine pancreas were obtained from Sigma-Aldrich (St. Louis, MO). The peptides LLEGDEKEYNADGGK and LLEGDEKEY(Phos)NADGGK were synthesized by GenScript USA Inc. (Piscataway, NJ).

Protein expression and liposome preparation

The protocols for Her4 protein expression and liposome preparation were previously reported (*6*). All protein and liposome preparations in this experiment followed this protocol without modification. Nickel-liposomes contained 95 mol% DOPC and 5 mol% Ni-NTA-DOGS. Control liposomes contained 100 mol% DOPC. Her4 residue numbering is based on the mature, full length protein minus its signal peptide and matches the numbering used by Qiu et al.(*5*).

Protein dimerization and phosphorylation

Protein stock solutions were prepared so that they contained 20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT (pH 8.0). Dimerization experiments started by equilibrating protein with nickel liposomes for 15 min on ice. Protein phosphorylation was initiated by adding ATP and $MgCl_2$ to the protein dimer solution (final concentration 100 μM ATP and 10 mM $MgCl_2$). After 5 min incubation on ice, phosphorylation samples were ready for modification without quenching.

Carboxyl group modification reaction

10 μL of 2.5 μM Her4 sample in 20 mM Tris-HCl, 100 mM NaCl, pH 7.5 in 0.5 mL Eppendorf protein LoBind tube (Eppendorf North America, Hauppauge, NY) were prepared for the modification reaction. 0.5 μL of GEE stock solutions (2 M in water) with 0.5 μL of EDC stock solution (50 mM in water) were mixed to initiate the reaction. The reaction was quenched at 10 min by adding an equal volume of 1 M ammonium

acetate. For chemical modification of a synthetic peptides, the peptides were kept at 5 nM in Tris (20mM Tris-HCl, 100mM NaCl, pH 7.5). The reaction was conducted under identical conditions as above. In titration experiments, Her4 samples with different protein concentrations (from 0.0375 to 1.25 μM) were used in the modification reaction.

Gel separation and in-gel trypsin digestion

Immediately after the carboxyl group modification,the reaction was quenched by ammonium acetate. A 5X SDS-PAGE sample loading buffer, containing DTT as the reducing reagent, was mixed with each sample and the mixture was boiled for 3 min. The samples were loaded into 10% SDS-PAGE gel. The gel was stained by SimplyBlue SafeStain (Invitrogen, Carlsbad, CA). In-gel digestion was performed as per the method of Shevchenko et al. (*22*) except the extraction of peptide digestion products was modified. Extraction buffer A (5% formic acid, 0.1% trifluoroacetic acid and 50% acetonitrile in water) was added to each tube (solution:gel = 2:1 v/v), and the tube submitted to shaking in Thermomixer (Eppendorf North America, Hauppauge, NY) at room temperature for 10 min. The supernatant was collected in an Eppendorf protein LoBind tubes. The second extraction was done the same way by using extraction buffer B (5% formic acid and 0.1% trifluoroacetic acid in acetonitrile). The supernatants were combined and dried with a speed vacuum.

ESI-MS of Her4 protein

ESI mass spectra were acquired in the positive-ion mode on a Waters (MicroMass) Q-TOF Ultima (Manchester, U.K.) equipped with a Z-spray ESI source. The Instrument parameter setup and trap column desalting for protein analysis were reported previously(*8*).

Circular Dichroism

Samples of Her4 for which the carboxyl group was either unmodified or modified were dialyzed into a CD Buffer (10 mM NaPO4, pH = 7.2, 5 mM NaCl) by using Amicon Ultra 10k spin-concentrators and centrifuging the samples to ~100 µL and filling the concentrator with CD Buffer three times. The samples, ~2 µM, were transferred to a 1 mm path length quartz cell. CD spectra were measured at room temperature from 195-260 nm wavelengths at 1 nm intervals in a JASCO J715 CD spectrometer (JASCO Analytical Instruments, Tokyo, Japan). Three scans were taken and averaged, then corrected with a buffer blank. The protein concentration was determined by the Bradford assay by using BSA as a standard.

LC-ESI-MS/MS of Her4 tryptic peptides

Peptide samples from the in-gel digestion were reconstituted with 20 µL water containing 0.1% trifluoroacetic acid. C18 zip tips (Millipore Co., Billerica, MA) were used to remove salts. The peptide sample was eluted with 75% acetonitrile 25% water 0.1% trifluoroacetic acid solution and dried by speed vacuum. Dry tryptic peptides were reconstituted with 15 µL solvent A (water, 0.1% formic acid). An aliquot (5 µL) was injected by Eksigent NanoLC-Ultra 1D (Eksigent Technologies, Inc. Livermore, CA) into a custom-packed nano column. This reverse-phase nano column was custom-built by packing C18 material (Magic, 5 µm, 120Å, Michrom Bioresources, Inc., Auburn, CA) into silica capillary tubing with a PicoFrit tip (75 µm × 200 mm, New Objective, Inc., Woburn, MA). The gradient was from 2% to 60% solvent B (acetonitrile, 0.1% formic acid) over 60 min at 260 nL/min. The gradient was followed by a 10 min 80% solvent B wash and a 12 min re-equilibration with 100% solvent A. A PicoView Nanospray Source

(PV550, New Objective, Inc., Woburn, MA) was used with an LTQ Orbitrap (Thermo-Scientific, San Jose, CA). The nano electro-spray parameters were tuned by direct infusion of a Angiotensin II solution (10 μg/mL, 70% solvent A, 30% solvent B). LC-MS data were acquired in standard data-dependent mode controlled by Xcalibur 2.0.7 software. Peptide mass spectra (*m/z* range: 350-2000) were acquired at high mass resolving power (60,000 for ions of *m/z* 400). The six most abundant ions were fragmented by CID (collision-induced dissociation) in the linear ion trap. CID experiment parameters were: isolation width, 2 Da; activation time, 30 ms; normalized collision energy, 35%; minimum ion counts, 1000. The mass calibration was checked and repeated regularly by using a standard calibration mixture of caffeine, short peptide MRFA and Ultramark 1621.

Database searching of LC-MS/MS results

Each product-ion mass spectrum was extracted from the raw data file by DTA creator of Rosetta Elucidator (v3.3.0.0.220, Rosetta Biosoftware, Seattle, WA). The combined dta files were searched by using Mascot (version 2.2.06, Matrix Science, London, UK) against a custom-built database containing His_6-tagged Her4 kinase domain sequence. The custom-built database was established by adding all analysis targets of our lab (30-40 entries) into the E.coli database (approximately 39,000 entries from NCBI nr database). Custom-built modification profiles (all carboxyl group modification products), phosphorylation (tyrosine, threonine and serine) and oxidation (methionine, tryptophan and histidine) were used as variable modifications. Carbamidomethylation of cysteine was considered in the database searching as a fixed modification. The other parameters in Mascot searching were: enzyme, trypsin;

maximum number of missed cleavage, one; peptide mass tolerance, 15 ppm; ^{13}C isotopic peak consideration, one; product-ion mass tolerance, 0.8 Da; instrument type, ESI-trap.

Data processing

Raw format data were loaded onto the Rosetta Elucidator (v3.3.0.0.220, Rosetta Biosoftware, Seattle, WA). The alignment retention time and calculating peak area were determined by using the Elucidator PeakTeller algorithm. Parameters in the algorithm were as in the default setup. Dynamic background subtraction and smoothing across the alignment window (both in retention time and *m/z* dimensions) were used. All features were assigned a feature ID; a product-ion mass spectrum was associated with a feature by judging the retention time and precursor *m/z*; the outcome of Elucidator was further processed by using an Excel macro written in our lab. Modification extents were calculated based on the peak areas of all extracted ion chromatogram features.

Equilibria Modeling

The inference of the Her4 dimerization association constant on the surface of the liposome relied on a bulk solution model, which was later reinterpreted to give the desired surface binding constants. Specific binding (the law of mass action applies) and the principle of microscopic reversibility(*23, 24*) were assumed in the model. The strategy for model calculation was an extension from one dimension to two dimensions of the strategy employed by Zhu(*25, 26*). The mass conservation law for the Her4 is shown in Eq. 1 where $[Her4]_{TOTAL}$ is the total analytical concentration of Her4 and $[Her4]$ is the solution concentration of the free monomer Her4.

$$[Her4]_{TOTAL} = [Her4] + \beta_{1,1}[B][Her4] + 2\beta_{2,2}[B]^2[Her4]^2 + \beta_{1,3}[P][Her4] \quad \text{(Eq.1)}$$

The overall Adair binding constant for attachment of the Her4 monomer to the liposome is given by $\beta_{1,1}$ ($\beta_{1,1} = K_{11}$; the term K_{11} was used in results section); the overall Adair binding constant for the dimerization of the Her4 on the liposome surface is $\beta_{2,2}$. The equivalent bulk concentration of the free nickel binding sites on the liposome surface is $[B]$. The model incorporated a provisional collection of parasitic binding sites for the Her4 that competed with the liposome binding and removed Her4 from the analytical path to detection. The equivalent bulk solution concentration of the free form of these sites is given by $[P]$. The total of the nickel binding sites on the liposomes has an equivalent bulk solution binding concentration of B_{MAX} as shown in Eq. 2 for the corresponding mass conservation law for the nickel binding.

$$B_{MAX} = [B] + \beta_{1,1}[B][Her4] + 2\beta_{2,2}[B]^2[Her4]^2 \qquad \text{(Eq. 2)}$$

The parasitic binding sites also have an equivalent total bulk solution concentration P_{MAX} and a corresponding mass conservation law as shown in Eq 3.

$$P_{MAX} = [P] + \beta_{1,3}[P][Her4] \quad \text{(Eq. 3)}$$

Eqs 1 and 2 were viewed as components of the vector Eq 4, which expresses $[Her4]_{TOTAL}$ and B_{MAX} as a function of $[Her4]$ and $[B]$.

$$\begin{bmatrix} [Her4]_{TOTAL} \\ B_{MAX} \end{bmatrix} = \vec{F}([Her4], [B]) \qquad \text{(Eq. 4)}$$

Calculation of model curves was achieved by evaluating the inverse function $\vec{F}^{-1}$ in Eq 5, which expresses $[Her4]$ and $[B]$ as a function of $[Her4]_{TOTAL}$ and B_{MAX}.

$$\begin{bmatrix} [Her4] \\ [B] \end{bmatrix} = \vec{F}^{-1}([Her4]_{TOTAL}, B_{MAX}) \quad \text{(Eq. 5)}$$

The evaluation follows the trajectory of the titration experiment as it moves along a path $\vec{l}(\tau)$ in the $[Her4]_{TOTAL} \times B_{MAX}$ domain as described in Eq 6.

$$\begin{bmatrix}[Her4]\\ [B]\end{bmatrix}(\tau) = \vec{F}^{-1}\left(\vec{l}(\tau)\right) \qquad \text{(Eq. 6)}$$

A total derivative with respect to the path variable τ, which marks off the progress of the titration, gives Eq 7.

$$\frac{d}{d\tau}\begin{bmatrix}[Her4]\\ [B]\end{bmatrix} = \left[\frac{\partial \vec{F}^{-1}}{\partial([Her4]_{TOTAL}, B_{MAX})}\right] \cdot \frac{d\vec{l}}{d\tau} \quad \text{(Eq. 7)}$$

By the inverse function theorem (*27*, *28*)

$$\frac{d}{d\tau}\begin{bmatrix}[Her4]\\ [B]\end{bmatrix} = \left[\frac{\partial \vec{F}}{\partial([Her4],[B])}\right]^{-1} \cdot \frac{d\vec{l}}{d\tau} \qquad \text{(Eq. 8)}$$

which can be viewed as a first-order, ordinary differential equation of a dynamical system and was solved with the "Rkadapt" function in Mathcad 14 (Parametric Technology Corporation, Needham, MA). The first leg of the path, which accounted for the addition of the binding sites on the liposomes to the system, moved from $(0,0)$ to $(0, B_{MAX})$ for the value of $([Her4]_{TOTAL}, B_{MAX})$ for which the solution of $([Her4],[B])$ moved trivially from the initial value of $(0,0)$ to $(0, B_{MAX})$. The more consequential second leg, which accounted for the titration with Her4, was given by (τ, B_{MAX}), where $\boldsymbol{\tau}$ varied over the domain interval of $[Her4]_{TOTAL}$ from zero to 17.3 uM. Rkadapt was set to record the solution for the second leg in a table of triples $(\tau, [Her4], [B])$ for 5000 equal length intervals. The accuracy of the solution was checked by recalculating $[Her4]_{TOTAL}$ by Eq 1 and B_{MAX} by Eq 2 from values for $[Her4]$ and $[B]$ extracted from the table. For the table triples, the relative error of the recalculated $[Her4]_{TOTAL}$ was typically less than 2×10^{-12}, and the relative error of the recalculated $[B]$ was less than 1×10^{-13}. For linear interpolations between table entries, the relative error of the recalculated $[Her4]_{TOTAL}$ typically varied from 2×10^{-3} at low concentrations to 1×10^{-6} at high concentrations for

Her4 and the relative error of the recalculated $[B]$ typically varied from 7×10^{-6} at low concentrations to 4×10^{-7} at high concentrations for Her4.

The signal function $Frac_{modified}$ (Eq. 9) was constructed to represent the transfer function of the mass spectrometer and the following spectrum analysis.

$$Frac_{modified} = \frac{[Her4]+\beta_{1,1}[B][Her4]+f\{2\beta_{2,2}[B]^2[Her4]^2\}}{\{[Her4]+\beta_{1,1}[B][Her4]+f\{2\beta_{2,2}[B]^2[Her4]^2\}\}+\{G(1-f)\{2\beta_{2,2}[B]^2[Her4]^2\}\}} \quad \text{(Eq. 9)}$$

It was assumed that the spectral signals varied in proportion to the concentrations of their corresponding solution species. The numerator in Eq 9 gives the concentrations of those species that were assumed to be modified by GEE. The factor f allows for a fraction of the Her4 dimer population to be modified in spite of the supposed protection. The remaining product term in the denominator represents the concentration of those Her4 molecules existing as dimers that were not modified. A relative sensitivity factor G accounts for the different sensitivities of the mass spectrometer to various peptides and their modified counterparts.

The inference of the Her4 dimerization association constant was accomplished by performing a nonlinear-least-square fit of a $Frac_{modified}$ curve to the experimentally determined extents of modification plotted as the curve of the Her4 titration experiment. The search minimizes the root mean square of the residuals (Residual RMS). In each trial of the search for the fit, Eq 8, was solved by using known fixed parameter values or trial parameter values posed by the Quasi-Newton algorithm of the "Minimize" function in Mathcad and a $Frac_{modified}$ curve was calculated. Fixed parameters were K_{11}, B_{MAX}, and G,. The parameters K_{11} and B_{MAX} were selected as described in the Results section. The parameter G was determined to have value of 1.1236 by using Arg ^{15}N labeled Her4 as standard. The searched parameters were f, $\beta_{2,2}$, $\beta_{1,3}$,

and P_{MAX}. The equivalent bulk solution dimerization association constant K_{22} was computed by evaluating Eq. 10 after the fit by $Frac_{modified}$.

$K_{22} = \beta_{22}/(\beta_{11}{}^2)$ (Eq. 10)

Results

Carboxyl Group Footprinting Method

In this MS-based footprinting, protein samples are modified by a coupling reaction between the carboxyl side chain of proteins and the primary amine of GEE, a reaction driven by EDC (Figure 4.1A). This reaction is performed at physiological pH and salt concentration and yields products of +85.0522 Da or +57.0209 Da. The modified protein can be directly analyzed by ESI-MS using a C18 trap column to desalt the sample, or can be digested by trypsin and analyzed by LC-ESI-MS/MS to quantify the site of modification. With accurate mass measurement, the percentage of modification (the modification extent) is determined with little ambiguity on the basis of the peak area of the modified peptide divided by the peak area of all forms of that peptide. The site of carboxyl group modification is located from the product-ion spectrum (Figure 4.1B).

Global information of modified Her4 protein

Multiply charged protein ions were observed upon ESI of the Her4 kinase domain protein (Figure 4.2A). The most abundant charge state was +51. Deconvolution ("de-chargiing") of the spectrum gave a protein molecular weight (MW) of 42,167 ± 3 Da. This measured protein MW agreed well with the theoretical MW provided the the protein N-terminus is acetylated (MW= 42,166 Da). A second series of multiple charged ions was clearly observed with extended view of charge state +51 (Figure 4.2B). The MWs of the species corresponding to the second and third peaks are incremented by +80 Da,

suggesting low levels of phosphorylation of Her4. This basal phosphorylation of Her4 is assessed in Figure 4.6.

To test the reactivity of carboxyl group modification and to limit the overall degree of modification of the protein, we measured the time course of carboxyl group modification of Her4 kinase domain by comparing the ESI mass spectra of modified proteins at different reaction times. In the extended view of charge state +51 (Figure 4.2B), the relative abundance of the ions of *m/z* = 829.4 and 831.0 increased with increasing modification time with GEE. These ions are consistent with addition of +85 Da ethyl ester groups to the carboxylic acid side chains. The relatively low abundance in the mass spectrum at zero time likely represents basal phosphorylation of the protein, as discussed above. The mass shifts induced by ethyl ester groups (+85 Da) and by phosphorylation (+80 Da) are similar and at the +51 charge state, this 5 Da nominal mass difference corresponds to an *m/z* difference of 0.098.

To insure that the carboxyl group modification performed here did not cause protein unfolding or major conformational change, we obtained circular dichroism (CD) spectra before and after carboxyl group modification. The CD spectra (Figure 4.2C) showed no difference between carboxyl group modified Her4 and unmodified Her4, demonstrating that carboxyl group modification caused no large conformation change in the protein. Because the carbodiimide can function as a cross linker, there is a concern that a coupling reaction occurred to form a covalent dimer. By adding GEE prior to the EDC, any intermolecular coupling was avoided as determined from a gel assay (Figure 4.3).

LC-MS/MS of Her4 kinase domain digest

Digestion gave 98% sequence coverage over the Her4 kinase domain sequence. The two missing regions correspond to two small peptides, each of which contained three or four amino acids. Beside common protein oxidation (on methionine, tryptophan and histidine), six phosphorylated residues (three on tyrosine, two on serine and one on threonine) were detected (Figure 4.4), and their significance is addressed below.

We found that 37carboxyl groups were modified: 20 on Glu and 17 on Asp. The Her4 kinase domain protein sequence contains a total of 47 Asp/Glu sites, and the carboxyl group modification conditions employed here modifies 79% of those residues. We verified all sites of carboxyl group modification reported here by manually inspecting the product-ion spectra. The modification assignments are problematic when applied for low- abundance ions (< 10,000 counts in the product-ion spectrum). Peaks corresponding to those low-abundance ions were not included among the peptides used for quantitative analysis. Modification extents data represent mean and standard deviation of three independent samples.

Footprinting between Her4 monomer and dimer

We generated Her4 dimers by binding the His_6-tagged Her4 kinase domain protein to a liposome containing the nickel chelating lipid Ni-NTA-DOGS, as described in Monsey et al. (*6*). The binding of the His_6-tagged Her4 to these "nickel liposomes" results in a high local concentration of Her4 that allows the dimers to form nearly spontaneously. The His_6-tagged Her4 kinase domain protein is predominantly a monomer in solution, as demonstrated by gel filtration analysis (data not shown). As an additional control, we tested Her4 incubated with control liposomes that lack the nickel-chelating lipid to evaluate for any non-specific effects from the liposomes on the

carboxyl group modification reaction. These three conditions (i.e., Her4 monomer in solution, Her4 monomer plus control liposomes, and Her4 dimer on nickel liposomes) were reacted with GEE in the presence of EDC in an identical way and analyzed by in-gel trypsin digestion and LC-MS/MS. Modification extents for the 37 Asp/Glu of Her4 kinase domain protein were compared (Figure 4.5A and Table 4.1). Decreases in modification extents of Her4 dimers compared to monomers were observed for several residues. Two N-lobe residues, Glu-690, Glu-692, one activation loop residue, Asp-853, and one C-lobe residue, Glu-912 all showed reduced modification extent in Her4 dimers. Based on the crystal structure of the Her4 kinase dimer(*5*), the C-lobe of one monomer (the donor monomer) contacts the N-lobe of the other monomer (the acceptor monomer) to form an asymmetric dimer (Figure 4.5B). Glu-690, Glu-692, and Glu-912 are all located in the dimer interface in the Her4 kinase domain crystal structure (Figure 4.5B). However, Asp-853, which also showed reduced carboxyl group modification, is not located near the dimerization interface; rather it is on the activation loop, which is a flexible loop that can adopt a number of different conformations(*5*, *29*). In fact, Asp-853 and its neighboring amino acids were not visualized on the Her4 crystal structure(*5*), further supporting the flexibility of this loop. Based on prior crystallographic studies of Her4 and its family member EGFR kinase, it is reasonable to expect that the conformation of the activation loop will be altered by dimerization(*5*, *30*). Therefore, the lower carboxyl group modification of Asp-853 likely represents an allosteric effect of dimerization.

One residue that shows increased carboxyl group modification is Asp-897 (Figure 4.5A), which is located on the C-lobe but away from dimer interface (> 15 Å). Similar to

the decreased modification of Asp-853, the increased modification of Asp-897 may involve allosteric changes induced by dimerization. No other residues showed any statistically significant increase in carboxyl group modification upon dimerization.

Footprinting on Her4 dimer autophosphorylation

A previous study of Her4 phosphorylation demonstrated that Tyr-850 in the Her4 activation loop can become phosphorylated(*31*). For a closely related kinase, Her2/neu, we previously showed that the homologous activation loop Tyr residue (Tyr-877) is also phosphorylated in cells(*32*). To study the conformation changes induced by phosphorylation, we first incubated Her4 monomers and Her4 dimers with ATP and $MgCl_2$. A low level of phosphorylation on six residues was detected even prior to ATP addition (Figure 4.6A), likely representing basal phosphorylation of the recombinant Her4 protein. Consistent with our previous report(*6*), phosphorylation levels of Her4 monomer showed only small changes when incubated with 100 μM ATP for 5 min at 4 ^{o}C. A dramatic increase in Tyr-850 phosphorylation, however, occurred when Her4 dimers were incubated with ATP under the same conditions. The phosphorylation extents were calculated in the same way as for the carboxyl group modification. Although these phosphorylation extents do not represent the absolute stoichiometry of phosphorylation in the sample, they do provide a reliable relative change in phosphorylation between samples.

We next compared modification extents of the Tyr-850 phosphorylated Her4 dimers to the Her4 dimers that had not been incubated with ATP. Given that the phosphorylation of Tyr-850 did not reach 100% stoichiometry, the Her4 dimer that was incubated with ATP represents a mixture of Tyr-850 phosphorylated and

unphosphorylated proteins. In a bottom-up approach, phosphorylated and unphosphorylated peptides can be separated by a LC and identified by MS based on the mass shift induced by phosphorylation, and their carboxyl group modification extents can be calculated separately. Measuring the modification extent for the unphosphorylated Tyr-850 containing peptides can be viewed as another control experiment. We found that, of the five possible modification sites in the peptide containing Tyr-850, four were modified (Figure 4.6B and Table 4.2). The modification extent was significantly decreased for Glu-847, Glu-849 and Asp-853 in the Tyr-850 phosphorylated state of Her4 as compared to the state for which Tyr-850 is unphosphorylated (Figure 4.6C).

To rule out that decreases in the carboxyl group modification are due simply to the proximity of the acidic phosphotyrosine group, we synthesized the tryptic peptide containing Tyr-850 in both its phosphorylated and unphosphorylated forms, and submitted them to carboxyl group modification. The carboxyl group modifications levels were similar whether the Tyr-850 was phosphorylated or unphosphorylated (Figure 4.6D). This indicates that the decrease in Glu-847, Glu-849, and Asp-853 modification upon activation loop phosphorylation of dimeric Her4 is due to a conformational change in this loop. For several other kinases, activation-loop phosphorylation results in formation of salt bridges between the phosphorylated residue and Arg side chains, and these bridges stabilize the activation loop structure(*33*, *34*). Similarly, we predict that the phosphorylated activation loop in Her4 will be stabilized and become less flexible; as a result, several of its residues will be protected from the carboxyl group modification. The results provide evidence that dynamic variations, which are usually not seen in a crystal structure of a protein, can be "trapped" by carboxyl group modification. The carboxyl

group modification takes place over minutes of time, thereby presenting an integrated view of protein dynamics that occur over that time.

Titration of Her4 kinase domain

To gain a deeper understanding of Her4 kinase domain dimerization, we designed a quantitative analysis of dimerization. Like we did in a previously reported H/D exchange method (SIMSTEX)(*14*), we used carboxyl group modification to determine the Her4 kinase domain dimer association constant. We implemented a titration experiment that was started by equilibrating nickel liposomes with different concentrations of the Her4 kinase domain protein. After reaching equilibrium, the samples were modified by EDC and GEE in an identical manner for each. By following the modification extent vs. protein concentration, we obtained titration curves and formulated mathematical models to fit the titration curve and afford the Her4 dimer association constant. Observed decreases in modification extents are in agreement with results of carboxyl group modification between Her4 dimers and monomers (Figure 4.5A and figure 4.7). The titration curve of Asp-847 in the presence of nickel liposomes (Figure 4.8A) was used in the dimerization modeling. As a control, we performed a similar titration experiment in the absence of liposomes, and we observed that the modification extent did not vary with Her4 concentration (data not shown). Therefore, we conclude that the titration curve (Figure 4.8A) is due to the dimerization equilibrium on the surface of the nickel liposomes.

We generated liposome-bound Her4 dimers by binding the His_6-tagged Her4 kinase domain protein to the nickel-liposome(*6*). This results in a high local

concentration of Her4, creating an equilibrium between Her4 monomers and dimers on the liposome surface (Figure 4.8B). We first assumed that all of Her4 was bound to the liposome, either as monomer or dimer and tested this to see that there is clear bias in this model; the fitted parameters *G* and *f* were clearly unreasonable (Figure 4.9). To adjust, we made a simple increment in model complexity that resulted in the removal of model bias. The increment incorporated four states for the Her4 kinase domain (Figure 4.8B): free monomer in solution, monomer on liposome surface, dimers on liposome surface, and binding of Her4 to plastic walls or other surfaces. The latter state, which is termed "parasitic binding sites" in the method section, removes Her4 from MS analysis. With this model, an equation giving the carboxyl group modification extents (Figure 4.8A) was established as $Frac_{modified}$ in Eq.9.

Choices for two parameters, K_{11} and B_{MAX}, were made based on the following considerations. Changes in K_{11}, the affinity of His_6-tagged Her4 to nickel-liposome produced little change in the shape of the fitted model curve (Figure 4.8C). Fits were calculated for K_{11} values of 3.3×10^5, 1×10^6, 3.3×10^6, and 1×10^7 M^{-1}, which is a reasonable range for the association of Her4 to nickel binding sites on the liposome surface (*6, 35*). Although the value for B_{MAX} could not be determined with precision, it was not likely to be greater than the value 4.2×10^{-6} M^{-1} that was determined in experiments performed under the same conditions by Monsey et al. (*6*). The minimum residual RMS for each K_{11} occurs along the left heavy line on the surface for which B_{MAX} ranges from 1.75×10^{-6} at the back to 1.28×10^{-6} M at the front of the surface (Figure 4.8C). Below these respective values, the model curves clearly exhibited bias with the curves turning up rapidly at the high end of the titration curve.

Based on this modeling, the Her4 kinase domain dimer association constant (K_{22}) is between 1×10^8 to 4.5×10^8 M^{-1}. Equilibrium constant in this range are not sensitive to changes in the liposome binding capacity (B_{MAX}) or the binding constant of Her4 to liposome (K_{11}) as shown in Figure 4.8C and 4.8D.

Discussion

Our current knowledge of the Her4 dimerization interface is based on protein crystallography studies. Although these crystal structures provide an excellent starting point, they have several limitations. First, the Her4 asymmetric dimer structures are inferred from the crystal packing of the kinase monomers, and had to be validated by mutational analysis. Second, these structures provides snapshots of one conformation of the active or inactive kinases and do not provide information about the dynamics or the transitions that the kinases move through in the liquid phase. Third, flexible regions of the kinase, particular the activation loop, frequently lack electron density and cannot be visualized in the crystal structure.

In contrast, MS methods are able to provide data on protein structure in solution and address questions of protein dynamics (*36*). MS can sample the ensemble of conformations of a protein in a time range extending from approximately 1 ms to many minutes. The nickel liposome system as a membrane-associated protein, provides a unique challenge for its study *in vitro* (*37*) because isolation of the protein following a footprinting reaction is time-demanding, making HDX a difficult approach. The successful implementation of another MS-based footprinting strategy for this *in vitro* study provides structural information for the first time for the protein complex in solution.

We expect that the reactivity of carboxylic acid side chains on Asp/Glu residues will be strongly influenced by their solvent accessibility. Therefore, protein-protein interactions that bury Asp and Glu side chains should result in decreased labeling of these residues. Crystal structure data on the Her4 kinase domain has shown that Her4 forms dimers (*5*). Two types of

interactions are involved in this dimer: interactions between hydrophobic residues form the major portion of the interface and are surrounded by several hydrogen bonds between the donor and acceptor monomers(*30*). The negative charged carboxylic acid side chains of Asp/Glu at physiological pH are major contributors to these hydrogen bonds. The decrease in modification extents for Glu-690, Glu-692 and Glu-912 confirms the existence of the Her4 asymmetric dimer on the liposome surface, as inferred from the crystal structure. Results presented here prove that this approach is sensitive to solvent accessibility of Asp/Glu. To probe solvent accessibility of hydrophobic residues, hydroxyl radical modification, like fast photochemical oxidation of protein (FPOP)(*38*), is required, and that study is under development in our group. We expect a comprehensive characterization of Her4 dimerization interface can be achieved by combining results from carboxyl group footprinting and FPOP.

To determine the effect of phosphorylation on the Her4 dimer and to demonstrate that footprinting can provide new structural information beyond what is already known from protein crystallography, we examined the effect of phosphorylation on Her4 kinase structure. Consistent with prior studies on Her2/neu and Her4, we observed activation loop phosphorylation on Tyr-850(*6, 31, 32*). In this experiment, a Tyr-850 phosphorylation-induced conformational change was probed by carboxyl group modification. Dynamic fluctuations of this activation loop are hard to study by X-ray crystallography. Based on our knowledge, the change of dynamic fluctuations on activation loop between phosphorylated and unphosphorylated is seen for the first time in solution phase by carboxyl group modification procedure. The results show clearly the stabilization of the Her4 activation loop induced by Tyr-850 phosphorylation.

To gain a deeper understanding of Her4 dimerization, a quantitative analysis of Her4 dimerization would be helpful. Given that Her4 is a transmembrane protein, measuring a dimerization association constant of Her4 protein is a challenge. Dimer contacts are known to exist in both the extracellular domain and intracellular kinase domain, and *in vitro* studies on the isolated extracellular domain of Her4 and other EGFR members were reported(*35*). To our

knowledge, however, no kinase domain dimerization affinity has ever been reported. MS- based methods including PLIMSTEX(*26*), SUPREX(*12*) and SPROX(*39*), demonstrate the potential of MS based approach to measure the dissociation constant or thermodynamics of protein-ligand interactions. We took a lead from those methods to adapt carboxyl group modification to study quantitatively the Her4 kinase domain dimerization. Mathematical modeling of Her4 kinase domain titration curve provides not only the kinase domain dimerization association constant to be between 1×10^8 to 4.5×10^8 M^{-1}, but also insights into the mechanism of kinase dimerization. The outcome is within a range that is comparable with association constants reported for extracellular domain association constants of EGFR family (*35*).

Acknowledgement

The authors thank Drs. Henry Rohrs and Reid Townsend for advice and help. Funding was provided by National Center for Research Resources of the NIH (2P41RR000954) to MLG and NIH (K22CA128951) to RB. Additional support was provided by Merck; MLG is a consultant for Merck.

References

1. Birchmeier, C. (2009) ErbB receptors and the development of the nervous system, *Exp Cell Res 315*, 611-618.
2. Pentassuglia, L., and Sawyer, D. B. (2009) The role of Neuregulin-1beta/ErbB signaling in the heart, *Exp Cell Res 315*, 627-637.
3. Gassmann, M., Casagranda, F., Orioli, D., Simon, H., Lai, C., Klein, R., and Lemke, G. (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor, *Nature 378*, 390-394.
4. Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M., and Leahy, D. J. (2005) The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand, *Proc Natl Acad Sci U S A 102*, 15024-15029.
5. Qiu, C., Tarrant, M. K., Choi, S. H., Sathyamurthy, A., Bose, R., Banjade, S., Pal, A., Bornmann, W. G., Lemmon, M. A., Cole, P. A., and Leahy, D. J. (2008) Mechanism of activation and inhibition of the HER4/ErbB4 kinase, *Structure 16*, 460-467.
6. Monsey, J., Shen, W., Schlesinger, P., and Bose, R. Her4 and Her2/neu tyrosine kinase domains dimerize and activate in a reconstituted in vitro system, *J Biol Chem 285*, 7035-7044.
7. Mendoza, V. L., and Vachet, R. W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry, *Mass Spectrom Rev 28*, 785-815.

8. Sperry, J. B., Shi, X., Rempel, D. L., Nishimura, Y., Akashi, S., and Gross, M. L. (2008) A mass spectrometric approach to the study of DNA-binding proteins: interaction of human TRF2 with telomeric DNA, *Biochemistry 47*, 1797-1807.

9. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass Spectrom Rev 25*, 158-170.

10. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem Rev 107*, 3514-3543.

11. Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *J Mass Spectrom 43*, 1021-1036.

12. Roulhac, P. L., Powell, K. D., Dhungana, S., Weaver, K. D., Mietzner, T. A., Crumbliss, A. L., and Fitzgerald, M. C. (2004) SUPREX (Stability of Unpurified Proteins from Rates of H/D Exchange) analysis of the thermodynamics of synergistic anion binding by ferric-binding protein (FbpA), a bacterial transferrin, *Biochemistry 43*, 15767-15774.

13. Zhang, J., Adrian, F. J., Jahnke, W., Cowan-Jacob, S. W., Li, A. G., Iacob, R. E., Sim, T., Powers, J., Dierks, C., Sun, F., Guo, G. R., Ding, Q., Okram, B., Choi, Y., Wojciechowski, A., Deng, X., Liu, G., Fendrich, G., Strauss, A., Vajpai, N., Grzesiek, S., Tuntland, T., Liu, Y., Bursulaya, B., Azam, M., Manley, P. W., Engen, J. R., Daley, G. Q., Warmuth, M., and Gray, N. S. Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors, *Nature 463*, 501-506.

14. Chitta, R. K., Rempel, D. L., Grayson, M. A., Remsen, E. E., and Gross, M. L. (2006) Application of SIMSTEX to oligomerization of insulin analogs and mutants, *J Am Soc Mass Spectrom 17*, 1526-1534.

15. Perdivara, I., Deterding, L. J., Przybylski, M., and Tomer, K. B. Mass Spectrometric Identification of Oxidative Modifications of Tryptophan Residues in Proteins: Chemical Artifact or Post-Translational Modification?, *J Am Soc Mass Spectrom 21*, 1114-1117.

16. Hebling, C. M., Morgan, C. R., Stafford, D. W., Jorgenson, J. W., Rand, K. D., and Engen, J. R. Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, *Anal Chem 82*, 5415-5419.

17. Pan, Y., Stocks, B. B., Brown, L., and Konermann, L. (2009) Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry, *Anal Chem 81*, 28-35.

18. Zhu, Y., Guo, T., Park, J. E., Li, X., Meng, W., Datta, A., Bern, M., Lim, S. K., and Sze, S. K. (2009) Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting, *Mol Cell Proteomics 8*, 1999-2010.

19. Hoare, D. G., Olson, A., and Koshland, D. E., Jr. (1968) The reaction of hydroxamic acids with water-soluble carbodiimides. A Lossen rearrangement, *J Am Chem Soc 90*, 1638-1643.

20. Hoare, D. G., and Koshland, D. E., Jr. (1967) A method for the quantitative modification and estimation of carboxylic acid groups in proteins, *J Biol Chem 242*, 2447-2453.

21. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting, *Proc Natl Acad Sci U S A 106*, 6134-6139.

22. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes, *Nat Protoc 1*, 2856-2860.

23. Tolman, R. C. (1925) The Principle of Microscopic Reversibility, *Proceedings of the National Academy of Sciences of the United States of America 11*, 436-439.

24. Lente, G. The connection between the second law of thermodynamics and the principle of microscopic reversibility, *Journal of Mathematical Chemistry 47*, 1106-1111.

25. Zhu, M. M., Rempel, D. L., and Gross, M. L. (2004) Modeling data from titration, amide H/D exchange, and mass spectrometry to obtain protein-ligand binding constants, *Journal of the American Society for Mass Spectrometry 15*, 388-397.

26. Zhu, M. M., Rempel, D. L., Du, Z., and Gross, M. L. (2003) Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX, *J Am Chem Soc 125*, 5252-5253.

27. Williamson, R. E., Crowell, R. H., and Trotter, H. F. (1968) *Calculus of Vector Functions*, 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

28. Krantz, S. G., and Parks, H. R. (2003) *The Implicit Function Theorem: History, Theory, and Applications*, Birkhauser, Boston.

29. Huse, M., and Kuriyan, J. (2002) The conformational plasticity of protein kinases, *Cell 109*, 275-282.

30. Zhang, X., Gureasko, J., Shen, K., Cole, P. A., and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor, *Cell 125*, 1137-1149.

31. Kaushansky, A., Gordus, A., Budnik, B. A., Lane, W. S., Rush, J., and MacBeath, G. (2008) System-wide investigation of ErbB4 reveals 19 sites of Tyr phosphorylation that are unusually selective in their recruitment properties, *Chem Biol 15*, 808-817.

32. Bose, R., Molina, H., Patterson, A. S., Bitok, J. K., Periaswamy, B., Bader, J. S., Pandey, A., and Cole, P. A. (2006) Phosphoproteomic analysis of Her2/neu signaling and inhibition, *Proc Natl Acad Sci U S A 103*, 9773-9778.

33. Russo, A. A., Jeffrey, P. D., and Pavletich, N. P. (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation, *Nat Struct Biol 3*, 696-700.

34. Hubbard, S. R. (1997) Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog, *EMBO J 16*, 5572-5581.

35. Ferguson, K. M., Darling, P. J., Mohan, M. J., Macatee, T. L., and Lemmon, M. A. (2000) Extracellular domains drive homo- but not hetero-dimerization of erbB receptors, *EMBO J 19*, 4632-4643.

36. Kiselar, J. G., and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting, *Journal of Mass Spectrometry*, DOI: 10.1002/jms.1808.

37. Shrout, A. L., Montefusco, D. J., and Weis, R. M. (2003) Template-directed assembly of receptor signaling complexes, *Biochemistry 42*, 13379-13385.

38. Hambly, D., and Gross, M. L. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin *International Journal of Mass Spectrometry 259*, 124-129.

39. West, G. M., Tang, L., and Fitzgerald, M. C. (2008) Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy, *Anal Chem 80*, 4175-4185.

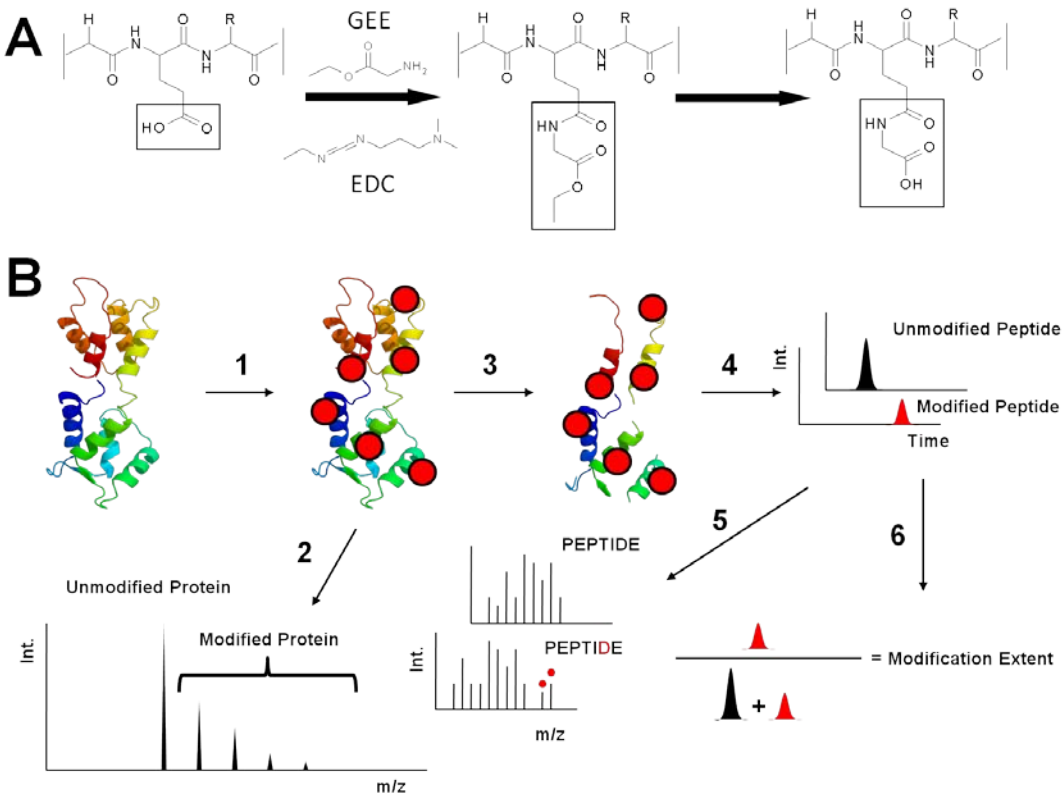


Figure 4.1. Carboxyl group modification based footprinting workflow and reaction. (A) Footprinting by carboxylic-acid, side-chain modificaiton. Carbodiimide (EDC) modifies acid side chain of proteins in physiological relevant buffer. In the presence of nucleophiles (glycine ethyl ester), reaction occur to give mass shifts 85.0522 Da and 57.0209 Da (hydrolysis product of ester). (B) Work flow of MS based carboxyl group protein footprint (1) Modification (2) ESI-MS analysis of modified protein (3) In-gel separation and proteolytic digestion (4) LC-MS analysis of peptides (5) LC-MS/MS analysis of peptides (6) calculation modified peptide level

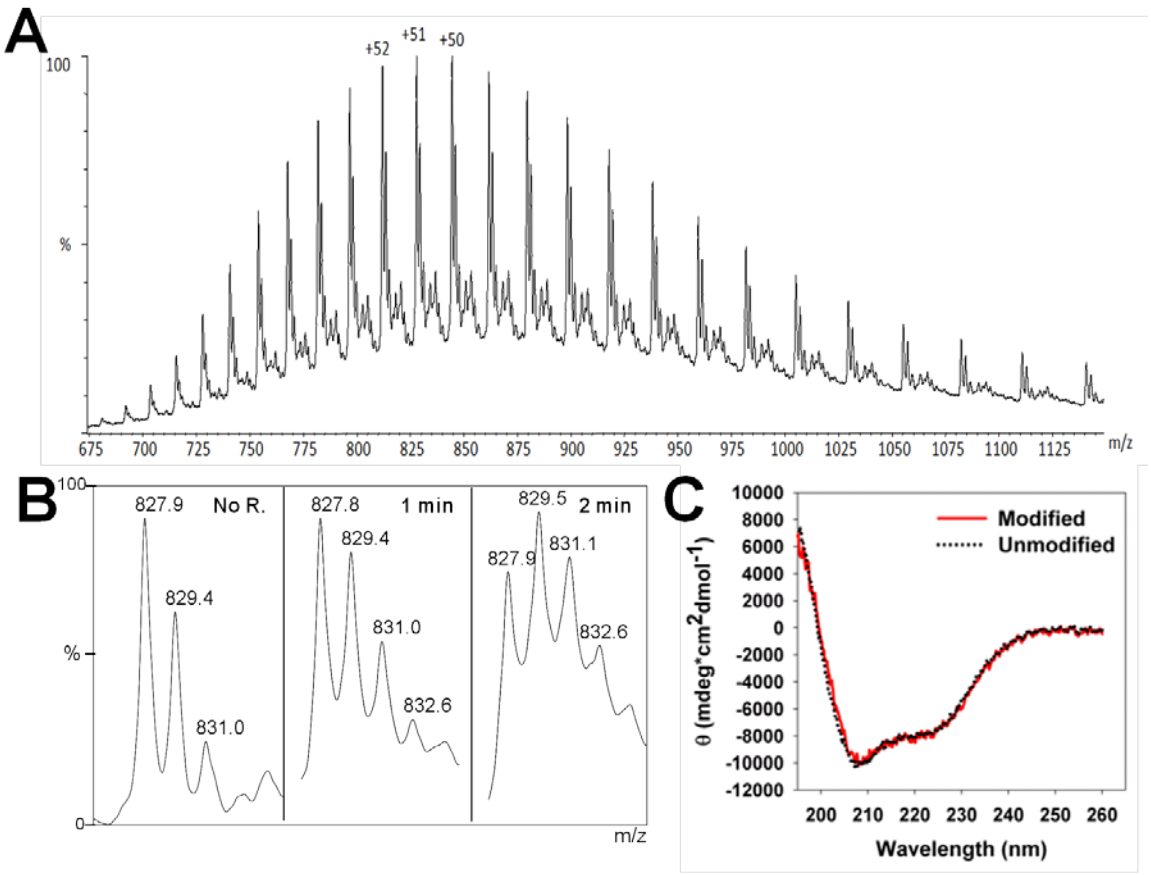


Figure 4.2. MS spectra of modified protein: (A) ESI-MS spectra of Her4 kinase domain protein. (B) Magnified view of +51 charge state with different reaction times. The ion of *m/z* 827.9 represents the unmodified ion peak. Two peaks separated by 80 Da (corrected by +51 charge state) were observed. (C) CD spectrum of protein with and without carboxyl group modification.

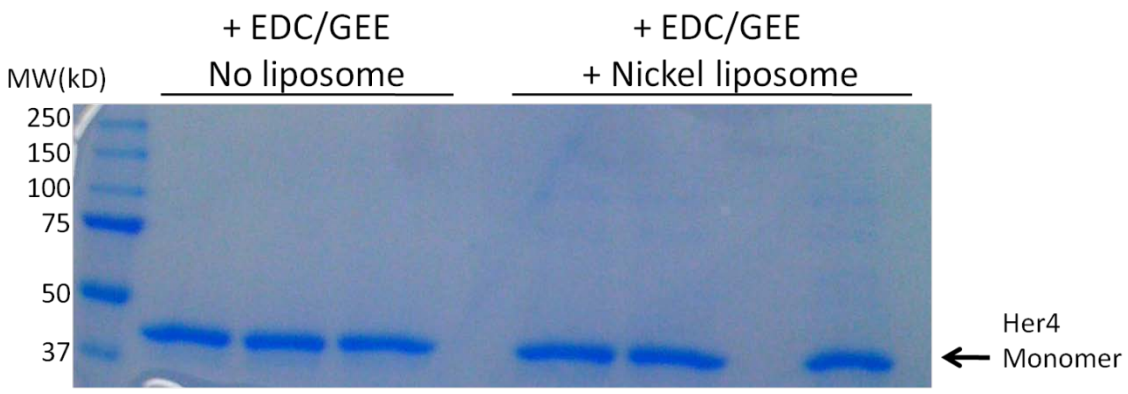


Figure 4.3. SDS PAGE of modified Her4 kinase domain protein. Only one band, corresponding to the Her4 monomer, was seen in each sample. No cross-linked Her4 dimer was detected.

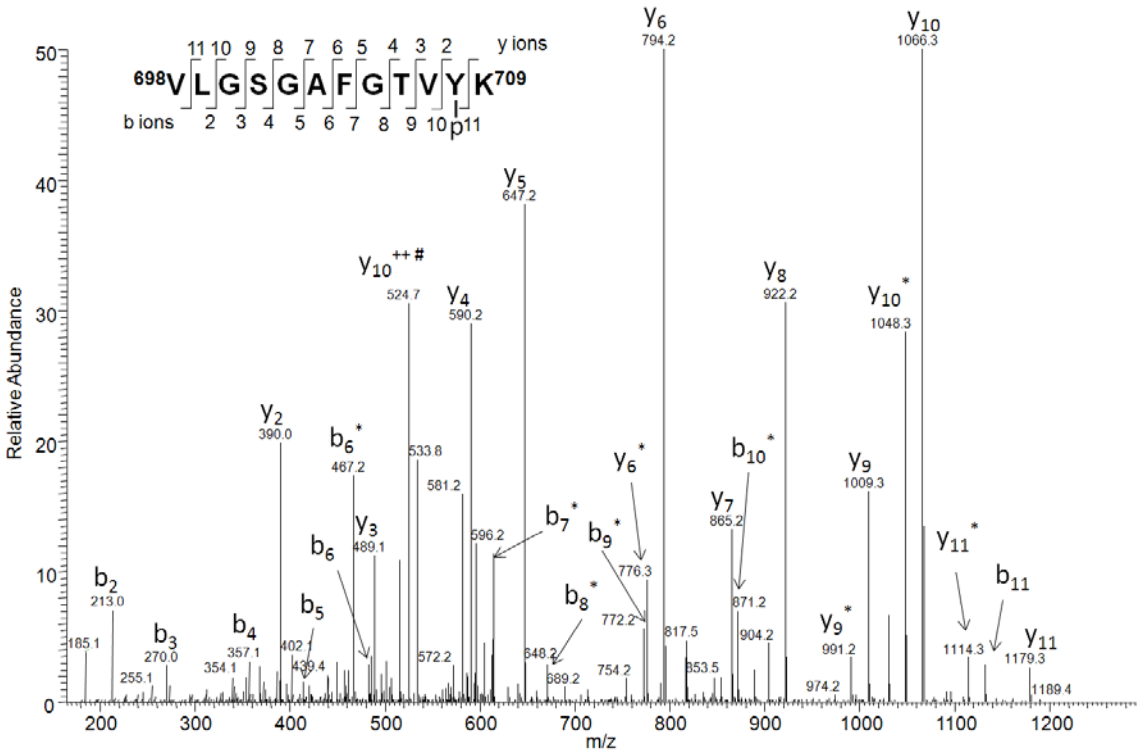


Figure 4.4a. The product ion spectrum of phosphorylated peptide 698-709. The phosphorylation site was located on Try-708. Peaks corresponding to ammonia loss are labeled with octothorpes (#), and those corresponding to water loss are labeled with asterisks (*).

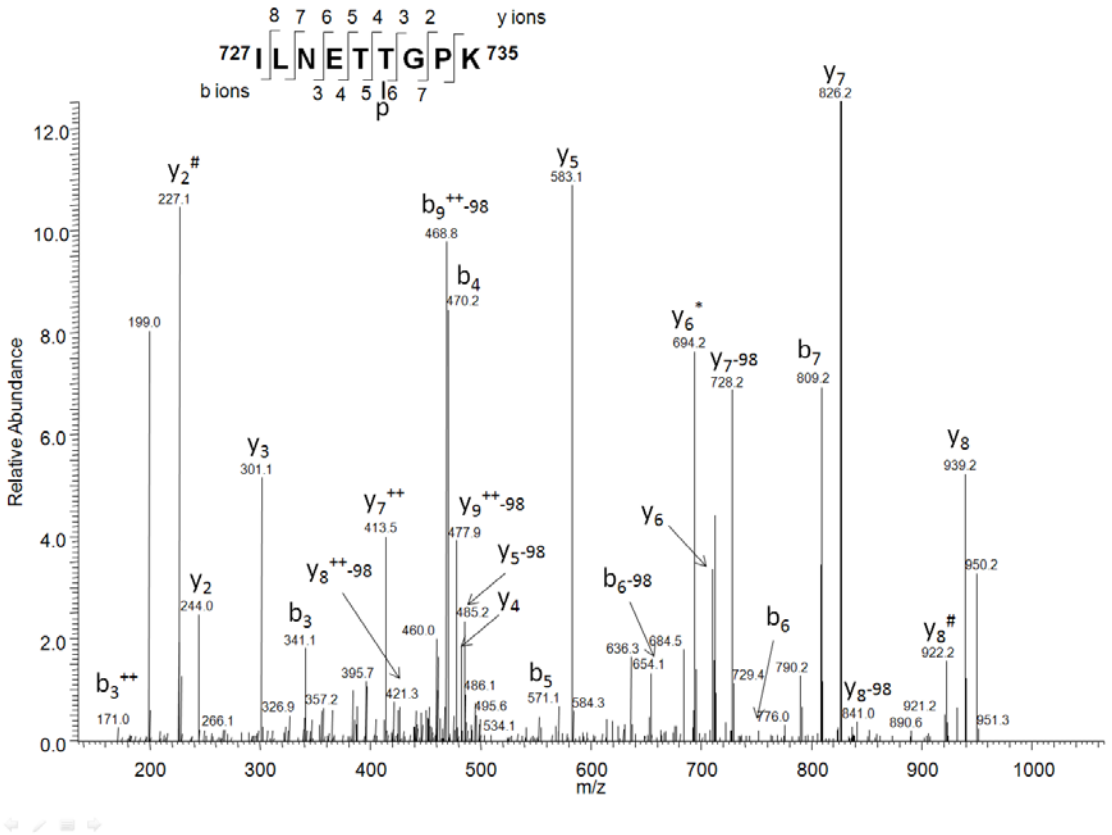


Figure 4.4b. The product ion spectrum of phosphorylated peptide 727-735. The phosphorylation site was located on Thr-732. Peaks corresponding to ammonia loss are labeled with octothorpes (#), and those corresponding to water loss are labeled with asterisks (*).

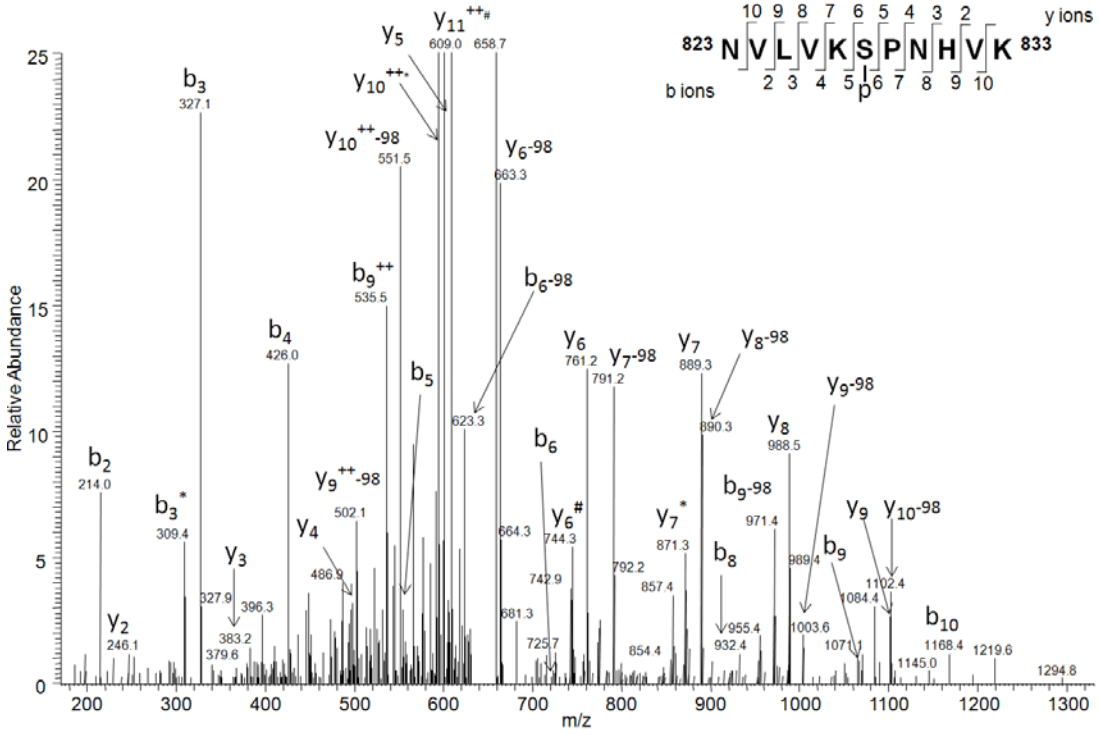


Figure 4.4c. The product ion spectrum of phosphorylated peptide 823-833. The phosphorylation site was located on Ser-828. Peaks corresponding to ammonia loss are labeled with octothorpes (#), and those corresponding to water loss are labeled with asterisks (*).

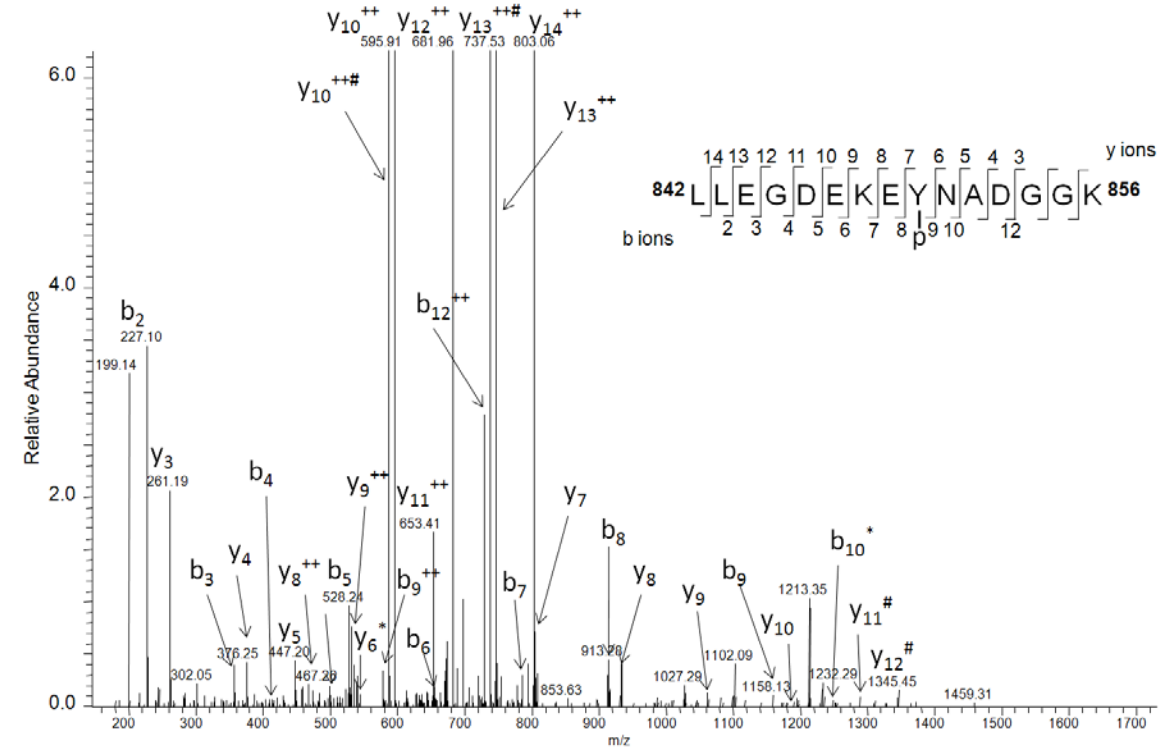


Figure 4.4d. The product ion spectrum of phosphorylated peptide 842-856. The phosphorylation site was located on Tyr-850. Peaks corresponding to ammonia loss are labeled with octothorpes (#), and those corresponding to water loss are labeled with asterisks (*).

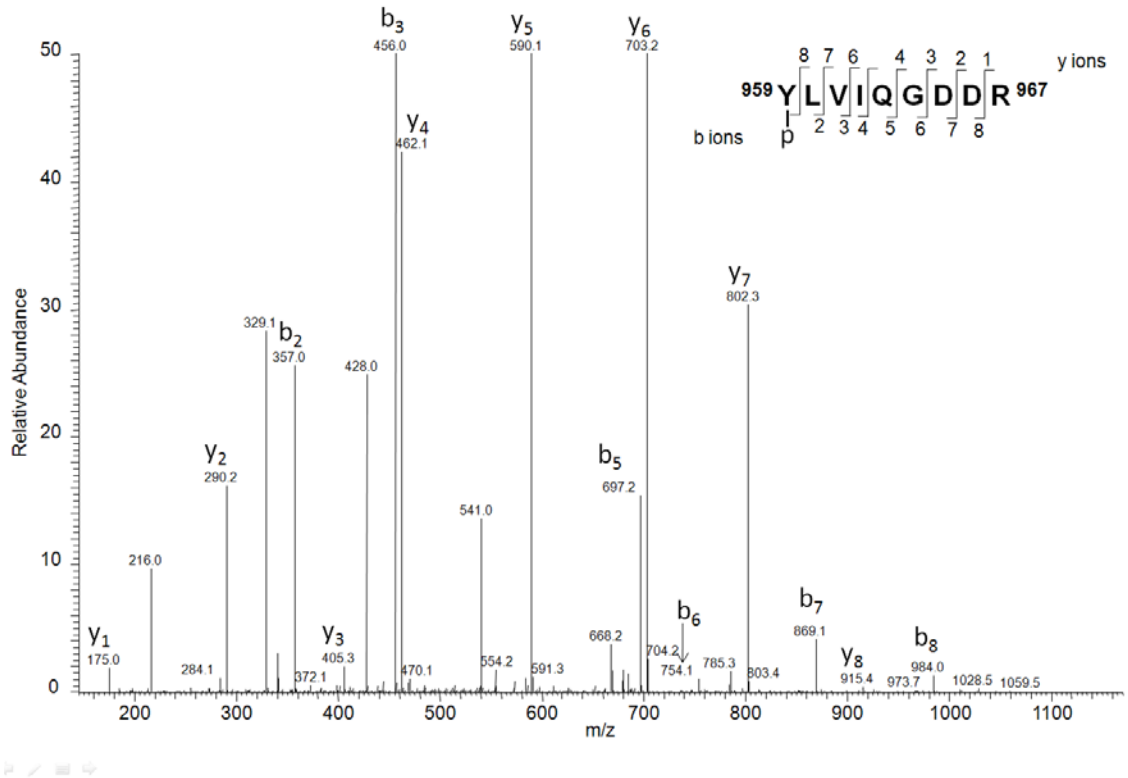


Figure 4.4e. The product ion spectrum of phosphorylated peptide 959-967. The phosphorylation site was located on Tyr-959. Peaks corresponding to ammonia loss are labeled with octothorpes (#), and those corresponding to water loss are labeled with asterisks (*).

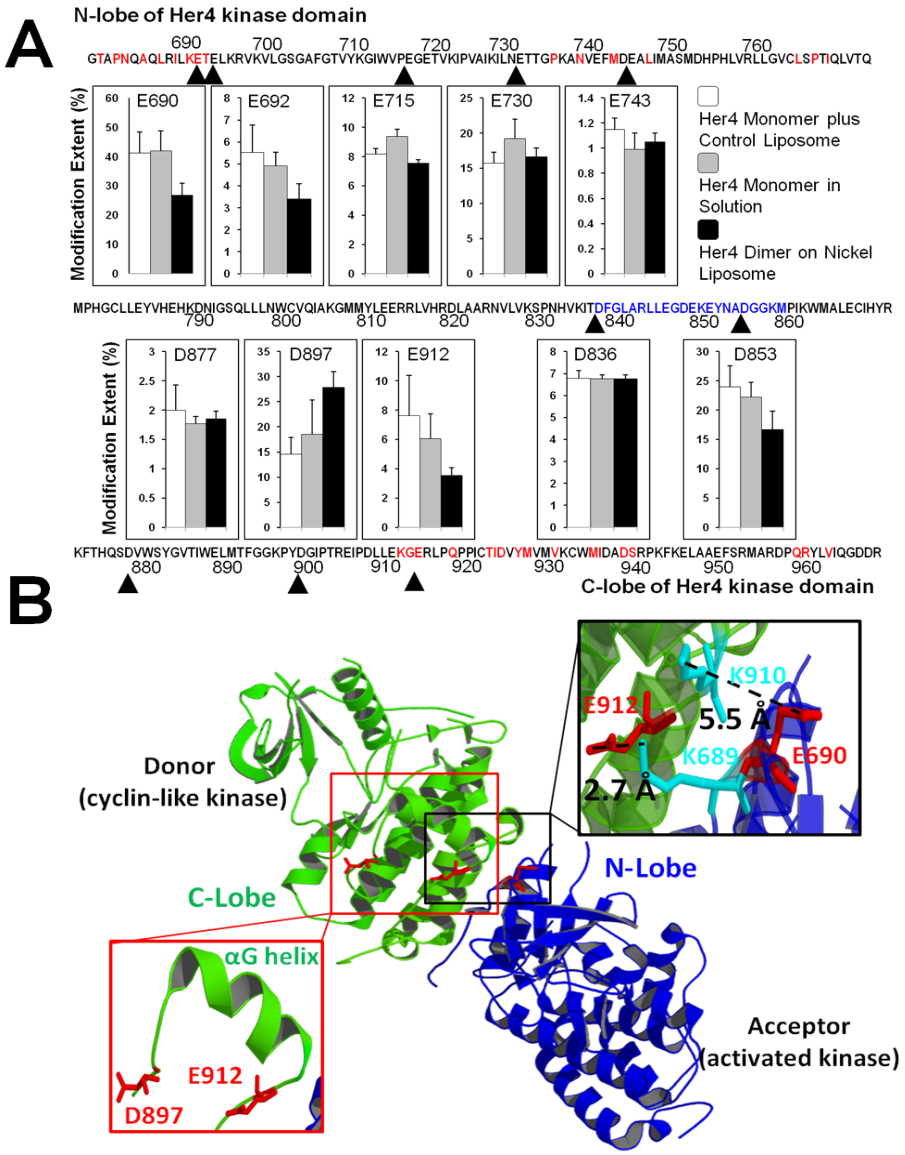


Figure 4.5. Carboxyl group modification extents: (A) Carboxyl group modification extents for Her4 kinase domain protein. Her4 residue numbering is based on the mature, full length protein minus its signal peptide and matches the numbering used in Qiu et al. The residues involved in dimer interface in crystal structure are labeled in red. The residues from activation loop are labeled in blue. Data represent mean and standard deviation of three independent samples. (B) Her4 kinase domain asymmetry dimer. Based on carboxyl group footprinting results, two residues E690 and E912 are labeled in red. (PDB ID: 3BCE).

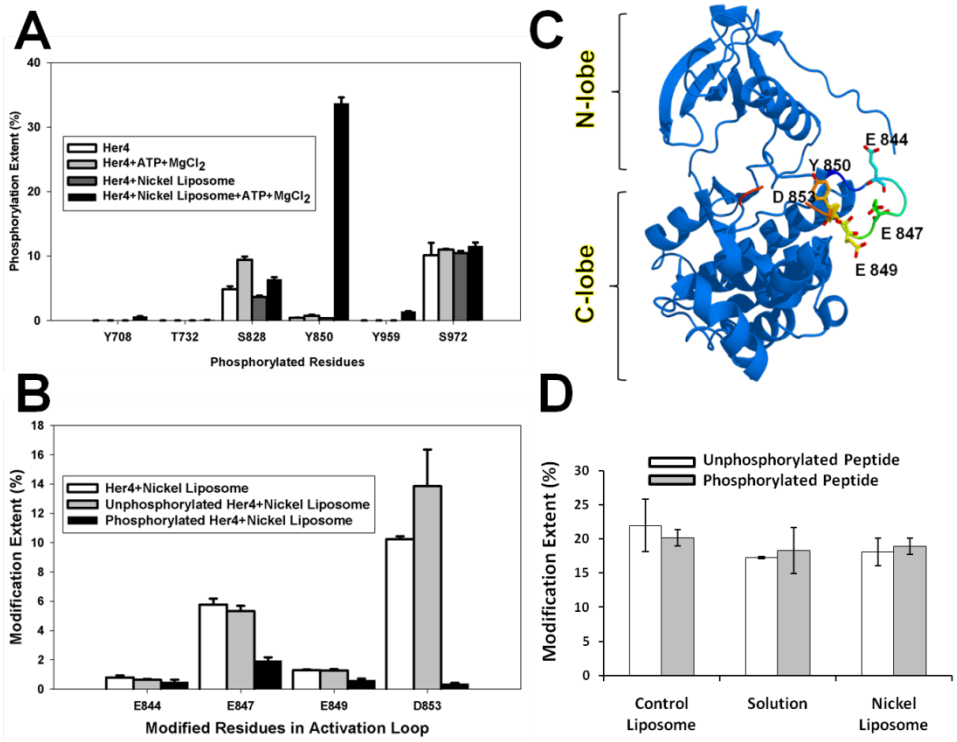


Figure 4.6. Phosphorylation and carboxyl group modification in activation loop region: (A) Normalized phosphorylation levels based on MS signals corresponding to all six phosphorylated residues detected in LC-MS/MS experiment. (B) Carboxyl group modification extents of residues in activation loop. A total of four modified residues were detected by LC-MS/MS. The distinction between modification of unphosphorylated tryptic peptide and phosphorylated tryptic peptide was based on mass difference and product-ion spectrum of each peptide ion. (C) Activation loop in Her4 kinase domain (crystal structure, PDB id: 3BCE). The loop is labeled in rainbow colors. Three Glu residues and the phosphorylated residue Tyr-850 are highlighted in stick mode. The Asp residue Asp-853 is not seen in the crystal structure. (D) Phosphorylation effects on carboxyl group modification. Two synthetic peptides (unphosphorylated vs.

phosphorylated form) with same sequence as the Tyr-850-containing tryptic peptide were used as a control to check the effects of phosphorylation on carboxyl group modification.

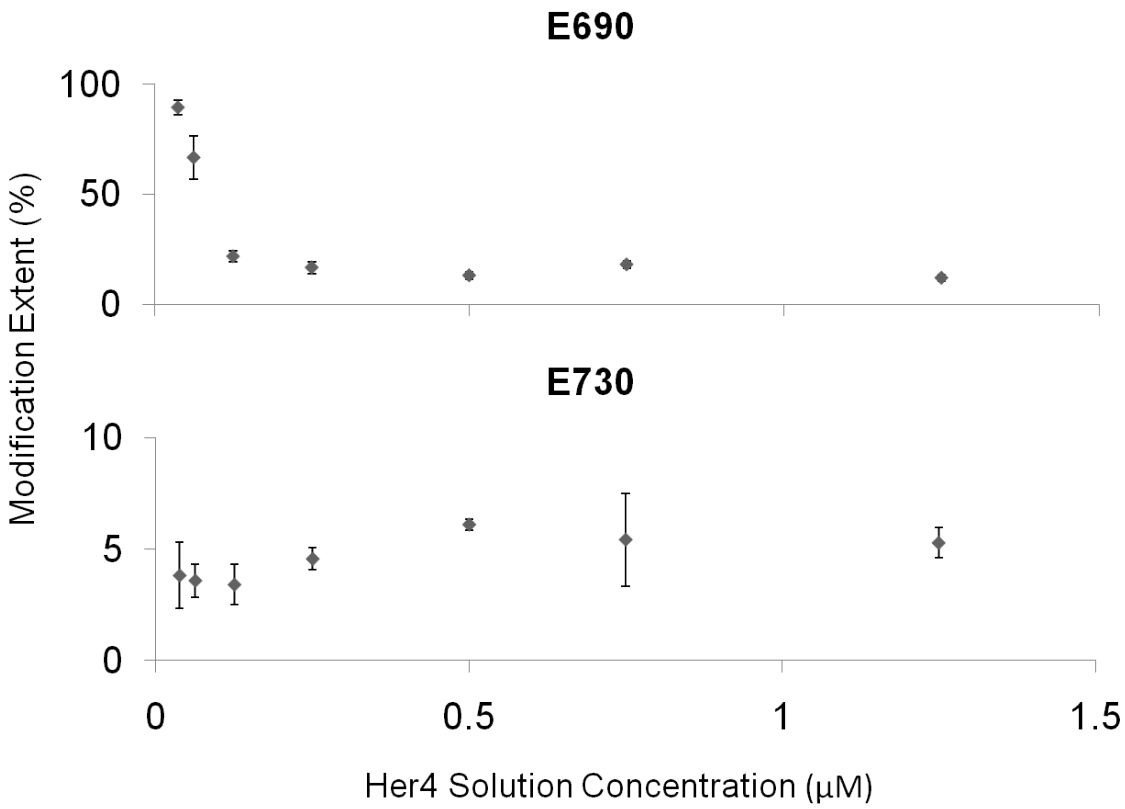


Figure 4.7. Titration experiment results of Glu-690 and Glu-730. The Her4 kinase domain protein solution concentration was listed on X axis. The black dots are modification extents on different protein concentrations.

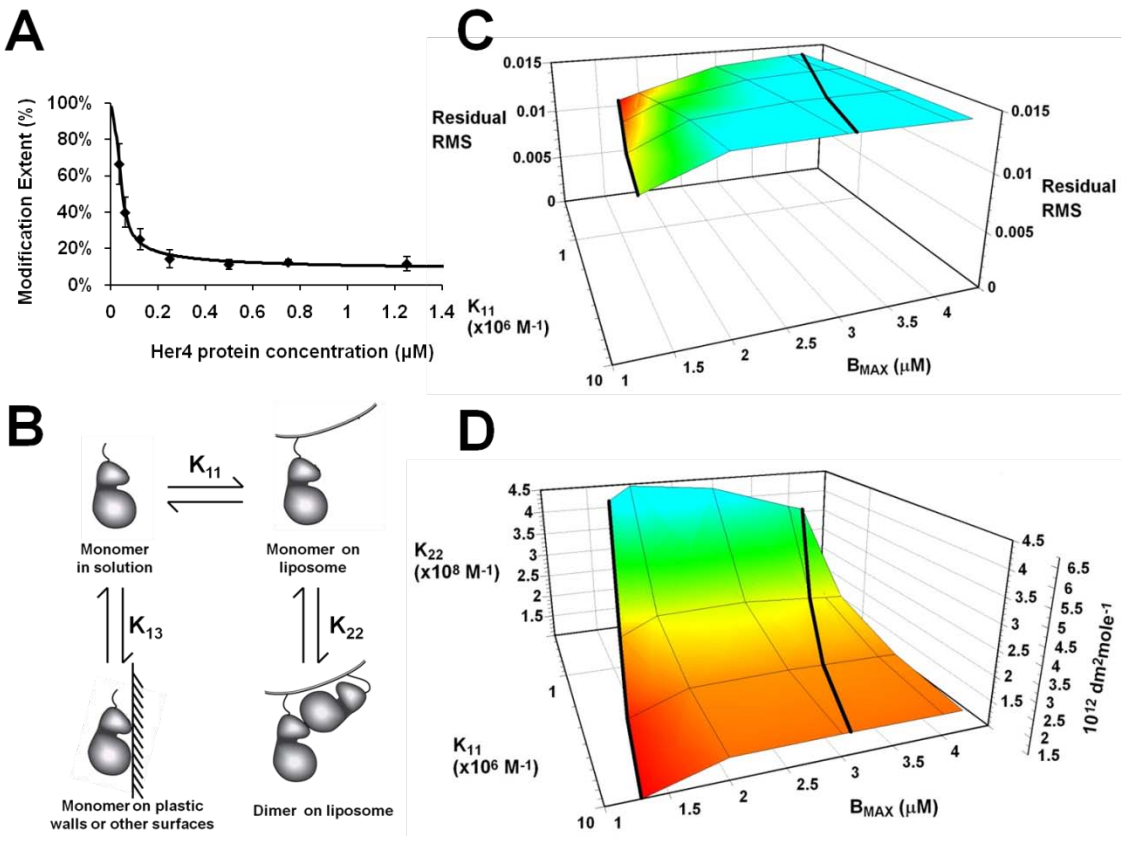


Figure 4.8. Model of Her4 dimerizaiton: (A) Titration curve of Asp-847. The Her4 kinase domain protein solution concentration was listed on X axis. The black dots are modification extents of Asp-847 on different protein concentrations. Asp-847 titration curve modeling by Her4 dimer on Ni liposome with parasitic binding site model. Black line is the curve generated by the best fit model. (B) Her4 kinase domain exists in four possible states: free monomer in solution, monomer on liposome surface, dimers on liposome surface, and binding of Her4 to plastic walls or other surfaces. (C) The root mean square of the residuals from the model fits as a function of monomer-to-liposome association constant (K_{11}) and liposome binding capacity (B_{max}). Results for B_{max} equal to 1×10^{-6} M are not shown because the model curves clearly show bias and the RMS is greater than 0.034 for these fits. The association constants K_{11} are given in M^{-1}. (D) Her4 dimer association equilibrium constant (K_{22}) vs monomer-to-liposome association constant (K_{11})

and liposome binding capacity (B_{max}). Results for B_{max} equal to 1×10^{-6} M are not shown because the model curves clearly show bias. The resulting value for K_{22} varies from 4.8×10^{14} to 23×10^{8} for these fits. The association constants are given in M^{-1}.

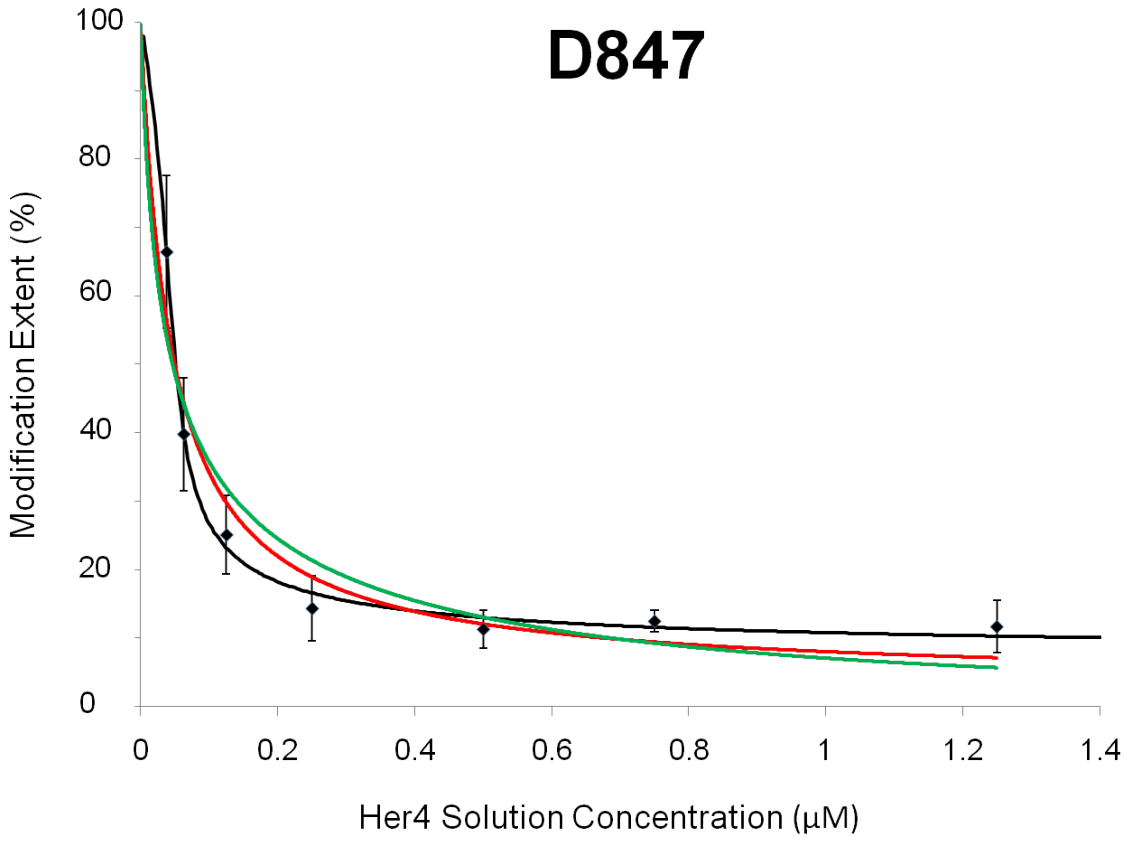


Figure 4.9. Asp-847 titration curve modeling by different Her4 dimer models. Her4 dimer on Ni liposome with parasitic binding site model (black curve). Simple Her4 dimer model (green curve) and Her4 dimer on Ni liposome model (red curve).

Table 4.1. Percentage of carboxyl group modification for Her4 kinase domain protein.

Residue	% of Modification								
	Monomer in Liposome			Monomer in Buffer			Dimer in Ni-Liposome		
E690	41.25%	±	7.13%	42.01%	±	6.69%	26.82%	±	4.15%
E692	5.52%	±	1.25%	4.90%	±	0.63%	3.40%	±	0.69%
E715	8.16%	±	0.41%	9.36%	±	0.50%	7.54%	±	0.25%
E717	9.03%	±	0.93%	11.73%	±	1.15%	10.82%	±	0.95%
E730	15.76%	±	1.50%	19.19%	±	2.83%	16.64%	±	1.22%
E739	4.00%	±	0.44%	3.17%	±	0.32%	3.56%	±	0.14%
D742	0.27%	±	0.04%	0.26%	±	0.03%	0.23%	±	0.03%
E743	1.15%	±	0.09%	0.99%	±	0.13%	1.05%	±	0.07%
D751	10.39%	±	1.67%	8.45%	±	0.09%	8.65%	±	0.75%
E810	0.71%	±	0.02%	0.63%	±	0.07%	0.74%	±	0.04%
E811	0.08%	±	0.02%	0.10%	±	0.03%	0.17%	±	0.06%
D836	6.79%	±	0.35%	6.77%	±	0.16%	6.76%	±	0.17%
E844	0.19%	±	0.03%	0.13%	±	0.02%	0.23%	±	0.04%
D846	5.88%	±	0.37%	5.81%	±	0.82%	5.46%	±	0.28%
E847	7.96%	±	0.58%	7.60%	±	0.76%	7.51%	±	0.57%
E849	1.04%	±	0.03%	1.12%	±	0.11%	1.00%	±	0.07%
D853	23.96%	±	3.63%	22.20%	±	2.58%	16.66%	±	3.13%
E865	3.97%	±	0.21%	4.94%	±	0.51%	4.34%	±	0.09%
D877	2.00%	±	0.43%	1.77%	±	0.12%	1.85%	±	0.13%
E887	1.25%	±	0.40%	0.19%	±	0.13%	0.04%	±	0.01%
D897	14.56%	±	3.34%	18.53%	±	6.85%	27.91%	±	3.05%
E903	0.27%	±	0.07%	0.22%	±	0.05%	0.52%	±	0.10%
D906	0.09%	±	0.01%	0.05%	±	0.01%	0.11%	±	0.02%
E909	2.47%	±	0.24%	2.00%	±	0.09%	2.41%	±	0.23%
E912	7.63%	±	2.73%	6.04%	±	1.71%	3.53%	±	0.55%
D923	0.08%	±	0.03%	0.17%	±	0.01%	0.19%	±	0.04%
D935	24.69%	±	9.76%	31.94%	±	3.94%	30.05%	±	4.93%
D937	11.70%	±	6.10%	10.79%	±	3.00%	17.36%	±	2.29%
E944	3.95%	±	0.03%	4.15%	±	0.44%	4.03%	±	0.05%
E948	0.24%	±	0.04%	0.15%	±	0.02%	0.17%	±	0.03%
D965	0.19%	±	0.01%	0.14%	±	0.02%	0.19%	±	0.01%
D966	1.00%	±	0.07%	1.02%	±	0.03%	1.28%	±	0.08%
D975	3.11%	±	0.05%	3.01%	±	0.03%	3.06%	±	0.08%
D984	10.84%	±	0.79%	15.67%	±	4.76%	15.87%	±	4.08%
E989	1.09%	±	0.11%	1.00%	±	0.31%	0.88%	±	0.37%
D990	2.30%	±	0.77%	2.46%	±	0.39%	2.70%	±	1.12%
D993	7.90%	±	2.21%	11.51%	±	0.90%	9.51%	±	0.04%

Table 4.2. LC-MS results of modified residues in the activation loop

Peptide	Sequence	MS2 ID'd Modification*	Theoretical Mass	Mass Centroid	Error (ppm)	Net Mass Change	Peak Retention Time (min)
842-848	LLEGDEK	E847	859.429	859.429	-0.1	57.0215	16.51
842-848	LLEGDEK	E847	887.46	887.459	-0.8	85.0528	21.56
842-856	LLEGDEKEYNADGGK	E847	1693.78	1693.78	-1.1	57.0215	20.94
842-856	LLEGDEKEYNADGGK	E844	1721.81	1721.81	-2.6	85.0528	23.69
849-856	EYNADGGK	E849	909.383	909.381	-1.5	57.0214	10.4
849-856	EYNADGGK	E849	937.414	937.412	-2.2	85.0527	17.58
849-860	EYNADGGKMPIK	M857	1337.63	1337.63	-0.7	15.9949	16.59
849-860	EYNADGGKMPIK	D853, M857	1394.65	1394.65	-0.4	73.0164	16.6
849-860	EYNADGGKMPIK	D853	1378.66	1378.65	-1.5	57.0215	21.06

*Note: +15.9949 Da, methionine oxidation; +57.0215 Da, ester hydrolysis product of carboxyl group footprinting reaction; + 73.0164 Da, methionine oxidation and carboxyl group modification, +85.0527 Da, ethyl ester product of carboxyl group footprinting reaction. The unmodified peptide peak usually has similar retention time with the +57.0215 Da modified peptide.

Chapter 5

Mass Spectrometry-Based Studies of Protein Assemblies

Abstract

Most functional units in biology are high-order assemblies formed of proteins(*1*). These assemblies have become a major target in structural biology. Besides traditional tools for protein structural studies, X-ray crystallography and NMR spectroscopy, mass spectrometry (MS) is becoming a complementary tool that can interrogate protein assemblies in their near native states(*2*). This is a remarkable development given that the gas phase is so different than solution. Native ESI of protein assemblies has showed its ability within last two decades to provide structural insights that complement those from traditional structural biology approaches(*3*). In this chapter, the details of native ESI of protein assemblies and recent advances in the area of protein assemblies are reviewed. As one of major approaches in probing protein assemblies in the gas phase, tandem mass spectrometry may become a powerful tool in obtaining structural information.

Mass Spectrometry of Protein Assemblies

Protein assemblies

Proteins are of importance in almost all cellular processes, from DNA replication to protein synthesis and degradation(*4*). Significant structural-biology efforts are ongoing to study the relationship between protein structure and function. Furthermore, many cellular processes are performed by large protein complexes or assemblies(*5*). Structural biology is now focused on higher order structures, protein assemblies, which are composed of proteins, DNA, and cofactors. Motivated by their importance, traditional and novel structure biology methods continue to be developed and dedicated to studies of protein assemblies.

Methods for study of protein assemblies

High resolution methods (e.g., X-ray crystallography and NMR spectroscopy), low resolution methods (e.g., cryo-electron microscopy and small-angle X-ray scattering (SAXS), proteomics-based tandem affinity purification (TAP), and protein footprinting (described in former chapters) are the principal tools for investigation of protein assemblies(*1*). Among those methods, MS is becoming an important complementary technique. Although MS cannot provide high resolution data, such as that from X-ray or NMR spectroscopy, it has an advantage that protein assemblies can be interrogated in their near native state with consumption of relatively small amounts of sample.

Since the first experiments of MS-based approaches to protein assembly, more and more applications have been reported(*6*). Progress of this area has been reviewed periodically(*2, 7, 8*). A brief summary about MS-based studies of protein assemblies, from fundamental MS to critical applications, will be presented here.

Generating Ions of Protein Assemblies

The first step in an MS-based study of protein assemblies is generating ions of the protein assemblies. It was not possible before the 1980s to generate ions in any routine way from nonvolatile and labile biomolecules much less from biomolecule assemblies. Traditional ionization methods, like EI and CI, are suitable for volatile small molecules and Fast Atom Bombardment (FAB) are suitable for small nonvolatile molecules(*9*), but are inappropriate for analyzing nonvolatile and large biomolecules. After the introduction of the new ionization methods, electrospray ionization (ESI) and matrix assistant laser desorption ionization (MALDI), MS became a more important method for

maromolecular studies owing to its sensitivity and tolerance of sample heterogeneity. Proteins are analyzed in solution (ESI) or as solids formed by co-crystallization with UV absorbing organic acids (MALDI). Successful studies of protein assemblies were reported when using both ionization methods(*10*). For example, ESI MS can analyze mega-Dalton protein within a modest *m/z* range by generating multiply charged protein ions. In fact, most of the reports of MS-based studies of protein assemblies come from the ESI experiments(*3*). For MALDI, preserving non-covalent interaction of protein assemblies in solid phase is problematic, and the singly charged large protein ions have *m/z* values usually beyond the optimum range of most commercial instruments except for time-of-flight spectrometers. These factors make MALDI less-ideal for study of protein complex than is ESI(*3*). In this thesis, we focus on the application of ESI in studies of protein assemblies.

ESI and multiply charged protein ions

A standard ESI setup involves flowing an analyte solution to the end of capillary held of high electrical potential (Figure 5.1). A parallel gas flow surrounds the capillary to aid the nebulization of emerging analyte solution. Because the electrical potential is high, positive ions (in the positive-ion mode of ESI) accumulate at the tip of the solution, causing the solution to form a “Taylor cone”. The steam of solution is drawn out as small charged droplets that move away each other owing to electrostatic repulsion of the excess positive (or negative) ions within each droplet. Solvent evaporation continually reduces the size of charged droplet until the Coulombic repulsions between crowded positive ions overcome the droplet surface tension (called the Rayleigh limit). Droplets undergo fission

and form even smaller droplets. Evaporation and fission are repeated several cycles until gas-phase ions are produced(*11*).

Generation of multiply charged protein ions is the usual outcome of ESI. Although a general mechanism for ESI is well accepted, a detailed mechanism for generation of protein multiple charged ions remains unclear. Two models have been developed to explain the mechanism of ESI(*12, 13*): one is the charge-residue model (CRM); the other is the ion-evaporation model (IEM). In the CRM, evaporation and fission continue until a single molecule of analyte is left in the droplet. Complete evaporation results in a charged analyte ion. Based on the CRM model, the final charge state of protein ions can be predicted by the charge at the Rayleigh limit of the droplet. According to the IEM model, repulsions between analyte ions with other charged ions overcome the solvent force and analyte ions are ejected directly from the droplet (desorbed by the high electric field surrounding the droplets). Based on the IEM model, the charge state of analyte ion is a function of analyte properties. Previous experimental evidences with small molecules are in agreement with the IEM model, whereas ESI of large molecules (e.g., proteins) can be explained well by the CRM model. Recently, a revised version of the charged residue field emission model was developed to interpret the native state protein charge distribution in ESI(*14, 15*). Multiply charged droplets are produced by ESI. The excess charges in each droplet are carried by small ionic species or clusters. The final charge state of protein is determined by emission of small ionic species or clusters from ESI droplet prior to complete solvent evaporation. The experimental observations of different size of proteins in their near-native states were examined.

Results show that charge states of protein ions are in good agreement with predicted values on the basis of charge carrier field emission [15].

Native ESI of protein assemblies

There are several factors including solvent pH, concentration of non-volatile salts, and percentage of organic solvent that affect protein signal in ESI. Using a typical ESI solvent (50% Acetonitrile, 50% Water, 0.1% Formic Acid, pH = 1~2.), we find the majority of proteins are in a denatured state, here the proteins are largely unfolded and have exposed all their basic side chains to solvent. The ESI spectra of denatured proteins shows a broad charge state distributions (charge envelop) centered at low *m/z* (Figure 5.2). Non-covalent interactions between proteins and ligands are usually destroyed in their denatured state, therefore, preserving a non-covalent protein complex is a problem when using a typical ESI solvent. Such solvents are not particularly useful for exploring protein-ligand interactions.

To preserve the native state and noncovalent interactions of a protein, it must be in aqueous solution at physiological pH and appropriate physiological ionic strength. An ESI approach called native ESI (or native MS) was developed to meet these requirements(*2*). Proteins in a volatile aqueous buffer or a salt solution (e.g., ammonium acetate) are directly sprayed by ESI. Native-like proteins carry fewer charges than denatured or unfolded states owing to the fewer exposed basic residues in the folded form. As a consequence, one achieves a narrow spread of the charge envelop for protein ions; the *m/z* of the protein in this case falls at a higher value (Figure 5.3). Non-covalent

interactions are thus preserved. Native ESI has become a suitable approach for studies of protein assemblies in their near-native states.

Another advance in ESI that benefits studies of protein assemblies is nano ESI (nESI)(*16*). nESI generates smaller droplets than does regular ESI by spraying the sample through a capillary with a smaller diameter (e.g., ~ 1 μm inner diameter) than normally used for ESI. Ionization conditions (spray voltage, capillary temperature, nebulizer gas flow) are more gentle for evaporating smaller droplets than for the larger drops of conventional ESI. nESI also dramatically reduces the amount of protein sample (e.g., from ~50 to 1-3 μL) required for MS analysis. More importantly, nESI has relative high tolerance for nonvolatile salts that are often required to maintain the native state of a protein assembly.

In native ESI, incomplete desolvation can cause problems in determining the molecular weight of a protein(*17*). In fact, the observed mass from native ESI usually is higher than the calculated mass based on the protein sequence(*18*). Because desolvation is incomplete, extra solvent molecules or other small MW substances in the solution can reside in the final droplet with the protein ions. These species associate with the protein, resulting in broad protein peaks in the native ESI mass spectrum. Desolvation can be improved by applying collision energy in the source or collision-cell regions of the mass spectrometer. The collision energy used in "cleaning up" native ESI mass spectra needs to be carefully tuned to avoid complex unfolding and dissociation. Practically, the most accurate mass assignments for protein assemblies come from the spectra acquired under condition just below the dissociation energy of the assemblies.

Another issue of native ESI is non-specific oligomerization(*19*). Non-native oligomers may form in native ESI; on the basis of the CRM model, high-order oligomers can form from association of two or more complexes that are present in the same droplet. Because the interactions of these non-specific oligomers are weak and concentration dependent, they can be overcome by applying collision energy in the source or in the collision-cell region or reducing the protein concentration. Alternatively, they can be recognized by data processing, as reported by Robinson and co-workers(*20*), and removed "computationally." The number of non-specific oligomers can be deduced through predicting the occupancy of the droplets in ESI. A model of the occupancy droplets was developed by incorporating the predicting into a Monte Carlo simulation. This droplet occupancy model was trained and validated by several protein complexes and applied in study the aggregation of amyloid-related protein transthyretin (TTR). The specific WT TTR tetramer and nonspecific higher order oligomers of L55P variant form of TTR were differentiated by the droplet occupancy model.

Transmitting and Analyzing Ions of Protein Assemblies

After successfully generating ions of protein assemblies, transmitting and mass analyzing these ions become a challenge for MS. Since the first MS-based study of protein assembly nearly 20 y ago, advances in both method and instrumentation have accelerated the study of protein assemblies. One major advance is the "collisional focusing" (or "collisional cooling") for transmitting large protein complex ions in the source region (ion-guide region). It is the first breakthrough that demonstrated that increasing pressure of the source region in the first vacuum stage resulted in signal improvement of high *m/z* ions(*21*). Systematic studies showed that increasing low

frequency collisions between high kinetic-energy protein ions and relatively small MW gas molecules can focus ions and improve ion transmission in the source region.

A number of instrument modifications have also been made, and they have benefited studies of protein assemblies(*22-24*). Similar "collisional focusing" effects are used in the later stages of mass spectrometers (quadrupole and collision cell regions). Carefully adjusting the pressure in mass spectrometer during ion transmission becomes an essential step for MS-based studies of protein assemblies.

High *m/z* protein-complex ions are usually observed at a high *m/z* range, which is beyond the range of most commercial mass spectrometers (maximal *m/z* 3000-4000). For example, protein or protein complexes with molecular weight over 60 KDa may form ions having an *m/z* greater than 4000. One can make use of the quadrupole analyzer, which is a major ion transmission component in hybrid mass spectrometers, to transmit the ions. When a quadrupole is operating in the RF-only mode, ions can be focused by RF frequency and transmitted to the detector. The maximal *m/z* is determined by the principles of a quadrupole: three parameters, RF amplitude, inner radius of the quadrupole, and RF frequency, can be adjusted to increase its maximal *m/z*. In practice, only the RF frequency is reduced to fulfill this purpose. Reports(*25, 26*) of Q-TOF instruments equipped with low-frequency quadrupoles described extended mass ranges up to *m/z* 22,000.

Analyzing Ions of Protein Assemblies

Two major mass analyzers can satisfy the analysis of high *m/z* ions of protein complexes: time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FTICR).

The TOF mass analyzer has a relatively simple design and high sensitivity. The majority of protein-assembly studies have been conducted with hybrid quadrupole time-of-flight (Q-TOF) mass spectrometers because these instruments can transmit and detect large protein complexes of MDa masses. The mass measurement accuracy that pertains to ions from protein complexes is more dependent on complete ion desolvation than on instrument mass resolving power. Even the high resolving power provided by FTICR MS is less important than the quality of ionization in measuring the mass of intact protein complexes. Advantages of FTICR MS, however, also include a variety of dissociation approaches, and these have important applications for the study of protein assemblies. The applications of FTICR MS and ECD as a top-down approach in studies of protein assemblies will be addressed in the following chapter.

Ion-Mobility Measurement

Another fast-growing area for protein-assembly studies is one that emerges from the coupling of ion mobility (IM) with MS(*27*). In IM, ions from ion complexes are injected into a region containing neutral gas molecules (gas pressure 100-1000 mbar) and having an electric field. Driven by this electric field (~ 10-30 V/cm), ions can be separated based on their shape. Large ions experience more collisions and take more time to arrive at the detector than do smaller ions or ones of small cross sections. The collision cross sections (CCS) of ions can be calculated based on their drift times and charge state, which are measured in the combined instrument. These provide structural insights as to the size and shape of protein assemblies. IM measurement, thus, has the potential to be an important extension of MS-based studies of protein assemblies. In the early applications,

IM has provided critical evidence that protein native or nearly native structures can be preserved in the gas phase.

Solution-Phase Properties of Protein Assemblies

Stoichiometry of Protein Assemblies

Studies of the functional unit formed by oligmerization of proteins is a crucial part of contemporary biology. One major application of native ESI is the determination of stoichiometry of protein assemblies; that is, the number and nature of subunits. Stoichiometry information of large protein complexes (e.g., the chaperone complex GroEL(*28*) or hemoglobin complexes with variable compositions(*29*), were already reported by various research groups.

In the typical work flow of protein-assembly studies, the identification of a subunit still relies heavily on a proteomics-based, bottom-up LC/MS experiment(*30*). Protein complexes are proteolytically digested, and peptide mixtures are analyzed by LC/MS to identify the subunits. Sequence information of each subunit is obtained from partial MS-based sequencing followed by database searching of peptide product-ion spectra that give the sequence information. The stoichiometry of the protein assembly is deduced from results of native ESI MW measurements and LC-MS experiments. The intact mass of protein assembly is obtained by native ESI. Subunits from the protein assembly are identified by LC/MS. Integrating information of the intact mass of protein assembly and the subunit mass can provide the stoichiometry of the protein assembly.

Subunits that comprise the protein complex can be released by adjusting the ion-source temperature or introducing denaturants into sample solution. The molecular

weight of subunit released from protein complex is measured by the mass spectrometer. Temperature variations can also initiate the dissociation of the complex, and these experiments can tell us about the thermodynamic stability of complexes.(*31*) Robinson and co-workers (*32*) reported, using a modified ESI source, the stability of heat-shock proteins. Alternatively, introducing denaturants (urea, organic solvents, and acids) can release various subunits from the protein complex(*33*). Analyzing samples submitted to gradually increased amounts of denaturants allows subunit molecular weights to be measured. However, this approach requires a number of steps in sample preparation and analysis, and, thus, can consume considerable protein sample. If we can directly dissociate protein complexes by tandem mass spectrometry, we would have a promising approach for studies of protein complexes(*34*, *35*).

Gas-Phase Properties of Protein Assemblies

Because native ESI probes protein assemblies in the gas phase, one must consider the relationship between a gas-phase and a solution structure. Given that water molecules stabilize protein structure in solution, the ongoing argument has been whether there is any similarity between the native-state protein structure in solution and its structure in the gas phase. In another words, the question is whether noncovalent interactions that join together the protein subunits can be preserved in the gas phase. There are two noncovalent interactions involved in protein assemblies: hydrophobic and electrostatic interactions. In theory, the electrostatic interactions are strengthened in the gas phase, in the absence of solvent, whereas hydrophobic interactions are weakened by dehydration and introduction into the gas phase.

IM measurements can provide critical structural insights on protein assemblies in the gas phase. For example the protein assembly, trp RNA binding protein, forms a ring structure with 11 subunits. Evidence from IM measurements indicate that the ring structures are preserved in the absence of bulk solvent(*36*). IM shows that large protein assemblies can structurally survive in the gas phase on the time scale of an MS experiment. In another example, important evidence comes from the measurement of hydrophobic interaction kinetics in the gas phase by black-body infrared dissociation (BIRD). An example deals with the hydrophobic interaction between beta lactoglobin and its ligand (a fatty acid)(*37*). The kinetic data of this protein-ligand interaction show that hydrophobic interactions are preserved in the gas phase.

Outcomes from various experiments indicate strongly that the some solution-phase interactions as well as some protein near-native structures can be preserved in the gas phase. That evidence serves as a growing foundation that native ESI of protein assemblies containing multiple hydrophobic interactions (e.g., membrane proteins) can successfully be introduced to the gas phase without major perturbation of structure. For example, one might expect that an assembly arrangement would be preserved even though some remodeling of the structures of individual protein subunits occurs.

For native ESI of protein assemblies, membrane protein assemblies are the biggest challenge owing to their hydrophobicity. Robinson and co-workers(*38*) first examined the native ESI of a membrane protein assembly by using micelles to protect the complex during its introduction to the gas phase. The micelles, formed by detergents, were removed by collisional activation by using small inert gas molecules as collision

partners, allowing the stoichiometry to be determined and structural insights of membrane complex to be gained.

Tandem Mass Spectrometry of Protein Assemblies

Once a protein assembly has been introduced to the gas phase by native ESI, dissociation of the assembly can be achieved by tandem MS(*34*). Protein complexes gain internal energy during collisions with inert gas in the source or collision-cell region. When the overall internal energy is sufficient to break the non-covalent interaction, the protein complex releases some subunits. The process of dissociation is controlled in native ESI by adjusting the collision energy. An early example of collision-induced dissociation (CID) in native ESI studies is the work of Smith and co-workers(*39*) who reported the release of a single monomer from the tetrameric concanavalin A. Interestingly, they learned that the charge partitioning between ejected monomer and the rest of complex is asymmetric based on the charge per mass. Several other groups reported the dissociation of protein complexes into highly charged monomers and relatively less charged oligomers missing one subunit (Figure 5.4)(*34*). Systematic studies of structure and charge effects on protein assembly dissociation are now being conducted. Recently, new evidence shows that the loss of a high charged monomer is not the only pathway of dissociation of protein assemblies(*40*).

Although the majority of tandem MS experiments are conducted by using CID, new dissociation techniques, including BIRD and SID, are becoming more important. Slow heating a protein complex trapped in FTICR cell by absorption of blackbody photons (via BIRD) is a unique approach in studies of protein-ligand interactions, as

shown in the work of Klassen and co-workers(*41*) . The temperature-dependent rate constant can be extracted from the dissociation data. Results from a study of fatty-acid binding protein and its ligand provide critical evidence for the preservation of hydrophobic interactions in the gas phase.

Wysocki and co-workers(*42*) first reported an extension of CID, surface induced dissociation (SID). Compared with CID, SID provides a large mass collision partner (a solid surface instead of small inert gas molecule) for protein assemblies, thus increasing the center-of-mass energy for a given acceleration potential. Interestingly, symmetric charge partitioning occurs in the SID of the protein assembly. More structural insight of protein assemblies should now become available by integrating fragmentation behaviors from different dissociation approaches.

Scope of the following chapter

In tandem MS studies of protein assemblies, disruption of noncovalent interactions commonly occur giving information on connectivity of the proteins. An important goal would be to break the covalent bonds along the polypeptide chain while maintaining the noncovalent interactions between subunits. This top-down approach would afford sequence information of subunits without carrying out an independent bottom-up LC-MS experiment. The next chapters focus on the development of ECD top-down approach in studies of protein assemblies that addresses achieving this goal.

References

1. Ban, N., and Egelman, E. H. Structure and function of large cellular assemblies, *Curr Opin Struct Biol 20*, 207-209.
2. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom Rev 16*, 1-23.
3. Lorenzen, K., and Duijn, E. V. (2010) Current Protocols in Protein Science, In *UNIT 17.12 Native Mass Spectrometry as a Tool in Structural Biology* (Coligan, J. E., Dunn, B. M., Speicher, D. W., and Wingfield, P. T., Eds.), John Wiley & Sons, Inc.
4. Petsko, G. A., and Ringe, D., (Eds.) (2004) *Protein Structure and Function*, New Science Press Ltd, London.
5. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell, *Nature 450*, 973-982.
6. Ganem, B., li, Y.-T., and Henion, J. D. (1991) Observation of Noncovalent Enzyme-Substrate and Enzyme-Product Complexes by Ion-Spray Mass Spectrometry, *J Am Chem Soc*, 7818-7819.
7. Heck, A. J., and Van Den Heuvel, R. H. (2004) Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom Rev 23*, 368-389.
8. Benesch, J. L., Ruotolo, B. T., Simmons, D. A., and Robinson, C. V. (2007) Protein complexes in the gas phase: technology for structural genomics and proteomics, *Chem Rev 107*, 3544-3567.

9. Tomer, K. B. (1989) The development of fast atom bombardment combined with tandem mass spectrometry for the determination of biomolecules, *Mass Spectrom Rev 8*, 445-482.

10. Kiselar, J. G., and Downard, K. M. (2000) Preservation and detection of specific antibody--peptide complexes by matrix-assisted laser desorption ionization mass spectrometry, *J Am Soc Mass Spectrom 11*, 746-750.

11. Kebarle, P. (2000) A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, *J Mass Spectrom 35*, 804-817.

12. Cech, N. B., and Enke, C. G. (2001) Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals, *Mass Spectrom Rev 20*, 362-387.

13. Kebarle, P., and Verkerk, U. H. (2009) Electrospray: from ions in solution to ions in the gas phase, what we know now, *Mass Spectrom Rev 28*, 898-917.

14. Hogan, C. J., Jr., Carroll, J. A., Rohrs, H. W., Biswas, P., and Gross, M. L. (2008) Charge carrier field emission determines the number of charges on native state proteins in electrospray ionization, *J Am Chem Soc 130*, 6926-6927.

15. Hogan, C. J., Jr., Carroll, J. A., Rohrs, H. W., Biswas, P., and Gross, M. L. (2009) Combined charged residue-field emission model of macromolecular electrospray ionization, *Anal Chem 81*, 369-377.

16. Wilm, M., and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source, *Anal Chem 68*, 1-8.

17. El-Faramawy, A., Guo, Y., Verkerk, U., Thomson, B. A., and Siu, M. (2008) Evaluation of IR Multiphoton Dissociation as a Method for High Mass Protein

Clean Up, In *56th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, Co.

18. McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes, *J Am Chem Soc 128*, 11433-11442.

19. Sun, N., Sun, J., Kitova, E. N., and Klassen, J. S. (2009) Identifying nonspecific ligand binding in electrospray ionization mass spectrometry using the reporter molecule method, *J Am Soc Mass Spectrom 20*, 1242-1250.

20. Lane, L. A., Ruotolo, B. T., Robinson, C. V., Favrin, G., and Benesch, J. L. (2009) A Monte Carlo Approach for Assessing the Specificity of Protein Oligomers Observed in Nano-electrospray Mass Spectra, *Intl. J. Mass Spectrom. 283*, 169-177.

21. Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W., and Standing, K. G. (1998) Studies of Noncovalent Complexes in an Electrospray Ionization/time-of-flight Mass Spectrometer, *J Am Soc Mass Spectrom 9*, 569-579.

22. Rostom, A. A., and Robinson, C. V. (1999) Detection of the intact GroEL Chaperonin Assembly by Mass Spectrometry, *J Am Chem Soc 121*, 4718-4719.

23. Schmidt, A., Bahr, U., and Karas, M. (2001) Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in nano-ESI MS, *Anal Chem 73*, 6040-6046.

24. Chernushevich, I. V., and Thomson, B. A. (2004) Collisional cooling of large ions in electrospray mass spectrometry, *Anal Chem 76*, 1754-1760.

25. Sobott, F., Benesch, J. L., Vierling, E., and Robinson, C. V. (2002) Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry, *J Biol Chem 277*, 38921-38929.

26. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies, *Anal Chem 74*, 1402-1407.

27. Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J., and Robinson, C. V. (2008) Ion mobility-mass spectrometry analysis of large protein complexes, *Nat Protoc 3*, 1139-1152.

28. Rostom, A. A., and Robinson, C. V. (1999) Disassembly of intact multiprotein complexes in the gas phase, *Curr Opin Struct Biol 9*, 135-141.

29. Kuchumov, A. R., Loo, J. A., and Vinogradov, S. N. (2000) Subunit distribution of calcium-binding sites in Lumbricus terrestris hemoglobin, *J Protein Chem 19*, 139-149.

30. Zhou, M., and Robinson, C. V. When proteomics meets structural biology, *Trends Biochem Sci 35*, 522-529.

31. Geels, R. B., Calmat, S., Heck, A. J., van der Vies, S. M., and Heeren, R. M. (2008) Thermal activation of the co-chaperonins GroES and gp31 probed by mass spectrometry, *Rapid Commun Mass Spectrom 22*, 3633-3641.

32. Benesch, J. L., Sobott, F., and Robinson, C. V. (2003) Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry, *Anal Chem 75*, 2208-2214.

33. Janis, J., Pasanen, S., Rouvinen, J., and Vainiotalo, P. (2008) Characterization of the pH-dependent dissociation of a multimeric metalloprotein Streptomyces rubiginosus xylose isomerase by ESI FT-ICR mass spectrometry, *J Mass Spectrom 43*, 1376-1380.

34. Benesch, J. L. (2009) Collisional activation of protein complexes: picking up the pieces, *J Am Soc Mass Spectrom 20*, 341-348.

35. Duijn, E. V. (2010) Current limitations in native mass spectrometry based structural biology, *J Am Soc Mass Spectrom 21*, 971-978.

36. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water, *Science 310*, 1658-1661.

37. Benesch, J. L., and Robinson, C. V. (2009) Biological chemistry: Dehydrated but unharmed, *Nature 462*, 576-577.

38. Barrera, N. P., Di Bartolo, N., Booth, P. J., and Robinson, C. V. (2008) Micelles protect membrane complexes from solution to vacuum, *Science 321*, 243-246.

39. Light-Wahl, K. J., Schaedler, T. A., and Smith, R. D. (1994) Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry, *J Am Chem Soc 116*, 5271-5278.

40. Boeri Erba, E., Ruotolo, B. T., Barsky, D., and Robinson, C. V. Ion mobility-mass spectrometry reveals the influence of subunit packing and charge on the dissociation of multiprotein complexes, *Anal Chem 82*, 9702-9710.

41. Felitsyn, N., Kitova, E. N., and Klassen, J. S. (2001) Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative

dissociation. Investigating the origin of the asymmetric dissociation behavior, *Anal Chem 73*, 4647-4661.

42. Jones, C. M., Beardsley, R. L., Galhena, A. S., Dagan, S., Cheng, G., and Wysocki, V. H. (2006) Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions, *J Am Chem Soc 128*, 15044-15045.

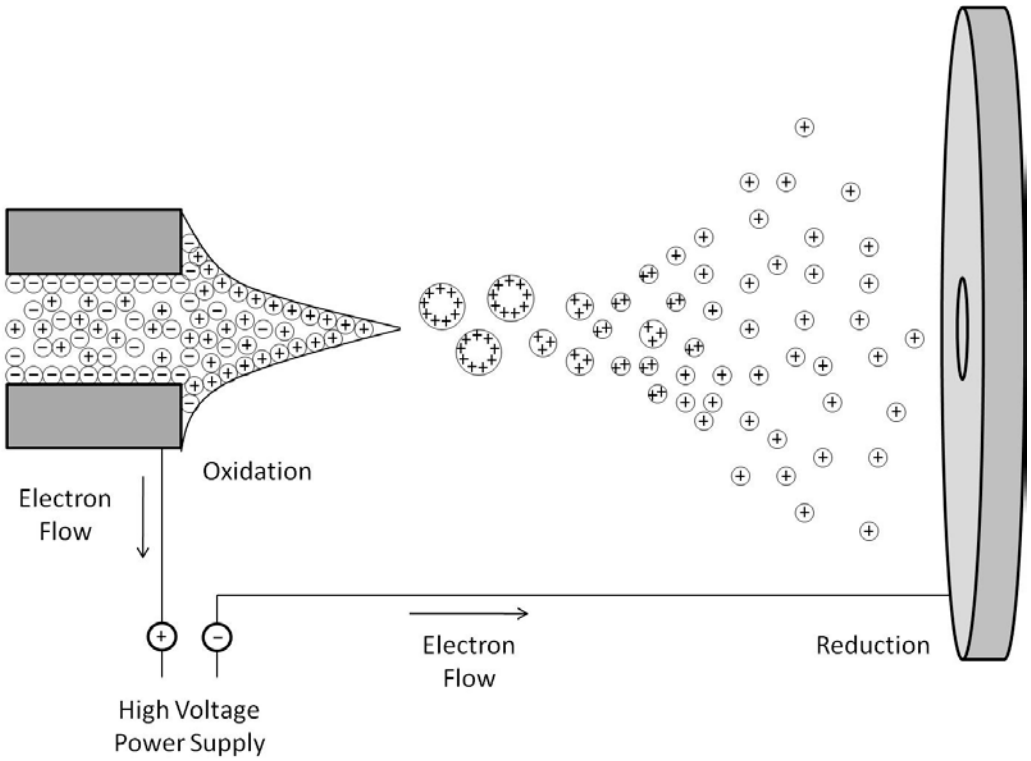


Figure 5.1. The ESI process. The analyte solution is pushed to the end of capillary held of high electrical potential. Positive ions accumulate at the tip of the solution. The steam of solution is drawn out as small charged droplets that move away each other owing to electrostatic repulsion of the excess positive ions within each droplet.

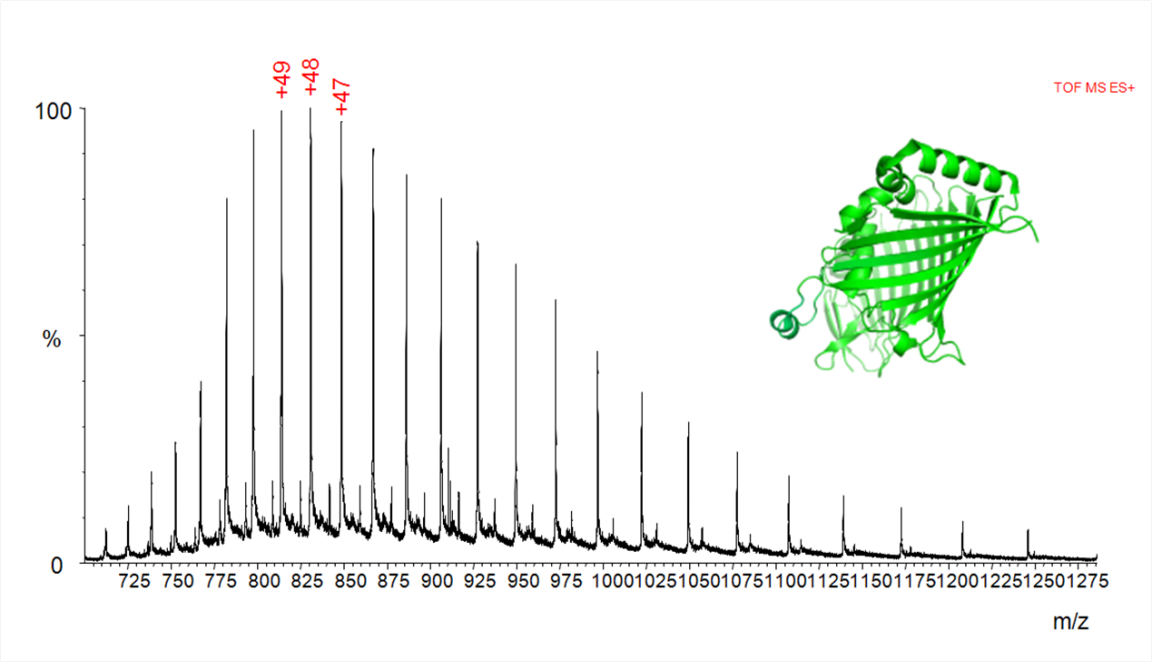


Figure 5.2. Regular ESI spectrum of FMO protein. FMO protein was sprayed in 70% Acentonitrile 30% Water and 0.1% Formic Acid. FMO protein unfolded in this solution. All its pigments (BChl a) were missing.

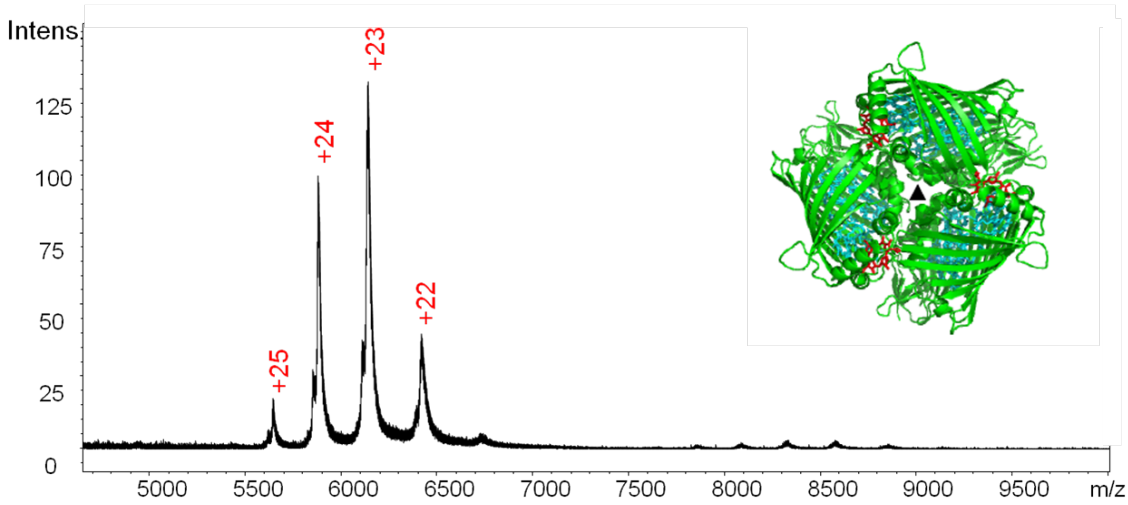


Figure 5.3. Native ESI spectrum of FMO protein. FMO protein in 1 M ammonium acetate solution and was sprayed by nESI. The trimeric complex was preserved and all pigments attached.

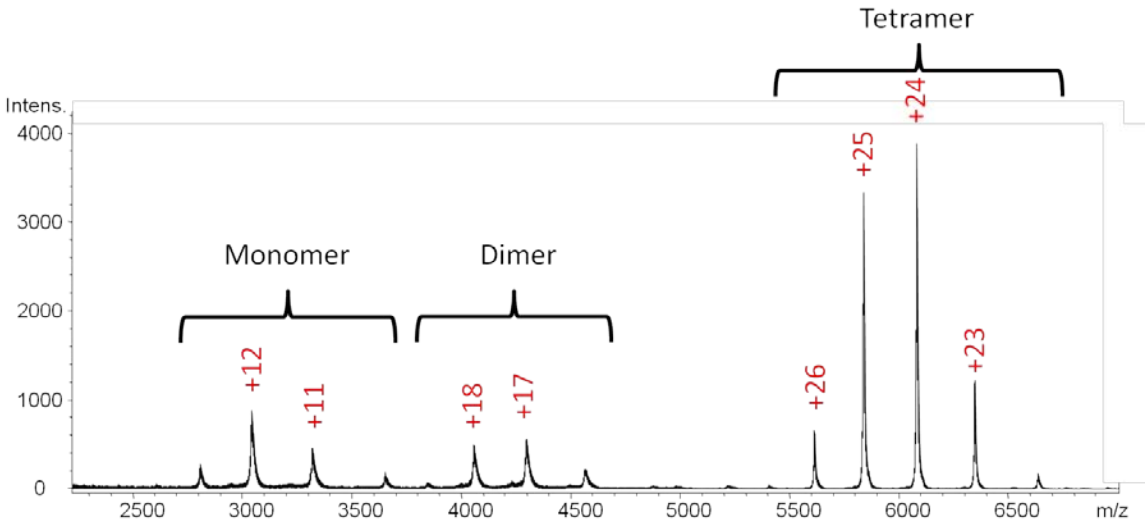


Figure 5.4. Tandem mass spectrum of L-lactate Dehydrogenase. Tetrameric form of lactate dehydrogenase was observed at 5500-6500 *m/z*. By adjusting the collision energy in the collision cell of QTOF, monomeric and dimeric forms of lactate dehydrogenase were observed at low *m/z* range.

Chapter 6

Native Electrospray and Electron-Capture Dissociation in FTICR Mass Spectrometry Provide Top-Down Sequencing of a Protein Component in an Intact Protein Assembly

This chapter is based on recent publication:

H. Zhang, W. Cui, J. Wen, R.E. Blankenship, M.L. Gross, Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly, J Am Soc Mass Spectrom, 21 (2010) 1966-1968.

Author contributions: H.Z., W.C., J.W., R.E.B., M.L.G. designed research; H.Z. W.C., performed research; H.Z., W.C. contributed analytic tool; H.Z., W.C., M.L.G. analyzed data; H.Z., W.C., M.L.G. discussed results; and H.Z., W.C. wrote the paper

Abstract

The intact yeast alcohol dehydrogenase (ADH) tetramer of 147 kDa was introduced into a FTICR mass spectrometer by native electrospray. Electron capture dissociation of the entire 23+ to 27+ charge state distribution produced the expected charge-reduced ions and, more unexpectedly, 39 c-type peptide fragments that identified N-terminus acetylation and the first 55 amino acids. The results are in accord with the crystal structure of yeast ADH, which shows that the C-terminus is buried at the assembly interface whereas the N-terminus is exposed, allowing ECD to occur. This remarkable observation shows promise that a top-down approach will be effective for characterizing their components, inferring their interfaces, and obtaining both proteomics and structural biology information in one experiment.

Introduction

Mass spectrometry (MS) is evolving as an important approach for investigating intact, large protein assemblies in the gas phase(*1*), augmenting other approaches(*2*). Advances in quadrupole time-of-flight instruments and ion mobility have underpinned this new role in structural biology(*3*-*5*). Determinations of molecular weight, stoichiometry, and assembly patterns of essential biological assemblies as large as MDa hepatitis B virus assemblies are now possible(*6*).

To identify the constituents of a protein complex, information about the subunits of a complex is needed. Collisionally activated dissociation (CAD)(*1*), black-body infrared dissociation (BIRD)(*7*), electron-capture dissociation (ECD)(*8*), infrared multiphoton dissociation (IRMPD)(*9*), and surface-induced dissociation (SID)(*10*) can

yield some sequence information for proteins in large assemblies. For example, multiple collisions at keV laboratory energy demonstrate that native protein complexes can be disassembled into peptide fragments(*11*). ECD(*12*), now an established tool for top-down sequencing and for determining post-translational modifications, can successfully characterize, in part, proteins of MW up to 200 kDa(*13*) and noncovalent protein-ligand complexes(*14*). Applications of ECD to protein–protein assemblies(*8*), however, are sparse.

ECD and FT (Fourier Transform) ion cyclotron resonance (ICR) MS combine as an appealing approach for study of protein assemblies because, in principle, the masses of the complex and of the subunits, as well as some sequence of the constituents can be acquired in a single experiment. Indeed, we report here a successful ECD-based FT ICRMS top-down approach that affords the composition, stoichiometry, and partial sequence of a 147 kDa noncovalent protein assembly; namely, the yeast alcohol dehydrogenase (ADH) tetramer.

Experimental

Fresh yeast alcohol dehydrogenase (ADH) complex (Sigma, St. Louis, MO, USA) (1.7 μM) was prepared by buffer exchange before every experiment and sprayed from aqueous NH_4OAc (1 M). Custom spray tips were pulled from Polymicro silicon tubing (360 μm o.d., 150 μm i.d., Phoenix, AZ, USA) by using a microcapillary puller (Sutter Instrument Co., Novato, CA, USA). The sample solution was infused at 25-100 nL/min (Harvard PHD Ultra syringe pump, Instech Laboratories, Inc., Plymouth Meeting, PA, USA). A Bruker SolariX™12 T FTICR mass spectrometer with capabilities for CAD, ETD, ECD and sustained off-resonance irradiation (SORI) was used for analysis.

Results and Discussion

We successfully sprayed (native ESI) the yeast alcohol dehydrogenase (ADH) assembly, obtaining significantly lower charged proteins (23+ to 27+, Figure. 6.1A) compared to what would be observed by non-native ESI. Charge deconvolution gave a MW of 147.5 kDa, verifying that we had introduced the tetramer.

We conducted collisional activation in the front end of the hybrid instrument (prior to the FTICR trap) and found it to be inadequate to fragment the complex, suggesting tight binding of the protein constituents. We also were unable to detect any fragments by using electron-transfer dissociation (ETD). Given the large *m/z* difference of the complex and ETD reagent ion, the precursor ions may not have optimally situated for the ETD reaction. We considered front-end selection by the quadrupole mass analyzer, but this is not yet possible for ions in this *m/z* range. When we attempted in-trap isolation and ECD of the most abundant 26+ charge state, only charge reduction was observed owing to poor dynamic range. The observation matches that of Geel (*8*) on ECD of the 84 kDa gp31 heptamer; only charge reduction and no peptide fragments were produced. The high MW ion packet isolated inside the ICR trap was likely displaced from the cell center owing to perturbations from the waveforms used for isolation. This led to poorer overlap between the electron beam for ECD and the stored ion packet than when activating lower *m/z* ions(*15*).

When we submitted the entire charge-state distribution to ECD, without isolation, we obtained a remarkable result (Fig. 1B) whereby the precursor ions were completely depleted, and two types of product ions formed: (1) a set of peptide fragments of *m/z* < 2000 and (2) charge-reduced precursors down to at least 10+, accompanied presumably

by high m/z ions that are complements to the peptide fragments. To our knowledge, this is the first observation of peptide fragments produced directly by ECD from a protein constituent of a protein assembly.

Analysis of the low *m/z* pattern shows that the first residue is acetylated serine, not methionine. Moreover, 39 c-type ions up to the 55^{th} residue form in the fragmentation from the N terminus (Figure 6.2). Missing in the pattern are ions from chain cleavages at Pro23, 25, 27, and 55, which cannot be seen because there is no mass change when the 5-membered ring is cleaved to give a c-type ion. Nevertheless, those fragments that are produced identify the protein (Figure 6.2).

Moreover, the fragments are consistent with the X-ray crystal structure (Figure 6.3) that shows the N terminal region is free and available for fragmentation whereas the C terminal region is buried at the interface. In this case, the transition to the gas phase preserves sufficient solid-state structure to enable these phenomena. This complex is a dimer of dimers(*16*), held together by nearly a score of salt bridges(*17*). The ionic forces that maintain protein higher order structure and hold together assemblies of this nature become stronger in the gas phase, suggesting why ECD affords sequence information rather than disrupts the assembly. It also explains why we can sequence through the first 55 residues of the N terminus, which is not involved in the interface. Further, the complementary z-ions are likely seen for each of the charge-reduced species at slightly lower *m/z,* dispersed by their complex isotope patterns.

One means to characterize large protein complexes by MS is a “bottom-up” strategy whereby the proteins are denatured leading to release of the subunits, the components separated and proteolyzed, or simply proteolyzed, and LC/tandem MS and

database searching are applied to identify the constituents. More refinement comes from high performance QToF MS of macromolecular assemblies, revealing stoichiometry and connectivity of subunits(*4*). Our observation suggests tha a top-down approach for characterization of protein complexes directly by tandem mass spectrometry will be successful.

Assuming the constituents of this complex are unknown and that top-down sequencing can be successful, we tested whether the sequence from ECD was sufficient to identify the protein. Although there are a number of ways to accomplish this, we chose to generate sequence tags(*18*) (7-20 amino acid) by using Bruker™ Biotools and extended them with consideration of the proline gaps. We then submitted the three longest tags to Mascot searching against the NCBI database with a mass tolerance of 0.02 Da to identify the published Chain A, yeast alcohol dehydrogenase I (gi 112491285) containing 347 residues and having MW of 36.7 kDa. The Mascot peptide score was 294 (expectation value 1.5e-16). Given that the measured molecular weight of the complex is 147.5 kDa, it is straightforward to conclude that four ADH monomers make up the whole assembly.

Conclusion

This approach using FTICR MS applied to the yeast ADH assembly afforded the MW of the complex. More importantly, ECD generated sufficient sequence information to identify reliably the constituent protein. Given the long run of ECD fragments, one can pinpoint that region of the protein that is relatively free in the ADH tetramer, and likely to be remote from the interface of this gas-phase assembly.

More generally, this “top-down” approach may offer a more efficient and simpler procedure than a “bottom-up” approach for characterization of each constituent in a macromolecular assembly. Furthermore, as a one-step protocol, it demonstrates the integration of MS-based proteomics and structural biology in one platform and satisfies a sought-after goal(*19*). In future experiments, we will focus on methodology; namely development of front-end preselection of precursor ions for ECD and addition of IRMPD to the platform for better fragmentation efficiency. We are also studying other homogeneous and ultimately heterogeneous assemblies, the results of which will be the subject of future publications.

Acknowledgement

Research funding was from the National Center for Research Resource of the NIH (Grant P41RR000954), and instrumentation funding was from the High-End Instrument program of the NCRR (Grant 1S10 025101)

References

1. Benesch, J. L., Ruotolo, B. T., Simmons, D. A., and Robinson, C. V. (2007) Protein complexes in the gas phase: technology for structural genomics and proteomics, *Chem Rev 107*, 3544-3567.
2. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell, *Nature 450*, 973-982.
3. Benesch, J. L. (2009) Collisional activation of protein complexes: picking up the pieces, *J Am Soc Mass Spectrom 20*, 341-348.
4. Sharon, M. (2010) How far can we go with structural mass spectrometry of protein complexes?, *J Am Soc Mass Spectrom 21*, 487-500.
5. van Duijn, E. (2010) Current limitations in native mass spectrometry based structural biology, *J Am Soc Mass Spectrom 21*, 971-978.
6. Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C., and Heck, A. J. (2008) High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids, *Proc Natl Acad Sci U S A 105*, 9216-9220.
7. Felitsyn, N., Kitova, E. N., and Klassen, J. S. (2001) Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior, *Anal Chem 73*, 4647-4661.
8. Geels, R. B., van der Vies, S. M., Heck, A. J., and Heeren, R. M. (2006) Electron capture dissociation as structural probe for noncovalent gas-phase protein assemblies, *Anal Chem 78*, 7191-7196.

9. El-Faramawy, A., Guo, Y., Verkerk, U., Thomson, B. A., and Siu, M. (2008) Evaluation of IR Multiphoton Dissociation as a Method for High Mass Protein Clean Up, In *56th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, Co.

10. Jones, C. M., Beardsley, R. L., Galhena, A. S., Dagan, S., Cheng, G., and Wysocki, V. H. (2006) Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions, *J Am Chem Soc 128*, 15044-15045.

11. Benesch, J. L., Ruotolo, B. T., Sobott, F., Wildgoose, J., Gilbert, A., Bateman, R., and Robinson, C. V. (2009) Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments, *Anal Chem 81*, 1270-1274.

12. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, *J. Am. Chem. Soc. 120*, 3265-3266.

13. Han, X., Jin, M., Breuker, K., and McLafferty, F. W. (2006) Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons, *Science 314*, 109-112.

14. Xie, Y., Zhang, J., Yin, S., and Loo, J. A. (2006) Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites, *J Am Chem Soc 128*, 14432-14433.

15. Guan, S., and Burlingame, A. L. High mass selectivity for top-down proteomics by application of SWIFT technology, *J Am Soc Mass Spectrom 21*, 455-459.

16. Powers, E. T., and Powers, D. L. (2003) A perspective on mechanisms of protein tetramer formation, *Biophys J 85*, 3587-3599.

17. Casadio, R., Martelli, P. L., Giordano, A., Rossi, M., and Raia, C. A. (2002) A low-resolution 3D model of the tetrameric alcohol dehydrogenase from Sulfolobus solfataricus, *Protein Eng 15*, 215-223.

18. Mortz, E., O'Connor, P. B., Roepstorff, P., Kelleher, N. L., Wood, T. D., McLafferty, F. W., and Mann, M. (1996) Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases, *Proc Natl Acad Sci U S A 93*, 8264-8267.

19. Zhou, M., and Robinson, C. V. (2010) When proteomics meets structural biology, *Trends Biochem Sci*, DOI:10.1016/j.tibs.2010.1004.1007

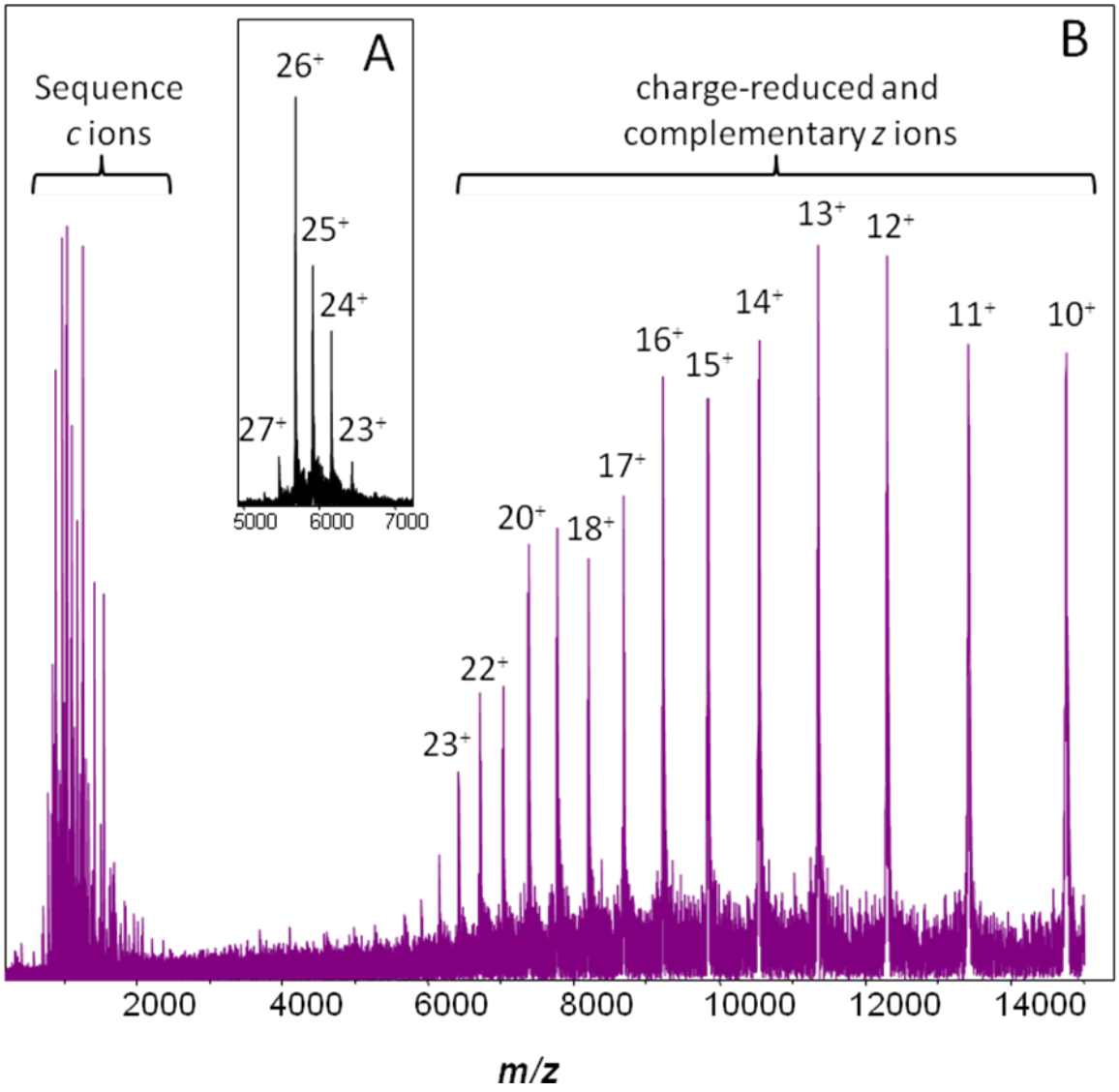


Figure 6.1. Mass spectra of ADH complex: (A) Mass spectrum and (B) ECD product-ion spectrum of the yeast ADH complex.

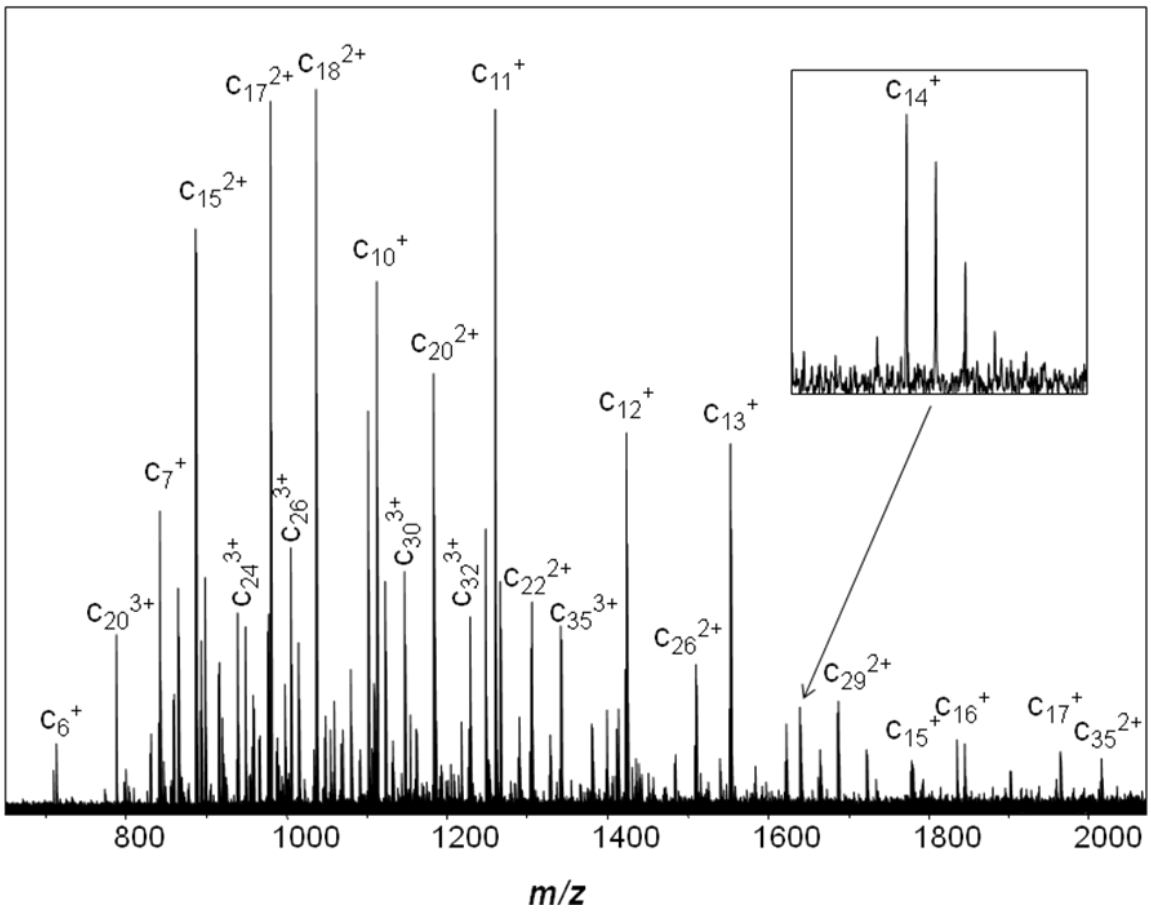


Figure 6.2. ECD mass spectrum of the ADH complex over the low m/z range of Figure 1

ac IPETQKGVIFYESHGKLEYKDIPVPKPKANELLINVKYSGVCH
TDLHAWHGDWPLPVKLPLVGGHEGAGVVVGMGENVKGWKIGDYA
GIKWLNGSCMACEYCELGNESNCPHADLSGYTHDGSFQEYATAD
AVQAAHIPQGTDLAEVAPVLCAGITVYKALKSANLMAGHWVAIS
GAAGGLGSLAVQYAKAMGYRVLGIDGGEGKEELFRSIGGEVFID
FTKEKDIVGAVLKATDGGAHGVINVSVSEAAIEASTRYVRANGT
TVLVGMPAGAKCCSDVFNQVVKSISIVGSYVGNRADTREALDFF
ARGLIKSPIKVVGLSTLPEIYEKMEKGQIVGRYVVDTSK

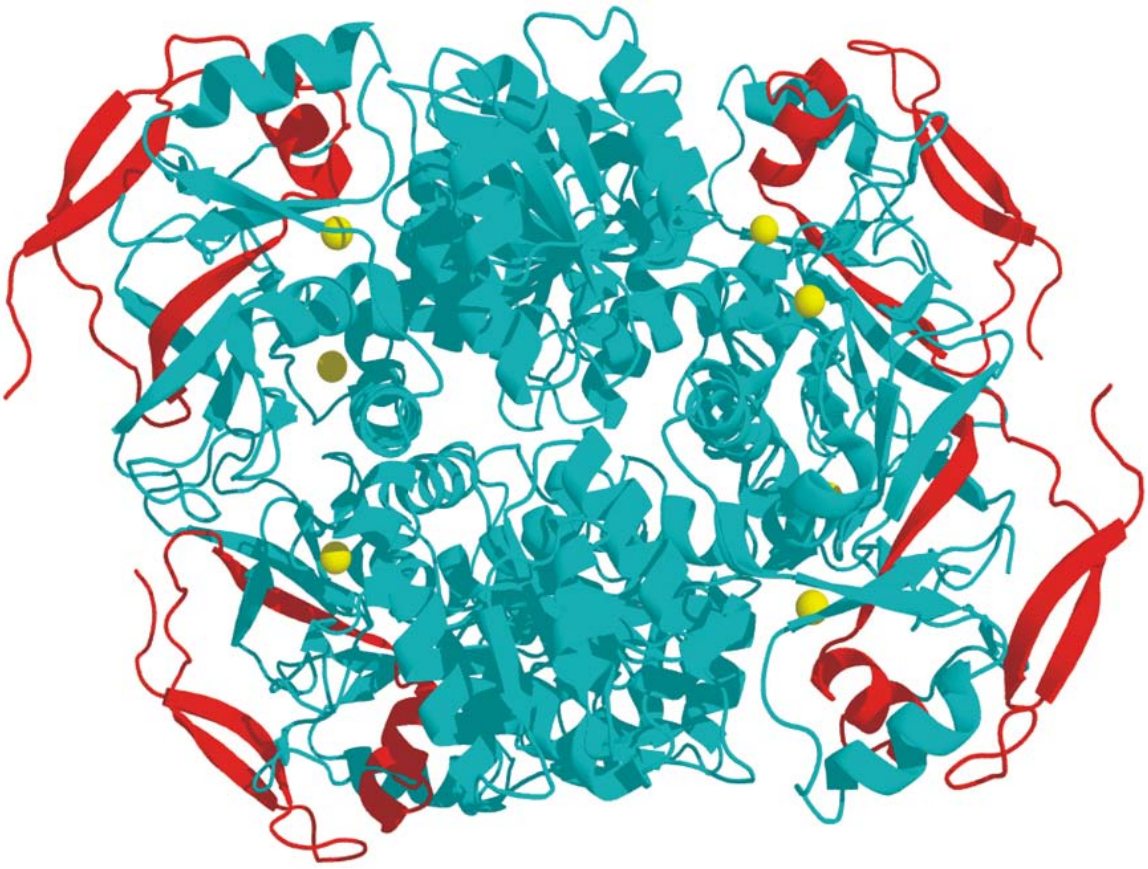


Figure 6.3. ADH tetramer: ADH sequence showing the sites of ECD cleavages (top) and crystal structure of the yeast ADH tetramer (2HCY in PDB) with the N-terminal 55 residues that were sequenced highlighted in red (bottom).

Chapter 7

Top-down Approach of Protein Assembly Studies: Native Electrospray and Electron-Capture Dissociation in FTICR Mass Spectrometry

Abstract

Native ESI combined with mass spectrometry is complementary to other biophysical methods owing to its high sensitivity, extended mass range, and fast data acquisition/processing. Protein assemblies with molecular masses up to MDa can now be investigated. Most approaches have used quadrupole/time-of-flight tandem mass spectrometry sometimes coupled with ion mobility to reveal the stoichiometry, shape, and dissociation of protein assemblies. The amino-acid sequence of the subunits, however, still relies heavily on independent bottom-up proteomics. We report here the integration of an electron-capture dissociation (ECD) top-down approach with native ESI to study protein assemblies in a 12 tesla FTICR mass spectrometer. The results from yeast alcohol dehydrogenase (ADH, 147 kDa), concanavalin A (ConA,103 kDa), and photosynthetic Fenna-Matthews-Olsen protein complex (FMO, 140 kDa) show that relatively free and flexible regions of the subunits can be sequenced by ECD or activated ion ECD. Furthermore, non-covalent metal-binding sites can also be determined for the ConA assembly. Most importantly, the regions that undergo fragmentation, either from one of the termini by ECD or from the middle on a protein as initiated by CID, correlate well with the B-factor from X-ray crystallography. This factor is a measure of to what extent an atom can move away from its coordinated position as a function of temperature or crystal imperfections. The approach provides not only top-down proteomics information of the complex subunits but also structural insights complementary to those obtained by ion mobility.

Introduction

We reported recently that a large number of consecutive backbone cleavages occur for the 147 kDa yeast alcohol dehydrogenase tetramer upon ECD in a 12 tesla FTICR mass spectrometer (*1*). We now follow up this observation with a report on a more detailed and general development and application of ECD and CID to several different protein assemblies. We show that the masses, subunit identities, metal-ion binding sites, and some structural information can be obtained in one experiment. The ECD fragmentation patterns of the protein complexes activated at different collision energies show preferred fragmentations of one terminus over the other, or of the middle region, indicating that the fragmentations are structurally significant. To correlate these fragmentations, we utilize a B-factor parameter from X-ray crystallography, a factor that is predictive of the flexible regions of a protein.

Material and Methods

Chemicals and Proteins

Ammonium acetate, water, yeast alcohol dehydrogenase (ADH) form *Saccharomyces cerevisiae*, concanavalin A (ConA) from *Canavalia ensiformis* (Jack bean) were purchased from Sigma-Aldrich (St. Louis, MO). The FMO protein from green sulfur bacterial *Chlorobium tepidum* was purified as previously described(*2*).

Sample preparation for Native ESI

Lyophilized protein powder was dissolved in 100 mM ammonium acetate (pH = 6.5-7) to afford an assembly concentration at 5 μM. The sample was washed three times with equal volume 100 mM ammonium acetate buffer in a Vivaspin 500 concentrator with 10,000 or 30,000 molecular weight cut off (MWCO) (Vivaproducts Inc., Littleton,

MA). Buffer exchange of purified FMO protein sample was conducted by Vivaspin 500 conentrators (30,000 MWCO) before the native ESI experiment.

MS of Protein complexes

Protein sample was delivered by syringe pump (Harvard PHD Ultra syringe pump, Instech Laboratories, Inc., Plymouth Meeting, PA) at flow rate 5-300 nL/min to a nano spray source, which has a custom-pulled nano spray tip (Sutter Instrument Co., Novato, CA) of silica capillary tubing (360 μm o.d., 150 μm i.d., Polymicro Technologies, Phoenix, AZ). Mass spectra were acquired with a Bruker Solarix 12T FTICR mass spectrometer (Bruker Daltonics, Bremen, Germany). The capillary voltage was 0.9-1.3 kV. The drying-gas temperature was 100 ℃; its flow was 2.5 L/min. The voltage for in-source fragmentation (ISCID) was varied from 0 to 100 V depending on the application. The ion-funnel RF amplitude was 300 V_{pp}, and the ion-funnel voltages were 200 V (funnel 1) and 18 V (funnel 2). RF frequencies used in all ion-transmission regions were the lowest available value: multipole1 (2 MHz), quadrupole (1.4 MHz) and transfer line (1 MHz). The collision voltage for the collision cell was varied from 0 to 50 V, depending on the application. Ions were accumulated for 500 ms in the RF-hexapole ion trap before being transmitted to the infinity ICR trap. The time-of-flight was ~ 2.5 ms for the protein-assembly ions. The source region (PS1) pressure was 2.3 mbar; the quadrupole region (PS4) pressure was 4.36e-6 mbar and trap-chamber pressure (PS6) was 1.3e-9 mbar. The typical ECD pulse length was 0.06 s, ECD bias 0.6 V and ECD lens 10 V. The ECD hollow cathode heater current was 1.6 A. MS parameters were slightly modified in each individual sample to obtain an optimized signal. Calibration was done by ESI of cesium perfluoroheptanoic acetate up to *m/z* 8500.

Data Analysis

Peak picking and spectra deconvolution were performed with Bruker DataAnalysis software (Bruker Daltonics, Bremen, Germany). Bruker Biotools software was used for mapping the measured peak mass list with calculated fragment mass list from the protein sequences. To validate that the sequence information is sufficient to identify the constituent subunits, sequence tags were generated with the Biotools software based on the deconvoluted spectra and submitted to Mascot searching against NCBI database as previous reported(*1*). Alternative peak matching was conducted by Prosight PTM (v1.0, https://prosightptm.northwestern.edu)(*3*).

Results and Discussion

ECD of protein complexes

In FTICR MS, electrons from a cathode emitter interact with trapped protein-assembly ions to initiate fragmentation. We could select the ions based on their *m/z* before fragmentation either prior to the ICR trap, using a selection quadrupole or in the FTICR trap itself (Figure 7.1). Ion selection allows the study of each species or charge state. When using native ESI, some signal intensity is lost during ion pre-selection in commercial FTICR instruments. The upper limit on the instrument used here is ~ 6000 *m/z* for selection by the quadrupole. Large protein-assembly ions in our experiment had *m/z* values close to 6000 because native spray produces lower charge states than normal ESI, where the proteins are denatured. Preselecting high *m/z* ions by the quadrupole was inefficient, giving poor signal intensity. To select ions in the FTICR trap, the excitation waveforms used to excite an ion of interest also excite ions of nearby *m/z*. This excitation

causes movement of high *m/z* ion packets from the trap's central axis, reducing the overlap between protein-assembly ions and the emitted electrons needed for ECD(*4*).

Because preselection for ECD is not highly efficient, we chose to avoid preselection and to submit the entire charge distribution to ECD. This is not a problem here because native ESI generated a narrow charge-state distribution extending over ~ 4 charge states with one as most significant. When we communicated preliminary results, we also did no preselection and obtained, using ECD, a clean set of c ions from the ADH protein assemblies(*1*). A report from the Langridge-Smith group demonstrated a similar improvement in overall efficiency by applying ECD to several of the most abundant charge states of a protein-ligand complex(*5*).

Having established a protocol, we conducted top-down ECD of protein assemblies with different sizes and oligomer states. In addition to the ADH assembly, ConA and the FMO antenna assemblies fragment to give c or z ions and charge reduction (Figure 7.2). The results suggest that ECD top-down fragmentation of protein assemblies will be useful for other assemblies. The approach affords subunit sequence information and MW information in single experiment.

Subunit sequence and metal binding sites

With ECD of protein assemblies, both ions representing the full assembly and fragment ions are observed. Information extracted from this single experiment may give subunit identification and stoichiometry of protein complexes. As for the previously reported ECD of the yeast ADH assembly, the ECD fragmentation extent is sufficient to identify the constituent subunit by sequence-database searching(*1*). The stoichiometry of

the assembly can be determined on the basis of the molecular mass of the assembly and that of its subunits.

For ConA, the intact mass of the assembly (103 kDa) is read out directly from spectrum of native ESI. By applying ECD to the assembly, c fragment ions are formed, and they can be assigned within 20 ppm mass accuracy. Sequence information matched ConA (237 amino acid, 25,539 Da, UniProtKB database ID: CONA_CANVI). A total 38 fragment ions, with or without metal adducts, covered the N terminus of ConA up to 57^{th} residue from the N terminus. Having obtained subunit identity and intact mass of complex, we found that it is straightforward to conclude that ConA forms a tetramer.

ECD is capable of identifying labile modifications and non-covalent ligand binding sites(*6*). Important biological information including phosphorylation and drug-binding sites, can be elucidated. For example, Loo and co-workers demonstrated that ECD in a top-down format can locate non-covalent drug binding sites(*7, 8*).

In the case of ConA, metal binding is preserved in the gas phase upon native ESI. Each lectin ConA protein has two metal-binding sites for two bivalent metal ions (Mn^{2+} for the S1 site and Ca^{2+} for the S2 site)(*9*); these sites are essential for ConA's interaction with glycogen. Previous metal-binding studies demonstrated that several bivalent metal ions can bind to ConA as substitutes for Mn^{2+} and Ca^{2+} (Co^{2+}, Ni^{2+}, Cd^{2+}, Mg^{2+} for the S1 site, Cd^{2+} for the S2 site)(*10-12*). Fragment ions containing Ca^{2+} and Mg^{2+} are preserved in the ECD top-down study of tetrameric ConA, and residues Asp-19 and Asp-10 can be identified as major Mg^{2+} and Ca^{2+} binding sites on the basis of ECD fragmentation (Figure 7.3a); these assignments are in agreement with the S1 and S2 binding sites seen

in the crystal structure(*13*). We note that an ECD top-down experiment with denatured ConA failed to provide metal binding information owing to the loss of metal ion upon denaturation or ESI (Figure 7.3b).

In most protein-assembly studies by MS, the mass of the assembly is measured by native ESI whereas subunit identity and PTM information are provided by bottom-up LC-MS of the proteolytically digested assembly. Our results show that these two independent experiments can be integrated into one by using a top-down approach with fragmentation induced by ECD. More importantly, this approach should be applicable to the study of non-covalent ligand binding and labile PTMs (e.g., phosphorylation and glycosylation) of a protein assembly.

Structural information elucidated by ECD experiment

One rapid growing area of MS-based investigations of protein assemblies focuses on their nature of dissociation in the gas phase(*14*). Protein assemblies can be dissociated by applying additional accelerating potential or heating during the transmission of the ions from the source to the detector(*15, 16*). The dissociation results in ejection of a few subunits from the assembly, providing a view of the arrangement of the assembly(*17*). Interestingly, ejected subunits usually carry more charges per mass than the remainder of the assembly. IM shows that as a region unfolds, it allows protons to move from the core, causing it to charge disproportionately; dissociation of that segment then occurs(*14*).

Subunit unfolding can be investigated by ECD top-down approach. This “peel-an-onion” process occurs because the internal energy of the assembly increases, providing in principle information about relatively flexible and unstructured arms of the constituent proteins. In our experiment, internal energy of protein complex was increased by using a

larger acceleration voltage to afford ion-source CID (ISCID). We used nitrogen as collision gas. The unfolding process induced by increasing the internal energy can be monitored by the locating the sites of ECD fragmentation. The tetrameric assembly of yeast ADH was activated by varying the acceleration voltage from 0 to 100 V (Figure 7.4). Various c ions of higher *m/z* become more abundant with increasing acceleration voltage (Figure 7.5a). The results indicate that protein unfolding starts at the N-terminus and extends further to the core of the assembly.

Correlation of fragmentation with the B-factor

There are clearly opportunities for new structural insights from ECD; its pattern may serve as an indicator of flexibility in protein structure. In the x-ray crystal structure, the atomic displacement parameter (B-factor) reflects the flexibility and dynamics of a polypeptide chain(*18, 19*); for example, a large B-factor indicates high mobility of individual atoms and side chains. When we plot the B-factor as a function of the location of the amino-acid residue for the N-terminal region of ADH, we see that the extent of fragmentation correlates with it; the highest B-factor is on the end of terminal region among the four chains (Figure 7.5b). We interpret that this subunit unfolds first, moving away from the protein core, adopting charge, and undergoing subsequent fragmentation (Figure 7.6).

There is evidence from previous IM studies on protein assemblies that subunit unfolding occurs before dissociation. A more detailed unfolding process can now be proposed by considering the ECD outcome and locating regions of structural flexibility. Given that protein flexibility is highly correlated with protein function, ECD in a top-

down format may be highly informative for the study of not only of structure but also of flexibilty of biologically important protein assemblies.

Structural elucidation by top-down ECD and CID

Although an ECD-based top-down approach to protein assemblies appears to identify the terminal regions of high flexibility, a flexible region in the middle of protein sequence may not be sampled in the fragmentation. We decided to integrate other fragmentation approaches (i.e., CID) to see if the combination of CID and ECD can provide more comprehensive structural information. In the study of the trimeric FMO antenna protein, only z ions from C terminal of FMO are produced in the ECD top-down experiment (Figure 7.7), and these results are consistent with the C terminal region of FMO being more flexible than the N terminal region, which is heavily involved in forming the protein interface.

Interestingly, an unusual set of fragment ions with different charge states were observed in a CID experiment (Figure 7.7). A database search indicates the fragment ion is not from the N or C terminal regions. Using the high mass resolving power and accurate mass measurement capabilities of FTICR MS, we were able to identify this ~ 10 kDa species as originating from 201-295 in the middle of FMO protein sequence (Figure 7.8). Previous studies of other protein assemblies showed that unfolding and ejection of a subunit is the major pathway upon CID.

The unique fragment patterns are consistent with the assembly's structure. The FMO protein trimer contains seven bacteriochlorophyll a (BChl a) pigments inside, forming a tight trimeric entity by salt bridges(*20*). For each subunit, a series of beta sheets form two parallel walls, like a "taco shell", holding seven BChl a pigments. The

open end of the “taco shell” contains several alpha helixes and two BChl a pigments (BChl 1 and 2) and points to the center of the trimeric complex; all the BChl a pigments are buried inside. Dissociation of the assembly will destroy those salt bridges and expose BChl a pigments, a process that requires high energy. Thus, no subunit ejection occurred upon CID. Alternatively, a loop of a beta sheet within the region 201-295 unfolded first, fragmented upon collisional activation, and lost a large peptide from the middle of the protein.

Based on the crystal structure of trimeric FMO complex(*21*) (Figure 7.9), the C terminus of the protein is at the bottom of this complex, while the N-terminal is in a middle beta-sheet structure forming the side wall of the assembly. The region that undergoes CID is on the top of protein complex. Given our supposition that the locale for fragmentation is related to the flexibility of protein structure, we applied the B-factor as a potential correlate for the fragment pattern. Indeed, the region undergoing CID is one that has one of the highest B factors in the protein sequence. Moreover, the region 201-295 released upon CID overlaps with 275-366 that releases upon ECD. The combined outcome of ECD and CID is evidence that the unfolding starts at the middle region of protein. This initial unfolding also increased the flexibility of the polypeptide chain extending to the C-terminus, which undergoes ECD.

Conclusions

ECD in a top-down format appears to be an alternative for the structural studies of protein assemblies and offers a unique view of the dynamics of a protein assembly. The results from three protein complexes show that unique fragmentations occur (upon both CID and ECD) and that they are indicative of their specific structures. We propose that

careful analysis of the fragmentation patterns obtained with this top-down approach will benefit the structural studies of protein assemblies. This proposal is supported by the results from three model assemblies: yeast ADH tetramer, FMO antenna complex of green sulfur bacterial, and concanavalin A. Specifically, we could obtain in one experiment sequence information, non-covalent metal-binding sites, assembly stoichiometry, and structural insight that pinpoints flexible regions. This approach is complementary to not only the traditional methods of crystal structure and NMR spectroscopy but also commonly applied approaches of native ESI, QToF MS, and ion mobility.

Although this FT-ICR based ECD top-down approach is still in its infancy, its promise will increase developments to improve efficiency and selection capability and to integrate other fragmentation method including IRMPD and other photochemical activation schemes, and in-trap, high-energy collisional activation. In another direction, different and important protein complexes of heterogeneous composition need to be investigated to demonstrate a broader application of this approach; some of these developments are under investigation in our laboratory.

References

1. Zhang, H., Cui, W., Wen, J., Blankenship, R. E., and Gross, M. L. Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly, *J Am Soc Mass Spectrom 21*, 1966-1968.
2. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting, *Proc Natl Acad Sci U S A 106*, 6134-6139.
3. LeDuc, R. D., Taylor, G. K., Kim, Y. B., Januszyk, T. E., Bynum, L. H., Sola, J. V., Garavelli, J. S., and Kelleher, N. L. (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry, *Nucleic Acids Res 32*, W340-345.
4. Guan, S., and Burlingame, A. L. High mass selectivity for top-down proteomics by application of SWIFT technology, *J Am Soc Mass Spectrom 21*, 455-459.
5. Clarke, D. J., Murray, E., Faull, P. A., Hupp, T., Barran, P., Langridge-Smith, P., and Mackay, C. L. (2010) Investigating Protein-Peptide Binding by 'Top-Down' FT-ICR MS, Ion-Mobility MS and Hydrogen/Deuterium Exchange, In *58th ASMS Conference on Mass Spectrometry and Allied Topics*, Salt Lake City, Utah.
6. Cooper, H. J., Hakansson, K., and Marshall, A. G. (2005) The role of electron capture dissociation in biomolecular analysis, *Mass Spectrom Rev 24*, 201-222.
7. Yin, S., and Loo, J. A. Elucidating the site of protein-ATP binding by top-down mass spectrometry, *J Am Soc Mass Spectrom 21*, 899-907.

8. Xie, Y., Zhang, J., Yin, S., and Loo, J. A. (2006) Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites, *J Am Chem Soc 128*, 14432-14433.

9. Hardman, K. D., and Ainsworth, C. F. (1972) Structure of concanavalin A at 2.4-A resolution, *Biochemistry 11*, 4910-4919.

10. Kalb, A. J., and Levitzki, A. (1968) Metal-binding sites of concanavalin A and their role in the binding of alpha-methyl d-glucopyranoside, *Biochem J 109*, 669-672.

11. Young, N. M. (1983) Magnesium as a natural substitute for manganese in concanavalin A and other lectins, *FEBS Lett 161*, 247-250.

12. Sanders, J. N., Chenoweth, S. A., and Schwarz, F. P. (1998) Effect of metal ion substitutions in concanavalin A on the binding of carbohydrates and on thermal stability, *J Inorg Biochem 70*, 71-82.

13. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1972) The covalent and three-dimensional structure of concanavalin A, *Proc Natl Acad Sci U S A 69*, 2580-2584.

14. Benesch, J. L., and Robinson, C. V. (2006) Mass spectrometry of macromolecular assemblies: preservation and dissociation, *Curr Opin Struct Biol 16*, 245-251.

15. Benesch, J. L., Sobott, F., and Robinson, C. V. (2003) Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry, *Anal Chem 75*, 2208-2214.

16. Geels, R. B., Calmat, S., Heck, A. J., van der Vies, S. M., and Heeren, R. M. (2008) Thermal activation of the co-chaperonins GroES and gp31 probed by mass spectrometry, *Rapid Commun Mass Spectrom 22*, 3633-3641.

17. Benesch, J. L. (2009) Collisional activation of protein complexes: picking up the pieces, *J Am Soc Mass Spectrom 20*, 341-348.

18. Daggett, V., and Levitt, M. (1992) A model of the molten globule state from molecular dynamics simulations, *Proc Natl Acad Sci U S A 89*, 5142-5146.

19. Parthasarathy, S., and Murthy, M. R. (2000) Protein thermal stability: insights from atomic displacement parameters (B values), *Protein Eng 13*, 9-13.

20. Tronrud, D. E., Wen, J., Gay, L., and Blankenship, R. E. (2009) The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria, *Photosynth Res 100*, 79-87.

21. Camara-Artigas, A., Blankenship, R. E., and Allen, J. P. (2003) The structure of the FMO protein from Chlorobium tepidum at 2.2 A resolution, *Photosynth Res 75*, 49-55.

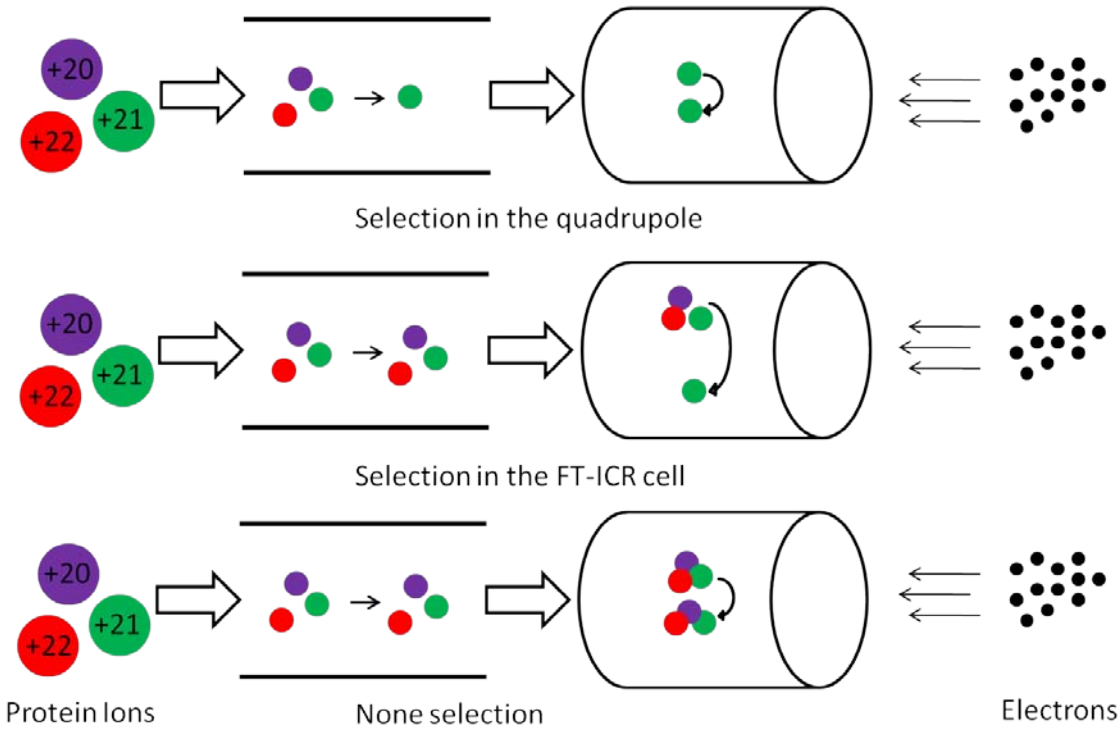


Figure 7.1. ECD top-down approach by FTICR MS showing three ways for ion selection: in the quadrupole region before entering the FTICR trap (top); in the FTICR trap (middle); and without pre-selection (bottom).

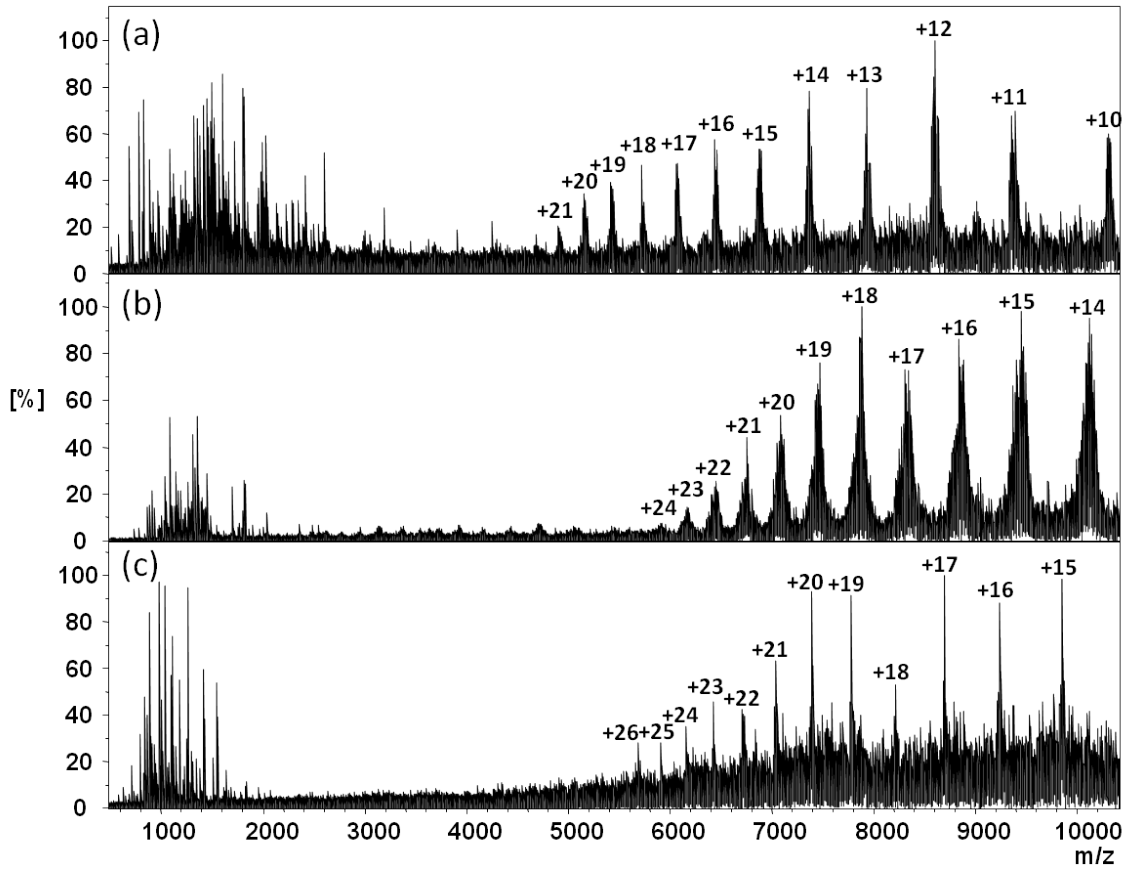
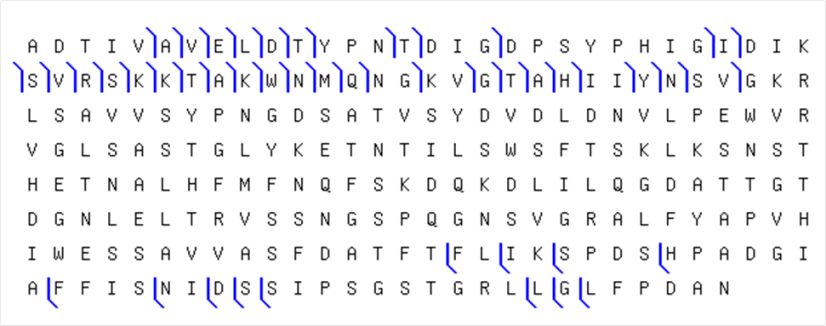


Figure 7.2. ECD spectra of intact protein complexes: (a) concanavalin A (ConA) from *Canavalia ensiformis* (Jack bean); (b) FMO antenna protein from green sulfur bacteria *Chlorobium tepidum*; (c) yeast alcohol dehydrogenase (ADH) from *Saccharomyces cerevisiae* .

Top-down experiment of denatured ConA



Top-down experiment of ConA from native ESI

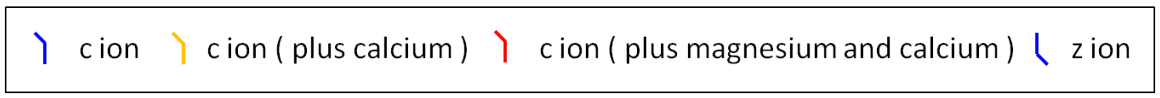


Figure 7.3. Sequence coverage of ConA by top-down ECD for denatured ConA (in 50% acetonitrile 50% water 1% formic acid) (top and for ConA in its near native state (in 100 mM ammonium acetate, pH 7).

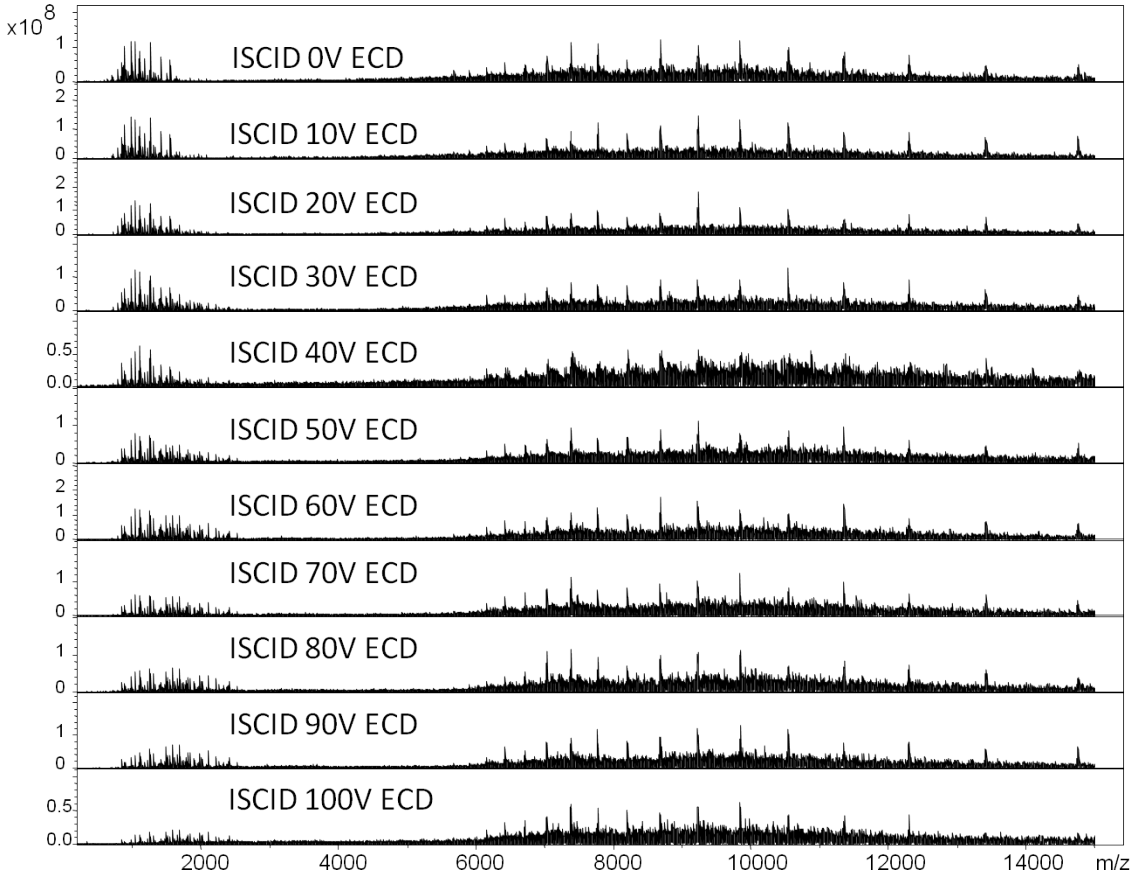


Figure 7.4. ECD spectra of ADH with different acceleration potentials (ISCID). Each spectrum was a summary of 100 scans with the same instrument parameters (except ISCID). The ECD fragment ions were observed at low *m/z* and are c ions from the N-terminal region of ADH. The charge-reduced ions and complimentary z ions were observed at high *m/z* region of the spectra.

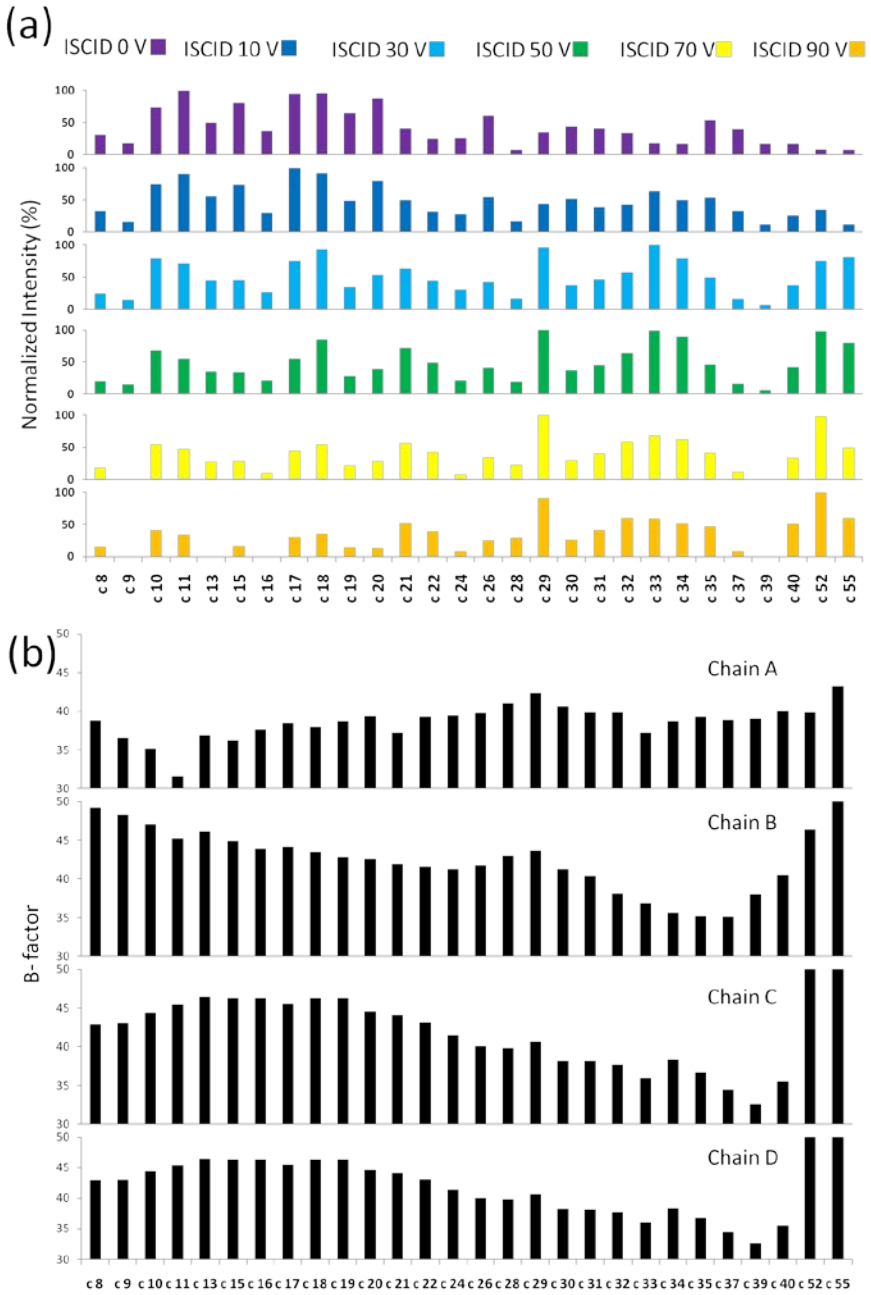


Figure 7.5. ECD fragmentation and B-factor: (a) Normalized intensity of ECD fragment ions (c ions) from the ADH assembly at different acceleration potentials (ISCID). (b) Atomic displacement parameter (B-factor) plot for the C terminal region of ADH. The B-factor values are from the crystal structure of yeast ADH (PDB id: 2HCY). In the crystal structure, the protein has four different conformations represented by Chain A-D.

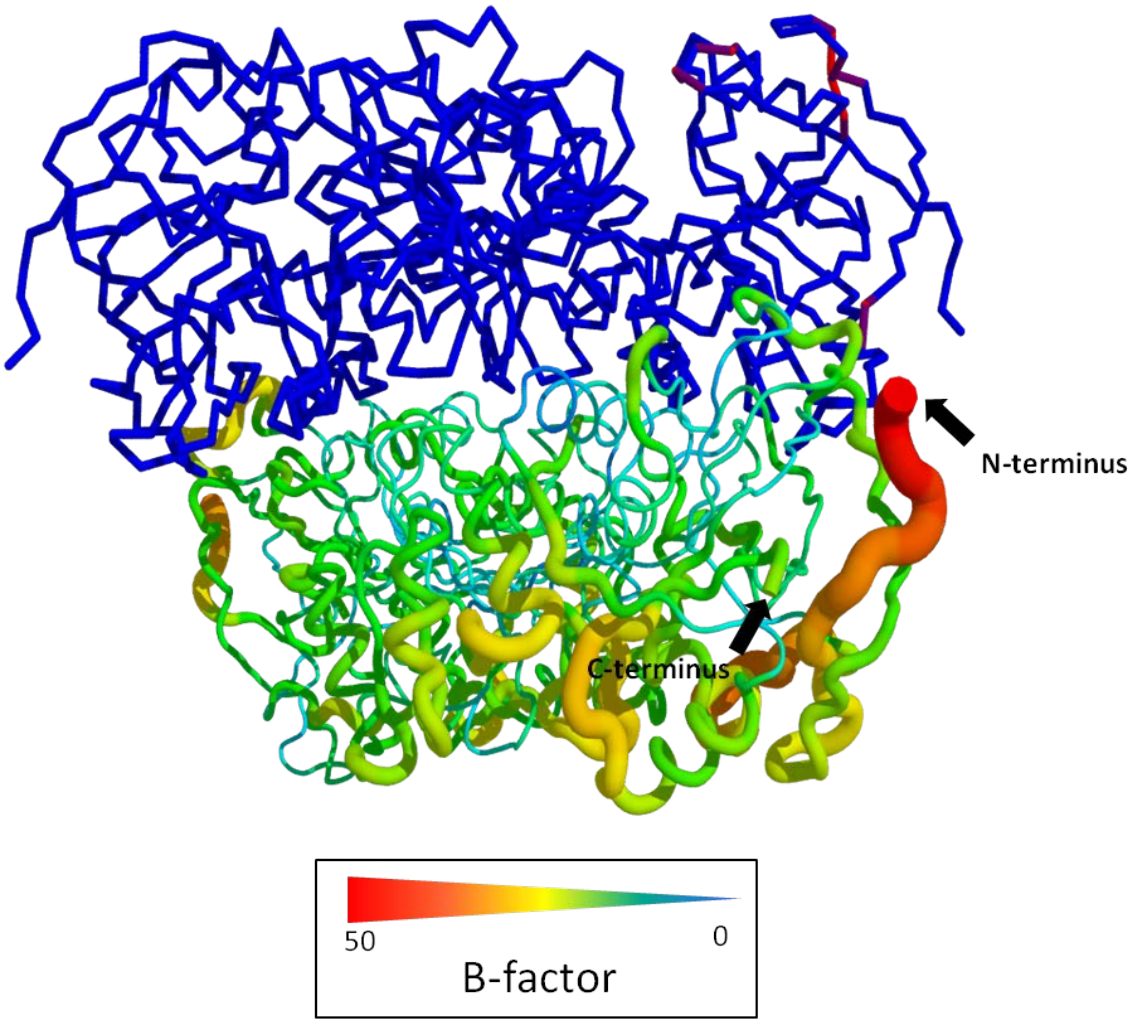


Figure 7.6. Tetrameric ADH crystal structure color-coded to show the B factor extent. Tetrameric ADH complex was assembled by crystal packing (PDB id: 2HCY). Dimer on the bottom of ADH complex is displayed with the B-factor scheme (the color and width represent the value of B factor) from crystal data (Chain B).

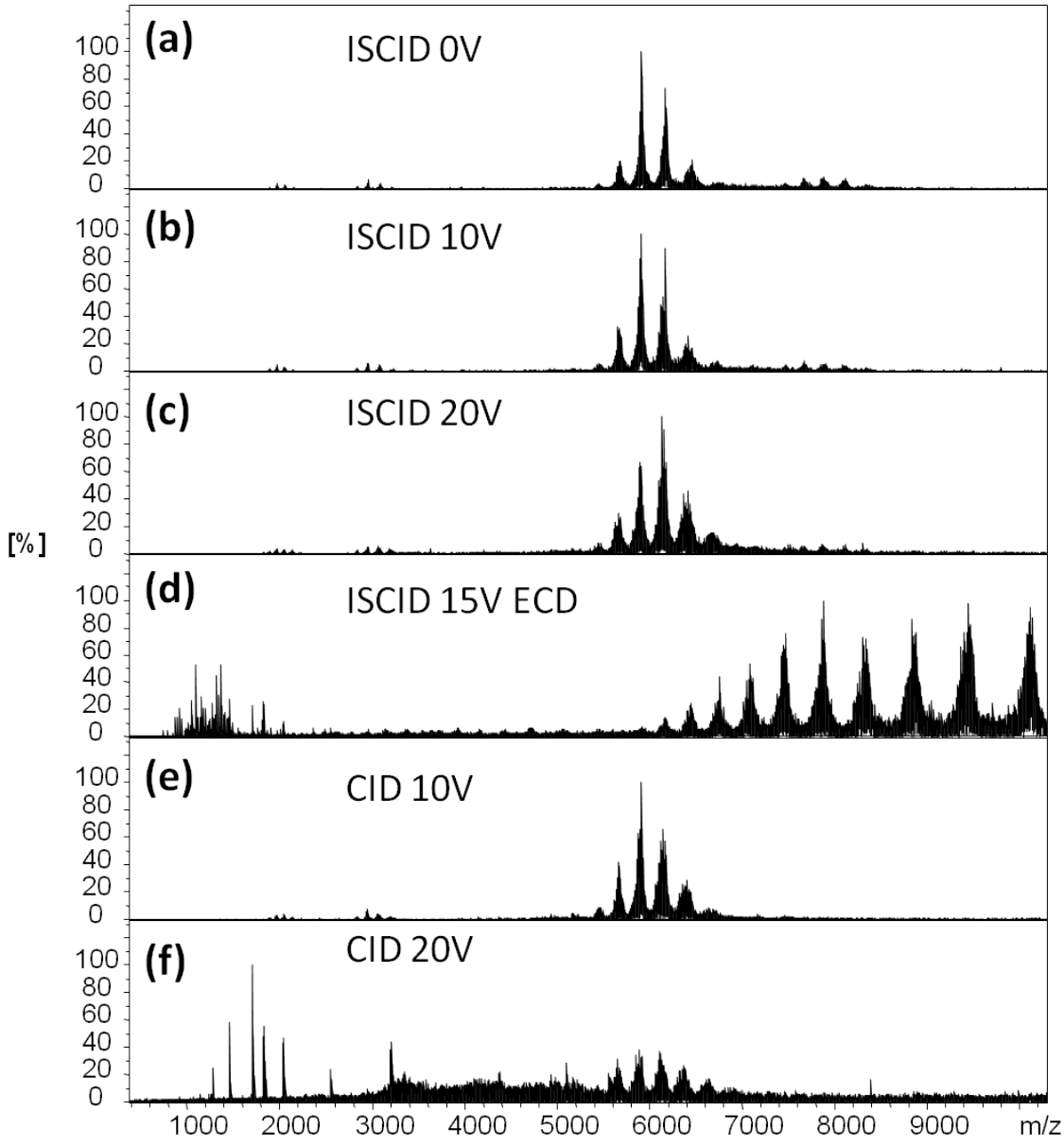


Figure 7.7. Native ESI and ECD/CID top-down spectra of the FMO antenna protein complex. (a-c) native ESI spectra of FMO with 0, 10, 20 V ISCID. No fragment ions were observed. (d) ECD spectrum of FMO protein with 15 V ISCID. (e) CID (10 V) spectrum of FMO; no fragment ions were observed. (f) CID (20 V) spectrum of FMO. Multiply charged fragment ions was observed at low *m/z*.

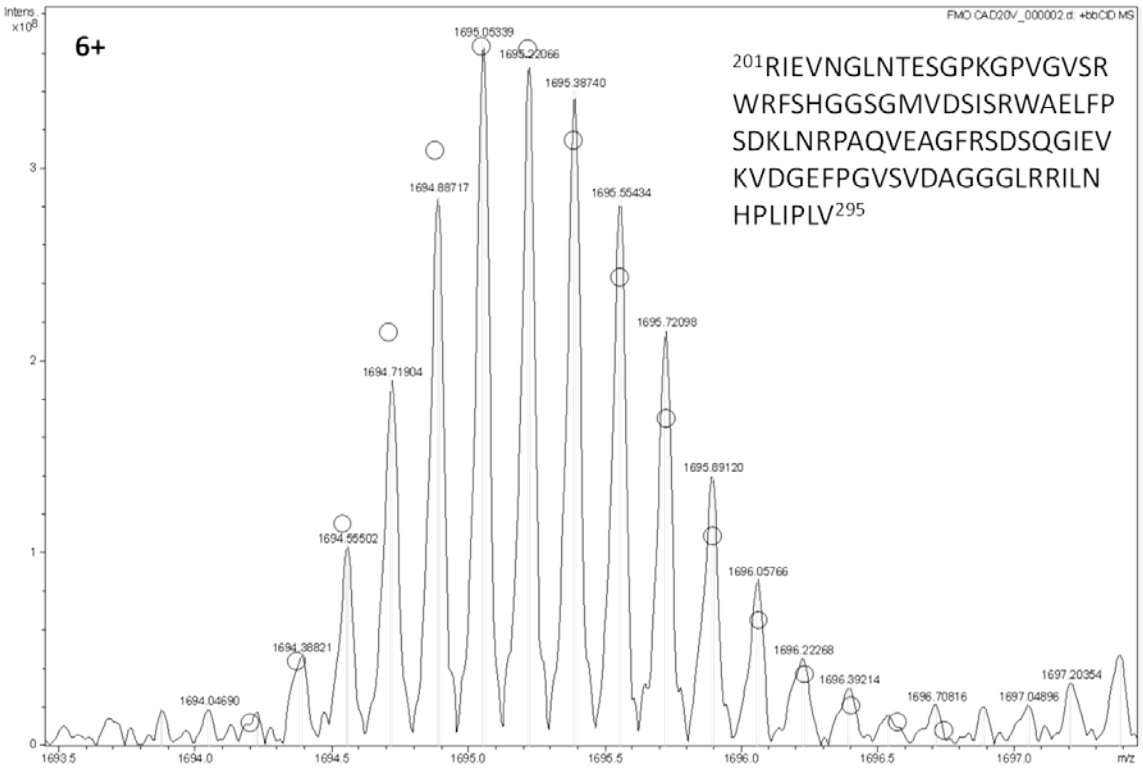


Figure 7.8. Product-ion spectrum of the FMO antenna protein produced upon top-down CID). The 6+ charge state was expanded and aligned with theoretical calculated *m/z* values and abundances (open circles).

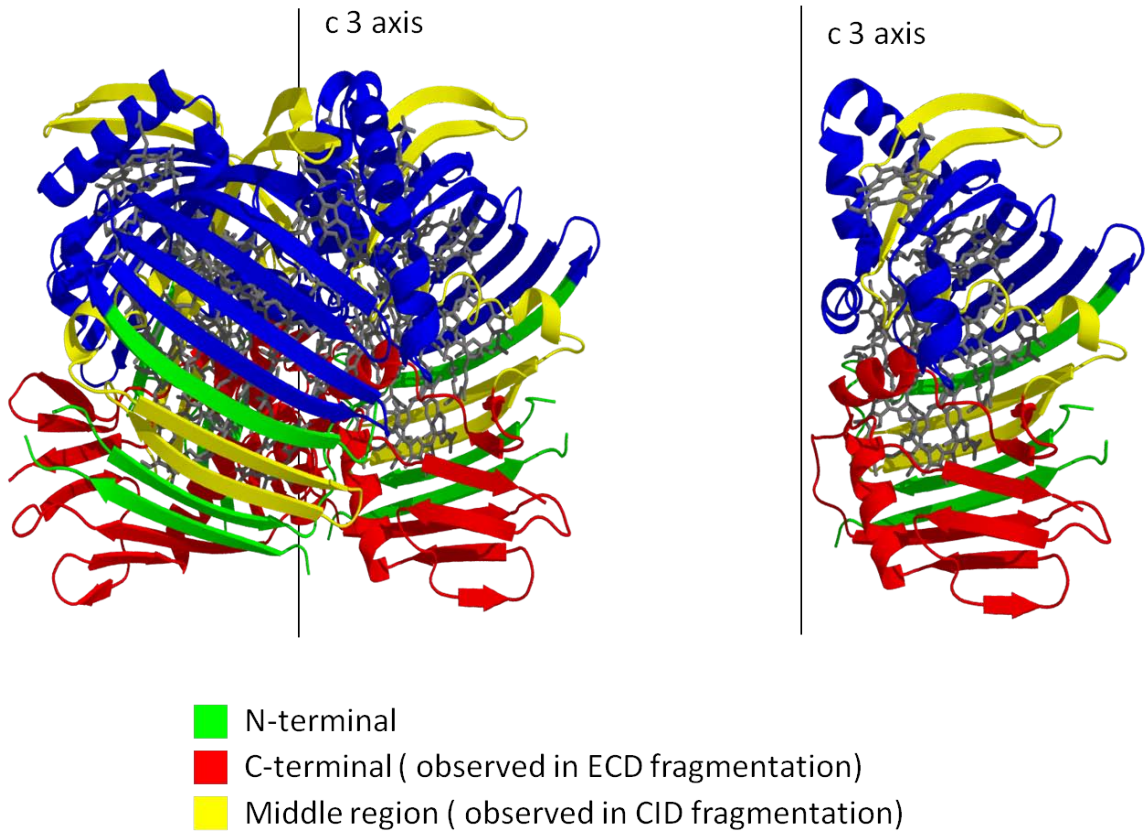


Figure 7.9. Crystal structure of the FMO antenna protein complex; the protein is displayed in cartoon mode; BChl a is depicted as a stick mode with gray color.