**This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.**

Structural proteomics : elucidating in vivo structural dynamics of integral membrane proteins by hydroxyl radical footprinting

Zhu, Yi

2012

Zhu, Y. (2012). Structural proteomics : elucidating in vivo structural dynamics of integral membrane proteins by hydroxyl radical footprinting. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/48373

https://doi.org/10.32657/10356/48373

Downloaded on 20 Mar 2024 20:05:46 SGT

NANYANG
TECHNOLOGICAL
UNIVERSITY

STRUCTURAL PROTEOMICS: ELUCIDATING *IN VIVO* STRUCTURAL DYNAMICS OF INTEGRAL MEMBRANE PROTEINS BY HYDROXYL RADICAL FOOTPRINTING

ZHU YI

SCHOOL OF BIOLOGICAL SCIENCES

2012

Structural Proteomics: Elucidating *in vivo* Structural Dynamics of Integral Membrane Proteins by Hydroxyl Radical Footprinting

ZHU YI

A thesis submitted to the
School of Biological Sciences
Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2012

Acknowledgements

My utmost gratitude goes to my Ph.D. supervisor, Dr. Siu Kwan Sze for allowing me to join his group, for his research insight, expertise, and most of all, for his kindness. With his enthusiasm, his inspiration and his great efforts, he quickly became for me the role model of a successful researcher in the field.

I’m very grateful to Dr. Alex Law for his guidance and direction for the second part of my work. From him, I also learned how to do research in a wise and scientific way.

My thanks and appreciation also goes to my thesis advisory committee members, Dr. Grüber Gerhard and Dr. Sze Chun Chau, for their invaluable discussions and suggestions, which helped me a lot during the study.

I would deeply express my sincere thanks to all the previous and present members of our lab. With whom I feel like home-like warmness. From whom I have learned a lot to make me competent as a Ph.D. student in this research field. I’d also like to thank other friends from Dr. Alex Law’s lab for their kind help and generous contributions to my project. Mr Sng Jin Ern Rueben for his great help with cell culture, preparation of antibodies and protein A beads, and optimization of immunoprecitation. Dr. Xiao Zhou, Mr. Cher Toh Hui Bernard and Miss Guan Siyu for valuable discussions and suggestions on the integrin project. Ms. Manisha Cooray for her help in preparing some materials needed in the experiment.

I owe my warmest gratitude to my parents and my younger sister for their endless love and supports at every crossroads of my life. My deepest appreciation also goes to my husband, also my labmate, Guo Tiannan. Without his love and support, it would have been impossible for me to pursue my research career abroad. I feel so lucky that he is with me during my PhD study.

And last but not least, I’d like to thank Nanyang Technological University for the financial support for my doctoral study. This work was supported by Nanyang Technological University (URC: RG61/06), Ministry of Education of Singapore (AcRF: T206B3211), and the Agency for Science, Technology and Research (BMRC: 08/1/22/19/575) of Singapore.

Table of Contents

Acknowledgements I

Table of Contents III

Summary 1

List of Abbreviations 4

Chapter 1. General Introduction 7

Current structural studies of Integral Membrane Proteins 7
Protein footprinting-a complementary mass spectrometry based technique 8
Generation of hydroxyl radicals in solution 11
Fenton and Fenton-like reactions 11
Radiolysis chemistry 13
Laser photolysis of hydrogen peroxide 14
Reactions of hydroxyl radicals and amino acids side chains 16
Footprinting MS data analysis by use of bioinformatics tools 20
Objectives and the overview of the thesis 23

Chapter 2. Elucidating *in vivo* Structural Dynamics in Integral Membrane Protein by Hydroxyl Radical Footprinting 26

Introduction 26
Materials and methods 32
Materials 32
Cell culture 32
Oxidation of living E. coli bacteria under the physiological condition 32
Footprinting study of voltage gating in vivo 33
Extraction and delipidation of E. coli outer membrane proteins 34
One dimensional gel electrophoresis (SDS-PAGE) and in-gel digestion 34
LC-MS/MS analysis 35
Bioinformatics analysis, peak assignment and quantitation of oxidation 36
Calculation of solvent-accessible areas of amino acids 39
Results and discussion 40
Outer membrane proteome of E. coli 40
Identification/Detection of oxidized peptides of OmpF by LC-MS/MS analysis .. 42
Consistency of footprinting and crystallographic data for porin OmpF 50
Footprinting results proved the ion gating of OmpF porin channel in vivo 52

Footprinting results revealed the molecular basis of voltage gating of OmpF porin 55
Conclusion 58

Chapter 3. Revealing Conformational Changes of αLβ2 Integrin Directly on Cell Surfaces by Hydroxyl Radical Footprinting........59

Introduction 59
Materials and methods 63
Antibodies and reagents 63
Cell Culture and Mg^{2+}/EGTA Activation 63
Flash oxidation of Molt-4 cells using a nanosecond pulsed laser 64
Immunoprecipitation (IP) 65
In-gel digestion 66
LC-MS/MS analysis 67
Bioinformatics analysis 68
Modeling of αLβ2 integrin 70
Calculation of solvent-accessible areas of amino acids 70
Results and discussion 71
Detection of oxidized peptides of αLβ2 integrin by LC-MS/MS analysis 71
Interpretation of footprinting MS/MS data 80
Exposure of αL tailpiece upon Mg^{2+}/EGTA activation 86
Release of thigh-I-EGF 3 interface upon Mg^{2+}/EGTA activation 88
Changes to β-propeller domain of αL subunit 89
Release of both αI and βI domains of integrin αLβ2 90
The β2 genu and PSI/hybrid/I-EGF1 complex 92
Conclusion 94

Chapter 4. Hydroxyl Radical Footprinting Study of EGF-EGFR Interaction Directly on Living Cells Reveals the Role of Glycosylation on EGF Acitvation..95

Introduction 95
Materials and methods 101
Antibodies and reagents 101
Cell Culture and Transfection 102
Western blotting 103
EGF stimulation 103
Flash oxidation of living cells using a nanosecond pulsed laser 103
Immunoprecipitation 104
In-gel digestion 105

LC-MS/MS analysis 105
Bioinformatics analysis 107
Calculation of solvent-accessible areas of amino acids 108
Results and discussion 110
Detection of oxidized peptides of EGFR by LC-MS/MS analysis 110
Domain I and domain III are involved in EGF binding 114
The two EGFR subpopulations show different EGF binding affinity 118
Different glycosylation levels of specific glycosylation sites from the two EGFR subpopulations differentiates the low and high affinity EGFRs 122
Conclusion 126

Chapter 5. Conclusions **127**

References **132**

Appendix A: Publications **147**

Appendix B: Conference **148**

Summary

We have developed a novel protein footprinting technique to probe the *in vivo* structural dynamics of integral membrane proteins directly on the living cell surface. This method utilized *in situ* generation of hydroxyl radicals to oxidize and covalently modify biomolecules on living cell surface. The modified amino acid residues of the protein were identified with tandem mass spectrometry to map the solvent accessible surface that will form the footprint of *in vivo* structure of the protein. In addition, the oxidation level of each amino acid side chain was accurately measured by quantitative proteomic approaches. Since the level of oxidation for each peptide depends primarily on the solvent-accessibility of the peptide side chains which is in turn dependent on the conformation of protein, oxidation level and changes in its level for each peptide can therefore be used to determine conformation and conformational changes of the protein.

In the first study, *E. coli* bacteria cell was used as a model for the method development. We used the Fenton oxidation coupled with LC-MS/MS to investigate the structural details of the porin OmpF. Our data demonstrated the voltage gating of OmpF *in vivo* for the first time and supported the proposed mechanism that the local electrostatic field changes in the eyelet region may alter the channel to switch the porin channels.

To further improve the accuracy of the protein conformational mapping, the oxidation of protein surface must occur in sub-microsecond time frame since

conformational changes can be induced by oxidation. We applied the ultrafast laser induced hydroxyl radical footprinting technique to study the conformational change of αLβ2 integrin directly on the living mammalian cell surface in the second study. The nanosecond pulse laser photochemically generates a burst of hydroxyl radicals to snap freeze the footprints of a protein in its native biological state. We confirmed the activation and the overall extension of αLβ2 integrin heterodimer directly on the plasma membrane surface in this study.

We further extended this technique to study the EGF and EGFR interaction *in vivo*. Two EGFR subpopulations were characterized to show different EGF binding affinity, which was differentiated by different glycosylation levels at two sites. This observation helped to understand the nature of low and high affinity EGFRs.

Taken together, we have developed a novel hydroxyl radical protein footprinting method for investigating the structure-function relationship of three distinct yet common classes of membrane proteins, a porin protein (OmpF) involved in voltage gating, a heterodimer (integrin αLβ2) important in cell adhesion and signaling, and a receptor-ligand interaction (EGF-EGFR) of a typical receptor tyrosine kinase essential for cell growth and signalingsignaling. With all these findings, we can make a conclusion that the novel *in vivo* hydroxyl radical footprinting technique is a promising approach to study the structural dynamics of the integral membrane proteins directly in the native environment on the cell surfaces, and furthermore, to understand the biological function of this important class of proteins that is notoriously difficult if impossible to be studied by other structural biological methods.

List of Abbreviations

2D, two dimensional

3D, three dimensional

ACN, acetonitrile

AUC, area under the curve

CID, collision-activated dissociation

DTT, dithiothreitol

EDTA, ethylenediaminetetraacetic acid

EGF, epidermal growth factor

EGFR, epidermal growth factor receptor

EGTA, ethylene glycol tetraacetic acid

EM, electron microscopy

FA, formic acid

H/DX, hydrogen/deuterium exchange

HMW, higher molecular weight

HPLC, high performance liquid chromatography

IAA, iodoacetamide

I-EGF, integrin-epidermal growth factor fold

IMPs, integral membrane proteins

IP, immunoprecipitation

LC, liquid chromatography

LMW, lower molecular weight

LTQ, linear quadrupole ion trap

M9, minimum salts medium

MIDAS, metal ion dependent adhesion sites

MS, mass spectrometry

MS/MS, tandem mass spectrometry

MW, molecular weight

NSAF, the normalized spectral abundance factor

OMPs, outer membrane proteins

PN, unique peptide number

PSI, plexin-semaphorin-integrin domain

ROS, reactive oxygen species

RTK, receptor tyrosine kinase

SC, spectral count

SDS, sodium dodecyl sulfate

SIn, normalized spectral index

Vc, critical voltage

VWA, von Willebrand factor A

XCorr, cross correlation

XIC, extracted ion chromatogram

βTD, β tail domain

Chapter 1. General Introduction

Current structural studies of Integral Membrane Proteins

Every living cell is separated and protected from its environment by a lipid bilayer that contains integral membrane proteins (IMPs). IMPs display two membrane spanning tertiary structural motifs, i.e. α-helix bundles in the cytoplasmic membrane and β-barrels in the outer membrane (1). The overall character of these proteins is amphipathic, i.e., they possess a large hydrophobic surface area that interacts with the membrane alkyl chains, whereas loops extending into the aqueous phase on both sides are mostly hydrophilic.

About 30% of all proteins encoded by the human genome are estimated to be membrane proteins (2), and they are usually critical mediators of material and information transfer between cells and their environment and are targets for many growth factors and pharmacologically active compounds (3). Signals from binding of ligands or drugs to receptors are transduced through conformational changes in the receptors. Therefore, understanding the dynamic conformational changes in the structure of membrane proteins is essential to the understanding of many biological processes and has important implications in human health.

Doubtlessly, assessment of protein structure is critical for understanding the functions of proteins at a molecular level. Protein structures at the atomic level of resolution are determined by techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. While due to technical limitations of atomic resolution techniques, the direct determination of protein three dimensional (3D) structure, especially the 3D structures of IMPs remains one of the most challenging problems (4). IMPs are usually insoluble low-abundance proteins for which expression, purification, and crystallization are generally too inefficient to generate sufficient materials for conventional structural analysis techniques such as NMR and X-ray. Generally, their hydrophobic exterior tends to result in aggregation and precipitation in solution; it also interferes with the growth of crystals that are amenable to X-ray structure determination. As a result of these difficulties, about only 1% of all known protein structures in databases is attributed to be IMPs (2).

Additionally, IMPs in living cells constantly interact with different molecules depending on the external environment and biological states of the cell such that the conformation of IMPs is highly dynamic. Thus it remains one of the most challenging problems to reveal the correlation of the dynamic 3D structural changes in membrane proteins to their biological functions. And monitoring real-time conformational changes of membrane proteins by NMR or X-ray method is theoretically infeasible and alternative methods are needed for investigating the structural dynamic of membrane proteins.

Protein footprinting-a complementary mass spectrometry based technique

“Footprinting” refers to assays that examine ligand binding and conformational changes by determining the solvent accessibility of the backbone, bases, or side-chain structures of macromolecules through their sensitivity to chemical or enzymatic cleavage or modification reactions (5). A fundamental feature of footprinting technique is the covalent modifications of the solvent accessible sites of the macromolecule. The basis of this technique is to monitor the change in accessibility of susceptible residues as a function of relevant conformational changes (5). Protein footprinting methods, especially the modification based protein footprinting methods, have been developed to examine protein structure and conformational changes in this way.

Recent progress in mass spectrometry (MS) has enabled a novel MS-based protein footprinting technique for studying protein structure, dynamics, interactions, and function. Due to its high sensitivity and high-throughput capabilities, MS provides structural biologists with kinds of methods that are complementary to traditional techniques such as NMR and X-ray. Protein footprinting based on MS has rapidly developed over the past two decades (6). This can be achieved using traditional “bottom-up” approaches, where proteolytic peptides are analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS); or by “top-down” strategies that involve the dissociation of intact proteins in the gas phase (7-9).

Major approaches include amide hydrogen/deuterium exchange (H/DX) (10) and hydroxyl (OH) radical footprinting (11-12), while H/DX usually suffers limitations due to significant back exchange accompanying the demanding post-labeling purification. The development of the OH radical as a modification reagent for footprinting and its application in conjunction with MS were directly inspired by the development of H/DX MS methods (5). It was initially developed by Tullius' group (13-14) for the folding study of DNA/RNA molecules in solution more than twenty years ago. Several groups have since extended this method in combination with MS for the mapping of protein surface, and OH radical represents a probe that is being widely used for the oxidative labeling of proteins (15-18).

Theoretically in this method, OH radical oxidizes amino acid residues located on the protein surface, which is solvent accessible, and produce stable covalent modifications to side chains more frequently than causing backbone cleavages (19). Due to the very small size and nonspecific reactivity of OH radical, this is a random process dependent only on the solvent-accessibility and the intrinsic chemical reactivity of the amino acids. The oxidized protein is subsequently proteolytically digested and sequenced by LC-MS/MS to locate the oxidized amino acid residues. The oxidized amino acid residues provide the surface information of the protein, and thus the surface topology can be mapped. At the same time, the oxidation level of each side chain can be accurately measured by quantitative proteomic methods. Since the level of oxidation for each digested peptide depends primarily on the solvent-accessibility of the peptide side chains which is in turn dependent on the conformation of protein, oxidation level and changes in its level for each peptide can therefore be used to determine conformational changes of the protein.

OH radical footprinting approach has been shown to be a powerful complimentary technique in understanding ligand-induced conformational changes when coupled with existing structural data that are either experimentally or computationally generated (11), such as the 3D structures generated by X-ray crystallography, NMR spectroscopy, homology modeling, and *de novo* folding.

Generation of hydroxyl radicals in solution

A variety of methods are available for OH radical production, as reviewed in references (5, 20-21). Among them Fenton reaction, synchrotron radiolysis of water and laser photolysis of hydrogen peroxide (H_2O_2) are the most frequently used methods to generate OH radicals.

Fenton and Fenton-like reactions

Fenton's reaction was invented in the 1890s by H.J.H. Fenton, who discovered that ferrous iron(II) catalytically promoted the oxidation of tartaric acid by H_2O_2 (13). Later, many transition metal ions in their lower oxidation states (e.g., Fe(II), Cu(I), Ti(III), Cr(II), and Co(II)) and their complexes were found to display similar oxidative features to those of the Fenton reagents, and the metal-catalyzed oxidations were called "Fenton-like" reactions (5). These reactions are the main source of reactive oxygen species (ROS) in the cells, and are known to damage DNA *in vitro* and *in vivo* (22). The reactive species generated in Fenton reaction is the OH radical, produced as in equation 1 below (14).

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + \cdot OH + OH^- \quad (1)$$

A popular Fenton system was developed by Tullius and coworkers to generate OH radical for biological footprinting to map the protein binding sites on DNA (14, 23-24). The system includes three essential components, Fe(II)-EDTA, H_2O_2, and ascorbate. In this system, EDTA increases the solubility of the metal ions, thus allowing the reaction to be carried out at neutral pH. Chelation of Fe(II) by EDTA also prevents the transition metal ion from binding to the macromolecules being studied (14). EDTA also increases the efficiency of Fe(III)/Fe(II) to catalyze the generation of OH radicals from H_2O_2, since EDTA sequesters Fe(III) much more strongly than Fe(II) (5). The ascorbate helps reductively cycle Fe(III) back to Fe(II) to further react with the excessive H_2O_2. This catalysis allows reagent concentrations that are micromolar in Fe-EDTA and millimolar in H_2O_2 and ascorbate to be used with reaction times of several to tens of minutes (23-25). The Fe(II)-EDTA/ H_2O_2/ascorbate Fenton system has become a standard method for studying nucleic acid-ligand interaction and nucleic acid dynamic processes (23).

Based on this standard system, Brenowitz and coworkers developed a fast Fenton footprinting system allows time-resolved OH radical footprinting to be conducted using widely available three-syringe quench-flow mixers (26). It’s a Fe(II)-EDTA/ H_2O_2 system, and the reaction doesn’t need ascorbate. Under the conditions of this protocal, Fe(II)-EDTA is rapidly oxidized to Fe(III)-EDTA as OH radical is concomitantly produced, and Fe(II) is not reductively recycled by ascorbate. Results show that by using high concentrations of Fe(II)-EDTA and H_2O_2, in mM concentration, a large burst of OH radicals can be quickly generated within 2 ms.

Radiolysis chemistry

The time frame for protein folding/unfolding kinetics ranges from microseconds to milliseconds (27). Footprinting of proteins by OH radicals generated on the minute timescale such as Fenton chemistry to probe protein surfaces suffers from the uncertainty that radical reactions cause the protein to unfold, exposing residues that are solvent inaccessible in the native protein (18). Radical-induced unfolding in the super-secondary structure will give misleading oxidations and false positive results on solvent accessibility. Chance and coworkers (28) have demonstrated in an extensive set of articles a more rapid, pulsed method, in which they use synchrotron radiation to form radicals from water molecules on the millisecond timescale. The radiolysis of water is an effective method for generating OH radicals. Exciting water (at 55 mM concentration) with photons generates a water ion and an electron within nanoseconds, and the water ion further reacts with another water molecule to yield •OH , a hydronium ion, and a hydrated electron, as shown in equation 2 (22). Although the synchrotron method is fast enough to label proteins, it is hardly a routine radiation source for most researchers because of the cost.

$$H_2O \overset{h\gamma}{\rightarrow} H_2O^{+} + e^{-}_{dry} \xrightarrow{H_2O} H_3O^{+} + \cdot OH + e^{-}_{aq} \quad (2)$$

Laser photolysis of hydrogen peroxide

A relatively new pulsed-laser photochemically generated OH radical labeling approach for protein surface footprinting has been developed later (15, 18). The generation of OH radicals using pulsed-laser irradiation can be achieved in nanosecond timescale, which is even faster than the synchrotron method.

Photolysis results from the absorption of photons by H_2O_2 directly, the subsequent excitation and cleavage are represented in equation 3 (15). The former reaction competes with the diffusion-controlled self-quenching reaction which proceeds with a rate constant of 4.7×10^{9} M^{-1} S^{-1} at 25 ℃ (equation 4) (5). And it’s also reported that the half-life of OH radicals ranges from nanoseconds to microseconds, depending on the composition of the solution (29-30).

$$H_2O_2 \overset{h\gamma}{\rightarrow} 2\cdot OH \quad (3)$$

$$2\cdot OH \rightarrow H_2O_2 \quad (4)$$

Our lab reported a nanosecond laser pulse-labeling technique for protein surface mapping with MS (15). In this method, OH radicals are formed in 3-5 ns duration by a pulse Nd:YAG laser operating at 266 nm wavelength, 30 Hz. Since the half-life of OH radicals is short enough, with a single pulse of laser, formation of OH radicals as well as their subsequent oxidation and quenching happens within the timescale of the protein conformational change. Ubiquitin and apomyoglobin were used as two model proteins in this work. The final concentration of H_2O_2 for photolysis ranged from 0.3 to 1%. Immediately after irradiation, the samples were quenched by freezing in liquid nitrogen and subsequently lyophilizing in a vacuum to remove the residual H_2O_2. The protein samples were then reconstituted, digested by trypsin, and analyzed by LC-MS/MS to map the oxidized amino acid residues. The result indicated that a single laser shot can generate extensive protein surface oxidations. It was also found that the mono-oxidized ubiquitin was more susceptible to further oxidation than un-oxidized one, indicating that oxidation may induce some conformational change of proteins. Therefore it is important to perform the experiment with one single laser pulse to ensure the accuracy of the labeling approach.

Hambly and Gross (18) also developed laser flash photolysis of H_2O_2 to oxidize protein surface residues at the same time. A 17-ns KrF excimer laser operating at 248 nm with a 50 mJ/pulse energy output was applied to cleave H_2O_2 into OH radicals. Apomyoglobin was used as model protein in this work. The final concentration of H_2O_2 was as low as 0.04% when compared with the above study. The exposed sample solution was incubated with agarose-immobilized catalase to remove residual H_2O_2. The reaction duration and protein exposure to radical were tuned by adding glutamine as a free radical scavenger. The time required to remove 99% of the generated OH radicals in the absence of a radical scavenger was calculated as 20 μs based on the OH radical self-quenching reaction, while the oxidation interval was calculated to decrease to 0.5 μs in the presence of 20 mM glutamine as radical scavenger. This study emphasized the necessity of using OH radical scavengers to ensure the oxidation happens before protein conformation changes.

Reactions of hydroxyl radicals and amino acids side chains

Besides generating backbone cleavages via direct attack on the protein main chain α-carbon, OH radicals modify the side chains of proteins quite efficiently (19). OH radicals attack preferentially at the side chains of amino acids residues due to the steric hindrance of main chain α-carbon under typical reaction conditions (31). The rate of the abstraction at backbone C_α-H is at ~10^7 $M^{-1}S^{-1}$, which is much slower than the rate of reaction with most side chains at 10^8-10^9 $M^{-1}S^{-1}$ (11). Furthermore, sulfur containing, aromatic and aliphatic side chains provide targets that are more sensitive to attack than the protein backbone. Thus these events are more frequent than the protein backbone cleavage initiated by abstraction of hydrogen from the α-carbons at main chains (5).

Oxidation schemes and pathways vary among amino acid residues of different side chains, i.e., aliphatic side chains, basic side chains, acidic side chains, aromatic side chains and sulfur-containing side chains (5) and we won’t discuss the detailed chemical mechanisms here. It has been reported that there are a number of types of side chain oxidation products in OH radical protein footprinting experiments, however, not all of these oxidation products are common events as reviewed (5, 11).

In general, as to almost all the amino acids, the oxidation produces a hydroxyl group at their side chains and the formation of alcohol group as major products, resulting in +16 Da mass increases. As to amino acids with aliphatic side chains, the minor products with the formation of carbonyl (aldehyde/ketone) group, with mass increases of +14 Da, are sometimes observed along with the relevant major products. Products with multiple oxygen additions can also be generated, especially to those amino acid residues with higher reactivity, i.e., amino acids with aromatic side chains and sulfur-containing side chains. We summarized the primary radiolytic oxidation products for all amino acid side chains based on a chemical review paper (5) and listed the details in Table 1.1 below.

Oxidation Pattern	Monoisotopic	Composition	Side chain Modification	Side chain in one letter code
O16+	15.994915	O(1)	hydroxy-	DEFGHIKLMN PQRSTVWY
O14+	13.979265	H(-2)O	carbonyl	EIKLPQRV (aliphatic)
O32+	31.989829	O(2)	sulfone, sulfinic acid	CM (sulfur-containing)
			sulfinic acid	Cystine (sulfur-containing)
			hydroxy-	FWY (aromatic)
O48+	47.984744	O(3)	sulfonic acid	C, cystine (sulfur-containing)
			hydroxy-	W (aromatic)
O16-	-15.994915	O(-1)	hydroxy-	C (sulfur-containing)
O32-	-32.008456	C(-1)H(-4)OS(-1)	aldehyde	M (sulfur-containing)
O30-	-30.010565	C(-1)H(-2)O(-1)	decarboxylation	DE (acidic)
O2-	10.987705	H(-2)	carbonyl	ST (hydroxyl containing)
O43-	-43.053433	C(-1)H(-5)N(-3)O	deguanidination	R (basic)
O10-	-10.031969	C(-1)H(-2)N(-2)O(2)	ring-open	H (basic)
O5+	4.97893	C(-1)H(-1)N(-1)O(2)	ring-open	
O22-	-22.031969	C(-2)H(-2)N(-2)O(2)	ring-open	

Table 1.1 Primary products and corresponding mass changes for different amino acid side chains

The extent of OH modification of protein amino acid side chain residues in aqueous media is determined by two factors: the inherent reactivity relative to its neighboring residues and its solvent accessibility. Researchers have examined the reaction rate constants of the 20 common amino acids with OH radical at neutral pH by eliminating the steric effects, as listed in Table 1.2 (32). The relative reactivity of the 20 amino acids approximately parallels their rate of initial reactivity with OH radicals.

Amino acids	**Rate($M^{-1}s^{-1}$)**	**pH**
Cys	$3.5x10^{10}$	7.0
Trp	$1.3x10^{10}$	6.5-8.5
Tyr	$1.3x10^{10}$	7.0
Met	$8.5x10^{9}$	6-7
Phe	$6.9x10^{9}$	7-8
His	$4.8x10^{9}$	7.5
Arg	$3.5x10^{9}$	6.5-7.5
cystine	$2.1x10^{9}$	6.5
Ile	$1.8x10^{9}$	6.6
Leu	$1.7x10^{9}$	~6
Val	$8.5x10^{8}$	6.9
Pro	$6.5x10^{8}$	6.8
Gln	5.4×10^{8}	6.0
Thr	$5.1x10^{8}$	6.6
Lys	$3.5x10^{8}$	6.6
Ser	$3.2x10^{8}$	~6
Glu	$2.3x10^{8}$	6.5
Ala	$7.7x10^{7}$	5.8
Asp	$7.5x10^{7}$	6.9
Asn	$4.9x10^{7}$	6.6
Gly	$1.7x10^{7}$	5.9

Table 1.2. Rate constants for reaction of amino acids with OH radical in aqueous media.

Chance M. R.’s group further did extensive studies and found that a reactivity order of Cys, cystine>Met>Trp>Tyr>Phe>His>Leu, Ile>Arg, Lys, Val>Ser, Thr, Pro>Gln, Glu>Asp, Asn>Ala>Gly in MS based experiments, when the oxidation source is generated by ionizing radiation (11, 17, 33). This fundamental finding provides researchers in this field very insightful clues and guidelines when investigating the MS data generated.

Based on these data together, Ala and Gly are too unreactive to be useful probes in typical experiments, and Asp and Asn are also less likely. Although Ser and

Thr are more reactive, their oxidation products are not easily detectible, and only trival amounts of +16 Da and -2 Da product were found (5). Therefore, it seems that not all of the 20 amino acid side chains can serve as the useful footprinting probes in experiments.

Footprinting MS data analysis by use of bioinformatics tools

MS data analysis is the final step of any proteomic workflows. Several software applications have been developed to date to automate this process. The most common programs used are: SEQUEST (34), Mascot (35), X!Tandem (36) and OMSSA (37). Generally, these programs generate theoretical MS/MS spectra based on enzyme cleavage and peptide fragmentation of the target protein sequences. The experimental spectra are then searched against the theoretical database and assigned a score. The match with the highest score is generally the most probable assignment. These programs, however, become slow and lose sensitivity when allowing peptide modifications because enumerating all possible modifications in the database makes the database prohibitively large (38-40).

MS data analysis is essential in protein footprinting study because only by this way people can interpret the seemly meaningless MS spectra readouts into meaningful protein structural features. High-resolution footprinting necessitates not only the identification of the sites of oxidation, but also the quantitation of the level of oxidation at each site. To analyze OH footprinting data, oxidative modifications of peptides on specific residues need to be built in first for database searching, in the meantime, the intensities of modified and unmodified peptide species need to be calculated and extracted. However, the interpretation of high volumes of data resulting from such experiments represents the biggest bottleneck for the overall experiments. As seen in Table 1.1, the majority of amino acid side chains exhibit variable kinds of oxidation and multiple oxidation states when reacting with OH radicals, and this complicated oxidation pattern results in a large number of possible combinations for peptide matches. Besides, this large number of combinations complicates the data interpretation and can lead to false spectral matches. This introduces great challenge in data analysis by the softwares above. Manual processing of protein footprinting data is also impossible as tens of thousands of mass spectra can be generated in one LC-MS/MS run.

Therefore several programs such as InsPect (38), ByOnic (39) and ProtMapMS (40) have been developed to improve database searching allowing for a large number of modifications, including the multiple oxidative modifications. InsPect adopts *de novo* sequencing, and uses peptide sequence tags as efficient filters that reduce the size of the database by a few orders of magnitude while retaining the correct peptide with high probability. It offers speed and robustness while requires high quality spectra, hence, is less sensitive to low-abundance peptides. ByOnic also uses a small amount of *de novo* analysis to identify likely *b*- and *y*- ion peaks that can then be used to extract candidate peptides from the database, with the number of candidate tunable to fit a computing capacity. The hybrid approach does not require a ladder of spectral peaks and can identify spectra containing modified peptides or peptide mixtures. It is reported to be more sensitive than *de novo* sequence tagging. ProtMapMS is an integrated suite of algorithms aimed at the automated analysis of protein structure studies using synchrotron radiolysis footprinting technique. It allows for localizing modification sites of the oxidized peptides using multistage MS data and quantification of the oxidation extent as a function of exposure time to synchrotron radiation. Although none of these programs is free access and none can be applied directly to our own studies and a number of labor intensive steps are required followed by the database search, they provide us clues and guidelines on how to write our own in-house program to analyze oxidative protein footprinting data.

Oxidized peptides can be detected (but not fully characterized) by inspecting the peptide extracted ion chromatogram (XIC) and its corresponding full mass spectrum (11). Normally the oxidized peptides which are less hydrophobic due to the addition of oxygen will be eluted slightly earlier than the unoxidized counterparts when peptides are separated by reverse-phase HPLC. Thereby, abundance of unoxidized peptide and its relevant oxidized counterparts can be obtained. It’s a classical and accurate method widely used in label free quantification in the past years. However, the disadvantage is that this ideal separation is somewhat challenging to be achieved because of the complexity of biological samples. Other approaches including spectral count (SC, number of MS/MS spectra per peptide), the normalized spectral abundance factor (NSAF), exponentially modified protein abundance index (emPAI), unique peptide number (PN) are also used in label free quantification, but none of them are universally used (41-42).

Normalized spectral index (SIn) quantification was recently presented as an accurate method of label-free quantitation, which improved spectral counting by incorporating the unique peptide number, and the intensities of peptide MS/MS fragment ions into the calculation of protein abundance (42). Because multiple replicate MS measurements of a sample are always required in the analysis of large-scale shotgun proteomic profiling to achieve statistically significant comprehensiveness in protein identifications, while the inherent biases and variations in such data set induce challenges for quantitative comparative analysis. SIn has been proved to largely eliminate variances between replicate MS measurements, and permit quantitative reproducibility in replicate MS measurements of the same and distinct samples. It has been proved to accurately predict protein abundance more often than the methods above (42).

Objectives and the overview of the thesis

Similar to other analytical techniques, the application of MS to membrane protein continues to be much less common than studies on their water-soluble counterparts (43), and researchers commonly resort to surfactant-solubilized proteins. And also because *in vivo* studies are often not feasible, current surface mapping techniques are therefore mostly limited to purified proteins which are stable in solution, and are not amenable to the conformation determination of most IMPs or their complexes. As a result of these difficulties, the study of structural dynamics of IMP remains elusive.

We have developed a novel adaptation of oxidative footprinting coupled with MS technique to study the *in vivo* structural dynamics of IMPs. In this method, generally, cells are first exposed to *in situ* generated OH radicals, and the membrane proteins on the cell surfaces are oxidized. The targeted protein is purified, followed by proteolysis to generate peptides for analysis and identification by LC-MS/MS and bioinformatics analysis. As discussed above, OH radicals can be generated by kinds of methods including Fenton chemistry, laser photolysis of hydrogen peroxide, and radiolysis of water with a high-energy X-ray synchrotron beam, etc. In the first study as illustrated in chapter 2, Fenton chemistry was adopted to study the *in vivo* voltage gating of the matrix porin OmpF. *E. coli* cells were exposed to OH radicals generated from an *in situ* Fenton reaction between H_2O_2 and Fe-bound ethylenediaminetetraacetic acid (EDTA) for oxidation. We adopted Fenton chemistry in this study because the 266 nm laser beam cannot penetrate the yellowish and cloudy *E. coli* bacterial culture, the laser photolysis method is not applicable. In the second study as illustrated in chapter 3, we tried to implement the application of the nanosecond pulse laser-induced photolytic hydroxide radical footprinting method developed previously in our lab (15) to study the conformational change of native αLβ2 integrin on plasma membrane of the surface of a mammalian cell. In this method, a high concentration of OH radicals is generated in 5~7 ns via photo-dissociation of a H_2O_2 solution by a pulse Nd:YAG laser. In the last study as illustrated in chapter 4, we further extended the application of the nanosecond laser photochemical footprinting method to study the protein complex interaction *in vivo*, using EGFR as a model protein because EGF binding of EGFR has been well and extensively characterized. Finally we summarized the results of the studies in the conclusions chapter.

Chapter 2. Elucidating *in vivo* Structural Dynamics in Integral Membrane Protein by Hydroxyl Radical Footprinting

Introduction

One of the most challenging problems in biological sciences is the correlation of the dynamic three dimensional (3D) structural changes in membrane proteins to their biological functions. Comprising about 30% of the human proteome, membrane proteins are critical mediators of material and information transfer between cells and their environment and are targets for many growth factors and pharmacologically active compounds. Signals from binding of ligands or drugs to receptors are transduced through conformational changes in the receptors. Therefore, understanding the dynamic conformational changes in the structure of membrane proteins is essential to the understanding of many biological processes and has important implications in human health.

Although some methodologies including NMR and X-ray have been developed to study protein structures, the determination of the structures of IMPs remains one of the most challenging problems in biological sciences (4). IMPs are usually insoluble low-abundance proteins for which expression, purification, and crystallization are generally too inefficient to generate sufficient materials for conventional structural analysis techniques such as NMR and X-ray. Additionally, IMPs in living cells are constantly interacting with different molecules depending on the external environment and biological states of the cell such that the conformation of IMPs is highly dynamic. Hence monitoring real-time conformational changes in these proteins by NMR and X-ray methods is infeasible and alternative methods are needed.

Recent progress in MS has enabled a novel MS-based protein oxidative footprinting technique to determine structural information by mapping of oxidation induced by OH radicals. This method is an adaptation of the OH radical footprinting first developed by Tullius' group (13-14) for the folding study of DNA/RNA molecules in solution. Several groups have since extended this method in combination with MS for the mapping of protein surface (15-16, 18, 44-45). In these studies, OH radicals oxidize amino acid residues located on the protein surface, and produce stable covalent modifications to side chains without causing backbone cleavages. Due to the very small size and nonspecific activity of OH radicals, this is a random process dependent only on the solvent-accessible surface and the chemical properties of the exposed amino acids. It has been reported that there are about twelve possible types of side-chain oxidation products in protein-footprinting experiments (5), however, not all of these oxidation products are common events as reviewed. The most common event results in formation of alcohol group for almost all the amino acid residues with mass increases of +16 Da. Another one is the formation of aldehyde/ketone group for eight residues V, I, L, K, R, P, E and Q with mass increases of +14 Da (5). Others include +32 Da on C, M, W, F, Y; +48 Da on C, W; +5 Da, -10 Da, -22 Da, and -23 Da on H; -30 Da on D and E; -16 Da on C; -32 Da on M; and -2 Da on S and T (5).

The oxidized protein is subsequently sequenced by tandem mass spectrometry to locate the oxidized amino acid residues. The oxidized amino acid residues provide the surface information of the protein, and thus the surface topology can be mapped. At the same time, the oxidation level of each side chain can be accurately measured by quantitative liquid chromatography-coupled MS. Since the level of oxidation for each tryptic peptide depends primarily on the solvent-accessibility of the peptide side chains which is in turn dependent on the conformation of protein, oxidation level and changes in its level for each peptide can therefore be used to determine conformation and conformational changes of the protein. This approach has been shown to be a powerful technique in understanding ligand-induced conformational changes when coupled with existing structural data that are either experimentally or computationally generated (11).

Despite recent advances, most current surface mapping techniques still require purified proteins and are therefore not amenable to the conformation determination of most IMPs or their complexes. In particular, the current techniques still cannot determine real-time *in vivo* structural changes of IMPs which are important in understanding the functions of IMPs. Therefore, the structural dynamics of IMP remains elusive.

IMPs display two membrane spanning tertiary structural motifs, i.e. α-helix bundles in the cytoplasmic membrane and β-barrels in the outer membrane (1**).** In Gram negative bacteria, the outer membrane forms a protective permeability barrier around the cells and serves as a molecular filter for hydrophilic substances (46). To facilitate this, channel-forming proteins are embedded in the outer membrane to mediate the transport of nutrients and ions across the membrane into the periplasm. The most abundant proteins in this class are the porins that form aqueous passive channels with a physical diameter about 1 nm for the transport of ions and relatively small polar molecules across the outer membrane of bacteria (47-49). Porins form β-barrels with 16 or 18 membrane-spanning anti-parallel β-strands which are connected by short turns on the periplasmic side named T1, T2 etc, and long loops on the extracellular side named L1, L2 etc (50-51). In all porins, the constriction at the barrel centre is formed by an inserted loop L3 (47, 50, 52), which is not exposed to the cell surface but folds back into the channel and contributes significantly to the permeability of the pore. Functional porins are homotrimers of these β-barrel subunits with each subunit producing a channel in the outer membrane and the trimer therefore contains three channels (48). Currently known porins fall into two distinct groups: 16-stranded general diffusion porins such as the matrix porin OmpF which transports ions and small molecules (<600Da) without much selectivity and 18-stranded specific porins such as maltoporin which has a precise selectivity for a defined substrate (49, 51, 53-56). General porin trimers exhibit symmetrical or asymmetrical voltage gating when reconstituted into planar lipid bilayers in electrophysiological studies (57-61). By measuring the ion conductance *in vitro*, porin trimers were found to be able to exhibit at least two functional states—an open and a closed one—in response to changes in the transmembrane potential difference, as is named as ‘voltage gating’, and the voltage above which the porin channel will be closed is called critical voltage, Vc. While existing data generally support voltage gating *in vitro*, there is no data to support voltage gating *in vivo* (59, 61).

We describe here a novel adaptation of oxidative footprinting and MS technique to study *in vivo* conformational changes of IMPs in living cells. We used *E. coli* as a model to study the structural dynamics of the outer membrane proteins, with an emphasis on the matrix porin OmpF. Considering the yellowish and cloudy *E. coli* bacteria culture would interfere with the laser photolysis of H_2O_2, because H_2O_2 only absorbs laser with the wavelength around ~250nm and this cloudy media would inhibit the laser penetration (15), we adopted Fenton reaction and not the pulsed laser photolysis method to oxidize the living bacteria. Fenton oxidation is not guaranteed for many proteins especially the highly dynamic ones because this method needs relatively longer incubation time to several minutes for the reagents to generate OH radicals (5, 13, 15, 18). This time scale is too long because most of the conformational changes of macromolecules take place only in the timescale of milliseconds to seconds (11). However, Fenton chemistry is suitable to study the outer membrane porins because they are of extremely stable structures which are restricted by both the intra-molecular hydrogen bond and the hydrophobic interaction with the lipid bilayer wall, and the great conformational change of the protein is considered unlikely (48, 54).

In this method, living *E. coli* cells were first exposed to OH radicals generated from an in situ Fenton reaction between H_2O_2 and Fe-bound ethylenediaminetetraacetic acid (EDTA). The outer membrane proteins which have large solvent accessible surface area would be preferentially oxidized by the OH radicals. The outer membrane proteins were then isolated and enriched from the cell lysate. The porin OmpF was subsequently purified by 1D SDS-PAGE electrophoresis followed by proteolysis to generate peptides for analysis and identification by LC-MS/MS and bioinformatics analysis. The structural information of porin OmpF as interpreted from our oxidative footprinting and MS study was in agreement with the X-ray crystallography structure (54). Highly reproducible oxidized amino acid residues were widely distributed in the external loops, transmembrane β-strands and periplasmic turnings of the protein and all of them were validated as solvent accessible according to the X-ray structure of porin OmpF.

We then extended this method to probe for voltage gating of porin OmpF under different conditioned circumstances to assess the robustness of its application in complicated biological systems. Here, *E coli* bacteria were stimulated first with either a low ionic strength solution or a low pH buffer after harvest, and exposed to OH radicals generated by Fenton reagents in solution for three minutes. Our data showed that the extent of oxidation of two polypeptides from extracellular loops remained constant independent of conditions that induced either channel-closing or -opening. However, we observed that oxidation of peptides in β-stranded areas deep inside the porin channel was reduced when the environment changed from one that induced channel-opening to another that induced channel-closing. These data not only supported the voltage gating of porin OmpF *in vivo* but also revealed the molecular basis underlying voltage gating of porin OmpF.

Materials and methods

Materials

HPLC grade solvents were purchased from J.T. Baker (Phillipsburg, NJ, USA). Complete Protease Inhibitor Cocktail Tablets were from Roche Diagnostics (Indianapolis, IN, USA). All remaining laboratory chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless stated otherwise.

Cell culture

E. coli K-12 strain WT31 from ATCC were grown aerobically in 450mL LB broth medium at 37 ℃ with shaking at 225 r.p.m. in 2L baffled flasks (Nalgene, NY, USA) from a 1:100 dilution of overnight culture until mid-log phase (OD600 = 0.9–1.1). Cells were collected by centrifugation at 3000 g, 4 ℃ for 10 min and washed three times in fresh minimum salts medium M9 (pH 7.4). The total amount of cells from one flask culture was aliquoted into three fractions and each fraction was resuspended into 15mL of fresh M9 medium, pH 7.4, for subsequent experiments.

Oxidation of living E. coli bacteria under the physiological condition

For the oxidation study, *E. coli* cells were divided into two groups, control group and oxidation group, with each group containing 3 fractions of the suspension of cells as mentioned above. The oxidation procedure was performed as described (24, 26) with some modifications. Briefly, in the oxidation group, the cells were incubated with a freshly prepared Fenton solution of Fe-EDTA (final concentration: 10mM $Fe(NH_4)_2(SO_4)_2$ and 25mM Na_2EDTA) in M9 (pH 7.4) at 37 ℃ for 10min. A certain amount of 3% H_2O_2 was then added into the cell suspension to a final concentration 0.3% and incubated for 3 min to allow the oxidation reaction which was initiated by OH radicals generated from H_2O_2 and Fe(II). The oxidation was stopped by removing the oxidative solution (OH radicals) using centrifugation at 5000 g, 4 ℃ for 5 min, and further quenched with washing with 50 mM ice-cold Tris-HCl (pH 8.0) for three times (5, 15). Experiments were done in duplicates.

Footprinting study of voltage gating in vivo

In this study, cells in mid-log growth phase were harvested, centrifuged and washed as described above. They were stimulated by resuspending in different buffers for half an hour before being exposed to OH radicals generated by Fenton reagents. The different buffers were: a low pH buffer (300 mM KCl, 20 mM citric acid, pH 3.0), a low ionic strength buffer (5 mM KCl, pH 7.4) and a low pH and low ionic strength (5 mM KCl, 20 mM citric acid, pH 3.0). All three buffers chosen here are not OH radical scavengers and are suitable for OH radicals initiated oxidation reaction (5).

Extraction and delipidation of E. coli outer membrane proteins

The enrichment of outer membrane proteins was carried out as described (1, 62) with some modifications. Briefly, cell pellet was resuspended in 10 mL ice cold 50 mM Tris-HCl (pH 8.0) supplemented with 100 mM Dithiothreitol (DTT) and Complete Protease Inhibitor Cocktail Tablets. The cells were ruptured by liquid nitrogen cracking with the unbroken cells being removed by centrifugation at 4 ℃, 5000 g for 15 min. The supernatant was collected and lyophilized to dryness, and resuspended in 50 mL ice cold 0.1 M sodium carbonate (pH 11) and incubated at 4 ℃ for 1 h with gentle stirring. The sodium carbonate treated membranes were then pelleted down by ultracentrifugation at 125,000 g, 4 ℃ for 1h in an Optima^{TM} L-100 XP ultracentrifuge (Beckman Coulter, Inc., Fullerton, CA, USA). The membrane pellet was washed twice with Milli-Q water using centrifugation at 125,000 g, 4 ℃ for 30 min.

Delipidation of membrane proteins was achieved via protein precipitation with cold acetone at -20 ℃ for 2 h. After precipitation, the protein pelletes were redissolved in 1% sodium dodecyl sulfate (SDS) and protein concentration was determined with a Bicinchoninic Acid Kit (Sigma-Aldrich).

One dimensional gel electrophoresis (SDS-PAGE) and in-gel digestion

150 µg of proteins from different treatment regimes were dissolved in 60 µL 1X LDS Nalgene gel-loading buffer (Bio-Rad Laboratories, Hercules, CA, USA)

supplemented with 100 mM DTT, and was denatured by boiling at 95 ℃ for 5 min. 1D gel electrophoresis was carried out using 12% SDS-PAGE. The gel was stained by Coomassie blue staining method.

For the identification of the outer membrane proteome of *E. coli,* a whole lane of the control group was cut off into 10 bands according to the molecular weight (MW) of the protein located, followed by reduction with 10 mM DTT in 25 mM ammonium bicarbonate at 56 ℃ for 1 h, alkylation with 55 mM iodoacetamide (IAA) in 25 mM ammonium bicarbonate solution at room temperature in dark for 45 minutes and digested by sequencing grade modified trypsin (Promega Corporation, Madison, WI. USA) digestion in a 1:100 (trypsin: protein) mass ratio at 37 ℃ overnight.

For the specific study of porin OmpF, the 37 kDa (OmpF) gel bands from each group were excised, followed by DTT reduction, IAA alkylation, and tryptic digestion as described above. Tryptic digested peptides were then analyzed by LC-MS/MS.

LC-MS/MS analysis

The LC-MS/MS analysis was done in LTQ Orbitrap (Thermo Fisher Scientific Inc., Bremen, Germany) as described previously with some modifications (15, 63). The tryptic peptides were reconstituted to 100 μl 0.1% formic acid (FA) in HPLC water. The samples were then injected by an autosampler and on-line desalted in a Zorbax peptide trap (Agilent, Pola Alto, CA). The peptides were separated by a HPLC (Shimadzu, Kyoto, Japan) using a home-packed nanobored C18 column (75 μm i.d x 10 cm, 5 μm particles) directly into a picofrit nanospray tip (New Objective, Wubrun, MA, USA), operating at a flow rate of about 200 nl/min after a splitter. Buffer A (99.9%H_2O/0.1%FA) and buffer B (99.9%ACN/0.1%FA) were used for LC gradient. The 90 min gradient was ramped from 5% ACN to 30% ACN in 65 min, then to 60% ACN in 10 min, to 80% ACN over 2 min and kept at 80% ACN for 3 min and ramped back to 5% ACN for last 10 min. The LTQ was operated in a data-dependent mode by performing MS/MS scans for the maximum ten most intense peaks (ion selection threshold of 500 counts) from each Orbitrap MS scan. Samples were injected into the MS with an electrospray potential of 1.8 kV without sheath or auxiliary gas flow. Ion transfer tube temperature was 180 ℃ and collision gas pressure was 0.85 mTorr. The MS scan range was 350–2000 m/z. The AGC target for the FT analyzer was set at 5e+05 and precursor ion charge state screening was activated in the Orbitrap MS scan. Dynamic exclusion was activated for the MS/MS scan, with a repeat count of 1 and exclusion duration of 20 s. Isolation width was 2 Da, and default charge state was 5. For CAD fragmentation, normalized collision energy was set to 35%, activation Q was set to 0.25, and activation time 30 ms. Spectra were acquired in centroid format in raw data files with XCalibur (version 2.0 SR2).

Bioinformatics analysis, peak assignment and quantitation of oxidation

MS/MS data for *E. coli* outer membrane proteome was analyzed by MASCOT (version 2.2.04, Matrix Sciences, London, UK) against NCBI *E.Coli* database (version 4, including 54,969 entries). Carbamidomethyled cysteine was set as a fixed modification, while oxidation at methionine residue was set as variable modification. Peptide tolerance was 8 ppm. Full enzymatic cleavage by trypsin was selected, with maximum 2 missed cleavages allowed.

MS/MS data for footprinting experiments were submitted to SEQUEST protein sequence database search engine for peptide/protein identification against an in-house *E. coli* database. For SEQUEST protein ID, the raw data was batch searched with Bioworks Browser (version 3.3, Thermo Fisher Scientific Inc.). Maximum of 8 post-translational modifications were allowed for single peptide. Peptide tolerance was 8 ppm, and fragment ions tolerance was 1AMU.

Full enzymatic cleavage by trypsin was selected, with maximum 2 missed cleavages allowed. As discussed in Chapter 1, the oxidation of amino acids by OH radical generally results in the formation of alcohol group as a major product for almost all the amino acid residues, and the formation of aldehyde/ketone group as a minor product for mainly the aliphatic amino acid side chains (Table 1.1). Therefore in this study, a mass increase of +16 Da group for all the amino acid residues and +14 Da for eight residues V, I, L, K, R, P, E and Q were included as variable modifications. For OmpF where the aromatic amino acids and acidic ones play important roles in both its structural and functional features, we took two more variable modification types into account. One caused a mass increase of +32 Da by adding two oxygen atoms to the side chain of five residues, C, F, M, W, Y and the other was specific to the side chain of two residues D and E, resulting in a mass shift of -30 Da (Table

1.1). Because OmpF has no cysteine residue in its sequence, we did not take carbamidomethylated cysteine into account.

Matched peptides were filtered by requiring peptide probability cutoff of 0.05, and XCorr cutoffs of 1.0, 2.0, and 2.5 for +1, +2, and +3 charged peptides, respectively. Quantitation of peptide was based on the integrated area of the extracted ion chromatogram (XIC) of precursor ions with mass tolerance of 0.05 Da by Bioworks Browser software. The XIC of precursor ions were confirmed by manual extraction from the MS raw data with mass tolerance of 5 ppm of the calculated m/z values.

A label free approach was employed here to quantify the oxidation levels of the oxidized peptide products of each tryptic peptide arising from Fenton reaction with OH radicals. As shown in the equation 5 below, by integrating the XIC chromatographic peak areas of both the un-oxidized peptide ions (A_0) and its oxidized counterparts (A_i, i=1, 2…), we could get the relative abundance of each species and to quantify the oxidation level.

$$Oxi_{(i)}\% = \frac{A_i}{A_0 + \sum_i A_i} \times 100\% \quad (5)$$

The data were also analysed by ByOnic (64), which includes an option for OH radical surface mapping (44). This option simultaneously turns on all the twelve possible modifications observed in OH radical mapping (5), as listed in the introduction part above. Like SEQUEST, ByOnic sets a limit on the number of modifications per peptide, but the limit depends upon the residue content. Any reactive residue (C, M, W, F, Y, and H) can be modified, but at most two non-reactive residues per peptide may be modified. In ByOnic, we also enabled the following non-oxidative modifications: pyro-glu on N-terminal E and Q, deamidated N and Q, and (on some of the samples) sodiation on any one residue. Precursor mass tolerance was 8 ppm and fragment tolerance, after recalibration using a quadratic regression curve computed from initial confident identifications, was 0.5 Da.

Calculation of solvent-accessible areas of amino acids

To determine whether oxidized residues were solvent accessible, exact accessible surface area of molecular surface for macromolecules was calculated by the program Surface Racer 4.0 (65) based on the X-ray crystal structural data of OmpF (1OPF).

Results and discussion

Outer membrane proteome of E. coli

1D SDS-PAGE method was utilized to separate the OMPs of *E. coli*. The main advantages of MS-based protein structure determination are the superb sensitivity and high throughput. Using 150 µg sample enriched in *E. coli* outer membrane proteins in the control group, we identified over 600 proteins. Of these, about one hundred proteins were annotated as OMPs. The top ten ranking OMPs based upon Mowse scores are listed in Table 2.1. Some proteins of typical β-barrel structure have been well documented previously. OmpF is a general β-barrel porin for transportation of polar solutes, while LamB, also named maltoporin, is a specific porin which facilitates movement of maltodextrins and other maltooligosaccharides. OmpA plays a fundamental role for the integrity of the bacterial cell surface. It is composed of an N-terminal membrane embedded domain of 170 residues, serving as a membrane anchor with 8-stranded β-barrel structure and a C-terminal 155 residue domain, which is located in the periplasmic space and has been proposed to interact specifically with the peptidoglycan layer (48, 66). TolC is also of this typical structure, and it functions as an energy dependent transporter (48, 67).

Index	Protein	Mowse score	Total sequence of amino acids	Sequence identified by MS	Sequence coverage (%)
1	P02931|OMPF_ECOLI	64235	362	334	92.27
2	P0A910|OMPA_ECOLI	24378	346	309	89.31
3	P69776|LPP_ECOLI	18111	78	75	96.15
4	P0A940|YAET_ECOLI	9080	810	637	78.64
5	P02930|TOLC_ECOLI	4824	493	375	76.06
6	P31554|OSTA_ECOLI	4430	784	500	63.78
7	P02943|LAMB_ECOLI	2595	446	325	72.86
8	P0A927|TSX_ECOLI	1705	294	212	72.11
9	P77774|YFGL_ECOLI	1691	392	322	82.14
10	P10384|FADL_ECOLI	1042	446	295	66.14

Table 2.1. **OMPs with their Mowse scores ranked in top 10 are listed.** The sequence coverage of each protein was calculated according to the ratio of the amino acids sequence identified by MS analysis and the whole sequence of the protein.

Although using a much lower amount of starting material (150 μg of membrane proteome) as compared with X-ray and NMR method, for which milligrams of pure single protein would be needed, the average sequence coverage of the top 10 OMPs identified was as high as 79%, indicating the possibility of using MS-based technique to elucidate the protein structure for these proteins. The protein OmpF, a structurally and functionally well characterized protein, was identified with higher than 90% sequence coverage and was thus chosen as our model protein for further structural analysis. A total of nineteen peptides which consists of 92% sequence coverage of OmpF were repeatedly identified by LC-MS/MS (Table 2.2). Several tryptic peptides (in grey shadow in Table 2.2) were not detected, which contributed to a more than 6% sequence loss. These peptides were relatively short, most of them consisting of only two to four amino acids, which might be missed by the precursor ion scan from 350-2000 Da. In addition, short tryptic peptides are relatively hydrophilic and results in the poor retention in the C18 reverse phase peptide trap or column.

Sequence	Peptide
23-28	AEIYNK
29-38	DGNKVDLYGK
39-47	AVGLHYFSK
48-64	GNGENSYGGNGDMTYAR
65-68	LGFK
69-102	GETQINSDLTGYGQWEYNFQGNNSEGADAQTGNK
103-104	TR
105-111	LAFAGLK
112-122	YADVGSFDYGR
123-154	NYGVVYDALGYTDMLPEFGGDTAYSDDFFVGR
155-162	VGGVATYR
163-182	NSNFFGLVDGLNFAVQYLGK
183-185	NER
186-189	DTAR
190-218	RSNGDGVGGSISYEYEGFGIVGAYGAADR
219-231	TNLQEAQPLGNGK
232-241	KAEQWATGLK
242-257	YDANNIYLAANYGETR
258-265	*NATPITNK*
266-275	FTNTSGFANK
276-299	TQDVLLVAQYQFDFGLRPSIAYTK
300-303	SKAK
304-327	DVEGIGDVDLVNYFEVGATYYFNK
328-345	NMSTYVDYIINQIDSDNK
346-362	LGVGSDDTVAVGIVYQF

Table 2.2. Tryptic peptides of OmpF. The peptides labeled in grey shadow were the short ones that could not be detected by LC-MS/MS. Nineteen peptides were clearly identified, covering 92% of the OmpF amino-acid sequence.

Identification/Detection of oxidized peptides of OmpF by LC-MS/MS analysis

The schematic representation of the *in-vivo* membrane structure probing process was shown in Figure 2.1. The overall schematic diagram of the LC-MS/MS setup was shown in Figure 2.2.

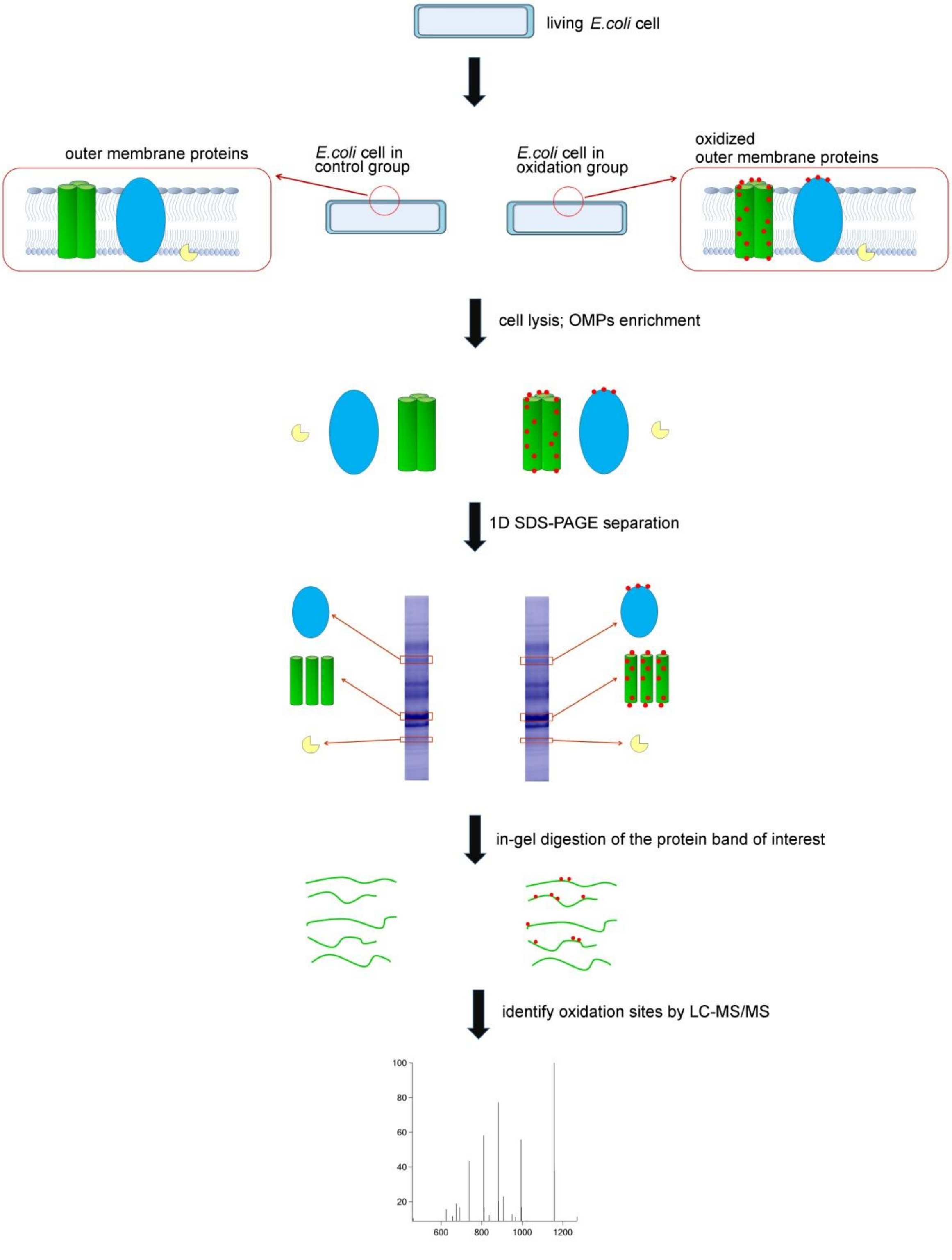

living E.coli cell
outer membrane proteins
E.coli cell in control group
E.coli cell in oxidation group
oxidized outer membrane proteins
cell lysis; OMPs enrichment
1D SDS-PAGE separation
in-gel digestion of the protein band of interest
identify oxidation sites by LC-MS/MS
100
80
60
40
20
600
800
1000
1200

Figure 2.1. The schematic representation of the footprinting process of outer membrane proteins in living *E.coli* cells, as described in the experimental procedures.

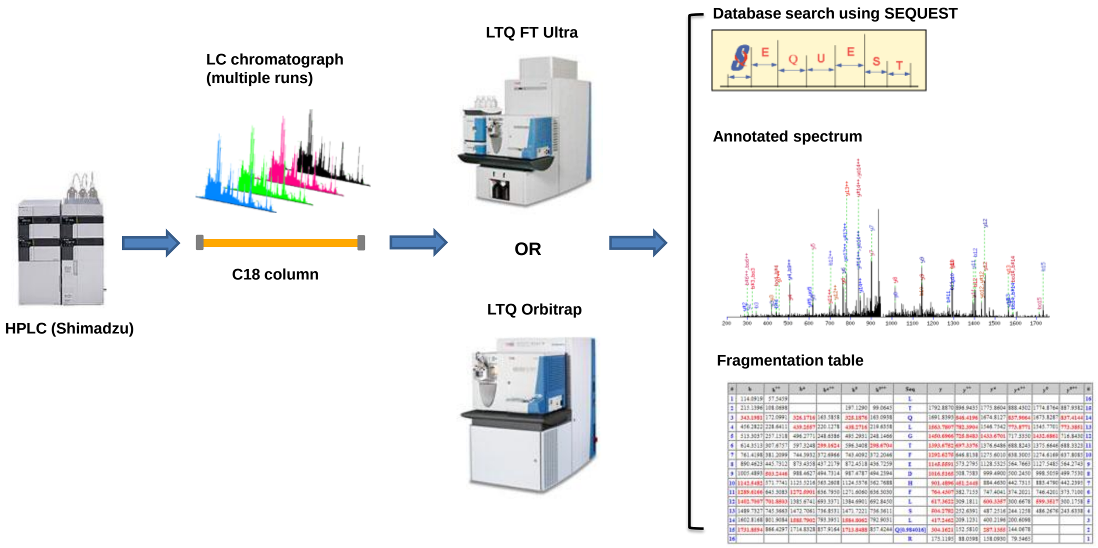

Database search using SEQUEST
S E Q U E S T
LTQ FT Ultra
LC chromatograph
(multiple runs)
Annotated spectrum
C18 column
OR
HPLC (Shimadzu)
LTQ Orbitrap
Fragmentation table

Figure 2.2. The schematic overview of the LC-MS/MS setup. The detailed setting and parameters were demonstrated in the ‘materials and methods’ part. The mass spectrometry used in this study was LTQ Orbitrap, and TLQ FT Ultra was applied in the next two studies for the same usage.

Tryptic digested peptides of OmpF were analyzed quantitatively by LC-MS/MS. For the positive identification of modification sites, two different database-search tools, SEQUEST and ByOnic were used. For SEQUEST database searching, we included four mass modifications +14 Da, +16 Da, +32 Da and -30 Da as variable modifications. To minimize false positives, only common modifications identified by both programmes were counted, even though ByOnic considers a suite of 12 possible oxidative modifications, including some quite uncommon ones.

The six most reactive residues, C, M, W, F, Y, and H, oxidize very readily, and thus serve as both positive and negative probes: a cysteine or methionine that does not react is probably not exposed. The less reactive residues serve mainly as positive probes, because even highly exposed residues may oxidize at too low a rate to gather meaningful statistics (5).

Oxidized peptides can be detected (but not fully characterized) by inspecting the peptide extracted ion chromatogram (XIC) and its corresponding full mass spectrum. Oxidized peptides are generally less hydrophobic due to the addition of oxygen and are expected to elute slightly earlier than the unoxidized counterparts when peptides are separated by reverse-phase HPLC. Thereby, abundance of oxidized and unoxidized peptides derived from different regions in OmpF is obtained.

Identification of modification sites was further confirmed by manually checking the mass spectra. Special attention was paid to the *b*-type fragment ions, which retain the original N terminus of the peptides, and *y*-type fragment ions, which retain the original C terminus. A *b*-type ion that retains only the N-terminal residue would be noted as *b1* and its sister ion with two residues would be *b2* and so on, same as y-type ions *y1*, *y2* and so on from the C terminus. The mass difference between b2 and b1 provided information on the identity of the second residue from the N terminus. Thus, the sequence of the fragmented peptide can be inferred to the extent that sufficient fragment ions are observed. Exact placement of a +16 Da modification is sometimes quite difficult with either SEQUEST or ByOnic, and even with manual analysis, and some spectra may in fact represent a mix of two or more isoforms, the same peptide with a single +16 Da modification in several different sites. Exact placements are more reliable if found independently by two different protein sequence database search engines, and confirmed by manual inspection.

By this way, from the nineteen tryptic peptides of OmpF identified, six peptides and nineteen amino acids from these six peptides were reproducibly detected to be oxidized in the oxidation group both by SEQUEST and ByOnic searching methods, as shown in Table 2.3 and Table 2.4 below. Table 2.3 listed both the oxidation pattern and the corresponding oxidation extent of each peptide derived from XIC. The extent of oxidation in the modified peptides was averaged from two independent experiments. In Figure 2.3, the extracted ion chromatograms of oxidized and unoxidized doubly charged peptide ions from tryptic peptide 328-345 'NMSTYVDYIINQIDSQNK' are shown. The chromatographic areas of peptide ions were integrated and used for quantitation of the extent of modification. In Figure 2.4, the MS/MS spectrum of the modified peptide sequence 'NM*STYVDY*IINQIDSDNK' was shown. Oxidation sites were confirmed by inspection of the *b* and *y* ion series. The different m/z value between y17 and y16 ions clearly shows a +16 Da oxidation to the 329M. In the same way, the gap between b9 and b7 ions, as well as the difference between y11 and y 10 ions, indicates the +16 oxidation to the 335Y, but not other residues.

Location	**Sequence**	**Peptide**	**Oxidation**	**Control**	**Oxidation**
			Pattern	**Oxi%**	**Oxi%**
L1	48-64	GNGENSYGGNGDMTYAR	O16	0	100
β4	105-111	LAFAGLK	O16	0	8.1
L3	123-154	NYGVVYDALGYTDMLPEFGGD	O16	1	100
β7	163-182	TAYSDDFFVGR	O16	0	6.4
L6	266-275	NSNFFGLVDGLNFAVQYLGK	O16	0	9.7
β15	328-345	FTNTSGFANK	O16	31.8	35.7
		NMSTYVDYIINQIDSDNK	2 O16	0	9.8

Table 2.3. Oxidation pattern and the responding oxidation extent of the six peptides derived from TIC. The oxidation extent of modified peptides was averaged from duplicate experiments, with the standard deviations seeing in Figure 2.5.

As shown in Table 2.3, five peptides with their amino-acid sequences 48-64, 105-111, 123-154, 163-182, 266-275, respectively, were oxidized in the sample as compared to control, while another one with its amino-acid sequences 328-345 displayed a significant level of oxidation in both control and oxidation group. However, the oxidation level of the modified products, especially of the di-oxidation products, was significantly higher in the sample as compared with control, as shown in Figure 2.3A and 2.3B. The oxidation background in control group was mainly from the oxidation of the amino acid 329M, because methionine is inherently highly reactive and easily to be oxidized without much external oxidative stress (5).

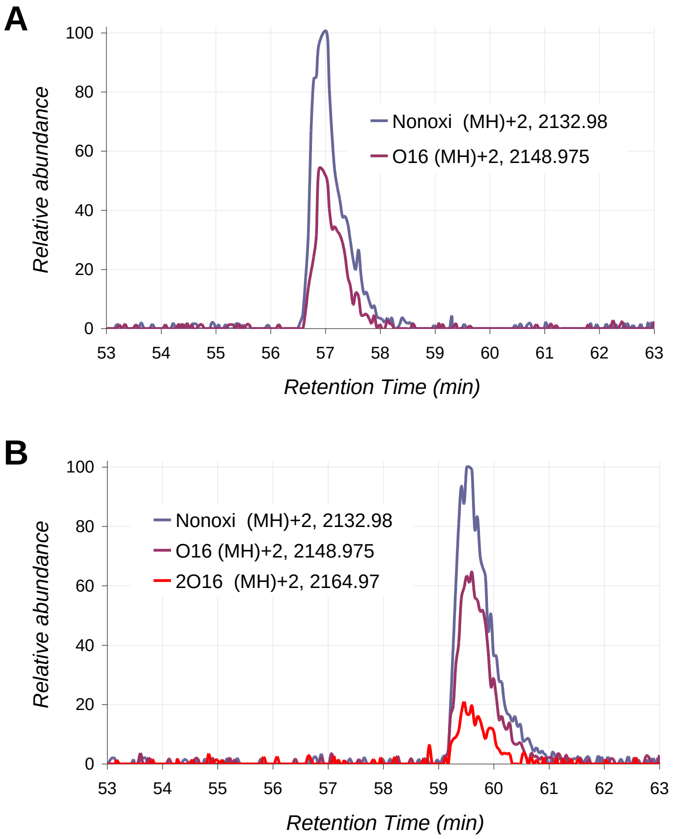
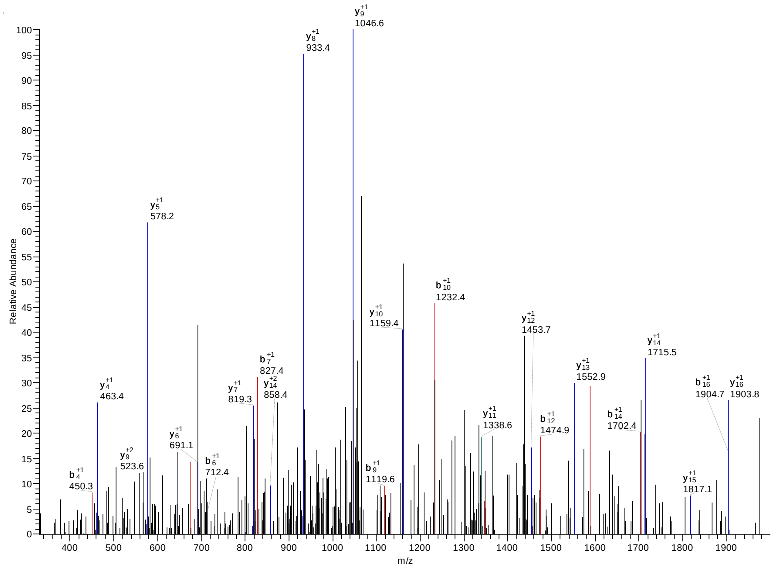


Figure 2.3. **The selected ion chromatograms of oxidized and unoxidized doubly charged peptide ions from tryptic peptide 328-345**. A. The peptides from negative control group without oxidation, 90min HPLC gradient. B. The peptides from oxidation group bathed with Fenton reagents in M9 medium, 90min HPLC gradient.

Some modified peptides might be missed by the current data analysis method because the oxidation products are very complicated and only a sub-set of possible modifications has been included in the SEQUEST database research. Interestingly, in our study, both SEQUEST and ByOnic identified similar modification sites even though their scoring algorithms differ substantially, indicating that identified modification sites are highly reliable.

Location	Sequence	Peptide	Oxidized AA	Mass Shift (Da)	SASA by X-ray (A^2)
L1	48-64	GNGENSYGGNGDMTYAR	59D	16	69.91
			60M	16	59.28
			62Y	16	27.28
β4	105-111	LAFAGLK	110L	16	93.93
L3	123-154	NYGVVYDALGYTDMLPEFGGDTAYSDDFFVGR	133Y	16	14.77
			135D	16	52.46
			136M	16	72.52
			137L	16	16.54
β7	163-182	NSNFFGLVDGLNFAVQYLGK	171D	16	123.86
			175F	16	54.38
			177V	16	38.97
			179Y	16	119.55
L6	266-275	FTNTSGFANK	266F	16	138
			268N	16	121.68
			270S	16	47.12
β15	328-345	NMSTYVDYIINQIDSDNK	329M	16	58.33
			332Y	16	15.04
			333V	16	69.4
			335Y	16	76.6

Table 2.4. Number of oxidations detected by LC/MS and reactive amino acid residues identified by LC/MS/MS for OmpF. The solvent accessible surface areas were calculated using Surface Racer 4.0 using X-ray structure of OmpF. Site locations of the amino acids were listed in column 1.



		B	Y	
1	N	115.05	-	18
2	M*	262.09	2050.93	17
3	S	349.12	**1903.89**	16
4	T	**450.17**	**1816.86**	15
5	Y	613.23	**1715.81**	14
6	V	**712.3**	**1552.75**	13
7	D	**827.32**	**1453.68**	12
8	Y*	1006.38	**1338.65**	11
9	I	**1119.47**	**1159.6**	10
10	I	**1232.55**	**1046.51**	9
11	N	**1346.59**	**933.43**	8
12	Q	**1474.65**	**819.38**	7
13	I	**1587.74**	**691.33**	6
14	D	**1702.76**	**578.24**	5
15	S	1789.79	**463.21**	4
16	D	**1904.82**	376.18	3
17	N	2018.86	261.16	2
18	K			

Fig. 2.4 The MS/MS spectrum of the modified peptide sequence 'NM*STYVDY*IINQIDSDNK'. Oxidation sites can be confirmed by inspection of the *b* and *y* ion series. * means the +16 Da mass increase. Values in bold represent the relative ions detected by MS/MS, with red referring *b*- ions and blue referring *y*- ions, respectively.

Consistency of footprinting and crystallographic data for porin OmpF

Among the oxidized peptides, the ones annotated as L1, L6 are loops locating on the extracellular side of the bacteria, which are in contact with the aqueous medium directly. Although L3 folds back into the center of the water channel, it locates right at the entrance of the eyelet and it remains completely solvent accessible. Although the polypeptides in the β-stranded area interact with the lipid layer on one side, their other sides are in contact with water inside the channel and are thus susceptible to oxidization by the Fenton reagents. In comparison to the external loops, the oxidation level of the less accessible internal β-barrels should be theoretically lower and our observation confirmed that the oxidation level of identified β-sheets was indeed far lower than the identified external loops, as shown in Table 2.3.

Interestingly, several loops still remained un-oxidized even in oxidation group e.g. L2 and L4. According to the crystallographic structure of OmpF (68), loops L2 and L4 are important for monomer-monomer interaction in the non-polar core of the porin trimer, and will therefore be inaccessible to the solvent and protected from oxidation by the Fenton reagents in the solution. Consistent with this hypothesis, neither L2 nor L4 was found to be oxidized in this study.

The nineteen amino acids from the six peptides were widely distributed in the external loop area and internal β-strand area, as shown in the reconstituted two dimensional (2D) topology of OmpF in Figure 2.5 and 3D structure of OmpF in

Figure 2.6. We further calculated the solvent accessible surface area of these amino acids according to their X-ray structure (1OPF) using Surface Racer 4.0. As displayed in Table 2.4, consistent with our expectation, all amino acid residues were validated as solvent accessible.

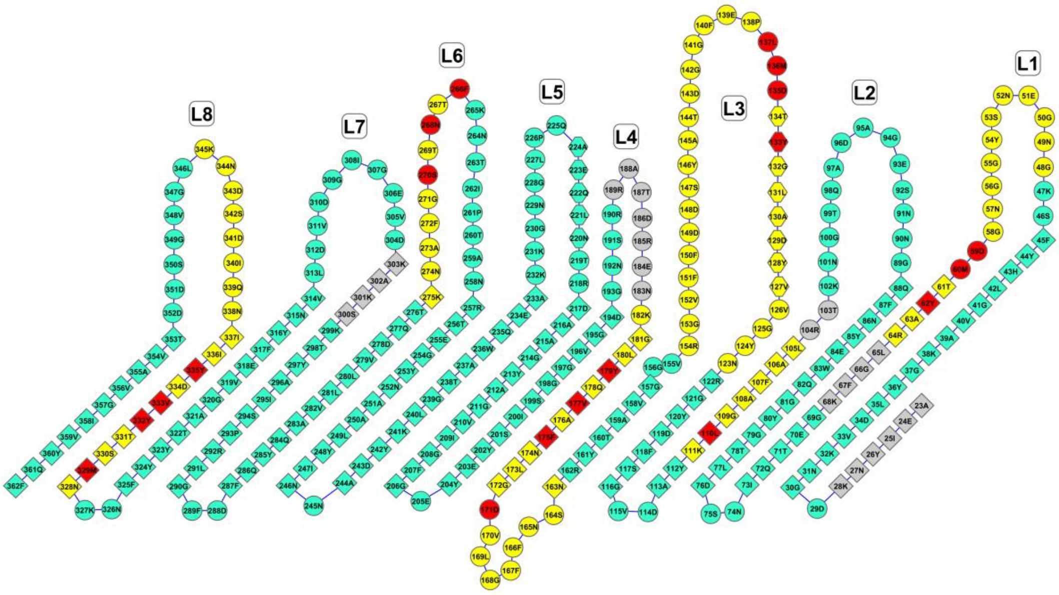


Figure 2.5. 2D topology of OmpF porin with amino-acid sequence in one-letter code. The view is from outside the 16-stranded anti-parallel β-barrel, which is unrolled. Secondary structural elements are indicated by diamonds for barrel β-strands, hexagons for α-helices and circles for turns and loops. Amino-acid sequences labeled in grey color were lost and unable to be detected by MS; the ones labeled in yellow were tryptic peptides oxidized in experiments, while the ones in cyan were un-oxidized. Amino acids labeled in red color were recognized by LC-MS/MS analysis as the oxidized sites.

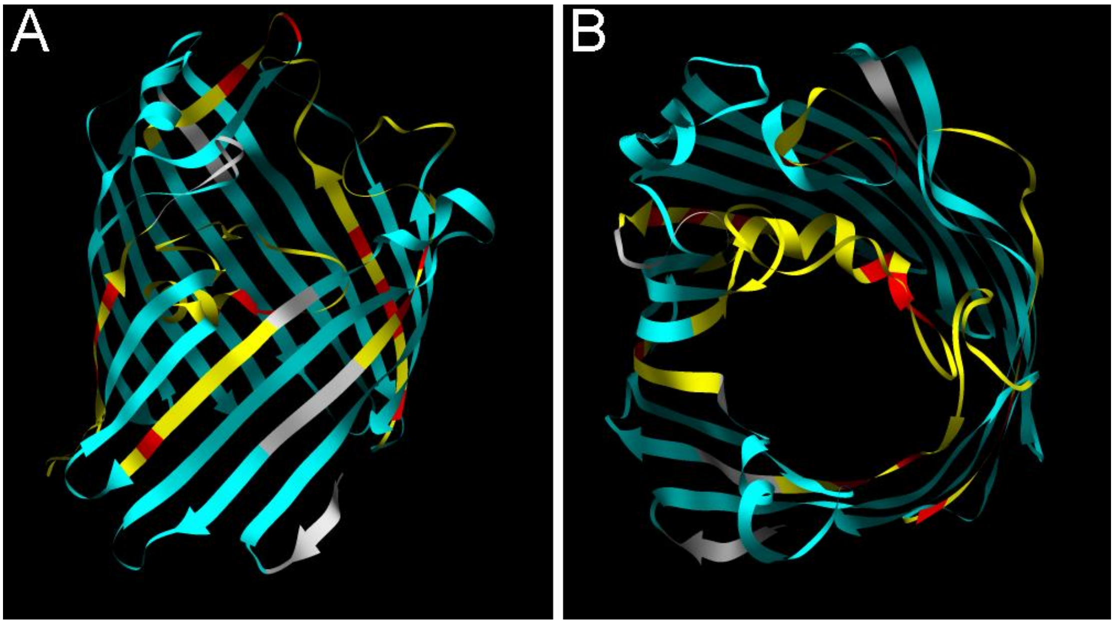


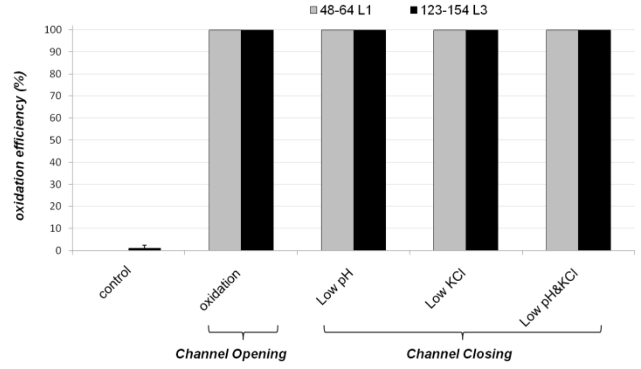
Figure 2.6. 3D structures of OmpF monomer, from side view A and top view B, respectively, according to the footprinting data as well as X-ray data (1OPF for OmpF). Peptides detected to be oxidized in experiments are labeled in yellow, with the specific oxidation sites in red color. Un-oxidized peptides are labeled in cyan and peptides unable to be detected in grey.

Footprinting results proved the ion gating of OmpF porin channel in vivo

It has been reported that the voltage gating of general porins depends on environmental factors like pH, low ionic strength buffer, poly-cations, polysaccharides, membrane-derived oligosaccharides and pressure (60, 69-73), implying that the environment mediated voltage gating of general porins may be a natural protective mechanism of Gram negative bacteria. Thus we coupled footprinting and environmental factors to probe the surface topology of porin OmpF to determine if voltage gating occurs *in vivo*. If voltage gating occurs *in vivo*, we expect that there will be some dynamic changes in OmpF in response to changes in ionic environment. Low pH and ionic strength buffer were investigated as two main environmental factors for i*n vivo* study of ion gating.

Figure 2.7 showed the extent of oxidation of the five tryptic peptides (the tryptic peptide 266-275 ‘FTNTSGFANK’ with its description of L6 was lost in some conditions in the experiment, thus we didn’t take it into account) under different conditions: normal physiological medium (M9 medium), low pH (300 mM KCl, pH 3.0)), low ionic strength (5 mM KCl, pH 7.4), or a combination of low pH and low ionic strength (pH 3.0, 5 mM KCl). Under low ionic strength, low pH or a combination of both, oxidation level in loops L1 and L3 remained unchanged from that in normal physiological condition while that in β-strand areas decreased significantly below that in normal physiological condition and were just above background level in the control group. As to peptides L1 and L3, only their oxidized counterparts could be easily detected by MS analysis in all four conditions, indicating that they remained highly solvent accessible. As to the amino acid residues of the β-stranded area deep inside the channel, the extent of oxidation significantly decreased. As to the tryptic peptide 328-345, the population of its di-oxidation products became much smaller and even disappeared. These results demonstrated that few OH radicals could reach this internal β-stranded region to oxidize the amino acids located there.

The significant decrease in the oxidation efficiency of these β-sheet peptides when the cells were exposed to low ionic strength, low pH or a combination of both is consistent with channel closure of the pore, i.e., voltage gating. In this study, the OH radicals used to oxidize the amino-acid side chains inside the β-barrel water channels were mainly generated by the reaction of Fe(II) with H_2O_2 in situ in the channel. When the channel closed in response to low ionic strength, low pH, or a combination of both, there was a “switch off” of Fe(II) current into the channel and effectively inhibited Fenton reaction and the subsequent formation of OH radicals. Also, the high reactivity and short half-life of OH radical make it impossible for any OH radical that is generated outside the channel to passively diffuse into the channel. Therefore when the channel was closed, Fe(II) current would be switched off, Fenton reaction would not occur, and no OH radical would be generated. Consequently, oxidation could not take place.


A. Peptides Outside the Channel

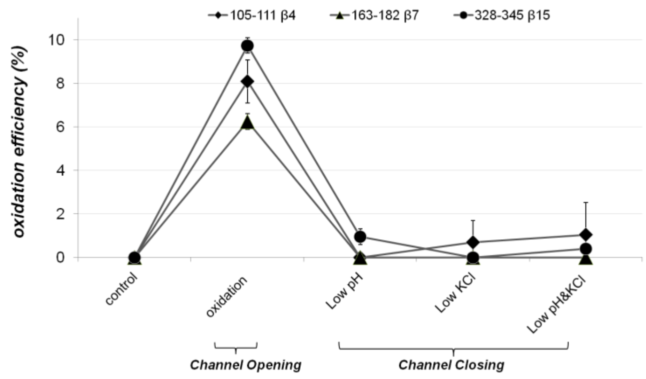

B. Peptides Inside the Channel

Figure 2.7. Oxidation level of the peptides from different areas, loop area outside the ion channel and β-stranded area inside the ion channel in different conditions, as shown in X axis. A. The extent of oxidation of Loops L1 and L3 kept constant in both channel opening and closing conditions. B. The oxidation extent of peptides in β-strand areas fell back to almost the level of control group in channel closing conditions.

Footprinting results revealed the molecular basis of voltage gating of OmpF porin

It has been proposed that large deformation–closing of the solid structure of the porin channel is impossible. In a complete barrel of β-strands, all of the polar main chain atoms are engaged in inter-strand hydrogen bonds, and the secondary elements are fixed in a rigid network of hydrogen bonds which can’t be easily broken, structural change of the barrel thus have been considered to be energetically forbidden (54). Therefore porins must have other features relevant to their gating. Several hypotheses for the molecular basis of porin voltage gating have been proposed by researchers (59, 74).

First, the eyelet has been considered by some researchers as the likely site of voltage gating with structural change in the barrel considered unlikely. Based on its crystallographic structure, the relatively narrower eyelet passageway within the OmpF porin is formed by folding back of the surface loop L3 into the barrel, thus L3 plays a key role in ion selectivity and transport of larger molecules (54). It is also responsible for voltage gating of porins, as has been proved by previous studies (75-78). Two possible mechanisms of voltage gating relevant to loop L3 were proposed, the first is the electrostatic feature of the eyelet region and the second is the large motion of L3 which closes the channel lumen (59, 74). However, the movement of L3 has been proven to be unlikely by previous mutagenesis experiments (75, 79-81). Therefore the former proposed mechanism: the local electrostatic field changes in the eyelet region in response to environmental changes may alter the channel to switch the conductivity of ions, has become widely accepted. Another proposed mechanism is the movement of surface loops folding back into porins under certain circumstance. According to Muller and Engel's study based on AFM (57), the height of OmpF above the lipid matrix is 0.6 nm at low pH or high transmembrane voltage, but the height is 1.3 nm at neutral pH or without transmembrane voltage and this conformational change is reversible. Based on this observed phenomenon, they further hypothesized that extracellular loops act as the door of the porin channel, and when loops protrude outside the channel, the door is open, and once they fold back inside the channel, the door would be closed, thereby inhibiting the ion current across the channel.

With the deformation of the solid structure of β-barrel and large motion of loop L3 considered unlikely, our results are opposite to the door-closing hypothesis above. According to our footprinting data *in vivo,* the interior L3 polypeptide still remained oxidized even in the 'closed state' of the channel, which indicated that L3 was surrounded by Fe(II) and H_2O_2 in the pore and that the entrance of the channel had not been blocked by extracellular loops. If the extracellular loops indeed folded back to close the entrance of the channel, L3 should be protected and remained as un-oxidized, meanwhile the oxidation efficiency of the superficial loop L1 polypeptide would decrease to some extent rather than remaining the same in this case.

To our current knowledge, acidic residues D135 and E139 of L3 (113D and 117E according to X-ray structure because the N-terminal 1-22 amino-acid sequence is absent in the crystallography data) and a cluster of basic residues

R64, R104 and R154 (R42, R82, R132 in X-ray structure) on the barrel wall opposite the loop give rise to a transverse electrostatic field, which is essential for ions diffusion (82-83). The electrostatic field strength has a close relationship with the membrane-bathing solution, especially with its salt concentration (84). Replacing of the minimum salts medium M9 with the salts concentration 130 mM with 5 mM KCl in our study weakened the electrostatic field and affected Vc (critical voltage above which the ion channel will be switched off), which caused the channel closure. Meanwhile, the charged state of the residues in the eyelet region can also have a significant effect on Vc, and changing the charge state of these residues may change the Vc thus affecting voltage gating according to previous mutagenesis studies (66, 76-79, 85-86). In our study, when the bacteria were bathed in a pH 3 buffer, the acidic residues (D113 and E117) of L3 would be neutralized by protonation of the carboxylic groups and the charged states of these two residues thus changed. These were of the same effect as the mutations of physiologically negatively charged D (E) with either neutralized N (Q) or positively charged K and our results were consistent with the mutagenesis studies (77-78, 85).

Conclusion

In conclusion, we have adapted the hydroxyl radical surface mapping and its associated MS analysis to probe the surface protein of living cell. The robustness of this adaptation was verified by the consistency of our surface mapping data of porin OmpF with its current crystallographic structure. In addition, this adaptation also enabled the detection of *in vivo* voltage gating of porin OmpF for the first time. These results suggest that the novel cell surface mapping method coupled with MS analysis is useful to study outer membrane proteins *in vivo.* And we believe that this method is not only suitable for the structural study of OMPs in bacteria, it has the potential of being extended to probing the conformational changes of IMPs in mammalian cells.

Chapter 3. Revealing Conformational Changes of αLβ2 Integrin Directly on Cell Surfaces by Hydroxyl Radical Footprinting

Introduction

Integrins are non-covalently associated transmembrane αβ heterodimers and to date 18 α and 8 β subunits are known to form 24 known αβ pairs in vertebrates (87-88). Nine of the α subunits contain an additional von Willebrand factor A (VWA) or inserted (I) domain (αI domain) in the extracellular segment, which has been shown as the principle divalent-cation-dependent ligand binding domain when present. The αI-lacking integrin subgroup binds ligand in part through a highly homologous αI-like domain (βI domain) present in every β subunit. Integrins use bidirectional signalingsignaling to integrate the intracellular and extracellular environments. Integrins play critical roles in the immune system and are critical for almost all cell migration and adhesive interactions.

Integrin functions are tightly linked to its molecular conformations. The structural basis of integrin affinity modulation and its interaction with ligand have been the interest and subject of recent studies (89). However, global structural information to date is limited to ectodomains of only three integrins, two of which lack αI domain, αVβ3 (90-91) and αIIbβ3 (92-93), and one with αI domain, αXβ2 (94). Until now only three predominant integrin conformations have been established together by crystal and electron microscopy (EM) studies—the bent conformation, the extended conformation with a closed headpiece, and the extended conformation with an open headpiece (Figure 3.1)—which represent different states of integrin affinity—inactive, intermediate, and active (91, 93-97). The integrin αLβ2 (LFA-1, leukocyte function-associated antigen-1) undergoes similar conformational transition as shown in previous *in vitro* EM study (97). Activation by Mg^{2+}/EGTA or Mn^{2+} is required to convert αLβ2 to a form that can bind some ligands with high affinity (98). Studies based on antibody binding affinity of αLβ2 (99) found that Mg^{2+}/EGTA treatment may convert αLβ2 into an intermediate affinity state which has its hybrid domain re-oriented, while such rearrangement is not great enough for the ligand binding to the αI domain, and the αLβ2 conformer on the cell surface adopts an extended conformation while with a still closed headpiece (Figure 3.1) in the presence of Mg^{2+}/EGTA (99-100).

Although the snapshots of different *in vitro* conformational states have been intensively studied, current techniques like NMR, X-ray, EM method cannot directly probe the conformational movements of integrins on cell surfaces which are essential to understanding their functions (4). Alternative methods are required to determine the real-time conformational changes in these proteins. Studies of natural and engineered mutations within integrins and analyses of the epitopes for function-altering monoclonal antibodies shed light on the structural basis of integrin regulation from one side (101). And also, work with a class of small molecules termed α/β I-like allosteric antagonists suggests analogous conformational changes in integrins (89). Except for these functional studies, there is no direct experimental evidence of native integrin to confirm the *in vitro* structural observations above.

Recent progress in mass spectrometry has enabled MS-based OH radical protein footprinting technique to determine the protein surface topology (15-17). This technique uses OH radical to oxidize amino acid residues located on the protein surface to produce stable covalent modifications to side chains without causing backbone cleavages. Due to the small size and nonspecific activity of OH radicals, this is a random process dependent only on the solvent-accessible surface and the chemical properties of the exposed amino acids. The oxidized amino acid residues provide the surface information of the protein, and thus the surface topology can be mapped. This approach has been shown to be a powerful technique in understanding ligand-induced conformational changes when coupled with existing structural data that are either experimentally or computationally generated (11, 102-103). Recently we have extended its application to study the *in vivo* structural dynamics of IMPs, and successfully illustrated the *in vivo* voltage gating of porin OMPF of *E. coli* (104).

Here we further implemented the application of the nanosecond pulse laser-induced photolytic hydroxyl radical footprinting method (15) to study the conformational change of native αLβ2 integrin on plasma membrane of the surface of a mammalian cell. In this method, a high concentration of OH radicals is generated in 5~7 ns via photo-dissociation of a H_2O_2 solution by a pulse Nd:YAG laser. The OH radicals immediately oxidize amino acid residues located on the protein surface to produce stable covalent modifications. The half-life of OH radicals is in nanosecond timescale, despite half-life of some secondary radicals generated by reacting with OH radical may be longer, which should be quenched in milli-second time scale. Because of the nanosecond time scale of the formation of OH radicals as well as their subsequent oxidation and quenching, this method allows snap frozen the solvent-accessible surface or footprints of a protein in its native biological state(15).

The footprinting MS data demonstrated the activation and overall extension of αLβ2 heterodimer upon Mg/EGTA activation. The findings are consistent well with what have been discovered in other studies. Therefore, the hydroxyl radical footprinting technique has the potential application in both the structural and funtional studies of plasma membrane proteins *in vivo*.

Materials and methods

Antibodies and reagents

Purified IgGs of MHM23 (a dimer specific mAb) and H52 (a β2-specific mAb) were prepared from hybridoma supernatants using Hi-Trap protein-G columns (GE Healthcare). Protein A-Sepharose CL-4B beads were obtained from GE Healthcare.

Complete protease inhibitor cocktail tablets were from Roche Diagnostics (Indianapolis, IN, USA). H_2O_2 was from Merck (Singapore). HPLC grade solvents were purchased from J.T. Baker (Phillipsburg, NJ, USA). All the other chemicals were purchased from Sigma-Aldrich (St.Louis, MO, USA) unless otherwise indicated.

Cell Culture and Mg^{2+}/EGTA Activation

MOLT-4 cells (T Lymphoblast) (ATCC) were grown in complete medium: RPMI 1640 with L-glutamine (Thermo Scientific HyClone), 10% (v/v) heat-inactivated fetal bovine serum (HI-FBS) (Thermo Scientific HyClone) with penicillin-streptomycin (100 IU/ml; 100 µg/ml). There were three groups of cell samples in this study. The first was the control group (abbreviated as Ctrl) containing cells without any treatments and oxidation. In the second group, cells were in the inactive physiological resting state (abbreviated as PhyR) and they were oxidized *in vivo* by laser footprinting. Cells in the third group were first activated with Mg^{2+}/EGTA reagent, and were then oxidized with laser footprinting. This group is thus named as the MgAc in short.

For MgAc group, briefly, Molt-4 cells were washed once in PBS, and were then incubated in Mg^{2+}/EGTA (5 mM $MgCl_2$ and 1.5 mM EGTA) in RPMI media containing 5% (v/v) HI-FBS and 10 mM HEPES (pH 7.4) for 30 min at 37 ℃ (105).

Flash oxidation of Molt-4 cells using a nanosecond pulsed laser

Cells from PhyR group were washed twice in PBS and re-suspended into fresh ice cold PBS (cell density $\sim 1\times 10^7$) for laser footprinting experiment. As to cells from MgAc group, after activation, they were washed twice with ice cold PBS supplemented with Mg^{2+}/EGTA (5 mM $MgCl_2$ and 1.5 mM EGTA) and re-suspended into ice cold PBS supplemented with 5 mM $MgCl_2$ (cell density $\sim 1\times 10^7$).

All samples were prepared in triplicates to minimize the biological variation among each group. About 3×10^8 of MOLT-4 cells was used for each duplicate sample in the same group. About 2.7 billion cells were used in each time experiment.

The PBS cell suspension for each group was then aliquoted into a 12-well plate, with each well containing 1.5 ml of suspension (cell density $\sim 1\times 10^7$). H_2O_2 was added to the cell suspension to a final concentration of 1% before laser irradiation. Cells were then oxidized with OH radicals generated by exposing the suspension to a pulse Nd: YAG laser (LOTIS TII LS-2134UTF, Minsk, Belarus) operating at 266 nm and 30 mJ. The 12-well plate was held horizontally, and each well opening was properly aligned with the laser beam (Figure 3.1). Immediately after one laser shot, the oxidation was quenched by addition of appropriate amount of catalase to remove the residual H_2O_2. Cells were collected subsequently by spinning down at 500g, 4 ℃ for 10 min.

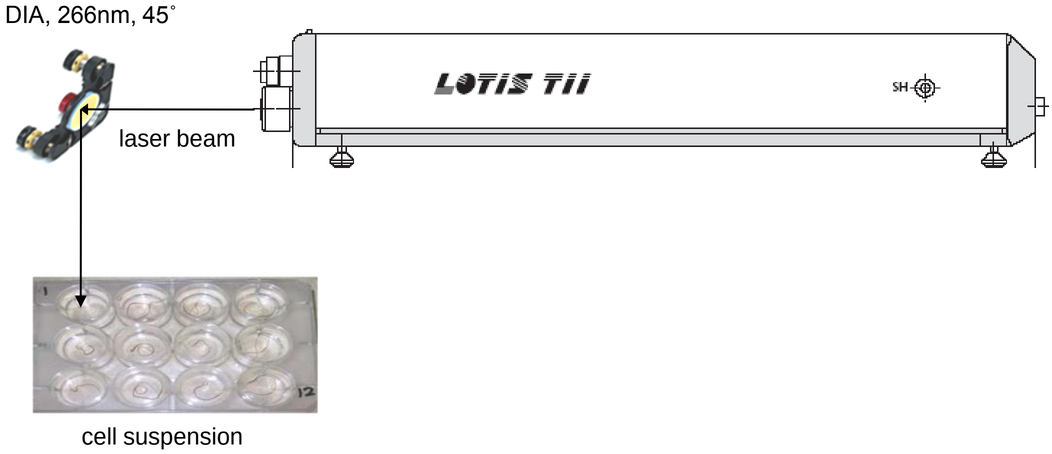


Figure 3.1 The schematic diagram of the setup for nanosecond pulse laser-induced photolytic hydroxyl radical footprinting method.

Immunoprecipitation (IP)

Cells were lysed in lysis buffer [10 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1% (v/v) NP-40] containing appropriate protease inhibitors cocktail. Cell debris was removed by spinning down at 13,000 ×g, 4 ℃ for 15 min. The supernatant was pre-cleared with rabbit anti-mouse IgG coupled to protein A sepharose beads first. MHM23 antibody was added to the precleared supernatant and incubated with end to end rotation in cold room for 3 h followed by further incubation with the rabbit anti-mouse IgG coupled to protein A sepharose beads for 1 h. After collecting the precipitated beads-protein complex by centrifugation, H52 antibody was added to the supernatant, following the same IP procedure as described above.

Before elution, beads were washed once with lysis buffer containing 500 mM NaCl, followed by another wash with lysis buffer, and then were loaded onto a 30% (w/v) sucrose cushion. After centrifugation for 3 min at 12,000 *g* at 4 ℃, beads were washed once with 10 mM Tris-HCl (pH 8.0) containing 0.05% (w/v) SDS (105). Bound proteins were eluted with 3×200 µl of glycine-HCl (pH 1.9). Protein samples were subjected to 7.5% SDS-PAGE of 1.5mm thickness.

Immunoprecitated proteins from three identical duplicates of each group of samples described above were pooled together for further deglycosylation, in gel digestion and subsequent LC-MS/MS analysis. The experiment was repeated three times.

In-gel digestion

Integrin protein bands from the SDS page were cut out according to their molecular weight after coomassie blue staining. Since the protein is highly glycosylated, deglycosylation is necessary. The glycans were enzymatically released from the gel bands by incubation with PNGase F (500 units/µL) in 25 mM NH_4HCO_3 at 37 ℃ overnight after destaining. Samples were reduced with

10 mM dithiothreitol in 25 mM NH_4HCO_3 at 56 ℃ for 1 h, followed by alkylation with 55 mM iodoacetamide in 25 mM NH_4HCO_3 solution at room temperature in dark for 45 minutes and digested by sequencing grade modified trypsin (Promega Corporation, Madison, WI. USA) in a 1:100 (trypsin: protein) mass ratio at 37 ℃ overnight.

LC-MS/MS analysis

The LC-MS/MS setup can be referred to Figure 2.2 in Chapter 2. The LC-MS/MS analysis was done in an LTQ-FT Ultra mass spectrometer (Thermo Electron) coupled with a prominence^{TM} HPLC unit (Shimadzu). The tryptic peptides were reconstituted to 150 µl of 0.1% formic acid (FA) in HPLC grade water. The samples were then injected by an autosampler with an injection volume of 50 µl each time and on-line desalted in a Zorbax peptide trap (Agilent, Pola Alto, CA). Each sample was analyzed by LC-MS/MS independently for three times. The peptide separation was performed in a home-packed C18 column (75 µm inner diameter×10cm, 5 µm particles, 300 Å pore size, Michrom BioResources, Auburn, CA, USA). Buffer A (99.9% H_2O, 0.1%FA) and buffer B (99.9% ACN, 0.1%FA) were used for the LC gradient. The 90-min gradient was ramped from 5% ACN to 30% ACN in 65 min, then to 60% in 10 min and to 80% ACN over 2 min and then was kept at 80% ACN for 3 min and ramped back to 5% ACN for the last 10 min. The sample was ionized into the mass spectrometer through an ADVANCE™ CaptiveSpray™ Source (Michrom BioResources) with an electrospray potential of 1.5 kV operating at a flow rate of about 200nl/min after a splitter. The nitrogen gas flow was set at 2, ion transfer tube temperature at 180℃ and collision gas pressure at 0.85 mTorr. The LTQ-FT Ultra was set to perform data acquisition in the positive ion mode as described previously (106). Briefly, a full MS scan (350-2000 m/z range) was acquired in the FT-ICR cell at a resolution of 100,000 and a maximum ion accumulation time of 1000 msec. The AGC target for FT was set at 1e+06 and precursor ion charge state screening was activated. The linear ion trap was used to collect peptides and to measure peptide fragments generated by collision-activated dissociation (CAD). The default AGC setting was used (full MS target at 3.0e+04, MS^n1e+04) in linear ion trap. The 10 most intense ions above a 500 counts threshold were selected for fragmentation in CAD (MS^2), which was performed concurrently with a maximum ion accumulation time of 200 msec. Dynamic exclusion was activated for the process, with a repeat count of 1 and exclusion duration of 20 s. Single charged ion is excluded from MS/MS. Isolation width was 2 Da, and default charge state was 5. For CAD, normalized collision energy was set to 35%, activation Q was set to 0.25, and activation time 30 ms. Spectra were acquired in centroid format in raw data files with XCalibur (version 2.0 SR2).

Bioinformatics analysis

The analysis followed the same way as described in the previous study (104) with some modifications. MS/MS data for αLβ2 integrin footprinting experiments were submitted to SEQUEST protein sequence database search engine (BioworksBrowser 3.3, Thermo) for peptide/protein identification against an in-house integrin database. Peptide tolerance was 8 ppm, and fragment ions tolerance was 1 AMU. Full enzymatic cleavage by trypsin was selected, with maximum 2 missed cleavages allowed. Two predominant kinds of OH radical oxidations, one with a mass increase of +16 Da for almost all the amino acid residues, and the other with a mass increase of +32 Da for five residues C, F, M, W and Y, were included as variable modifications. We also enabled the following non-oxidative modifications, carbamidomethylated cysteine, and deamidated N and Q as variable modifications. Maximum of six modifications were allowed for a single peptide.

Identified peptides were subject to correlation analysis as described below to remove low-quality identifications (107). For each unique peptide, all the spectra matched were pooled, and divided into two groups according to their XCorr (cross correlation) values. The qualified group contains spectra with minimal XCorr cutoffs of 1.0, 1.5, and 2.0 for +1, +2, and ≥+3 charged peptides, respectively (108). As to those spectra with lower XCorr values, they were subsequently correlated with each of the qualified spectrum by correlation analysis using a perl package Statistics::LineFit (version 0.06, from Comprehensive Perl Archive Network, CPAN). P values of correlation were calculated using another perl package Statistics::Distributions (version 1.02). For each spectrum, the top 10 peaks, where available, from every 100 m/z range, were used to simplify the spectrum. The top 50 peaks from the resulting spectrum were submitted for correlation analysis. Spectra with fewer than 50 peaks were excluded. Only spectra with minimum R-square value of 0.81, correlation between 0.8 and 1.25, and p values<0.05 were "rescued" and included into the qualified group. After identification and qualification, normalized spectral index (SIn) was calculated to represent the abundance of oxidized and un-oxidized peptides (42). SIn values of peptides were grouped based on the longest peptide sequence without miss-cleavage, and presented using box-plot in statistical platform R. Student *t* test was used to calculate significance for difference between experiment conditions.

Modeling of αLβ2 integrin

The 3D structure of αLβ2 integrin was obtained by homology modeling by using the SWISS-MODEL protein modeling (109), inputting the αLβ2 integrin sequence as "target" and the 3D structure of αXβ2 as "template". Sequence alignment was further manually adjusted according to the returned SWISS-MODEL report using Swiss-PdbViewer to get the energy favored 3D structure. UCSF Chimera (110) was used in this study to display the 3D structure.

Calculation of solvent-accessible areas of amino acids

The solvent-accessible surface areas of amino acids were calculated by using the computer program Surface Racer 4.0 with a probe radius of 1.4 Å (65) based on the structural model of αLβ2 and the crystal structure of αXβ2 (3k6s).

Results and discussion

Detection of oxidized peptides of αLβ2 integrin by LC-MS/MS analysis

Both integrin αL and β2 subunits are type I transmembrane proteins with large extracellular domain, a single transmembrane domain and a short cytoplasmic tail (Figure 3.2A) (87). The N-terminal portions of the αL and β2 subunits fold into the globular headpiece, which is connected through αL and β2 tailpiece domains to the membrane (Figure 3.2B). The αL ectodomain is composed of 1090 amino acids and β2 ectodomain is composed of 701 amino acids.

The expression level of αLβ2 integrin is quite low (10^4-10^5 copies/cell) in MOLT-4 cells. Thus we used about 3×10^8 of MOLT-4 cells for each duplicate sample to generate enough amount of αLβ2 integrin and did large-scale immnoprecipitation to enrich the protein for LC-MS/MS analysis. The amount of protein was estimated to several µg for each sample. We used two monoclonal antibodies, a dimer specific mAb MHM23 and a β2 specific mAb H52, in the IP step to enrich the two protein subunits. In our preliminary experiments, initially we only used the dimer specific mAb MHM23 for the IP. It works well to the cell lysate from the control group. But we found that β2 band is very faint if we used MHM23 alone after oxidation when compared with the β2 band got from the control group in our preliminary experiments, thus we included another mAb H52 for β2 subunit to further recover it.

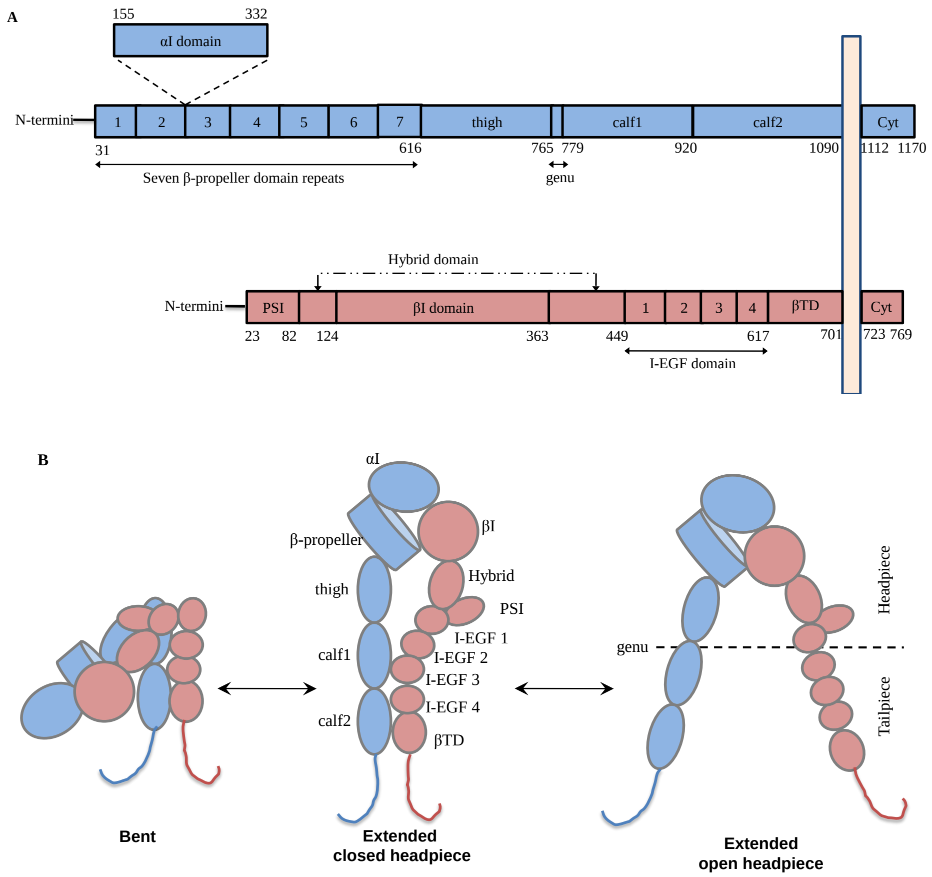

A
155
332
αI domain
N-termini
1
2
3
4
5
6
7
thigh
calf1
calf2
Cyt
31
616
765
779
920
1090
1112
1170
Seven β-propeller domain repeats
genu
Hybrid domain
N-termini
PSI
βI domain
1
2
3
4
βTD
Cyt
23
82
124
363
449
617
701
723
769
I-EGF domain
B
αI
β-propeller
βI
thigh
Hybrid
PSI
I-EGF 1
I-EGF 2
calf1
I-EGF 3
I-EGF 4
calf2
βTD
genu
Headpiece
Tailpiece
Bent
Extended closed headpiece
Extended open headpiece

Figure 3.2. Integrin architecture. A. Organization of domains with the primary structures. αI domain which inserted between the 2nd and 3rd β-propeller repeats is denoted by the dashed lines. The αL ectodomain contains five domains, αI domain, β-propeller domain, thigh domain, calf1 domain and calf2 domain; the β2 subunit contains eight domains, βI domain, hybrid domain, PSI domain, I-EGF1, 2, 3, 4 domains, and β tail domain (βTD). The domain-domain boundary amino acid is labeled by numbers along the primary structure. B. Schematic with cartoons illustrating the rearrangement of domains during activation of integrin αLβ2 based on Luo et al (87) . The headpiece is composed of αI domain, β-propeller domain, thigh domain from αL subunit and PSI domain, Hybrid domain, βI domain and I-EGF1 domain from β2 subunit. PSI, plexin-semaphorin-integrin domain; I-EGF, integrin-epidermal growth factor fold; βTD, β tail domain; Cyt, cytoplasmic domain.

When investigating the distribution of theoretical tryptic digestion sites Lys (K) and Arg (R) along the primary sequence, we find that there are many peptide fragments with less than five amino acids in their sequence, and there are also several extremely long sequences with more than thirty amino acids. These peptides are either challengingly short or long to be identified via LC-MS/MS analysis and may further induce extra missed cleavages, resulting in certain extent of sequence loss during the LC/MS/MS analysis.

In the present study, footprinting MS method still gives satisfying coverage of both integrin αL and β2 subunits for elucidating conformational changes. As to αL subunit, the sequence coverage of it in the three groups-Ctrl, PhyR, and MgAc-are 75%, 72% and 72%, respectively, as calculated from MS/MS results. The sequence coverage of β2 subunit in the three groups-Ctrl, PhyR, and MgAc-are calculated to be 82%, 78%, and 81%, respectively.

As described above, two most frequent modifications, +16 Da and +32 Da, were included as variable oxidative modifications in this study. Oxidized peptides can thus be detected by the increased peptide mass with an integral multiple of +16 Da or +32 Da, and identification of modified sites could be further confirmed by manually inspection of the mass spectra.

58 peptides from αL subunit and 46 peptides from β2 subunits were identified from the MS/MS datasets. Among these peptides identified, most of them showed no significant difference in their oxidation levels among the three groups, while 14 peptides from αL subunit and 12 peptides from β2 subunit were found to have significant changes in the relative abundance of their relevant oxidized counterparts after Mg^{2+} activation (Table 3.1, Figure 3.3, and Figure 3.4).They are widely distributed along the two subunits, from the globular headpiece to the tailpiece. Table 3.1 also lists the location along the subunits of each peptide.

Some modified peptides might be missed by the current analysis method as well as the data set because the oxidation can generate many different oxidative products, and only a subset of possible modifications has been included in the SEQUEST database search.

A

domain	**start site**	**peptide sequence**	**change of oxidation extent after Mg activation**
αI	153	K.GNVDLVFLFDGSMSLQPDEFQK.I	↑
	175	K.ILDFMK.D	↑
	258	K.VLIIITDGEATDSGNIDAAK.D	↑
	331	K.IYVIEGTSK.Q	↓
β-propeller	49	R.VLQVGNGVIVGAPGEGNSTGSLYQCQSGTGHCLPVTLR.G	↑
	134	R.QNLQGPMLQGRPGFQECIK.G	↑
	494	R.VFIYQR.R	↑
	569	R.IEGTQVLSGIQWFGR.S	↓
thigh	667	R.LVANLTYTLQLDGHR.T	↑
calf1	777	K.CEANLR.V	↑
	783	R.VSFSPAR.S	↑
	858	R.ALSCNVSSPIFK.A	↑
calf2	1027	R.QEILVQVIGTLELVGEIEASSMFSLCSSLSISFNSSK.H	↑
	1064	K.HFHLYGSNASLAQVVMK.V	↑

B

domain	**start site**	**peptide sequence**	**change of oxidation extent after Mg activation**
βI	175	K.TVLPFVNTHPDK.L	↑
	197	K.ECQPPFAFR.H	↑
	210	K.LTNNSNQFQTEVGK.Q	↑
	258	R.LLVFATDDGFHFAGDGK.L	↑
	275	K.LGAILTPNDGR.C	↑
	311	K.LAENNIQPIFAVTSR.M	↑
	340	K.SAVGELSEDSSNVVQLIK.N	↑
hybrid	367	R.VFLDHNALPDTLK.V	↑
	380	K.VTYDSFCSNGVTHR.N	↑
	398	R.GDCDGVQINVPITFQVK.V	↓
I-EGF	489	R.SSQELEGSCR.K	↑
βTD	642	K.NCSAACPGLQLSNNPVK.G	↓

Table 3.1. Peptides identified showing different oxidation extents. A. αL subunit. B. β2 subunit. Their sequences and locations are listed. The peptides from N-terminal to C-terminal of each protein subunit are displayed from the top to the bottom. "↓" means decrease of oxidation extent after Mg activation, and "↑" means increase of oxidation extent after Mg activation.

Furthermore, 143 amino acids from these 26 oxidized peptides are detected and annotated to be probable oxidation sites by manually checking the MS/MS spectra. We also found that there were mixtures of oxidized counterparts to each un-oxidized peptide. As to amino acids with sulfur-containing and aromatic side chains such as Cys (C), Met (M), Phe (F), Trp (W), and Tyr (Y), especially Phe (F), they are highly reactive and readily to be oxidized with +32 Da modification besides the +16 Da modification product (Table 3.2 A, B). We calculated the solvent accessible surface area of these amino acids using Surface Racer 4.0. As to integrin β2 subunit, the calculation was based on the crystal structure of integrin αXβ2 (PDB code 3k6s) which is in the physiological resting condition and adopts the bent conformation (94). As to αL subunit, since there is no intact crystal structure available, we built a computational structural model based on its family member, integrin αX subunit (PDB code 3k6s), which shares 36.7% sequence identity with the αL subunit, using Swiss-Model.

As shown in Table 3.2, consistent with our expectation, among the 143 amino acids detected, most of them are validated as solvent accessible when the protein is in the bent conformation. Some theoretically solvent inaccessible sites are also found to be modified, which are mainly Phe and Tyr (Table 3.2). These two amino acids are among the most reactive residues, Cys, Met, Trp, Phe, and Tyr, which oxidize very readily even not exposed (5). The solvent accessible surface area of these amino acids in turn proves the rationality of the computationally modeled 3D structure.

A.

Domain	Start site	Peptide	Oxidized AA	SASA by X-ray	Mass Shift
				Å^2	Da
αI	153	K.GNVDLVFLFDGSMSLQPDEFQK.I	L167	1.38	16
			M183	2.50	16
	175	K.ILDFMK.D	F178	0.00	16
			K180	79.79	16
			K184	158.39	16
	258	K.VLIIITDGEATDSGNIDAAK.D	L259	9.65	16
			D264	28.90	16
			E266	67.82	16
			T268	135.12	16
			S270	132.78	16
			N272	143.49	16
			I273	91.19	16
			D274	72.93	16
			K277	191.53	16
	331	K.IYVIEGTSK.Q	Y332	83.10	16
			I334	38.75	16
			E335	120.31	16
			S338	125.91	16
			K339	140.70	16
β-propeller	49	R.VLQVGNGVIVGAPGEGNSTGSLYQCQSGTGHCLPVTLR.G	N54	142.63	16
			Y71	30.16	16
			R86	215.26	16
	134	R.QNLQGPMLQGRPGFQECIK.G	Q134	194.17	16
			L136	145.21	16
			Q137	106.16	16
			P139	73.80	16
			M140	49.79	16
			L141	30.47	16
			Q142	88.95	16
			R144	57.98	16
			P145	110.53	16
			F147	171.21	16, 32
			I151	110.39	16
			K152	145.46	16
	494	R.VFIYQR.R	F495	11.50	16
	569	R.IEGTQVLSGIQWFGR.S	E570	42.01	16
			Q573	72.53	16
			L575	27.36	16
			I578	20.90	16
			W580	44.07	32

			F581	2.03	16, 32
thigh	667	R.LVANLTYTLQLDGHR.T	L667	68.86	16
			V668	30.34	16
			Y673	0.00	16, 32
			Q676	69.27	16
			D678	3.07	16
			H680	189.61	16
			R681	79.54	16
			T682	146.81	16
calf1	777	K.CEANLR.V	C777	44.29	32
			E778	91.07	16
			N780	51.24	16
			L781	10.38	16
			R782	149.61	16
	783	R.VSFSPAR.S	F785	16.40	16
			P787	70.65	16
			R789	225.74	16
	858	R.ALSCNVSSPIFK.A	L859	2.34	16
			S860	39.09	16
			V863	60.47	16
			S865	94.65	16
			P866	91.54	16
			I867	75.46	16
			F868	16.06	16, 32
			K869	116.32	16
calf2	1027	R.QEILVQVIGTLELVGEIEASSMFSLCSSLSISFNSSK.H	Q1032	38.77	16
			F1049	34.56	16
	1064	K.HFHLYGSNASLAQVVMK.V	F1065	142.11	16, 32
			H1066	54.81	16
			L1067	62.75	16
			Y1068	3.62	16
			V1077	65.93	16
			M1079	106.36	16
			Y1085	32.58	16

B.

Domain	Start site	Peptide	Oxidized AA	SASA by X-ray	Mass Shift
				Å^2	Da
βI	175	K.TVLPFVNTHPDK.L	V176	42.21	16
			L177	82.20	16
			P178	61.47	16
			F179	2.12	16, 32
			N181	28.74	16
			T182	28.99	16
			H183	128.80	16
			P184	104.67	16
			D185	76.09	16
	197	K.ECQPPFAFR.H	E197	156.17	16
			C198	10.13	32
			P201	10.68	16
			F202	7.97	16
			F204	0.37	16
			R205	89.53	16
	210	K.LTNNSNQFQTEVGK.Q	L210	5.89	16
			N215	101.04	16
	258	R.LLVFATDDGFHFAGDGK.L	L258	14.86	16
			D265	19.39	16
			F269	85.05	16, 32
			D272	34.26	16
	275	K.LGAILTPNDGR.C	L275	120.55	16
			L279	122.96	16
			T280	80.55	16
	311	K.LAENNIQPIFAVTSR.M	E313	105.78	16
			N314	30.79	16
			F320	10.13	16
			T323	14.47	16
			R325	162.18	16
	340	K.SAVGELSEDSSNVVQLIK.N	V342	58.07	16
			E344	95.42	16
			S346	68.40	16
			D348	45.57	16
			S350	65.45	16
			N351	57.96	16
			V353	28.00	16
			Q354	112.95	16
			Y360	53.84	16
			K362	98.42	16
hybrid	367	R.VFLDHNALPDTLK.V	F368	1.37	16, 32
			D370	37.24	16
			H371	50.60	16
			L374	36.53	16
			P375	57.21	16
			L378	2.93	16

	380	K.VTYDSFCSNGVTHR.N	V390	52.84	16
			T391	82.21	16
	398	R.GDCDGVQINVPITFQVK.V	D399	71.61	16
			D401	67.57	16
			Q404	108.20	16
			I405	103.96	16
			N406	81.82	16
			V407	73.59	16
			P408	65.53	16
			T410	54.05	16
			F411	0.00	16, 32
			Q412	51.37	16
I-EGF	489	R.SSQELEGSCR.K	E492	151.46	16
			E494	51.50	16
βTD	642	K.NCSAACPGLQLSNNPVK.G	N642	94.64	16
			S644	83.35	16
			C647	6.49	16, 32
			P648	80.18	16
			L650	131.16	16
			Q651	107.14	16
			L652	41.37	16
			N654	128.56	16
			N655	90.68	16
			P656	112.72	16

Table 3.2. Reactive amino acid residues identified by LC/MS/MS for αLβ2. The solvent accessible surface areas were calculated using Surface Racer 4.0 according to the structural model based on the crystal structure of αXβ2 (PDB code: 3k6s). Site locations of the amino acids are listed in column 1. AA, amino acid; SASA, solvent accessible surface area. **A**. αL subunit; **B**. β2 subunit.

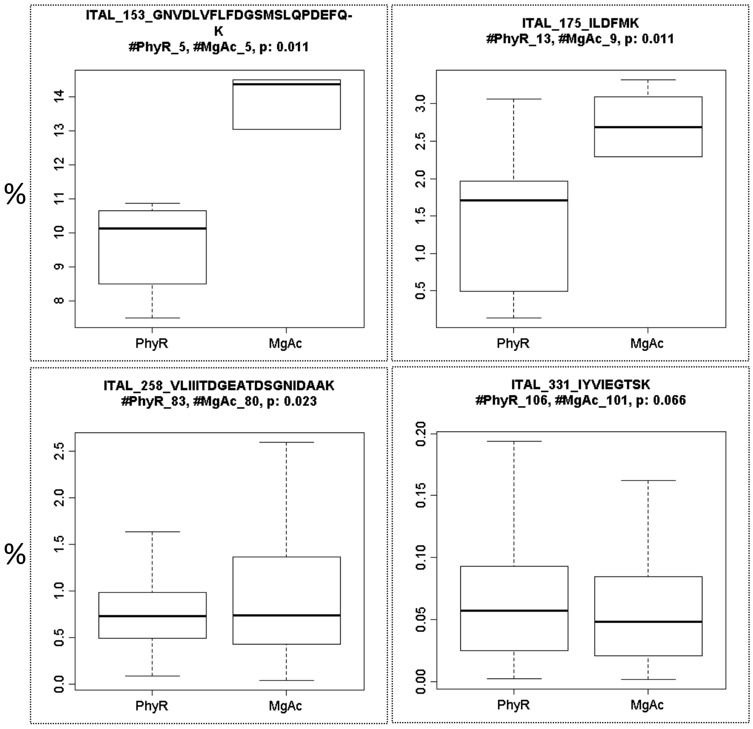
Interpretation of footprinting MS/MS data

In this study, we used tandem mass spectrometry to sequence the tryptic peptide and to locate the oxidized amino acid residues. At the same time, the relative abundance of oxidized counterparts of the tryptic peptide, or the oxidation extent, were measured by SIn method (42). Multiple MS measurements of a sample are required to achieve statistically significant comprehensiveness in proteomics studies. As described above, nine LC-MS/MS replicates of the experiment were done to obtain multiple datasets in this study. SIn largely eliminated variances between replicate MS measurements, permitting quantitative reproducibility in replicate MS measurements of the same and distinct sample (42). It has been proved to be more accurately and statistically robust than other label-free quantification methods including spectral count (SC), unique peptide number (PN), the normalized spectral abundance factor (NSAF), and area under the curve (AUC) methods. Compared with the AUC method, it has the convenience of getting rid of the complexity of peak area measurements (42).

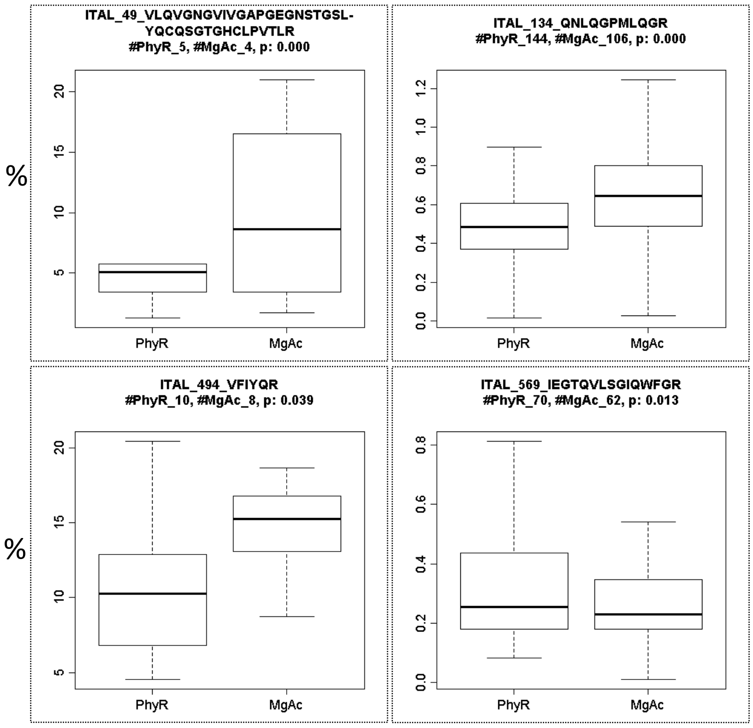
The oxidized amino acid residues provide the surface information of the protein, and thus the surface topology can be mapped. Furthermore, since the oxidation extent depends primarily on the solvent-accessibility of the peptide side chains which is in turn dependent on the conformation of protein, the changes in the relative abundance of oxidized counterparts for each peptide can therefore be used to determine conformation and conformational changes of the protein.

As to αL subunit, fourteen peptides were detected to have significant differences in the relative abundance of their relevant oxidized species after Mg activation. There were four peptides from αI domain, three of them were found to have increased oxidation extent after Mg activation and one with decreased oxidation extent (Figure 3.3A). β-propeller domain showed the similar pattern, three peptides were found to be with increased oxidation extent and one with decreased extent (Figure 3.3B). There was one peptide from thigh domain found to be oxidized to greater extent after Mg activation (Figure 3.3C). As to the αL tailpiece, five peptides from calf1 and calf2 domains were found to have increased oxidation extent in the MgAc group (Figure 3.3C).

A. Peptides from αI domain



B. Peptides from β-propeller domain



C. Peptides form thigh domain, calf1 domain and calf2 domain

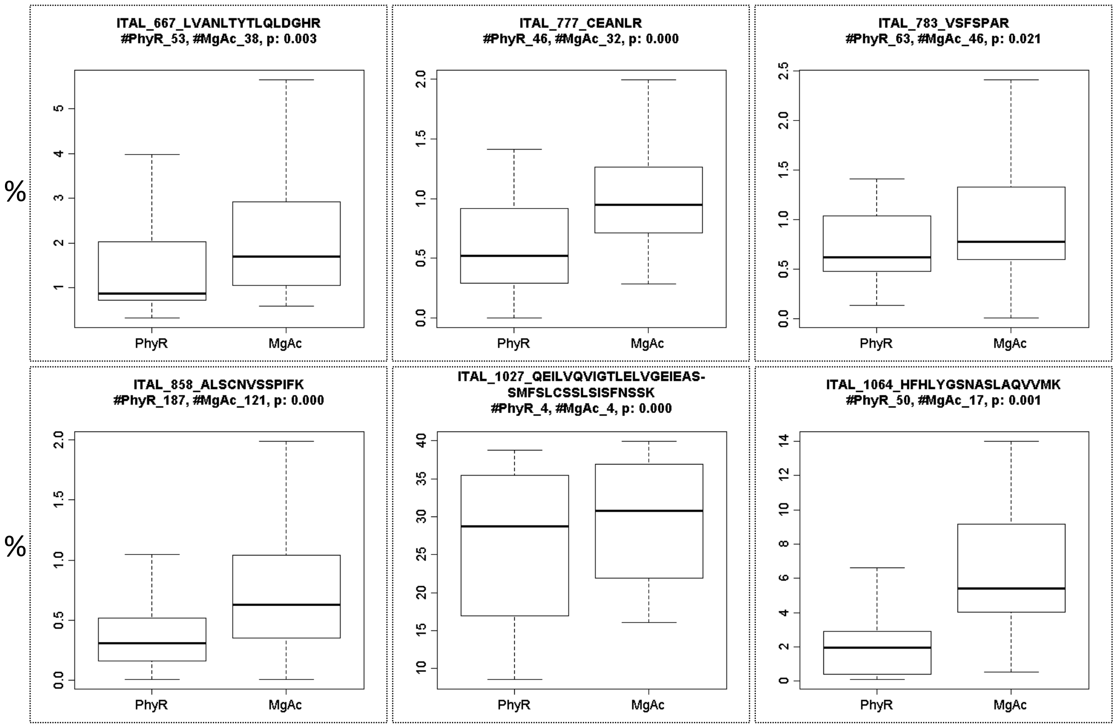
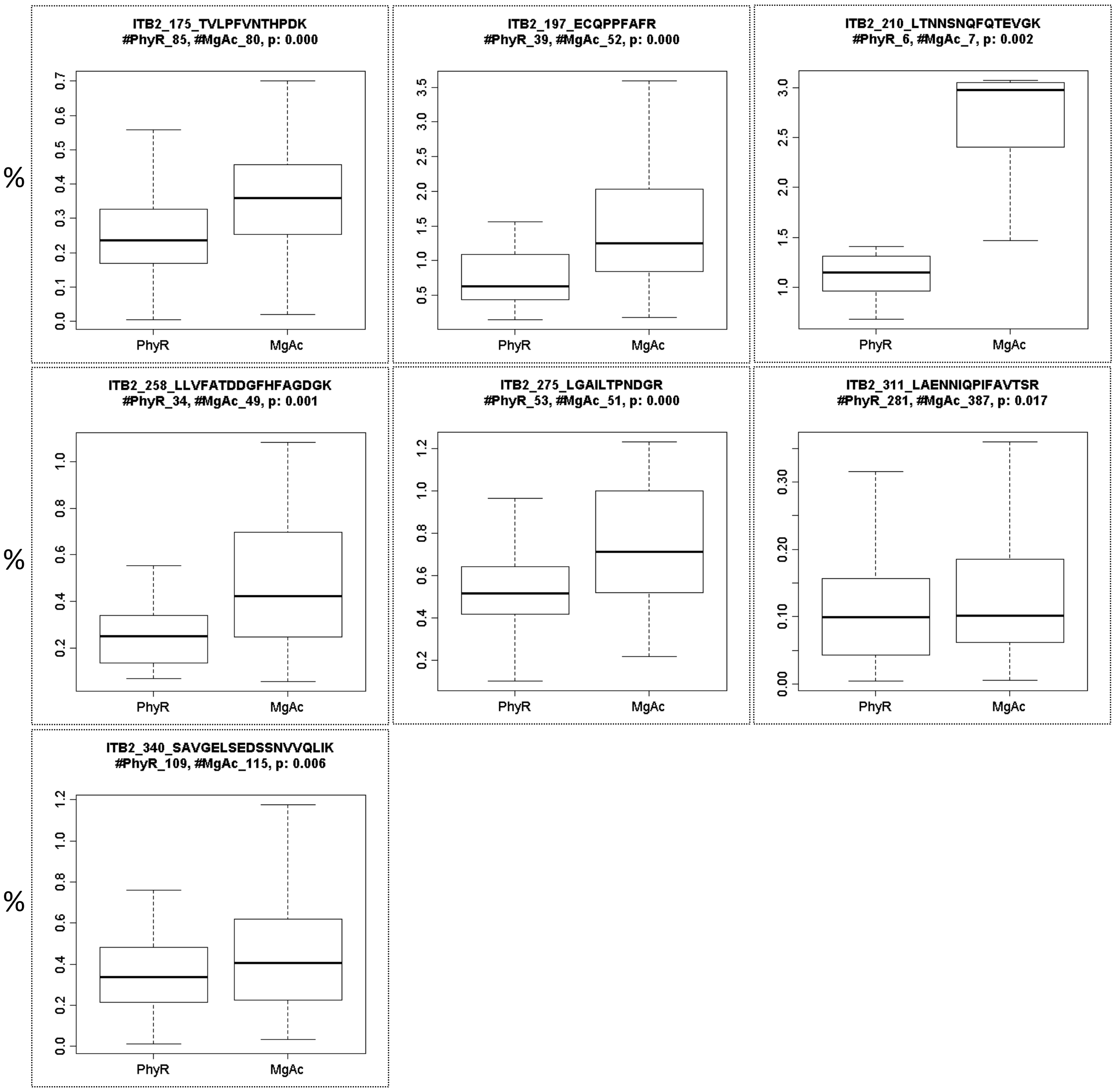


Figure 3.3 Box plot display of the distribution of the relative abundance of the oxidized species of each peptide from αL subunit. The title of each box contains the information of the short protein name shown in Swiss-Prot database, the site of the first amino acid of each peptide, the peptide sequence, and the number of spectra of oxidized species in each group. The Y-axis shows the relative abundance of each oxidized species calculated from the MS2 intensity in percentage (%). A. Peptides from αI domain. B. Peptides from β-propeller domain. C. Peptides from thigh domain, calf1 and calf2 domains.

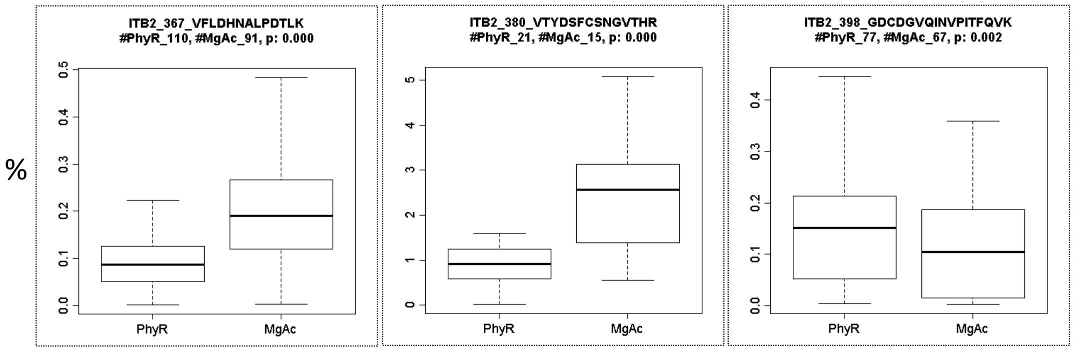
Twelve peptides from β2 subunit were found to have significant differences in the relative abundance of their relevant oxidized species after Mg activation, as shown in Figure 3.4. Seven peptides from βI domain were found to be with increased oxidation extent after Mg activation (Figure 3.4A). As to the hybrid domain, two peptides were found to have higher abundance of oxidized species while one was found to have decreased level of abundance after Mg activation (Figure 3.4B). Two peptides from β2 tailpiece were identified, one with increased oxidation extent and the other with decreased oxidation level after Mg activation (Figure 3.4C).

We can also learn from Figure 3.3 and Figure 3.4 that the oxidation extent varies from peptide to peptide, which is mainly due to the solvent accessibility as well as the chemical reactivity of the amino acid itself.

A. Peptides from βI domain



B. Peptides from hybrid domain



C. Peptides from I-EGF domain and βTD domain

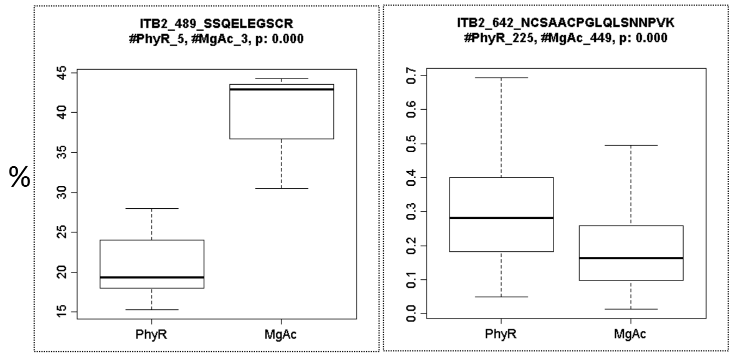


Figure 3.4 Box plot display of the distribution of the relative abundance of the oxidized species of each peptide from β2 subunit. The title of each box contains the information of the short protein name shown in Swiss-Prot database, the site of the first amino acid of each peptide, the peptide sequence, and the number of spectra of oxidized species in each group. The Y-axis shows the relative abundance of each oxidized species calculated from the MS2 intensity in percentage (%). A. Peptides from βI domain. B. Peptides from hybrid domain. C. Peptides from I-EGF domain and βTD domain.

Exposure of αL tailpiece upon Mg^{2+}/EGTA activation

The αL tailpiece is composed of two domains, calf1 and calf2, while the β2 tailpiece is composed of I-EGF 2, 3, 4 and βTD domains (Figure 3.2B). The β-propeller-βI domain interface is the main contact between the two subunits (91, 94), while additional contacts are also seen between the β-propeller-I-EGF 3 and 4 domains, thigh- I-EGF 3, calf2-I-EGF 4, and calf2-βTD when the protein adopts the bent conformation, as is shown in Figure 3.5C, 3.5D.

Footprinting results in this study found that five peptides from αL tailpiece become exposed in the MgAc group after Mg^{2+} activation, as listed in Table 3.1 and Figure 3.3C. Among them 'CEANLR', 'VSFSPAR' and 'ALSCNVSSPIFK' are from calf1 domain and 'QEILVQVIGTLELVGEIEASSMFSLCSSLSISFNSSK' and 'HFHLYGSNASLAQVVMK' are from calf2 domain (Figure 3.5A). αLβ2 integrin from the PhyR group adopts a bent conformation in the physiological resting condition. The two-tailpieces not only interact with each other extensively, but also have interactions with the headpieces as demonstrated above. Thus the peptides from this region are protected from oxidation by OH radicals in the bent conformation.

The increased abundance of the oxidized counterparts of these peptides indicated the exposure of the αL tailpiece upon activation, which would be probably induced by both the separation from the other tailpiece and the separation from the headpiece. As discussed above, there are extensive contacts between the calf1-I-EGF 3, calf2-I-EGF 4, and calf2-βTD when the protein adopts the bent conformation, which is shown in Figure 3.5C, 3.5D. These regions will be exposed to the solvent when the two tailpieces become separated. There is also one observation to support the separation between the headpiece and tailpiece, which is the exposure of the peptide 'ALSCNVSSPIFK' from calf1 domain. We can know from the crystal structure that it's in the crevice between the thigh domain (which corresponds to the headpiece) and calf1 domain (which is part of the tailpiece) in bent conformation. The footprinting MS results showed that the oxidation level of this peptide increased in the activating state (Figure 3.3C), indicating of the release of this region.

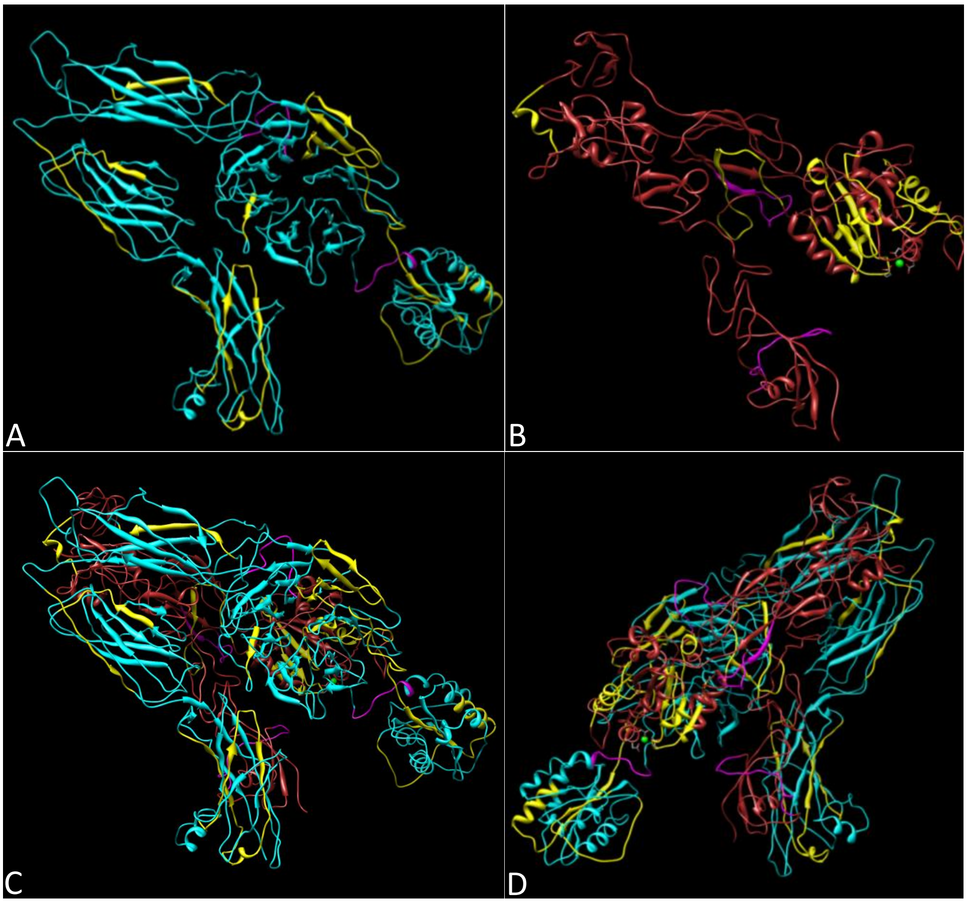


Figure 3.5. 3D Structural model of integrin αLβ2 based on the crystal structure of αXβ2 (PDB code: 3k6s). Ca^{2+} is shown as a green ball in the structure. αL subunit is in cyan and β2 subunit in brown. Peptides labeled in yellow are found to be exposed upon Mg^{2+}/EGTA activation, the ones in magenta are detected to be buried in the activation state. A. αL subunit. B. β2 subunit. C. Interaction of αLβ2 heterodimer from the view of αL side. D. Interaction of αLβ2 heterodimer from the view of β2 side.

Release of thigh-I-EGF 3 interface upon Mg^{2+}/EGTA activation

I-EGF 3 domain interacts extensively with thigh domains of αL subunit in the bent conformation. The peptide ‘LVANLTYTLQLDGHR’ was found to be with greater oxidation extent after Mg activation in this study (Figure 3.3C).

We can know from the structure that it is located at the outer side of the thigh domain which interacts with both I-EGF 2 and I-EGF 3domains (Figure 3.5A,D). The result indicated the separation of the thigh and I-EGF 2/3 domains, and also the release of of thigh-I-EGF 3 interface.

Changes to β-propeller domain of αL subunit

Most parts of β-propeller domain are buried in the bent conformation because of the large interface induced by the headpiece-headpiece interaction as well as the headpiece-tailpiece interaction (Figure 3.5). After extension, some parts of the β-propeller domain will still interact with the βI domain, and some parts will still have contacts with αL- thigh domain and αI domain. In this study, three peptides were found to be with increased oxidation level upon activation by Mg^{2+}/EGTA (Figure 3.2B). They are ‘VLQVGNGVIVGAPGEGNSTGSLYQCQSGTGHCLPVTLR’, ‘QNLQGPMLQGRPGFQECIK’ and ‘VFIYQR’, as shown in Figure 3.5A. Meanwhile, one peptide ‘IEGTQVLSGIQWFGR’ was found to be with decreased oxidation level upon activation by Mg^{2+}/EGTA (Figure 3.3B, 3.5A).

Taken these findings together, Mg^{2+}/EGTA activation induces great changes to the propeller-βI domain interface. The upper part of propeller domain becomes more exposed while the lower part became shielded by the newly formed propeller-βI domain interface.

αI domain was found to be greatly exposed to solvent after Mg activation. Three peptides from this domain were found to be with increased oxidation level after Mg activation (Figure 3.3A). The αI domain is inserted between blades 2 and 3 of the αL β-propeller domain (Figure 3.2A and 3.5A), and αI domain itself has no contact with the β-propeller (94). But since the legs were severely bent, the headpiece was closely juxtaposed to the membrane-proximal portions of the legs. It's challenging for OH radicals to access this region when the protein adopts the bent conformation and peptides from this domain would be protected under this condition. After Mg activation, the overall extension happens and the αI domain becomes exposed to the solvent, thus peptides from this domain can be oxidized.

Seven peptides from βI domain were also found to be exposed upon activation by Mg^{2+}/EGTA (Figure 3.4A). The βI domain is almost buried by the βI–β-propeller interface and the headpiece-tailpiece interaction in the bent conformation. Footprinting results showed that it becomes greatly exposed after Mg^{2+}/EGTA activation, which demonstrated the overall extension of the protein from one side.

αI and βI domains have structurally homologous α/β folds, that is, a central β-sheet surrounded by α-helices and a MIDAS at the C-terminal ends of the central β-strands, i.e., the 'top' face (91). In αI domain containing integrins such as αLβ2, the αI domain binds the acidic residue of the ligand through its

MIDAS, whereas the βI domain regulates the ligand binding by the αI domain (111). The βI MIDAS of αI integrins may bind to an intrinsic ligand, that is to an invariant Glu residue that corresponds to the first residue in the αI C-linker. Based on our current knowledge, residue αL-GLu-310 (namely Glu-335 in the primary sequence) locates in the αI C- linker that follows the C-terminal α7-helix of αI domain. When activated, the βI domain MIDAS binds αL-GLu-310 and exerts a downward pull on the C-terminal α7-helix of the αI domain that activates the αI domain (112). The interaction of αL-GLu-310 and the metal of βI domain MIDAS induces the high-affinty intermediate conformation of αI domain.

The two linkers connecting the αI and β-propeller domains are flexible, and αI domain itself has no contact with the β-propeller (94). In αXβ2 integrin, flexibility of the N-terminal (N) linker to the αI domain is limited by its short length; however, the long, 10-residue Ser- and Thr-rich C-terminal (C) linker is highly flexible. This is similar in the αI domain of αLβ2 integrin. The αI domain C-linker peptide 331-339 ‘IYVIEGTSK’with the intrinsic invariant Glu (E) residue was identified in this study. The oxidation extent of this peptide was found to be slightly decreased in MgAc group (Figure 3.3A, 3.5). Although not that significant ($p=0.066$), the result suggested certain interaction of the linker with the βI due to the coordination of αL Glu-310 and the β2 MIDAS (Figure 3.5C, D).

The observations together support the overall αLβ2 activation. Upon Mg^{2+}/EGTA activation, a new interface surrounding the α7-helix-linker and the βI-MIDAS region will form. The newly formed interface takes the role of the previous one between the β-propeller and βI domains and induces regional changes to them. The allostery of other domains which connect to these domains change accordingly to the motions.

The β2 genu and PSI/hybrid/I-EGF1 complex

In integrin extension, interdomain motions play a key role in the structural changes through the molecule. By far the largest changes in interdomain orientation occur at the α- and β- genu (94, 97, 113), which transit from highly bent to extended conformations (Figure 3.2B). Within β2 the greatest interdomain variation is seen at hybrid/I-EGF1 and I-EGF1/I-EGF2 interfaces at the β genu. Some studies proposed that ligand binding may involve the outward swing of the hybrid domain with respect to the βI domain and that the hybrid domain forms rigid complex with the PSI domain (87, 94, 114-115).

Peptides 'VFLDHNALPDTLK' and 'VTYDSFCSNGVTHR' from hybrid domain were found to become more exposed upon activation (Figure 3.4B). These two peptides locate at the inner side of hybrid domain, which interacts extensively with the β-propeller, I-EGF2, I-EGF3 domains and is protected by the large interface (Figure 3.5B). These observations together indicate the release of hybrid domain during the extension.

Peptide with the sequence 489-499 'SSQELEGSCRK' is found to be exposed upon activation (Figure 3.4B). It's from the I-EGF 2 domain and locates right at the I-EGF1/I-EGF2 interfaces at the β genu (Figure 3.5B). This area interacts with the I-EGF2 and thigh domain in the bent conformation.

As to β2 subunit, there were also two peptides found to be protected after Mg activation. One is from the hybrid domain, with the sequence ‘GDCDGVQINVPITFQVK’ (Figure 3.4B) and the other one ‘NCSAACPGLQLSNNPVK’ is from βTD domain (Figure 3.4C). Based on our current knowledge, it seems difficult to explain why the two peptides show the decreased oxidation levels when the protein adopts the extended conformation. It may be because that there is certain degree of rotation of β2 subunit during the extension. More work need to be done in future to work out how exactly the conformational change of αLβ2 integrin happens upon Mg/EGTA activation.

Conclusion

In summary, we applied the nanosecond pulse laser photolytic hydroxyl radical footprinting method to study the conformational change of αLβ2 integrin upon Mg^{2+}/EGTA activation on the cell surface. It is well known that in the resting bent conformation, the headpiece of the integrin folds over its legs at the genu and faces downward the membrane, the legs were severely bent, generating a V-shaped topology in which the headpiece was closely juxtaposed to the membrane-proximal portions of the legs. Upon activation, the legs extend upward in a switch-blade like opening, and dramatic rearrangements occur in the orientation of different domains of integrin. Our *in vivo* footprinting technique showed these phenomena from one side and demonstrated the activation and overall extension of αLβ2 heterodimer. The technique is proved to be helpful in both the structural and funtional studies of plasma membrane proteins of live cells.

Chapter 4. Hydroxyl Radical Footprinting Study of EGF-EGFR Interaction Directly on Living Cells Reveals the Role of Glycosylation on EGF Acitvation

Introduction

Epidermal growth factor (EGF) receptor (EGFR) belongs to the ErbB family of receptor tyrosine kinases (RTKs) responsible for cell proliferation, differentiation and migration in embryogenesis and carcinogenesis (116). The protein tyrosine kinase activities of the EGFR family of RTKs are normally subject to multiple regulatory mechanisms before and after growth factor stimulation. EGFR stimulation is proposed to be preceded by receptor dimerization, a conformational change in the ectodomain that exposes a loop (dimerization arm) required for receptor association (117-118). Although the fact that dimerization of EGFR is known to be essential for signal transduction, the detailed process of dimerization has not been fully understood.

It was noted that a certain portion of EGFRs is present as dimers even before interaction with EGF (119). Molecular forms of EGFR present in cells include monomers, predimers, heterodimers, and homodimers, or signaling dimers (120-121). A heterodimer is formed when an EGF molecule activates one of the monomers in a predimer by binding in its extracellular domain, and a signaling dimer is formed when both the monomers in a predimer are bound with two EGF molecules. In recent years, developments in single cell imaging analysis and single molecule imaging technique have revealed the existence of a minor population of EGFR predimers and the fluctuation between the monomers and predimers, and further found that the formation of the predimeric binding sites of EGFR makes the EGF signaling very rapid and sensitive (119, 122-126). These techniques, together with mathematical modeling, revealed that the binding of EGF to an EGFR dimer is more than ten times, in some cases two orders of magnitude faster than to an EGFR monomer (125, 127-129). Thus, the formation of homo-and heterodimers of EGFRs is regarded as an essential feature of the signaling mechanism and the notion that predimers are primed for EGF binding and signaling is currently widely accepted by researchers (122, 124-125).

It was also noted that only a small fraction of EGFR needs to be participated to induce EGF signaling (125-126). Quantitative studies of ^{125}I-labeled EGF binding to living cells bearing EGFRs have demonstrated that EGFRs exist on the surface in two different affinity classes for more than two decades (130). The first is a majority of low-affinity receptors that bind EGF with a K_D of ≈2-5 nM, accounting for 95~98% of the receptors. The second is a minority of high affinity with a K_D of ≈10-100 pM which accounts for 2~5% of EGF receptors. Cells sensitive to EGF typically express hundreds of thousands of, or millions of EGFRs on the plasma membrane, while it was found that EGF-induced global cellular responses occurred immediately after application of EGF at less than 1% occupancy of EGFRs on single cell surface (122, 126, 131). Single cell imaging studies showed that the high affinity receptor subpopulation is present in a dimerized state in intact cells, even in the absence of the EGF (119, 132). Studies also showed that the high affinity receptor subpopulation itself is required and sufficient for EGF function (133-135). The combined observations that dimerization of the EGFR parallels receptor activation, that receptor dimerization increases the affinity for EGF, and that signal transduction can be effected via the high affinity receptors alone have pointed to a linkage between high affinity EGF binding, receptor dimerization and activation. However, the detailed underlying molecular mechanism remains uncertain.

X-ray crystal structures of the EGFR extracellular ligand-binding domain revealed two distinct conformational states: an intramolecularly tethered monomeric autoinhibited configuration and an extended dimeric configuration (136-138). EGFR is a type I membrane protein with a single polypeptide chain that spans the membrane once via an α-helical segment of 23 hydrophobic amino acids (139). The intracellular domain (542 amino acids) is the locus of a latent tyrosine kinase. The extracellular region (621 amino acids without the signal peptide) contains four subdomains, designated domains I, II, III, and IV (Figure 4.1A) (140). Domains I and III (alternatively called L1 and L2) share approximately 37% amino acid sequence identity, while domains II and IV (also termed CR1 and CR2) are cysteine rich domains, and share 17% sequence identity (141). In the tethered conformation, a loop in domain II, termed the dimerization arm, interacts with a disulfide-bonded loop termed the tether loop in domain IV, thus domain II is completely occluded by intramolecular interactions with domain IV (Figure 4.2A). In the extended dimeric conformation, the complex contains two EGF molecules and two EGFR molecules, and EGF is bound by primary contacts with receptor domains I and III of the same receptor molecule and is distant from the dimer interface (Figure

4.2B). The dimerization arm in domain II interacts with the one of the adjacent receptor of the complex. On the basis of both the functional and structural studies of the ectodomain of EGFR, it was proposed that the tethered configuration corresponds to the low-affinity receptor, whereas the extended configuration accounts for the high affinity EGFRs on intact cells.

As to the nature of low and high affinity EGFRs on living cells, although there is limited experimental evidence, it is generally accepted by researchers that it is related to glycosylation, either the location of the glycosylation site or the sequence of the glycan (142-143). Glycans mediate and/or modulate many cellular interactions, and are involved in a broad range of biological processes, including intercellular adhesion, signaling and microbial attachment (144). A major function of protein-linked glycans is to provide additional recognition epitopes for protein receptors, and it was found that once the EGFR is glycosylated, it acquires ligand binding ability (145-147). Potential N-glycosylation sites can be identified by the presence of the Asn-X-Ser/Thr (where X is any amino acid except proline) sequon in peptide sequence. The EGFR extracellular domain for ligand binding is heavily glycosylated, and twelve N-glycosylation sites have been characterized, with the glycans contributing nearly 40kD to the 170kD molecular weight (Figure 4.1B) (142). Matured EGFR was also reported to contain both complex-type and high mannose-type N-glycans (147-148). The large number of glycosylation sites as well as different types of glycans suggests that EGFR exist as many different forms in terms of glycosylation, which may induce the low and high affinity subpopulations of EGFR.

Hydroxyl radical protein footprinting technique has been shown to be a powerful method for understanding ligand-induced conformational changes when coupled with existing structural data that are either experimentally or computationally generated (11, 102-103). Recently we have extended its application to study the *in vivo* structural dynamics of IMPs, and successfully illustrated the *in vivo* voltage gating of porin OMPF of *E. coli* using Fenton chemistry (104). Here in this study we further implemented the nanosecond laser photochemical hydroxyl radical footprinting method (15) to study the structure-functional relationship of EGF binding of EGFR directly on the living cell surface. The EGF binding of EGFR has been well and extensively characterized that provides the framework for interpretation of our footprinting results. In this study, we separated two different glycosylated EGFR protein subpopulations by SDS-PAGE, characterized the glycosylation sites and investigated the relevant glycosylaton levels of each population, to help to understand the relationship of glycosylation and EGFR function.

Footprinting results demonstrated that domain I and domain III of EGFR were involved in EGF binding. In addition, two EGFR subpopulations with different glycosylation levels have been characterized, and it was further determined that the two EGFR subpopulations showed different EGF binding affinity. Furthermore, two glycosylation sites from domain III of EGFR, N413 and N444, were quantified to have lower glycosylation level in the minority lower molecular weight population of EGFR, which helps to differentiate the two EGFR subpopulations. These observations not only consist with what have been discovered in other studies of ErbB family members, but also help to reveal the difference of low and high EGF binding affinity populations of EGFR. Thus the nanosecond laser photochemical footprinting method is proved to be helpful in both the structural and funtional studies of plasma membrane proteins of live cells.

Materials and methods

Antibodies and reagents

Recombinant human EGF (Cell Signaling Technology) was from Millipore. EGF Receptor (EGFR1) mouse monoclonal antibody (#2256, Cell Signaling Technology) used for western blot in this study was from Millipore. This antibody raises against the extracellular domain of human EGF receptor. It specifically detects endogenous EGF receptors from various cell lysates, and does not cross-react with other EGF receptor family members. EGFR mouse monoclonal antibody (sc-71032, Santa Cruz Biotechnology) used for IP was from vCell Science Pte Ltd. This antibody raised against amino acids 985-996 of EGFR of human origin. Protein G agarose beads (Upstate) was from Millipore.

Heat-inactivated fetal bovine serum (FBS, Hyclone) was from Thermo Fisher Scientific, Waltham, MA. Phosphate buffered saline (PBS, GIBCO) was from Invitrogen. Dulbecco’s modified Eagle’s medium (DMEM) was obtained from Sigma. Penicillin and streptomycin were from Invitrogen. FreeStyle 293 Expression Medium (GIBCO) was obtained from Invitrogen. The pcDNA3-EGFR vector for expression of EGFR was kindly provided by Professor Yosef Yarden (Weizmann Institute of Science) (149).

Complete protease inhibitor cocktail tablets were from Roche Diagnostics (Indianapolis, IN, USA). Sequencing grade modified trypsin for protein digestion was from Promega. The 4-12% Bis-Tris gradient gel was from Invitrogen. HPLC grade solvents were purchased from J.T. Baker (Phillipsburg, NJ, USA). H_2O_2 was from Merck, Singapore. All the other chemicals were purchased from Sigma-Aldrich unless otherwise indicated.

Cell Culture and Transfection

293-F cells were maintained under an atmosphere of 5% CO_2 in FreeStyle 293 expression medium in a shaking incubator (150 rpm) at 37 ℃. The day before transfection, determine the number of cells needed for transfection and pass cells at 6×10^5 cells/ml. On the day of transfection, determine the density and viability of cells to make sure the cell density is about 10^6cells/ml and the viability of cells is over 90%. For expression of EGFR, F293 cells were transfected for 24 h with the corresponding pcDNA3-EGFR plasmid and 293Fectin; the ratio of plasmid to 293Fectin is optimized to be 1:3.

A431 cells were cultured up to ~95% confluence in DMEM medium supplemented with 10% heat-inactivated FBS and penicillin-streptomycin in a 5% CO_2 humidified atmosphere at 37 ℃. The cells were trypsinized and washed with ice cold PBS.

For each cell line, there were three groups of cell samples in this study. The first was the control group (abbreviated as ‘Ctrl’) containing cells without any treatments and oxidation. Cells in the second group were harvested and then oxidized *in vivo* by laser footprinting, which we named as the normal group (abbreviated as ‘Norm’). Cells in the third group were first activated with EGF, and were then oxidized with laser footprinting. This group is thus named as the ‘EGF’.

Western blotting

Western blotting was performed using the primary anti-EGFR mAb (#2256) at the dilution of 1:1000, and the secondary antibody at the dilution of 1:5000.

EGF stimulation

EGF was added in ice cold PBS supplemented with 50 mM glucose, 5 mM $MgCl_2$, 8.7 mM $CaCl_2$ at a final concentration of 100 nM (132). In the experiments on the EGF activation of EGFR, the cells were prewashed two times with PBS and incubated with the EGF solution described above with gentle agitation for 1 h at 4 ℃.

Flash oxidation of living cells using a nanosecond pulsed laser

In the EGF activation group, cells were washed with ice cold PBS once following EGF activation, and resuspended in ice cold PBS for laser irradiation (cell density ~$3\text{-}5\times10^6$). In the normal group without EGF activation, cells were washed and resuspended in ice cold PBS for laser irradiation (cell density ~3-5×10^6).

The cell suspension for each group of cells was then aliquoted into a 12-well plate, with each well containing 1.5 ml of suspension. H_2O_2 was added to the cell suspension to a final concentration of 1% before laser irradiation. Cells were then oxidized with OH radicals generated by exposing the suspension to a pulse Nd: YAG laser (LOTIS TII LS-2134UTF, Minsk, Belarus) operating at 266 nm and 30 mJ. The 12-well plate was held horizontally, and each well opening was properly aligned with the laser beam. The schematic overview of the flash oxidation was shown in Figure 3.1 in Chapter 3. Immediately after one laser shot, appropriate amount of catalase was added to remove the residual H_2O_2. Cells were collected subsequently by spinning down at 500g, 4 ℃ for 10 min.

Samples were prepared in triplicates to minimize the biological variation among each group. About 3×10^7 cells were used for each sample. The experiment was repeated for five times.

Immunoprecipitation

Cells were lysed in lysis buffer [10 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1% (v/v) NP-40] containing appropriate protease inhibitors cocktail. Cell debris was removed by spinning down at 13,000×g, 4 ℃ for 15 min. The supernatant was pre-cleared with protein G beads first. EGFR antibody was added to the precleared supernatant and incubated with end to end rotation in cold room for 3 h followed by further incubation with the protein G beads for 1 h. Before elution, beads were washed with ice cold lysis buffer for five times. Bound proteins were eluted with 3×150 µl of glycine-HCl (pH 1.9). The amount of enriched EGFR protein was estimated around several µg for each sample. Protein samples were subjected to 4-12% gradient gel for resolving.

In-gel digestion

EGFR protein bands from the SDS page were cut out according to their molecular weight after coomassie blue or silver staining. The glycans were enzymatically released from the gel bands by incubation with PNGase F (500 units/µL) in 25 mM NH_4HCO_3 at 37 ℃ overnight. Samples were reduced with 10 mM dithiothreitol in 25 mM NH_4HCO_3 at 56 ℃ for 1 h, followed by alkylation with 55 mM iodoacetamide in 25 mM NH_4HCO_3 solution at room temperature in dark for 45 minutes and digested by sequencing grade modified trypsin (Promega Corporation, Madison, WI. USA) in a 1:100 (trypsin: protein) mass ratio at 37 ℃ overnight.

LC-MS/MS analysis

The LC-MS/MS setup can be referred to Figure 2.2 in Chapter 2. The LC-MS/MS analysis was done in an LTQ-FT Ultra mass spectrometer (Thermo Electron) coupled with a prominence™ HPLC unit (Shimadzu). The tryptic peptides were reconstituted to 100 μl of 0.1% formic acid (FA) in HPLC grade water. The samples were then injected by an autosampler and on-line desalted in a Zorbax peptide trap (Agilent, Pola Alto, CA). The peptide separation was performed in a home-packed C18 column (75 μm inner diameter × 10cm, 5 μm particles, 300 Å pore size, Michrom BioResources, Auburn, CA, USA). Buffer A (99.9% H_2O, 0.1%FA) and buffer B (99.9% ACN, 0.1%FA) were used for the LC gradient. The 90-min gradient was ramped from 5% ACN to 30% ACN in 65 min, then to 60% in 10 min and to 80% ACN over 2 min and then was kept at 80% ACN for 3 min and ramped back to 5% ACN for the last 10 min. The sample was ionized into the mass spectrometer through an ADVANCE™ CaptiveSpray™ Source (Michrom BioResources) with an electrospray potential of 1.5 kV operating at a flow rate of about 200nl/min after a splitter. The nitrogen gas flow was set at 2, ion transfer tube temperature at 180 ℃ and collision gas pressure at 0.85 mTorr. The LTQ-FT Ultra was set to perform data acquisition in the positive ion mode as described previously (106). Briefly, a full MS scan (350-2000 m/z range) was acquired in the FT-ICR cell at a resolution of 100,000 and a maximum ion accumulation time of 1000 msec. The AGC target for FT was set at 1e+06 and precursor ion charge state screening was activated. The linear ion trap was used to collect peptides and to measure peptide fragments generated by collision-activated dissociation (CAD). The default AGC setting was used (full MS target at 3.0e+04, MS^n1e+04) in linear ion trap. The 10 most intense ions above a 500 counts threshold were selected for fragmentation in CAD (MS^2), which was performed concurrently with a maximum ion accumulation time of 200 msec. Dynamic exclusion was activated for the process, with a repeat count of 1 and exclusion duration of 20 s. Single charged ion is excluded from MS/MS. Isolation width was 2 Da, and default charge state was 5. For CAD, normalized collision energy was set to 35%, activation Q was set to 0.25, and activation time 30 ms. Spectra were acquired in centroid format in raw data files with XCalibur (version 2.0 SR2).

Bioinformatics analysis

MS/MS data for EGFR footprinting experiments were submitted to SEQUEST protein sequence database search engine (BioworksBrowser 3.3, Thermo) for peptide/protein identification against an in-house EGFR database. Peptide tolerance was 8 ppm, and fragment ions tolerance was 1AMU. Full enzymatic cleavage by trypsin was selected, with maximum 2 missed cleavages allowed. Two predominant kinds of OH radical oxidations, one with a mass increase of +16 Da for almost all the amino acid residues, and the other with a mass increase of +32 Da for five residues C, F, M, W and Y, were included as variable oxidative modifications. We also enabled the following non-oxidative modifications, carbamidomethylated cysteine (+57 Da), and deamidated N and Q (+0.98 Da) as variable modifications. Maximum of three oxidative modifications and in total six modifications were allowed for a single peptide. Identified peptides were subject to correlation analysis as described below to remove low-quality identifications (107). For each unique peptide, all the spectra matched were pooled, and divided into two groups according to their XCorr (cross correlation) values. The qualified group contains spectra with minimal XCorr cutoffs of 1.0, 1.5, and 2.0 for +1, +2, and ≥+3 charged peptides, respectively (108). As to those spectra with lower XCorr values, they were subsequently correlated with each of the qualified spectrum by correlation analysis using a perl package Statistics::LineFit (version 0.06, from Comprehensive Perl Archive Network, CPAN). P values of correlation were calculated using another perl package Statistics::Distributions (version 1.02). For each spectrum, the top 10 peaks, where available, from every 100 m/z range, were used to simplify the spectrum. The top 50 peaks from the resulting spectrum were submitted for correlation analysis. Spectra with fewer than 50 peaks were excluded. Only spectra with minimum R-square value of 0.81, correlation between 0.8 and 1.25, and p values<0.05 were "rescued" and included into the qualified group. After identification and qualification, normalized spectral index (SIn) was calculated to represent the abundance of oxidized and un-oxidized peptides (42). SIn values of peptides were grouped based on the longest peptide sequence without miss-cleavage, and presented using box-plot in statistical platform R. Student *t* test was used to calculate significance for difference between experiment conditions.

Calculation of solvent-accessible areas of amino acids

Exact accessible surface area of molecular surface for macromolecules was calculated by the program Surface Racer 4.0 (65) based on the crystal structure of EGFR (1nql for tethered monomer).

Results and discussion

Detection of oxidized peptides of EGFR by LC-MS/MS analysis

We used the suspension cell, 293-F cell, in this study for transient transfection of EGFR for EGF binding. For EGFR expressed by 293-F cell line, the sequence coverage of EGFR detected in the three groups-Ctrl, Norm, and EGF- are 91%, 84% and 86%, respectively as calculated from LC-MS/MS results.

In this study, tandem mass spectrometry was used to sequence the oxidized protein and to locate the oxidized amino acid residues. As described above, two most frequent modifications +16 Da, +32 Da were included as variable oxidative modifications. Oxidized peptides thus were detected by the increased peptide mass with an integral multiple of +16 Da or +32 Da, and the modified sites were then confirmed by manually inspection of the mass spectra. The relative abundance of oxidized counterparts of the tryptic peptide, or the oxidation extent, can be measured by SIn method (42). Five replicates of the experiment were performed to achieve statistically significant comprehensiveness. SIn largely eliminated variances between replicate MS measurements, permitting quantitative reproducibility in replicate MS measurements of the same and distinct samples (42). It has been proved to be more accurately and statistically robust than other label-free quantification methods including spectral count (SC), unique peptide number (PN), the normalized spectral abundance factor (NSAF), and area under the curve (AUC) methods. Compared with the AUC method, it has the convenience of without the complexity of peak area measurements.

To validate the footprinting results, we further included another cell line, the A431 cell which highly expresses EGFR, to study the EGF binding. As to A431 cell line, LC-MS/MS result showed that the sequence coverage of EGFR in the three groups-Ctrl, Norm, and EGF-are 58%, 56%, and 60%, respectively. The cause of the difference lies in the different starting amount of cell lysate for IP. As to 293-F cell, there was around 20 mg of lysate used in IP for each sample in each time experiment. While as to A431 cells being an adhesion cell, detaching the cells from the cell culture disk is necessary, more cells broke and were lost during the further oxidation process because of the steps of trypsinization and scrapping. And only around 10 mg of A431 cell lysate was obtained for each sample in each time experiment. The use of both 293-F cells and A431 cells in this study can help to minimize false positives.

Among the tryptic peptides identified, most of them showed no difference in the abundance of oxidized species, while eight peptides from the two datasets (293-F and A431cells) were found to show large changes in the relative abundance of their relevant oxidized species after EGF binding, as shown in Table 4.1 and Figure 4.3. They are mainly located in domains I and III of EGFR.

Domain	Start site	Peptide sequence	Change of oxidation extent after EGF binding	Cell line
I,25-189	38	LTQLGTFEDHFLSLQR	↓	293-F
	81	TIQEVAGYVLIALNTVER	↑	293-F
	99	IPLENLQIIR	↓	293-F
	150	FSNNPALCNVESIQWR	↓	293-F, A431
III,335-504	336	VCNGIGIGEFK	↓	A431
	361	NCTSISGDLHILPVAFR	↓	293-F
	400	EITGFLLIQAWPENR	↓	A431
	415	TDLHAFENLEIIR	↓	293-F

Table 4.1. Peptides identified showing different oxidation extents. Their sequences and locations are listed. The peptides from N-terminal to C-terminal of each protein subunit are displayed from the top to the bottom. "↓" means decrease of oxidation extent after EGF binding, and "↑" means increase of oxidation extent after EGF binding.

Furthermore, 35 amino acids from these eight oxidized peptides are detected and annotated to be probable oxidation sites by manually checking the MS/MS spectra. We also found that there were mixtures of oxidized counterparts to each un-oxidized peptide. As to amino acids with sulfur-containing and aromatic side chains such as Cys (C), Met (M), Phe (F), Trp (W), and Tyr (Y), especially Phe (F), they are highly reactive and readily to be oxidized with +32 Da modification besides the +16 Da modification product (Table 4.2). We calculated the solvent accessible surface area of these amino acids using Surface Racer 4.0. The calculation was based on the crystal structure of the tethered EGFR (PDB code 1nql). As shown in Table 4.2, consistent with our expectation, among the 35 amino acids detected, most of them are validated as solvent accessible when the protein adopts the tethered conformation. Some theoretically solvent inaccessible sites are also found to be modified, which are mainly Cys, Phe and Tyr (Table 4.2). These three amino acids are among the most reactive residues, Cys, Met, Trp, Phe, and Tyr, which oxidize very readily even not exposed (5). We would like to point out that the conformation of the native full length EGFR on cell surface may be slightly different from the tethered EGFR structure, this may also account for the difference between the footprinting result and the crystal structure result.

Domain	Start site	Peptide	Oxidized AA	SASA by X-ray	Mass Shift
				Å^2	Da
I, 25-189	38	LTQLGTFEDHFLSLQR	L38	3.17	16
			Q40	18.66	16
			E45	100.39	16
			H47	9.12	16
			S50	13.35	16
	81	TIQEVAGYVLIALNTVER	T81	44.31	16
			E84	15.15	16
			Y88	4.70	16, 32
			T95	12.71	16
			V96	4.93	16
	99	IPLENLQIIR	P100	6.24	16
			N103	62.43	16
	150	FSNNPALCNVESIQWR	F150	0.00	16, 32
			P154	36.36	16
			E160	99.17	16
			Q163	50.63	16
III, 335-504	336	VCNGIGIGEFK	V336	78.96	16
			C337	11.96	16
			I342	91.64	16
			E344	104.40	16
			F345	17.16	16, 32
			K346	56.48	16
	361	NCTSISGDLHILPVAFR	C362	0.00	32
			T363	15.12	16
			I372	27.54	16
			P373	55.42	16
			V374	21.19	16
			F376	48.03	16
	400	EITGFLLIQAWPENR	T402	16.01	16
			F404	9.73	16, 32
	415	TDLHAFENLEIIR	H418	48.19	16
			F420	0.00	32
			N422	65.16	16
			I425	36.00	16

	R427	83.50	16

Table 4.2. Reactive amino acid residues identified by LC/MS/MS for tethered EGFR. The solvent accessible surface areas were calculated using Surface Racer 4.0 according to the crystal structure of EGFR (PDB code: 1nql). Site locations of the amino acids are listed in column 1. AA, amino acid; SASA, solvent accessible surface area.

Domain I and domain III are involved in EGF binding

As discussed above, there are eight peptides showing significant changes in the oxidation extent after EGF binding. We can know from the backbone sequence that four of them are from domain I and another four peptides are from domain III (Table 4.1, Figure 4.1). Most of the peptides were detected to have a significantly decreased relative abundance of their oxidized counterparts after EGF binding except for one (Figure 4.3), which indicated that these regions became buried after EGF binding and were protected.

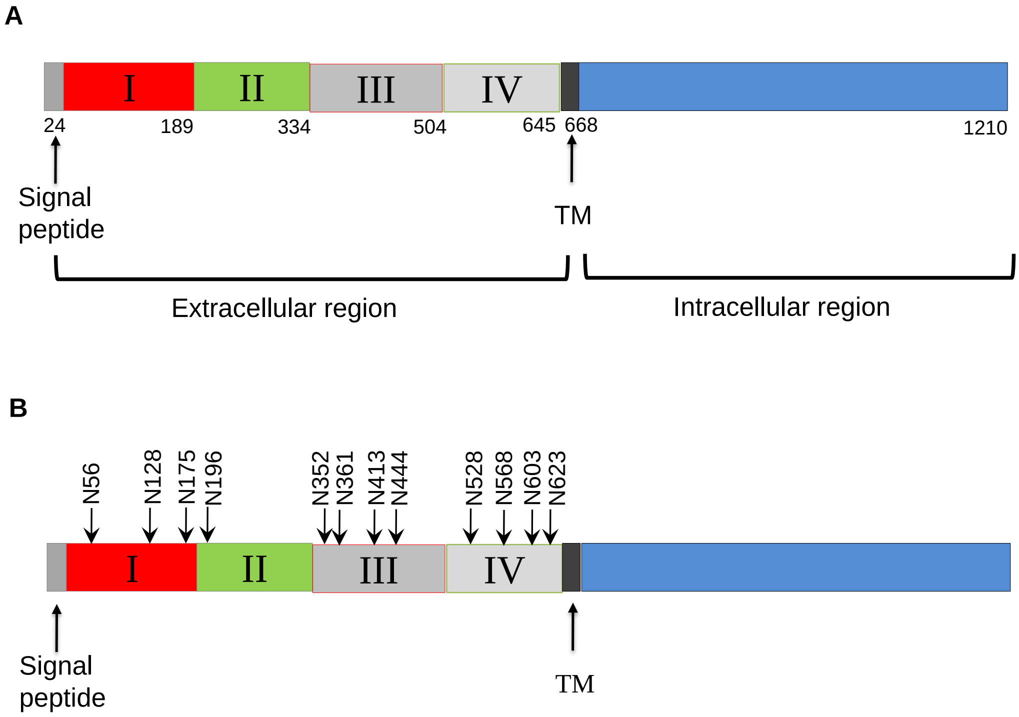

A
I
II
III
IV
24
189
334
504
645
668
1210
Signal peptide
TM
Extracellular region
Intracellular region
B
N56
N128
N175
N196
N352
N361
N413
N444
N528
N568
N603
N623
I
II
III
IV
Signal peptide
TM

Figure 4.1. Structural feature and potential N-glycosylation sites in EGFR. A. Organization of domains with the primary structures of EGFR. I, II, III, and IV denote the four subdomains in the extracellular region. Domain I is in red color, and domain III sharing 37% sequence identity with it is in grey with the red border. Domain II is in green, and the other cysteine rich domain IV is in grey with the green border. The approximate domain-domain boundary amino acid is labeled by numbers along the primary structure. The sequence of signal peptide is included in the primary sequence of the EGFR protein. TM is the transmembrane region. B. Summary of N-glycosylation sites of EGFR extracellular region. Twelve potential sites for N-glycosylation are schematically represented.

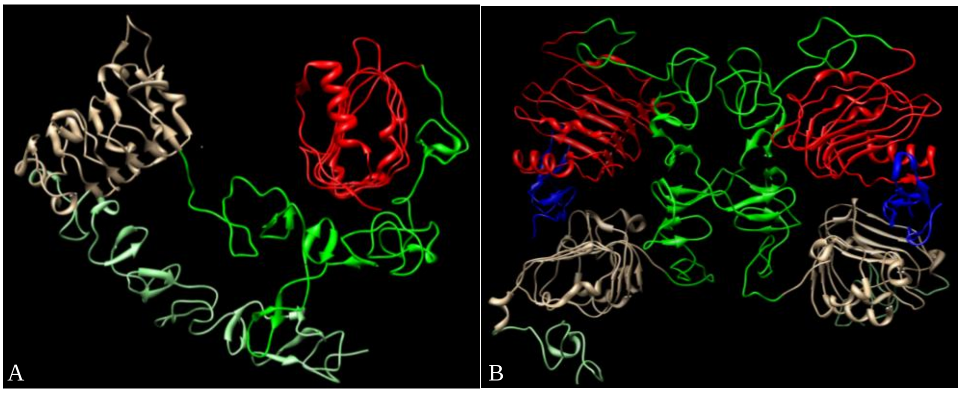


Figure 4.2. 3D Structure of EGFR based on the crystal structure (PDB code: 1nql for tethered monomer, and 1ivo for extended dimer). Domain I is in red color, and domain III sharing 37% sequence identity with it is in tan. Domain II is in green, and the other cysteine rich domain IV is in light green. The EGF in the complex is shown in blue color. A. The structure of the tethered EGFR monomer. B. The structure of the extended (EGF:EGFR)$_2$ dimer.

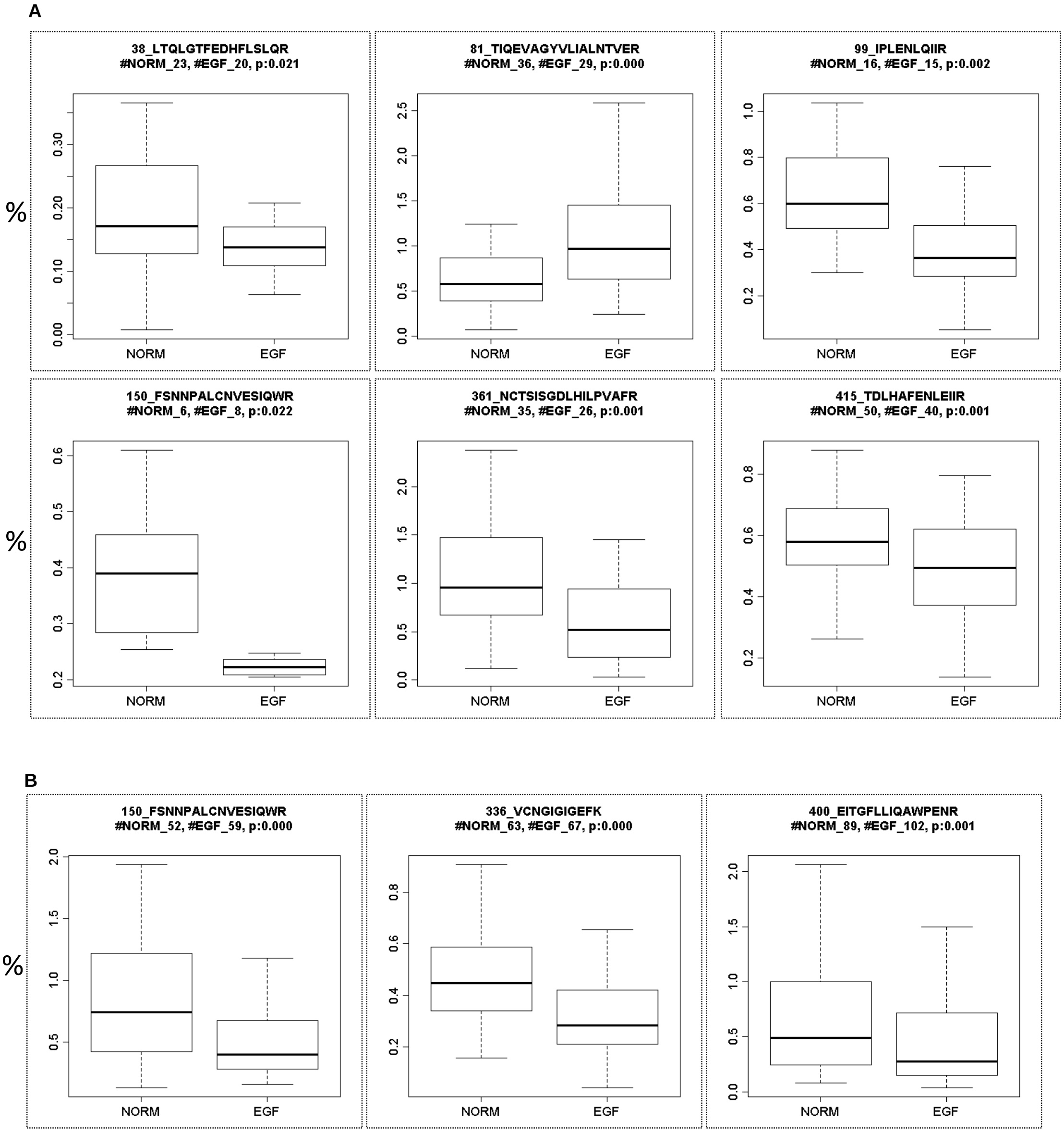


Figure 4.3. Box plot display of the distribution of the relative abundance of the oxidized species of each peptide from EGFR. The title of each box contains the start site of the first amino acid of each peptide, the peptide sequence, and the number of spectra of oxidized species in each group. The Y-axis shows the relative abundance of each oxidized species calculated from the MS2 intensity in percentage (%). **A**. Six peptides of EGFR identified in the 293-F cell data set. **B**. Three peptides of EGFR identified in the A431 cell data set.

From the crystal structure of (EGF:EGFR)$_2$ complex (Figure 4.2B), we can map the EGF binding sites on EGFR. Three EGF binding sites on domains I and III of the receptor have been confirmed by X-ray structural study, and a series of amino acid side chains are involved in these sites (136). Interestingly, in this study, three peptides showing decreased abundance of oxidized species after EGF binding contain the amino acid side chains which are involved in EGF binding, as shown in Figure 4.4. The first is the peptide ‘LTQLGTFEDHFLSLQR’ from domain I, which contains three amino acid residues reported in EGF binding site 1: Leu38, Gln40 and Gly42. The second is the peptide with the sequence ‘NCTSISGDLHILPVAFR’ from domain III, which contains Val374 in the binding site 2 for interaction with EGF. The third one with the sequence ‘EITGFLLIQAWPENR’ is also from the domain III, along which Leu406 and Gln408 are from the binding site 3 and take part in the interaction with EGF.

The crystal structure (Figure 4.4) shows that the side chain of Leu38 in domain I, the side chain of Val374 and Leu406 in domain III of EGFR hydrophobically interacts with the side chain of Met21, Leu15, and Leu47 of EGF; respectively (136). Besides, the Gln40 side chain of EGFR hydrogen bonds with the Asn32 side chain of EGF; and the Gln408 side chain of EGFR hydrogen bonds with the main chain carbonyl and amide groups of Gln43 and Arg45 of EGF, respectively (136). Additionally, residues 40-42 of EGFR and residues 31-33 of EGF form a short parallel β sheet (136). What footprinting MS/MS data revealed is consistent with the X-ray crystal structure, and laser photochemical footprinting technique coupled with MS analysis confirmed that domain I and domain III of EGFR play important roles in EGF binding.

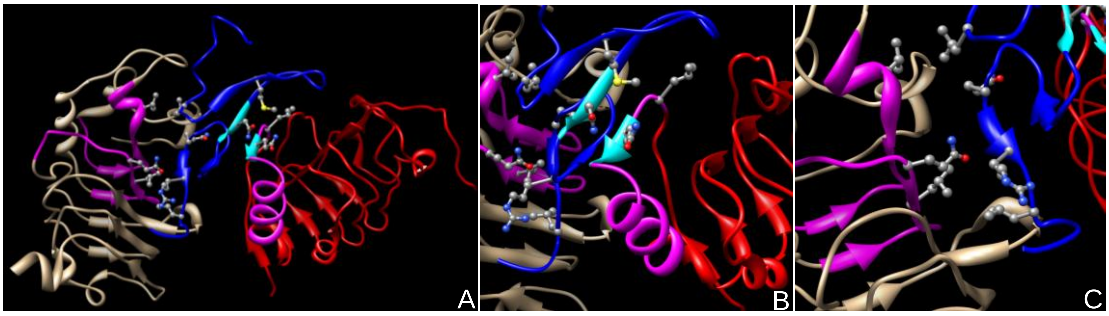


Figure 4.4. 3D schematic representation of EGF interaction with domain I and domain III of EGFR based on the crystal structure (PDB code: 1ivo for extended dimer). Domain I is in red color, and domain III sharing 37% sequence identity with it is in tan. The EGF in between is shown in blue color. The three peptides showing decreased abundance of oxidized species after EGF binding are shown in magenta. Amino acids found to be involved in the interaction were shown as ball and stick atomic models, and they were colored in default by element.The short parallel β sheet formed by the residues 40-42 of EGFR and residues 31-33 of EGF was labelled in cyan color. **A**. The overall display of EGF interaction with domain I and domain III of EGFR. **B**. Interaction of EGF with domain I of EGFR. **C**. Interaction of EGF with domain III of EGFR.

The two EGFR subpopulations show different EGF binding affinity

We know from the EGF binding affinity studies that EGFRs exist on the cell surface in two different affinity classes for more than two decades (130). Our western blot of EGFR expressed by 293-F cells also showed two visible separated bands (Figure 4.5A), one is a very thick band above the 150kD molecular weight marker, and the other is a thinner one below the 150kD marker. This indicates the existence of two EGFR subpopulations with different molecular weight, which is probably caused by different glycosylation levels between them.

A431 cell expresses native EGFR protein. We resolved the EGFR protein bands by running the immunoprecitated protein in a 4-12% gradient gel for around six hours in cold room till the small proteins (with the molecular weight less than

37kD) come out of the gel to obtain better separation for the relatively bigger proteins. When the gel was stained with Coomassie blue, a strong band with the MW above 150kDa (Figure 4.5B) and a cluster of barely observable faint bands below 150kDa were obtained. The gel was destained and re-stained with silver staining method. Three high MW bands became clearly visible after silver staining (Figure 4.5C). The first band with the MW above 150 kD appeared at the very location of the original blue band, and the other two bands very near to each other correspond to the cluster of very faint bands in the Coommassie blue staining picture and their MW is below 150 kD (Figure 4.5C). We cut out the gel bands and classified them into two subpopulations according to their abundance: the major population with higher molecular weight (HMW, the band above 150kD marker) and the minor population with lower molecular weight (LMW, the two bands below 150 kD marker). These protein bands were confirmed as to be EGFR proteins by MS analysis.

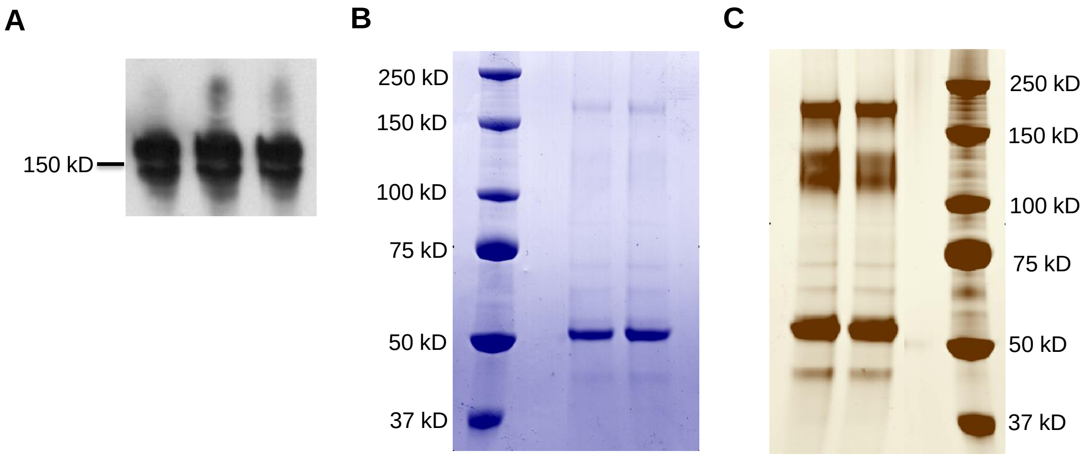


Figure 4.5. Different subpopulations of EGFR as revealed by western blot and immunoprecipitation. A. Western blot of EGFR EGFR expressed by 293-F cells. Two visible separated bands were detected (7.5% SDS page). B. Coomassie blue staining picture of the resolved immunoprecipitated EGFR (4-12% gradient gel). One protein band above 150kD could be clearly seen. **C**. Silver staining picture of the resolved immunoprecipitated EGFR (4-12% gradient gel). One protein band above 150kD could be clearly seen. Another cluster of two bands below 150kD also appeared.

As showed in Table 4.1, three peptides from A431 footprinting LC-MS/MS dataset were detected to have decreased oxidation extent after EGF binding (also as shown in Figure 4.3B). We further investigated the distribution of the relative abundance of the relevant oxidized species of these three peptides from each subpopution of EGFR, the major HMW subpopulation and the minor LMW subpopulation, respectively.

As shown in Figure 4.6, from the distribution of the relative abundance of the oxidized species of each peptide, we can see that peptides from the minor LMW population of EGFR showed greater oxidation extent than those from the HMW population, as indicated by the significantly greater Y-axis scales shown in the figure. It has been generally accepted that the minority high affinity EGFR presents as an extended dimer upon EGF binding (Figure 4.2B) and activates the intracellular signal transduction. The extended EGFR dimer is more liable to OH radicals in solution than the tethered monomer. The higher relative abundance of the oxidized peptide species from the minor LMW population of EGFR suggested that this subpopulation may adopt a more extended conformation than the EGFR from the major HMW population. Furthermore, peptides from the minor LMW population of EGFR showed greater significant decrease in the oxidation extent than the ones from the major HMW population on EGF binding, indicating the minor LMW population of EGFR has higher affinity for EGF. Thus the minor LMW population of EGFR probably represents the minority high affinity population for EGF binding, and the major HMW population corresponds to the majority low affinity population for EGF binding.

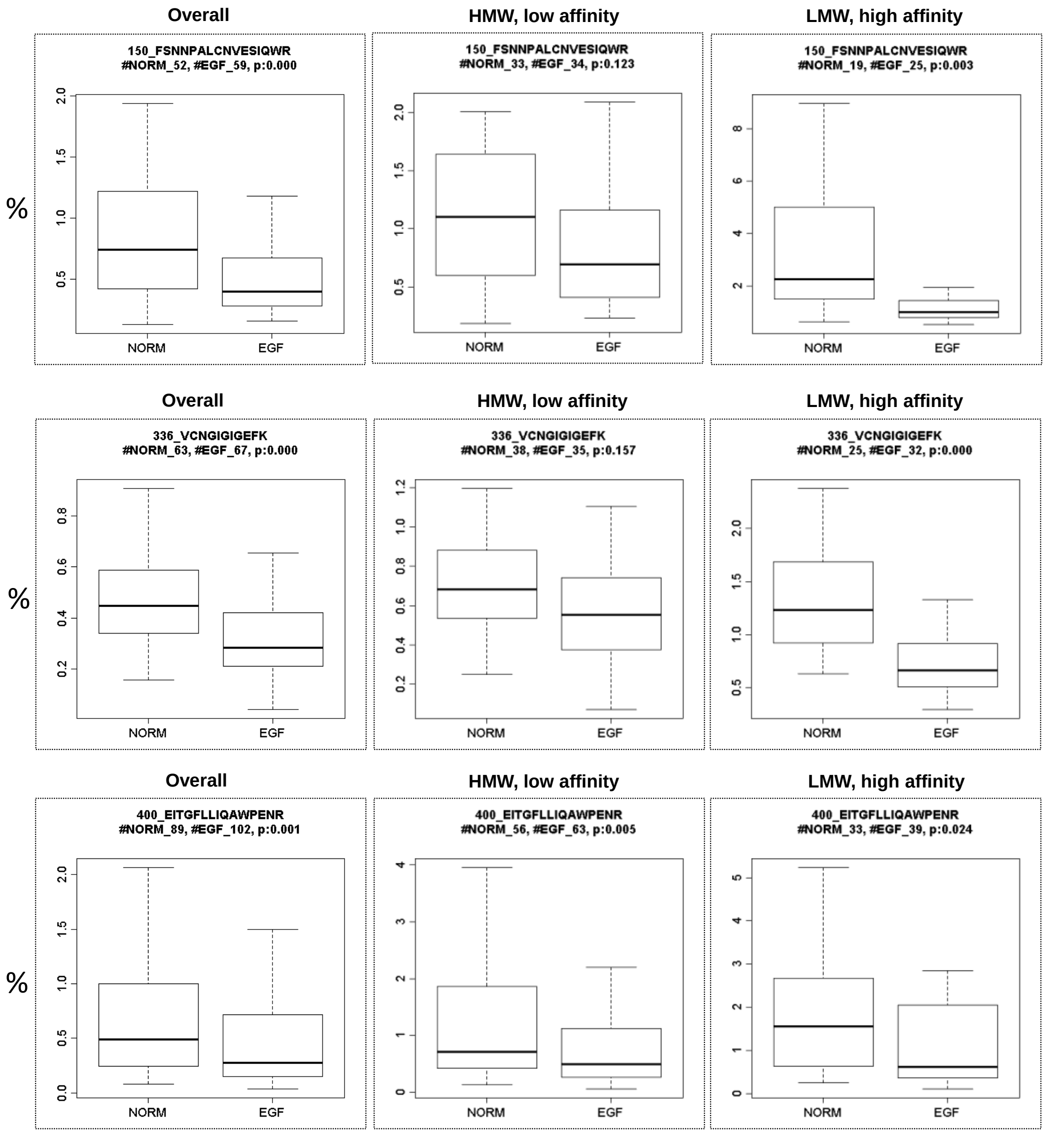


Figure 4.6. Box plot display of the distribution of the relative abundance of the oxidized species of the three peptides identified from different subpopulations of EGFR in the A431 dataset. The box plot charts in every row denotes the same peptide, while from different subpopulations. They are the 'Overall' population, the 'HMW' population, and the 'LMW' population, respectively, from left to right in each row. The title of each box contains the start site of the first amino acid of each peptide, the peptide sequence, and the number of spectra of oxidized species in each group. The Y-axis shows the relative abundance of each oxidized species calculated from the MS2 intensity in percentage (%).

Different glycosylation levels of specific glycosylation sites from the two EGFR subpopulations differentiates the low and high affinity EGFRs

As reviewed in the introduction, EGFR exists as different forms in terms of glycosylation, which may induce different subpopulations of EGFR. And it's generally accepted that the difference of low and high affinity EGFRs on living cells lies in glycosylation (142-143). Previous studies suggested that glycosylation of EGFR is involved in dimerization and endocytosis. It was found that glycosylation at N568 in domain IV may affect proper folding of domain III, because EGFR lacking the N-glycan was not recognized by the conformationally sensitive antibody C225 (150). Another study on analysis of epitope accessibility suggested that the lack of glycosylation at N603 in domain IV weakens auto-inhibitory tether interactions, however, they found that untethering, while necessary, is not sufficient to drive dimerization (151). As to sites in domain I, no functional difference between EGFR lacking the N-glycan at the noncanonical N56 site and that containing the N-glycan attachment was detected (152). As to glycosylation sites in domain III, it has been found that N-444-Q mutant EGFR undergo spontaneous dimerization, resulting in the phosphorylation of the receptor in the absence of ligand, thus N-glycan on N444 of EGFR appears to prevent the spontaneous dimerization and activation of EGFR (142). Interestingly, it has also been found that the loss of N-glycan on N437 of ErbB3 (corresponds to N444 of EGFR) leads to spontaneous dimerization of the receptor (143).

In the A431 dataset, five N-glycosylation sites from the twelve reported glycosylation sites in EGFR extracellular domain have been identified and quantitatively characterized (Figure 4.7). They are N352, N361, N413 and N444 from domain III of EGFR and N603 from domain IV of EGFR (Figure

4.1B). We quantified the relevant N-glycosylation levels of these sites identified from the major HMW population of EGFR (which corresponds to the majority low affinity population of EGFR) and the minor LMW population of EGFR (which corresponds to the minority high affinity population of EGFR), respectively. After that, we also calculated the overall N-glycosylation level of each site from all the populations.

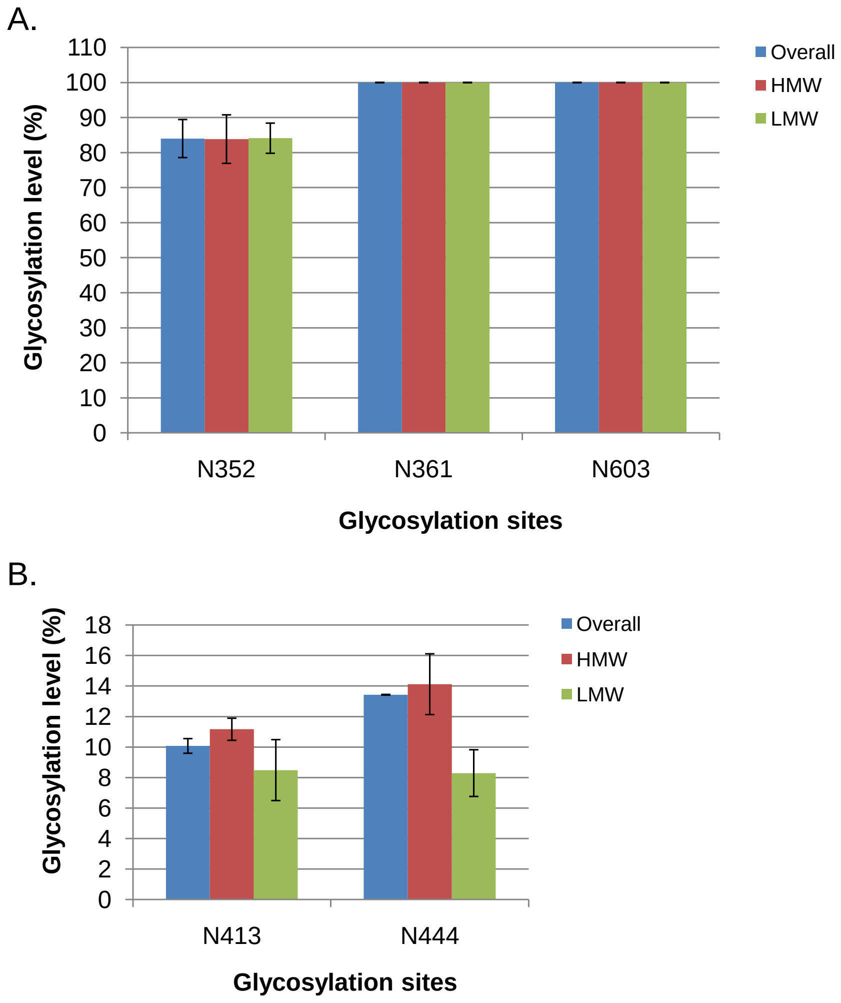


Figure 4.7. **The glycosylation level of the five N-glycosytion sites identified from the A431 data set.** 'Overall' means the whole population, 'HMW' means the majority EGFR subpopulation with higher molecular weight, and 'LWM' represents the minority EGFR subpopulation with lower molecular weight. A. The three sites showing no difference among different subpopulations. B. Two sites showing significantly lower glycosylation level in the minority EGFR population.

As shown in Figure 4.7A, N352, N361 and N603 showed comparable glycosylation levels among the overall population, the low affinity population (HMW), and the high affinity population (LMW) of EGFR. In the meantime, as shown in Figure 4.7B, LC-MS/MS results revealed much lower glycosylation levels of N413 and N444 from the high affinity population (LMW) than from the low affinity population (HMW) of EGFR. N-Glycosylation alters the Asn side chain torsion angle distribution and reduces its flexibility and influences local protein conformation near the glycosylation sites (153). The lower glycosylation level of both N413 and N444 in the minor LMW population of EGFR (which corresponds to the minority high affinity population of EGFR) causes a more flexible conformer for EGF binding and activation of EGFR. This observation is in agreement with the mutagenesis studies of N444 of EGFR and N437 of ErbB3 discussed above, suggesting that the glycosylation of N413 and N444 may differentiate the low and high EGF binding affinity subpopulations of EGFR.

Conclusion

In conclusion, we applied the ultrafast laser photochemical hydroxyl radical footprinting method to study the EGF binding of EGFR directly on live cell surface. Footprinting results revealed that domain I and domain III of EGFR were involved in EGF binding. In addition, two EGFR subpopulations with different glycosylation levels have been characterized. The results further indicated that the two subpopulations of EGFR showed different EGF binding affinity. The minor LMW subpopulation of EGFR showed higher EGF binding affinity, thus it corresponds to the minority high affinity EGF binding population of EGFR. More profoundly, two glycosylation sites from domain III of EGFR, N413 and N444, were quantified to have lower glycosylation level in the minority population of EGFR, which allows to form a more flexible conformer for EGF binding and activation of EGFR. These observations not only consist with what have been discovered in other studies of ErbB family members, but also help to reveal the difference of low and high EGF binding affinity populations of EGFR. Thus this technique is proved to be helpful in both the structural and funtional studies of plasma membrane proteins of live cells.

Chapter 5. Conclusions

Recent progress in mass spectrometry has enabled a novel MS-based protein oxidative footprinting technique to determine structural information by mapping of oxidation induced by OH radicals. This method is an adaptation of the OH radical footprinting in combination with mass spectrometry detection for the mapping of protein surface (15-16, 18, 44-45). It has been shown to be a powerful technique in understanding ligand-induced conformational changes when coupled with existing structural data that are either experimentally or computationally generated (11).

Similar to other analytical techniques, the application of MS to membrane protein continues to be much less common than studies on their water-soluble counterparts (43), and researchers commonly resort to surfactant-solubilized species. And also because *in vivo* studies are often not feasible, current surface mapping techniques are therefore mostly limited to purified proteins which are stable in solution, and are not amenable to the conformation determination of most IMPs or their complexes. As a result of these difficulties, the study of structural dynamics of IMP remains elusive.

This thesis reports the development and application of a novel adaptation of oxidative footprinting coupled with MS technique for investigating the *in vivo* structural dynamics of IMPs. In this method, generally, cells are first exposed to a burst of *in situ* generated OH radicals, and the membrane proteins on the cell surfaces are oxidized. The targeted protein is purified, followed by proteolysis to generate peptides for analysis and identification by LC-MS/MS and bioinformatics analysis.

In chapter 2, we used *E. coli* as a model to study the structural dynamics of the outer membrane proteins, with an emphasis on the matrix porin OmpF (104). Briefly, *E coli* cells are first bathed in a stimulation buffer to induce the porin channel closure after harvest. Cells are then exposed to OH radicals generated from an *in situ* Fenton reaction between H_2O_2 and Fe-bound ethylenediaminetetraacetic acid (EDTA) for oxidation. Footprinting results show that the surface mapping data of OmpF are consistent with its current crystallographic structure. In addition, this technique also enables the detection of *in vivo* voltage gating of porin OmpF for the first time.

As discussed in the introduction part, OH radicals can be generated by various kinds of methods including Fenton chemistry, laser photolysis of hydrogen peroxide, and radiolysis of water with a high-energy X-ray synchrotron beam, etc. In the first study, because the 266 nm laser beam cannot penetrate the yellowish and cloudy *E. coli* bacterial culture, the laser photolysis method does not work in this system. In addition, the OmpF voltage gating is specially related to small cations and has nothing to do with small molecules like H_2O, H_2O_2, and OH radicals, etc. Thereby, we adopt a Fenton reaction to oxidize the living bacterial cells. We think that Fenton chemistry is suitable to study the outer membrane porins, because they are extremely stable structures that are restricted by both the intra-molecular hydrogen bond and the hydrophobic interaction with the lipid bilayer wall and are unlikely to have a great conformational change of the protein (48, 54). However, Fenton chemistry is not guaranteed for many other proteins, especially dynamic ones, because this method needs a relatively longer incubation time of several minutes for the reagents to generate OH radicals (154), which may be too long for many macromolecules as conformational changes take place on the order of milliseconds to seconds (11). For a conformational study of those highly dynamic proteins, the laser photolysis of H_2O_2 or the radiolysis of water with a high-energy X-ray synchrotron beam may be a better method because of the ultrafast generation of OH radicals in nanosecond to microseconds (12, 155). We have already developed an ultra-fast laser photolysis technique to rapidly and accurately detect solvent-accessible regions of proteins in their native states (15). Our laser induced protein surface footprinting technique enables investigation of the receptor-ligand interaction by determining the footprints left behind earlier on the plasma membrane surface of a living cell (*in vivo*). The photochemically generated hydroxide radicals in the culture medium will oxidize any molecule exposed to the culture medium on µs timescale. Extracellular domains of membrane proteins which are in contact with the medium will be oxidized by the OH radicals. The solvent accessible surface of the extracellular domain at the moment of ligand-binding can be frozen in the snapshot of that laser flash. By comparing the solvent accessible surfaces of the protein and protein-ligand complexes on the cellular plasma membrane surface, the conformational changes as well as the ligand and drug recognition sites on the protein can be elucidated *in vivo*.

In chapter 3, we further implemented the application of the nanosecond pulse laser-induced photolytic hydroxide radical footprinting method (15) to study the conformational change of native αLβ2 integrin on plasma membrane of the surface of a mammalian cell. In this method, a high concentration of OH radicals is generated in 5~7 ns via photo-dissociation of a H_2O_2 solution by a pulse Nd:YAG laser at 266 nm. The OH radicals immediately oxidize amino acid residues located on the protein surface to produce stable covalent modifications. The half-life of OH radicals is in nanosecond timescale, despite half-life of some secondary radical generated by reacting with OH may be longer. They should be quenched in micro-second time scale. Because of the nanosecond time scale of the formation of OH radicals as well as their subsequent oxidation and quenching, this method allows snap frozen the solvent-accessible surface or footprint of a protein in its native biological state (15). The footprinting MS data demonstrated the activation and overall extension of αLβ2 heterodimer upon Mg/EGTA activation. The findings are consistent well with what have been discovered in other studies.

We reported in chapter 4 the further application of the nanosecond laser photochemical footprinting method (15) to study the receptor-ligand interaction *in vivo*, using EGFR as a model protein because EGF binding of EGFR has been well and extensively characterized in numerous previous studies. In the meantime, we separated two different glycosylated EGFR protein subpopulations by SDS-PAGE, characterized the glycosylation sites and investigated the relevant glycosylaton levels of each population, to help to understand the relationship of glycosylation related EGFR function.

Footprinting results demonstrated in this study that domain I and domain III of EGFR were involved in EGF binding. In addition, two EGFR subpopulations with different glycosylation levels have been characterized, and it was further found that the two EGFR subpopulations showed different EGF binding affinity. Furthermore, two glycosylation sites from domain III of EGFR, N413 and N444, were quantified to have lower glycosylation level in the minority lower molecular weight population of EGFR, which helps to differentiate the two EGFR subpopulations. These observations not only confirmed what have been discovered in other previous studies of ErbB family members, but also helped to reveal the relationship of glycosylation and EGFR function, and further understand the nature of low and high affinity EGFRs.

In conclusion, the hydroxyl radical footprinting technique coupled with MS analysis can be a promising approach and has the potential application for investigating the structure-functional relationship of integral membrane proteins directly on plasma membrane surface of the living cells.

References

1. Speers, A. E., and Wu, C. C. (2007) Proteomics of integral membrane proteins--theory and application. *Chem Rev* 107, 3687-3714.
2. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R., 3rd (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 21, 532-538.
3. George, S. R., O'Dowd, B. F., and Lee, S. P. (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat Rev Drug Discov* 1, 808-820.
4. Punta, M., Forrest, L. R., Bigelow, H., Kernytsky, A., Liu, J., and Rost, B. (2007) Membrane protein prediction methods. *Methods* 41, 460-474.
5. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem Rev* 107, 3514-3543.
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. Zubarev, R. A., Zubarev, A. R., and Savitski, M. M. (2008) Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? *J Am Soc Mass Spectrom* 19, 753-761.
8. 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.
9. Zhou, Y., Dong, J., and Vachet, R. W. (2011) Electron Transfer Dissociation of Modified Peptides and Proteins. *Curr Pharm Biotechnol*.
10. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* 25, 158-170.
11. Guan, J. Q., and Chance, M. R. (2005) Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem Sci* 30, 583-592.

12. 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.

13. Tullius, T. D., and Dombroski, B. A. (1985) Iron(II) EDTA used to measure the helical twist along any DNA molecule. *Science* 230, 679-681.

14. Pogozelski, W. K., McNeese, T. J., and Tullius, T. D. (1995) What species is responsible for strand scission in the reaction of $[Fe^{II}EDTA]^{2-}$ and H_2O_2 with DNA? *Journal of the American Chemical Society* 117, 6428-6433.

15. Aye, T. T., Low, T. Y., and Sze, S. K. (2005) Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry. *Anal Chem* 77, 5814-5822.

16. Sharp, J. S., Becker, J. M., and Hettich, R. L. (2003) Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry. *Anal Biochem* 313, 216-225.

17. Xu, G., and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Anal Chem* 77, 4549-4555.

18. Hambly, D. M., and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J Am Soc Mass Spectrom* 16, 2057-2063.

19. Schuessler, H., and Schilling, K. (1984) Oxygen effect in the radiolysis of proteins. Part 2. Bovine serum albumin. *Int J Radiat Biol Relat Stud Phys Chem Med* 45, 267-281.

20. 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.

21. McClintock, C., Kertesz, V., and Hettich, R. L. (2008) Development of an electrochemical oxidation method for probing higher order protein structure with mass spectrometry. *Anal Chem* 80, 3304-3317.

22. Halliwell, B., and Gutteridge, J. M. (1985) The importance of free radicals and catalytic metal ions in human diseases. *Mol Aspects Med* 8, 89-193.

23. Tullius, T. D., and Greenbaum, J. A. (2005) Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr Opin Chem Biol* 9, 127-134.

24. Tullius, T. D., and Dombroski, B. A. (1986) Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc Natl Acad Sci U S A* 83, 5469-5473.

25. Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A., and Tullius, T. D. (1991) Hydroxyl radical footprinting. *Methods Enzymol* 208, 380-413.

26. Shcherbakova, I., Mitra, S., Beer, R. H., and Brenowitz, M. (2006) Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins. *Nucleic Acids Res* 34, e48.

27. Dai, S. Y., Gardner, M. W., and Fitzgerald, M. C. (2005) Protocol for the thermodynamic analysis of some proteins using an H/D exchange- and mass spectrometry-based technique. *Anal Chem* 77, 693-697.

28. Kiselar, J. G., Maleknia, S. D., Sullivan, M., Downard, K. M., and Chance, M. R. (2002) Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry. *Int J Radiat Biol* 78, 101-114.

29. Fessenden, R. W., and Verma, N. C. (1978) A time-resolved electron spin resonance study of the oxidation of ascorbic acid by hydroxyl radical. *Biophys J* 24, 93-101.

30. LaVerne, J. A. (2000) OH radicals and oxidizing products in the gamma radiolysis of water. *Radiat Res* 153, 196-200.

31. Hawkins, C. L., and Davies, M. J. (2001) Generation and propagation of radical reactions on proteins. *Biochim Biophys Acta* 1504, 196-219.

32. Buxton, G. V., Greenstock, C. L., Helman, W. P., and Ross, A. B. (1988) Critical-Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen-Atoms and Hydroxyl Radicals (.Oh/.O-) in Aqueous-Solution. *J Phys Chem Ref Data* 17, 513-886.

33. Xu, G. Z., and Chance, M. R. (2005) Radiolytic modification of sulfur-containing amino acid residues in model peptides: Fundamental studies for protein footprinting. *Analytical Chemistry* 77, 2437-2449.

34. Eng, J. K., Mccormack, A. L., and Yates, J. R. (1994) An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectr* 5, 976-989.

35. 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.

36. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466-1467.

37. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3, 958-964.

38. Tanner, S., Shu, H. J., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., and Bafna, V. (2005) InsPecT: Identification of posttransiationally modified peptides from tandem mass spectra. *Analytical Chemistry* 77, 4626-4639.

39. Bern, M., Cai, Y. H., and Goldberg, D. (2007) Lookup peaks: A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Analytical Chemistry* 79, 1393-1400.

40. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated Algorithms for High-Throughput Examination of Covalently Labeled Biomolecules by Structural Mass Spectrometry. *Analytical Chemistry* 81, 8141-8149.

41. Trudgian, D. C., Ridlova, G., Fischer, R., Mackeen, M. M., Ternette, N., Acuto, O., Kessler, B. M., and Thomas, B. (2011) Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics* 11, 2790-2797.

42. Griffin, N. M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J. A., and Schnitzer, J. E. (2010) Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* 28, 83-89.

43. 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.

44. Charvatova, O., Foley, B. L., Bern, M. W., Sharp, J. S., Orlando, R., and Woods, R. J. (2008) Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: application to galectin-1. *J Am Soc Mass Spectrom* 19, 1692-1705.

45. Bohon, J., Jennings, L. D., Phillips, C. M., Licht, S., and Chance, M. R. (2008) Synchrotron protein footprinting supports substrate translocation by ClpA via ATP-induced movements of the D2 loop. *Structure* 16, 1157-1165.

46. Ruiz, N., Kahne, D., and Silhavy, T. J. (2006) Advances in understanding bacterial outer-membrane biogenesis. *Nat Rev Microbiol* 4, 57-66.

47. Schulz, G. E. (1993) Bacterial porins: structure and function. *Curr Opin Cell Biol* 5, 701-707.

48. Galdiero, S., Galdiero, M., and Pedone, C. (2007) beta-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. *Curr Protein Pept Sci* 8, 63-82.

49. Schirmer, T. (1998) General and specific porins from bacterial outer membranes. *J Struct Biol* 121, 101-109.

50. Schulz, G. E. (2000) beta-Barrel membrane proteins. *Curr Opin Struct Biol* 10, 443-447.

51. Wimley, W. C. (2003) The versatile beta-barrel membrane protein. *Curr Opin Struct Biol* 13, 404-411.

52. MacKinnon, R. (1995) Pore loops: an emerging theme in ion channel structure. *Neuron* 14, 889-892.

53. Schulz, G. E. (1996) Porins: general to specific, native to engineered passive pores. *Curr Opin Struct Biol* 6, 485-490.

54. Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N., and Rosenbusch, J. P. (1992) Crystal structures explain functional properties of two E. coli porins. *Nature* 358, 727-733.

55. Dutzler, R., Schirmer, T., Karplus, M., and Fischer, S. (2002) Translocation mechanism of long sugar chains across the maltoporin membrane channel. *Structure* 10, 1273-1284.

56. Schirmer, T., Keller, T. A., Wang, Y. F., and Rosenbusch, J. P. (1995) Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* 267, 512-514.

57. Muller, D. J., and Engel, A. (1999) Voltage and pH-induced channel closure of porin OmpF visualized by atomic force microscopy. *J Mol Biol* 285, 1347-1351.

58. Samartzidou, H., and Delcour, A. H. (1998) E.coli PhoE porin has an opposite voltage-dependence to the homologous OmpF. *EMBO J* 17, 93-100.

59. Bainbridge, G., Gokce, I., and Lakey, J. H. (1998) Voltage gating is a fundamental feature of porin and toxin beta-barrel membrane channels. *FEBS Lett* 431, 305-308.

60. Brunen, M., and Engelhardt, H. (1993) Asymmetry of orientation and voltage gating of the Acidovorax delafieldii porin Omp34 in lipid bilayers. *Eur J Biochem* 212, 129-135.

61. Lakey, J. H., and Pattus, F. (1989) The voltage-dependent activity of Escherichia coli porins in different planar bilayer reconstitutions. *Eur J Biochem* 186, 303-308.

62. Fujiki, Y., Hubbard, A. L., Fowler, S., and Lazarow, P. B. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J Cell Biol* 93, 97-102.

63. Sze, S. K., de Kleijn, D. P., Lai, R. C., Khia Way Tan, E., Zhao, H., Yeo, K. S., Low, T. Y., Lian, Q., Lee, C. N., Mitchell, W., El Oakley, R. M., and Lim, S. K. (2007) Elucidating the secretion proteome of human embryonic stem cell-derived mesenchymal stem cells. *Mol Cell Proteomics* 6, 1680-1689.

64. Bern, M., Cai, Y., and Goldberg, D. (2007) Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem* 79, 1393-1400.

65. Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem* 23, 600-609.

66. Hong, H., Szabo, G., and Tamm, L. K. (2006) Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nat Chem Biol* 2, 627-635.

67. Koronakis, V. (2003) TolC--the bacterial exit duct for proteins and drugs. *FEBS Lett* 555, 66-71.

68. Cowan, S. W., Garavito, R. M., Jansonius, J. N., Jenkins, J. A., Karlsson, R., Konig, N., Pai, E. F., Pauptit, R. A., Rizkallah, P. J., Rosenbusch, J. P., and et al. (1995) The structure of OmpF porin in a tetragonal crystal form. *Structure* 3, 1041-1050.

69. Sen, K., Hellman, J., and Nikaido, H. (1988) Porin channels in intact cells of Escherichia coli are not affected by Donnan potentials across the outer membrane. *J Biol Chem* 263, 1182-1187.

70. delaVega, A. L., and Delcour, A. H. (1995) Cadaverine induces closing of E. coli porins. *EMBO J* 14, 6058-6065.

71. Schindler, H., and Rosenbusch, J. P. (1981) Matrix protein in planar membranes: clusters of channels in a native environment and their functional reassembly. *Proc Natl Acad Sci U S A* 78, 2302-2306.

72. Delcour, A. H., Adler, J., Kung, C., and Martinac, B. (1992) Membrane-derived oligosaccharides (MDO's) promote closing of an E. coli porin channel. *FEBS Lett* 304, 216-220.

73. Le Dain, A. C., Hase, C. C., Tommassen, J., and Martinac, B. (1996) Porins of Escherichia coli: unidirectional gating by pressure. *EMBO J* 15, 3524-3528.

74. Robertson, K. M., and Tieleman, D. P. (2002) Molecular basis of voltage gating of OmpF porin. *Biochem Cell Biol* 80, 517-523.

75. Eppens, E. F., Saint, N., Van Gelder, P., van Boxtel, R., and Tommassen, J. (1997) Role of the constriction loop in the gating of outer membrane porin PhoE of Escherichia coli. *FEBS Lett* 415, 317-320.

76. Mobasheri, H., and Lea, E. J. (2002) Biophysics of gating phenomena in voltage-dependent OmpC mutant porin channels (R74C and R37C) of Escherichia coli outer membranes. *Eur Biophys J* 31, 389-399.

77. Phale, P. S., Philippsen, A., Widmer, C., Phale, V. P., Rosenbusch, J. P., and Schirmer, T. (2001) Role of charged residues at the OmpF porin channel constriction probed by mutagenesis and simulation. *Biochemistry* 40, 6319-6325.

78. Miedema, H., Vrouenraets, M., Wierenga, J., Eisenberg, B., Schirmer, T., Basle, A., and Meijberg, W. (2006) Conductance and selectivity fluctuations in D127 mutants of the bacterial porin OmpF. *Eur Biophys J* 36, 13-22.

79. Bainbridge, G., Mobasheri, H., Armstrong, G. A., Lea, E. J., and Lakey, J. H. (1998) Voltage-gating of Escherichia coli porin: a cystine-scanning mutagenesis study of loop 3. *J Mol Biol* 275, 171-176.

80. Phale, P. S., Schirmer, T., Prilipov, A., Lou, K. L., Hardmeyer, A., and Rosenbusch, J. P. (1997) Voltage gating of Escherichia coli porin channels: role of the constriction loop. *Proc Natl Acad Sci U S A* 94, 6741-6745.

81. Saint, N., Lou, K. L., Widmer, C., Luckey, M., Schirmer, T., and Rosenbusch, J. P. (1996) Structural and functional characterization of OmpF porin mutants selected for larger pore size. II. Functional characterization. *J Biol Chem* 271, 20676-20680.

82. Im, W., and Roux, B. (2002) Ions and counterions in a biological channel: a molecular dynamics simulation of OmpF porin from Escherichia coli in an explicit membrane with 1 M KCl aqueous salt solution. *J Mol Biol* 319, 1177-1197.

83. Schirmer, T., and Phale, P. S. (1999) Brownian dynamics simulation of ion flow through porin channels. *J Mol Biol* 294, 1159-1167.

84. Alcaraz, A., Nestorovich, E. M., Aguilella-Arzo, M., Aguilella, V. M., and Bezrukov, S. M. (2004) Salting out the ionic selectivity of a wide channel: the asymmetry of OmpF. *Biophys J* 87, 943-957.

85. Bredin, J., Saint, N., Mallea, M., De, E., Molle, G., Pages, J. M., and Simonet, V. (2002) Alteration of pore properties of Escherichia coli OmpF induced by mutation of key residues in anti-loop 3 region. *Biochem J* 363, 521-528.

86. Van Gelder, P., Saint, N., Phale, P., Eppens, E. F., Prilipov, A., van Boxtel, R., Rosenbusch, J. P., and Tommassen, J. (1997) Voltage sensing in the PhoE and OmpF outer membrane porins of Escherichia coli: role of charged residues. *J Mol Biol* 269, 468-472.

87. Luo, B. H., Carman, C. V., and Springer, T. A. (2007) Structural basis of integrin regulation and signaling. *Annu Rev Immunol* 25, 619-647.

88. Xiong, J. P., Goodman, S. L., and Arnaout, M. A. (2007) Purification, analysis, and crystal structure of integrins. *Methods Enzymol* 426, 307-336.

89. Shimaoka, M., and Springer, T. A. (2003) Therapeutic antagonists and conformational regulation of integrin function. *Nat Rev Drug Discov* 2, 703-716.

90. Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L., and Arnaout, M. A. (2002) Crystal structure of the

extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 296, 151-155.

91. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L., and Arnaout, M. A. (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294, 339-345.

92. Springer, T. A., Zhu, J., and Xiao, T. (2008) Structural basis for distinctive recognition of fibrinogen gammaC peptide by the platelet integrin alphaIIbbeta3. *J Cell Biol* 182, 791-800.

93. Xiao, T., Takagi, J., Coller, B. S., Wang, J. H., and Springer, T. A. (2004) Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* 432, 59-67.

94. Xie, C., Zhu, J., Chen, X., Mi, L., Nishida, N., and Springer, T. A. Structure of an integrin with an alphaI domain, complement receptor type 4. *EMBO J* 29, 666-679.

95. Takagi, J., Petre, B. M., Walz, T., and Springer, T. A. (2002) Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* 110, 599-511.

96. Takagi, J., Strokovich, K., Springer, T. A., and Walz, T. (2003) Structure of integrin alpha5beta1 in complex with fibronectin. *EMBO J* 22, 4607-4615.

97. Nishida, N., Xie, C., Shimaoka, M., Cheng, Y., Walz, T., and Springer, T. A. (2006) Activation of leukocyte beta2 integrins by conversion from bent to extended conformations. *Immunity* 25, 583-594.

98. Salas, A., Shimaoka, M., Kogan, A. N., Harwood, C., von Andrian, U. H., and Springer, T. A. (2004) Rolling adhesion through an extended conformation of integrin alphaLbeta2 and relation to alpha I and beta I-like domain interaction. *Immunity* 20, 393-406.

99. Tang, R. H., Tng, E., Law, S. K., and Tan, S. M. (2005) Epitope mapping of monoclonal antibody to integrin alphaL beta2 hybrid domain suggests different requirements of affinity states for intercellular adhesion molecules (ICAM)-1 and ICAM-3 binding. *J Biol Chem* 280, 29208-29216.

100. Shimaoka, M., Xiao, T., Liu, J. H., Yang, Y., Dong, Y., Jun, C. D., McCormack, A., Zhang, R., Joachimiak, A., Takagi, J., Wang, J. H., and

Springer, T. A. (2003) Structures of the alpha L I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* 112, 99-111.

101. Humphries, M. J., Symonds, E. J., and Mould, A. P. (2003) Mapping functional residues onto integrin crystal structures. *Curr Opin Struct Biol* 13, 236-243.

102. 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.

103. Kamal, J. K., Benchaar, S. A., Takamoto, K., Reisler, E., and Chance, M. R. (2007) Three-dimensional structure of cofilin bound to monomeric actin derived by structural mass spectrometry data. *Proc Natl Acad Sci U S A* 104, 7910-7915.

104. 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.

105. Cheng, M., Foo, S. Y., Shi, M. L., Tang, R. H., Kong, L. S., Law, S. K., and Tan, S. M. (2007) Mutation of a conserved asparagine in the I-like domain promotes constitutively active integrins alphaLbeta2 and alphaIIbbeta3. *J Biol Chem* 282, 18225-18232.

106. Guo, T., Lee, S. S., Ng, W. H., Zhu, Y., Gan, C. S., Zhu, J., Wang, H., Huang, S., Sze, S. K., and Kon, O. L. (2010) Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome. *Cell Mol Life Sci*.

107. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Anal Chem* 81, 8141-8149.

108. MacCoss, M. J., Wu, C. C., and Yates, J. R., 3rd (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* 74, 5593-5599.

109. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4, 1-13.

110. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.

111. Lu, C., Shimaoka, M., Zang, Q., Takagi, J., and Springer, T. A. (2001) Locking in alternate conformations of the integrin alphaLbeta2 I domain with disulfide bonds reveals functional relationships among integrin domains. *Proc Natl Acad Sci U S A* 98, 2393-2398.

112. Yang, W., Shimaoka, M., Salas, A., Takagi, J., and Springer, T. A. (2004) Intersubunit signal transmission in integrins by a receptor-like interaction with a pull spring. *Proc Natl Acad Sci U S A* 101, 2906-2911.

113. Xie, C., Shimaoka, M., Xiao, T., Schwab, P., Klickstein, L. B., and Springer, T. A. (2004) The integrin alpha-subunit leg extends at a Ca2+-dependent epitope in the thigh/genu interface upon activation. *Proc Natl Acad Sci U S A* 101, 15422-15427.

114. Shi, M., Sundramurthy, K., Liu, B., Tan, S. M., Law, S. K., and Lescar, J. (2005) The crystal structure of the plexin-semaphorin-integrin domain/hybrid domain/I-EGF1 segment from the human integrin beta2 subunit at 1.8-A resolution. *J Biol Chem* 280, 30586-30593.

115. Shi, M., Foo, S. Y., Tan, S. M., Mitchell, E. P., Law, S. K., and Lescar, J. (2007) A structural hypothesis for the transition between bent and extended conformations of the leukocyte beta2 integrins. *J Biol Chem* 282, 30198-30206.

116. Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103, 211-225.

117. Schlessinger, J. (1988) Signal Transduction by Allosteric Receptor Oligomerization. *Trends Biochem Sci* 13, 443-447.

118. Ullrich, A., and Schlessinger, J. (1990) Signal Transduction by Receptors with Tyrosine Kinase-Activity. *Cell* 61, 203-212.

119. Sako, Y., Minoghchi, S., and Yanagida, T. (2000) Single-molecule imaging of EGFR signaling on the surface of living cells. *Nat Cell Biol* 2, 168-172.

120. Jorissen, R. N., Walker, F., Pouliot, N., Garrett, T. P., Ward, C. W., and Burgess, A. W. (2003) Epidermal growth factor receptor: mechanisms of activation and signaling. *Exp Cell Res* 284, 31-53.

121. Schlessinger, J. (2002) Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* 110, 669-672.

122. Chung, I., Akita, R., Vandlen, R., Toomre, D., Schlessinger, J., and Mellman, I. (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature* 464, 783-787.

123. Lidke, D. S., Nagy, P., Heintzmann, R., Arndt-Jovin, D. J., Post, J. N., Grecco, H. E., Jares-Erijman, E. A., and Jovin, T. M. (2004) Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction. *Nat Biotechnol* 22, 198-203.

124. Ozcan, F., Klein, P., Lemmon, M. A., Lax, I., and Schlessinger, J. (2006) On the nature of low- and high-affinity EGF receptors on living cells. *Proc Natl Acad Sci U S A* 103, 5735-5740.

125. Teramura, Y., Ichinose, J., Takagi, H., Nishida, K., Yanagida, T., and Sako, Y. (2006) Single-molecule analysis of epidermal growth factor binding on the surface of living cells. *EMBO J* 25, 4215-4222.

126. Uyemura, T., Takagi, H., Yanagida, T., and Sako, Y. (2005) Single-molecule analysis of epidermal growth factor signaling that leads to ultrasensitive calcium response. *Biophys J* 88, 3720-3730.

127. Hurwitz, D. R., Emanuel, S. L., Nathan, M. H., Sarver, N., Ullrich, A., Felder, S., Lax, I., and Schlessinger, J. (1991) Egf Induces Increased Ligand-Binding Affinity and Dimerization of Soluble Epidermal Growth-Factor (Egf) Receptor Extracellular Domain. *Journal of Biological Chemistry* 266, 22035-22043.

128. Lemmon, M. A., Bu, Z. M., Ladbury, J. E., Zhou, M., Pinchasi, D., Lax, I., Engelman, D. M., and Schlessinger, J. (1997) Two EGF molecules contribute additively to stabilization of the EGFR dimer. *Embo Journal* 16, 281-294.

129. Sorokin, A., Lemmon, M. A., Ullrich, A., and Schlessinger, J. (1994) Stabilization of an active dimeric form of the epidermal growth factor receptor by introduction of an inter-receptor disulfide bond. *J Biol Chem* 269, 9752-9759.

130. Lax, I., Bellot, F., Howk, R., Ullrich, A., Givol, D., and Schlessinger, J. (1989) Functional analysis of the ligand binding site of EGF-receptor utilizing chimeric chicken/human receptor molecules. *EMBO J* 8, 421-427.

131. Wiley, H. S., Walsh, B. J., and Lund, K. A. (1989) Global Modulation of the Epidermal Growth-Factor Receptor Is Triggered by Occupancy of Only a Few Receptors - Evidence for a Binary Regulatory System in Normal Human-Fibroblasts. *Journal of Biological Chemistry* 264, 18912-18920.

132. Gadella, T. W. J., and Jovin, T. M. (1995) Oligomerization of Epidermal Growth-Factor Receptors on A431 Cells Studied by Time-Resolved Fluorescence Imaging Microscopy - a Stereochemical Model for Tyrosine Kinase Receptor Activation. *J Cell Biol* 129, 1543-1558.

133. Defize, L. H. K., Boonstra, J., Meisenhelder, J., Kruijer, W., Tertoolen, L. G. J., Tilly, B. C., Hunter, T., Henegouwen, P. M. P. V., Moolenaar, W. H., and Delaat, S. W. (1989) Signal Transduction by Epidermal Growth-Factor Occurs through the Subclass of High-Affinity Receptors. *J Cell Biol* 109, 2495-2507.

134. Bellot, F., Moolenaar, W., Kris, R., Mirakhur, B., Verlaan, I., Ullrich, A., Schlessinger, J., and Felder, S. (1990) High-affinity epidermal growth factor binding is specifically reduced by a monoclonal antibody, and appears necessary for early responses. *J Cell Biol* 110, 491-502.

135. Yarden, Y., and Schlessinger, J. (1987) Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. *Biochemistry-Us* 26, 1443-1451.

136. Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J. H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110, 775-787.

137. Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H. S., Leahy, D. J., and Lemmon, M. A. (2003) EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Molecular Cell* 11, 507-517.

138. Li, S. Q., Schmitz, K. R., Jeffrey, P. D., Wiltzius, J. J. W., Kussie, P., and Ferguson, K. M. (2005) Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* 7, 301-311.

139. Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., and et al. (1984) Human epidermal growth factor receptor cDNA sequence and aberrant

expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 309, 418-425.

140. Bajaj, M., Waterfield, M. D., Schlessinger, J., Taylor, W. R., and Blundell, T. (1987) On the tertiary structure of the extracellular domains of the epidermal growth factor and insulin receptors. *Biochim Biophys Acta* 916, 220-226.

141. Ward, C. W., Hoyne, P. A., and Flegg, R. H. (1995) Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins* 22, 141-153.

142. Takahashi, M., Tsuda, T., Ikeda, Y., Honke, K., and Taniguchi, N. (2004) Role of N-glycans in growth factor signaling. *Glycoconj J* 20, 207-212.

143. Takahashi, M., Yokoe, S., Asahi, M., Lee, S. H., Li, W., Osumi, D., Miyoshi, E., and Taniguchi, N. (2008) N-glycan of ErbB family plays a crucial role in dimer formation and tumor promotion. *Biochim Biophys Acta* 1780, 520-524.

144. Varki, A. (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* 446, 1023-1029.

145. Slieker, L. J., and Lane, M. D. (1985) Post-translational processing of the epidermal growth factor receptor. Glycosylation-dependent acquisition of ligand-binding capacity. *J Biol Chem* 260, 687-690.

146. Slieker, L. J., Martensen, T. M., and Lane, M. D. (1986) Synthesis of epidermal growth factor receptor in human A431 cells. Glycosylation-dependent acquisition of ligand binding activity occurs post-translationally in the endoplasmic reticulum. *J Biol Chem* 261, 15233-15241.

147. Cummings, R. D., Soderquist, A. M., and Carpenter, G. (1985) The oligosaccharide moieties of the epidermal growth factor receptor in A-431 cells. Presence of complex-type N-linked chains that contain terminal N-acetylgalactosamine residues. *J Biol Chem* 260, 11944-11952.

148. Stroop, C. J., Weber, W., Gerwig, G. J., Nimtz, M., Kamerling, J. P., and Vliegenthart, J. F. (2000) Characterization of the carbohydrate chains of the secreted form of the human epidermal growth factor receptor. *Glycobiology* 10, 901-917.

149. Levkowitz, G., Waterman, H., Zamir, E., Kam, Z., Oved, S., Langdon, W. Y., Beguinot, L., Geiger, B., and Yarden, Y. (1998) c-Cbl/Sli-1 regulates

endocytic sorting and ubiquitination of the epidermal growth factor receptor. *Genes Dev* 12, 3663-3674.

150. Tu, B., Ziemann, R. N., Tieman, B. C., Hawksworth, D. J., Tyner, J., Scheffel, J., Pinkus, M. S., Brophy, S. E., Werneke, J. M., Gutierrez, R., and White, M. (2010) Generation and characterization of chimeric antibodies against NS3, NS4, NS5, and core antigens of hepatitis C virus. *Clin Vaccine Immunol* 17, 1040-1047.

151. Whitson, K. B., Whitson, S. R., Red-Brewer, M. L., McCoy, A. J., Vitali, A. A., Walker, F., Johns, T. G., Beth, A. H., and Staros, J. V. (2005) Functional effects of glycosylation at Asn-579 of the epidermal growth factor receptor. *Biochemistry-Us* 44, 14920-14931.

152. Sato, C., Kim, J. H., Abe, Y., Saito, K., Yokoyama, S., and Kohda, D. (2000) Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of recombinant human epidermal growth factor receptor. *J Biochem* 127, 65-72.

153. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14, 103-114.

154. Balasubramanian, B., Pogozelski, W. K., and Tullius, T. D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc Natl Acad Sci U S A* 95, 9738-9743.

155. 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.

Appendix A: Publications

1. **Zhu Y**, Guo T, Park JE, Li X, Meng W, Datta A, Bern M, Lim SK, Sze SK., *Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting.* Mol Cell Proteomics. 2009 Aug;8(8):1999-2010.

2. **Yi Zhu**, Tiannan Guo, and Siu Kwan Sze, *Elucidating Structural Dynamics of Integral Membrane Proteins on Native Cell Surface by Hydroxyl Radical Footprinting and NanoLC-MS/MS.* Nanoproteomics, Methods in Molecular Biology. Vol. 790, Chapter 22, 287-303.

3.Tiannan Guo, **Yi Zhu**, Chee Sian Gan, Sze Sing Lee, Jiang Zhu, Haixia Wang, Xin Li, James Christensen, Shiang Huang, Oi Lian Kon* and Siu Kwan Sze*. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of MET Kinase Inhibitor in Cancer Cells. Molecular & Cellular Proteomics. 2010. 9(12):2629-41.

4. Tiannan Guo, Sze Sing Lee, Wai Har Ng, **Yi Zhu**, Chee Sian Gan, Jiang Zhu, Haixia Wang, Shiang Huang, Siu Kwan Sze*, Oi Lian Kon*. Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome. Cellular and Molecular Life Sciences. 2011. 68 (11): 1983-2002.

Appendix B: Conference

Yi Zhu, Tiannan Guo, and Siu Kwan Sze, *Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting.* Human Proteome Organisation's (HUPO) 9th Annual World Congress, 19-23 Sep 2010, Sydney, Australia. (Oral presentation)