Fall 2010

Characterization of Metastasis-Associated Cell Surface Glycoproteins in Prostate Cancer

Lifang Yang
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/biomedicalsciences_etds
Part of the Biochemistry Commons, Cell Biology Commons, Molecular Biology Commons, and the Organic Chemistry Commons

Recommended Citation
Yang, Lifang. "Characterization of Metastasis-Associated Cell Surface Glycoproteins in Prostate Cancer" (2010). Doctor of Philosophy (PhD), dissertation, Chemistry and Biochemistry, Old Dominion University, DOI: 10.25777/fpy3-tc43
https://digitalcommons.odu.edu/biomedicalsciences_etds/142

This Dissertation is brought to you for free and open access by the College of Sciences at ODU Digital Commons. It has been accepted for inclusion in Theses and Dissertations in Biomedical Sciences by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
CHARACTERIZATION OF METASTASIS-ASSOCIATED CELL SURFACE GLYCOPROTEINS IN PROSTATE CANCER

by

Lifang Yang
M.D., 1998, Beijing Medical University
M.S., 2001, Academy of Military Medical Sciences

A Dissertation Submitted to the Faculty of
Eastern Virginia Medical School and Old Dominion University
in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

BIOMEDICAL SCIENCES

EASTERN VIRGINIA MEDICAL SCHOOL
OLD DOMINION UNIVERSITY
December 2010

Approved by:

Q. John Semmes (Director)
Edward M. Johnson (Member)
Richard A. Britten (Member)
Richard R. Drake (Member)
ABSTRACT

CHARACTERIZATION OF METASTASIS-ASSOCIATED CELL SURFACE GLYCOPROTEINS IN PROSTATE CANCER

Lifang Yang
Eastern Virginia Medical School and Old Dominion University, 2010
Director: Dr. O. John Semmes

Prostate cancer (PCa) is a major health problem in males in the United States. Its lethality is mostly attributed to the primary tumor metastasizing to distant sites that are highly resistant to conventional therapies. Serum Prostate Specific Antigen (PSA) is the only protein biomarker used in clinic for prediction of prostate cancer recurrence following local therapies. Nonetheless, PSA lacks the ability to predict the behavior of an individual tumor in an individual patient. Therefore, development of reliable biomarkers for detection of metastatic potential in primary tumors, as well as discovery of new therapeutic targets, is in a great need for improved disease survival and management. Tumor metastasis is a multistep process involving extravasation of a cancer cell subsequent invasion and expression at a site distal to the primary tumors. Cell surface glycoproteins play pivotal roles as recognition molecules in a range of cell communication and adhesion events. Aberrant cell surface glycosylation has been reported in various cancers including PCa, and strongly correlated with prognosis and metastasis. However, the staggering complexity of glycans renders their analysis extraordinarily difficult. This research project aims to develop a mass spectrometry-based glycoproteomic approach for the selective isolation and identification of cell surface glycoproteins from cellular samples, and apply this technology to the discovery of new glycoprotein biomarkers which are indicative of prostate cancer progression and metastasis. To this end, cell surface glycosylation patterns were characterized by lectin
flow cytometry and lectin cytochemistry on a human syngeneic PCa cell metastatic model, PC3 and its two variants with different metastatic potentials. It was found that metastatic potentials of PC3 variants were inversely correlated with cell surface α2-6 sialic acid levels. Targeted to cell surface sialoglycoproteins, a new glycoproteomic approach was successfully developed, which combined selective metabolic labeling of cell surface sialyl glycans, chemically probing the labeled sugar with a biotin tag, affinity purification of sialylated proteins, SDS-PAGE separation, and subsequent LC-MS/MS for protein identification. Application of this methodology in our prostate cancer model system resulted in unique identification of a total of 80 putative cell surface sialoglycoproteins differentially expressed between PC3 variants. After prioritization of the candidate biomarkers, one cell-based prioritized biomarker CUB-domain-containing protein 1 (CDCP1) was verified in prostate cancer cell lines and clinical samples, including tissues and body fluids, by immunoassays. Results indicated that expression of CDCP1 protein is dysregulated in prostate cancer and it has potential utility as a therapeutic target and a diagnostic marker for PCa progression. Overall, the data from this research project provided the proof-of-principle evidence for our targeted glycoproteomic approach, which we believe will help expedite the discovery of new cancer biomarkers and therapeutic targets in diseases and delineation of signal transduction pathways on a global scale.
This dissertation is dedicated to my husband, Siqi Guo, for his consistent encouragement and patience as I worked on my PhD. I also dedicate this dissertation to my beloved family: my parents, Rujin Yang and Xiuying Wang, my brother Lijiang Yang, and my sister Ruifang Yang for their everlasting love and support.
ACKNOWLEDGMENTS

First of all, I would like to express my sincerest appreciation to my research advisor Dr O. John Semmes, for his constant support, guidance, and encouragement throughout my graduate study. Dr Semmes has offered me great freedom in developing this project and has always been very supportive of me when I had difficulties. The diverse, interdisciplinary nature of his research program, the inspirational discussion at group meetings, and his widespread collaborations make the group an excellence place to grow as a scientist. I have really enjoyed my time and learned much by being in such an environment.

I am also grateful to Dr Richard R. Drake, who provided invaluable insight and care on my project. I feel fortunate to do research work with his help.

I would also like to thank the other members of my dissertation committee: Dr Edward M. Johnson and Dr Richard A. Britten. They have given me a lot informative suggestions and constructive criticism that further contributed to the quality of this work. Their profound knowledge, professional research skills, and great passion for science have had a lasting influence on my career.

Finally, I wish to acknowledge the researchers in our laboratory and the staff of the Department of Microbiology and Molecular Cell Biology. Thanks for their support and friendship. I have had much pleasure of working with all of them.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
</tbody>
</table>

## Chapter

### I. INTRODUCTION

1.1 PROSTATE CANCER PATHOGENESIS AND BIOMARKERS .................................................. 1
1.2 PROTEIN GLYCOSYLATION ......................................................................................... 16
1.3 ALTERED CELL SURFACE GLYCOSYLATION AND CANCER ........................................... 21
1.4 GLYCOPROTEOMICS IN CANCER BIOMARKER DISCOVERY .......................................... 28
1.5 MASS SPECTROMETRY-BASED GLYCOPROTEOMICS .................................................. 32

### II. DISSERTATION RATIONALE AND SUMMARY OF AIMS ................................. 50

### III. AIM I: CHARACTERIZATION OF CELL SURFACE LECTIN-BINDING PATTERNS ASSOCIATED WITH METASTASIS IN HUMAN PROSTATE CANCER CELLS .......................................................... 55

3.1 INTRODUCTION ........................................................................................................ 55
3.2 MATERIALS AND METHODS .................................................................................... 58
3.3 RESULTS .................................................................................................................. 62
3.4 DISCUSSION ............................................................................................................. 80

### IV. AIM II: TARGETED IDENTIFICATION OF METASTASIS-ASSOCIATED CELL SURFACE SIALOGLYCOPROTEINS IN PROSTATE CANCER ........................................................... 84

4.1 INTRODUCTION ........................................................................................................ 84
4.2 MATERIALS AND METHODS .................................................................................... 90
4.3 RESULTS .................................................................................................................. 96
4.4 DISCUSSION ............................................................................................................. 130

### V. AIM III: VERIFICATION OF CELL-BASED BIOMARKERS IN CLINICAL SAMPLES ................................................................................................................ 135

5.1 INTRODUCTION ........................................................................................................ 135
5.2 MATERIALS AND METHODS .................................................................................... 139
5.3 RESULTS .................................................................................................................. 145
5.4 DISCUSSION ............................................................................................................. 170
VI. CONCLUSIONS AND FUTURE DIRECTIONS ............................................. 179

6.1 AIM I: CHARACTERIZATION OF CELL SURFACE LECTIN-BINDING PATTERNS ASSOCIATED WITH METASTASIS IN HUMAN PROSTATE CANCER CELLS .............................................. 179

6.2 FUTURE DIRECTIONS OF AIM I ...................................................... 179

6.3 AIM II: TARGETED IDENTIFICATION OF METASTASIS-ASSOCIATED CELL SURFACE SIALOGLYCOPROTEINS IN PROSTATE CANCER .......................................................... 180

6.4 FUTURE DIRECTIONS OF AIM II .................................................... 182

6.5 AIM III: VERIFICATION OF CELL-BASED BIOMARKERS IN CLINICAL SAMPLES ................................................................. 184

6.6 FUTURE DIRECTIONS OF AIM III .................................................... 186

6.7 CONCLUDING REMARKS .............................................................. 189

REFERENCES .......................................................................................... 190

VITA ........................................................................................................ 214
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TNM staging system of prostate cancer</td>
<td>10</td>
</tr>
<tr>
<td>2. D'Amico et al risk stratification for clinically localized prostate cancer</td>
<td>13</td>
</tr>
<tr>
<td>3. Some of the FDA-approved cancer glycobiomarkers</td>
<td>30</td>
</tr>
<tr>
<td>4. Lectins and the optimized protocol for flow cytometric analysis</td>
<td>74</td>
</tr>
<tr>
<td>5. Total number of proteins identified for each experimental group</td>
<td>113</td>
</tr>
<tr>
<td>6. Top 10 abundant glycoproteins identified in N2 and ML2 cells ranked by unique peptides</td>
<td>114</td>
</tr>
<tr>
<td>7. Top 10 abundant glycoproteins identified in N2 and ML2 cells ranked by emPIA scores</td>
<td>116</td>
</tr>
<tr>
<td>8. Cell surface N-linked Glycoproteins differentially expressed in N2 Cells</td>
<td>122</td>
</tr>
<tr>
<td>9. Cell surface N-linked Glycoproteins differentially expressed in ML2 Cells</td>
<td>124</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>The subclasses and oligosaccharide structures of N- and O-linked glycoproteins.</td>
<td>19</td>
</tr>
<tr>
<td>3.</td>
<td>The process of tumor metastasis.</td>
<td>24</td>
</tr>
<tr>
<td>4.</td>
<td>The workflows of mass spectrometry-based glycoproteomic analysis.</td>
<td>34</td>
</tr>
<tr>
<td>5.</td>
<td>The nomenclature of peptide and glycan fragment ions from a tandem mass spectrum.</td>
<td>42</td>
</tr>
<tr>
<td>6.</td>
<td>The properties of prostate cancer PC3 cell line and its sublines.</td>
<td>63</td>
</tr>
<tr>
<td>7.</td>
<td>Flow cytometric assessment of different cell dissociation methods.</td>
<td>65</td>
</tr>
<tr>
<td>8.</td>
<td>Comparison of different blocking buffers on SNA binding.</td>
<td>68</td>
</tr>
<tr>
<td>9.</td>
<td>The effect of different staining buffers on lectin binding.</td>
<td>69</td>
</tr>
<tr>
<td>10.</td>
<td>The effect of cell confluence on SNA staining.</td>
<td>70</td>
</tr>
<tr>
<td>11.</td>
<td>Titration of SNA against PC3 cells.</td>
<td>72</td>
</tr>
<tr>
<td>12.</td>
<td>Titration of LFA against ML2 cells.</td>
<td>73</td>
</tr>
<tr>
<td>13.</td>
<td>The carbohydrate specificities of lectin binding.</td>
<td>75</td>
</tr>
<tr>
<td>14.</td>
<td>Cell surface carbohydrate profile analysis by lectin flow cytometry.</td>
<td>77</td>
</tr>
<tr>
<td>15.</td>
<td>Cell surface carbohydrate expression analyzed by lectin cytochemistry.</td>
<td>79</td>
</tr>
<tr>
<td>16.</td>
<td>Schematic representation of sialic acid biosynthetic pathway in human cells.</td>
<td>87</td>
</tr>
<tr>
<td>17.</td>
<td>Bioorthogonal chemical reactions for sialic acid.</td>
<td>88</td>
</tr>
<tr>
<td>18.</td>
<td>Outline of Aim II experimental work-flow.</td>
<td>97</td>
</tr>
<tr>
<td>19.</td>
<td>Efficient and selective labeling of sialoglycoproteins on the cell surface.</td>
<td>99</td>
</tr>
</tbody>
</table>
Fluorescent imaging of intracellular sialylation .................................................101
Optimization of azide labeling and click reaction ..............................................102
Detection of azide-tagged sialoglycoproteins in cell lysates ...............................105
Schematic representation of streptavidin (SAv) pull down assays .......................106
The effect of different protein/beads ratios in SAv pull down assays ...................107
Efficient capture and enrichment of azide-tagged sialoglycoproteins ..................109
Separation of proteins enriched by SAv pull down assays .................................110
Number of proteins identified in PC3 variants by LC-MS/MS ...........................111
Improved capture of known glycoproteins .......................................................117
Subcellular location of identified proteins .......................................................118
Overlap of our cell surface proteins with prostate cancer secretome .................120
Functional pathway analysis of glycoproteins unique to N2 cells .....................127
Functional pathway analysis of glycoproteins unique to ML2 cells ...................129
The work-flow of verification of cell-based biomarkers ....................................137
Identification of CDCP1 in ML2 cell surface sialoglycoprotein enriched fraction .....................................................................................................................147
Verification of CDCP1 expression in N2 and ML2 cells .................................151
Verification of CDCP1 expression in prostate cell lines ....................................152
Characterization of CDCP1 glycosylation in vitro ...........................................154
Characterization of CDCP1 glycosylation in vivo .............................................157
Fucosylation of CDCP1 in N2 and ML2 cells ..................................................158
Cellular expression of CDCP1 in N2 and ML2 cells .......................................160
Analysis of CDCP1 subcellular location by colocalization ...............................161
42. Analysis of CDCP1 ER subcellular location by Immunolocalization..........163
43. Analysis of CDCP1 Glogi subcellular location by Immunolocalization..........164
44. Expression of CDCP1 in prostate patient tissues..................................................166
45. Shedding of CDCP1 ectodomain into the extracellular compartment..........169
CHAPTER I
INTRODUCTION

1.1 Prostate Cancer Pathogenesis and Biomarker

1.1.1 Incidence and mortality

Prostate cancer (PCa) accounts for 9.7% of malignant tumors in men and is the most commonly diagnosed cancer among males in the Western world\(^1\). The highest PCa incidence rates (>100.0 per 100,000/year) are found in North America and the lowest in Asia\(^2\). The American Cancer Society estimates there are 217,730 newly diagnosed cases and 32,050 deaths due to PCa in 2010\(^3\). Incidence of PCa increases with age more rapidly than any other cancer. Autopsy studies have shown invasive carcinomas in 8% of men in their 20s rising to 80% for men in their 70s\(^4\). Based on National Cancer Institute (NCI)'s Surveillance and Epidemiology and End Result (SEER) program statistics review, the average annual age-adjusted incidence rate in the United States has experienced 2.6% and 16.5% growth from 1975–1988 and 1988–1992, respectively. This dramatic increase in the late 1980s reflects improvements in detection and diagnosis through widespread use of prostate-specific antigen (PSA) testing, which received initial U.S. Food and Drug Administration approval in 1986. Since the early 1990s, PCa incidence has been declining (Fig. 1). From 2003-2007, the age-adjusted incidence

---

\(^1\) The model journal for this dissertation is *Nature Biotechnology*
Figure 1 Annual age-adjusted prostate cancer incidence and mortality rate among males, US, 1975 to 2007. Figure was generated from Surveillance, Epidemiology, and End Result (SEER) program website (www.seer.cancer.gov). *Incidence source: SEER 9 areas (Connecticut, Iowa, Hawaii, New Mexico, Utah, Atlanta, Detroit, Seattle, San Francisco). Mortality source: US Mortality Files, National Center for Health Statistics, CDC. Rates are age-adjusted to 2000 US-standard population.
rate was 156.9 per 100,000 men per year. The mean age at PCa diagnosis is 67 years, and 91% of the cases are diagnosed after 54 years and less than 1% before 45 years. PCa incidence also differs for different racial and ethnic populations. The rate is highest in Blacks and lowest in Asians. Black Americans have the highest PCa incidence in the world and are considered at high risk for the disease. In the United States, PCa is the second most frequent cause of cancer-related death in men. Overall, annual age-adjusted mortality remains relatively stable in last thirty years although the rate has declined slightly since the early 1990s (Fig. 1). The age-adjusted death rate was 24.7 per 100,000 men per year and the median age at death was 80 years of age during 2003-2007 period.

1.1.2 Etiology and risk factors

Despite the high prevalence of PCa, the pathogenesis of PCa remains unclear. Most experts believe the etiology of prostate cancer is multi-factorial, with advancing age, diet, race, and family history most often implicated. After age and ethnic background, the strongest epidemiological risk factor for PCa is a positive family history. 5-10% of PCa are reported to be hereditary or familial. A man with first degree relatives (father, brother, or son) who have had prostate cancer is two to three times more likely to develop the disease himself and generally associated with early onset. Several genes such as RNaseL, E1AC2, MSR1, BRCA2 have been identified as candidate susceptibility genes. But epidemiologic evidence suggests that familial PCa is a genetically heterogeneous disease and many gene loci rather than a specific major susceptibility gene predispose to familial PCa. Moreover, gene-environment interactions may play a crucial role in PCa development especially for these genes with low-penetrance and polymorphisms. Recent studies have indicated that chronic inflammation and atrophy are very important
in PCa pathogenesis. In particular, there is growing evidence that asymptomatic chronic prostatic inflammation, defined as category IV prostatitis, is prevalent in cases of PCa. Several investigators have proposed a causal link between inflammation and PCa, but currently no firm statement on causality can be made. Dietary factors such as animal fat consumption and calcium from dairy foods are associated with increased risk of PCa, whereas anti-oxidative nutrients like selenium, vitamin E and tomato-based products containing lycopene probably protects against PCa. The elevated insulin-like growth factor (IGF) has been proposed to represent a link between western lifestyle or obesity and PCa, however, a contradictory association was also reported. Androgens are well known to regulate prostate growth and function and may also speed tumor growth, but no definite link between increased circulating androgen levels and PCa has been found. The etiology of PCa is thus poorly understood compared to other common human cancers.

1.1.3 Pathological characteristics and natural history

The prostate gland is located just beneath the bladder and in front of the rectum. It contains three major glandular regions—the peripheral zone, the central zone, and the transition zone. Each glandular zone has specific architectural and stromal features. But in all zones, both ducts and acini are lined by androgen-dependent glandular PSA-secreting cells. Beneath it, there is a proliferative compartment of androgen-independent basal cells, which are recognized by their location, morphology and expression of high molecular weight cytokeratins and p63 as opposed to luminal cells. Endocrine-paracrine cells are irregularly distributed in the gland. Most prostate adenocarcinomas arise in
the peripheral zone, and are characterized by multi-focality and morphological and molecular heterogeneity, typically lacking basal cells\textsuperscript{7,19,20}. There are two types of prostatic intraepithelial neoplasia (PIN) one of which is considerate to be a forerunner of cancer. The first is low grade (LGPIN) which is not associated with PCa. The other, high grade (HGPIN) predicts risk of subsequent cancer\textsuperscript{21,22}. One characteristic of HGPIN is the retention of the basal cell layer which harbors over-proliferative phenotype. One study discovered that HGPIN coexists with PCa in 85\% of prostate surgery specimens, and that PIN may appear as soon as ten years before cancerous cells develop\textsuperscript{23}. The prostate lesions, named proliferative inflammatory atrophy (PIA), are usually located in the periphery of the gland near the prostate carcinoma, have been proposed as precursors for PIN or carcinoma\textsuperscript{24}, but the hypothesis presently remains controversial\textsuperscript{25}.

The natural history of PCa is highly variable\textsuperscript{26}. Some tumors simply do not continue to grow or only grow at an extremely slow rate and never form metastases. Autopsies of men who died of unrelated causes have demonstrated that latent PCa is present in about 70\% of men in their 60s\textsuperscript{4,27}. However, the majority of PCa grows and expands around the urethra, causing urinary problems similar to benign prostatic hypertrophy (BPH). When the cancer is still confined to the prostate itself, it is called a "localized disease"\textsuperscript{28}. By the time the tumor is large enough to cause symptoms, it has often spread beyond its capsule. PCa may invade surrounding fat and tissue, the seminal vesicles, and/or the neck of the bladder. The cancer cells may invade lymph nodes in the pelvic region. Later, PCa can spread to the bones, primarily those in the lower spine, pelvis, and femur. Other organs such as the liver, brain, or lungs can also be metastatic sites\textsuperscript{29}. Once the cancer has spread distantly from the prostate, it is difficult to
effectively treat and completely cure. If the cancer spreads without intervention, or if treatment has failed, PCa can ultimately result in death.

1.1.4 Diagnosis and treatment

Two standard screening examinations, digital rectal examination (DRE) and serum prostate-specific antigen (PSA), are the first step in diagnosing prostate cancer, no matter symptoms are present or not. If a man's PSA level is elevated and/or his digital rectal exam is abnormal, ultrasound directed needle biopsy and histological examination are usually performed. The widespread use of PSA monitoring for early detection has greatly changed the way PCa is diagnosed. The increase of PCa detected due to increased screening examinations has revealed problems with PSA, such as lacking specificity and without lower limit to completely exclude PCa, and cause some dilemmas in clinical practice. For example, there has been an increasingly large number of men with elevated PSA and an initial set of negative biopsies. The negative result does not exclude the presence of cancer. A high percentage of these men require repeat biopsies. However, the threshold level of PSA that should indicate biopsy, the optimal number of biopsy cores for detecting PCa in prostate biopsy, and when to stop biopsying remains controversial. The absence of clear guidelines in this field can delay diagnosis and treatment of PCa patients, as well as cause cancer-free men anxiety and side effects. For patients who have been diagnosed with PCa, various diagnostic imaging tests such as MRI, CT, and PET are the options to determine the extent of the tumor in the prostate and whether cancer cells have spread to surrounding tissues or other parts of the body.
Treatment for PCa depends on many factors: the patient’s age, the risk of the disease, the presence of other medical conditions, and the patient’s overall health. To date, there are five major treatment options for patients with PCa: watchful waiting, radical prostatectomy, radiotherapy (external and internal radiation), hormone therapy, and chemotherapy. For localized PCa, radical prostatectomy is standard treatment for patients with a life expectancy >10 years, whereas radiotherapy is an alternative. The randomized studies comparing these two therapies suggest that they provide similar rates of cancer control at 10 years, and that except for the youngest patients, choice of therapy should be based on toxicity and quality of life. Patients with a life expectancy <10 years and low-moderate grade tumors may be offered "watchful waiting", in which patients are under active surveillance. For locally advanced PCa, combined radiotherapy and neoadjuvant total androgen blockade (LH-RH agonist plus anti-androgen) is the preferred treatment. The role of radical prostatectomy in this population has been debated. Patients with metastatic disease are treated with individual or combined hormonal therapies, chemotherapy (docetaxel), radiotherapy, surgical treatments, and antiresorptive medications.

1.1.5 Prognosis and risk stratification

PCa is a markedly heterogeneous disease with considerable variation in clinical course ranging from asymptomatic disease to a rapidly progressing fatal malignancy. Sophisticated risk assessment is required in clinic. Reliable methods to identify high-risk patients are essential to help clinicians make therapeutic decisions to achieve maximal cancer control. Currently, three preoperative clinicopathologic parameters are used to
assess PCa risk. They are clinical tumor stage, biopsy grade and volume parameters, and preoperative serum PSA.

PCa is staged anatomically. Stage I (A) corresponds to clinically inapparent tumors usually found microscopically. Stage II (B) refers to a palpable tumor confined with the prostate. Stage III (C) indicates that cancer extends outside of the capsule of prostate, but only barely. Stage IV (D) signifies involvement of cancer to lymph nodes or distant tissue. Since 1992, the consensus TNM staging (Table 1) has gained general acceptance and has replaced the old A-B-C-D system. The stage level is an important prognostic factor in PCa. In particular, seminal vesicle invasion and regional lymph node involvement are inversely associated with survival. Also, capsular penetration is associated with increased risk of recurrence. However, the prognostic role of substaging in organ-confined disease is less informative.

PCa is also staged pathologically using the Gleason system, which is the method of predicting a prognosis based on the histological appearance of cells obtained from either a needle biopsy or a prostatectomy. Small samples of prostate cancer cells are reviewed under the microscope and assigned a pattern number from 1 to 5. Patterns 1 and 2 correspond to well-differentiated cancers. Pattern 3 indicates a moderately differentiated cancer, and pattern 4 and 5 correspond to poorly differentiated or anaplastic lesions. The advantage of Gleason grading over other grading system is that it accounts for the remarkable heterogeneity of PCa. It records two predominant patterns and adds them together to give a Gleason score range from 2 to 10. If only one pattern is present, that pattern number is doubled to obtain the Gleason sum score (for example 4+4=8). The higher the score, the more aggressive the cancer. In many multivariate analyses, the
Gleason score proves to be an independent predictor of both pathologic tumor stage and time to biochemical recurrence.\(^{44-46}\) Gleason score is commonly stratified as: score 2–4 (well differentiated); 5–6 (moderately differentiated); 7 (moderately to poorly differentiated); and 8–10 (poorly differentiated). Score 7 tumors can further be subclassified into 3+4 or 4+3, depending on which grade is most prevalent in the cores.\(^{22}\) This type of Gleason score classification and subclassification has been validated to predict postoperative outcomes.\(^{47,48}\) However, the Gleason grading system has been hampered by poor reproducibility.\(^{49,50}\) Intra- and inter-pathologist variations of interpretation are common, reported in up to 64% of cases.\(^{51}\) Another limitation of Gleason score system is poorly correlation of scores from needle biopsies with Gleason scores of specimens obtained surgically during prostatectomies. Discordance rates of 43% to 46%, with upgrading in 35% of cases, have been reported.\(^{52}\)

Tumor volume is associated with adverse pathologic findings (high grade, extracapsular extension, seminal vesicle involvement, positive margins, lymph node metastasis, and ultimate disease progression) after a radical prostatectomy.\(^{53}\) The prognostic impact of tumor volume however, is variable, largely depending how to quantitatively measure tumor volume.\(^{54}\) Several studies report the percentage of the prostate involved by the tumor and maximum tumor diameter to be an independent predictor of recurrence.\(^{55,56}\) In needle biopsies, tumor volume as measured by number of positive cores or the percentage of positive cores, is also associated with recurrence or progression after surgery or radiation therapy.\(^{57}\) Tumor volume probably has little prognostic independence once the Gleason score and pathological stage are known.\(^{22}\)
Table 1 TNM staging system of prostate cancer

<table>
<thead>
<tr>
<th>T = Tumors that is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0: No evidence of primary tumor</td>
</tr>
<tr>
<td>T1: Tumor present, but undetectable using DRE or imaging</td>
</tr>
<tr>
<td>T1a: Incidental in the resected specimen (&lt;5% tissue involvement)</td>
</tr>
<tr>
<td>T1b: Incidental in the resected specimen (&gt;5% tissue involvement)</td>
</tr>
<tr>
<td>T1c: Found by biopsy with rising PSA</td>
</tr>
<tr>
<td>T2: Tumor palpable and localized tumor</td>
</tr>
<tr>
<td>T2a: Half or less than half of one lobe involved</td>
</tr>
<tr>
<td>T2b: More than half of one lobe, but not both involved</td>
</tr>
<tr>
<td>T2c*: Both lobes involved</td>
</tr>
<tr>
<td>T3: Extended to the capsule, but not beyond prostate</td>
</tr>
<tr>
<td>T3a: Extracapsular involvement (unilateral, bilateral)</td>
</tr>
<tr>
<td>T3b: Seminal vesicle involvement</td>
</tr>
<tr>
<td>T4: Fixed or invading adjacent organs (such as bladder or rectum)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N = Lymph node metastasis in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>N0: No spread to the regional lymph nodes</td>
</tr>
<tr>
<td>N1: Spread to the regional lymph nodes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M = Distant metastasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0: No distant metastasis</td>
</tr>
<tr>
<td>M1: Distant metastasis</td>
</tr>
<tr>
<td>M1a: Metastasis to lymph nodes beyond the regional ones</td>
</tr>
<tr>
<td>M1b: Metastasis to bone tissue</td>
</tr>
<tr>
<td>M1c: Metastasis to other organs (regardless of bone involvement)</td>
</tr>
</tbody>
</table>


* "T2c" implies a tumor which is palpable in both lobes of the prostate. Tumors which are found to be bilateral on biopsy only but which are not palpable bilaterally should not be staged as T2c.
Despite limitations as a screening tool, PSA testing may still supply valuable clues to the risk of tumor recurrence and treatment response. Numerous studies have confirmed elevated serum PSA was associated with larger tumor volume, higher clinical stage, and Gleason grade \(^{58}\). PSA seems to be a strong predictor of aggressive PCa. A Malmo Preventive Medicine study with 21,277 men participating showed that PSA level at age 44-50 was highly predictive of advanced prostate cancer diagnosed up to 25 years later, with advanced cancer defined as T3, T4 or metastasis at diagnosis. But the prognostic accuracy decreased with age \(^{27,59}\). Preoperative PSA is also a sensitive indicator of recurrence after radical prostatectomy. In a large study with a cohort of 1582 men who underwent radical prostatectomies, lower preoperative serum PSA was reported to be associated with a decreased incidence of positive surgical margins, extracapsular disease, seminal vesicle invasion and lymph node involvement. Multivariate analysis suggested that only serum PSA levels and the biopsy Gleason score predicted time to recurrence \(^{60}\). However, contrasting results are reported. Stamey's group has shown PSA values may only presage pathologic Gleason scores if PSA values are either extremely high or low, and less reliable as a prognosticator when PSA level was between 2 and 9 ng/mL \(^{61,62}\). These results support a previous study in which no clinically useful relationship was found between preoperative PSA and biochemical progression in large volume cancers \(^{63}\). Therefore, preoperative PSA seems insufficient to allow useful predictions for individual patients based on PSA as a single marker. Predictions can be improved by PSA derivatives such as PSA doubling time (PSADT) and PSA velocity (PSAV), as they have demonstrated the significant relationship with PCa aggressiveness \(^{64-66}\). However, the use of these indications in the pretreatment setting is more
controversial, and none of the measures has been shown to increase diagnostic and
prognostic accuracy better than a simple PSA test$^{27,67}$.

Recognition of the limitations of above single prognostic factors has led to the
development of multivariable tools that incorporate preoperative clinicopathologic factors
in an array in an attempt to better predict pathological tumor features and treatment
outcome. Current diagnosis employs a mix of biopsy Gleason score, clinical tumor stage
and preoperative PSA levels$^{62,68}$. Currently, the most commonly used stratification
system was developed by D'Amico et al. which divides patients into low, intermediate,
and high risk (Table 2)$^{47}$. This tool has been validated for use in men treated with
radical prostatectomy and radical radiotherapy$^{47,69}$. Other combined prognostic tools
include Partin Tables, and statistical model nomograms, which have different predictive
significances and limitations, and their universal application still needs to be validated$^{62,68}$.

1.1.6 Current challenges in clinical PCa management

The introduction of prostate-specific antigen (PSA) into the paradigm of early
detection, as well as a lower PSA threshold (in the range of 2.5-3.0 ng/mL rather than
traditional 4.0 ng/mL) used in clinic for recommending prostate biopsy, have resulted in
profound changes in stage and grade of PCa. As expected due to increased PSA
screenings, stage has migrated towards more localized, and curable diseases (T1c-T2a)$^{70}$. Results from the SEER database also showed that the percentage of PCa patients with
metastatic disease at the time of diagnosis dropped from 16% in the 1980s to 4% in 2003
$^{71}$. Unexpectedly, grade has shifted heavily toward moderate differentiation (Gleason 5-
Table 2 D’Amico et al. risk stratification for clinically localized prostate cancer

<table>
<thead>
<tr>
<th>PCa Risk</th>
<th>Serum PSA (ng/mL)</th>
<th>Gleason Score</th>
<th>Clinical Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk</td>
<td>&lt;= 10</td>
<td>&lt;= 6</td>
<td>T1c, T2a</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td>10-20</td>
<td>7</td>
<td>T2b</td>
</tr>
<tr>
<td>High risk</td>
<td>&gt; 20</td>
<td>8-10</td>
<td>T2c, T3</td>
</tr>
</tbody>
</table>

PSA = prostate-specific antigen

The risk stratification scheme, developed by D'Amico et al. is currently in widespread use for prognostic assessment of clinically localized prostate cancer. Patients are divided into three risk groups (low, intermediate, or high) of occult micrometastases and relapse by the combination of three preoperative clinicopathologic factors: biopsy Gleason score, serum PSA, and clinical stage. When there is a conflict in three parameters, the worse factor wins and determines the risk category assignment. For example, a patient with a low Gleason Score of 6 combined with a PSA of 25 ng/mL is considered high risk. Table adapted from D’Amico et al. J Urol. (1998).
7), whereas rates of poorly (Gleason 8-10) and well-differentiated (Gleason 2-4) disease have remained relatively stable \(^7\). According to D'Amico's risk categories, most diagnoses of PCa belong to the low-risk group, the proportion of which increased from 30% in 1990 to 45% in 2000 \(^7\). The low-risk tumors, however, can be either indolent or aggressive. There is currently no means that allow urologist/oncologist to differentiate the minority of men with potentially lethal cancer from a large proportion of those whose PCa pose no threat to life or health. The net result is an unfortunate trend to over-diagnosis and over-treatment of clinically insignificant PCa \(^7\). This can create the potential for patient's mental burden and the morbidity and mortality associated with aggressive local therapies. On the other side, patients with potential aggressive disease may omit adjuvant intervention and miss a crucial opportunity to be cured. There is a significant need for an improved risk-stratification system that incorporates novel biomarkers to better determine the risk of recurrence and allow patients to make more informed decisions about treatment.

For patients with significant disease, the major challenge is the prediction of clinical progression and survival after defined local therapy. It is reported that about 25% of men with PCa ultimately will experience subsequent recurrent disease after radical prostatectomy within 15 years. A portion of these men will go on to experience clinically expressed extracapsular disease, with as many as two-thirds having evidence of bone involvement \(^29\). Monitoring of post-treatment PSA provides a means of detecting recurrent disease long before the tumor is detected by other means such as imaging or symptoms. Increasing PSA levels after treatment, also called 'biochemical recurrence", generally indicates the presence of tumor cells. However, It is noteworthy that PSA
recurrence merely represents a surrogate endpoint marker, as only 34% of these patients develop distant metastases after a median follow up of 8 years. In line with this, many patients will experience a post-treatment PSA elevation that confers no prognostic information, referred to PSA only progression. In addition, there is considerable debate on the best criteria for defining biochemical recurrence. No single PSA value is invariably associated with clinical metastasis or cancer-specific survival. Moreover, PCa can recur in the absence of a PSA increase or detectable PSA level. Hence, biochemical failure after local therapy does not perform well in this field, and there is an equally urgent need for new biomarkers which can serve as surrogate endpoints for clinical recurrences or cancer specific survival.

The first choice to manage distant metastatic disease is hormonal modulation. But nearly all cancers will eventually acquire androgen independence, becoming refractory to androgen deprivation therapy (ADT). The median survival of a patient after developing hormone-refractory prostate cancer (HRPC) is approximately 12 months, and few palliative treatment options are available which have shown benefit at this stage. PSA tests are frequently used to monitor the effects of the treatment in men with HRPC. However, PSA level in patients with such an advanced stage is only moderately associated with survival, and PSA can't accurately reflect the treatment response. Currently the greatest need for stage IV patients is to find more accurate makers of response as well as develop novel agents to prolong survival.

In summary, PCa is a major public health issue in the United State. The development of PSA for PCa detection and prognostication has revolutionized the
management of this disease. Nevertheless, the limitations of PSA, as described above, underscore the necessity to find more accurate diagnostic and prognostic biomarkers that can predict the natural history of the disease, assess recurrent risk, and facilitate stratification of patients to different treatment protocols aimed at improving survival outcome.

1.2 Protein Glycosylation

Covalent attachment of carbohydrates to proteins is one of the most common and versatile post-translational modifications with more than 50% of eukaryotic proteins thought to be glycosylated. Protein glycosylation is prevalent in membrane and secreted proteins where they play a decisive role in cellular recognition events involved in cell adhesion, cell-to-cell communication, and receptor-ligand interactions. Glycans also critically influence the physicochemical properties of proteins that impact protein folding, solubility, and turnover. Hence, glycans expand the biological information carried by proteins. Glycans also confer glycoproteins the nature of complexity which arises from both the structural diversity and elaborate biosynthetic processes.

The glycans associated with human proteins are composed of various combinations of seven monosaccharide units. The monosaccharide units are glucose (Glc), mannose (Man), galatose (Gal), fucose (Fuc), N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc) and sialic acid (SA). Free monosaccharides can exist in an open form, but in nature they are usually found as five or six sided ring structures. For carbohydrates in a cyclic form, the configuration is given as α or β anomers.
Monosaccharides are commonly joined together to form di-, tri-, oligo-, and polysaccharides by linking the anomic carbon atom of one monosaccharide to any hydroxyl group on another monosaccharide. This reaction determines the type of linkage (α- or β-bond), which is essential for enzyme recognition. Furthermore, bonds can be made between the different carbon atoms of the two monosaccharides, for example, carbon 1 of one monosaccharide and carbon 3 of the adjacent monosaccharide, denoted as (1→3). In addition, the monosaccharides can form both linear and branching chains. Therefore, a combination of the number and sequences of monomeric units, their linkage, a number of conformations, as well as occurrence of branch points means there are theoretically astonishingly large number of different structures possible. Although not all possible oligosaccharides are observed in nature, glycan structures are still extremely varied and complex. Moreover, glycan structures can be modified by diverse substitutions including sulfation, phosphorylation, methylation, and acetylation, which also contribute to the structural diversity.

The point of attachment of the sugar chain to the polypeptide is another level of complexity. Glycosylation can occur at different amino acid residues in the protein sequence. Several types of glycosylation are known: (i) the N-linked glycans which attached to the amide group of asparagine residues in a consensus Asn-X-Ser/Thr sequence (X can be any amino acid except proline) via an N-acetylglucosamine (GlcNAc) residue; (ii) the O-linked glycans where the sugar is commonly attached to the hydroxyl group on serine/threonine residues or rarely to other amino acids, e.g. tyrosine, hydroxyproline and hydroxylysine. (iii) glycosylphosphatidylinositol anchors which are attached to the carboxyl terminus of some membrane-associated proteins; (iv) C-
glycosylation that occurs on tryptophan residues in certain membrane-associated and secreted proteins \(^{89}\); (v) S-linked glycosylation where the sugar is attached through a sulfur atom on cysteine or methionine \(^{90}\). The most common and widely studied forms are N-linked and O-linked glycosylation. Fluctuations can occur in both the occupancy rate of O- and N-linked sites on a protein and in the substructures of carbohydrates attached to a specific site. The former, referred to macroheterogeneity, results from the fact that not all potential glycosylation sites are actually glycosylated. The later, called microheterogeneity, is normally a range of glycan structures associated with each site. These phenomena are a consequence of the competitive action of diverse enzymes during biosynthesis \(^{87}\).

Glycan structural diversity requires a finely tuned glycosylation process to accomplish. N-linked glycosylation is a highly specific cotranslational reaction which occurs during protein synthesis before the protein is folded into its final conformation. When the nascent protein enters the endoplasmic reticulum (ER), a block of core precursor sugars (Glc\(_3\)Man\(_9\)GlcNAc\(_2\)), which is a highly conserved evolutionary process in ER, is transferred to the amino group in the asparagine side chain by the oligosaccharyltransferase complex \(^{91}\). The glycan is later processed by stepwise trimming and stepwise addition of new monosaccharide residues by a number of enzymes in the ER and Golgi, generating the three main classes of N-glycans found in mature glycoproteins with a common conserved pentasaccharide core Man\(_3\)GlcNAc\(_2\) (Fig. 2A). The three main classes of glycans are: (i) High mannose – only mannose residues are attached to the core structure. (ii) Complex N-glycans – characterized by the
Figure 2 The subclasses and oligosaccharide structures of N- and O-linked glycoproteins. (A) The representative three types of N-linked oligosaccharides. N-glycans are attached to the protein backbone by linkage to asparagine in the consensus sequence Asn-Xaa-Ser/Thr. They share a universal pentasaccharide core sequence Man$_3$GlcNAc$_2$, and are classified into three types: (1) high mannose, in which only mannose residues are attached to the core; (2) complex, in which various antennae are attached to the core; and (3) hybrid, in which only mannose residues are attached to the Man$_{1-6}$ arm of the core and one or two antennae are on the Man$_{1-3}$ arm. (B) The representative four types of O-glycan cores. The O-glycans are commonly attached to the protein via serine or threonine residues. The most basic O-glycan core structures are observed linked to a single Gal residue, however it can be extended by lactosamine repeats, fucosylation and sialylation with different linkages. In both figure, directionality is from nonreducing end at the top to the reducing end at the bottom with the arrows indicating the extension at the nonreducing end. Linkages between monosaccharides contain the anomeric configuration of the monosaccharide (α, alpha and β, beta). Glycan symbol nomenclature is indicated in the low panel. In the case of complex N-linked glycans, the common terminal motifs attached to Gal are shown in a dotted box. Adapted from Raman, R. et al. Nat Methods 2, 817 (2005).
presence of variable numbers of antennae as well as N-acetyllactosamine (LacNAc, Gaβ1-3/4GlcNAc) in their antennal region. They can be further capped by fucose and sialic acid residues to form a variety of oligosaccharide structures. These N-glycans constitute the most abundant class of mammalian oligosaccharide. (iii) Hybrid N-glycans—containing both a mannose residue and N-acetyllactosamine attached to the trimannosyl chitobiose core, sharing structural features of the high mannose type and complex type. The presence or absence of an α-fucosyl residue on the core mannose of three types of N-glycans further adds to the structural diversity of the oligosaccharides.

In contrast to N-glycans, the biosynthesis of O-glycans is a post-translational event that takes place in Golgi after protein N-glycosylation, folding and oligomerization. Because there is no clear consensus sequence analogous for O-glycans, the synthesis is simpler than for N-glycans. An example of the process is the most prevalent form mucin-type of O-glycosylation. Generally, the process involves sequential single monosaccharide transfer steps. The first step is normally the transfer of N-acetylgalactosamine (GalNAc) to Ser/Thr residues of the protein backbone, by the action of specific polypeptide GalNAc transferases (ppGalNAcTs). To date, more than 24 ppGalNAcTs have been described, with differential tissue expression and overlapping but different specificities. Subsequent addition of galactose (Gal) and/or GlcNAc by specific transferases leads to the formation of the common O-glycan core structures (Fig. 2B). There are at least eight core structures, four of which are particularly widespread in human. The basic core structures can be further extended and terminated by the transfer of lactosamine repeats, sialic acid and fucose residues to give a heterogeneous range of glycans. Besides GalNAc, single GlcNAc can also be attached to Ser/Thr
residues to ubiquitously modify cytoplasmic and nuclear O-linked glycoproteins. O-GlcNAcylation has been found to be as abundant and dynamic as phosphorylation to regulate cellular signal transduction.  

Carbohydrates are constructed stochastically by a complicated pathway. Unlike genes and protein, carbohydrates are generated through non-template-driven biosynthesis and regulated by multiple factors, including metabolic levels of sugar nucleotides, competitive expression and localization of glycosylation enzymes (glycosyltransferases, glycosylhydrolases), and protein-trafficking mechanisms. Due to these features, glycans are expressed in a cell type-specific fashion and subjected to alterations under different physiological and pathological settings.

1.3 Altered Cell surface Glycosylation and Cancer

During malignant transformation and cancer progression, tumor cells express a unique repertoire of glycosylation patterns. The common types of aberrant glycosylation associated with cancer are: (i) hypo, hyper or newly glycosylated protein glycosylation sites; (ii) change in the glycan biosynthetic machinery including upregulation of fucosyltransferase (FucTs), sialyltransferases (SiaTs), N-acetylglucosaminyltransferase V (GnTV) which is involved in the elaboration of highly branched N-linked glycans; (iii) increased levels of fucosylation; (iv) alterations in sialylation; (v) addition of polylactosamine units; (vi) increased β1-6 GlcNAc branching in N-linked structures; (vii) truncated glycoforms such as terminal N-acetylgalactosamine (GalNAc) or Tn and STn antigens in O-linked structures; (viii) inappropriate synthesis of blood group related carbohydrate antigens such as histo-
blood group A/B antigens\textsuperscript{105,106}, or Lewis antigens (i.e. sialyl-Le\textsuperscript{x}, sialy-Le\textsuperscript{a}, and Le\textsuperscript{y})\textsuperscript{107,108} in either N-linked or O-linked structures. Among these changes, altered sialylation of tumor cell surface glycoconjugates is most often correlated with a metastatic tumor phenotype. Changes have been reported reflecting the amount, type, distribution, and bonding of sialic acids to adjacent molecules\textsuperscript{86,109}. Several reports showed a positive correlation between the levels of cell surface sialylation and the metastatic ability of various experimental tumors\textsuperscript{110,111}. For instance, an increase in the amount of N-acetylneuraminic acid (Neu5Ac) on glycan chains has been reported in colorectal cancers and associated with the metastatic spread of colon cancer cells\textsuperscript{112}. However, the correlation between cell surface sialylation and metastatic potentials of tumor cells is not absolute. Sometimes the distribution of sialic acid on specific N- or O-linked oligosaccharides\textsuperscript{113} or alterations in the attachment of sialic acid to its neighboring monosaccharide\textsuperscript{86}, rather than the total quantity of sialic acid, may affect the metastatic properties of these cells. For example, Neu5Ac is usually attached to adjacent monosaccharides via an $\alpha2\rightarrow3$ or an $\alpha2\rightarrow6$ bond. However, the attachment in an $\alpha2\rightarrow8$ bond results in the structure called polysialic acids, which are at an increased levels in lung cancer\textsuperscript{114} and breast cancer\textsuperscript{115} and are associated with pancreatic tumor cell invasion\textsuperscript{116}. Moreover, altered sialylation can also results from other glycosylation changes. One example is increased $\beta1$-$6$ branching can increase the number of Gal$\beta1$-$4$GlcNAc structures that can be sialylated, leading to a total increase in cell surface sialylation\textsuperscript{117}. Therefore, alternation in cell surface sialylation is a complex and dynamic process.
Alterations of the carbohydrate structures have important clinical and functional relevance. Clinicopathological studies showed that high expression of some glycosyl epitopes promotes invasion and metastasis, leading to a shorter 5-10 year survival rate for patients, whereas expression of the glycosyl epitopes suppress tumor progression, leading to higher postoperative survival rates. The concept of glycosylation-dependent promotion or inhibition of tumor progression has been further supported by the data from human tumor cell lines and experimental animal studies. Modification of surface carbohydrate expression by specific strategies such as glycosylation inhibition, enzymatic treatment, gene transfection, and gene knockout mice, brings direct evidence that cell surface carbohydrates are very important in the regulation of metastatic behavior of tumor cells, as detailed below.

Metastasis is defined as the growth of a cancerous cell in an environment noncontiguous with its tissue of origin. It is a complex process occurring in a stepwise fashion (Fig. 3). At the beginning, individual primary tumor cells gain the capacity to degrade and migrate through the basement membrane and extracellular matrix (ECM). Then the tumor cells access and intravasate into the local vasculature, interacting with host cells such as endothelial cells, platelets, and leukocytes. After traveling around the body, tumor cells lodge in the small vessels of distant organs and exit from the circulation by extravasation. Eventually tumor cells survive and proliferate at the secondary sites to become a pathological metastasis. The mechanisms of each event are functionally maintained by a combination of defined molecules involved in the process, each of which is affected directly or indirectly by glycosylation. Particularly tumor cell proliferation, invasion, and extravasation of metastatic cascade are of high relevance in glycobiology.
Figure 3 The process of tumor metastasis. Metastasis is a complex and multiple-stage process. (A) Tumor proliferation at the origin site is crucial at early stages of progression. (B) During invasion, tumor cells gain the capacity to degrade, and migrate through, basement membranes and extracellular matrix. (C) Tumor cells access the local vasculatures and intravasate into the bloodstream or lymphatic glands, where they aggregate with host cells such as platelets and lymphocytes and eventually lodge in the small vessels of distant organs. (D) During extravasation, tumor cells adhere to capillary wall and escape from the blood vessel. (E) In the distal organ, tumor cells proliferate to form micrometastases. (F) Once new angiogenesis is set up, tumor cells are able to form pathological macrometastases. Adapted from Alberts, B. et al. Molecular Biology of the Cell, 4th ed., New York: Garland, 1325 (2002).
Tumor proliferation is very important in the early stages of progression and after metastasis. A broad spectrum of glycans affects tumor proliferation. For example, many studies indicate that N-glycans of many growth factor receptors regulates signal transduction by modulating receptor functions. N-glycosylation of the epidermal growth factor receptor (EGFR) is required for receptor dimerization and transforming activity. Treating U251 glioma and BXPC3 pancreatic adenocarcinoma cells with N-glycosylation inhibitor tunicamycin can disrupt EGFR signaling and enhance radiosensitivity. Similarly, O-lined glycans and mucins play a role in tumor cell proliferation. Mucins are glycoproteins that contain numerous O-glycans in clustered domains along the core protein. Most tumor cells express mucins as transmembrane proteins on the cell surface or as secreted proteins. MUC4 is reported to be involved in the pathogenesis of pancreatic cancer and aberrantly expressed in many other epithelial carcinomas. MUC4 contains an EGF-like domain on its extracellular part, which can directly interact with ErbB2 initiating phosphorylation of the receptor tyrosine kinase in the absence of typical ErbB2 ligands. Therefore, MUC4 can modulate HER2/ErbB2 signaling and is a determinant of therapeutic outcome of Herceptin-based therapy for patients with breast cancer. High expression of another mucin family member MUC1 has been also implicated in many solid tumors including prostate cancer. Recent investigation suggests that MUC1 can be associated with intracellular signal molecules such as Ras and β-catenin to regulate tumor cell growth.

During invasion, tumor cells detach from each other, invade, and migrate through the neighboring tissues. This requires the coordinated remodeling of cell surface adhesion molecules and secretion of degradative enzymes (proteases and glycosidase).
Glycans are involved in each of these stages. The presentation of cancer-related glycans generally reduces cell-cell adhesion and promotes migration. Studies show structural modification of N-glycans on N-cadherin with increased branched structures by enzyme GnTV, led to decreased cell-cell adhesion, promoting tumor cell detachment and invasion. However, small interfering RNA-directed knockdown of GnTV in HT1080 fibrosarcoma cells resulted in enhanced N-cadherin-mediated cell-cell adhesion as well as decreased cell migration and invasion. Expression of sialylated glycans is also reported to promote dissociation of tumor cells. Transfection of breast cancer cells with enzyme ST6Gal-I, which is responsible for sialic acid capping of terminal galactose residues on N-linked glycans, increased cell migration and reduced cell to cell adhesion. In contrast, transfection with antisense ST6Gal-I RNA enhances homotypic cell-cell adhesion. Enhanced expression of sialic acid is also found on the O-linked sugars such as sialyl Tn antigen (STn), which is commonly overexpressed in the mucin-rich surfaces of cancer cells. Transfection of breast cancer cells with cDNA that encodes the human sialyltransferase responsible for STn biosynthesis (ST6GalNAc-I) reduced cell to cell interaction and increased the migration potential of transfected cells. Besides altered cell to cell interaction, local ECM remodeling is essential for tumor cell invasion. Integrins represent a very important class of cell surface adhesion receptors mediating attachment to ECM protein ligands, such as fibronectin and laminin. Numerous integrin biological processes including expression, heterodimerization, ligand binding, and downstream signaling, are modulated by glycosylation. For instance, sialylation led to integrin activation, increased adhesion to ECM ligands, and augmented migration and invasion of colon cancer cell lines in vitro. In addition to sialylation, integrins are also modified
by GnTV, and as a result, carry high level of β1,6-branched structures\textsuperscript{137}. Increased
GnTV expression disrupts the ability of integrins to cluster on the tumor cell membrane,
which in turn, reduces the formation of tumor cell focal adhesion and increases tumor
motility and invasion. Similarly, \textit{in vivo}, spontaneous mammary tumors, which
developed in a GnTV-deficient mouse model, exhibited decreased growth, metastasis,
and signaling downstream of the focal adhesion kinase\textsuperscript{138}.

Like invasion, extravasation requires a series of steps. After entry into the
vasculature, tumor cells, stimulated by some blood-borne factors, aggregate with
activated platelets, and/or leukocytes to form microvascular occlusions in the small
vessels of distant organs, initiating extravasation\textsuperscript{123}. The molecular mechanisms by
which cancer cells traverse the vasculature is thought to occur through “hematopoietic
mimicry”, an identical traffic control axis used by hematopoietic progenitor cells (HPC)
into bone marrow and lymphocytes homing to sites of inflammation. This shear stress-
dependent process begins with rolling on endothelia cells, which leads to tethering and
integrin-dependent stable adhesion. The first step of extravasation, rolling of tumor cells
along the endothelium, is regulated by glycosylation\textsuperscript{96}. Rolling is mediated by the
adhesion of selectins to carbohydrate ligands on the surface of tumor cells. The selectin
family consists of three members each exhibiting differential expression patterns.
Endothelial cells display E-selectin, platelets express P-selectin, and leukocytes present
L-selectin\textsuperscript{139}. It is thought that tethering and rolling of tumor cells occurs through
selectin-mediated interaction between malignant cells and the endothelium. However,
interactions between malignant cells and host cells, such as circulating platelets and
leukocytes, are also indicated to participate in tumor cell extravasation\textsuperscript{140}. Carbohydrate
ligands for selectins typically consist of clustered arrangements of oligosaccharides that bear sialic acid, fucose, and sulfate, predominantly present at the outmost layer of O-linked glycans. One class of carbohydrate ligands are Lewis type blood group antigens, sialyl Lewis X (sLe\(^X\)) and sialyl Lewis A (sLe\(^A\)). Tumor cells frequently overexpress sLe\(^X\) and sLe\(^A\) bearing glycoforms of cell surface molecules such as glycoprotein CD44, CEA, PCLP, MCAM, and glycolipids. The presence of these ligands corrects with metastatic formation and patient survival for several types of cancers. Recent experimental animal studies confirmed a clinical correlation is a reflection of a causal relationship. Mice lacking L-selectin, P-selectin, or both, exhibit greater resistance to the metastasis of colon carcinoma cells. The expression pattern of E-selectin also governs metastasis. In a mouse model in which E-selectin was overexpressed in the liver, melanoma cells expressing E-selectin ligands formed liver metastasis but not tumor cells lacking E-selectin ligands.

In summary, altered glycosylation is a universal feature of malignant transformation and tumor progression. The highly selective glycosylation changes have the greatest functional consequences in the metastatic cascade. Exploiting these differences would likely afford excellent opportunities to identify sensitive and specific cancer biomarkers, as well as uncover potential new cancer therapeutic targets.

1.4 Glycoproteomics in Cancer Biomarker Discovery

Cancer related changes in glycosylation can lead to biomarker discoveries. In fact, many of the clinical tests currently used to diagnose and manage the treatment of cancer
have exploited changes in glycosylation that accompany the disease process. **Table 3** lists some of the FDA approved cancer biomarkers that are glycosylated. These include many glycoproteins, such as carcinoembryonic antigen (CEA), commonly used as a marker of colon cancer \(^{145}\); cancer antigen 125 (CA-125), frequently used to diagnose ovarian cancer \(^{146}\); \(\alpha\)-fetaprotein (AFP), usually used in screening and diagnosing liver cancer \(^{147}\); and PSA, a marker for prostate cancer as described earlier. Other clinically useful cancer biomarkers directly detect glycan epitopes. One of the examples is CA19-9 which binds to sialy-Le\(^a\) and commonly used for diagnosing and monitoring pancreatic cancer \(^{148}\). Despite the fact that these serum biomarkers remain the diagnostic gold standard in clinic, their diagnostic value is more frequently questioned, due to poor sensitivity and specificity of the assays \(^{27,145,147-149}\). Alternatively, researchers have found capitalizing upon changes in cancer-related glycoforms affords the possibility of increasing diagnostic sensitivity and specificity. The separation of serum AFP by lectin LCA highlights this point. Based on the binding capacity of AFP towards LCA, AFP has three different glycoforms with different diagnostic function: AFP-L1, AFP-L2 and AFP-L3. While AFP-L1, known as the non-LCA-binding fraction, is the most abundant isoform of AFP in serum of patients with benign liver disease, chronic hepatitis and liver cirrhosis; AFP-L2 has intermediate affinity to LCA and is present in maternal serum during pregnancy and frequently derived from yolk sac tumors. AFP-L3 or the LCA-reactive AFP glycoform, carrying an additional \(\alpha1-6\) fucose residue, appears to be produced by cancer cells. It is the major glycoform in serum of HCC patients and a marker for biologic malignancy of HCC \(^{147}\). Overall, detection of these glycoprotein
Table 3 Some of the FDA-approved cancer glycobiomarkers

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Assay type</th>
<th>Source</th>
<th>Cancer type</th>
<th>Clinical use</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-Fetoprotein</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Nonseminomatous testicular</td>
<td>Staging</td>
</tr>
<tr>
<td>Human chorionic gonadotropin-(\beta)</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Testicular</td>
<td>Staging</td>
</tr>
<tr>
<td>CA19-9</td>
<td>Carbohydrate</td>
<td>Serum</td>
<td>Pancreatic</td>
<td>Monitoring</td>
</tr>
<tr>
<td>CA125</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Ovarian</td>
<td>Monitoring</td>
</tr>
<tr>
<td>CEA</td>
<td>Protein</td>
<td>Serum</td>
<td>Colon</td>
<td>Monitoring</td>
</tr>
<tr>
<td>Epidermal growth factor receptor</td>
<td>Protein</td>
<td>Colon</td>
<td>Colon</td>
<td>Therapy selection</td>
</tr>
<tr>
<td>KIT</td>
<td>Protein (IHC)</td>
<td>Tissue</td>
<td>Gastrointestinal stromal tumor</td>
<td>Diagnosis/Therapy selection</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Protein</td>
<td>Serum</td>
<td>Thyroid</td>
<td>Monitoring</td>
</tr>
<tr>
<td>PSA</td>
<td>Protein</td>
<td>Serum</td>
<td>Prostate</td>
<td>Screening/Monitoring/Diagnosis</td>
</tr>
<tr>
<td>CA15-3</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Breast</td>
<td>Monitoring</td>
</tr>
<tr>
<td>CA27-29</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Breast</td>
<td>Monitoring</td>
</tr>
<tr>
<td>HER2/NEU</td>
<td>Protein (IHC), Protein</td>
<td>Tissue, Serum</td>
<td>Breast</td>
<td>Prognosis/Therapy selection/Monitoring</td>
</tr>
<tr>
<td>Fibrin/FDP</td>
<td>Protein</td>
<td>Urine</td>
<td>Bladder</td>
<td>Monitoring</td>
</tr>
<tr>
<td>BTA</td>
<td>Protein</td>
<td>Urine</td>
<td>Bladder</td>
<td>Monitoring</td>
</tr>
<tr>
<td>CEA and HMW-mucin</td>
<td>Protein</td>
<td>Urine</td>
<td>Bladder</td>
<td>Monitoring</td>
</tr>
</tbody>
</table>

BTA, bladder tumor-associated antigen; CA, cancer antigen; CEA, carcinoembryonic antigen; FDP, fibrin degradation protein; IHC, immunohistochemistry; PSA, prostate-specific antigen
glycoform changes is nearly all at the cytological and histological levels, as shown with lectins and glycoreactive antibodies, the availability of which is limited. Thus, sensitive, fast, and robust analytical methodologies are required to identify cancer-specific glycobiomarkers.

The complex and heterogeneous nature of carbohydrate structure and synthesis has constrained the pace of glycosylation study. Traditional analytical techniques including capillary electrophoresis, gel filtration, high-pressure liquid chromatography, and thin layer chromatography are time consuming and most have only been applied to a single glycoprotein. With recent advances in proteomics, chemical biology, and mass spectrometric instrumentation, glycoproteomics is rapidly emerging as a critical sub-field of proteomics. Mass spectrometry (MS), due to its high sensitivity and versatility, has been a state-of-art technology in glycoproteomics. While the knowledge of the mass of a glycoprotein or glycan provides initial information regarding composition, the fragmentation ions of glycopeptides and glycans can determine structural sequence directly. There are two basic levels of glycoproteomics analysis by MS: (1) examination of the glycoproteome (glycoproteomics), which includes identification of glycoproteins and mapping glycosylation sites; (2) survey of the glycome (glycomics), which is concerned characterizing glycan repertoires released from glycoproteins.

Protein biomarker discovery is a complex and challenging task. Five conceptual phases of biomarker development have been proposed: preclinical exploratory (I), clinical assay and validation (II), retrospective longitudinal repository studies (III), prospective screening studies (IV) and cancer control studies (V). The conventional pipeline for the preclinical exploratory phase involves four essential stages: initial
discovery of differentially abundant proteins; prioritization of qualified candidates; candidate verification; assay development and pre-clinical validation. MS-based glycoproteomics is now playing a principle role in the initial phase of discovering cancer-related glycobiomarkers. Glycoprotein enrichment techniques, coupled with and high-resolution MS have greatly increased the analytical sensitivity to reach the depth of glycobiomarker coverage in complex samples such as plasma, serum, other bodily fluids, tissue lysate or cell line homogenates. In addition, with the current drive towards larger panels of multiplexed biomarkers, candidate-based quantitative glycoproteomics platforms, such as multiple reaction monitoring (MRM), have been introduced recently. The new quantitative MS approaches perform full biomarker discovery, quantification, and verification on the same technology platform, allowing targeted detection of glycoprotein candidates in complex samples in a multiplexed fashion. Furthermore, the new approaches have greatly enhanced both the specificity and sensitivity to the point of rivaling immunoassay for some analytes, providing a complementary tool for glycoprotein biomarker verification in addition to antibody based approaches. It is clear that glycoproteomics is playing an increasingly important role in biomarker discovery.

1.5 MS-based Glycoproteomics Approaches

A comprehensive analysis of glycosylation in a complex biological sample requires a concerted approach. Technology platforms incorporating MS for glycobiomarker discovery typically consists of glycoprotein or glycopeptide enrichment, multidimensional protein or peptide separation, tandem MS analysis and bioinformatic data
interpretation (Fig. 4)\textsuperscript{153}. A bottom-up approach (the analytes introduced into the mass spectrometer are peptides) is the most mature and most widely used approach in proteomics, most glycoproteomics workflows adapt it into the platform. Proteolytic digestion can be performed before or after enrichment step, with different rationales for each strategy. Major advantages of the former, enrichment at glycopeptide level, include a lower possibility of nonspecific protein-protein interactions, a particular concern about glycoprotein capture efficiency\textsuperscript{157}. In contrast, enrichment at glycoprotein level can increase the sequence coverage from multiple peptides, which in turn facilitates glycoprotein identification\textsuperscript{158}. The separation procedures for glycoproteins or glycopeptides closely parallel the methodologies of mainstream proteomics. For glycoproteins, various fractional approaches, such as gel electrophoresis, multiple forms of chromatography (size exclusion, hydrophobic interaction, ion exchange, affinity chromatography), and chemical immobilization can be applied, depending on the specific workflow. For glycopeptides, multidimensional liquid chromatography (LC) and capillary electrophoresis (CE) favor their separation\textsuperscript{92, 159}. It is notable that the reliable analysis of MS data in glycoproteomic studies largely relies on available bioinformatics tools and glyco-related databases. The well-built proteomics softwares and databases allow the fully automatic assignment of fragment spectra of unknown peptides. Once protein sequence is identified, glycosylation status of the targeted proteins can be annotated by UniProt Knowledgebase (UniProtKB) database. The high-throughput tools for glycan characterization are evolving, and there are several software platforms such as GlycoMod, Cartoonist, GlycoPep DB, Peptoonist, and GlycoMiner that aid in the identification of glycans\textsuperscript{92}. Bioinformatic tools for direct glycopeptide analysis,
Figure 4 The workflows of mass spectrometry-based glycoproteomic analysis. The pipeline parallels the mainstream of proteomic analysis, typically consisting of four steps: glycoprotein or glycopeptide enrichment, glycoprotein or glycopeptide separation, tandem mass spectrometric analysis, and bioinformatic data interpretation. After enrichment step, the workflow can split into three directions: intact glycoprotein or glycopeptide analysis, glycan analysis and deglycosylated protein analysis. Adapted from Pan, S. et al. Mol Cell Proteomics. (2010).
however, are still primitive and require manual assignment and interpretation of the spectra. Data evaluation is therefore the bottleneck in this field. Looking back to the glycoproteomic pipelines, however, the key stages developed in recent years are the isolation of glycoproteins/glycopeptides and MS analysis of glycoproteomes.

1.5.1 Enrichment of glycoproteomes

Most mass spectrometers can only analyze samples in which the concentrations of protein cover three to four orders of magnitude, whereas the protein concentrations of biological samples span up to 10 to 12 orders of magnitude. Naturally occurring glycoproteins are often present at low levels. In a complex protein digest, the signal intensities of glycopeptides are relatively low since the extensive microheterogeneity of the carbohydrate gives rise to different glycoforms, each present at a low concentration in the total peptide pool. Moreover, glycopeptide signals generated by MS are often suppressed in the presence of other peptides, especially when the glycans are terminated with a negatively charged sialic acid moiety. Therefore, it is essential to enrich glycoproteins present in complex biological samples prior to MS analysis. Two main specific approaches for glycoprotein/glycopeptide enrichment have been widely applied: lectin affinity based enrichment methods and chemical-based methods.

1.5.1.1 Lectin affinity chromatography

Lectins are a class of proteins isolated from plants, fungi, bacteria, and animals that have a unique affinity towards carbohydrates. Lectin affinity enrichment is based on the specific and reversible binding interaction between a lectin and a distinct glycan
structure attached on a glycoprotein/glycopeptide. A variety of lectins can selectively bind to different oligosaccharide moieties. For example, concanavalin A (ConA) binds to high mannose, hybrid and biantennary complex-type N-glycans; wheat germ agglutinin (WGA) binds to N-acetyl-glucosamine and sialic acid; PNA is specific to T-antigen found commonly in O-glycans, and jacalin (JAC) specifically recognizes galactosyl (β→1,3) N-acetylgalactosamine and O-linked glycans. Lectin affinity enrichment has been applied to enrich glycoproteins/glycopeptides with specific glycan attachment from plasma, serum, urine, tissue and other biological samples of human and mice through affinity chromatography and other methods. The typical lectin affinity chromatography protocol involves immobilization of lectins onto appropriate matrices such as agarose, silica or magnetic beads in a number of chromatographic formats including tubes, columns and microfluidic channels, whereby protein sample or trypsin-digested peptides are loaded by gravity-flow mode or low-pressure pumps. The unbound material is removed by washing the column while the captured glycoproteins/glycopeptides are displaced from the column by elution buffer and identified by MS/MS. Multiple lectin species with distinct specificities can also be used in combination or in series to achieve a comprehensive enrichment of glycoproteins/glycopeptides from complex biological mixtures. For instance, our group analyzed human plasma glycoproteins captured by various lectins and found a difference between sera from subjects with benign prostatic hyperplasia and prostate cancer or between sera from hepatocellular carcinoma patients and control subjects. Likewise, Hancock and co-workers developed a multilectin affinity column which combines ConA, WGA and Jacalin to perform a comparative glycoproteome analysis of
sera from breast cancer patients and normal controls. The authors identified 813 glycoproteins in the serum samples including low-abundance components such as neuropilin-1 and pregnancy zone protein, and a number of proteins associated with lipid transport and cell growth were detected of changes in expression \(^{167}\). In related work, Qiu and Regnier utilized an extended strategy called serial lectin affinity chromatography (SLAC) for fractionation and comparison of glycan site heterogeneity on glycoproteins derived from human serum through a series of different lectins with precisely elucidated binding specificities \(^{170}\). To date, most of the work using lectin affinity for targeted glycoprotein enrichment has been limited to N-glycosylation because the binding specificity of lectin for O-glycosylation is less satisfactory. To overcome such caveat, efforts have been made using serial lectin columns of Con A and Jacalin in tandem to remove N-glycosylated species from a complex mixture with Con A before Jacalin selection, which greatly improves O-glycopeptide selection specificity and allow the identification of a broad variety of O-glycosylated proteins from human serum \(^{171}\).

Lectin-based affinity selection has successfully purified glycoproteins/glycopeptides with specific structures and identified some cancer biomarker candidates. Numerous advantages of the lectin affinity approach including its simplicity, cost-effectiveness, and capability of discrimination of glycan structures among different proteins and different glycoforms of the same protein. The weakness of this method does exist, such as the low binding affinity between oligosaccharides and most lectins (the affinity constant \(K_a\) in the low micromolar range), broad binding specificity for certain lectins, and the fact that the binding capacities of lectins are affected by the tertiary structure of glycoconjugate.

These disadvantages could result in some of the proteins/peptides captured by the lectins
actually have no designated sugar and linkage that is ascribed to lectin affinity. In this case, further validation of the oligosacchride structure of candidate proteins by other methods will be required.

1.5.1.2 Chemical-based Methods

Alternative approach for glycosylation characterization stems from the exploitation of glycan chemical reactivity. Two similar approaches, including hydrazide and boronic acid chemistry, capitalize on the cis-diols present in monosaccharides.

Aebersold and colleagues first described the use of hydrazide chemistry for the N-linked glycoproteome examination. In this scheme, carbohydrate cis-diols are converted to aldehydes by sodium periodate treatment, which sequentially react with the immobilized hydrazide groups. Bound glycoproteins are trypsin-digested in situ, and nonglycosylated/unbound peptides are removed by washing. Covalently coupled N-linked glycoproteins are then released by PNGaseF treatment and analyzed by LC-MS/MS. Using this approach, the author identified 145 glycopeptides from 57 unique proteins. The same technique was applied to examine the changes in the N-linked glycoproteome profile of sera from cancer-bearing and genetically identical normal mice. As a result, glycoproteome profiles of mouse sera from mice with cancer formed a distinct cluster from the profiles of normal mice as examined by unsupervised hierarchical clustering. Interestingly, when the entire proteome was examined, clustering analysis did not distinguish cancerous from non-cancerous sera samples, indicating the potential of glycoproteomics for cancer biomarker discovery. To date, hydrazide-based glycoprotein enrichment has been applied to various biological samples including cell
lines, body fluids, and tissues. For example, Soltermann et al. used the method to capture the glycoproteins from malignant pleural effusions of patients with lung cancer and controls. They were able to analyze the glycoproteins with low protein abundance (μg/mL to ng/mL) and identified several proteins associated with tumor progression or metastasis, e.g., CD44, CD166, CA-125, and lysosome-associated membrane glycoprotein 2 (LAMP-2). In 2007 the Aebersold group published a revised version of this method that captured glycopeptides rather than intact glycoproteins, resulting in a claim of an improved glycopeptide recovery (>90%). Later the same group further modified the protocol by substituting hydrazide functionalized beads with biotinylated hydrazide, and employing streptavidin beads to capture extracellular N-glycoproteins on live cells.

By integrating of quantitative proteomic workflows such as stable isotope labeling by amino acids in cell culture (SILAC), the group demonstrated the ability to monitor abundant changes of the cell surface glycoproteome between related cell types and upon perturbation of a specific cell type. The main advantage of hydrazide approaches is the high specificity. The method simplifies the peptide mixture by only isolating and analyzing on average one to two deglycosylated peptides from each glycoprotein. This character also raised issues such as low sequence coverage and potential false identification for glycoprotein. Another drawback of hydrazide methods is limited to only the isolation of broad classes of N-linked glycoproteins/glycopeptides regardless of the glycan structure, therefore, lacking the specificity and selectivity required for analysis of cancer-related glycosylation changes.

Recent studies have compared lectin affinity and hydrazide chemistry methods for performance and efficiency in isolating glycoproteins/glycopeptides from a complex
biological sample \textsuperscript{179-181}. The two methods are complementary in enriching glycoproteins due to their different mechanisms and the structural and complexity of a specific protein. When both methods were applied, it significantly improves the coverage of the glycoproteome, resulting in an increased number of glycoproteins identified.

The second chemical-based method takes advantage of the principle that heterocyclic boronic diesters formed by the reaction of geminal diol groups in glycans with boronic acid are stable under basic conditions. Using this method, Sparbier \textit{et al.} recently showed boronic-acid-functionalized magnetic beads could be used to capture low-abundance glycoproteins from human serum \textsuperscript{182}. In addition, Xu \textit{et al.} synthesized a novel diboronic acid functionalized mesoporous silica (FDU-12-GA) and successfully applied it for specific glycopeptide enrichment \textsuperscript{183}.

\textbf{1.5.2 MS Analysis of Glycoproteomes}

Central to all glycoproteomic strategies is the MS analysis of glycopeptides generated by the digestion of glycoproteins. The complete analysis of glycosylation includes the determination of the peptide sequence, the carbohydrate structure (sequence, branching points, and heterogeneity), and the sites of attachment to the protein backbone. However, it is technically challenging to simultaneously perform such comprehensive analysis in a complex biological sample. First, compared to peptides, glycans exhibit poorer ionization efficiencies. Second, glycosidic bonds are highly labile under conventional fragmentation condition so that different instrument parameters are required for peptide backbone and glycan analysis \textsuperscript{160}. For example, quadrupole-TOF MS/MS has been widely used for analysis of N-glycopeptides. With low collision-induced
dissociation (CID) fragmentation energy, B-type and Y-type fragmentation ions of the glycan moiety prevails and information on the composition and sequence of carbohydrates is predominantly achieved (Fig. 5A). At higher collision energy, glycan ions are hardly observed any more. Instead, a series of y- and b-type peptide ions dominate the MS/MS spectra (Fig. 5B). Third, mixed fragmentation ion spectra from the peptide backbone, the carbohydrate group and the combinations of both, complicates the interpretation and may be sometimes misleading. As a standard setup, most workflows employ deglycosylation before MS analysis and splits glycosylation study into two directions: glycan analysis and peptide analysis (Fig. 4). To define the total glycan repertoire of the samples, specific released glycan populations are derivatised, fractionated, and analyzed by MS, as described below. For peptide analysis, a standard proteomic pipeline is applied.

1.5.2.1 Glycan release from the glycoproteins/glycopeptides

Glycans can be released from glycoprotein/glycopeptide by using enzymatic and chemical methods. Each procedure has certain advantages and disadvantages. Enzymatic release is straightforward, but depending on the substrate specificity not all types of glycans are liberated. Chemical methods, can suffer from several limitations, such as low specificity, destruction of non-carbohydrate substituents and polypeptide backbone, and modifications of the amino acid residues. Hence, some glycosylation information (extent and sites) is lost.
Figure 5 The nomenclature of peptide and glycan fragment ions from a tandem mass spectrum. (A) Peptide fragmentation notation. Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either a, b or c. If the charge is retained on the C terminal, the ion type is either x, y or z. A subscript indicates the number of residues in the fragment. (B) Glycan fragmentation nomenclature. In this nomenclature, ions retaining the charge on the non-reducing terminus are named A, B and C, and the ions retaining charge on the reducing terminus are X, Y and Z. A and X correspond to cross-ring cleavages, whereas B, C, Y and Z correspond to glycosidic cleavages. Subscript numbers denote the cleavage position, starting at the reducing terminus for the X, Y and Z ions, and at the non-reducing terminus for the others. In the case of ring cleavages, superscript numbers are given to show the cleaved bonds. Adapted from Morelle, W. et al. Nat Protocols 2, 1585 (2007).
Intact N-glycans are usually released using enzymatic methods. Several enzymes are commercially available. The most popular is peptide N-glycosidase F (PNGase F), which cleaves the linkage between the core GlcNAc and the asparagine residues of all classes of N-glycans, with the exception of N-glycans containing a (α1→3)-linked fucose attached to the core GlcNAc residue $^{185}$. PNGase A, an enzyme found in almond emulsion, is sensitive to such PNGase F-resistant glycans $^{186}$. Two other enzymes for N-glycans cleave the glycosidic bond between the two GlcNAc residues within the chitobiose core, differ in their substrate specificities: endoglycosidase D (endo D) releases all classes of N-glycans, while endo H selectively cleaves oligomannose- and hybrid-type structures $^{87}$.

Since O-glycans lack a common core and may contain extended sugar chains, the enzymes are quite limited for their release. So far, only one enzyme, endo-α-N-acetylgalactosaminidase (O-glycanase) targeted to the core 1 O-glycan structure, has been described $^{187}$. Most assays rely on chemical methods, such as trifluoromethansulfonic acid, hydrazinolysis, β-elimination, and periodate oxidation $^{188}$. Among these, β-elimination has been widely adopted. Reductive β-elimination uses NaOH to cleave the glycosidic bond and NaBH₄ to reduce the free reducing end of the released glycan to an alditol form $^{189}$. This procedure has even been applied to glycoproteins separated on SDS-PAGE or blotted to PVDF membranes $^{190}$. However, the degradation of glycan at the high pH remains a problem. Another drawback is the reduction conversion prevents downstream glycan labeling and derivatization. A newly introduced method is ammonia-based non-reductive β-elimination. This method avoids these drawbacks and becomes the favored choice of many glycan analysis $^{191, 192}$.
1.5.2.2 Mapping glycosylation sites

In nature, not all potential glycosylation sites are actually occupied, and this characteristic (carbohydrate macroheterogeneity) also correlates the disease states. Since direct intact glycopeptide analysis is difficult, one way around this problem is to remove the sugar chains and concomitantly incorporate a MS-amenable tag at the glycosylation sites. This strategy is particularly suitable for mapping N-glycosylation sites because PNGase F not only cleaves all types of N-glycans, but also possesses additional amidase activity during this process. Therefore, PNGaseF converts asparagine to aspartic acid during the cleavage reaction. This results in a 1-Da mass shift detectable by high resolution mass spectrometers. Stable isotope labeling with $^{18}$O during deglycosylation can further enhance the precise of site identification. The study of Kaji et al. represents one of the earliest efforts to apply this strategy to glycosylation analysis. The approach, termed isotope-coded glycosylation-site-specific tagging (IGOT), involves the capture of glycopeptides by lectin affinity chromatograph, followed by PNGase F mediated incorporation of $^{18}$O into the N-glycosylation site and LC-MS/MS analysis. Using IGOT, Kaji et al. identified 250 glycoproteins containing 400 unique N-glycosylation sites. In a similar manner, Lewandrowski et al. coupled glycopeptides enrichment by lectin affinity chromatography and hydrazide beads with PNGase F deglycosylation to elucidate N-glycosylation sites on human platelet proteins. Over 70 glycosylation sites from 41 unique proteins were identified, and the majority had not been found in previous studies.

Deglycosylation can also be used to determine O-glycosylation sites. A method, called BEMAD- β-elimination followed by Michael addition with dithiothreitol (DTT) or
biotin pentylamine (BAP) yielded a unique tag on O-GlcNAc sites, allowing for enrichment via affinity chromatography as well as O-GlcNAc site identification by LC-MS/MS. Three previously known and three new O-GlcNAc sites on Synapsin 1 were identified by this strategy. Vosseller et al. used BEMAD in a quantitative manner, where two samples were compared by using deuterated (d6) DTT for one sample and unlabeled (d0) DTT for another sample. Combining thiol chromatography and LC-MS/MS, the team could identify the BEMAD-labeled peptides, determine the sites of modification, and perform comparative quantitation of large number of O-GlcNAc proteins in one study. However, BEMAD is not specific for O-GlcNAc glycosylation. Any modification of serines and threonines that is susceptible to β-elimination, such as O-phosphorylation and O-sulfation, may be targeted by this method. Therefore, extra experiments are required to distinguish between O-glycosylation and other types of modification.

1.5.2.3 Glycomics Methods

Various MS techniques can be employed to decipher glycan structures. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) and electrospray ionization (ESI) are the two major ionization techniques used for glycan analysis. MALDI-MS, usually with a time-of-flight (TOF) analyzer, can be used for monosaccharide composition analysis, in which predominantly singly charged ions are formed with very little fragmentation. The major advantage of MALDI-TOF-MS is the relative simplicity of instrumentation, the speed of analysis, the high mass range and the robustness towards low concentrations of contaminants. Together with automated
sample spotting, MALDI-TOF provides a valuable tool for high-throughout glycan profiling. However, this type of analysis only provides a snapshot of the most likely structures to be present, but does not allow the assignment of defined carbohydrate structures. ESI-MS is also applied to analysis of glycan mixtures. The advantage of ESI-MS over MALDI-MS is that ESI-MS can be coupled with glycan separation column online to enhance the detection sensitivity. But ESI-MS usually generates doubly or triply charged ions, thus leading to high complex mass spectra when heterogeneous mixtures of glycans are analyzed.

Monosaccharide sequence, branching, and linkages can be determined through fragmentation. With CID function, ESI-MS/MS, and MALDI-TOF/TOF are currently the most frequently used methods, and both can be performed in the positive or in the negative ion mode. In general, glycans show two types of fragmentation: cleavage of the glycosidic bonds between the sugar rings and cleavage within the rings, so-called cross-ring cleavages (Fig. 5B). Glycosidic cleavage ions provide information on carbohydrate sequence and branching as well as non-carbohydrate substitutes, whereas the cross-ring cleavage ions are usually weak and may be used to define linkage.

Recently, MS*-based analytical methods and new fragmentation technologies have been introduced to enhance the identification of structural isomers. In the case of MS* in iron trap, sequential fragmentation of glycan was found to result in cross-ring cleavages that are highly informative in the determination of linkage position and anomeric configuration identification.

Fragmentation analysis may be performed with underivatized or derivatized oligosaccharides. To ensure the highest sensitivity and to facilitate unambiguous
sequencing in MS/MS experiments, glycans are usually derivatized by permethylation, by methylesterification of sialic acids, or by reducing-end tagging before MS analysis\textsuperscript{203}. Permethyla-
tion, which generates methyl derivatives for each free OH and NH group and esterifies the carboxy group in the sialyl residues, often prevails in glycomic analysis (especially in MALDI-MS analysis) with some benefits. First, the addition of methyl groups stabilizes sialylated structures and enhances overall sensitivity. Second, the hydrophobic nature of methylated glycans facilitates their separation from salts and other impurities that may affect the MS analysis. Third, the fragmentation of methylated glycans is more predictable than that of their native counterparts, leading to accurate structural assignments\textsuperscript{204}. A recent multi-institutional study by the Human Proteome Organization revealed that MALDI-TOF-MS of permethylated N-glycans is consistent and comparable to HPLC analysis of reductive amination derivatives for quantitation\textsuperscript{205}.

1.5.2.4 Site-specific glycosylation analysis

One of the disadvantages of glycoproteomic analysis involving glycan release is that this strategy does not allow site-specific glycosylation profiling. For this purpose, the direct analysis of the intact glycopeptides is required. As described earlier, CID ESI-MS/MS, the most frequently employed technique, prefers to detect fragmentation ions of the glycan moiety other than that of peptide. This limitation can be overcome by using modern high-speed tandem MS (or MS\textsuperscript{n}) technologies. So far, a variety of instruments with MS\textsuperscript{n} capacity have been introduced, including ESI based ion trap (IT), quadrupole ion trap (QIT), Fourier transform ion cyclotron resonance (FT-ICR), ion trap/time-of-
flight (IT/TOF) and quadrupole/time-of-flight (Q/TOF); MALDI based Q/TOF, quadrupole ion trap/time-offlight (QIT/TOF) and tandem time-of-flight (TOF/TOF) mass spectrometers. However, different instruments require pronounced distinct fragmentation conditions, and provide different structural information for glycopeptides. For instance, ESI-IT-MS, which allows repetitive ion isolation / CID fragmentation cycles, has been recently shown to characterize both peptide and glycan in a single measurement. In this approach, the glycan fragment ions are collected at the first cycle (MS² stage), and peptide fragment ions carrying a single sugar are selected and subjected to a second ion isolation/fragmentation, providing information on the peptide sequence and glycosylation site. This technology has successfully applied to determine the three glycosylation sites and the glycoform structures in Thy-1, a GPI-anchored protein, and determined the structures of neutral and sialylated N-glycans attached to chicken egg yolk glycopeptides.

The invention of new fragmentation technologies of electron capture dissociation (ECD) in FT-ICR-MS, can also facilitate the single amino acid localization of labile modifications. ECD produces cleavage specifically at N-αC backbone of the peptide leaving the modification intact on the peptide backbone, thus allowing unambiguous assignment of glycosylation sites in N- or O-glycopeptides. In quadrupole ion traps, a similar type of fragmentation method to ECD, electron transfer dissociation (ETD), has been recently introduced. ETD uses the same principle as ECD and can generate the cleavage of peptide backbone while preserving labile modifications. Together ECD/ETD with CID complementary information can be thus obtained. Hogan et al. demonstrate the power of combination of CID and ETD in a 3D-quadrupole IT for the structural
characterization of a tryptic N-glycopeptide. CID experiments on the multiply protonated glycopeptide ions yielded, almost exclusively, cleavage at glycosidic bonds and therefore glycan structure information. ETD reactions resulted in cleavage of the peptide backbone with no loss of the glycan structure. Although the application of this complementary multistage fragmentation technique has been limited to model glycopeptides, it holds promise for complete chemical structural determination of glycopeptides in complex biological mixtures.

1.5.3 Concluding remarks

In the past decade, MS in combination with modern enrichment methodologies has become one of the most powerful and versatile techniques for glycosylation analysis. There are different strategies for characterization of glycosylation at the levels of intact proteins, glycopeptides, and free glycans. Therefore, MS-based glycoproteome analysis is a multidimensional task. There is no universal method for rapid and reliable interrogation of the global glycoproteome. Research goals must dictate the best method or combination of methods for analysis. Sometimes other non-MS-based approaches such as HPLC glycan analysis, exoglycosidase digestion and chemical treatment are necessary for full structural elucidation of glycoproteome.

Looking to the future, we envision that a large-scale, automatic, quantitative glycoproteomic analysis of complex biological samples is on the way, with the continued evolution in mass spectrometry, enrichment approach, separation technology and bioinformatics. There is no doubt that glycoproteomics is gaining momentum in biomarker research and clinical studies.
CHAPTER II

DISSERTATION RATIONALE AND SUMMARY OF AIMS

The objective of this research project is to develop a mass spectrometry-based glycoproteomic approach for the selective isolation and identification of cell-surface glycoproteins from cellular samples. This technology is applicable in the discovery of new glycoprotein biomarkers which are indicative of prostate cancer progression and metastasis.

Prostate cancer is the most prevalent cancer and second leading cause of cancer deaths in men in the USA. Its lethality is mostly attributed to the primary tumor metastasizing to distant sites. Cancer metastasis involves a series of sequential steps consisting of tumor invasion at the site of origin, followed by intravasation, survival in the bloodstream or lymphatic glands, arrest in a distal organ, extravasation, and finally growth and angiogenesis in the distal organ. The steps requiring interactions between tumor cells and the surrounding microenvironment are tightly regulated by cell surface mechanisms. Most membrane and secretory proteins are glycosylated. Glycans not only affect properties of the proteins but also regulate diverse biological functions through particular protein-carbohydrate interaction. At the cell surface, sugar moieties of glycoproteins contribute to molecular recognition events involved in cancer metastasis. Aberrant glycosylation of cell-surface glycoproteins have been repeatedly reported in a wide variety of cancers, including prostate cancer, and are strongly correlated with cancer prognosis, metastasis, and the organ specificity of metastasis. It is possible that a portion of these alternatively glycosylated molecules can reach the bloodstream and serve as
early sentinels that enable cancer detection. In fact, many of the oldest and most widely used clinical cancer biomarkers are based on glycoproteins or glycans. It is clear that glycoproteins become important targets for the development of cancer biomarkers. With recent advances in carbohydrate chemistry, analytical technologies, and mass spectrometric instrumentation, glycoproteomics is rapidly emerging as a sub-field of proteomics with high biological and clinical relevance. By focusing on the glycoprotein subproteome, glycoproteomic analysis greatly reduces sample complexity, which significantly increases the detection sensitivity for low abundant proteins. Therefore, the central hypothesis of this dissertation is that cell-surface glycoproteins with aberrant glycosylation, mediating the key events during prostate cancer metastasis, can be indentified using mass spectrometry-based glycoproteomics. An improved understanding of metastasis-associated cell-surface glycoproteins may not only provide diagnostic and prognostic indicators, but also enable the discovery of new therapeutic targets for prostate cancer. The central hypothesis will be tested and our overall research objective will be achieved by addressing the following specific aims:

Aim 1. Characterization of cell-surface lectin-binding patterns associated with metastasis in human prostate cancer cells. This aim entails:

A. Optimization of lectin staining protocols to cultured live cells. This includes finding the suitable cell dissociation methods which preserve both cell viability and cell-surface carbohydrates, and determining optimal staining conditions that allows the reliable quantitation of lectin binding.

B. A syngeneic PCa cell line model, which consists of human prostate cancer cell line PC3 and its two syngeneic variants with different metastatic potentials, will be profiled
by flow cytometry using a panel of lectins with different carbohydrate specificities to delineate the carbohydrate contents on their surface and identify different glycosylation patterns associated with metastasis.

C. The cells described above will be stained by lectin cytochemistry and observed under confocal microscopy to evaluate the distribution of the existing sugars.

**Aim II.** Targeted identification of metastasis-associated cell surface sialoglycoproteins in prostate cancer cell lines. This aim entails:

A. Development of a targeted glycoproteomic methodology which combines advanced carbohydrate labeling technologies with high performance mass spectrometry to selectively identify cell-surface glycoproteins.

B. Examining the sensitivity and specificity of cell-surface carbohydrate labeling by flow cytometry and confocal microscopy.

C. Optimization of carbohydrate labeling protocol in order to maximize carbohydrate labeling performance and maintain cell viability.

D. Evaluating the efficiency of glycoprotein capture and enrichment by our targeted glycoproteomic approach using in-blot streptavidin detection and 1D-PAGE. The efficiency will be assessed by comparing the number and intensity of protein bands before *versus* after enrichment and labeling *vs* control groups.

E. Application of the optimized glycoproteomic approach to our prostate cancer cell line model system and analysis of the recovered glycoproteins by LC-MS/MS in an attempt to identify panels of cell-surface glycoproteins as potential PCa metastatic
biomarkers. The identified proteins will be analyzed by bioinformation tools for their subcellular location and glycosylation status.

F. Evaluating the utility of this methodology for glycoprotein isolation and identification from a mass spectrometric perspective. It will be completed by comparing glycoproteins identified by our targeted glycoproteomic with those from total cell lysate proteome as well as other publications.

G. Functionally mapping differentially expressed glycoproteins between PC3 variants. Analysis will be performed by Ingenuity Pathway Analysis software across glycoprotein biomarker lists to prioritize biomarker candidates based on key biological characteristics and to shed light on mechanisms linking potential biomarkers to prostate cancer metastasis.

**Aim III.** Verification of expression of cell-based biomarkers in clinical samples. This aim entails:

A. Confirming the differential expression of glycoproteins identified in Aim II in our cell line models using available antibodies. We will employ several methods for analysis glycoproteins in cell lines. Initially, we will perform Western blot analysis on glycoprotein enriched fractions and whole cell lysates to determine their expression differences. Next, we propose to extend analysis of whole cell lysates to other prostate cancer cell line model systems. We will also examine candidate expression by antibody-based flow cytometry and confocal microscopy.
B. Characterization of glycosylation status of biomarkers in prostate cancer cells. The specific glycosidases and glycosylation inhibitors will be employed to investigate the carbohydrate residues on the biomarkers.

C. Verification of expression of biomarkers in clinical samples. Using a cohort of prostate cancer patient samples including tissues and Expression Prostate Secretion (EPS) urines with different risk factors but no extra-prostatic disease at the time of their prostatectomy, the clinical utility of biomarkers and their expression will be compared by immunohistochemistry/immunocytochemistry and/or Western blot techniques.
CHAPTER III

AIM I: CHARACTERIZATION OF CELL SURFACE LECTIN-BINDING PATTERNS ASSOCIATED WITH METASTASIS IN HUMAN PROSTATE CANCER CELLS

3.1 Introduction

Prostate cancer is the most prevalent cancer and second leading cause of cancer deaths in men in the USA. It affects men after middle age, and accounts for approximately 30% of all male malignancies. In nature PCa is pathologically and clinically heterogeneous ranging from indolent, low grade disease to aggressive high grade disease which frequently metastasizes. Its prognosis rate varies greatly for patients with different clinical stages and pathological grades. Despite early detection through testing for serum PSA and the efficacy of surgery and radiation therapy for treating clinically localized PCa, about 30% of men with PCa will experience relapse and develop a PSA recurrence within 10 years after radical prostatectomy. Metastatic PCa claimed the lives of 28,660 American men in 2008. PCa metastasis has organ tropism. Bubendorf et al. examined metastatic patterns of PCa from a postmortem study of 1,589 patients and reported hematogeneous metastases were present in 35% patients, with the most frequent locations being bone (90%), lung (46%), liver (25%), pleura (21%), and adrenals (13%). The mechanisms responsible for this phenomenon remain largely undefined. Most advanced or metastatic PCa show an initial response to androgen-ablation therapy in up to 80% of cases, but ultimately it progresses to hormone refractory prostate cancer (HRPC) after 18-24 months. At this stage, despite castrate...
levels of testosterone, the tumor will become independent of androgens resulting in death within a few years from diagnosis. Moreover, choices of therapeutic regimens for late stage patients are very limited and the major goal of the treatments is just to improve the quality of life. Eventually, these patients die.\textsuperscript{215,216}

The clinical fate of PCa patients is ultimately determined by the primary tumor's malignant potential, which is controlled by many factors: (i) apoptotic susceptibility of tumor cells; (ii) matrix-destroying protease activity; (iii) adhesion of tumor cells to target cells, particularly to platelets, leukocytes, and microvascular endothelia cells; (iv) the ability to extravasate into the target organ parenchyma; (v) the capacity to evoke angiogenesis in tumor cell surroundings; (vi) cross-talks with the host microenvironment.\textsuperscript{217} The underlined mechanisms of each determinant are functionally maintained by a combination of molecules, each of which can be affected directly or indirectly by N- or O-linked glycosylation.\textsuperscript{123} Glycosylation is critical in determining a protein's stability, conformation, folding, and binding affinity for other molecules.\textsuperscript{85} Moreover, glycans \textit{per se} are adhesion molecules directly involved in cell recognition events through protein-carbohydrate and carbohydrate-carbohydrate interactions.\textsuperscript{218} In mammals, most secreted and membrane proteins are glycosylated. Given their important location and interactive roles in biological and pathological processes, aberrant cell surface glycosylation, including altered sialylation and fucosylation, polylactosaminylation, higher-ordered N-linked branching, and truncated O-linked glycans, is often observed in a variety of tumors, and are strongly correlated with cancer prognosis, metastasis, and the organ specificity of metastasis.\textsuperscript{86,100} Nevertheless, the expression of cell surface carbohydrates is cell-type specific and developmentally regulated by the activity of glycosyltransferases.
Examining the relationship between cell surface oligosaccharide structures and metastatic phenotype variants within a defined tumor-cell population has been mostly done in breast cancer. In PCa, the exact nature of glycosylation changes accompanying metastasis is ill defined. Characterizing such patterns is thus critical both to understanding their role in PCa progression as well as to provide diagnostic tools and therapeutic targets to help clinical PCa management.

Specific carbohydrate-binding ligands such as lectins have proven to be invaluable tools for the analysis of complex glycoconjugates. These proteins, derived from non-immune origin, recognize and bind to specific carbohydrate structural epitopes. In most previous studies, lectins were tagged with a fluorophore or enzyme and used for histochemical analysis of formalin-fixed paraffin embedded (FFPE) tissues. Although such analysis is very powerful, it does suffer from some limitations. First, histological processing of tissues can render carbohydrates inaccessible due to protein denaturation and glycolipid losses. Second, visual assessment of such stained samples is subjective and probably with considerable error, particularly for complex samples with heterogeneous staining patterns. Third, the quantitation across multiple areas of tissues is difficult and fairly lengthy. In contrast, lectin flow cytometry offers several advantages. It can perform on a large number of live cells, and the extent of lectin binding can be expressed in a defined quantitative fashion within a short period.

In Aim I we refined and optimized a lectin staining protocol for prostate cultured cells with emphasis on compatibility with flow cytometric analysis. We then applied this approach to delineate the carbohydrate alterations on a human syngeneic metastatic PCa cell line model by using a battery of lectins with defined specificities. Additionally, we
evaluated the distribution of the existing sugars over PCa cell surface with lectin
cytochemistry.

3.2 Materials and Methods

Materials — Dulbecco's modified Eagle media (DMEM), RPMI 1640 medium,
penicillin, streptomycin, and fetal bovine serum (FBS) were from Invitrogen (Carlsbad,
CA, USA). Antibiotics-antimycotic, propidium iodide, nonenzymatic dissociation buffer,
fluorescein conjugated streptavidin (streptavidin-FITC), 4-12% NuPAGE® Bis-Tris gels,
and lithium dodecyl sulfate (LDS) buffer were from Invitrogen. Halt™ protease inhibitor
cocktail was from Pierce (Rockford, IL, USA). Cell proliferation reagent WST-1 (4-[3-
(4-iodophenyl)-2-(4-nitrophenul)-2H-5-tetrazolio]-1,3-benzene disulfonate) was from
Roche (Indianapolis, IN, USA). Prestained protein standard was from Bio-Rad
(Hercules, CA, USA). Murine anti-α-tublin mAb was provided by Sigma (St Louis, MO,
USA). Murine anti-androgen receptor mAb was from BD PharMingen (San Diego, CA,
USA). Rabbit anti-PSA polyclonal antibody was from Dako (Carpinteria, CA, USA).
IRDye 700 or IRDye 800 conjugated secondary reagents were from Li-COR Biosciences
(Lincoln, PA, USA). All chemicals were of analytic grade and were purchased from
Sigma.

Lectins — Seven biotin- or FITC-conjugated lectins were purchased from EY
Laboratories (San Mateo, CA, USA) or Vector Laboratories (Burlingame, CA, USA).
Common name, abbreviation, source, and normal sugar specificity were as follows:
anguilla anguilla lectin (AAA) from Fresh water eel: fucose in an (α-1,6) linkage;
concanavalin A agglutinin (ConA) from jack bean seeds: \(\alpha\)-mannose/\(\alpha\)-glucose; Helix pomatia lectin (HPA) from *Edible snail*: N-acetyl-galactosamine (GalNAc); limax flavus lectin (LFA) from *Garden slug*: sialic acid; *Maackia amurensis* seeds: sialic acid in an (\(\alpha\)-2,3) linkage; sambucus nigra lectin (SNA) from *elderberry* bark: sialic acid in an (\(\alpha\)-2,6) linkage; wheat germ agglutinin (WGA) from *Triticum vulgaris*: sialyloligosaccharide/N-acetyl-glucosamine (GlcNAc).

*Cell Lines* — PC3-N2 and PC3-ML2, two sublines of PC3 prostate cancer cells (originally obtained from a skeletal metastasis in a patient with primary prostate adenocarcinoma) were kindly provided to us by Dr. Stearns’ group at Drexel University. The N2 and ML2 cell lines have been developed on the basis of their invasiveness *in vitro* and metastatic potential *in vivo*. Both cells were tumorigenic when injected subcutaneously in SCID mice. However, N2 cells were unable to migrate through a Matrigel-coated membrane *in vitro* as well as induce metastases following tail vein injections in SCID mice, whereas ML2 cells were highly invasive *in vitro* and induced skeletal metastases in more than 80% of cases\(^{227}\). Human prostate bone metastasis PC3, prostate carcinoma lymph node metastasis LNCaP, and prostate carcinoma brain metastasis DU145, and HPV-18 immortalized prostate epithelial cell line RWPE-1 were obtained from the American Type Culture Collection (ATCC). PC3 and its sublines were routinely cultured in growth medium (DMEM medium, 10% FBS, 1% Antibiotic-antimycotic) at 37°C in a humidified atmosphere containing 5% carbon dioxide. DU145 and LNCaP cells were cultured in the growth medium and passaged according to the manufacturer’s instructions.
Cell Proliferation Assay — Cells were seeded into 96-well microplates under 200 μl/well complete medium. At the indicated time point, WST-1 reagent was added to 10% per well and incubated for 2h at 37°C in 5% CO₂. The absorbance was measured at 450 nm by ELISA reader (ELx800, Bio-Tek Instruments, Winooski, VT, USA).

Flow Cytometric Assessment of Lectin Binding — Cells were grown in 150-mm dishes to 90% confluence and detached by nonenzyme dissociation buffer at 37°C. After washing, 0.5×10⁶ cell suspension was incubated with lectins in appropriate buffers and at optimized concentrations for 30min at 4°C. Then cells were washed three times with PBS/TBS and resuspended in 0.5 ml PBS/TBS. Before subjected to flow cytometry, they were counterstained with 1 μg/ml PI for a minimum of 20 min at 4°C. For biotin-labeled lectins, streptavidin-FITC staining was applied (4°C, 30 min) before PI staining. A flow cytometer (FACScalibur, BD Biosciences, San Jose, CA) equipped with 488 nm excitation source was used for data collection. The fluorescent signal of FITC and PI was recorded in FL1 (525 nm band pass) and FL2 (585 nm band pass), respectively. Control samples including unstained cells, lectin-stained cells only and PI-stained only from each cell line, were prepared for flow cytometric compensation setting. Flowjo software (Tree Star Inc., Ashland, OR, USA) were employed for data analysis. To address the specificity of lectin binding, the inhibitory sugar was preincubated with lectin overnight and mixture was used for lectin staining.
Lectin Cytochemical Analysis — Cells were seeded onto 6-well plates containing glass coverslips and cultured in growth medium up to 80% confluence. After rinsing with ice cold PBS/TBS three times, cells were incubated with lectins in appropriate buffers and at indicated concentrations for 30 min at room temperature in dark. Then cells were fixed with 2% paraformaldehyde (PFA) in PBS/TBS for 12 min at room temperature. For biotin-labeled lectins, streptavidin-FITC staining was applied for 30 min at 4°C. Finally coverslips were mounted to the slides with VectorShield medium (Vector Labs, Burlingame, CA, USA), and sealed with nail polish. Fluorescent images were examined and captured under a confocal microscope (Carl Zeiss, Thornwood, NY, USA).

Western Blot Analysis — Whole cell lysates were collected in the M-PER lysis buffer containing 1x protease inhibitor cocktail. Protein concentration was measured by the BCA protein assay. Protein samples (20 μg each) were separated by electrophoresis through 4-12% SDS-PAGE and then transferred to Immobilon-FL PVDF membranes (Millipor, Billerica, MA, USA). Prestained protein standard was used to assess transfer efficiency and to determine the size of blotted proteins. Membranes were blocked in LiCor blocking buffer (Rockland Immunochemicals, Gilbertsville, PA, USA) diluted with PBS (1:1), and then incubated with anti-PSA primary antibodies (1:1000) or anti-androgen receptor (1:3,000) overnight at 4°C. Following extensive washing, membranes were incubated with species-specific IRDye700 or 800-conjugated secondary antibodies (goat anti-rabbit, 1:15,000; goat anti-rabbit, 1:15,000) for 1 h at room temperature, and visualized with a LiCor Odyssey infrared imager (LiCor, Lincoln, NE, USA). Consistent
protein loading was determined by reprobing membranes stripped in Restore Western blot stripping buffer with anti-α-tubulin antibody (1:4,000).

3.3 Results

3.3.1 The properties of prostate cancer PC3 cells and its sublines

To study cell surface glycosylation changes accompanying metastasis, we need a suitable model which accurately recreates the histopathological metastatic process observed in humans. Clinically, the lethal phenotypes of human PCa are characterized by their progression to HRPC and their propensity to form osseous metastases. Here we employed a syngeneic PCa cell line model which consists of two sublines derived from the widely used PCa PC3 cell line. PC3, originated from a patient bone lesion, is an androgen-independent cell line. PC3-N2 (non-metastatic) and PC3-ML2 (metastatic) was previously selected for their ability to both invade in vitro and generate skeletal metastases in SCID mice in vivo. To better support our following experiments, we first tested two basic characteristics of PC3 variants: proliferation and androgen dependence. Proliferation of PC3, N2, and ML2 cells was assessed using the cell proliferation reagent WST-1 (Fig. 6A). Compared to parent PC3 cells, N2 cells grow slower and ML2 faster. Immunoblot analysis of total cell lysates shows that these cells don’t express androgen receptor and androgen-regulated gene products such as PSA, therefore verifying that they are androgen-independent cell lines (Fig. 6B).

3.3.2 Optimization of cell dissociation methods
Figure 6 The properties of prostate cancer PC3 cell line and its sublines. (A) WST-1 assessment of cell proliferation of PC3, N2 and ML2 cells. (B) Western blot analysis of AR and PSA expression in PC3, N2, and ML2 total cell lysates. Androgen-dependent cell line LNCaP and androgen refractory cell line DU145, were included as positive and negative controls. α-tublin was used as a loading control.
Since PC3 cells and its sublines are adherent cells, and traditional cell dissociation method, trypsinization, cleaves or strips some cell surface glycoproteins, we need to find the best method to dislodge the cells from the plastic flask surface. Five cell detachment methods (Fig. 7) were compared on ML2 cells. They are enzyme preparation, trypsin-EDTA; mechanical disruption, gentle scraping; and three nonenzymatic reagents, commercial dissociation buffer, EDTA only, and sodium citrate only. The following three parameters were concerned: (1) cell dissociation efficiency; (2) cell viability; (3) cell surface carbohydrate retention. As expected, trypsinization lifted cells very efficiently with only 5-10 min incubation. In contrast, nonenzymatic reagents needed 20-30 min and gentle tapping of the flask to take effect. Cells removed by these three preparations resulted in some cell clumping, but all formed a single cell suspension after repeated pipetting. The integrity conservation of the dissociated cells was compared by flow cytometric analysis (Fig. 7A). Scraping pronouncedly changed cell light scatter properties with more than 50% cell debris generated, while other methods could keep cell physically intact. To assess cell viability, we incubated the dissociated cells with propidium idodide (PI) and discriminated viable cells by flow cytometric analysis. As shown in Figure 7B, trypsinization led to minimal cell death (about 9.5%) whereas cell scraping damaged cells the most (>50% cell death). Three nonenzymatic preparations resulted in different cell viability rate, and commercial buffer had the advantage over EDTA and sodium citrate both in terms of percentage of the main population (intact cells) and the total population (all detached events, including debris). These results agreed with trypan blue exclusion counts shortly after dissociation (data not shown). For cell surface carbohydrate retention, we used four FITC- or biotin-labeled lectins specific
Figure 7 Flow cytometric assessment of different cell dissociation methods. (A) Cell light scatter properties of ML2 cells dissociated by different methods. A polygon delimits the main population of intact cells, whereas events outside of main population consist of aggregated cells and cellular debris. Number indicates main population as a percentage of total events. (B) Cell viability within the main population. Positive PI staining indicates dead cells. Numbers indicate percentages of live (PI-) cells within main population. (C) Effect of dissociation methods on carbohydrate retention. ML2 cells were dissociated by different methods, incubated with FITC- or biotin-conjugated lectins (LFA, SNA, MAL, WGA), followed by streptavidin-FITC (only for biotin-lectins), and analyzed by flow cytometry. PI staining was used to gate live cells. Histogram for each lectin is shown by red line and nonspecific staining (unstained cells) is overlaid in grey line for comparison.
for sialic acid to measure cell surface carbohydrate levels on living cells. Flow cytometry data revealed a homogenous fluorescent staining for all examined lectins on cell surface. Scraping compromised most cell surface carbohydrates and trypsin selectively ate up some sugars such as WGA (Fig. 7C). Sodium citrate also exhibited poor carbohydrate retention. Conversely, commercial dissociation buffer preserved all carbohydrates best followed by EDTA procedure.

On the basis of above results, we concluded that the commercial dissociation buffer provides the best balance of dissociation efficiency, viability, and cell surface carbohydrate retention. The reagent will be used for the following experiments in this study.

3.3.3 Optimization of lectin staining protocols

Because of their extensive availability and the wide spectrum of carbohydrates that may be specifically bound, lectins have become essential reagents for glycosylation analysis. They bind non-covalently to carbohydrate residues and can be used in the same types of application as monoclonal antibodies. However, lectin binding has different characteristics. For example, some lectins at high concentration can induce apoptosis, resulting in global changes of cell surface glycosylation\textsuperscript{228}. In addition, some lectins require divalent cations (Ca\textsuperscript{2+} and Mg\textsuperscript{2+}) to take effect while the binding of other lectins may be interfered by phosphate ions\textsuperscript{229}. On the other side, cell surface glycans are expressed in a cell type and lineage-specific manner\textsuperscript{219}. Therefore, when applying lectin staining to cultured live cells, we need to work out a lectin binding protocol that allows
the reproducible determination of carbohydrate moiety profiles, as well as reliable quantitation of lectin binding.

A blocking reagent contains a high concentration of immunoglobulin and is often used in immunostaining to minimize the non-specific antibody-base staining. But for lectin staining, how to decrease the non-specific binding is largely unknown. Here we compared three different blocking buffers with respect to their preservation of cell viability and SNA binding to non-malignant prostate RWPE-1 epithelium cells (Fig. 8). Although blocking buffers either with 1% FBS or 1% BSA can effectively protect the cell death arisen from the high concentrations of SNA (Fig. 8B), they also significantly decreased cell surface fluorescent signals (Fig. 8A). It is not surprising since serum contains a lot glycoconjugates which can competitively inhibit lectin binding. BSA is not a glycoprotein, but it is often contaminated by serum glycoproteins during its commercial preparations.

Next we investigated the effect of two different buffers, PBS and TBS containing divalent cations Mg$^{2+}$ and Ca$^{2+}$, on lectin binding and cell death. As shown in Figure 9A, the addition of divalent cations induced RWPE-1 cell death as well as loss of MAL binding. Similar results have been obtained for lectins SNA, WGA, and AAA. There is no obvious change for ConA regarding binding signal and cell death between the two buffers (Fig. 9B). Since an improvement in binding was not gained in most cases, the addition of divalent cations is not recommended for most lectins.

As cell surface glycosylation is involved into cell-cell contact, the cell confluence status may influence the expression level of carbohydrate. Figure 10 summarized the relationship of SNA binding and the degree of PC3 cell confluence. In general, cells with
Figure 8 Comparison of different blocking buffers on SNA binding. Prostate epithelium RWPE-1 cells were stained with 2.5-50 μg/ml SNA-FITC (A) and PI (B) after blocking with three different buffers: PBS, 1% BSA/PBS, 1% FBS/PBS, and analyzed by flow cytometry. PI staining was used to gate live cells. Top panels: FITC (A) or PI (B) fluorescence histogram overlays of cell populations blocked with different buffers. Unstained cell and cells stained only with PI are also included as controls. Bottom panels: comparison of Mean fluorescence intensity (MFI) of SNA binding (A) and cell death (B) with different blocking buffers.
**Figure 9** The effect of different staining buffers on lectin binding. Prostate non-malignant epithelium RWPE-1 cells were stained with 100 μg/ml MAL (A) and 20 μg/ml ConA (B) in different buffer: PBS and TBS + CaCl$_2$+ MgCl$_2$, and analyzed by flow cytometry. PI staining was used to gate live cells. Dual parameter correlated dot plots display the fluorescence of FITC and PI. MFI of lectin staining for PI-population are indicated. Numbers at corner mean percentages of main population events within each quadrant.
Figure 10 The effect of cell confluence on SNA staining. PC3 cells with different confluence were stained with 10 μg/ml SNA in PBS, and analyzed by flow cytometry. PI staining was used to gate live cells. Left panel: FITC fluorescence histogram overlays of cells with different confluence rates. Nonspecific staining (unstained cells) is overlaid in grey filled line for comparison. Right panel: mean fluorescence intensity (MFI) of SNA binding upon different cell confluence rates.
low confluence showed lower fluorescent intensity than cells with higher confluence. However, the intensity was no different after cells reached 60% confluence.

Some lectins such as SNA, WGA and ConA are potential inducers of apoptosis. The optimal concentration of these lectins which can best distinguish positive from negative cell population and cause minimum cell death was determined. Figure 11 showed the example of SNA dose curve. The short incubation of PC3 cells (30 min at 4°C) with a series of diluted SNA generated a dose-dependent increase in fluorescent signal; however, it also produced dose-dependent cell agglutination and cell death. Balance of the signal intensity and cell death, SNA at 10 μg/ml was chosen for the subsequent flow cytometric experiments. By contrast, some lectins such as LFA had minimal cytotoxic effects on cells (Fig. 12). So the optimal concentrations for these lectins (100 μg/ml for LFA) were based on their best capacity to tell positive population. The same procedures were performed to optimize conditions for other lectins used in this study, and their optimal working concentrations and buffers were indicated in Table 4.

Lastly, we examined the specificity of lectin binding by their hapten sugars. Five free sugars were chosen to for the examined lectins: lactose (SNA, MAL), GlcNAc (WGA), fucose (AAA), GalNAc (HPA), and mannose (ConA). Representative histograms of the various blocking experiments can be seen in Figure 13. Preincubation of lectin with their corresponding competitive sugars exclusively blocked or reduced lectin binding (p<0.05, range 1.4-74.4 fold), confirming the binding specificity of these lectins. However, the degree of reduction was various. For instance, 0.2 M GalNAc almost completely reduced the fluorescence to the auto fluorescence level while the same concentration of lactose only decreased 28% MAL binding. The differences underscore
Figure 11 Titration of SNA against PC3 cells. PC3 cells were stained with 5-100 μg/ml SNA-FITC (A) in PBS and PI (B), and analyzed by flow cytometry. PI staining was used to gate live cells. Top panels: FITC (A) or PI (B) fluorescence histogram overlays of cell populations with different concentrations of SNA. Unstained cell and cells stained only with PI are also included as controls. Bottom panels: mean fluorescence intensity (MFI) of SNA binding (A) and cell death (B) upon different SNA concentrations.
Figure 12 Titration of LFA against ML2 cells. ML2 cells were stained with 5-100 μg/ml LFA-FITC (A) in TBS + CaCl$_2$+ MgCl$_2$ and PI (B), and analyzed by flow cytometry. PI staining was used to gate live cells. Top panels: FITC (A) or PI (B) fluorescence histogram overlays of cell populations with different concentrations of LFA. Unstained cell and cells stained only with PI are also included as controls. Bottom panels: mean fluorescence intensity (MFI) of LFA binding (A) and cell death (B) upon different LFA concentrations.
<table>
<thead>
<tr>
<th>Lectin</th>
<th>Carbohydrate Affinity</th>
<th>Staining Buffer</th>
<th>Concentration (µg/ml)</th>
<th>Inhibitory Sugars</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFA</td>
<td>Sialic acid</td>
<td>TBS+CaCl₂</td>
<td>100</td>
<td>0.2M NANA</td>
</tr>
<tr>
<td>SNA</td>
<td>α2-6 sialic acid</td>
<td>PBS</td>
<td>10</td>
<td>0.2M Lactose</td>
</tr>
<tr>
<td>MAL/MAA</td>
<td>α2-3 sialic acid</td>
<td>PBS</td>
<td>100</td>
<td>0.2M Lactose</td>
</tr>
<tr>
<td>WGA</td>
<td>sialyloligosaccharide, GlcNAc</td>
<td>PBS</td>
<td>20</td>
<td>0.2M GlcNAc</td>
</tr>
<tr>
<td>AAA</td>
<td>α-fucose</td>
<td>PBS</td>
<td>100</td>
<td>0.2M Fucose</td>
</tr>
<tr>
<td>HPA</td>
<td>GalNAc</td>
<td>TBS+CaCl₂</td>
<td>100</td>
<td>0.2M GalNAc</td>
</tr>
<tr>
<td>ConA</td>
<td>α-Man&gt; α-Glc</td>
<td>TBS+CaCl₂</td>
<td>20</td>
<td>0.2M Mannose</td>
</tr>
</tbody>
</table>

LFA, Limax Flavus Agglutinin; SNA, Sambucus Nigra Agglutinin; MAL, Maakia Amurensis Lectin; WGA, Wheat Germ Agglutinin; AAA, Anguilla Anguilla Agglutinin; HPA, Helix Pomatia Agglutinin; ConA, Concanavalin A. NANA, N-acetyl neuraminic acid; GlcNAc, N-acetyl-D-glucosamine; GalNAc, N-acetyl-D-galactosamine. Glc, D-glucose; TBS, Tris buffered saline; PBS, phosphate buffered saline.
Figure 13 The carbohydrate specificities of lectin binding. (A-F) Lectins were preincubated with their corresponding inhibitory sugars for overnight. The mixtures or lectins only were incubated with PC3 cells in PBS or TBS + CaCl$_2$ + MgCl$_2$. The lectin binding signal was analyzed by flow cytometry. PI staining was used to gate live cells. Fluorescence histograms of cells stained with lectin only (red), lectin+ inhibitory sugar (blue), and unstained (gray), are overlaid.
the application of optimal concentrations as well as binding conditions of inhibitory sugars.

### 3.3.4 Flow cytometric assessment of lectin-binding profiles

Having established an optimized lectin staining protocol (Table 4), we next compared the cell surface glycosylation profiles among PC3 and its metastatic variants by lectin flow cytometric analysis. Mean fluorescent intensity (MFI) of living cells (PI-) was used to quantitate lectin binding. The results of cell-binding assays for each of the three cell lines on the panel of 7 lectins are shown in Figure 14. In general, control unstained samples emitted very low levels of fluorescence (MFI average: PC3 6.63; N2 4.91; ML2 6.54). Four lectins (SNA, MAL, ConA, and WGA) showed very strong binding compared to control samples (p<0.05). The MFI for the remaining lectins was low, but still significantly different (p<0.05) from control value. Among cell lines, distinct binding between N2 and ML2 cells was observed for sialic acid. N2 cells showed higher binding than ML2 cells for SNA (2.14 fold, p<0.05) which recognizes α2-6 linked sialic acid, and for WGA (1.25 fold) which has the carbohydrate specificity for sialyloligosaccharide/ GlcNAc. In terms of MAL which binds preferentially to α2-3 linked sialic acid, ML2 cells showed a little higher binding (1.2 fold) than N2 cells, but the difference was not significant. Neither was those detected between cell lines for other lectins.

### 3.3.5 Carbohydrate analysis using lectincytochemistry
Figure 14 Cell surface carbohydrate profile analysis by lectin flow cytometry. PC3, N2, and ML2 cells were stained with seven lectins with different carbohydrate specificities under the optimal experiment conditions, and analyzed by flow cytometry. PI staining was used to gate live cells. Mean fluorescence intensity (MFI) of each lectin binding for three PCa cell lines was compared. Error bars represent the standard deviation of three independent experiments. Mann-Whitney U test was used to ascertain statistical differences. * denotes significant difference (p ≤ 0.05).
The spatial distribution of lectin binding was revealed by confocal microscopy. Because cells need to be fixed in this experiment, we first assessed the effect of PFA fixation on lectin binding. Lectin binding on postfixed cells (staining before fixation) was compared with that on prefixed cells (staining after fixation). In general, the staining patterns under two different experimental conditions were close except that prefixation resulted in loss of LFA binding to all three cell lines (data not shown). In fact, it was documented that fixation can change lectin binding patterns of cells. Unfavorable combinations of fixatives and buffers can cause a loss of more than 90% bound lectin \(^{231}\). For example, a loss of lectin mistletoe lectin-1 binding was observed during lectin cytochemistry analysis of fixed colon cancer HT 29 cells \(^{232}\). Based on these findings, postfixation of cells is more effective and we applied it for all coverslip staining.

Under confocal microscopy, similar subcellular staining patterns were observed across three cell lines for all lectins (Fig. 15). Binding was typically seen along the cell membranes, with long or short cell surface projections. In very few cells, granular uptake of the glycoconjugate complexes into the cytoplasm could be visualized as a granular fluorescence. WGA and ConA displayed the brightest fluorescence over the entire cell membrane, while SNA and MAL exhibited moderate signals. The binding sites for lectins LFA, AAA, and HPA, which showed low fluorescent intensities in flow analysis, were also evident during microscopic evaluation except LFA, which gave a totally negative binding to ML2. Due to the variations in capture conditions, we didn't quantitate the staining intensity across cell lines. But above results confirmed that lectin binding sites were almost exclusively on the cell surface.
Figure 15 Cell surface carbohydrate expression analyzed by lectin cytochemistry. PC3, N2, and ML2 cells were cultured on coverslips in 6-well plates. They were stained with seven lectins with different carbohydrate specificities under the optimal experiment conditions, fixed, and examined by confocal microscopy.
3.4 Discussion

In this aim we presented a study using a panel of lectins to characterize carbohydrate expression patterns on PC3 and its metastatic variants. Our approach was first to delineate the types and levels of carbohydrates using flow cytometry, and second to localize specific carbohydrate residues with respect to the subcellular regions. Because several factors such as blocking buffer, cell confluence, and specificity of lectin are important considerations, an optimized protocol and suitable controls were included to verify accurate data interpretation. For flow cytometry, counterstaining with an impermeable nuclear stain, propidium iodide, and analyzing the fluorescence signals of intact cells ensure that lectin binding is localized to the cell surface rather than intracellular sites. For microscopic analysis, fixation performed after lectin binding avoids the interference by various artifacts. By adopting these strategies, we have evaluated the carbohydrate profiles on PCa cells and identified the specific carbohydrates associated with PCa cell metastatic potentials.

Flow cytometry assessment of lectin binding revealed that cell surface carbohydrate of PC3 and its variants is composed mainly of glycoconjugates containing α-mannose/α-glucose, GlcNAc, and sialic acid as terminal saccharides, based on the high fluorescence intensity values obtained from SNA, MAL, ConA, and WGA. These results were not surprising since α-mannose/α-glucose and GlcNAc are present in all N-linked glycans, and sialic acid is capped on both O- and N-linked glycans. Few residues of fucose (α-1, 6 linkage) and GalNAc were detected by flow cytometry, suggesting that these carbohydrates were minor. But another possibility is that these sugars are present on the core structure of glycans and masked by the terminal sialic acid\(^{233}\). In this case,
neuraminidase treatment in combination of lectin flow cytometry can uncover more binding sites and reveal their real abundance. It is noteworthy that lectin LFA recognizing the same broad carbohydrate group as SNA, MAL, and WGA, however, provided the very low fluorescent intensity. This apparent conflict in carbohydrate identification can result from several factors affecting binding site access, which include subtle differences in binding specificity within carbohydrate groups, as well as steric configurations of lectins. In addition, the differences in fluorescence conjugated to each lectin also influence lectin performance. Therefore, a diversity of glycoconjugates containing unique sugar structures or sequences may be present on PCa cell surface, and a wider spectrum of reliable lectins is necessary to properly characterize glycan composition.

The most remarkable difference in cell surface glycan composition among three cell lines was observed for α2-6 linked sialic acid. SNA binding was greatly decreased in metastatic ML2 cells compared to non-metastatic N2 cells, suggesting a possible negative role for the expression of α2, 6-linked sialic acid in metastasis. Sialic acids are nine-carbon monosaccharide typically terminating the out ends of cell surface glycan chains. They are greatly diversified and more than 50 structural variations exist in nature. Increases in cell surface sialylation are often linked to transformation and metastatic phenotype. For instance, Yogeeswaran et al. reported the positive correlation between the quantity of cell surface total sialic acid and metastatic ability of murine tumor cells. Similarly, Sata et al. demonstrated an increased α2,6-sialylation on human colon cancer cell membranes using the lectin SNA. A high level of SNA reactivity was found to be an independent parameter for poor prognosis in colorectal carcinoma.
However, the opposite evidence was also documented. In a comparison of cell surface glycoconjugates between a highly metastatic murine melanoma cell line B16-BL6 and its positive sialidase transfectants, which showed markedly decreased metastasis in vivo and invasiveness in vitro, lectin flow cytometry and lectin blotting didn't revealed significant differences using either sialic acid-recognizing lectins or anti-sialyl Le\[^x\] antibody\(^{238}\). In addition, Takano \textit{et al.} described a WGA-resistant mutant of MDAY-D2 cell line (W16) which showed a small increase in total sialic acid, but accompanied by slower growing subcutaneous tumors and less liver metastasis\(^{239}\). Therefore, some researcher pointed out that metastatic ability does not correlate with the overall sialic acid content but only with specific molecules such as β1 integrin\(^{238,240}\). In another direction, there are observations that specific sialic acid linkage correlated with the metastatic growth of cancer cells. For example, analysis of B16 melanoma metastatic variants showed no difference in the total levels of sialic acid, but the highly metastatic subline contained more α2-3 sialylation in SAα2-3Galβ1-4GlcNAc oligosaccharides whereas low metastatic variants exhibited higher α2-6 sialylation in SAα2-6Galβ1-4GlcNAc chains\(^{113}\). These results provide supporting for our finding. Alternations in the distribution of sialic acid on cell surface glycoconjugates in cancers might lead to changes in cell-cell adhesion by the exposing of sub-terminal glycans, and/or inappropriate exposure of peptides. These changes may be important in the process of cancer cell dissemination and may influence the ability of cancer cells to adhere to endothelia at distant sites for the formation of metastasis\(^{86,234}\).

In summary, the study in Aim I showed that cell surface glycoconjugates of PC3 and its variants are extensively sialylated and contain residues of α-mannose/α-glucose, GlcNAc, fucose, and GalNAc in variable amounts. The differences in metastatic abilities
of PC3 variants did correlate with the relative levels of α2-6 sialic acid on their surfaces. Characterization of carbohydrate residues related to PCa metastasis represents the first step in our investigation. Studies would move forward to specifically isolate and identify the metastasis-associated cell surface sialoglycoproteins.
CHAPTER IV

AIM II: TARGETED IDENTIFICATION OF METASTASIS-ASSOCIATED CELL SURFACE SIALOGLYCOPROTEINS IN PROSTATE CANCER

4.1 Introduction

Altered cell surface glycosylation is a hallmark of cancer. Specific glycan structures have been identified on tumors and serve as diagnostic indicators of the malignant and metastatic phenotype. Sialic acid (SA) residues are located on the nonreducing terminus of the glycans on many cell surface glycoproteins where they play a pivotal role in viral infection, leukocyte-endothelial cell adhesion, neuronal development, immune cell activation, cancer metastasis, and many other normal and pathological processes. Our previous data from lectin staining demonstrated changes of cell surface sialylation in a syngeneic prostate cancer cell model. Evidence from both patient histochemical analysis and experimental tumor models also indicate that altered sialylation of tumor cell surfaces is associated with a metastatic tumor phenotype. As we described earlier, the surface sialylation changes have been reported reflecting the amount, type, distribution, and bonding of sialic acids to adjacent molecules. For instance, a positive correlation can be reported between the levels of cell surface sialylation and metastatic ability of various experimental tumors. In addition, the distribution of sialic acid on specific N- or O-linked oligosaccharides has been demonstrated to alter the metastatic potential of cancer cell lines. In some cases tumor development has been associated with overexpression of β1-6 branching that results in increased number of Galβ1-4GlcNAc structures available for sialylation. Therefore, exploring cell surface sialylation changes during tumor development and
disease progression likely affords excellent opportunities to identify sensitive and specific cancer biomarkers.

Elucidation of structural details of cell surface glycosylation by mass spectrometry is hampered by the limited relative abundance of surface proteins compared to cytosolic components, the complex and microheterogeneous nature of glycans, and the inherent complexities of deciphering low energy carbohydrate fragmentation ions versus higher energy peptide fragments in complex mixtures. The type of mass spectrometer and ionization energies to be used, and the complexity of the sample, are critical parameters for successful glycoprotein analysis. In recent years, lectin- and antibody-based affinity selection has been used with some success to purify glycoproteins/glycopeptides with specific structures. Other approaches for glycopeptide characterization arise from the exploitation of glycan physical and chemical reactivity. Larsen et al. took advantage of the high affinity of titanium dioxide microcolumns toward SA residues to isolate SA-containing peptides from serum under highly acidic conditions. Two similar approaches, involving hydrazide and boronic acid chemistry, capitalize on the cis-diols present in monosaccharides. Zhang et al. described the use of hydrazide chemistry for the purification by directly coupling of glycoproteins to a solid support. Similarly, Sparbier et al. used boronic acid-functionalized beads to covalently capture glycoproteins followed by elution with acid. Although these methods are effective at the enrichment and identification of broad classes of glycoproteins/glycopeptides, they still lack the specificity and selectivity required for analysis of specific cell surface glycoproteins that could serve as potential cancer biomarkers.
In recent years, a growing area of chemical biology strives to probe glycans in living systems by using bioorthogonal chemical reactions. These strategies, first described by Bertozzi's group, modulate cell surface glycan structures by biosynthetic introduction of unnatural sugars containing bioorthogonal chemical reporters, namely, functional groups that possess unique reactivity orthogonal to those of natural biomolecules. The process entails two steps. First, cells (or organisms) are incubated with a metabolic monosaccharide precursor adorned with a unique functional group—the chemical reporter. Once the chemical reporter is incorporated into the target glycans, it is treated in a second step with a probe molecule bearing complementary bioorthogonal functionality. In principle, a chemical handle such as a ketone, azide, or thiol carried by unnatural monosaccharides, can be metabolically incorporated into glycoproteins of cells or in living organisms. Azide has proven the most useful analogs owing to the small size, metabolic stability, selective reactivity profile, and lack of reactivity with natural biofunctional structures. In sialic acid biosynthesis, as shown in Figure 16, the first committed step is the conversion of UDP-N-acetylglucosamine (UDP-GlcNAc) to N-acetylmannosamine (ManNAc). Six subsequent steps, including five enzymatic transformations and one transport step, produce a sialic acid residue appended to a glycoprotein or glycolipid destined for cell surface or secretion. Interception of the pathway with N-azidoacetyl-D-mannosamine (ManNAz), an unnatural analog of ManNAc which competes with ManNAc in the sialic acid pathway, results in the cell surface presentation of the azide-labeled glycans (Fig. 17A). The azide can be chemically probed with phosphine via the Staudinger ligation, linear alkynes via the Cu-catalyzed azide-alkyne cycloaddition (CuAAC), and with a variety of
Figure 16 Schematic representation of sialic acid biosynthetic pathway in human cells. Cytosolic glucose is converted in several steps into UDP-GlcNAc, which serves as substrate for the bifunctional, rate-limiting, committed enzyme of sialic acid biosynthesis, UDP-GlcNAc 2-epimerase (GNE) (a). The GNE catalytic activity epimerizes UDP-GlcNAc to ManNAc, followed by the phosphorylation of ManNAc to ManNAc-6-phosphate (MaNAc-6-P) by ManNAc 6-kinase (b). ManNAc-6-P is then further converted into NeuAc (sialic acid) by the sequential action of sialic acid 9-phosphate synthase (c) and sialic acid 9-phosphatase (d). NeuAc is activated into cytidine monophosphate–sialic acid (CMP-NeuAc) by CMP-sialic acid synthetase (e) in the nucleus. Cytosolic CMP–NeuAc displays strong feedback inhibition (dotted line) of GNE enzymatic activity by binding to its allosteric site, thereby contributing to the tight regulation of intracellular sialic acid biosynthesis. CMP–NeuAc can subsequently be transported into the Golgi complex (f) as a substrate for the biosynthesis of sialyl-oligosaccharides by sialyltransferases (g). Adapted from Yarema, K.J. et al. Nat Biotechnol 19, 553 (2001).
Figure 17 Bioorthogonal chemical reactions for sialic acid. The process entails two steps: metabolic labeling (A) and click chemistry (B). (A) Cells are fed with peracetylated N-azidoacetylmannosamine (Ac$_4$ManNAz), an unnatural analog of N-acetylmannosamine (ManNAc). Ac$_4$ManNAz penetrates the cell-surface membrane easily due to the hydrophobic nature of the modified sugar. In the cytosol, it is enzymatically deacetylated and then metabolically converted to the corresponding N-azidoacetyl sialic acid (SiaNAz), which is subsequently incorporated into sialoglycoconjugates. (B) Glycoproteins selectively labeled with azide can be detected with a fluorescent or biotinylated alkyne via the Cu-catalyzed azide-alkyne cycloaddition or “click chemistry”. In this reaction, an azide and a terminal alkyne form a very stable triazole conjugate with the help of Cu (I) catalysis. Adapted from Invitrogen website (www.invitrogen.com).
cyclooctynes via the strain-promoted azide-alkyne cycloaddition\textsuperscript{255}. As a result, the reporting sugars have been coupled with a plethora of tags such as FLAG peptides, biotin, fluorescent or fluorogenic molecules, with the potential for visualization and isolation. Each chemicalselective reaction has its own advantages and disadvantages. The Staudinger ligation, which forms an amide bond between the azide and an esterdervatized phosphine, suffers from slow reaction kinetics and competing oxidation of the phosphine reagents\textsuperscript{249}. By contrast, CuAAC, often referred to classical click chemistry, offers high-sensitivity and fast rate by forming a triazole from an azide and a terminal alkyne with the help of Cu (I) catalysis (Fig. 17B). But it may hurt live cells by Cu cytotoxicity\textsuperscript{249}. The successful application of bioorthogonal chemical reactions to label and detect several types of cell surface glycans, such as sialiac acid\textsuperscript{253,256,257} and mucin-type O-linked glycans (GalNAC)\textsuperscript{258,259}, or cytosolic O-GlcNAcylated glycans (GlcNAc)\textsuperscript{260}, has been reported by some groups. However, their application for proteomics is limited to only O-GlcNAc modified cytosolic glycoproteins based on the Staudinger ligation reaction\textsuperscript{261,262}.

In this study, we described a novel glycoproteomic identification strategy for the selective detection, isolation and identification of cell surface sialoglycoproteins from cultured cell lines. The method utilizes the sialic acid biosynthetic pathway for the incorporation of monosaccarides bearing bioorthogonal functional handles (tetraacetylated N-azidoacetyl-D-mannosamine, AC\textsubscript{4}ManNAz) into cellular sialic acid. This reagent has primarily been used to label and visualize cell surface sialyl glycoconjugates by Staudinger ligation\textsuperscript{252,253}. We extended its application to detect cell surface azide-modified sialoglycoproteins by click chemistry. To illustrate the potential
of using this cell labeling procedure in biomarker discovery, we combined it with an MS-based proteomics approach as applied to our syngeneic metastatic PCa cell line model.

4.2 Materials and Methods

Materials — Complete™ protease inhibitors were purchased from Roche Applied Sciences (Indianapolis, IN, USA), sequencing grade trypsin was from Promega (Madison, WI, USA), and Immobilon-FL PDVF membrane was from Millipore (Billerica, MA, USA). Protein-free blocking buffer and high capacity streptavidin agarose resin was from Thermo Scientific (Rockford, IL, USA). Dulbecco's modified Eagle media (DMEM), fetal bovine serum (FBS), antibiotics-antimycotic, Click-iT™ ManNAz metabolic glycoprotein reagent, Click-iT™ Biotin Protein Analysis Detection Kit, propidium iodide, fluorescein conjugated streptavidin (streptavidin-FITC), 4-12% NuPAGE® Bis-Tris gels, and lithium dodecyl sulfate (LDS) buffer were from Invitrogen (Carlsbad, CA, USA). 2 x Laemmli buffer, 7.5% Criterion® Tris-HCl Gel, and Bio-Safe coomassie blue were from Bio-Rad (Hercules, CA, USA). The anti-β-actin mAb were provided by BD PharMingen (San Diego, CA, USA). The streptavidin-IR 800 and goat anti-mouse antibody conjugated to IR 800 were from Li-COR Biosences (Lincoln, PA, USA).

Cell Culture and Metabolic Labeling — PC3-N2 and PC3-ML2, as described in Aim I, were cultured in DMEM medium supplemented with 10% FBS and 1% antibiotics at 37 °C with 5% CO₂. For metabolic labeling, growth medium was replaced at 70% cell confluence with complete DMEM medium containing various concentrations of an azido-
modified sugar, tetraacetylated N-azidoacetyl-D- mannosamine (AC₄ManNAz), or a control sugar, N-acetyl-D-mannosamine (ManNAc), and cells were incubated for 1-3 days. Cells then were dissociated from the plastic surface by non-enzyme dissociation buffer.

Flow Cytometric Analysis of Cell surface Sialic Acid Labeling — After metabolic labeling, N2 and ML2 cells were harvested, washed with 0.1% FBS/PBS, and resuspended (10⁶ cells) in 100 µl click reaction solution with the indicated amount of each component. The reaction was incubated at room temperature for 30 min, and then cells were washed three times with 0.1% FBS/PBS. Cells were subsequently stained with streptavidin-FITC (1 µg/sample in 100 µl 2% FBS/PBS) for 30 min at 4°C, and washed with 2% FBS/PBS three times. Before subjecting to flow cytometric analysis, cells incubated with 1 µg/ml propidium iodide in 500 µl 2% FBS/PBS at 4°C for 20 min. Data was acquired by FACScalibur and analyzed by Flowjo software.

Confocal Microscopy Analysis of Cell surface Sialic Acid Labeling — N2 and ML2 cells were seeded onto 6-well plates containing glass coverslips and cultured in 10% FBS/DMEM. Growth medium was supplemented with 40 µM AC₄ManNAz or ManNAc for 3 days. Cells were washed with ice-cold PBS, fixed with 2% paraformaldehyde, and then subjected to click reaction solution (25 µl biotin-alkyne, 2.5 µl CuSO₄, 2.5 µl and 5 µl for component D and E, 65 µl PBS). Subsequently, the fixed and labeled cells were washed with PBS and stained with streptavidin-FITC (same as above) or streptavidin Alexa fluor 488. To visualize intracellular sialylation, permeabilization with 0.5% Triton
X-100 was performed after fixation and before click reaction. PI or TO-PRO3 was used to stain nuclei. Finally coverslips were mounted to the slides with VectorShield medium, and sealed with nail polish. Fluorescent images were examined and captured under a confocal microscope.

*Cell surface Sialoglycoprotein Labeling and Detection in Cell Extracts* — N2 and ML2 cells were seeded in 15-cm dishes and treated with the optimized labeling conditions for AC₄ManNAz and ManNAc (20 μM for 1 day) in growth medium. After cells were harvested with non-enzyme dissociation buffer, 5 × 10⁷ cells were collected and washed with 0.1% FBS/PBS. Subsequently, cells were subjected to click reaction solution as described above. Upon conjugation, cell pellets were washed with ice-cold PBS twice to remove unreacted reagents and then lysed in lysis buffer I (1% NP-40, 150 mM NaCl, protease inhibitor, 100 mM sodium phosphate, pH 7.5) using a Dounce homogenizer. The total cell lysate were further cleared by methanol-chloroform precipitation and resolved in lysis buffer II (1% SDS, protease inhibitor, 50 mM Tris-HCl, pH 8.0).

Protein concentrations were measured using the BCA protein assay (Pierce). To detect biotin-labeled sialoglycoproteins in cell extracts, 20 μg of labeled protein lysate was resolved by SDS-PAGE. Electrophoresed proteins were transferred onto PVDF membrane, blocked with Odyssey Blocking Buffer (Rockland Immunochemicals, Gilbertsville, PA, USA), probed with streptavidin-IR 800, and visualized and quantified using an Odyssey infrared imaging system.
Sialoglycoprotein Capture — Streptavidin (SAv) beads were pretreated with protein-free blocking buffer overnight at 4°C and washed five times with PBS. Cell lysate (2 mg) was incubated with 100 μl beads in 0.3% NP-40/ PBS overnight at 4°C with a rotating shaker. The captured glycoproteins were washed intensively with modified RIPA buffer (150 mM NaCl, 2% SDS, 1% NP-40, 1% Na Deoxycholate, 50 mM Tris-HCl, pH 7.5). Bound material was eluted by boiling for 10 min in 100 μl 2 × Laemmli sample buffer. For assessment of capture efficiency, 10 μl of the eluent, along with the input and flow-through fractions were separated by SDS-PAGE and visualized with streptavidin-IR 800 as stated above.

1D Gel Electrophoresis and In-gel Digestion — Captured glycoproteins (45 μl) were separated through on 1.0 mm 7.5% Tris-HCl polyacrylamide gels. Gels were stained with Coomassie blue and imaged on a Typhoon 9410 (GE Healthcare, Piscataway, NJ, USA). Twenty-seven equally-spaced gel pieces were excised from each lane, spanning the full height of the gel (40-300 kDa). Individual gel pieces were destained with 25 mM NH₄HCO₃ in 50% acetonitrile (ACN), reduced with 20 mM dithiothreitol in 25 mM NH₄HCO₃ for 45 min at 56°C, and alkylated with 55 mM iodoacetamide in 25 mM NH₄HCO₃ in the dark for 1 hr at room temperature. After washing, the gel pieces were dehydrated with ACN and dried using a speed vac. Trypsin (12.5 ng/μl in 25 mM NH₄HCO₃, pH 8.0) was added to each gel piece, and the gel pieces were allowed to swell on ice for 1 hr. Excess trypsin was removed, replaced with 25 mM NH₄HCO₃, 10% ACN, pH 8.0, and the mixture was incubated overnight at 37°C. The digest was collected
and peptides were extracted with 50% ACN/5% formic acid twice. The combined extracts were then dried using a Speed-Vac for subsequent LC-MS/MS analysis.

**Whole Cell Extract and In-solution Digestion**—For N2 and ML2 cell lysate experiments, both cell lines were cultured in complete growth medium to 80% confluence. Cells were then washed with PBS and the pellet resuspended in 160 µl dissolution buffer containing 100 mM NH₄HCO₃ and TFE (1:1 v/v). The samples were sonicated for 20 seconds three times and incubated at 60°C for 1 h. The lysates were centrifuged to remove cell debris and unbroken cells and the supernatant was collected. The protein concentration of the samples was determined by BCA assay and normalized for each sample and before reduction and alkylation with TCEP and iodoacetamide respectively. The concentration of TFE was reduced to 5% by the addition of 1.4 ml 100 mM NH₄HCO₃. Trypsin was added at a ratio of 1:50 protease to protein and the digestion proceeded at 37°C for 18 hrs with constant mixing. After digestion the sample was dried down in a Speed-Vac, desalted and lyophilized before LC-MS/MS analysis.

**LC-MS/MS analysis** — Digests were resuspended in 20 µl Buffer A (5% ACN, 0.1% Formic Acid, 0.005% heptafluorobutyric acid) and 15 µl loaded onto a 12-cm x 0.075 mm fused silica capillary column packed with 5 µM diameter C-18 beads (The Nest Group, Southboro, MA) using a programmed automatic injection. Peptides were eluted over 80 minutes, by applying a 0-80% linear gradient of Buffer B (95% Acetonitrile, 0.1% Formic Acid, 0.005% HFBA) at a flow rate of 200 µl/min with a pre-column flow splitter resulting in a final flow rate of ~300 nl/min directly into the source. In some
cases, the gradient was extended to 150 minutes to acquire more MS/MS spectra. A LTQ™ Linear Ion Trap (ThermoFinnigan, San Jose, CA, USA) was run in an automated collection mode with an instrument method composed of a single segment and five data-dependent scan events with a full MS scan followed by four MS/MS scans of the highest intensity ions. Normalized collision energy was set at 35, activation Q was 0.250 with minimum full scan signal intensity at $1 \times 10^5$ with no minimum MS$^2$ intensity specified. Dynamic exclusion was turned on utilizing a three-minute repeat count of 2 with the mass width set at 1.0 m/z. Sequence analysis was performed with MASCOT (Matrix Sciences, London GB) using the SwissProt 57.1 database with a human taxonomy filter enabled.

The database searches were performed with fixed modification as carbamidomethy (Cys) and variable modifications as oxidation (Met) and deamidation (Asn, Gln). Enzyme specificity was selected to typsin. The spectra from 27 LTQ runs in each gel lane were combined to search SwissProt. All the MS/MS spectra were manually inspected to verify the validity of the database search results. False discovery rates were estimated to be 0.25% on the protein level by searching a decoy version of the SwissProt protein database. The relative abundance of peptides was estimated by spectral counting.

**Bioinformatic Analysis** — Initial basic information on the structure of the identified proteins was obtained through literature reports. The ProteinID Finder (Proteome Solutions) program was employed to extract such available information from the UniProt database. Due to the limited annotation of protein glycosylation for most human proteins and the shortage of subcellular location information for hypothetical proteins and functionally uncharacterized proteins, we also subjected each identified protein to
prediction algorithms based on protein sequence analysis. For protein glycosylation we used NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) to predict the possible presence of the consensus NXS/T glycosylation motif. For subcellular location, all identified proteins were analyzed with two transmembrane prediction algorithms SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) indicating hydrophobic protein sequence regions. As a group, the total identified proteins were analyzed by Ingenuity Pathways Analysis (IPA) software to determine their molecular function and interacting networks.

To generate the lists of cell surface glycoproteins uniquely expressed in N2 or ML2 cells, identified proteins were required to meet the following 3 criteria: (a) identified from one cell line; (b) unique peptide ≥ 1, and NXS/T motif ≥ 1 or UniProt indicated N-linked or O-linked glycosylation; and (c) transmembrane region ≥ 1, or UniProt membrane subcellular location.

4.3 Results

4.3.1 Methodology Overview.

Our strategy for interrogation of cell surface sialoglycoproteins using selective chemical tagging followed by high-affinity enrichment and GeLC-MS/MS analysis is summarized in Figure 18. The method consists of several steps: 1) metabolic labeling of N2 and ML2 cells with the azide-containing mannose analogue, AC₄ManNAz; 2) chemoselective conjugation of azide sugars with a biotinylated alkyne capture reagent via Cu (I) catalyzed click chemistry in live cells; 3) affinity enrichment of the labeled cell
Figure 18 Outline of Aim II experimental work-flow. (1) metabolic labeling cells with an unnatural mannose analogue, peracetylated azido-mannose (AC₄ManNAz); (2) chemoselective conjugation of azido sugars with a biotinylated alkyne capture reagent via Cu (I) catalyzed click chemistry on live cells; (3) cell lysis; (4) streptavidin (SAv) affinity purification; (5) sialoglycoproteins elution; (6) SDS-PAGE for protein separation; (7) gel slice digestion and peptide release; (8) peptide analysis by LC-MS/MS; (9) database search and protein identification; (10) bioinformatic analysis.
surface sialylated proteins by streptavidin (SAv) capture; and 4) separation by 1D gel electrophoresis and identification by LC-MS/MS. When compared with other published approaches, the theoretical advantage of our approach is the targeted selectivity for sialyl glycosylated proteins on the cell surface.

4.3.2 Efficient Cell Surface Expression of AC₄ManAz-Sialoglycoproteins.

A critical parameter in the proposed strategy is the efficient labeling and surface expression of the azide-containing sialoglycoproteins. In order to evaluate the efficiency of surface expression of azide-containing sialoglycoproteins, N2 and ML2 cells were subjected to metabolic incorporation of AC₄ManNAz, conjugated with biotin with the click reaction, and subsequently stained with streptavidin-FITC to visualize the glycans harboring the azide group. We conducted fluorescence-assisted flow cytometric analysis of non-permeabilized cells for determining the efficiency of streptavidin-FITC labeling. As quantified by flow cytometry (Fig. 19A), the AC₄ManNAz–treated N2 and ML2 cells displayed 80-100 fold greater FITC-specific fluorescence intensity than the control ManNAc-treated and untreated cells. This result demonstrates that the azide moieties were efficiently incorporated into cellular glycans. We next conducted confocal microscopy analysis of streptavidin labeled cells to specifically assess the cellular location of the azide-modified glycans. As seen in Figure 19B, prominent FITC staining primarily at the cell membrane was observed when the cells were subjected to metabolic incorporation of ManNAz. Consistent with flow cytometric data, confocal microscopic analysis of cells treated with control ManNAc displayed much reduced staining, confirming the labeling via azide-containing glycans is specific. Furthermore, there was
Figure 19 Efficient and selective labeling of sialoglycoproteins on the cell surface. (A) Flow cytometry analysis of N2 and ML2 cells treated with AC₄ManNAz or ManNAc. Nonpermeabilized cells were biotin-tagged through click reaction, and stained with streptavidin-FITC. PI negative gated FACS histogram showing three populations of cells: control cells (red), cells treated with ManNAc (gray filled), and cells treated with AC₄ManNAz (green). (B) Confocal imaging of N2 and ML2 cells for visualization of the tagged cell-surface sialoglycoproteins. Cells were treated with AC₄ManNAz or ManNAc, labeled with biotin-alkyne via click reaction and stained with streptavidin-FITC (green) and PI (red) after fixation. Left panel: AC₄ManNAz-treated cells; right panel: ManNAc-treated control cells.
no evidence of cell surface labeling for the control groups.

### 4.3.3 Visualization of sialylated glycoconjugated inside the cell.

The biosynthesis of cell surface sialoglycoconjugates involves a lot of intracellular processes (Fig. 16). After confirming that azide-modified sialyl glycans resided on the cell surface, we were curious how the metabolism of AC₄ManNAz looked inside the cells. To this end, we changed cell staining protocol by permeabilizing cells before performing click reaction. As demonstrated in Figure 20, treatment with ManNAz resulted in membranous staining as well as a strong juxtanuclear pattern inside the cells after clicking on the biotin-alkyne and staining with Alexa 488-conjugated streptavidin. This pattern seems typical for the Glogi apparatus, but need verification in future study. In contrast, cells incubating with natural ManNAc and following the same staining procedures showed minimal fluorescence. These data documented that metabolic labeling in combination with click reaction is an option to study intracellular sialyl residues, likely those trafficking through the secretory pathway.

### 4.3.4 Optimization of metabolic labeling and click reaction.

We further optimized the azide labeling and click reaction to maximize cell viability/vigor (Fig. 21). The first concern was whether exogenous sugar analogs affected cell growth since cell toxicity may lead to global protein expression changes. We measured cell proliferation with WST-1 reagent after cells were treated with AC₄ManNAz. As shown in Figure 21A, the addition of 200 µM ManNAz greatly
Figure 20 Fluorescent imaging of intracellular sialylation. N2 and ML2 cells were treated with AC₄ManNAz or ManNAc, fixed and permeabilized. Then cells were conjugated with biotin-alkyne via click reaction and stained with streptavidin Alexa Fluor 488 (green) and TO-PRO3 (blue). Fluorescent images were captured by cofocal microscopy. Left panel: AC₄ManNAz-treated cells; right panel: ManNAc-treated control cells.
**Figure 21** Optimization of azide labeling and click reaction. (A) N2 cells were grown in a 96-well plate with different doses of AC$_4$ManNAz (0-200 µM) for up to 3 days. Cell viability and proliferation was determined using WST-1 reagent. Each point was tested in 6 replicates. (B) Dose-dependency of click reaction. ML2 cells were treated with 40 µM AC$_4$ManNAz for 3 days, probed with different volumes of click reagents, and stained with streptavidin-FITC. The fluorescent intensity was monitored by flow cytometry. (C) Dose curve of ManNAz labeling. ML2 cells were treated with 0-80 µM AC$_4$ManNAz for 3 days, probed with 25 µl biotin-alkyne. (D) Time course of siaNAz glycoconjugate labeling. The optimized conditions of 20 µM AC4ManNAz and reaction with 25 µl biotin-alkyne were used in a time course of incorporation and detection in ML2 cells.
inhibited cell growth up to 90% at day 1-3. However, ManNAz up to 50 μM had no or minimal toxicity on cells. We next optimized the click reaction with the emphasis on the dose of biotin-alkyne and copper. Both showed a dose-dependent increase of fluorescent signal on the cell surface (Fig. 21B). Considering the potential toxicity of Cu (I), we choose 25 μl of biotin-alkyne with 2.5 μl CuSO₄ for future experiment because this amount of reagents is enough to provide strong azide-alkyne reaction (50 fold greater than control). Dose-curve of ManNAz labeling (Fig. 21C) showed a dose-dependent increase of fluorescent signal up to 20 μM. After that the MFI decreased gradually. However, the MFI of ManNAz treated cells was still with 24-fold at 80 μM greater than control ManNAc. Time-course experiments (Fig. 21D) revealed the saturation of ManNAz incorporation within one day of incubation. Surprisingly, two-day incubation led to a decrease of azide-bearing sialyl glycans on the cell surface whereas the level of labeled glycans on day 3 restored to 55% level of day1. This result might reflect turnover status of cell surface sialyl glycans. Base on these cumulative results, the optimal concentration and time of ManNAz for labeling sialyl glycans fall between 20-40 μM for 1 day and 25 μl of biotin-alkyne reagent.

4.3.5 Targeted Isolation of Cell surface Sialoglycoproteins.

Having demonstrated efficient labeling of cell surface glycoconjugates, we continued with a general characterization of whole cell lysates of Biotin-labeled glycoconjugates via SDS-PAGE separation and visualization by reaction with streptavidin-IR800. Expression of β-actin was used for normalization. Overall sialylation, as judged by streptavidin staining, was restricted to extracts from the
ManNAz labeled cells (Fig. 22). In addition, there appeared to be greater sialylated glycoproteins in the ML2 cell line when compared to the N2 cell line. We also observed a major protein band that migrated just under 150 kDa and reacted with streptavidin. These observations were consistent with the prior fluorescence imaging analyses in demonstrating a specific uptake and labeling of the ManNAz treated cells.

We next examined the efficiency of our system for the enrichment of biotin-labeled sialoglycoproteins. The streptavidin-derivatized beads were used to isolate targeted proteins from N2 and ML2 cell lysates. The effect of various parameters on the binding of non-glycoproteins versus glycoproteins to the streptavidin beads was evaluated by 1D-PAGE and streptavidin-FITC reaction to develop an optimized enrichment (Fig. 23). For example, a variety of protein/streptavidin bead ratios were investigated for efficient capture of the targeted glycoproteins (Fig. 24). The capture efficiency was analyzed by 1D-PAGE followed by silver staining as well as in-blot streptavidin-FITC detection. Compared to the control groups, the number of bands and band intensity observed on silver-staining gels and streptavidin-FITC blot for treated groups were increased dramatically when the ratio of sample loading/SAv beads increased. 500 μg cell lysate on 20 μl bead slurry resulted in the maximum isolation. Using the similar criteria, we optimized other parameters including SAv beads preblocking buffer, binding buffer, washing buffer, and elution buffers. We found that targeted glycoprotein binding to SAv beads was optimal when the beads were preblocked with protein-free blocking buffer and binding was performed at 4°C with a detergent (0.3% NP-40) present in the binding buffer. Increasing detergent concentration to 4% in RIPA wash buffer significantly
Figure 22 Detection of azide-tagged sialoglycoproteins in cell lysates. After N2 and ML2 cells were metabolically labeled with ManNAc analogs and conjugated with biotin-alkyne under the optimized conditions, the cells were lysed and total protein was extracted as stated in Experimental procedures. Top panel: 20 μg of labeled protein was resolved in 4-12% gel by SDS-PAGE and visualized by incubation with streptavidin-IR 800. Reactive bands indicate sialylated glycoproteins. The same blot was probed with β-actin to verify equal protein loading. Bottom panel: the numbers showed relative expression of sialylated glycoproteins after normalization with the β-actin.
Figure 23 Schematic representation of streptavidin (SAv) pull down assays. SAv beads were pretreated with protein-free blocking buffer and washed with PBS. Then they were incubated with cell lysate in 0.3% NP-40/ PBS. The bound proteins were washed intensively with modified RIPA buffer to reduce nonspecific protein binding. Precipitated proteins were eluted from SAv beads with Laemmli buffer by boiling. The optimized steps such as SAv beads preblocking, protein/beads ratio, binding buffer, wash buffer, and elution buffer, were highlighted in blue.
Figure 24 The effect of different protein/beads ratios in SAv pull down assays. (A and B) ML2 cells were metabolically labeled with ManNAc or its analogs (ManNAz), conjugated with biotin-alkyne, and lysed. Different amounts of lysate (25-200 μg) were subjected to SAv pull down assays by incubating with 20 μl SAv beads slurry. The eluted proteins were resolved in 4-12% gel by SDS-PAGE and visualized by silver staining. (C) Blot staining was performed on the same set of samples with streptavidin-FITC to confirm the specific enrichment. The blot was scanned by typhoon imager.
reduced nonspecific protein binding. Efficient elution of bound glycoproteins was achieved by boiling samples in 2 x laemmli buffer.

With the optimized streptavidin binding protocol, we added 2 mg whole cell lysates to 100 μl bead slurry and performed pull down assay as described. The eluted sialoglycoproteins were subjected to in-blot streptavidin-IR 800 detection. In both cell types we observed a dramatic enrichment for biotin-conjugated proteins in the bound and eluted fractions from the ManNAz-treated cells (Fig. 25). In contrast there was dramatically reduced overall staining of biotin-labeled protein in the ManNAc-labeled control cells. We again observed an intense signal at sub-150 kDa in eluted fractions from cells treated with ManNAz, and absent or minimal signal in flow-through fractions and eluents from the control group. These results demonstrate efficient capture and enrichment of the azide-labeled proteins.

4.3.6 Identification of Surface Glycoproteins by Mass Spectrometry.

Since we were able to establish both efficient cell surface labeling and effective isolation of sialoglycoproteins, we next performed a proof-of-principle experiment for the application of this approach to the identification of sialoglycoproteins differentially expressed between metastatic (ML2) and non-metastatic (N2) prostate cancer cell lines. The cells were processed for metabolic labeling as described above, the tagged proteins were enriched with SAv resins, and the eluted proteins were separated by SDS-PAGE (Fig. 26) and analyzed by LC-MS/MS. In total, 536 non-redundant proteins were identified (Fig. 27). This total derived from 324 proteins from the N2 line and 372 proteins from the ML2 line when metabolically labeled with ManNAz. In contrast the
Figure 25 Efficient capture and enrichment of azide-tagged sialoglycoproteins. Azide-tagged and biotin-conjugated sialoglycoproteins from the total lysate of N2 and ML2 cells were captured by streptavidin beads, separated in 7.5% gel and visualized by streptavidin-IR 800. Shown are the results from 20 μg of post-click lysate (N for neat), 5% of the flow through material that did not bind to the beads (FT for flow through), and 5% of material captured from 2 mg of protein that bound to beads (E for eluent).
Figure 26 Separation of proteins enriched by SAv pull down assays. Azide-tagged and biotin-conjugated sialoglycoproteins from 2 mg total lysate of N2 and ML2 cells were enriched by streptavidin beads. The eluents were separated in 7.5% gel and visualized by Coomassie blue staining. Twenty-seven equally-spaced gel pieces spanning the full height of the gel (40-300 kD) were excised from each lane, in-gel digested by trypsin, and analyzed by LC-MS/MS.
Figure 27 Number of proteins identified in PC3 variants by LC-MS/MS. Venn diagrams showed the overlap of proteins identified in N2_ManNAz group, ML2_ManNAz group, and ManNAc_pool group which combined N2_ManNAc and ML2_ManNAc groups. Proteins identified with significant peptide scores (p<0.05) and expect scores (p<0.5), and at least one unique peptide were included in the analysis.
control cells lacking ManNAz yielded 132 proteins from the N2 line and 75 proteins from the ML2 line. The total number of proteins identified using a sliding scale for unique peptides is shown in Table 5. Approximately 60% of proteins identified contained two or more unique peptide hits. Another purpose of setting up control groups was to evaluate the extent of nonspecific interaction between the SAv resin and proteins in the N2 and ML2 cell lysates. Although a pre-blocking step and high detergent resin wash was carried out, a number of proteins were still identified in the control groups. Most of these proteins were high abundant house-keeping proteins, such as actin, tubulins, plectin-1, pyruvate carboxylase, pyruvate kinase and heat shock proteins. These proteins were subtracted from the list of total proteins identified in the treated groups, further narrowing the pool of proteins identified as cell surface molecules (216 proteins for N2 and 245 for ML2 group).

The enrichment efficiency for sialoglycoproteins is a key factor in characterization of the utility of this methodology. As shown in Figure 25, the presence of sialoglycoproteins was significantly enhanced after affinity chromatography by SAv beads when compared to raw cell lysate or processed control extracts (ManNAc). The specific enrichment of glycoproteins was further examined using the results from the mass spectrometry characterization. As an initial analysis we determined the top 10 most abundant glycoproteins from each cell line. The expression of these abundant proteins were estimated and ranked by spectral counts and unique peptides as shown in Table 6A and Table 6B. We then compared the relative abundance of these top 10 glycoproteins, as derived from the sialoglycoprotein-targeted enrichment/isolation, between the ManNAz and ManNAc labeled cells. Based upon this comparison, the top most
Table 5 Total number of proteins identified for each experimental group

<table>
<thead>
<tr>
<th>No. of Uniq_pep</th>
<th>N2_ManNAz</th>
<th>N2_ManNAc</th>
<th>ML2_ManNAz</th>
<th>ML2_ManNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>64</td>
<td>153</td>
<td>44</td>
</tr>
<tr>
<td>2</td>
<td>62</td>
<td>24</td>
<td>72</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>13</td>
<td>47</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>11</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>≥5</td>
<td>82</td>
<td>20</td>
<td>78</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>324</td>
<td>132</td>
<td>372</td>
<td>75</td>
</tr>
</tbody>
</table>
Table 6 Top 10 abundant glycoproteins identified in N2 and ML2 cells ranked by unique peptides

A. N2 cells

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc #</th>
<th>Prot_desc</th>
<th>N2_ManNAz</th>
<th></th>
<th>N2_ManNAc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spect_</td>
<td>Uniq_pep</td>
<td>Spect_</td>
<td>Uniq_pep</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>counts</td>
<td></td>
<td>counts</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>O75694</td>
<td>Nuclear pore complex protein Nup155</td>
<td>21</td>
<td>17</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>P05556</td>
<td>Integrin beta-1</td>
<td>64</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>P14625</td>
<td>Endoplasmin</td>
<td>17</td>
<td>10</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>P26006</td>
<td>Integrin alpha-3</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>P17301</td>
<td>Integrin alpha-2</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>P15144</td>
<td>Aminopeptidase N</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>P02786</td>
<td>Transferrin receptor protein 1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>P16144</td>
<td>Integrin beta-4</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>P16070</td>
<td>CD44 antigen</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>O00592</td>
<td>Podocalyxin-like protein 1</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* P=0.0015

B. ML2 cells

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc #</th>
<th>Prot_desc</th>
<th>ML2_ManNAz</th>
<th></th>
<th>ML2_ManNA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spect_</td>
<td>Uniq_pep</td>
<td>Spect_</td>
<td>Uniq_pep</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>counts</td>
<td></td>
<td>counts</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>P05556</td>
<td>Integrin beta-1</td>
<td>75</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>P15144</td>
<td>Aminopeptidase N</td>
<td>21</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>O75694</td>
<td>Nuclear pore complex protein Nup155</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>P17301</td>
<td>Integrin alpha-2</td>
<td>17</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>P16144</td>
<td>Integrin beta-4</td>
<td>10</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>P14625</td>
<td>Endoplasmin</td>
<td>16</td>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>P02786</td>
<td>Transferrin receptor protein 1</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>Q02413</td>
<td>Desmoglein-2</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>P35613</td>
<td>Basigin</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>P26006</td>
<td>Integrin alpha-3</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* P<0.0001
abundant glycoproteins were present in far higher incidence from the ManNAz-treated group. We observed a similar enrichment by the ManNAz group when we employed emPIA score to rank the glycoproteins (Table 7A and Table 7B). These results clearly demonstrate a substantial increase in the selective capture of the most abundant known cellular glycoproteins via this methodology.

As a final assessment of the extent of enrichment of glycoproteins on the cell surface and in the extracellular compartment, we compared our targeted method to a non-selective global proteomic analysis of whole cell lysates of the same cell model. Whole cell lysates were prepared from N2 and ML2 cells, directly digested by trypsin and subjected to LC-MS/MS as described in materials and methods. Three nano-LC-MS/MS runs of tryptic digests showed that a significant number of total proteins (310 for N2 and 280 for ML2) could be detected utilizing this simple non-fractionation approach.

When analyzed for protein content, we found that 3.8% of the N2 proteome and 2.9% of the ML2 proteome identified from this global approach were annotated as glycoproteins using the UniProt database. In comparison, our targeted approach yielded 9.3% and 13.5% respectively (Fig. 28). Although the use of the UniProt database to determine if a protein is glycosylated is notoriously under-representative, a four-fold enrichment was revealed. Most striking was that 10 of 12 internal glycoproteins, found to be abundant and identified in the targeted approach, were not identified in the whole cell proteome approach.

We next conducted a comparative analysis of the two approaches with specific emphasis on which cellular compartment the proteins were derived. Figure 29 displays the UniProt-designated subcellular distribution of proteins identified from the whole cell
Table 7 Top 10 abundant glycoproteins identified in N2 and ML2 cells ranked by emPIA scores

A. N2 cells

<table>
<thead>
<tr>
<th>Acc #</th>
<th>prot_desc</th>
<th>N2_ManNAz</th>
<th>N2_ManNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P05556 Integrin beta-1</td>
<td>2.71</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>P14625 Endoplasmin</td>
<td>1.01</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>O75694 Nuclear pore complex protein Nup155</td>
<td>0.75</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>P02786 Transferrin receptor protein 1</td>
<td>0.31</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>O00592 Podocalyxin-like protein 1</td>
<td>0.28</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>P26006 Integrin alpha-3</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>P16070 CD44 antigen</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Q8TCT9 Minor histocompatibility antigen H13</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>P17301 Integrin alpha-2</td>
<td>0.24</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>P15144 Aminopeptidase N</td>
<td>0.23</td>
<td>0</td>
</tr>
</tbody>
</table>

* P=0.0319

B. ML2 cells

<table>
<thead>
<tr>
<th>Acc #</th>
<th>prot_desc</th>
<th>ML2_ManNAz</th>
<th>ML2_ManNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P05556 Integrin beta-1</td>
<td>3.54</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>P35613 Basigin</td>
<td>0.91</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>P15144 Aminopeptidase N</td>
<td>0.88</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>P14625 Endoplasmin</td>
<td>0.73</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>O75694 Nuclear pore complex protein Nup155</td>
<td>0.51</td>
<td>0.06</td>
</tr>
<tr>
<td>6</td>
<td>P02786 Transferrin receptor protein 1</td>
<td>0.46</td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>P16070 CD44 antigen</td>
<td>0.33</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>P27797 Calreticulin</td>
<td>0.33</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Q9UNF1 Melanoma-associated antigen D2</td>
<td>0.24</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Q13740 CD166 antigen</td>
<td>0.23</td>
<td>0</td>
</tr>
</tbody>
</table>

* P=0.00207
Figure 28 Improved capture of known glycoproteins. Identified proteins were cross-referenced to the UniProt Knowledgebase for known glycoproteins. The percent of total proteins that were matched to known glycoproteins is shown, in comparison to that typically determined by global proteomic approaches.
Figure 29 Subcellular location of identified proteins. The cellular location was determined by cross-referencing the annotations from UniProtKB. Each protein was designated with only one subcellular location. The protein classes are shown as percent of total protein on the x axis. This analysis was conducted for both the surface glycoprotein and global methodologies.
proteome and the cell surface glycoproteome. We observed that about 13.2\% of proteins from the whole cell analysis of N2 and ML2 were classified as being extracellular and membrane-bound. This result is consistent with studies using whole cell proteome approaches to examine breast cancer cell and ovarian cancer cell proteome. In contrast, our cell surface glycoprotein approach resulted in a proteome for N2 and ML2 with 25.9\% and 29.8\% of proteins being secreted or cell surface proteins, respectively. These cell surface proteins include CD markers, cell surface receptors, membrane transport, and cell adhesion proteins, such as CD44, CD166, integrins, anexins, calnexin, etc. In contrast, a small portion of proteins identified using our targeted approach was classified as having ER/golgi location (~2\%). A large portion of proteins were classified as intracellular (>50\%), either cytoplasm or nucleus, whereas 18 proteins for N2 and 13 proteins for ML2 remain unclassified. Sardana et al. conducted a cellular secretome analysis in which the conditioned medium of three prostate cancer cell lines (including PC3) were examined to identify just over 2000 proteins of which 12\% were classified as known extracellular proteins. We observed 32 proteins that were in common with the “secretome” study and that 85\% of these were classified as membrane proteins (Fig. 30). These findings underscore the value of our targeted strategy to capture and isolate cell surface glycoproteins.

4.3.7 Selective Identification of Differentially Expressed Cell surface Sialoglycoproteins

Currently less than 5\% of proteins are annotated as glycoproteins in the UniProt database. This is a striking number when current estimates suggest that more than
Figure 30 Overlap of our cell surface proteins with prostate cancer secretome. Venn diagram shows the overlap between cell surface proteins identified by our targeted glycoproteomic approach and prostate cancer secretome reported by Diamandis' group (Sardana, Ref.32). Since subcellular location classification in prostate cancer secretome is redundant, the combination of 329 extracellular and 504 membraneous proteins results in total 730 proteins.
50% of eukaryotic proteins are glycosylated. Therefore, we assembled the identified proteins that could be attributed to membrane-bound and extracellular compartments by UniProt database or predicted to contain transmembrane domains (SOSUI, TMHMM). These were further analyzed for the prediction of sites for N-linked glycosylation (NetNGlyc). Using these criteria, we compared the surface sialoglycoprotein expression profiles of the metastatic variants. Ultimately, 36 and 44 cell surface glycoproteins were uniquely identified in N2 and ML2 respectively. A list of these proteins, their UniProt accession number, and molecular function is presented in Tables 8 and 9. Known functions of these proteins include cell-cell or cell-matrix adhesion, protein transporter, signaling in response to cytokines and growth factors, induction of coagulation, and proteolysis.

5.3.8 Functional Mapping of Differential Expressed Sialoglycoproteins

Among those 36 differentially expressed cell surface glycoproteins in non-metastatic N2 cells (Table 8), many proteins, such as, complement decay-accelerating factor (CD55), Catenin beta-1 (CTNNB1), focal adhesion kinase 1 (PTK2), Myosin-VI (MYO6), A kinase anchor protein 1 (AKAP1) and symplekin (SYMPK), can positively or negatively regulate apoptosis pathway (Fig. 31). The several proteins, including catenin beta-1, catenin delta-1(CTNND1), disks large homolog 5 (DLG5) and symplekin, are involved in regulation of cell adhesion. The protein transports, Sec16A, Sec23B and Sec24C are involved in transport from the endoplasmic reticulum to the Golgi apparatus. Whereas coatomer subunit gamma (COPG) reversibly associates with Golgi non-clathrin-coated vesicles and protein YIPF5 can assist transport between endoplasmic reticulum
### Table 8 Cell surface N-linked glycoproteins differentially expressed in N2 Cells

<table>
<thead>
<tr>
<th>Spect counts</th>
<th>Uniq pep</th>
<th>Acc#</th>
<th>Gene Name</th>
<th>Protein Name</th>
<th>Predicted MW (Da)</th>
<th>Observed MW (Da)</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>P28290</td>
<td>SSFA2</td>
<td>Sperm-specific antigen 2</td>
<td>139,554</td>
<td>170,000</td>
<td>Actin binding</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q92667</td>
<td>AKAP1</td>
<td>A kinase anchor protein 1, mitochondrial</td>
<td>98,250</td>
<td>120,000</td>
<td>Protein binding</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q00341</td>
<td>HDLBP</td>
<td>Vigilin</td>
<td>141,995</td>
<td>140,000</td>
<td>Sterol metabolism</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q8NFJ5</td>
<td>GPRC5A</td>
<td>Retinoic acid-induced protein 3</td>
<td>40,624</td>
<td>300,000</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Q9Y678</td>
<td>COPG</td>
<td>Coatomer subunit gamma</td>
<td>98,967</td>
<td>90,000</td>
<td>ER-Golgi vesicle transport</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P50416</td>
<td>CPT1A</td>
<td>Carnitine O-palmitoyltransferase 1, liver isoform</td>
<td>88,995</td>
<td>300,000</td>
<td>Lipid metabolism</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>P35222</td>
<td>CTNNB1</td>
<td>Catenin beta-1</td>
<td>86,069</td>
<td>90,000</td>
<td>Cell adhesion; signal transduction</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O60716</td>
<td>CTNND1</td>
<td>Catenin delta-1</td>
<td>108,674</td>
<td>100,000</td>
<td>Cell adhesion; signal transduction</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P08174</td>
<td>CD55</td>
<td>Complement decay-accelerating factor</td>
<td>42,400</td>
<td>75,000</td>
<td>Complement activation</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Q7L2E3</td>
<td>DHX30</td>
<td>Putative ATP-dependent RNA helicase DHX30</td>
<td>134,938</td>
<td>130,000</td>
<td>ATP binding; ATP-dependent helicase activity</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O60610</td>
<td>DIAPH1</td>
<td>Protein diaphanos homolog 1</td>
<td>141,942</td>
<td>140,000</td>
<td>Receptor binding</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q8TDM6</td>
<td>DLG5</td>
<td>Disks large homolog 5</td>
<td>215,446</td>
<td>240,000</td>
<td>Cell adhesion; signal transduction</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>P50570</td>
<td>DNM2</td>
<td>Dynamin-2</td>
<td>98,345</td>
<td>95,000</td>
<td>Vesicular trafficking</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9BW60</td>
<td>ELOVL1</td>
<td>Elongation of very long chain fatty acids protein 1</td>
<td>32,755</td>
<td>300,000</td>
<td>Fatty acid biosynthesis</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O95864</td>
<td>FADS2</td>
<td>Fatty acid desaturase 2</td>
<td>52,340</td>
<td>300,000</td>
<td>Lipid metabolism</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q5397</td>
<td>PTK2</td>
<td>Focal adhesion kinase 1</td>
<td>119,956</td>
<td>120,000</td>
<td>Integrin-mediated signaling pathways</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>Q92616</td>
<td>GCN1L1</td>
<td>Translational activator GCN1</td>
<td>294,953</td>
<td>300,000</td>
<td>Translation activator</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Q8TCT9</td>
<td>HM13</td>
<td>Minor histocompatibility antigen H13</td>
<td>41,747</td>
<td>300,000</td>
<td>Proteolysis; protein binding</td>
</tr>
<tr>
<td>ID</td>
<td>Rep.</td>
<td>Accession</td>
<td>Protein Name</td>
<td>Molecular Weight</td>
<td>pI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------------</td>
<td>----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q86YT6</td>
<td>MIB1</td>
<td>E3 ubiquitin-protein ligase MIB1</td>
<td>112,461</td>
<td>110,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9UM54</td>
<td>MYO6</td>
<td>Myosin-VI</td>
<td>150,965</td>
<td>150,000</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>Q5SRE5</td>
<td>NUP188</td>
<td>Nucleoporin NUP188 homolog</td>
<td>198,369</td>
<td>220,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O75781</td>
<td>PALM</td>
<td>Paralemmin</td>
<td>42,221</td>
<td>60,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O00408</td>
<td>PDE2A</td>
<td>cGMP-dependent 3',5'-cyclic phosphodiesterase</td>
<td>107,360</td>
<td>50,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P13796</td>
<td>LCP1</td>
<td>Plastin-2</td>
<td>70,815</td>
<td>70,000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Q8TEM1</td>
<td>NUP210</td>
<td>Nuclear pore membrane glycoprotein 210</td>
<td>205,895</td>
<td>300,000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>O75400</td>
<td>PRPF40A</td>
<td>Pre-mRNA-processing factor 40 homolog A</td>
<td>109,022</td>
<td>120,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q63HN8</td>
<td>RNF213</td>
<td>RING finger protein 213</td>
<td>378,474</td>
<td>300,000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>O15027</td>
<td>SEC16A</td>
<td>Protein transport protein Sec16A</td>
<td>234,855</td>
<td>250,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q15437</td>
<td>SEC23B</td>
<td>Protein transport protein Sec23B</td>
<td>87,393</td>
<td>80,000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>P53992</td>
<td>SEC24C</td>
<td>Protein transport protein Sec24C</td>
<td>119,779</td>
<td>120,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P46977</td>
<td>STT3A</td>
<td>Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A</td>
<td>81,104</td>
<td>300,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q92797</td>
<td>SYMPK</td>
<td>Symplekin</td>
<td>141,915</td>
<td>140,000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Q9NXF1</td>
<td>TEX10</td>
<td>Testis-expressed sequence 10 protein</td>
<td>106,349</td>
<td>90,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O94901</td>
<td>UNC84A</td>
<td>Protein unc-84 homolog A</td>
<td>90,806</td>
<td>300,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9H1C4</td>
<td>UNC93B1</td>
<td>Protein unc-93 homolog B1</td>
<td>66,930</td>
<td>300,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q969M3</td>
<td>YIPF5</td>
<td>Protein YIPF5</td>
<td>28,370</td>
<td>40,000</td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
- Notch signal pathway, ubiquitination regulation
- Protein transport; endocytosis
- Component of the nuclear pore complex (NPC)
- Cell shape
- Signal transduction
- T-cell activation; intracellular protein transport
- mRNA processing
- ER-Golgi transport
- N-linked glycosylation
- Cell adhesion; mRNA processing
- Unknown
- Protein binding; cytoskeletal anchoring
- Immunity; intracellular protein transport
- ER-Golgi transport
**Table 9 Cell surface N-linked glycoproteins differentially expressed in ML2 Cells**

<table>
<thead>
<tr>
<th>Spect_counts</th>
<th>uniqu_pep</th>
<th>Acc#</th>
<th>Gene Name</th>
<th>Protein Name</th>
<th>Predicted MW (Da)</th>
<th>Observed MW (Da)</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>P08195</td>
<td>SLC3A2</td>
<td>4F2 cell-surface antigen heavy chain (CD98)</td>
<td>58,023</td>
<td>80,000</td>
<td>Cell growth; amino-acid transport; carbohydrate metabolism</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P49748</td>
<td>ACADVL</td>
<td>Very long-chain specific acyl-CoA dehydrogenase</td>
<td>70,745</td>
<td>70,000</td>
<td>Fatty acid metabolism</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>O95573</td>
<td>ACSL3</td>
<td>Long-chain-fatty-acid-CoA ligase 3</td>
<td>81,338</td>
<td>85,000</td>
<td>Fatty acid metabolism</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Q5T9A4</td>
<td>ATAD3B</td>
<td>ATPase family AAA domain-containing protein 3B</td>
<td>73,098</td>
<td>70,000</td>
<td>ATP binding</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>P35613</td>
<td>BSG</td>
<td>Basigin</td>
<td>42,573</td>
<td>50,000-70,000</td>
<td>Cell surface receptor</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>P27797</td>
<td>CALR</td>
<td>Calreticulin</td>
<td>48,283</td>
<td>60,000</td>
<td>Molecular calcium binding chaperone</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q01518</td>
<td>CAP1</td>
<td>Adenyl cyclase-associated protein 1</td>
<td>52,222</td>
<td>60,000</td>
<td>Cell polarity maintenance; signal trasduction</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q96A33</td>
<td>CCDC47</td>
<td>Coiled-coil domain-containing protein 47</td>
<td>56,123</td>
<td>65,000</td>
<td>Protein binding</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>Q9H5V8</td>
<td>CDCP1</td>
<td>CUB domain-containing protein 1</td>
<td>94,241</td>
<td>130,000</td>
<td>Cell adhesion; tumor progression and metastasis</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>P12109</td>
<td>COL6A1</td>
<td>Collagen alpha-1(VI) chain</td>
<td>109,602</td>
<td>130,000</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>A8TX70</td>
<td>COL29A1</td>
<td>Collagen alpha-5(VI) chain</td>
<td>291,796</td>
<td>120,000</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q15046</td>
<td>KARS</td>
<td>Lysyl-tRNA synthetase</td>
<td>68,461</td>
<td>70,000</td>
<td>Protein biosynthesis</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9BW27</td>
<td>NUP85</td>
<td>Nuclear pore complex protein Nup85</td>
<td>75,826</td>
<td>75,000</td>
<td>Protein transport; mRNA transport</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9NZ01</td>
<td>TECR</td>
<td>Synaptic glycoprotein</td>
<td>36,410</td>
<td>300,000</td>
<td>Lipid/steroid biosynthesis Cell adhesion</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q02413</td>
<td>DSG1</td>
<td>Desmoglein-1</td>
<td>114,670</td>
<td>70,000</td>
<td>Angiogenesis; gluconeogenesis</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P06744</td>
<td>GPI</td>
<td>Glucose-6-phosphate isomerase</td>
<td>63,335</td>
<td>65,000</td>
<td>Unknown</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>A6NKF9</td>
<td>GPR89C</td>
<td>Putative protein GPR89C</td>
<td>36,760</td>
<td>300,000</td>
<td>Calcium ion binding</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P43304</td>
<td>GPD2</td>
<td>Glycerol-3-phosphate dehydrogenase</td>
<td>81,296</td>
<td>80,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acetolactate synthase-like protein</td>
<td>68,452</td>
<td>70,000</td>
<td>Transferase activity; magnesium ion binding Ion transport</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q14721</td>
<td>KCNB1</td>
<td>Potassium voltage-gated channel subfamily B member</td>
<td>96,672</td>
<td>60,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q8TF66</td>
<td>LRRC15</td>
<td>Leucine-rich repeat-containing protein 15</td>
<td>65,268</td>
<td>75,000</td>
<td>Protein binding</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P29966</td>
<td>MARCKS</td>
<td>Myristoylated alanine-rich C-kinase substrate</td>
<td>31,707</td>
<td>80,000</td>
<td>Substrate for protein kinase C; binds calmodulin, actin, and synapsin</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Q7Z434</td>
<td>MAVS</td>
<td>Mitochondrial antiviral-signaling protein</td>
<td>57,063</td>
<td>75,000</td>
<td>Innate immune defense against viruses</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q8NHP6</td>
<td>MOSPD2</td>
<td>Motile sperm domain-containing protein 2</td>
<td>60,051</td>
<td>60,000</td>
<td>Unknown</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q6PIF6</td>
<td>MYO7B</td>
<td>Myosin-VIIb</td>
<td>243,612</td>
<td>250,000</td>
<td>Actin binding; motor activity</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>O15118</td>
<td>NPC1</td>
<td>Niemann-Pick C1 protein</td>
<td>144,868</td>
<td>250,000</td>
<td>Intracellular trafficking of cholesterol Protein transport; mRNA transport</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>Q9UKX7</td>
<td>NUP50</td>
<td>Nucleoporin 50 kDa</td>
<td>50,512</td>
<td>60,000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q6V1P9</td>
<td>DCHS2</td>
<td>Protocadherin-23</td>
<td>323,744</td>
<td>55,000</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9H307</td>
<td>PNN</td>
<td>Pinin</td>
<td>81,179</td>
<td>130,000</td>
<td>Cell adhesion; mRNA processing</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>P14923</td>
<td>JUP</td>
<td>Junction plakoglobin</td>
<td>82,434</td>
<td>85,000</td>
<td>Junctional plaque formation</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q15063</td>
<td>POSTN</td>
<td>Periostin</td>
<td>93,883</td>
<td>70,000</td>
<td>Cell adhesion.</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>P18031</td>
<td>PTPN1</td>
<td>Tyrosine-protein phosphatase non-receptor type 1</td>
<td>50,505</td>
<td>55,000</td>
<td>Regulation of insulin signal pathways</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>P46060</td>
<td>RANGAP1</td>
<td>Ran GTPase-activating protein 1</td>
<td>63,958</td>
<td>80,000</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9GZS1</td>
<td>POLR1E</td>
<td>DNA-directed RNA polymerase 1 subunit RPA49</td>
<td>54,441</td>
<td>55,000</td>
<td>rRNA transcription</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>P04843</td>
<td>RPN1</td>
<td>Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 1</td>
<td>68,641</td>
<td>70,000</td>
<td>N-linked glycosylation</td>
</tr>
<tr>
<td>#</td>
<td>Replicates</td>
<td>ID</td>
<td>Description</td>
<td>MW (kDa)</td>
<td>MOL (kDa)</td>
<td>Function</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9Y265</td>
<td>RUVBL1 RuvB-like 1</td>
<td>50,538</td>
<td>60,000</td>
<td>Cell cycle; cell division; growth regulation</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q96QD8</td>
<td>SLC38A2 Sodium-coupled neutral amino acid transporter 2</td>
<td>56,332</td>
<td>300,000</td>
<td>Amino-acid transport</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q9Y512</td>
<td>SAMM50 Sorting and assembly machinery component 50</td>
<td>52,342</td>
<td>60,000</td>
<td>Mitochondrial outer membrane translocase complex assembly</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>P35498</td>
<td>SCN1A Sodium channel protein type 1 subunit alpha</td>
<td>230,876</td>
<td>300,000</td>
<td>Ion transport</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>O15269</td>
<td>SPTLC1 Serine palmitoyltransferase 1</td>
<td>53,281</td>
<td>50,000</td>
<td>Lipid metabolism</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>O00186</td>
<td>STXBP3 Syntaxin-binding protein 3</td>
<td>68,633</td>
<td>70,000</td>
<td>Vesicle docking in exocytosis</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>O94826</td>
<td>TOMM70A Mitochondrial import receptor subunit TOM70</td>
<td>68,096</td>
<td>75,000</td>
<td>Receptor for protein import into mitochondria</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Q15849</td>
<td>SLC14A2 Urea transporter, kidney</td>
<td>102,213</td>
<td>70,000</td>
<td>Urea transporter.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>P38606</td>
<td>ATP6V1A V-type proton ATPase catalytic subunit A</td>
<td>68,660</td>
<td>70,000</td>
<td>ATP synthesis coupled proton transport</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 31** Functional pathway analysis of glycoproteins unique to N2 cells. Ingenuity Pathway Analysis (IPA) was employed to functionally map the cell-surface glycoproteins unique to N2 cells. The network diagram shows the biological functions that these glycoproteins are associated with, in the context of disease. Apoptosis involved in tumor development has been listed as the top function, as reflected by the number of glycoproteins (highlighted in blue) connected with it.
and Golgi. Other candidates are involved in different functions, such as T cell activation (Plastin-2), N-oligosaccharyl transferase (dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A), mRNA and protein transporter (nucleoporin NUP188 homolog and nuclear pore membrane glycoprotein 210), fatty acid beta-oxidation (carnitine O-palmitoyltransferase 1, liver form), fatty acid biosynthetic process (elongation of very long chain fatty acids protein 1), unsaturated fatty acid biosynthetic process (fatty acid desaturase 2), and RNA splicing, mRNA processing (pre-mRNA-processing factor 40 homolog A), etc.

Among those 44 differentially expressed cell surface glycoproteins in metastatic ML2 cells (Table 9), a large number of proteins are involved in cell movement, cell-cell signaling and interaction, and cellular growth and proliferation. There are 11 proteins involved in cell movement (Fig. 32). Notably, 6 proteins, Basigin (BSG), Periostin (POSTN) and Glucose-6-phosphate isomerase (GPI), Calreticulin (CALR), Leucine-rich repeat-containing protein 15 (LRRC15), Myristoylated alanine-rich C-kinase substrate (MARCKS), are involved in the invasion of tumor cell lines. Other 5 proteins, Junction plakoglobin (JUP), Tyrosine-protein phosphatase non-receptor type 1 (PTPN1), Niemann-Pick C1 protein (NPC1), Sodium-coupled neutral amino acid transporter 2 (SLC38A2) and Adenylyl cyclase-associated protein 1 (CAP1), are involved in migration or movement of eukaryotic cells, embryonic cells or cancer cells. Of above proteins, SLC38A2, JUP, PTPN1, POSTN, CALR, BSG, and Pinin (PNN), CUB domain-containing protein 1 (CDCP1), DNA-directed RNA polymerase I subunit RPA49 (POLR1E), Collagen alpha-1(VI) chain (COL6A1), are involved in growth, colony formation of eukaryotic cells, leukocyte, fibroblast or cancer cell lines. Other
Figure 32 Functional pathway analysis of glycoproteins unique to ML2 cells. Ingenuity Pathway Analysis (IPA) was employed to functionally map the cell-surface glycoproteins unique to ML2 cells. The network diagram shows the biological functions that these glycoproteins are associated with, in the context of disease. Cell movement, invasion and migration have been listed as the top functions, as reflected by the number of glycoproteins (highlighted in blue) connected with them.
glycoproteins differentially expressed in ML2 are associated with cell signaling (Syntaxin-binding protein 3), cell assembly (Ran GTPase-activating protein 1), nuclear protein export (Nucleoporin 50 kDa), innate immune defense against viruses (Mitochondrial antiviral-signaling protein), et al. Interestingly, several proteins, Very long-chain specific acyl-CoA dehydrogenase (ACADVL), Long-chain-fatty-acid--CoA ligase 3 (ACSL3), Synaptic glycoprotein (TECR), Serine palmitoyltransferase 1 (SPTLC1) and Niemann-Pick C1 protein, are involved in unique lipid/steroid metabolism.

4.4 Discussion

Metastatic spread of cancer involves tumor cell extravasation and subsequent invasion of surrounding tissue, all processes that are tightly regulated by cell surface mechanisms \(^{123,140}\). The identification of surface-accessible glycoproteins, which are altered in the metastatic process, will reveal the underlying biological events that support cancer metastasis, may facilitate the discovery of new biomarkers for cancer diagnosis, prognosis, and help define new therapeutic targets. In this study, a novel surface glycoprotein targeted strategy, based on metabolic labeling and click chemistry, was developed for purification and characterization of sialoglycoproteins. Metabolic labeling of glycans with sugar analogues and subsequent probe capture via Click chemistry have been applied to cellular microscopy \(^{256}\). However, to our knowledge, our data is the first to demonstrate its application to targeted proteomics.

Click chemistry specifically refers to a bioorthogonal ligation event between azide and alkyne \(^{267}\). The position of azide-alkyne pair can be exchanged, that is, the same
reaction can be carried out using the alkyne as the chemical reporter. Wong and coworkers reversed the polarity of the reagents, using an alkynyl ManNAc derivative for metabolic labeling of cultured cells and CuAAC-mediated reaction with an azide-functionalized probe for imaging of sialylated glycoproteins. The primary advantage of Click chemistry over the Staudinger ligation is its faster rate, high-sensitivity, and high-selectivity. Our flow cytometry results and confocal images confirmed the high selectivity/specificity of click chemistry for labeling cell surface sialoglycoconjugates. In conjunction with LC-MS/MS, we applied this method to comparatively analyze the cell surface sialoglycoproteome of non-metastatic (N2) and metastatic (ML2) prostate cancer cell lines. Our findings reveal substantial changes in the abundance of proteins that have previously been associated with metastatic spread of cancer and the discovery of some new cell surface sialoglycoprotein with potential roles in metastasis. The enrichment for surface glycoproteins demonstrated by this method provides proof of concept of the utility of this approach in the discovery of sialic acid-containing glycoproteins of disease significance.

Although many cellular proteins are glycosylated, they are rarely identified by global proteomics approaches. Targeted labeling and enrichment of specific groups of glycoproteins simplifies the complexity of a cellular proteome and serves to reduce the dynamic range of protein expression. We demonstrate that the application of the described method is effective for cell surface glycoprotein identification. A substantial enrichment of cell surface glycoproteins of low abundance (based on spectral counts) was observed. These included adhesion proteins implicated in cancer development such as proocadherin-23 and CUB-domain containing protein 1, protein binding and signal
transduction molecules such as leucine-rich repeat-containing protein 15 and focal adhesion kinase 1.

Our approach resulted in a proteome that was enriched to approximately 29% for cell surface or excreted proteins. This compares favorably with other proteomic approaches using cell surface biotinylation and MudPIT analysis\textsuperscript{264,268}, in which cell surface protein enrichment of \textasciitilde30% was reported for human HT-1080 fibrosarcoma cell line and human CaOV3, OVCAR3, and ES2 ovarian cancer cell lines. More recently, on-bead tryptic digestion followed by the specific release of N-linked glycosylated peptides via peptide-N-glycosidase F (PNGase F) and LC-MS/MS has been reported to successfully identify N-linked glycopeptides\textsuperscript{173,178,180}. Because the authors assumed that all proteins identified by the approach were glycoproteins there was no report of target efficiency. Aebersold and co-workers combined this system with the periodate oxidation/hydrazide method for the specific analysis of the cell surface membrane glycoproteome\textsuperscript{173}. Although successful, this technology is still challenging due to a potential for false positive identification that arises from the dependence on single peptide analysis, 1 Da glycan shift and competition with natural Asn to Asp conversion. To eliminate the false positive identification of the N-linked peptides, MS data acquisition from more intense isotopic peaks and manual inspection of mass spectra are required. Because N-linked glycosylation predominates in secreted and cell surface proteins, we would suggest that combining a selective click chemistry approach as we have described with this gel-free system for the analysis of the cell surface glycoproteome would be a successful alternative strategy.
The significant number of proteins annotated as nuclear or cytoplasmic that were present among our cell surface sialoglycoproteome is noteworthy. To some extent, occurrence of these predominantly intracellular proteins may be due to their high abundance. Despite our efforts to optimize the labeling protocol and increase wash stringency, some proteins classified as intracellular were still precipitated by the SAν resin. High-salt (1.5M NaCl) and high-pH (0.1M Na₂CO₃) buffers have been reported to reduce such contamination and may reduce this background. Another possible reason for the contamination is limited cell death during the click reaction. Click reaction requires copper as a catalyst, which is toxic to the cell. Mammalian cells can survive low concentrations (below 500 μm) of copper(I) for 1 hour. However, certain cell death was still observed under our optimized click conditions. Fortunately, a new generation click chemistry, called stain-promoted cycloaddition, has been developed and has been shown to maximally preserve cell viability through bypassing the requirement of copper. Bertozzi’s group applied copper-free click chemistry for dynamic in vivo imaging, claiming comparable kinetics to the copper-catalyzed reaction and speed with no apparent toxicity. We have similarly observed identical labeling performance and selectivity with much less cytotoxicity when cells were labeled with these copper-free reagents (data not shown). Improving cell viability could improve proteome selectivity and would also allow for kinetic experiments post-reaction via click. Finally, one must consider existing evidence for the occurrence of cytoplasmic proteins bound to the extracellular membrane or exposed on the cell surface, particularly in aggressive cancer cells. Thus, a portion of these “cytoplasmic” proteins may be glycosylated and surface bound.
The cell surface glycoproteomics approach we describe can also be expanded to other types of glycoprotein targets. For instance, in addition to increased sialylation, other types of cancer-related glycosylation changes have been reported, such as increased cell surface fucosylation, the addition of polygalactosamine units, higher-ordered branching of N-linked glycans and mucin-type O-linked glycosylation. Currently, in addition to tetraacelated N-azidoacetyl-D-mannosamine for sialic acid, several click pair labeled molecules targeted for these cancer-related glycosylation alterations are commercially available. For example, tetraacetyl fucose alkyne for fucosylate proteins, tetraacelated N-azidoacetylgalactosamine for O-linked glycoproteins and tetraacelated N-azidoacetylgluctosamine for O-GlcNAC-modified proteins are readily available. In addition to cancer, this method is clearly also valuable for analysis of glycosylation changes in immune deficiencies, hereditary disorders, and neurodegenerative diseases.

In summary, we report a new glycoproteomic strategy for the analysis of specific subclasses of cell surface glycoproteins. The approach combines selective metabolic labeling of cell surface glycans, affinity purification of sialylated proteins, SDS-PAGE separation, and subsequent LC-MS/MS for protein identification. A total of 80 proteins differentially expressed between cell lines were uniquely identified, including many integral plasma membrane glycoproteins that were not reported by previous proteomic analyses. Given the importance of the cell surface glycoproteins for drug design, biomarker discovery, and the regulation of cellular behaviors, the described protocol would help expedite the discovery of new cancer biomarker and therapeutic targets in diseases and delineation of signal transduction pathways.
CHAPTER V

Aim III: VERIFICATION OF EXPRESSION OF CELL-BASED BIOMARKERS IN CLINICAL SAMPLES

5.1 Introduction

Biomarker development is a multistep process entailing several essential components: discovery, qualification, verification, research assay optimization, clinical validation and commercialization. With the application of high throughput genomic and proteomic technologies to biological samples, a large number of biomarker candidates with apparent differential expression have been reported as candidates for novel cancer detection assays. However, few biomarkers have been validated and successfully translated into clinical tests. This is in large part due to the incredible mismatch between the large numbers of biomarker candidates and the paucity of reliable assays, the cost, and the methods for validation studies. Therefore, the critical step in biomarker development lies in screening candidate lists to find those few with performance characteristics that merit the effort and expense of full validation.

Biomarker prioritization is often performed on the basis of case biological knowledge, the candidate’s performance in discovery datasets, and reagent availability. Depending on the stringency of the discovery process and the number of candidates identified, verification may be undertaken on only a subset of qualified candidates. In verification, independent methods should be employed to provide better quantification of candidates than is typical for discovery. In addition, verification, like validation, should be performed on clinical samples that closely represent the population in which a final
clinical test would be deployed. The expected output of prioritization and verification is a small number of highly credentialed candidates suitable for validation studies.

Having identified a panel of differentially expressed cell surface sialoglycoproteins on a syngeneic metastatic PCa model by our targeted glycoproteomic approach in Aim II study, we further prioritized and verified these metastasis-associated candidates for additional rationales. Previously, we used an *in vitro* cell culture model system. Cultured PCa cell lines are valuable tools to replicate the progression of the disease, especially for complex metastatic events which are difficult to analyze in clinical specimens. They are typically easy to establish, maintain and manipulate. However, cell line model systems lack the biological complexity and variability that is inherent in real patient samples. In addition, the bottom-up strategy employed in our previous study has the inherent weakness of unambiguously reassembling the peptide sequence into their precursor proteins. Therefore, it is necessary to verify their expression and determine their clinical utility to warrant further investment in clinical validation studies. Specifically, our prioritization was based on the following criteria: (1) candidates are significantly differentially expressed in PCa metastatic variants; (2) bioinformatics analysis indicates the cell surface location and potential function in cancer invasion and metastasis; (3) literature mining shows an inferred role in cancer progression; (4) specific antibodies are available since the immunoassays remain the gold standard for ultimate clinical application of validated biomarkers. Using this criteria, we selected two glycoprotein candidates from our list, CUB domain containing protein 1 (CDCP1) and basigin, for downstream experiments (*Fig. 33*). Here we present the important findings for CDCP1 glycoprotein.
Figure 33 The work-flow of verification of cell-based biomarkers. After candidate glycoproteins are identified, we prioritize them and confirm the differential expression in our cell line models using immunoassays. We also evaluate their expression in clinical samples including PCa tissues and Expression Prostate Secretion (EPS) urines with different aggressiveness.
CDCP1 is one of the candidates that were preferentially expressed by the highly metastatic ML2 cells. CDCP1 is a type I single transmembrane protein also known as subtractive immunization associated 135 kDa (SIMA135)\textsuperscript{280}, gp140\textsuperscript{281}, and transmembrane and associated with Src kinases (Trask)\textsuperscript{282}, and has been assigned the cluster of differentiation (CD) designation CD318\textsuperscript{283}. Physiologically, CDCP1 is expressed by stem or progenitor cells in the hematopoietic, mesenchymal, and neural tissues\textsuperscript{284,285}. It is also overexpressed in a number of solid tumors including breast\textsuperscript{286}, colon\textsuperscript{287,288}, kidney\textsuperscript{289}, pancreatic\textsuperscript{290}, and lung cancers\textsuperscript{291}. To date, its biological function has not been well defined. Some experimental data support the role of CDCP1 in cancer progression\textsuperscript{292,293}, ECM degradation\textsuperscript{290}, anchorage-independent signaling in epithelial cells\textsuperscript{294}, and cancer cell resistance to anoikis\textsuperscript{292}, a form of apoptosis triggered by the loss of cell survival signals generated from the interaction of cells with the extracellular matrix. In PCa, although the potential of CDCP1 as a therapeutic target as cancer treatment has been evaluated by an anti-CDCP1 monoclonal antibody generated from a phage display combinatorial antibody library\textsuperscript{295}, its expression and potential clinical significance have not been systematically studied.

In Aim III, we comprehensively examined CDCP1 protein expression in a battery of prostate cancer cell lines and clinical samples by applying multiple immuno-based methods. We also characterized the glycosylation status of CDCP1 by employing a series of glycosidases and glycosylation inhibitors. Importantly, we evaluated the potential clinical significance of CDCP1 expression in PCa patient samples, including surgical PCa tissues and Expressed Prostatic Secretion (EPS) urine fluids. Overall, our
studies suggest that CDCP1 is dysregulated in prostate cancer and it might be a useful diagnostic marker and therapeutic target for PCa patients.

5.2 Materials and Methods

Antibodies and Reagents — The following CDCP1-specific antibodies were used in the study: mAb 41-2, generated by subtractive immunization with a nonmetastatic population versus metastatic population of epidermoid HEp-3 carcinoma\textsuperscript{280}, and kindly provided by Dr James P. Quigley (The Scripps Research Institute, CA, USA); rabbit antibody CS4115, generated against the COOH-terminal peptide of CDCP1 (Cell Signaling, Danvers, MA, USA); and goat antibody AF2666 recognizing the extracellular domain of CDCP1 (R&D Systems, Minneapolis, MN, USA). Murine anti-β-actin (#612656) and anti-mAb (#610524) were provided by BD PharMingen (San Diego, CA, USA). Murine anti-Golgi 58 kD protein (ab27043) was purchased from Abcam (Cambridge, MA, USA). IRDye 700 or IRDye 800 conjugated secondary reagents and Alexa Fluor 488 or Alexa Fluor 594 conjugated secondary reagents were from Li-COR Biosciences (Lincoln, PA, USA) and Invitrogen (Carlsbad, CA, USA), respectively. Control immunoglobulins (IgGs) were from Vector lab (Burlingame, CA, USA). Complete\textsuperscript{TM} protease inhibitor cocktail and Halt\textsuperscript{TM} protease inhibitor cocktail were from Roche Applied Sciences (Indianapolis, IN, USA) and Pierce (Rockford, IL, USA). PNGase F, Neuraminidase, and Endoglycosidase H were from New England Biolabs (Beverly, MA, USA). High capacity streptavidin agarose resin and protein A/G PLUS-agarose were from Thermo Scientific (Rockford, IL, USA) and Santa Cruz Biotechnology (Santa Cruz, CA, USA). Dulbecco's modified Eagle media (DMEM) and fetal bovine serum (FBS) were from
Invitrogen. Antibiotics-antimycotic, Click-iT™ ManNAz metabolic glycoprotein reagent, Click-iT™ Biotin Protein Analysis Detection Kit, propidium iodide, TO-PRO3, 4-12% NuPAGE® Bis-Tris gels, and lithium dodecyl sulfate (LDS) buffer were from Invitrogen. 2 × Laemmli buffer, 7.5% Criterion® Tris-HCl Gel, Precision protein standards, and Bio-Safe coomassie blue were from Bio-Rad (Hercules, CA, USA). All other chemicals were from Sigma (St. Louis, MO, USA).

**Cell Lines** — Human prostate bone metastasis PC3, prostate carcinoma lymph node metastasis LNCaP, prostate carcinoma brain metastasis DU145, and prostate carcinoma xenograft 22Rv1 cell lines were obtained from the American Type Culture Collection (ATCC). PC3 two sublines denoted PC3-N2 (non-metastatic) and PC3-ML2 (highly metastatic) were selected based on differential invasiveness and ability to form metastatic tumors in SCID mice as described and kindly provided to us by Dr. Stearns’ group at Drexel University. The lineage-related LNCaP sublines C4, C4-2 and C4-2B were resulted from passage and hormone manipulation of LNCaP cells in vivo. They mimic the progression of PCa from androgen sensitive (LNCaP and C4) to androgen refractory (C4-2 and C4-2B), as well as the increasingly malignant phenotype from primary tumor (C4-2) to bone metastases (C4-2B) in a nude mouse host. These cells were developed by Dr. Chung’s group and provided to us via ViroMed Labs. HPV-18 immortalized prostate epithelial cell line RWPE-1 retains some “normal” cell attributes such as responsive to growth factors and androgen. The cells were exposed to mutagen and sublines WPE1-NB14 and WPE1-NB26 were derived with variable phenotype. The WPE1-NB14 and NB26 lines are both anchorage independent but differ in their
invasiveness. The WPE1-NB14 is moderately invasive in vitro whereas NK26 has a high invasive ability. These sublines were developed by Dr. Williams' group at the University of Michigan and purchased from ATCC. Another prostate carcinoma lymph node metastasis PacMetUT1 cell line was obtained from the University of Texas Health Science Center at San Antonio. All of the cells were cultured in the growth medium and passaged according to the developer's and manufacturer's instructions.

Expressed Prostatic Secretion (EPS) Urine Sample Preparation — EPS Urine samples were collected from men reporting to the clinic for a prostate biopsy procedure at the Urology of Virginia, Sentara Medical Group and Eastern Virginia Medical School between 2007 and 2009, under the informed consent and Institutional Review Board approved protocols as described. Briefly, following a gentle prostate massage during a digital rectal examination (DRE), the first voided urine (10-20 ml) containing the EPS was obtained from each individual. Then, they were centrifuged to remove the cell pellet/sediment, and the supernatant was aliquoted and stored at -80°C. In this study, individual EPS urine sample pools representative of non-cancer (n=15), low-grade PCa (n=16), and high-grade PCa (n=14) conditions were compiled. Non-cancer cases were defined as being biopsy negative for cancer, low-grade PCa samples were exclusively a Gleason grade of 6 (3+3), and high-grade PCa samples were with a Gleason score of 8 and above, predominantly 9 or 10. Each pool was concentrated to ~200 fold using an Amicon Ultra spin filter (10 kDa and 3 kDa cutoff; Millipore) and the aliquots were stored at -80°C. Protein concentration was measured by Bradford assay.
**Flow Cytometry Analysis** — Confluent cultures of N2 and ML2 cells were harvested from the culture dishes by non-enzyme dissociation buffer, washed with 2% FBS in PBS, and incubated with mouse-anti-CDCP1 mAb 41-2 antibody (6 µg/ml) for 1 h at 4°C. After two washes, cells were incubated with Alexa Fluor 488-conjugated goat-anti-mouse antibody (2 µg/ml) for 30 min at 4°C. Before subjected to flow cytometry analysis, cells were treated with propidium iodide (1 µg/ml) in 500 µl 2% FBS in PBS for 20 min at 4°C. Cells stained without the primary antibody were used as negative controls. Data was acquired by FACScalibur and analyzed by Flowjo software.

**Immunofluorescence Staining of Cultured Cell Lines** — Cells were seeded onto 6-well plates containing glass coverslips and cultured in growth medium to 80% confluence. They were washed with ice-cold PBS and fixed in 4% paraformaldehyde for 15 min at room temperature. For intracellular staining, the cells were permeabilized with 0.5% Triton X-100 in PBS for 10 min at room temperature or ice-cold 100% methanol for 10 min at -20°C. Cells were subsequently washed twice with PBS, blocked with 3% BSA in PBS for 1 h at room temperature. The cells were then incubated with mouse anti-CDCP1 mAb 41-2 (6 µg/ml) or rabbit-anti-CDCP1 CS4115 (1:100 dilution) polyclonal antibodies in the blocking buffer overnight at 4°C. After three washes, the cells were incubated with Alexa Fluor 488 or 594-conjugated secondary antibodies (2 µg/ml, goat anti-mouse or goat anti-rabbit IgG, respectively) in the blocking buffer for 1 h at room temperature. Nuclei were stained with TO-PRO3. For double staining with click labeling, cells were grown in medium supplemented with 20 µM AC4ManNAZ or ManNAc for 2 days, fixed as above, subjected to click reaction solution (25 µl biotin-
alkyne, 5 µl CuSO4, 5 µl and 10 µl for component D and E, 65 µl PBS, total 100 µl) for 30 min at room temperature. Cells were then permeabilized and incubated with anti-CDCP1 antibodies as above. The next day, cells were stained with Alexa Fluor 488-conjugated streptavidin (2 µg/ml) and Alexa Fluor 594-conjugated CDCP1 secondary antibody for 1 h at room temperature. For double staining with the ER and Golgi markers, cells were fixed and permeabilized as above, incubated with mouse-anti-calnexin (2.5 µg/ml) or anti-Golgi 58kDa protein (5 µg/ml) antibodies, and rabbit-anti-CDCP1 CS4115 (1:100 dilution) overnight at 4°C. The following day, cells were stained with Alexa Fluor 488 or 594-conjugated secondary antibodies as above. The cells were finally washed twice, mounted to the slides with VectorShield medium (Vector Labs, Burlingame, CA, USA), and sealed with nail polish. Cells treated in a similar way but in the absence of primary antibodies were included as negative controls. Labeled cells were visualized and photographed using a Zeiss LSM510 META confocal imaging system at the appropriate excitation/emission wavelengths (Carl Zeiss, Thornwood, NY, USA).

Tissue Samples and Immunofluorescence Staining — O.C.T.-embedded frozen PCa tissues from patients who underwent radical prostatectomy were obtained from our Virginia Prostate Center. Cryosectioning was done on a Microm HM 505E cryostat at −20°C. A serial cryosection (6 µm) was stained with Hematoxylin and Eosin (H&E) as a guide and analyzed by a pathologist to determine tissue morphology. Two additional serial sections were mounted on Superfrost microscope slides and used for immunofluorescence staining. Tissue sections were fixed immediately after cutting in 4% paraformaldehyde in PBS for 15 min, and permeabilized with 0.5% Triton X-100 in PBS
for 10 min. They were then blocked with 3% BSA in PBS for 1 h and incubated with mouse-anti-CDCP1 mAb 41-2 antibody (6 μg/ml) overnight at 4°C. After three washes, the cells were incubated with Alexa Fluor 488-conjugated secondary goat-anti-mouse antibody (2 μg/ml) in the blocking buffer for 1 h at room temperature. Nuclei were stained with propidium iodide (PI) (0.1 μg/ml) for 3 min before mounting and sealing. An adjacent section treated similarly but in the absence of primary antibody was used as the negative control. Slides were viewed and imaged by an Olympus BX51 fluorescent microscope fitted with a DP70 digital camera and a Zeiss LSM510 META confocal imaging system.

*Deglycosylation Assays* — To elucidate glycan components of CDCP1, enzymatic deglycosylation experiments were performed. Before PNGase F and Endo H treatment, cell lysates (20 μg) from PC3, N2, and ML2 cells were denatured and reduced for 10 min at 100°C in 0.5% SDS and 1% 2-mercaptoethanol. For PNGase F digestions, samples were adjusted to 1% Nonidet P 40 and 50 mM sodium phosphate, pH 7.5, and then incubated with PNGase F (1,000 NEB units) overnight at 37°C. For Endo H digestions, samples were adjusted to 50 mM sodium citrate, pH 5.5. After addition of Endo H (1,000 NEB units), the mixtures were incubated overnight at 37°C. Neuraminidase reactions were incubated overnight at 37°C in 50 mM sodium acetate, pH 6.0 and contained 50 NEB units neuraminidase per 20 μg of protein. Fetuin was used as the positive controls for PNFase F, Endo H, and neuraminidase deglycosylation. For drug treatment, near-confluent cells were washed twice with PBS and cultivated for 23 h in fresh culture media in the absence or presence of 5 μg/ml tunicamycin or 1 μg/ml swainsonine. Cells
were then washed twice before addition of M-PER lysis buffer containing 1x complete protease inhibitor. The cell lysates were collected with a rubber scraper, passed through 26-gauge needles, and sonicated by 3 times for 10 seconds. The whole cell lysates were centrifuged at 14,000g for 15 min to remove cell debris and unbroken cells. The supernatant was saved and protein concentration was tested by the BCA protein assay kit.

*Western blot analysis* — Whole cell lysates were collected in the M-PER lysis buffer containing 1x protease inhibitor cocktail. Protein concentrations were measured by the BCA protein assay. Cell lysates, streptavidin pull down fractions, and serum-free conditioned media were separated by electrophoresis through 4-12% or 7.5% SDS-PAGE and then transferred to Immobilon-FL PDVF membranes (Millipore, Billerica, MA, USA). Membranes were blocked in LiCor blocking buffer (Rockland Immunochemicals, Gilbertsville, PA, USA) diluted with PBS (1:1) and then incubated with anti-CDCP1 primary antibodies (CS4115, 1:1000; AF2666, 0.2 μg/ml) overnight at 4°C. Following extensive washing, membranes were incubated with species-specific IRDye700 or 800-conjugated secondary antibodies (goat anti-rabbit, 1:15,000; donkey anti-goat, 1:15,000) for 1 h at room temperature, and visualized with a LiCor Odyssey infrared imager (LiCor, Lincoln, NE, USA). Consistent protein loading was determined by reprobing membranes stripped in Restore Western blot stripping buffer with anti-actin antibody (0.05 μg/ml).

5.3 Results

5.3.1 Identification of CDCP1 in ML2 Cell surface Sialoglycoprotein Enriched Fraction
Having combined metabolic labeling and click chemistry with LC-MS/MS, we successfully identified a panel of cell surface sialoglycoproteins whose expression was differential between metastatic (ML2) and nonmetastatic (N2) human prostate cancer cells. CDCP1 is one of the candidates that were identified only from metastatic ML2 cells. Here we present the details about how CDCP1 was confidently identified by this technique. After using the streptavidin pull down assay in Aim II, we saw an intense signal at sub-150 kDa in eluted fractions from cells treated with ManNAz, which was absent in flow-through fractions and eluents from the control group (Fig. 25).

Specifically, in-blot streptavidin IR-800 detected a higher intensity of this area in the ML2 eluent than that of the N2 eluent (Fig. 34A, SAvB). The corresponding bands from a Coomassie stained gel (Fig. 34A, CB) were excised and analyzed by LC-MS/MS. Three unique peptides from the 135-kDa band of ML2 eluent were determined by mass spectrometry to be the recently identified membrane protein CDCP1 (Fig. 34A). The positions of these peptides in the sequence of CDCP1 were designated in Figure 34B. They spanned the amino acids within intracellular carboxy terminus (peptide 3:700-718), extracellular CUB1 (peptide 1:280-297) and CUB3 domains (peptide 2: 582-593).

Figure 34C showed the fragmentation pattern of nanoelectrospray ionization MS/MS of the doubly charged (m/z 910.3) tryptic peptide, which corresponded to the C-terminal tryptic fragment of human CDCP1. There was no CDCP1 identified in N2 or the control fractions, although the overall protein pattern and content was indistinguishable for N2 and ML2 eluents (Fig. 34A, CB).
Figure 34A-B Identification of CDCP1 in ML2 cell surface sialoglycoprotein enriched fraction. (A) N2 and ML2 cells were metabolically treated with AC4ManNAz and conjugated with biotin-alkyne on live cells via click reaction. The total cell lysate was extracted and pulled down with streptavidin resins. Aliquots of the purified sialoglycoproteins were examined by in-blot streptavidin-IR 800 (SAvB) detection, and the remaining samples were stained with Bio-Safe coomassie blue (CB). Three tryptic peptide fragments determined by LC-MS/MS (peptides 1 to 3) were indentified within the sequence of CDCP1. (B) Schematic representation of human CDCP1 structural features including N-terminal signal peptide, 3 extracellular CUB-like domains, 14 potential N-glycosylation sites, transmembrane domain, and intracellular cytoplasmic domain. Dark boxes indicate the positions of tryptic peptides identified by LC-MS/MS.
Figure 34C Identification of CDCP1 in ML2 cell-surface sialoglycoprotein enriched fraction. (C) Representative mass spectra of tryptic CDCP1 digestion. Top panel: MS spectrum of the CDCP1 tryptic peptides; arrow indicates ion at m/z=910.31 (2+) selected for MS/MS. Low panel: MS/MS spectrum of the peptide K.GPAVGIYNGNINTEMPR.Q, annotated using the y-ions series.
5.3.2 CDCP1 Structural Features

The schematic representation of human CDCP1 structure features was included in Figure 34B. Consistent with a predicted cell surface protein, CDCP1 contains a signal peptide at position 1-29, an extracellular domain at 30-665, a transmembrane domain at 666-686, and cytoplasmic domains at 687-836. It was reported that the extracellular domain harbors three regions with low homology to complement protein subcomponents C1r/C1s, urchin embryonic growth factor, and bone morphogenetic protein 1 (CUB) domains, which could be involved in protein:protein and protein:carbohydrate interactions\(^{301}\). Importantly, there are 14 consensus N-glycosylation sites predicted in the extracellular part, as well as a palmitylation motif and five tyrosine residues present in the intracellular domain, indicating the potential function of CDCP1 in outside-in signal transduction\(^{283}\). The predicted mass for full length CDCP1 is 92.9 kDa and 90.1 kDa for mature CDCP1 after signal peptide cleavage. However, the apparent molecular weight in Western blot is 135-140 kDa indicating extensive glycosylation\(^{280}\). UniProtKB database suggests that CDCP1 may have two truncated isoforms: CDCP2 (73 kDa) which is missing 187 amino acids from its N-terminal and CDCP3 (34 kDa) that lacks 502 amino acids from its C-terminal and is a secreted form. To date, there has been no evidence about the presence of these two isoforms.

5.3.3 CDCP1 Expression in Prostate Cell Lines

To verify our proteomic results, we compared CDCP1 levels in cell surface glycoprotein enriched fractions and whole cell lysates by Western blot. N2 and ML2 cells were metabolically labeled, conjugated with biotin-alkyne, and lysed as described
above. An equal volume of streptavidin pull down eluents was resolved by SDS-PAGE and detected by a rabbit polyclonal antibody generated against peptides near the C-terminus. Consistent with the observed LC-MS/MS differential of CDCP1 between PC3 variants, Western blot analysis of the cell surface glycoprotein pool demonstrated that CDCP1 levels were ~3.4-fold greater on the surface of ML2 compared with N2 cells (Fig. 35A). The level of CDCP1 (135 kDa) in ML2 total cell lysate was also higher (1.5-fold) than N2 lysate, albeit not as dramatically as in the surface fraction (Fig. 35B).

Besides 135 kDa full-length CDCP1, N2 and ML2 cells also manifested another truncated (70 kDa) species, a proteolysis product of 135-kDa species, as recently reported. Although LMW-CDCP1 presented marginal levels in both N2 and ML2 streptavidin pull down fractions (Fig. 35A), we could not identify this species by our targeted glycoproteomic approach, mostly like because its levels were beyond MS detection threshold. Interestingly, LMW-CDCP1 displayed a lower level (1.7-fold) in ML2 lysate than N2 lysate. Therefore, the ratio of HMW-CDCP1 to LMW-CDCP1 seemed to be correlated with metastatic ability of prostate cancer cells. There was a slight difference in the apparent molecular weight of this protein between N2 and ML2 cells, possibly reflecting different glycosylation status of CDCP1 in metastatic cells.

Since increased levels of CDCP1 were observed in some aggressive epithelial tumor cells, we set out to investigate the expression of CDCP1 in a panel of prostate cell lines. As shown in Figure 36, CDCP1 protein was detected in all 13 prostate cell lines examined but with various expression levels: 10 clearly expressed both forms of CDCP1: PC3 and its two sublines, LNCaP and its three sublines, RWPE1's two sublines, and DU145 prostate cancer cells; 3 expressed only HMW-CDCP1: immortalized RWPE-1,
Figure 35 Verification of CDCP1 expression in N2 and ML2 cells. (A) N2 and ML2 cells were metabolic labeled with AC₄ManNAz, and conjugated with biotin-alkyne via click chemistry. Then total cell lysates were extracted and subjected to streptavidin pull-down assays as described previously. The eluted fractions were analyzed by Western blot with CDCP1 antibody (pAb4115). Cells treated with natural ManNAc were used as controls by following the same experimental procedures. (B) Western blot analysis of total cell lysates of RWPE1, PC3, N2, and ML2 with CDCP1 antibody. β-actin was included as a loading control.
Figure 36. **Verification of CDCP1 expression in prostate cell lines.** Whole cell lysates were prepared from different prostate cell lines and their sublines. Western blot analysis was performed with primary antibody specific to CDCP1 (pAb4115). β-actin was included as a loading control. PC3 & Sublines: PC3 (androgen-independent), developed from a skeletal metastasis; N2 and ML2, selected from PC3 cell line; N2 is non-metastatic while ML2 is metastatic. LNCaP & Sublines: LNCaP (androgen-dependent), isolated from the lymphnode of a confirmed metastatic PCa patient; C4, C4-2 and C4-2B, developed from LNCaP; C4-2, androgen-independent and tumorigenic; C4-2B, androgen-refractory and bone metastatic. RWPE-1 & Sublines: RWPE-1, HPV-18 immortalized prostate epithelial cell line response to growth factors and androgen; WPE-NB-14, tumorigenic and moderately invasive; WPE-NB-26, tumorigenic and highly invasive. DU145 (androgen-independent), isolated from a brain lesion of a metastatic PCa patient. PacMetUT1, isolated from a lymph node of PCa patient.
PacMetUT1, and 22Rv1 prostate cells. Among these cells, the highest CDCP1 expression was found in androgen-independent prostate cancer cell lines of the PC3 lineage followed by DU145 cells. Lower CDCP1 expression was found in androgen dependent LNCaP cells, its androgen refractory C4-2 and C4-2B sublines, as well as RWPE1 subline cells. In LNCaP and RWPE1 prostate cancer progression models, the correlation of ratio of HMW-CDCP1 to LMW-CDCP1 with invasive activity appeared not strong in panels of tumor cell line from same origins but with more aggressive phenotypes.

5.3.4 Characterization of CDCP1 Glycosylation

CDCP1 was identified in our cell surface sialoglycoprotein enriched fractions. The apparent molecular mass of CDCP1 in SDS-PAGE (135 kDa, Fig. 35A) exceeded that predicted by primary sequence analysis (90 kDa). Moreover, there are 14 potential N-linked glycoprotein sites in the extracellular part of human CDCP1 (Fig. 35B). All of these observations highlight the possibility of CDCP1 glycosylation. However to date, no study about the carbohydrate structure of CDCP1 has been reported.

To address CDCP1 glycosylation, we first employed three glycosidases with different specificities to enzymatically digest CDCP1 in vitro. Whole cell extracts were treated with N-glycosidase F (PNGase F) that cuts off entire N-linked glycans, neuraminidase which hydrolyzes terminal sialic acid units, or Endoglycosidase H (Endo H) that preferentially cleaves high mannose-containing and hybrid oligosaccharides (Fig. 37A). Then proteins were separated by SDS-PAGE, and subjected to Western blot analysis. As shown in Figure 37B, PNGase F treatment resulted in the disappearance
Figure 37 Characterization of CDCP1 glycosylation in vitro. (A) Schematic representation of cleavage sites for the three highly specific glycosidases, PNGase F, Endo H, and neuraminidase, respectively on a complex type of N-linked oligosaccharide. (B-D) Western blot analysis of untreated and PNGase F (B), neuraminidase (C), and Endo H (D) hydrolyzed lysates from PC3, N2, and ML2 cells. The blots were probed with specific CDCP1 antibody (pAb4115). β-actin was included as a loading control.
of 135 kDa and 70 kDa bands and generation of lower molecular weight bands of approximately 90 kDa and 60 kDa, respectively, the former corresponding to the mass of deglycosylated HMW-CDCP1 form. After digestion with neuraminidase, the size of the HMW band decreased slightly, whereas the changes of LMW band were not detectable (Fig. 37C). These results indicated higher sialic acid content in N-glycans of the HMW-CDCP1 than the LMW species, which was consistent with the previous cell surface glycoproteomic data. In that experiment, we could only detect the marginal levels of LMW-CDCP1 expression compared to the significantly high levels of HMW-CDCP1 in the cell surface sialoglycoprotein enriched fractions (Fig. 35A). Deglycosylation with Endo H led to a mass shift of 25 kDa for HMW-CDCP1 (Fig. 37D), less than what would be expected for the full cleavage of N-glycans (45 kDa). This result demonstrated that high-mannose or hybrid oligosaccharide chains only contributed to a portion of the N-glycans present on HMW-CDCP1. In contrast, Endo H treatment of LMW-CDCP1 produced several broad bands closer to or even lower than that after N-glycan removal (Fig. 37D), suggesting the presence of different glycoforms of LMW-CDCP1 containing a higher extent of mannose or hybrid glycosylation, and implying occurrence of O-glycans on this CDCP1 species.

To further study the glycosylation status of CDCP1 in vivo, we treated cells with a N-linked glycosylation inhibitor tunicamycin (TM) or a mannosidase inhibitor II swainsonine (SW) under conditions optimal for drug activity but minimal for cytotoxicity. Whole cell extracts were then obtained and examined by Western blot analysis. Addition of TM, a nucleoside antibiotic which prevents the formation of the dolichol intermediate necessary for oligosaccharide addition of the nascent polypeptide
chain caused a significant decrease in cellular HMV-CDCP1, which was similar to the results produced with the PNGase F treatment (Fig. 38A). Surprisingly, TM did not affect the migration of LMW-CDCP1 in gel. Addition of SW, which blocks converting N-glycans from hybrid-type to complex-type during mannose processing in Golgi, slightly decreased the masses of HMW-CDCP1, but had no effect on LMW-CDCP1 levels (Fig. 38B). This result reflected the addition of complex glycosylation to HMW-CDCP1 but not LMW-CDCP1, consistent with results from the Endo H treatment.

Like sialylation, fucosylation is another form of glycosylation commonly present on cell surface glycoproteins. Altered fucosylation has been also associated with cancer progression and metastasis. To determine whether CDCP1 is a carrier of cell surface fucosylation, we applied our targeted glycoproteomic strategy for cell surface fucosylated glycoproteins. N2 and ML2 cells were metabolically labeled with alkyne fucose and probed with biotin-azide via click chemistry. Then, total cell lysates were subjected to a streptavidin pull-down assay and the eluted fractions were analyzed by Western blot with CDCP1 antibody. As shown in Figure 39, the same levels of HMW-CDCP1 were detected in N2 and ML2 treated fractions (fuc_alkyne), but absent in control groups (fuc), indicating fucosylation of HMW-CDCP1 to the same degree in N2 and ML2 cells. Conversely, there was no detectable fucosylation on LMW-CDCP1.

Taking together, the above data suggested that although both CDCP1 species contained heavy N-linked glycans and were modified by fucose, the type and extent of carbohydrate moieties attached were different; the carbohydrate present on the HMW-CDCP1 harbored three types of N-glycosylation and was terminated by a significant amount of sialic acid residues. In contrast, high-mannose/hybrid-type N-glycans and
Figure 38 Characterization of CDCP1 glycosylation in vivo. (A-B) PC3, N2, and ML2 cells were treated with or without tunicamycin (5 μg/ml, A) or swainsonine (1 μg/ml, B) for 23 h. The total cell lysate was extracted and analyzed by Western blot with CDCP1 antibody (pAb4115). β-actin was included as a loading control.
Figure 39 Fucosylation of CDCP1 in N2 and ML2 cells. N2 and ML2 cells were metabolic labeled with alkynl fucose (fuc_alkyne) and probed with biotin-azide via click chemistry. Then total cell lysates were extracted and subjected to streptavidin pull down assays. The eluted fractions were analyzed by Western blot with CDCP1 antibody. Cells treated with natural fucose (fuc) were used as controls by following the same experimental procedures.
probably O-glycans were primarily present on LMW-CDCP1.

### 5.3.5 Cellular Expression of CDCP1 in N2 and ML2 Cells

CDCP1 expression in N2 and ML2 cells was further substantiated by analysis of CDCP1 protein expression on the cell surface by flow cytometry (Fig. 40A) with a mouse monoclonal antibody (mAb 41-2) generated in subtractive immunization. Both living N2 and ML2 cells showed high levels of cell surface CDCP1, but ML2 cells (Mean fluorescence intensity, MFI, 377) seemed to have more CDCP1 on the cell surface than N2 cells (MFI, 290). Immunofluorescence staining of permeabilized N2 and ML2 cells, using the same antibody, showed a clear plasma membrane localization of CDCP1 (Fig. 40B). Surprisingly, some intracellular juxtanuclear staining was also found in both N2 and ML2 cells. No positive immunostaining of N2 and ML2 cells was observed in the absence of the primary antibody indicating the specificity of the antibody for human CDCP1. A similar finding was observed staining using another antibody CS4115 (data not shown).

Since the glycosylation processing and transport pathway of a cell surface glycoprotein usually go via the ER and Golgi apparatus, we further examined the subcellular location of CDCP1 in N2 and ML2 cells by co-localization studies. First, we wanted to confirm the plasma membrane localization of CDCP1. As expected, simultaneously applying glycan chemical labeling and immunofluorescence staining with CDCP1 antibody revealed complete cell surface co-localization (Fig. 41, yellow coloring), consistent with the initial technique used to identify CDCP1. We next investigated which particular intracellular compartments CDCP1 was located, using
Figure 40 Cellular expression of CDCP1 in N2 and ML2 cells. (A) Flow cytometry analysis of CDCP1 expression in N2 and ML2 cells. N2 and ML2 cells dissociated from the flasks were incubated with CDCP1 antibody mAb 41-2 (closed histograms) or blocking buffer (open histograms) and then with goat anti-mouse IgG conjugated with Alexa 488. PI staining was used to gate living cells. Data was acquired by FACS calibur and analyzed by Flowjo software. (B) Immunofluorescence analysis of CDCP1 expression in N2 and ML2 cells. N2 and ML2 cells growing on coverslips were fixed with 4% paraformaldehyde and permeabilized with ice-cold 100% methanol. They were incubated with CDCP1 antibody mAb 41-2 (top panel) or blocking buffer (bottom panel), stained with goat anti-mouse IgG conjugated with Alexa Fluor 488 (green). The nuclei were stained with TO-PRO3 (blue). Fluorescent images were captured by Zeiss confocal microscopy.
Figure 41 Analysis of CDCP1 subcellular location by colocalization. N2 cells were treated with AC₄ManNAz for 2 days, fixed with 4% paraformaldehyde, and conjugated with biotin-alkyne via click reaction. Cells were then permeabilized with 0.5% Triton X-100, and stained with streptavidin-Alexa Fluor 488 (green) and anti-CDCP1 antibody (pAb 4115), Alexa Fluor 594-conjugated secondary reagents (red). The nuclei were stained with TO-PRO3 (blue). Fluorescent images were captured by Zeiss confocal microscopy.
double immunostaining of CDCP1 and an ER-resident marker calnexin or a Golgi complex marker 58 kDa protein in N2 cells. As shown in Figures 42 and 43, cells expressing endogenous CDCP1 displayed a perinuclear and associated reticular staining typical of the ER (Fig. 42) along with a juxtanuclear pattern typical of the Golgi apparatus (Fig. 43). Merging images with CDCP1 signal resulted in considerable punctuate overlapping patterns, indicating CDCP1 was localized and processed in ER and Golgi complex. Taking together, immunofluorescent and co-localization data indicated that CDCP1 went along with the secretory pathway and mainly localized on the plasma membrane in prostate cancer cells.

5.3.6 CDCP1 Expression in Prostate Cancer Tissues

The expression pattern of CDCP1 protein in four frozen sections of human prostate cancer tissues was examined by immunofluorescence with the use of mAb 41-2 (Fig. 44). CDCP1 was present on normal prostate epithelial cells as well as on malignant cells. However, the staining intensity and subcellular staining pattern were significantly different between malignant and adjacent normal epithelial cells. In contrast to a previous report that found the staining of malignant glands was generally less prevalent and less intensive than benign glands, we observed that human prostate epithelial cancer cells had overall higher reactivity than the matched normal ones (Fig. 44A, b-c), although CDCP1 expression levels varied among tumor cells in the same case. Malignant epithelial cells expressing CDCP1 protein usually belonged to a single acinus or to a few adjacent acini (Fig. 44A, h-i). Staining was predominantly localized on the plasma membrane in cancerous epithelial cells as shown by the confocal images (Fig.
Figure 42 Analysis of CDCP1 ER subcellular location by Immunolocalization. N2 cells growing on coverslips were fixed with 4% paraformaldehyde, and permeabilized with ice-cold 100% methanol, and stained with anti-CDCP1 antibody (pAb 4115), Alexa Fluor 488-conjugated secondary reagents (green), and anti-calnexin antibody, Alexa Fluor 594-conjugated secondary reagents (red). The nuclei were stained with TO-PRO3 (blue). Fluorescent images were captured by Zeiss confocal microscopy. Arrow indicated the punctuate colocalization patterns of CDCP1 and calnexin.
Figure 43 Analysis of CDCP1 Golgi subcellular location by Immunolocalization. N2 cells growing on coverslips were fixed and permeabilized with 0.5% Triton X-100, and stained with anti-CDCP1 antibody (pAb 4115), Alexa Fluor 488-conjugated secondary reagents (green) and anti-Goldi 58 kD protein antibody, Alexa Fluor 594-conjugated secondary reagents (red). The nuclei were stained with TO-PRO3 (blue). Fluorescent images were captured by Zeiss confocal microscopy. Arrow indicated the punctuate colocalization patterns of CDCP1 and Golgi58 kD protein.
The basal surface contained more concentrated CDCP1 than the apical and lateral membranes. In contrast, in adjacent normal cells the majority of the CDCP1 targeted was in a discrete juxtanuclear compartment, close to basal side. The staining on the plasma membrane appeared to a much lesser degree than in cancer cells (Fig. 44C, a-c). The proportion and intensity of cancerous cells expressing CDCP1 protein in a given tissue section were heterogeneous, but it seemed to increase with the high Gleason score (Fig. 44D). Other cell types including blood cells, endothelium, vascular smooth muscle, neuroendocrine cells, and prostatic stoma fibroblasts, were negative. Control sections that were incubated with the secondary, but not the primary antibody, were free of staining (Fig. 44B). In summary, the expression of CDCP1 proteins in prostate cancer tissues was altered as it presented extensively on the plasma membrane of malignant epithelial cells with elevated level of expression.

5.3.7 Shedding of CDCP1 Ectodomain from Conditioned Media and EPS Samples

A number of integral cell surface proteins, such as c-met and CD44 are also produced as soluble molecules \(^{303,304}\). Western blot analysis, probing with an antibody generated against the CDCP1 extracellular domain (AF2666), was employed to examine whether PCa cells produce a soluble form of CDCP1. The concentrated serum-free conditioned media (SFCM) from DU145 cells was used as a positive control because a 65 kDa CDCP1 ectodomain was recently reported in DU145 SFCM \(^{302}\). As shown in Figure 45A, the antibody detected several immunoreactive bands (55 kDa, 42 kDa, 33 kDa, and 28 kDa) in EPS urine, possibly the ectodomain shedding of CDCP1. After normalization of total protein loading, as shown in Figure 45B, band at 55 kDa was
Expression of CDCP1 in prostate patient tissues. (A) Immunofluorescent analysis of CDCP1 expression in matched normal (N) and cancerous (T) prostate tissues from a PCa patient (Gleason score = 3+4). Successive sections (6 μm) were stained with H&E (a, d, g) or anti-CDCP1 antibody (mAb41-2) and Alexa Fluor 488-conjugated secondary reagents (green, b-c, e-f, h-i). Nuclei were stained with PI (red, b, e, h). d-f and g-i showed high magnification (x400) of the normal glands (N) and tumor glands (T) marked in the top panels (white box), respectively. Images were captured by Olympus fluorescent microscopy. Magnification for top panel x100, for middle and low panels x400. (B) Negative control for CDCP1 staining. Slides stained in the absence of primary CDCP1 antibody were used as negative controls.
Figure 44C Expression of CDCP1 in prostate patient tissues. (C) Immunofluorescent analysis of CDCP1 subcellular location in matched normal and cancerous prostate tissues from a PCa patient (Gleason score = 3+4). PCa sections were stained with anti-CDCP1 antibody (mAb41-2, green) and/or PI (red) and representative images for normal glands (a-c) and tumor glands (d-e) were demonstrated. b-c and e-f showed high magnification of the normal glands and tumor glands marked in the left panels (white box), respectively. Images were captured by Zeiss cofocal microscopy. Magnification x100 (a, d), x880 (b, c), x1040 (e, f).
**Figure 44D** Expression of CDCP1 in prostate patient tissues. (D) Three more PCa tissues were analyzed for CDCP1 expression. Top panels, tissue specimens were stained with H&E and examined by the pathologist. Bottom panels, immunofluorescent staining of CDCP1 in the circled tumor areas of H&E sections. Middle panels, immunofluorescent staining of CDCP1 in matched normal areas. Red fluorescence showed PI staining for nuclei. Representative PCa Gleason scores are indicated.
Figure 45 Shedding of CDCP1 ectodomain into the extracellular compartment. (A) Detection of CDCP1 ectodomain in the conditioned media and expressed prostatic secretion (EPS) samples. Western blot analysis of conditioned media from DU145 cell lines and EPS samples from non-cancer, non-aggressive PCa, and aggressive PCa were performed with a goat antibody generated against the extracellular domain of CDCP1 (AF2666). The detected low molecular weight signals (55 kDa, 42 kDa, 33 kDa, and 28 kDa) in EPS urine may indicate the ectodomain shedding forms of CDCP1. (B) Sypro Ruby stained gel of samples shown in the left panel indicating the total protein loading.
present at a higher level in the aggressive PCa EPS urine pool than in the non-aggressive PCa and non-cancer pools.

5.4 Discussion

PCa is the most common occurring malignancies in men and advanced metastatic PCa is currently incurable. Identifying the mediators that drive metastasis may offer insights into how to monitor or prevent PCa progression. By utilizing a novel cell surface glycoproteomic approach and a PCa metastatic cell model derived from PC3 cells, we uniquely identified CDCP1 protein in the metastatic ML2 cells, indicating a potential role of CDCP1 in PCa progression. The function of CDCP1 in cancer progression is also highlighted in other independent studies to identify CDCP1 by employing several different biased genetic and protein based approaches. Using cDNA chip hybridization techniques to search for genes preferentially expressed in solid tumors relative to normal tissues, Scherl-Mostageer et al. identified CDCP1 as a novel human tumor-associated gene in 2001, which is highly transcribed in lung and colon cancer derived cell lines. Two years later, Hooper et al. first isolated the complete CDCP1 protein sequence using subtractive immunization intended to generate antibodies towards cell surface epitopes preferentially expressed by highly metastatic relative to non-metastatic carcinomas. In 2005, aiming at searching for cell-cycle phase-specific mediators of activated Src family kinases (SFKs), Bhatt et al. identified CDCP1 as a mitotic substrate of SFKs in human MDA-MB-468 breast cancer cells. They also showed that overexpression of CDCP1 led to cell loss of adhesion phenotype. Subsequently, Uekita et al. also identified CDCP1 as a SFK interacting protein in human A459 lung adenocarcinoma cells. In these
cells, phosphorylated CDCP1 was required to overcome anoikis and permit in vitro anchorage-independent growth. In addition, CDCP1 was identified as a molecule associated with important signaling proteins such as protein kinase Cδ (PKCδ)\textsuperscript{305}. Supporting the role of CDCP1 in cancer progression, upregulated CDCP1 expression has been observed in breast\textsuperscript{286}, colon\textsuperscript{287,288}, kidney\textsuperscript{289}, pancreatic\textsuperscript{290}, and lung cancer\textsuperscript{291} patients and a number of epithelial cancer cell lines\textsuperscript{280,282,292,293,306}.

Further characterization of CDCP1 in this study showed that there are two forms of CDCP1 protein in most PCa cell lines, the 135 kDa full-length and a shorter 70 kDa species. Similar sized CDCP1 species have been observed in many other human epithelial cell lines such as breast cancer MDA-453 and MDA-468 cells, pancreatic cancer BxpC3 cells, esophageal cancer TE7 cells, and several lung and colon cancer cell lines\textsuperscript{306}. As reported, the LMW-CDCP1 can be expressed endogenously or induced through the action of exogenous serine proteases such as trypsin\textsuperscript{281,302}, plasmin\textsuperscript{281} and matriptase\textsuperscript{282,302}. Currently the endogenous mechanisms and functional relevance of this cleavage is not clear. Here our data showed that the ratio of HMW-CDCP1 to LMW-CDCP1 is highly relevant to the metastatic potential of PCa cells. Supporting this proposal, we observed a trend that membranous expression of CDCP1 was more intense in high-grade aggressive PCa tissues than low-grade PCa (Fig. 44D). Other data coming from in vitro experiments and animal models also strengthened the involvement of HMW-CDCP1 in cancer progression. In another congenital metastatic epidermoid carcinoma cell model, the expression levels of 135 kDa CDCP1 were much higher in the highly metastatic variant of Hep3 over non-metastatic variants\textsuperscript{280}. Also overexpression of CDCP1 in HeLa cells conferred these relatively non-aggressive cells with higher
efficiency in experimental metastasis in mice. Moreover, Uekita et al. demonstrated that reduction of 135 kDa CDCP1 by siRNA significantly inhibited the metastasis of A549 lung adenocarcinoma cells in vivo, but had no effect on primary tumor growth in mice. Recently, this group showed that HMW-CDCP1 promoted migration, invasion, and anchorage independence of highly invasive gastric cancer cell lines in vitro as well as enhanced the peritoneal dissemination in mice of these cells. Taking advantage of the function-blocking mAb 41-2 which mainly targets HMW-CDCP1, and chick embryo metastasis model, Quigley's group elegantly showed that CDCP1 functions as an anti-apoptotic molecule and facilitates tumor cell survival at the late stages of the metastatic cascade, likely during or soon after extravasation. Unlike HMW-CDCP1, there is little accumulative data for the function of LMW-CDCP1 in cancer progression. In our cell model system, we found that the expression levels of LMW-CDCP1 are negatively associated with the malignant abilities of PCa cells. Some studies showed that LMW-CDCP1 was tyrosine phosphorylated under several cellular events including proteolysis and cell adhesion. The phosphorylation promoted the formation of the CDCP1:Src:PKCδ multiprotein complex, suggesting a role of LMW-CDCP1 in signaling transduction. However, the contribution of such phosphorylation for disease progression, especially for cancer metastasis, remains undefined. Therefore, further exploration of the essential roles of two CDCP1 forms in cancer progression should be of high priority.

For other lineage-related PCa cell pairs, the trend seemed not as pronounced, as only low levels of HMW-CDCP1 were detected in the invasive cell lines. This can be explained because the different cell models reflect the distinct properties of PCa
progression, for example, the LNCaP derivative sublines mainly mimic the progression of PCa hormone independent, whereas RWPE-1and its sublines represent the early stages of tumorigenesis and invasiveness. Thus it is not surprising to see the weak trend in these models. Additional congenically matched pairs of high and low metastatic tumor cell lines as well as accurate *in vivo* animal models will have to be examined to establish whether a clear link exists between two species of CDCP1 expression and the extent of malignant progression.

The existence of 14 consensus N-glycosylation sites in the extracellular domain and 5 tyrosine residues in the intracellular COOH terminus of CDCP1 suggests functional involvement of CDCP1 in outside-in signal transduction. Compared to extensive studies of CDCP1 phosphorylation, the glycosylation status and its roles are not yet understood. The work reported here for the first time characterizes the glycosylation of CDCP1. Deglycosylation of CDCP1 with PNGase F and Endoglycosidase H showed that both CDCP1 forms contain N-glycans of high-mannose or hybrid types. Despite these common glycosylation features, the CDCP1 forms differ in extent and type of certain glycosylation modifications. We previously identified HMW-CDCP1 by virtue of it being predominant in the cell surface sialoglycoprotein enriched fraction of metastatic ML2 cells, indicating sialylation of CDCP1. Neuraminidase treatment confirmed the presence of sialic acid. Therefore, these two independent experiments both indicated that HMW-CDCP1 carries more sialic acid in its N-glycans than LMW-CDCP1. In another one of our recent projects aimed at the identification of cell surface fucosylated glycoproteins, HMW-CDCP1 rather than LMW-CDCP1 was also identified in cell surface fucosylated glycoprotein enriched fractions of N2 and ML2 cells, suggesting the
existence of core-fucosylated N-glycans on HMW-CDCP1. More interestingly, we observed that the extent of sialylation of HMW-CDCP1 also correlated with the metastatic potentials of PCa cells consistent with the increasing evidence that hypersialylation is one of the most common changes of N-glycosylation during malignant transformation and progression\textsuperscript{98,110}. Our data raises the possibility that glycosylation signatures, especially sialylation, affect roles of two CDCP1 species in cancer progression. It is well known that glycosylation is critical in determining a protein's stability, conformation, folding, and binding affinity for other molecules\textsuperscript{85}. In addition, glycans \textit{per se} are adhesion molecules directly involved in cell recognition\textsuperscript{84}. The negative charge and the terminal position of sialic acid can determine the ability of a protein to interact with other molecules and even cells. As reported, CDCP1 interacted with a number of adhesion and matrix proteins including cadherins and syndecans\textsuperscript{282}. Actually, CDCP1 overexpression stabilizes a detached cellular phenotype and has a poor ability to spread plastic, which was ascribed to the intracellular constitutive phosphorylation by SFKs\textsuperscript{282,294}. Therefore, it is attractive to think that the glycosylation of CDCP1 is the extracellular signal. Together with intracellular phosphorylation, it may modulate cell adhesion as well as epithelial tumor attributes such as anchorage-independent growth, anoikis resistance, and metastasis. Current detailed N-glycan profiling of two CDCP1 forms as well as elucidation of the possible roles of sialylation of CDCP1 in relation to cancer progression are being pursued in our laboratory.

Another interesting finding from this study is the dysregulated expression of CDCP1 in primary PCa tissues. By using sensitive immunofluorescent staining instead of routine immunohistochemistry, we observed significant alterations in expression levels
and subcellular locations of this protein in PCa tissues compared to adjacent normal tissues. Moreover, although only a small number of patient samples were examined, it appeared that the expression pattern of CDCP1 was more obvious in high-grade aggressive PCa. Our result is opposite of a previous report in which Siva et al. found lower CDCP1 mRNA and protein expression in cancerous PCa cells than normal ones. Since CDCP1 is a protein that is processed by multiple post-translational modifications including glycosylation, phosphorylation, and proteolysis, mRNA levels can not accurately reflect the amount of functional protein. The difference in protein staining levels may mainly result from the usage of two distinct antibodies. Here we used a murine monoclonal antibody mAb 41-2, which was generated by subtractive immunization and demonstrated functional blocking of experimental and spontaneous metastasis of Hela-CDCP1 and PC3-hi/diss cells in mice and chicken embryo model systems. In contrast, Siva et al. made the mouse-anti-CDCP1 antibody 25A11 by whole cell panning and subtraction with red blood cells. This antibody decreased PC3 cells migration and invasion in vitro and exhibited inhibitory effects on primary tumor growth and metastasis of PC3 cells in a mouse xenograft model, when coupled to the cytotoxin saporin, either directly or via a secondary antibody. The mechanistic differences between these two antibodies are unclear and await further exploration. Other variations such as prostate tissues, and staining technology (immunofluorescence vs immunohistochemistry) could also contribute to the difference. However, consistent with our results, dysregulated CDCP1 expression has been also observed in a number of cancers and is reported to have important clinical utility. In 2008, Awakura et al. analyzed CDCP1 expression by immunohistochemistry in 230 renal cell carcinomas and
showed that 33.5% patients had high levels of CDCP1 expression which inversely correlated with the disease-specific and recurrent-free survival. In another study, Ikeda et al. performed the similar examination in 200 lung adenocarcinoma patient samples and demonstrated that 30.0% cases were classified as expressing moderate to high levels of CDCP1. High CDCP1 expression is associated with the occurrence of lymph node metastasis, high relapse rate, and poor prognosis. Recently Miyazawa et al. evaluated 145 pancreatic cancer samples and indicated the elevated expression of CDCP1 was significantly correlated with decreased overall survival of these patients.

Taken together, these clinicopathological studies and cell-based investigations strongly supported that CDCP1 can be used as a cancer prognostic marker and could potentially be targeted to disrupt epithelial cancer progression. Indeed the murine anti-CDCP1 antibodies, exhibiting function-blocking activity of metastasis as mentioned above, have underscored the potential of CDCP1 as a therapeutic target as cancer treatment. More recently, the human antibody specifically against CDCP, obtained from human single chain Fv (scFv) phage displayed libraries by subtractive panning, was shown to effectively inhibit in vivo metastasis and in vitro colony formation of human carcinoma cells without toxin conjugation. The properties of immuno-tolerance and efficiency of anti-metastasis make the human anti-CDCP1 antibody a promising candidate for further therapeutic development. Although significant for several epithelial cancers, the clinical relevance of CDCP1 expression in PCa has not been reported as of now and a large-scale analysis of CDCP1 expression in PCa samples will warrant future investigation.

The detection of CDCP1 fragments in PCa patient EPS urine samples is a significant finding. Proximal fluids, including EPS urine, are found adjacent to a given tissue or
organ and represent a repertoire of secreted proteins and shed cells reflective of the physiological and pathological states of that tissue. Thus, they are potential sources of proteins and genetic biomarkers for cancer. In this study, EPS urine was collected in the clinic by forced (or expressed) ejection of prostatic fluids into the urethra following prostate massage. The presence in vivo of a 55 kDa CDCP1 ectodomain shed from cell surface, and the relatively higher levels of CDCP1 fragment in aggressive PCa EPS urines, indicates that CDCP1 has the potential as a diagnostic maker for prostate cancer. Hooper et al. detected a membrane-shed moiety of CDCP1 of approximately 110 kDa in the serum-free conditioned medium of human epidermoid carcinoma cell line, which would be predicted to contain most of the extracellular domain. They also described strong immunohistochemical staining in the goblet cells of normal colon crypts, the lumen of glands and in the glandular mucus of colon carcinoma, suggesting the production of a soluble form of CDCP1. Bhatt et al. and He et al. reported that CDCP1 undergoes proteolytic digestion by the serine proteases after amino acids R368 or K369 and this cleavage is linked to intracellular signaling pathways. The presence of CDCP1 ectodomain in clinical samples prioritizes CDCP1 as a diagnostic marker worth future validation.

In summary, most PCa cell lines expressed the 135 kDa full-length and a truncated 70 kDa species of CDCP1 protein. Both CDCP1 forms contain heavy N-linked oligosaccharides rich in high-mannose or hybrid-type chains and terminated by sialic acid residues. However, the ratio of high to low molecular weight CDCP1, as well as the glycosylation levels of 135 kDa CDCP1 correlated to the metastatic status of variant PC3 cell lines. Flow cytometry and immunocytochemistry data confirmed elevated levels of
CDCP1 in metastatic ML2 cells, and also demonstrated that this protein is processed by ER/Golgi network and concentrated on the plasma membrane. Immunofluorescent staining of human PCa tissues indicated that CDCP1 exhibited increased intensity and extensive cell surface staining in malignant epithelial cells. In contrast, CDCP1 was targeted to a discrete juxtanuclear compartment and the plasma membrane staining in matched normal cells appeared much weaker. Shedding of CDCP1 ectodomain was detected by Western blot at elevated levels in aggressive PCa patient EPS urine samples. Our data indicate that expression of CDCP1 protein is dysregulated in PCa and it has the potential utility as a therapeutic target and a diagnostic marker for PCa progression.
CHAPTER VI
CONCLUSIONS AND FUTURE DIRECTIONS


A. Our optimization of lectin staining protocols determined that the use of non-enzymatic dissociation buffer for cell suspension preparation provides the best balance of dissociation efficiency, cell viability, and cell-surface carbohydrate retention. Additionally, our results show that blocking buffer, staining buffer, cell confluence, and the concentrations of lectin are potential sources of variation affecting successful lectin flow cytometry analysis. Lectin staining of unfixed cells is superior to fixed cells for cytohistochemistry analysis for representation of the true surface binding patterns.

B. Lectin flow cytometry and microscopy revealed that the cell-surface glycoconjugates of PC3 and its variants contain a diversity of carbohydrate groups and are extensively sialylated.

C. Total cell-surface carbohydrate profiling by lectin binding showed that the metastatic potential of PC3 variants is inversely correlated with the level of cell-surface \( \alpha_2-6 \) linked sialic acid.

6.2 Future Directions of Aim I

In Aim I, the predominant carbohydrate detected by lectin flow cytometry for PC3 and its variants was sialic acid while other residues showed relatively weaker signals, initially suggesting that these carbohydrates were low contents on the PCa cells. However, terminal sialic acid can mask other lectin binding sites and the extent of
masking effect is various\textsuperscript{111}. One aspect of future studies will use a neuraminidase for enzymatic removal of sialic acid to reveal the real abundance of underlined sugars.

Besides sialylation, fucosylation, and N-acetylgalactosamine (GalNAc) studied in this Aim, well-characterized lectins for other types of aberrant cancer-related glycosylation are also available. For example, lectin L-PHA recognizes β1-6 branching of N-linked oligosaccharides, a common feature of most cancers\textsuperscript{86}. Lectins AHA and ECA react with two different linkages of T antigen, Gal beta 1-3GalNac, and Gal beta 1-4GlcNac, respectively. The expression of these carbohydrates was reported to correlate with metastatic sites of prostate cancer\textsuperscript{310}. Therefore, the future employment of a wider spectrum of and linkage-specific lectins will be useful to intensively characterize cell-surface glycosylation patterns associated with PCa metastasis.

Our data reported the use of flow cytometry as an effective technique for detecting differences in cell-surface lectin binding in fresh cultured cells. To make more clinical relevant conclusions, in the future we can extend it to spilled cells from fresh prostate cancer tissues or fine-needle aspirates, and compare lectin binding patterns of primary PCa tumor to metastases from other sites such as the lymph nodes.

6.3 AIM II (Chapter IV): Targeted identification of metastasis-associated cell-surface sialoglycoproteins in prostate cancer.

A. We have developed a robust MS-based glycoproteomic methodology for the selective identification of specific subclasses of cell-surface glycoproteins. The approach combined metabolic labeling of cell-surface sialyl glycans, chemical probing of the
labeled sugar with biotin-alkyne via click reaction, affinity purification of sialylated proteins, SDS-PAGE separation, and subsequent LC-MS/MS for protein identification.

B. Azide-modified sialic acid metabolic precursor (ManNAc) was efficiently and specifically incorporated into cell-surface sialylated glycans in PCa cancer cells, as demonstrated by flow cytometry and confocal microscopic analysis. Click-activated conjugation of the biotin tag following metabolic labeling allows visualization of cell-surface glycoproteins as well as, observation of intracellular localization of sialylated glycoconjugates.

C. Our optimization of metabolic labeling and click reaction experiments determined that ManNAz treatment with 20-40 μM for 1 day and click reaction with 25 μl of biotin-alkyne with 2.5 μl CuSO₄ can provide maximum cell viability/vigor.

D. Specific azide-modified sialoglycoprotein signals were detected by streptavidin-IR 800 in post-click cell extracts treated with ManNAz, demonstrating the feasibility of labeling individual glycoproteins by using this probing system. In addition, there appeared to be a major protein band (under 150 kDa) that reacted with streptavidin and greater sialylated glycoproteins in the ML2 cell line when compared to the N2 cell line.

E. The bound and eluted fractions from streptavidin beads produced strong distinct glycoprotein patterns visualized by in-blot streptavidin-IR 800 detection and 1D-PAGE analysis, indicating efficient capture and enrichment of the azide-labeled proteins.

F. Application of the described targeted glycoproteomic approach to our syngeneic PCa cell line model allowed for the identification of total 536 non-redundant proteins. Using conservative criteria, 64 proteins (26%) from N2 and 72 proteins (29%) from ML2 were classified as extracellular or membrane-associated glycoproteins. A selective
enrichment of sialoglycoproteins was observed at the affinity chromatography step. When compared to global proteomic analysis of the same cells, the proportion of identified glycoprotein and cell-surface proteins were on average 3-fold higher in the AC₄ManNAz labeled cells. Functional clustering of differentially expressed proteins by Ingenuity Pathway Analysis demonstrated that the vast majority of the glycoproteins over-expressed in the metastatic ML2 subline were involved in cell motility, migration, and invasion.

G. A total of 80 cell-surface glycoproteins differentially expressed between cell lines were uniquely identified and proposed as PCa biomarkers specific to metastatic competence.

6.4 Future Directions of Aim II

In Aim II, our targeted glycoproteomic approach coupled glycan labeling with traditional gel-based liquid chromatography tandem mass spectrometry (GeLC-MS/MS) technology for glycoprotein identification. GeLC-MS/MS offers a number of benefits for protein identification such as increasing the depth of analysis by spreading out the complex protein mixture over 20-30 gel slices, removing low molecular weight impurities which are often detrimental for MS sequencing, and providing observed mass of proteins which is very useful to predict post-translational modifications. Alternatively, MS-based shotgun technology which directly analyzes complex protein mixtures using multidimensional high-pressure liquid chromatography (LC/LC) in conjunction with tandem mass spectrometry (MS/MS), has advantages in speed, sensitivity, and automation. In the future, we would combine glycan labeling
technology with shotgun proteomic platform for the analysis of the cell-surface glycoproteome. Specifically, glycoproteins enriched by streptavidin beads as we have described will be serially digested with Lys-C and trypsin, yielding a first peptide pool representing all of the bound proteins. The peptide pool will be subjected to shotgun proteomic analysis and the resulting proteins will be searched against the UniprotKB human protein database for glycoprotein assignment. The remaining tagged N-linked glycopeptides bound to streptavidin beads will be released via peptide-N-glycosidase F (PNGase F), yielding a secondary glycopeptide pool. The glycopeptide pool will be analyzed by LC-MS/MS and glycoprotein identification will be read directly from the conversion of Asn residues to Asp. Such strategies will offer two extra advantages: an increase in sequence coverage of the glycoproteins by crossing-referencing between the two pools; and the determination of N-linked glycosylation sites. This would be of particular interest in studies where changes of glycosylation occupancy are suspected to link with the disease states.

One possible reason for the significant number of proteins annotated as nuclear or cytoplasmic that was present among our cell-surface sialoglycoproteome is copper cytotoxicity during the click reaction. Fortunately, a new generation of copper-free click chemistry can resolve this problem. Improving cell viability could allow for subcellular fractionation to increase membrane glycoprotein identification. With collaboration with Dr. Binghe Wang at Georgia State University, the development of copper-free click chemistry-based glycoproteomic methodology is currently underway.

The main accomplishment of Aim II was the identification of 80 cell-surface glycoproteins differentially expressed between cell lines by our targeted glycoproteomic
approach. Here "differentially expressed glycoproteins" refer to glycoproteins which were uniquely identified in one cell line. Some glycoproteins identified in both cell lines but possibly with differential levels were lost. To overcome such limitations, quantitative proteomic techniques are necessary to be incorporated into our glycoproteomic approach. In the past decade, several quantitative proteomics techniques have been widely employed by the whole proteomics community, including stable isotopic or isobaric labeling using chemical reactions (e.g. ICAT and iTRAQ)\textsuperscript{313,314}, metabolic incorporation (e.g. SILAC)\textsuperscript{315}, and enzymatic reactions (e.g. \textsuperscript{18}O labeling)\textsuperscript{316}, as well as less quantitatively accurate label-free approaches\textsuperscript{317}. These techniques are readily adapted into our glycoproteomic analysis at different stages. For example, the SILAC technique can incorporate stable isotope labeled amino acids into proteins during the cell culturing process while the iTRAQ technique can introduce isobaric tags after tryptic digestion of glycoproteins. \textsuperscript{18}O can be incorporated into the N-glycopeptides during enzymatic reactions, such as tryptic digestion (incorporation of two \textsuperscript{18}O into the peptide carboxyl terminal) and PNGase F-mediated hydrolysis (incorporation of one \textsuperscript{18}O into the asparagine of N-glycosylation sites). Therefore, future attempts will be made to incorporate these quantitative techniques into our targeted glycoproteomics to facilitate glycoprotein isolation, identification, and quantification in one experiment.

6.5 AIM III (Chapter V): Verification of expression of cell-based biomarkers in clinical samples.

A. CUB-domain-containing protein 1 (CDCP1) was successfully identified from the prominently sialylated band (under 150 kDa) observed in Aim II by GeLC-MS/MS.
B. Western blot analysis of the cell-surface glycoprotein enriched fractions as well as whole cell lysates confirmed the proteomic data, demonstrating the over-expression of CDCP1 in metastatic PCa cell line (ML2). Additionally, the 135 kDa full-length and a truncated 70 kDa species of CDCP1 protein were found to be expressed in most PCa cell lines.

B. Analysis of deglycosylation cell lysates indicated that both CDCP1 species contains heavy N-linked glycans and are modified by sialylation and fucosylation, where the type and extent of carbohydrate moieties attached are different. The carbohydrate present on the HMW-CDCP1 harbors three types of N-glycosylation, and is terminated by a significant amount of sialic acid residues. In contrast, high-mannose/hybrid-type N-glycans and probably O-glycans are primarily present on LMW-CDCP1 and they are less sialylated.

C. Western blot analysis of our syngeneic PCa cell line model suggested that the ratio of high to low molecular weight CDCP1, as well as glycosylation levels of the 135 kDa CDCP1 is correlated to the metastatic competence of PC3 variants.

D. Flow cytometry and immunocytochemistry data confirmed elevated levels of CDCP1 in metastatic ML2 cells, and also demonstrated that this protein is processed by the ER/Golgi network and concentrated on the plasma membrane.

E. Immunofluorescence microscopy of human PCa tissues revealed dysregulated expression of CDCP1 proteins in prostate cancer tissues. CDCP1 exhibits an increased intensity of cell surface presentation in malignant gland cells whereas it is confined to a discrete perinuclear compartment with only weak plasma membrane staining in matched
normal cells. Importantly, CDCP1 expression levels appeared to increase with increasing Gleason score.

F. Shedding of putative CDCP1 ectodomains was detected by Western blot at elevated levels in aggressive PCa post-DRE urine samples.

6.6 Future Directions of Aim III

The prioritization and verification of candidate biomarkers provides an important bridge between discovery and validation. Traditionally, proteins of interest are verified using immunoassays, as we did in Aim III, which require high-quality antibodies that already exist. For many or most novel candidates, such antibodies will not be available. The cost and time associated with the generation and optimization of a specific antibody is considerable and makes this option both prohibitively expensive and inefficient for high numbers of proteins. For panels of protein candidates, verification by multiple reaction monitoring MS (MRM-MS) using a triple quadrupole mass spectrometer presents a compelling alternative. MRM on the triple quadrupole mass analyzer is the gold standard for developing MS-based quantitative methods. In the MRM mode, an ion of interest (the precursor) is preselected with the first mass filter Q1, and induced to fragment in Q2 by collisional excitation with a neutral gas in a pressurized collision cell. The resulting ‘product ions’ (that is, the fragmentation products of the selected precursor) are mass analyzed using the third quadrupole Q3. MRM-MS has much higher throughput, accuracy, and sensitivity and allows for substantial multiplexing. Importantly, addition of isotopically labeled internal standards allow for absolute quantification. By rapid and continuous monitoring exclusively for the specific ions
of interest, MRM-MS enhances the lower detection limit for peptides by up to 100-fold. Moreover, a moderate number of candidate proteins (30–100) can be simultaneously targeted and measured in a single experiment, particularly when using scheduled MRM, which incorporates prior knowledge of peptide elution times into the LC-MS/MS program\textsuperscript{152}. MRM-MS has been applied to estimate the levels of give minor glycoproteins in human serum\textsuperscript{155}, to measure the levels of biomarker candidates in a mouse model of breast cancer\textsuperscript{318}, and to study the N-linked glycosylation reaction in congenital disorders of glycosylation type-I serum\textsuperscript{193}. Thus, MRM-MS is an extremely powerful for biomarker discovery and candidate validation. With panels of cell-surface sialoglycoprotein candidates on hand and triple quadrupole mass spectrometry available at our lab, we will be devoted to develop MRM-MS assays for our future candidate biomarker verification.

CDCP1 is type I cell-surface protein with extensive glycosylation. Compared to extensive studies of CDCP1 phosphorylation, the glycosylation status and its roles are not yet understood. The work reported in Aim III for the first time characterizes the glycosylation of CDCP1 by several glycosidases and glycosylation inhibitors. To uncover the underlying mechanisms of CDCP1 glycosylation in PCa progression and metastasis and discover potential glycan-based biomarkers, a survey of CDCP1 glycome is necessary. The complete analysis of CDCP1 glycosylation includes mapping its glycosylation sites, determination of sequence and linkage of glycans, and site-specific carbohydrate characterization. We have multiple mass spectrometers including MALDI-TOF/TOF, ESI-LC-MS/MS, and triple quadrupole MS at our lab with the capabilities of performing multiple levels of glycan analysis. For example, the monosaccharide
composition of CDCP1 can be analyzed by MALDI after the glycan is released from purified CDCP1. Glycosylation sites of CDCP1 can be mapped by PNGase F treatment in conjunction with LC-MS/MS analysis. Determination of the glycan sequence of CDCP1 and site-specific glycan analysis can be performed by tandem MS experiments by the use of the triple quadrupole MS. Therefore, detailed glycan profiling of two CDCP1 forms as well as elucidation of the possible roles of glycosylation of CDCP1 in relation to cancer progression will be an interesting project and it will be pursued in our laboratory.

In a small set of PCa tissues, we observed that the expression of CDCP1 in PCa is dysregulated and elevated membranous expression of CDCP1 appears to correlate with PCa aggressiveness. To further determine the clinical significance of CDCP1 expression, a larger cohort of samples will be used to examine CDCP1 expression in PCa tissues using immunohistochemistry (IHC). We have a large bank of formalin-fixed paraffin-embedded (FFPE) PCa tissue blocks representing all stages and grades. The analysis of these archived samples with complete pathology and clinical outcome will allow us to establish a clearer link between CDCP1 expression and the extent of malignant progression. Currently, optimization of IHC conditions for FFPE PCa tissues is being pursued at our laboratory.

The detection of CDCP1 fragments in PCa EPS urine samples patients is significant since the finding indicated that the ectodomain shedding of CDCP1 can reach the body fluids and serve as early sentinel for cancer detection. Future studies will confirm the Western blot results by immunoprecipitation of CDCP1 from EPS with antibody
targeting to the N-terminal portion of CDCP1 and analysis of the purified CDCP1 fragments by mass spectrometry.

6.7 Concluding Remarks

Altered glycosylation has been associated with tumor development and disease progression. Given the growing evidence that glycans of cell-surface glycoproteins are involved in each and every aspect of tumor progression, glycoproteins are becoming important targets for the development of biomarkers for cancer diagnosis, prognosis, and therapeutic response to drugs. With the guidance of cancer-related changes in glycosylation, the marriage of efficient glycoprotein enrichment and sophisticated MS technologies offers great opportunities for biomarker research. The studies described herein provide an example of glycobiology-driven biomarker discovery. By selectively tagging cancer-related cell-surface glycans, we have been able to exclusively target the pool of cell surface-expressed glycoproteins for subsequent in-depth MS analysis. The successful identification of membrane glycoprotein CDCP1 whose expression is dysregulated in PCa provides strong proof-of-principal evidence for our developed glycoproteomic approach. There are numerous roles of CDCP1 that may play in cancer metastasis including inhibiting apoptosis and aiding in extravasations. Additional, CDCP1 may have soluble/shed isoforms which could provide additional critical information concerning PCa progression. Much work remains to be done to translate the identified biomarkers to validation assays and subsequently in high-throughput clinical assays.
REFERENCE


163. Sharon, N. Lectin-carbohydrate complexes of plants and animals: an atomic view. 

explore and analyze hepatocellular carcinoma-related glycoprotein markers. 

165. Wei, X. & Li, L. Comparative glycoproteomics: approaches and applications. 

166. Mao, X. _et al._ Integrated lectin affinity microfluidic chip for glycoform 

167. Yang, Z. & Hancock, W.S. Approach to the comprehensive analysis of 

168. Mechref, Y., Madera, M. & Novotny, M.V. Glycoprotein enrichment through 

169. Drake, R.R. _et al._ Lectin capture strategies combined with mass spectrometry for 

170. Qiu, R. & Regnier, F.E. Use of multidimensional lectin affinity chromatography 

171. Durham, M. & Regnier, F.E. Targeted glycoproteomics: serial lectin affinity 
chromatography in the selection of O-glycosylation sites on proteins from the 


173. Zhang, H., Li, X.J., Martin, D.B. & Aebersold, R. Identification and 
quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope 

glycopeptide capture and liquid chromatography mass spectrometry. _Mol Cell 

175. Zhang, H. Glycoproteomics using chemical immobilization. _Curr Protoc Protein 

176. Soltermann, A. _et al._ N-glycoprotein profiling of lung adenocarcinoma pleural 


VITA

LIFANG YANG

Department of Microbiology and Molecular Cell Biology
Eastern Virginia Medical School (EVMS)
Norfolk, VA 23501

Education:

2010  Ph.D., Biomedical Sciences
       Eastern Virginia Medical School
       and Old Dominion University
       Norfolk, VA

2001  M.S., Academy of Military Medical Sciences
       Beijing, China

1998  M.D., Beijing Medical University
       Beijing, China

Work Experience:

2002-2004  Resident, Internal Medicine
           Beijing Cancer Hospital, Beijing, China

2001-2002  Resident, Internal Medicine
           Affiliated Hospital of Academy of Military Medical Sciences,
           Beijing, China

Publications:


