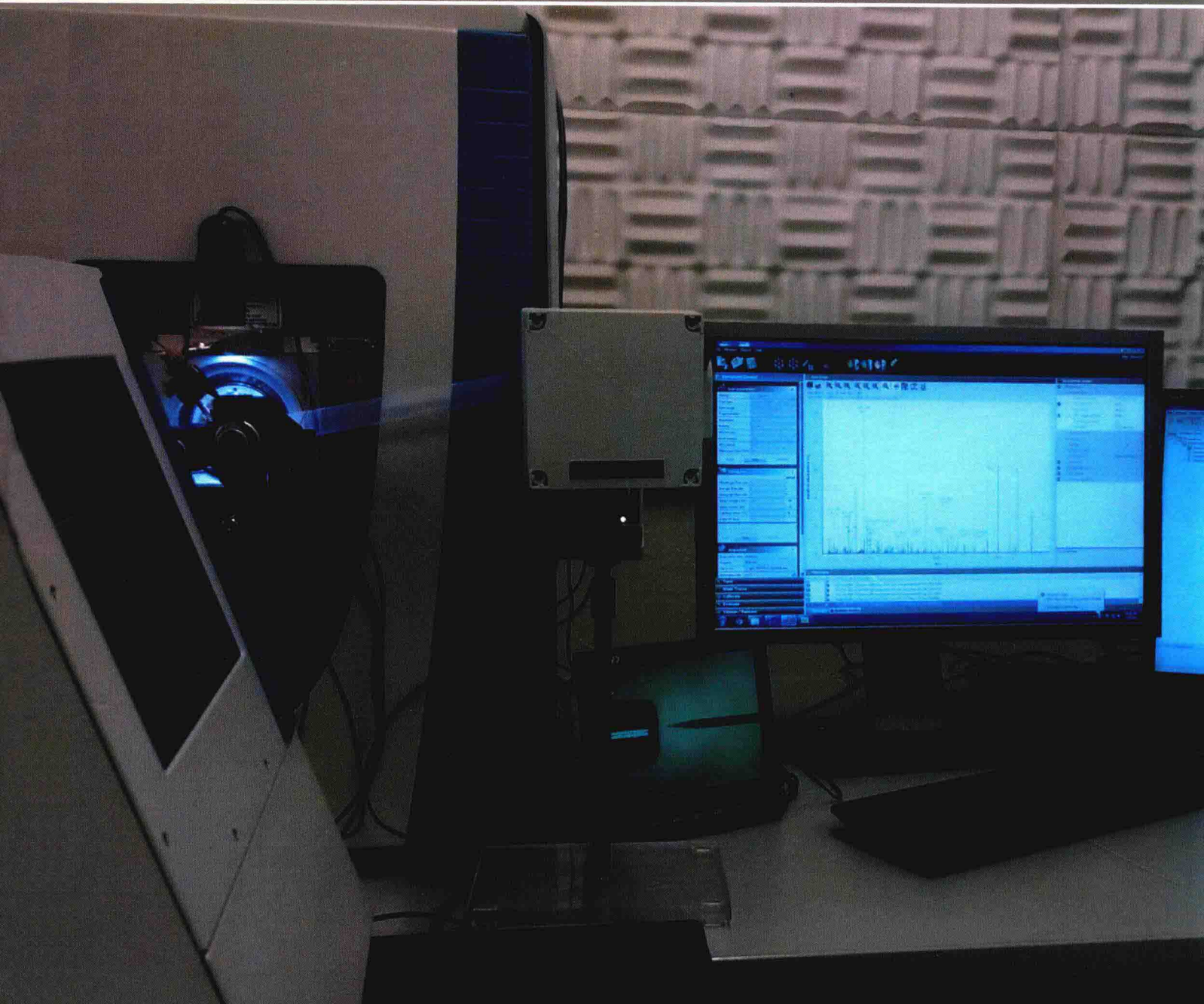
Proteomics

Advanced Concepts and Perspectives

Arthur Handley



Proteomics: Advanced Concepts and Perspectives

Edited by **Arthur Handley**

RCallisto Reference
New York

Published by Callisto Reference,
106 Park Avenue, Suite 200,
New York, NY 10016, USA
www.callistoreference.com

Proteomics: Advanced Concepts and Perspectives
Edited by Arthur Handley

International Standard Book Number: 978-1-63239-526-9 (Hardback)

Printed in China.

Proteomics: Advanced Concepts and Perspectives

Preface

Advances in the fields of biology and computational science during the past twenty years have opened new possibilities. One such discipline which has emerged is Omics fields. Due to a lack of correlation between mRNA and proteins, and growing significance of post transcriptional regulations and protein modifications in protein functions and human diseases, proteomics has become an increasingly important research field. Proteomics is the study of the proteome in the cell, which represents the complete set of proteins encoded by the genome. Proteomics is different and more complicated as compared to genomics. This is because while an organism's genome is more or less constant, the proteome differs across cells from time to time.

Protein research has expanded from its biochemical characterization of individual proteins, to the high throughput proteomics analysis of a cell, complex cell populations, and even an entire organism. This extremely important development highlights the potential of using proteomics methods in studies related to protein functions and human diseases. An important application of proteomics is in in the field of oncology, which is the discovery and validation of prognostic and predictive biomarkers, and plays a fundamental role in personalized therapy for cancer patients. Though Proteomics essentially focuses on proteins, it often finds its applications in areas such as protein purification and mass spectrometry.

I especially wish to acknowledge the contributing authors, without whom a work of this magnitude would clearly not have been realizable. I thank all the contributing authors for allocating much of their scarce time to this project. Not only do I appreciate their participation, but also their adherence as a group to the time parameters set for this publication.

I hope that this book proves to be a resourceful guide for both basic and advanced concepts in proteomics.

Editor

Contents

	Preface	**IX**
Chapter 1	**High Mass Accuracy Phosphopeptide Identification Using Tandem Mass Spectra** Rovshan G. Sadygov	**1**
Chapter 2	**Identification of a Novel Biomarker for Biliary Tract Cancer Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry** Shintaro Kikkawa, Kazuyuki Sogawa, Mamoru Satoh, Hiroshi Umemura, Yoshio Kodera, Kazuyuki Matsushita, Takeshi Tomonaga, Masaru Miyazaki, Osamu Yokosuka, and Fumio Nomura	**6**
Chapter 3	**A Comprehensive Subcellular Proteomic Survey of *Salmonella* Grown under Phagosome-Mimicking versus Standard Laboratory Conditions** Roslyn N. Brown, James A. Sanford, Jea H. Park, Brooke L. Deatherage, Boyd L. Champion, Richard D. Smith, Fred Heffron, and Joshua N. Adkins	**14**
Chapter 4	**Serum Biomarkers Identification by Mass Spectrometry in High-Mortality Tumors** Alessandra Tessitore, Agata Gaggiano, Germana Cicciarelli, Daniela Verzella, Daria Capece, Mariafausta Fischietti, Francesca Zazzeroni and Edoardo Alesse	**26**
Chapter 5	**Proteomic Analysis of the Ontogenetic Variability in Plasma Composition of Juvenile and Adult *Bothrops jararaca* Snakes** Karen de Morais-Zani, Kathleen Fernandes Grego, Aparecida Sadae Tanaka, and Anita Mitico Tanaka-Azevedo	**41**
Chapter 6	**Current Status and Advances in Quantitative Proteomic Mass Spectrometry** Valerie C. Wasinger, Ming Zeng, and Yunki Yau	**50**
Chapter 7	**Advantageous Uses of Mass Spectrometry for the Quantification of Proteins** John E. Hale	**62**

Chapter 8 **Cathepsin D Expression in Colorectal Cancer: From Proteomic Discovery through Validation Using Western Blotting, Immunohistochemistry, and Tissue Microarrays** 68
Chandra Kirana, Hongjun Shi, Emma Laing, Kylie Hood, Rose Miller, Peter Bethwaite, John Keating, T. William Jordan, Mark Hayes, and Richard Stubbs

Chapter 9 **A Miniaturized Ligand Binding Assay for EGFR** 78
Jochen M. Schwenk, Oliver Poetz, Robert Zeillinger, and Thomas O. Joos

Chapter 10 **Top-Down Characterization of the Post-Translationally Modified Intact Periplasmic Proteome from the Bacterium *Novosphingobium aromaticivorans*** 83
Si Wu, Roslyn N. Brown, Samuel H. Payne, Da Meng, Rui Zhao, Nikola Tolić, Li Cao, Anil Shukla, Matthew E. Monroe, Ronald J. Moore, Mary S. Lipton, and Ljiljana Paša-Tolić

Chapter 11 **Miniaturized Mass-Spectrometry-Based Analysis System for Fully Automated Examination of Conditioned Cell Culture Media** 93
Emanuel Weber, Martijn W. H. Pinkse, Eda Bener-Aksam, Michael J. Vellekoop, and Peter D. E. M. Verhaert

Chapter 12 **An Internal Standard-Assisted Synthesis and Degradation Proteomic Approach Reveals the Potential Linkage between VPS4B Depletion and Activation of Fatty Acid β-Oxidation in Breast Cancer Cells** 101
Zhongping Liao, Stefani N. Thomas, Yunhu Wan, H. Helen Lin, David K. Ann, and Austin J. Yang

Chapter 13 **Proteomic Analysis and Label-Free Quantification of the Large *Clostridium difficile* Toxins** 114
Hercules Moura, Rebecca R. Terilli, Adrian R. Woolfitt, Yulanda M. Williamson, Glauber Wagner, Thomas A. Blake, Maria I. Solano, and John R. Barr

Chapter 14 **Human Myocardial Protein Pattern Reveals Cardiac Diseases** 124
Jonas Bergquist, Gökhan Baykut, Maria Bergquist, Matthias Witt, Franz-Josef Mayer, and Doan Baykut

Chapter 15 **Plasma Fractionation Enriches Post-Myocardial Infarction Samples Prior to Proteomics Analysis** 141
Lisandra E. de Castro Brás, Kristine Y. DeLeon, Yonggang Ma, Qiuxia Dai, Kevin Hakala, Susan T. Weintraub, and Merry L. Lindsey

Chapter 16 **Characterization of the Phosphoproteome in Human Bronchoalveolar Lavage Fluid** 149
Francesco Giorgianni, Valentina Mileo, Dominic M. Desiderio, Silvia Catinella, and Sarka Beranova-Giorgianni

Chapter 17 **Rapid Screening of the Epidermal Growth Factor Receptor Phosphosignaling Pathway via Microplate-Based Dot Blot Assays** 157
Amedeo Cappione III, Janet Smith, Masaharu Mabuchi, and Timothy Nadler

Chapter 18 **Seven-Signal Proteomic Signature for Detection of Operable Pancreatic Ductal Adenocarcinoma and Their Discrimination from Autoimmune Pancreatitis** 164
Kiyoshi Yanagisawa, Shuta Tomida, Keitaro Matsuo, Chinatsu Arima, Miyoko Kusumegi, Yukihiro Yokoyama, Shigeru B. H. Ko, Nobumasa Mizuno, Takeo Kawahara, Yoko Kuroyanagi, Toshiyuki Takeuchi, Hidemi Goto, Kenji Yamao, Masato Nagino, Kazuo Tajima, and Takashi Takahashi

Chapter 19 **Application of iTRAQ Reagents to Relatively Quantify the Reversible Redox State of Cysteine Residues** 175
Brian McDonagh, Pablo Martínez-Acedo, Jesús Vázquez, C. Alicia Padilla, David Sheehan, and José Antonio Bárcena

Chapter 20 **Functional Proteomic Profiling of Phosphodiesterases Using SeraFILE Separations Platform** 184
Amita R. Oka, Matthew P. Kuruc, Ketan M. Gujarathi, and Swapan Roy

Chapter 21 **Optimization of an Efficient Protein Extraction Protocol Compatible with Two-Dimensional Electrophoresis and Mass Spectrometry from Recalcitrant Phenolic Rich Roots of Chickpea (*Cicer arietinum* L.)** 192
Moniya Chatterjee, Sumanti Gupta, Anirban Bhar, and Sampa Das

Permissions

List of Contributors

High Mass Accuracy Phosphopeptide Identification Using Tandem Mass Spectra

Rovshan G. Sadygov

Sealy Center for Molecular Medicine, Department of Biochemistry and Molecular Biology, The University of Texas Medical Branch, Galveston, TX 77555, USA

Correspondence should be addressed to Rovshan G. Sadygov, rgsadygo@utmb.edu

Academic Editor: Qiangwei Xia

Phosphoproteomics is a powerful analytical platform for identification and quantification of phosphorylated peptides and assignment of phosphorylation sites. Bioinformatics tools to identify phosphorylated peptides from their tandem mass spectra and protein sequence databases are important part of phosphoproteomics. In this work, we discuss general informatics aspects of mass-spectrometry-based phosphoproteomics. Some of the specifics of phosphopeptide identifications stem from the labile nature of phosphor groups and expanded peptide search space. Allowing for modifications of Ser, Thr, and Tyr residues exponentially increases effective database size. High mass resolution and accuracy measurements of precursor mass-to-charge ratios help to restrict the search space of candidate peptide sequences. The higher-order fragmentations of neutral loss ions enhance the fragment ion mass spectra of phosphorylated peptides. We show an example of a phosphopeptide identification where accounting for fragmentation from neutral loss species improves the identification scores in a database search algorithm by 50%.

1. Introduction

The reversible phosphorylation of proteins regulates many aspects of cell life [1–3]. Phosphorylation and dephosphorylation, catalyzed by protein kinases and protein phosphatases, can change the function of a protein, for example, increase or decrease its biological activity, stabilize it or mark it for destruction, facilitate or inhibit movement between subcellular compartments, initiate or disrupt protein-protein interactions [1]. It is estimated that 30% of all cellular proteins are phosphorylated on at least one residue [4]. Abnormal phosphorylation is now recognized as a cause or consequence of many human diseases. Several natural toxins and tumor promoters produce their effects by targeting particular protein kinases [5, 6] and phosphatases. Protein kinases catalyze the transfer of the γ-phosphate from ATP to specific amino acids in proteins; in eukaryotes, these are usually Ser, Thr, and Tyr residues.

Mass-spectrometry-based proteomics has emerged as a powerful platform for the analysis of protein phosphorylations [7]. In particular, the shotgun proteomics [8], using liquid chromatography coupled with mass spectrometry (LC-MS), has been successfully employed for comprehensive analysis of global phosphoproteome [6, 9, 10]. The advances in the phosphoproteomics were driven by developments in mass spectrometry (high resolution and mass accuracy), peptide/protein separation, phosphopeptide/protein enrichment, peptide fragmentation [11, 12], quantification, and bioinformatics data processing, Figure 1. Currently, thousands of the phosphopeptides can be detected and quantified in just one experiment. Excellent recent reviews describe experimental procedures involved in phosphoproteomics [13, 14]. Bioinformatics processing is recognized as an integral part of phosphoproteome analysis. Several applications have been developed for phosphopeptide identifications [15, 16], phosphorylation site localization [17, 18], and quantification [19]. Tandem mass spectra are searched for phosphopeptides from protein sequences with potential modifications on Ser, Thr, and Tyr residues. The searches are not targeted. Every modifiable residues can be either modified or unmodified. The effective peptide search space increases exponentially leading to computational complexity

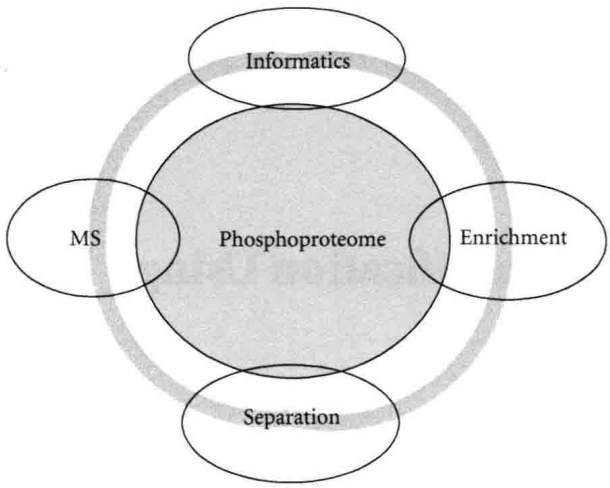


FIGURE 1: Phosphoproteomics and its constituent parts.

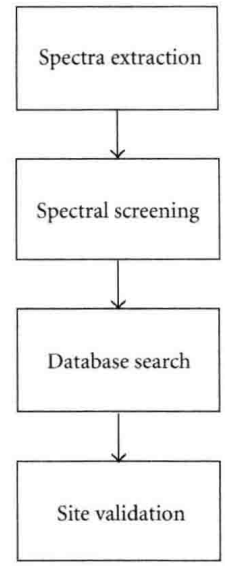


FIGURE 2: General informatics flowchart of a phosphoproteomics analysis.

as well as possible false identifications. High mass accuracy afforded by the modern mass spectrometers enables reducing the complexity of the search space by applying tighter bounds on peptide masses.

Lu and coworkers [20, 21] have developed models based on support vector machine (SVM) to screen for phosphopeptide spectra and validate their identifications. Their approach accurately explains spectra from phosphorylated peptides. However, SVM also acts like a black box, and it is difficult to gain insights into specifics of its decision making. Another development had used dynamic programming to relate spectra of modified and unmodified forms of a peptide [22]. This approach identifies modified peptides by comparing their tandem mass spectra with the annotated tandem mass spectra of unmodified peptides. The search space is restricted to peptides positively identified in unmodified form.

Here, we describe the informatics aspects of phosphopeptide identifications using protein sequence databases and mass spectral data from high mass accuracy and resolution instruments. Database identifications of phosphorylated peptides are done in a dynamic mode—assuming that in a peptide sequence Ser, Thr, and Tyr may or may not be are modified. For database searches, it effectively means exponential increase in the size of database. About 17% of amino acid residues (of which Ser 8.5%, Thr 5.7%, Tyr 3.0%) [23] in human proteome can potentially be phosphorylated. In general, if there are N amino acid residues which can potentially be phosphorylated, the effective database size could increase by as much as 2^N times.

2. Informatics Aspects of the Phosphoproteomics

2.1. Spectra Extraction. LTQ-Orbitrap mass spectrometer [24] stores the mass spectra in a proprietary "raw" file format (ThermoFisher Scientific, San Jose, CA). extract_msn algorithm extracts spectral information from the raw file and converts it into text file format for further data processing. It uses a built-in module to evaluate isotopic envelope of mass species. From the isotope distribution, extract_msn determines the monoisotopic mass and charge state of a peptide. Both of these are critically important and used by database search algorithms.

Normally, the full MS scan is recorded in the Orbitrap mass analyzer which is a high resolution and mass accuracy mass analyzer. The routine mass accuracy of intact peptides is in the range of ±5–10 part-per-million (ppm). This is a very high mass accuracy and is very important for reducing false discovery rates of peptide identifications. The accuracy of the intact peptide's mass affects the number of candidate peptides from the database that will be considered in matching to the spectra. The candidate peptides are filtered based on the mass of the intact peptide and accuracy with which the mass has been measured. The higher the accuracy, the smaller the number of candidate peptides, and as a result the smaller the possibility of false positives. Fragment ion masses are recorded in ion trap mass analyzer. This is a very sensitive mass analyzer. However, the mass accuracy of measured ions is nominal, and normally in the range of ±0.5 Da. Figure 2 summarizes the informatics flowchart of a phosphoproteomics analysis.

2.2. Database Searching. Peptide identification using tandem mass spectra and protein databases is an integral part of proteomics. It is important that peptide assignments are determined with high accuracy and are verifiable. In high-throughput experiments, when thousands of tandem mass spectra are searched, it is not practical for an expert user to manually assign every spectrum and the assignments are made by software. The software uses a concept, either heuristic or probabilistic model, to measure similarity between experimental tandem mass spectrum and an amino acid sequence. For high quality spectra, when signal-to-noise ratio is high and spectra contains clearly defined ion

theoretical ions matched the tandem mass spectrum in the original model (ignoring neutral loss fragments). The results on this and other phosphorylated peptide spectra showed that a realistic model of product ions of phosphorylated peptides needs to account for the fragments resulting from neutral (phosphoric acid group) loss of the b- and y-ions. The procedure has been automated and is used in the case of dedicated MS^n experiments to enhance fragmentation spectra of phosphopeptides.

4. Conclusion

Increased mass accuracy for precursor ions combined with enhanced fragmentation pathways helps bioinformatics methods to improve phosphopeptide identifications from tandem mass spectra and protein sequence databases. Normally identifications of phosphorylated peptides yield small cross-correlation scores. This has partially to do with the theoretical fragmentation models, which take into account only b- and y-ions generated from the peptide bond fragmentations of phosphorylated precursor peptides. We augmented the fragmentation pattern (in SEQUEST) [26] introducing theoretical peaks for b- and y-ions from neutral loss precursors and fragments. Cross-correlation scores of phosphorylated peptides increased by up to 50% using the enhanced fragmentation model.

Abbreviations

Da: Dalton
ppm: Parts per million
Ser: Serine
Thr: Threonine
Tyr: Tyrosine
CID: Collision induced dissociation
LC: Liquid chromatography
MS: Mass spectrometry.

Acknowledgments

This work was supported, in part, by UL1RR029876 UTMB CTSA (ARB), HHSN272200800048C NIAID Clinical Proteomics Center (ARB), and NIH-NLBIHHSN268201000037C NHLBI Proteomics Center for Airway Inflammation (Alex Kurosky, UTMB).

References

[1] P. Cohen, "The origins of protein phosphorylation," *Nature Cell Biology*, vol. 4, no. 5, pp. E127–E130, 2002.

[2] J. A. Ubersax and J. E. Ferrell Jr., "Mechanisms of specificity in protein phosphorylation," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 7, pp. 530–541, 2007.

[3] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, "The protein kinase complement of the human genome," *Science*, vol. 298, no. 5600, pp. 1912–1934, 2002.

[4] P. Cohen, "The regulation of protein function by multisite phosphorylation—a 25 year update," *Trends in Biochemical Sciences*, vol. 25, no. 12, pp. 596–601, 2000.

[5] J. N. Andersen, S. Sathyanarayanan, A. Di Bacco et al., "Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors," *Science Translational Medicine*, vol. 2, no. 43, Article ID 43ra55, 2010.

[6] A. Moritz, Y. Li, A. Guo et al., "Akt-RSK-S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases," *Science Signaling*, vol. 3, no. 136, article ra64, 2010.

[7] H. Zhou, J. D. Watts, and R. Aebersold, "A systematic approach to the analysis of protein phosphorylation," *Nature Biotechnology*, vol. 19, no. 4, pp. 375–378, 2001.

[8] A. J. Link, J. Eng, D. M. Schieltz et al., "Direct analysis of protein complexes using mass spectrometry," *Nature Biotechnology*, vol. 17, no. 7, pp. 676–682, 1999.

[9] E. L. Huttlin, M. P. Jedrychowski, J. E. Elias et al., "A tissue-specific atlas of mouse protein phosphorylation and expression," *Cell*, vol. 143, no. 7, pp. 1174–1189, 2010.

[10] B. Zhai, S. A. Beausoleil, J. Mintseris, and S. P. Gygi, "Phosphoproteome analysis of Drosophila melanogaster embryos," *Journal of Proteome Research*, vol. 7, no. 4, pp. 1675–1682, 2008.

[11] J. J. Coon, B. Ueberheide, J. E. P. Syka et al., "Protein identification using sequential ion/ion reactions and tandem mass spectrometry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9463–9468, 2005.

[12] M. P. Jedrychowski, E. L. Huttlin, W. Haas, M. E. Sowa, R. Rad, and S. P. Gygi, "Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics," *Molecular & Cellular Proteomics*, vol. 10, no. 12, article M111, 2011.

[13] F. Wang, C. Song, K. Cheng, X. Jiang, M. Ye, and H. Zou, "Perspectives of comprehensive phosphoproteome analysis using shotgun strategy," *Analytical Chemistry*, vol. 83, no. 21, pp. 8078–8085, 2011.

[14] C. L. Nilsson, "Advances in quantitative phosphoproteomics," *Analytical Chemistry*, vol. 84, no. 2, pp. 735–746, 2012.

[15] B. E. Ruttenberg, T. Pisitkun, M. A. Knepper, and J. D. Hoffert, "PhosphoScore: an open-source phosphorylation site assignment Tool for MS^n data," *Journal of Proteome Research*, vol. 7, no. 7, pp. 3054–3059, 2008.

[16] J. K. Eng, A. L. McCormack, and J. R. Yates III, "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database," *Journal of the American Society for Mass Spectrometry*, vol. 5, no. 11, pp. 976–989, 1994.

[17] S. A. Beausoleil, J. Villén, S. A. Gerber, J. Rush, and S. P. Gygi, "A probability-based approach for high-throughput protein phosphorylation analysis and site localization," *Nature Biotechnology*, vol. 24, no. 10, pp. 1285–1292, 2006.

[18] T. Taus, T. Kocher, P. Pichler et al., "Universal and confident phosphorylation site localization using phosphoRS," *Journal of Proteome Research*, vol. 10, no. 12, pp. 5354–5362, 2011.

[19] J. Cox, I. Matic, M. Hilger et al., "A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics," *Nature protocols*, vol. 4, no. 5, pp. 698–705, 2009.

[20] B. Lu, C. Ruse, T. Xu, S. K. Park, and J. Yates III, "Automatic validation of phosphopeptide identifications from tandem mass spectra," *Analytical Chemistry*, vol. 79, no. 4, pp. 1301–1310, 2007.

[21] B. Lu, C. I. Ruse, and J. R. Yates III, "Colander: a probability-based support vector machine algorithm for automatic screening for CID spectra of phosphopeptides prior to database

search," *Journal of Proteome Research*, vol. 7, no. 8, pp. 3628–3634, 2008.

[22] D. Tsur, S. Tanner, E. Zandi, V. Bafna, and P. A. Pevzner, "Identification of post-translational modifications by blind search of mass spectra," *Nature Biotechnology*, vol. 23, no. 12, pp. 1562–1567, 2005.

[23] N. Echols, P. Harrison, S. Balasubramanian et al., "Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes," *Nucleic Acids Research*, vol. 30, no. 11, pp. 2515–2523, 2002.

[24] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. G. Cooks, "The Orbitrap: a new mass spectrometer," *Journal of Mass Spectrometry*, vol. 40, no. 4, pp. 430–443, 2005.

[25] B. Lu, D. B. McClatchy, Y. K. Jin, and J. R. Yates III, "Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry," *Proteomics*, vol. 8, no. 19, pp. 3947–3955, 2008.

[26] R. G. Sadygov, J. Shofstahl, and A. Humer, "Improvements to the database search algorithm SEQUEST for accurate mass support and improved phosphorylation searching," in *Proceedings of the Annual Conference on Mass Spectrometry and Allied Topics*, 2005.

[27] B. Paizs and S. Suhai, "Fragmentation pathways of protonated peptides," *Mass Spectrometry Reviews*, vol. 24, no. 4, pp. 508–548, 2005.

[28] R. G. Sadygov, F. M. Maroto, and A. F. R. Hühmer, "ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces," *Analytical Chemistry*, vol. 78, no. 24, pp. 8207–8217, 2006.

Identification of a Novel Biomarker for Biliary Tract Cancer Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

Shintaro Kikkawa,[1] Kazuyuki Sogawa,[2] Mamoru Satoh,[2] Hiroshi Umemura,[3] Yoshio Kodera,[2,4] Kazuyuki Matsushita,[3] Takeshi Tomonaga,[2,5] Masaru Miyazaki,[6] Osamu Yokosuka,[1] and Fumio Nomura[2,3]

[1] *Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba, Chiba City 260-8670, Japan*
[2] *Clinical Proteomics Center, Chiba University Hospital, 1-8-1 Inohana, Chuo-ku, Chiba, Chiba City 260-8670, Japan*
[3] *Department of Molecular Diagnosis, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba, Chiba City 260-8670, Japan*
[4] *Department of Physics, School of Science, Kitasato University, 1-15-1 Kitasato, Minami-ku, Kanagawa, Sagamihara City 228-8555, Japan*
[5] *Laboratory of Proteome Research, National Institute of Biomedical Innovation, 7-6-8 Saito Asagi, Osaka, Ibaraki City 567-0085, Japan*
[6] *Department of General Surgery, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba, Chiba City 260-8670, Japan*

Correspondence should be addressed to Fumio Nomura, fnomura@faculty.chiba-u.jp

Academic Editor: Terence C. W. Poon

Early diagnosis of biliary tract cancer (BTC) is important for curative surgical resection. Current tumor markers of BTC are unsatisfactory in terms of sensitivity and specificity. In a search for novel biomarkers for BTC, serum samples obtained from 62 patients with BTC were compared with those from patients with benign biliary diseases and from healthy controls, using the MALDI-TOF/TOF ClinProt system. Initial screening and further validation identified a peak at 4204 Da with significantly greater intensity in the BTC samples. The 4204 Da peak was partially purified and identified as a fragment of prothrombin by amino acid sequencing. The sensitivity of the 4204 Da peptide for detection of stage I BTC cancer was greater than those for CEA and CA19-9. Also, serum levels of the 4204 Da peptide were above the cut-off level in 15 (79%) of 19 cases in which the CEA and CA19-9 levels were both within their cut-off values. Receiver operating characteristic analysis showed that the combination of the 4204 Da peptide and CA19-9 was significantly more sensitive for detection of stage I BTC cancer compared to CEA and CA19-9. These results suggest that this protein fragment may be a promising biomarker for biliary tract cancer.

1. Introduction

Biliary tract cancer (BTC) is a neoplasm that accounts for 3% of all gastrointestinal cancers and 15% of all primary liver cancers. Over the last two decades, the incidence of BTC has risen, mainly due to an increase in the intrahepatic form [1, 2], which has a particularly high incidence in Northern Thailand [3]. Surgical resection is the only curative treatment and this requires an early diagnosis. Even in cases in which surgical resection with negative histological margins is achieved, the 5-year survival rates range from 20% to 40% [4, 5]. The mean one-year survival rate for unresectable cases is only 6 months [4]. Therefore, there is a need to establish a tool for early diagnosis of BTC. Currently, diagnosis of

Table 1: Clinical characteristics of patients with biliary tract cancer or benign biliary disease and healthy volunteers.

Healthy volunteers	Benign biliary disease	Biliary tract cancer	
No. of patients	30	30	62
Male/female	18/12	18/12	36/26
Mean age	65.5 ± 4.5	64.4 ± 39.0	64.7 ± 37.4
CEA (ng/mL)	3.1 ± 3.5	2.5 ± 3.6	37.6 ± 1577.3
CA19-9 (U/mL)	15.3 ± 18.6	70.4 ± 603.9	5033.6 ± 190367.2

BTC depends on imaging of the biliary tree using computed tomography (CT), ultrasonography, and endoscopic retrograde cholangiography (ERC) in symptomatic subjects. Brush cytology during ERC can lead to morphological diagnosis, but the sensitivity is limited because of the highly desmoplastic reaction of BTC [5, 6]. For these reasons, tumor markers that can detect BTC with high diagnostic efficiency are urgently needed. Carcinoembryonic antigen (CEA) and carbohydrate antigen 19.9 (CA19-9) are tumor markers that are used for diagnosis of BTC, but their sensitivity and specificity are unsatisfactory [2, 7].

Proteome analysis is increasingly being applied to cancer biomarker discovery. Surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) is a proteomics technique used for high-throughput fingerprinting of serum proteins [8]. We have used this technology to identify diagnostic markers for alcohol abuse [9] and a prognostic marker for pancreatic cancer [10]. SELDI-TOF MS can be used to analyze many samples rapidly and simultaneously, but has drawbacks of high cost and difficulty with protein identification. More recently, high-throughput workflow with matrix-assisted laser desorption/ionization-time of flight/time of flight-mass spectrometry (MALDI-TOF/TOF MS) has been established for discovery and identification of serum peptides [11]. This method uses magnetic beads with different chemical chromatographic surfaces, instead of ProteinChip arrays. Proteins selectively bound to the magnetic beads are eluted and analyzed by MALDI-TOF/TOF MS. Compared with the SELDI-TOF MS ProteinChip system, the cost is low, and subsequent protein identification is relatively easy. We recently used the ClinProt system for MALDI-TOF/TOF MS to detect novel biomarkers for alcohol abuse that could not be detected using SELDI [12]. In the present study, we carried out a serum peptidome study to identify novel biomarkers for biliary tract cancer using the MALDI-TOF/TOF MS ClinProt system.

2. Methods

2.1. Patients and Samples. Serum samples were obtained from 62 patients with BTC (36 males, 26 females; median age 64.7 years old, range 27–81 years old), 30 age-matched healthy controls (18 males, 12 females; median age 65.5 years old, range 61–69 years old), and 30 age-matched patients with benign biliary disease (18 males, 12 females; median age 64.4 years old, range 27–90 years old). Clinicopathological data for all the subjects are shown in Tables 1, 2, and 3. The

Table 2: Characteristics of patients with biliary tract cancer.

Item	Number of patients
Location	
Extrahepatic	16
Intrahepatic	17
Klatskin	7
Ampulla of Vater	6
Gallbladder	16
UICC stage	
Stage I	6
Stage II	10
Stage III	16
Stage IV	30

Table 3: Characteristics of patients with benign biliary disease (n = 30).

Item	Number of patients
Cholelithiasis	26
Benign fibrous stricture	2
Primary sclerosing cholangitis	2

BTC group (Table 2) included cases of intrahepatic cholangiocarcinoma (n = 17), Klatskin tumor (n = 7), extrahepatic cholangiocarcinoma (n = 16), tumor of the ampulla of Vater (n = 6), and gallbladder tumor (n = 16). The pathological stages of the BTC patients were defined according to the Union Internationale Contre le Cancer tumor node metastasis classification [13]. The patients with benign biliary diseases (Table 3) included cases of cholelithiasis (n = 24), benign fibrous stricture (n = 4), and primary sclerosing cholangitis (n = 2). All the cases of BTC were diagnosed by radiological imaging. In 58 cases, cytology was also compatible with the diagnosis. All of the patients with benign biliary disease were diagnosed by endoscopic retrograde cholangiopancreatography and were followed up for more than 12 months to confirm that they had no malignancy. Serum samples were obtained and processed under standardized conditions that we have described elsewhere [14] and were stored at −80°C until analysis. Written informed consent was obtained from all the subjects. The study was approved by the Ethics Committee of Chiba University School of Medicine.

2.2. Serum Pretreatment with Magnetic Beads Using the ClinProt Robot. We used weak cation exchange (WCX) magnetic

beads (Bruker Daltonics) and performed serum peptidome fractionation according to the manufacturer's protocol. A 5 μL serum sample was mixed with 10 μL of binding buffer to which 5 μL of WCX beads was added, and the solution was carefully mixed. The peptides in the serum were then allowed to bind to the WCX beads for 5 min. The tube was then placed in a magnetic bead separator (Bruker Daltonics) for separating unbound beads, and the supernatant was removed. The beads were washed three times with 100 μL of washing buffer, and the proteins as well as peptides were then eluted from the magnetic beads with 10 μL each of elution and stabilization buffer. Thereafter, 2 μL of peptide elution solution was mixed with 20 μL of alpha-cyano-4-hydroxy-cinnamic acid matrix (Bruker Daltonics). Then 0.8 μL of this mixture was spotted onto an AnchorChip target plate (Bruker Daltonics) and crystallized. Each sample was duplicated, and quadruplicate spotting was performed using each eluate; eight spots were developed from each sample. The mean spectra from these eight spots were used for data analyses. These procedures from bead fractionation to spotting were performed automatically using the ClinProt robot (Bruker Daltonics) under strictly controlled humidity, as we previously described [14].

2.3. Mass Spectrometry. The AnchorChip target plate was placed in an AutoFlex II TOF/TOF mass spectrometer (Bruker Daltonics) controlled by Flexcontrol 2.4 software (Bruker Daltonics). The instrument was equipped with a 337 nm nitrogen laser, delayed-extraction electronics, and a 25 Hz digitizer. All acquisitions were generated by an automated method included in the instrument software and based on averaging of 1000 randomized shots. Spectra were acquired in positive linear mode in the mass range of 600–10000 Da. Peak clusters were completed using second pass peak sections (signal to noise ratio > 5). The relative peak intensities of *m/z* between 600 and 10000 normalized to a total ion current were expressed in arbitrary units. Calibration was performed using Peptide Calibration Standard II (Bruker Daltonics). All MALDI-TOF MS spectra from *m/z* 1000 to 10000 were analyzed with FlexAnalysis 2.1 and ClinProtools 2.1 software (Bruker Daltonics).

2.4. Protein Identification. A CM ceramic Hyper DF spin column (Bio-Rad Laboratories, Irvine, CA, USA) was washed 3 times with 400 μL of MB-WCX binding solution (Bruker Daltonics). Serum samples (320 μL) were diluted 5-fold with binding buffer and the diluted sample (1600 μL) was applied to the spin column. The sample was allowed to bind at 4°C for 1 h on a shaker and then the spin column was washed 3 times with 400 μL of binding buffer. Finally, 320 μL of MB-WCX stabilization solution (Bruker Daltonics) was added to the spin column for elution. Ten volumes of ice cold acetone were added to the eluate. Peptides/proteins were allowed to precipitate at −20°C for 2 h and then obtained by centrifugation at 13000 g for 10 min at 4°C. After decanting the acetone, the peptides/proteins were allowed to air dry. The dried pellets were resuspended in buffer (0.1% trifluoroacetic acid in water, vol/vol) and further separated by reversed-phase HPLC in an automated HPLC system (Shiseido Nanospace SI-2, Shiseido Fine Chemicals, Tokyo, Japan). The concentrated flow-through sample (75 μL) was directly loaded onto an Intrada WP-RP column (Imtakt, Kyoto, Japan). The reversed-phase separations for each flow-through fraction were performed using a multisegment elu-k tion gradient with eluent A (0.1% trifluoroacetic acid in water, vol/vol) and eluent B (0.08% trifluoroacetic acid in 90% acetonitrile, vol/vol). The gradient elution program consisted of three steps with increasing concentrations of eluent B (5% B for 5 min, 5% to 95% B for 23 min, and 95% B for 11 min) followed by 5% B for 21 min for reequilibration of the column at a flow rate of 0.40 mL/min for a total run time of 60 min. Based on the chromatogram recorded by measuring the absorbance of the eluate at 280 nm, fractions eluted at retention times between 19.1 and 39.1 min were collected in 40 0.2 mL aliquots at a fraction size setting of 0.5 min. Fractions including objective peaks were confirmed by MALDI-TOF MS. N-terminal amino acid sequence analysis was performed using a Procise 494 cLC protein sequencing system (Applied Biosystems, Foster City, CA, USA).

Table 4: Discriminatory peaks and *P* values in the training set.

Higher in biliary tract cancer group		Lower in biliary tract cancer group	
m/z	*P* value	*m/z*	*P* value
1207	<0.0001	1944	<0.0001
1466	<0.0001	2669	<0.001
3261	<0.001	2931	<0.0001
3950	<0.001	3239	<0.0001
4202	<0.001	3272	<0.001
4635	<0.001	3878	<0.01
4654	<0.001	4051	<0.001
5791	<0.0001	4086	<0.0001
5890	<0.0001	4276	<0.0001
5929	<0.001	6414	<0.001
9246	<0.001		
9285	<0.0001		

2.5. Statistical Analysis. Univariate analysis of individual peaks was performed using a nonparametric Mann-Whitney *U* test, with $P < 0.05$ considered significant. Discriminatory power for putative markers was further evaluated by receiver operating characteristic (ROC) analysis and the area under the curve (AUC) using IBM SPSS Statistics 18 (SPSS Inc., Ill, USA).

3. Results

3.1. MALDI-TOF-MS Analysis of Peptides in BTC Sera. As a first step, we compared the peptide profiles of serum samples obtained from BTC patients ($n = 30$) with those from healthy controls ($n = 12$) as a training set (Table 4). Totally 134 peaks were detected and compared in the MALDI

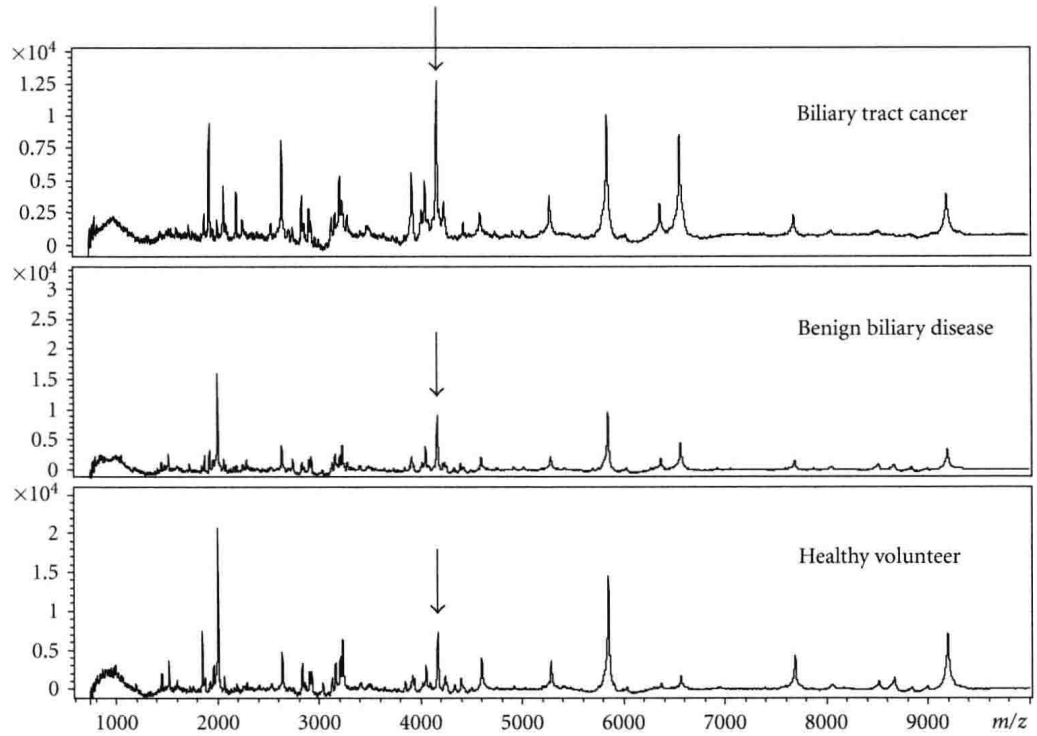


FIGURE 1: The protein mass profile between *m/z* 0 and 10000 highlighting the differentially expressed peaks in serum from healthy volunteers, patients with benign biliary disease, and BTC patients. The *m/z* 4204 peak (indicated by arrows) intensity was higher in cancer patients compared with patients with benign disease and healthy volunteers.

proteomic profile. Total of 22 peak intensities differed significantly between the BTC patients and healthy controls, including 12 that were higher and 10 that were lower in the BTC group. In typical spectra for serum samples from each group (Figure 1), the intensity of the 4204 *m/z* peak was higher in BTC samples compared with those from patients with benign biliary diseases and from healthy controls.

Next, we tested whether the differences observed in the 22 peaks in the training set were reproducible in another set of samples (test set) (Table 5). 32 BTC patients, 30 benign biliary patients, and 18 healthy controls were included in the test set. Out of these 22 peaks, the intensities of 2 peaks (3272 *m/z* and 4204 *m/z*) were again significantly different between the BTC and control groups. Out of these 2 peaks, the intensity of one peak (4204 *m/z*) was also found to be significantly higher in the BTC group compared to the benign disease group. The relative intensities of the 4204 *m/z* peak in sera obtained from the three groups of subjects are summarized in Figure 2.

3.2. Identification of the 4204 Da Peptide as a Fragment of Prothrombin. Partial purification of the peptide corresponding to the 4204 *m/z* peak was conducted as outlined in Section 2. N-terminal amino acid sequencing of trypsin digests of the final preparation containing the 4204 Da peptide revealed that it was a fragment of prothrombin (Figure 3).

3.3. Diagnostic Value of the 4204 Da Peptide Compared with Conventional Markers. Patients with BTC were divided into 4 groups based on clinical stage. The sensitivities of CEA, CA19-9, and 4204 Da in the BTC patients were determined (Figure 4). The optimal cut-off point for CEA, CA19-9, and the 4204 Da peptide were selected based on mean + 2SD in healthy subjects. The cut-off levels for CEA, CA19-9, and the 4204 Da peptide were set at 6.4 ng/mL, 33.5 U/mL, and 372.1 AU, respectively. The sensitivities of CEA, CA19-9, and the 4204 Da peptide in stage IV patients were 33.3%, 80.0%, and 66.7%, and the specificities of CEA, CA19-9, and the 4204 Da peptide were 93.3%, 93.3%, 96.7%, respectively. In contrast, these sensitivities in stage I patients were 0.0%, 16.7%, and 50.0%, and these specificities were 93.3%, 93.3%, 96.7%. The sensitivity of the 4204 Da peptide was also greater than those of CEA and CA19-9 in stage II patients.

The ROC curves for the 4204 Da peptide, CEA, and CA19-9 as single markers and combinations are shown in Figure 5. The sensitivities were determined from the results for the 62 patients with BTC and specificities were based on the 60 non-BTC subjects. The AUCs for the 4204 Da peptide, CEA, and CA19-9 as single markers were 0.75, 0.60, and 0.732, respectively. The AUC for the combination of the 4204 Da peptide and CA19-9 was significantly greater than

TABLE 5: Discriminatory peaks detected in the training set and test set.

Higher in biliary tract cancer group		Lower in biliary tract cancer group	
m/z	*P* value	*m/z*	*P* value
4202	<0.001	3272	<0.001

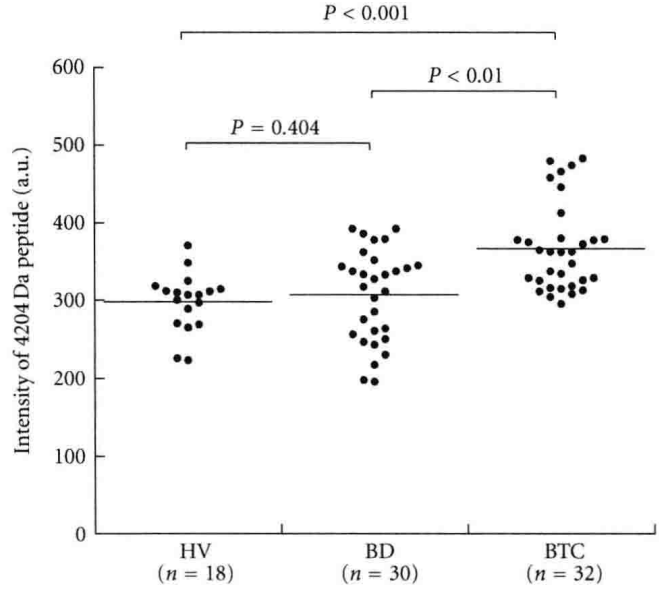


FIGURE 2: Normalized intensities of the peak corresponding to the 4204 Da peptide in serum from healthy volunteers ($n = 18$), patients with benign biliary disease ($n = 30$), and BTC patients ($n = 32$). The peak intensity was significantly higher in sera obtained from patients with biliary tract cancer (BTC) compared with sera from healthy volunteers. There was no significant difference in intensity between BTC sera and benign biliary disease (BB) sera (Mann-Whitney U test).

Prothrombin fragment detected in this study

MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRRANTFLEE VRKGNLEREC
VEETCSYEEA FEALESSTAT DVFWAKYTAC ETARTPRDKL AACLEGNCAE GLGTNYRGHV
NITRSGIECQ LWRSRYPHKP EINSTTHPGA DLQENFCRNP DSSTTGPWCY TTDPTVRRQE
CSIPVCGQDQ VTVAMTPRSE GSSVNLSPPL EQCVPDRGQQ YQGRLAVTTH GLPCLAWASA
QAKALSKHQD FNSAVQLVEN FCRNPDGDEE GVWCYVAGKP GDFGYCDLNY CEEAVEEETG
DGLDEDSDRA IEGRTATSEY QTFFNPRTFG SGEADCGLRP LFEKKSLEDK TERELLESYI
DGRIVEGSDA EIGMSPWQVM LFRKSPQELL CGASLISDRW VLTAAHCLLY PPWDKNFTEN
DLLVRIGKHS RTRYERNIEK ISMLEKIYIH PRYNWRENLD RDIALMKLKK PVAFSDYIHP
VCLPDRETAA SLLQAGYKGR VTGWGNLKET WTANVGKGQP SVLQVVNLPI VERPVCKDST
RIRITDNMFC AGYKPDEGKR GDACEGDSGG PFVMKSPFNN RWYQMGIVSW GEGCDRDGKY
GFYTHVFRLK KWIQKVIDQF GE

FIGURE 3: N-terminal amino acid sequence of the purified fraction. Red letters identify the N-terminal sequence. The molecular weight of the underlined region is 4204 Da.

that for CEA and CA19-9 ($P < 0.01$). The sensitivity and specificity for combination of the 4204 Da and CA19-9 were 59.8% and 84.0%.

The 62 patients with BTC were also classified into 8 groups based on their tumor marker status, as shown in Table 6. The cut-off values for CEA and CA19-9 were set at 5 ng/mL and 37 U/mL, respectively. The optimal cut-off point for the 4204 Da peptide was selected based on the ROC analysis. The 4204 Da peptide level was greater than the cut-off value in 15 (79%) of 19 cases in which the CEA and CA19-9 levels were within their respective cut-off values.

TABLE 6: Positive or negative status of 4204 Da peptide, CEA, and CA19-9 in patients with biliary tract cancer.

CEA (≥5 ng/mL)	CA19-9 (≥37 U/mL)	4204 Da peptide (≥322 A.U.)	Number of patients
−	−	−	4
−	−	+	15
−	+	−	6
+	−	−	2
+	+	−	3
+	−	+	2
−	+	+	18
+	+	+	12

4. Discussion

The sequencing of the human genome has opened the door for comprehensive analysis of all mRNAs (transcriptome) and proteins (proteome). However, the levels of mRNAs are not necessarily predictive of the corresponding protein levels. Indeed, a recent report indicated that the consistency between cDNA microarray and proteome-based profiles is limited for identification of candidate biomarkers in renal cell carcinoma [15]. Therefore, proteome analysis is a prerequisite for identification of novel biomarkers.