Protein Cell 2012, 3(9): 669–674
DOI 10.1007/s13238-012-2934-4

Protein & Cell

COMMUNICATION

Comprehensive analysis of the N and C terminus of endogenous serum peptides reveals a highly conserved cleavage site pattern derived from proteolytic enzymes

Fangjun Wang, Jun Zhu, Lianghai Hu, Hongqiang Qin, Mingliang Ye, Hanfa Zou✉

CAS Key Lab of Separation Sciences for Analytical Chemistry, National Chromatographic Research and Analysis Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
✉ Correspondence: hanfazou@dicp.ac.cn
Received April 12, 2012 Accepted May 7, 2012

ABSTRACT

The human serum proteome is closely associated with the state of the body. Endogenous peptides derived from proteolytic enzymes cleaving on serum proteins are widely studied due to their potential application in disease-specific marker discovery. However, the reproducibility of peptidome analysis of endogenous peptides is significantly influenced by the proteolytic enzymes within body fluids, thereby limiting the clinical use of the endogenous peptides. We comprehensively investigated the N and C terminus of endogenous peptides using peptidomics. The cleavage site patterns of the N and C terminus and adjacent sites from all the identified endogenous peptides were highly conserved under different sample preparation conditions, including long-term incubation at 37°C and pretreatment with repeated freeze-thaw cycles. Furthermore, a distinguishable cleavage site pattern was obtained when a different disease serum was analyzed. The conserved cleavage site pattern derived from proteolytic enzymes holds potential in highly specific disease diagnosis.

KEYWORDS human serum, endogenous peptide, N and C termini, disease diagnosis

INTRODUCTION

The number of proteins that can be analyzed by proteomics has increased dramatically in recent years due to the development of such novel techniques as nanoflow reversed-phase liquid chromatography coupled with tandem mass spectrometry (nano-RPLC-MS/MS). Currently, it is feasible to identify and quantify thousands of proteins in one analysis (Cox and Mann, 2008; Nilsson et al., 2010). Although mainstream proteomics has primarily focused on proteins, a subfield of proteomics named peptidomics (Schrader and Schulz-Knappe, 2001), which studies endogenous peptides that occur in cell lines, tissues, and body fluids, has recently emerged. The underlying hypothesis of peptidomics is that endogenous peptides reflect the biological processes occurring in a system, are particularly related to the specificity and activity of proteolytic enzymes, and thus might be used as markers of the state of the system. Since 2000, peptidomics has been applied in different studies, such as neuroendocrine research as well as biomarker and drug discovery (Diamandis, 2006; Soloviev and Finch, 2006). Serum is the most convenient clinical sample, and it is believed to contain various proteins from different tissues of the body. Therefore, serum is widely used in endogenous peptide analysis for biomarker discovery.

Although endogenous peptides reflect the state of the biological system, all the manipulation required for their analysis can lead to biased results. Recent research has shown that the collection, transportation, storage, and freeze-thaw cycles, among others, of serum samples can induce variations in endogenous peptidome analysis (West-Nielsen et al., 2005). This has prompted the Human Proteome Organization to provide guidelines for blood sampling and handling, such as depleting the platelets in plasma, aliquoting and storing the sample in liquid nitrogen, and adding prote-

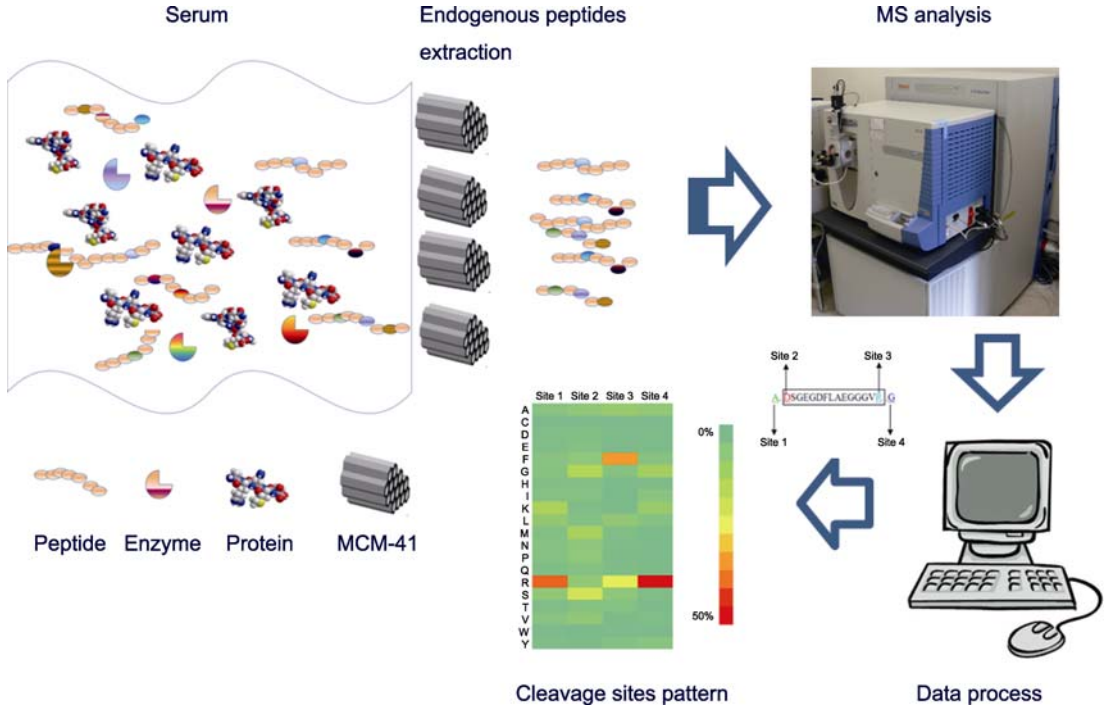

Figure 1. Schematic diagram of the experimental procedures.

ase inhibitors; however, controlling serum quality remains a great challenge (Rai et al., 2005; Ransohoff, 2005; Villanueva et al., 2005). Moreover, it is difficult to obtain consistent peptidome analyses using current proteomic technologies because (1) the presence of some endogenous peptides with high abundance, such as fibrinogen α, in serum samples severely hinders the detection of those with low abundance through ion suppression effects (Koomen et al., 2005) and (2) shotgun MS identification of peptides is always biased for most low-abundance ions in data-dependent analysis. These make it difficult to enhance the performance of serum endogenous peptide analysis in identifying peptide candidates as potential biomarkers for clinical use.

Variations in detection prevent the use of endogenous peptides as biomarkers; therefore, it is essential to develop a more robust way of investigating the nature of different serum samples. We postulated that protein and peptide degradation can be attributed mainly to the proteolytic enzymes present in serum and that the type and level (including cleavage specificity and activity) of these enzymes can differ during disease (Koomen et al., 2005; Villanueva et al., 2006; Doucet et al., 2008). Thus, the cleavage sites observed by MS might be more reliable differentiators of disease than the identified peptides. We combined peptidome analysis of sera by nano-RPLC-MS/MS with bioinformatic analysis of the two amino acids nearest the cleavage sites at the N and C termini of the observed endogenous peptides. Briefly, the N-terminal pre-cleavage site (site 1), N terminus (site 2), C terminus (site 3), and C-terminal post-cleavage site (site 4) of all the identified peptides were used to investigate the nature of proteolytic enzymes within specific human sera (Fig. 1). The percentage of every terminal amino acid in each site was calculated with spectral counts of the corresponding peptides, and the combined pattern of all percentages in the four cleavage sites gave the specific identity of one type of serum (Fig. 1). The cleavage site pattern was highly conserved and could be used to differentiate different types of sera.

Results

Endogenous peptidome analysis on different days

We first investigated the N and C termini as well as the presence of specific cleavage sites within the endogenous peptides of normal human sera. Briefly, 100 randomly selected human healthy sera were pooled, aliquoted, and stored at −80°C before the experiments. One aliquot of 200 μL of the pooled sera was used for endogenous peptide enrichment by an MCM-41 mesoporous material as previously described (Tian et al., 2007). One-dimensional nano-RPLC-MS/MS analysis of one-fifth of the enriched endogenous peptides was performed, and it identified 942 serum endogenous peptides (285 unique peptides). The identified peptide sequences could be clustered into different sets of ladders from different parent peptides with loss of one or more amino acid residues from the N or C terminus (Table

S1), which is consistent with previous reports (Koomen et al., 2005). Triplicate analyses of different aliquots of the pooled healthy sera were performed over approximately 2 months. Individual analysis identified 328 unique endogenous peptides on average (RSD = 47%), and overlaps of the unique peptides identified across duplicate and triplicate analyses were only 24% and 8%, respectively (Fig. 2). These results confirm that the variation in the endogenous peptides within sera over time is extensive due to differences in proteolysis and technology even under stringent conditions of collection and storage. In contrast, bioinformatic analysis of these data sets revealed that the cleavage site patterns remained very similar across the triplicate analyses even after the 2-month study period (Figs. 3 and S1). We believe that the fluctuations in the identified peptides were caused by the remaining protease activity within the serum that triggered endogenous peptides to be further degraded into a series of new peptides over time (Table S1). However, as the cleavage specificity of the proteases was stable for one type of serum, the cleavage site patterns were reproducible even if the endogenous peptides have degraded.

Cleavage site patterns under different sample preparation conditions

Long-term storage at room temperature as well as frequent freezing and thawing of serum samples are believed to significantly affect the results of peptidome analysis by shifting the balance from parent peptides to daughter peptides. To investigate the influence of long-term storage, we incubated the sera at 37°C for different times before endogenous peptide extraction. At first, we compared the results obtained from sera that were incubated at 37°C for 0, 24, and 108 h,

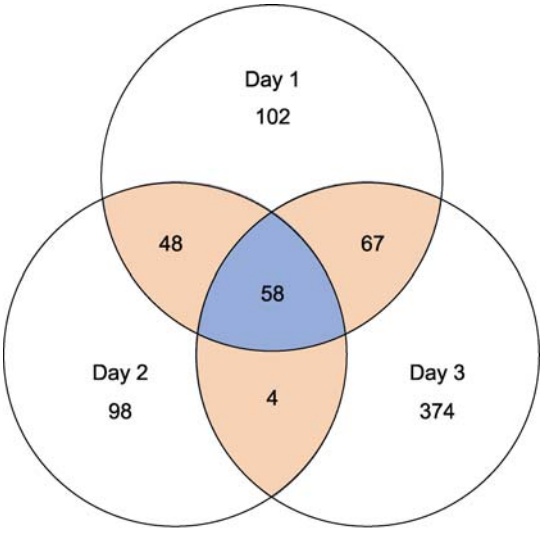

Figure 2. Peptide identification overlap in the thrice-daily analysis of endogenous peptides extracted from healthy human serum by nano-RPLC-MS/MS.

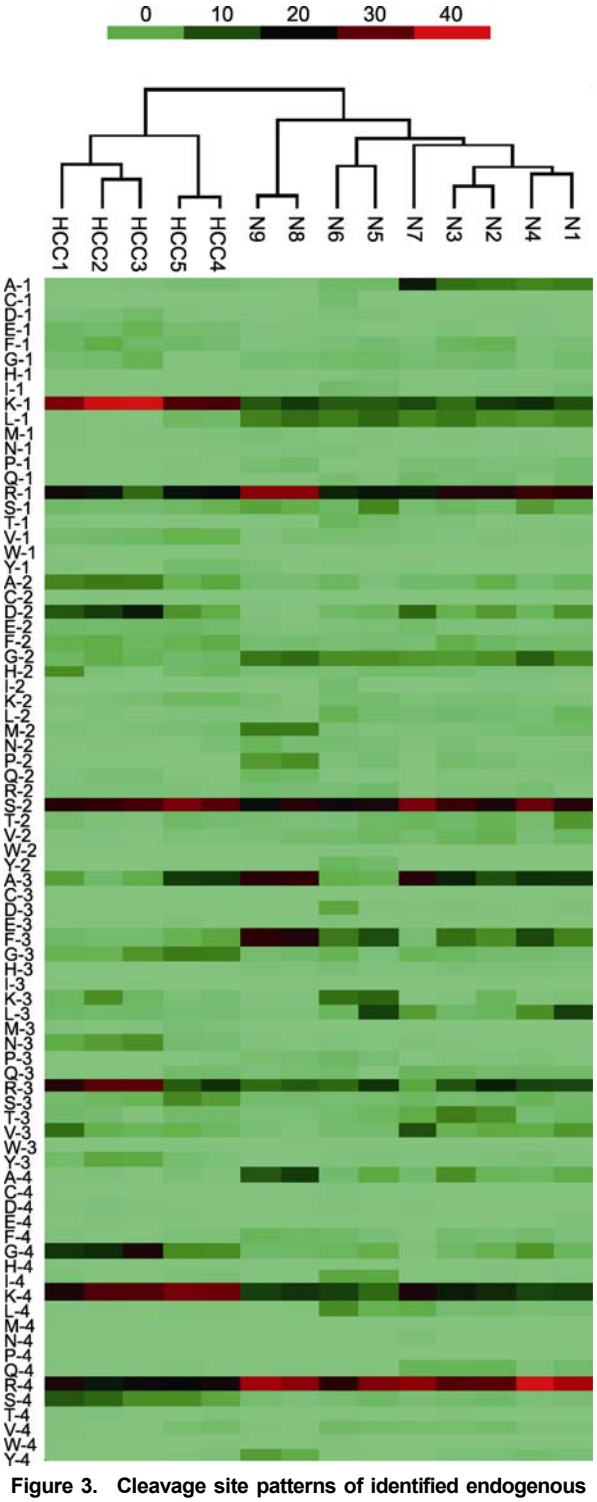

Figure 3. Cleavage site patterns of identified endogenous peptides in different human serum samples. A-x to Y-x, amino acid residues A to Y in cleavage site x; N1–N3, three day-to-day replicate analyses of healthy human serum; N4–N6, three replicate analyses of healthy human serum incubated at 37°C for 24 h; N7, analysis of healthy human serum incubated at 37°C for 108 h; N8 and N9, parallel analyses of healthy human serum with 20 cycles of freezing and thawing; H1–H3, three replicate analyses of HCC human serum; and H4 and H5, parallel analyses of HCC human serum with 20 freeze-thaw cycles.

for which the peptide identification overlap was only 17%. Subsequently, we performed 20 cycles of freezing and thawing prior to analysis and detected an endogenous peptide identification overlap of only 18%, compared with the peptidome of fresh serum. These results demonstrated that parent peptides rapidly degrade once the serum is thawed, which is consistent with previous reports (Rai et al., 2005; West-Nielsen et al., 2005). However, the combined cleavage site patterns still exhibited much better reproducibility than the identified endogenous peptides' sequences under different conditions (Figs. 3, S2 and S3). These data indicate that the cleavage site pattern is highly conserved and may better represent the cleavage specificity and activity of serum proteolytic enzymes under different sample preparation conditions compared with the sequences of endogenous peptides.

Cleavage site pattern of hepatocellular carcinoma (HCC) sera

Twenty HCC sera were also pooled, and the endogenous peptides were extracted and analyzed following the same procedure as described above. The combined cleavage site patterns of the identified endogenous peptides were checked (Figs. 3 and S4). Highly reproducible cleavage site patterns were obtained in different analyses for the HCC serum sample. We found that the cleavage site patterns for the healthy and HCC serum samples are differed. Briefly, compared with the healthy serum, the HCC serum had more K residues in sites 1 and 4, more S and G residues in site 4, less L and R residues in sites 1 and 4, and less F residues in site 3. These results suggest that proteolytic enzymes within HCC sera have different types and levels of specificity and that the protease activity also exhibits some difference. The cleavage site pattern has the potential to differentiate between different types of sera.

To test if the cleavage site patterns of different types of sera can be differentiated, we categorized nine analyses of healthy serum samples and five analyses of HCC serum samples in different sample preparation conditions as follows: N1–N3, three replicate analyses of healthy human serum; N4–N6, three replicate analyses of healthy human serum incubated at 37°C for 24 h; N7, analysis of healthy human serum incubated at 37°C for 108 h; N8 and N9, parallel analyses of healthy human serum with 20 freeze–thaw cycles; H1–H3, three replicate analyses of HCC human serum; and H4 and H5, parallel analyses of HCC human serum with 20 freeze–thaw cycles. The cleavage site patterns were then clustered using Perseus (v1.1.1.36) (Fig. 3). The nine healthy sera analyses could be clearly distinguished from the five HCC sera analyses. One-way ANOVA was then used to test the two groups of analyses in each amino acid residue among the four cleavage sites. The percentages of the 37 amino acid residues among the 80 residues significantly differed between groups ($P < 0.05$) (Table S2).

To further demonstrate if the cleavage site patterns can be used to differentiate between different types of disease sera, we mixed and analyzed seven breast cancer (BC) sera according to the above-described procedure. The BC serum exhibited more F residues in site 3 and more R residues in sites 1 and 4 (Fig. S5). The cleavage site patterns of three BC sera could be clearly distinguished from the HCC and healthy sera in the cluster analysis. Therefore, the cleavage site pattern is a unique characteristic of a specific serum, and it has the potential to differentiate between different types of sera.

Discussion

Although proteomics for cell lines and tissues has been developed greatly in recent years, peptidomics for body fluids is still hindered by poor reproducibility in sample collection, transportation, storage, and preparation. In each step, peptide degradation might be introduced by proteases within the body fluid. Many proteins associated with cancer have been identified using this method, such as apolipoproteins, complement components, and inter-α-trypsin inhibitors (Koomen et al., 2005; Zhu et al., 2010). In contrast, endogenous peptides originating from these proteins have not been used as cancer biomarkers due to the low disease specificity and large variations between samples. As it is difficult to obtain peptide markers for a specific serum status, we attempted to use the cleavage site pattern to represent the inherent property of the proteases within the serum. This strategy greatly increases the reproducibility of peptidome analysis and exhibits promising capability in discriminating between different types of serum samples and in disease diagnosis.

MATERIALS AND METHODS

Materials

Daisogel ODS-AQ (5 μm, 12 nm pores) was purchased from DAISO Chemical CO., Ltd. (Osaka, Japan). PEEK tubing, sleeves, microtees, and microcrosses were obtained from Upchurch Scientific (Oak Harbor, WA, USA). Fused silica capillaries with 75 and 200 μm I.D. were purchased from Polymicro Technologies (Phoenix, AZ, USA). All water used in the experiments was purified using a Mill-Q system from Millipore (Bedford, MA, USA). Dithiothreitol, iodoacetamide, and trypsin were obtained from Sigma (St. Louis, MO, USA). Formic acid (FA) was obtained from Fluka (Buches, Germany), whereas acetonitrile (ACN; HPLC grade) was from Merck (Darmstadt, Germany).

Endogenous peptide extraction

The serum samples used in this study were obtained from the Second Affiliated Hospital of Dalian Medical University (Dalian, China) and handled in compliance with the guidelines of the hospital's ethics committee. MCM-41 with a pore size of 2.1 nm was prepared as reported previously (Kresge et al., 1992; Tian et al., 2007). In total,

10 mg of MCM-41 particles well dispersed in 0.5 mL of water was mixed with 0.5 mL of diluted human serum (0.2 mL human serum diluted with 0.3 mL water) and shaken at room temperature for 1 h. After centrifugation at 15,000 *g* for 5 min, the supernatant was removed. The resulting mesoporous particles were washed with 1 mL of water three times, and the enriched peptides were eluted by 1 mL of 50% ACN aqueous solution, followed by lyophilization and redissolving into 50 μL of 0.1% FA aqueous solution (Zhu et al., 2010). The samples were stored at −30°C until use. For incubation testing, the serum samples were incubated at 37°C for 24 or 108 h before endogenous peptide enrichment. For freeze-thaw testing, the serum samples were thawed in room temperature and then frozen at −30°C 20 times, followed by endogenous peptide enrichment as described above.

Nano-RPLC separation and MS detection conditions

The nano-RPLC separation was operated on a quaternary Surveyor MS pump (Thermo, San Jose, CA), and the flow rate after splitting was adjusted to 200 nL/min. A homemade 75 μm I.D. capillary column with an electrospray tip was packed with 12-cm-long C18 AQ particles (3 μm, 120 Å) and then used as separation column. A 200 μm I.D. capillary column with a homemade monolithic frit was packed with 4-cm-long C18 AQ particles and used for automated sample injection as described previously (Wang et al., 2010). Ten microliters of the enriched peptides was injected in all of our experiments. Next, 0.1% FA aqueous solution (mobile phase A) and ACN with 0.1% FA (mobile phase B) were used for RP binary gradient separation. The binary gradient was set as follows: 0%–10% mobile phase B for 2 min, 10%–35% for 90 min, and 35%–80% for 5 min. After elution with 80% mobile phase B for 10 min, the column was equilibrated with mobile phase A for 13 min. The separation column was coupled using LTQ-Orbitrap MS (Thermo, San Jose, CA) directly by the electrospray tip, and a 1.8 kV spray voltage was applied. The temperature of the ion transfer capillary was 200°C, and the normalized collision energy was set to 35.0%. One microscan was recorded for each MS and MS/MS. The MS/MS spectra were acquired in data-dependent analysis mode, and the six most intense peaks in full MS (400–2000) were selected for MS/MS fragmentation. The dynamic exclusion function was set as follows: repeat count, 2; repeat duration, 30 s; and exclusion duration, 90 s.

Database search and data processing

The *.raw files collected by LTQ-Orbitrap were converted to *.mgf files using DTASupercharge (v2.0a7) and then searched against the IPI Human v3.52 database using Mascot Version 2.3 (Matrix Science). No enzyme and static modification was set for the database search, and oxidation M was selected as a variable modification. The mass tolerance levels were 5 ppm for parent ions and 0.8 Da for fragment ions. After the database search, the filter criteria were established as follows: rank 1, bold red required, $P < 0.05$, and ion scores > 30 to control false-positive discovery rates of less than 1% for the peptides. The filtered results for the peptides were exported in *.CSV file, and the cleavage site pattern was analyzed using Microsoft Excel. Cluster analysis and one-way ANOVA were performed using Perseus (v1.1.1.36).

ACKNOWLEDGEMENTS

This study was supported by the Creative Research Group Project of National Natural Science Foundation of China (Grant No. 21021004), the National Basic Research Program (973 Program) (Nos. 2012CB910601, 2012CB910101), and the Analytical Method Innovation Program of MOST (No. 2010IM030500) (H.Z.), as well as the "Hundred Talent Young Scientist Program" of DICP (F.W.).

REFERENCES

Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech 26, 1367–1372.

Diamandis, E.P. (2006). Peptidomics for cancer diagnosis: present and future. J Proteome Res 5, 2079–2082.

Doucet, A., Butler, G.S., Rodriguez, D., Prudova, A., and Overall, C.M. (2008). Metadegradomics toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Mol Cell Proteomics 7, 1925–1951.

Koomen, J.M., Li, D.H., Xiao, L.C., Liu, T.C., Coombes, K.R., Abbruzzese, J., and Kobayashi, R. (2005). Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. J Proteome Res 4, 972–981.

Kresge, C.T., Leonowicz, M.E., Roth, W.J., Vartuli, J.C., and Beck, J.S. (1992). Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. Nature 359, 710–712.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., Bairoch, A., and Bergeron, J.J.M. (2010). Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Meth 7, 681–685.

Rai, A.J., Gelfand, C.A., Haywood, B.C., Warunek, D. J., Yi, J., Schuchard, M.D., Mehigh, R.J., Cockrill, S.L., Scott, G.B. I, Tammen, H., et al. (2005). HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5, 3262–3277.

Ransohoff, D.F. (2005). Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 5, 142–149.

Schrader, M. and Schulz-Knappe, P. (2001). Peptidomics technologies for human body fluids. Trends Biotechnol 19, 55–60.

Soloviev, M. and Finch, P. (2006). Peptidomics: bridging the gap between proteome and metabolome. Proteomics 6, 744–747.

Tian, R.J., Zhang, H., Ye, M.L., Jiang, X.G., Hu, L.H., Li, X., Bao, X.H., and Zou, H.F. (2007). Selective extraction of peptides from human plasma by highly ordered mesoporous silica particles for peptidome analysis. Angew Chem Int Ed 46, 962–965.

Villanueva, J., Philip, J., Chaparro, C.A., Li, Y.B., Toledo-Crow, R., DeNoyer, L., Fleisher, M., Robbins, R.J., and Tempst, P. (2005). Correcting common errors in identifying cancer-specific serum peptide signatures. J Proteome Res 4, 1060–1072.

Villanueva, J., Shaffer, D.R., Philip, J., Chaparro, C.A., Erdjument-Bromage, H., Olshen, A.B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J., Sanchez-Carbayo, M., Holland, E.C., Cordon-Cardo, C., Scher, H.I., and Tempst, P. (2006). Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin

Invest 116, 271–284.

Wang, F.J., Chen, R., Zhu, J., Sun, D.G., Song, C.X., Wu, Y.F., Ye, M.L., Wang, L.M., and Zou, H.F. (2010). A fully automated system with online sample loading, isotope dimethyl labeling and multi-dimensional separation for high-throughput quantitative proteome analysis. Anal Chem 82, 3007–3015.

West-Nielsen, M., Høgdall, E.V., Marchiori, E., Høgdall, C.K., Schou, C., and Heegaard, N.H.H. (2005). Sample handling for mass spectrometric proteomic investigations of human sera. Anal Chem 77, 5114–5123.

Zhu, J., Wang, F.J., Dong, X.L., Ye, M.L., and Zou, H.F. (2010). A strategy with label-free quantification of the targeted peptides for quantitative peptidome analysis of human serum. Sci Chin A-Chem 53, 759–767.

