Quantitative detection of single amino acid polymorphisms by targeted proteomics

Zhi-Duan Su 1,2, Liang Sun 3, Dan-Xia Yu 3, Rong-Xia Li 1, Huai-Xing Li 3, Zhi-Jie Yu 3, Quan-Hu Sheng 1, Xu Lin 3,4, Rong Zeng 1,4,*, and Jia-Rui Wu 1,2,4,5,*

Introduction

Single-nucleotide polymorphisms (SNPs) are recognized as one kind of major genetic variants in population scale. However, polymorphisms at the proteome level in population scale remain elusive. In the present study, we named amino acid variances derived from SNPs within coding regions as single amino acid polymorphisms (SAPs) at the proteome level, and developed a pipeline of non-targeted and targeted proteomics to identify and quantify SAP peptides in human plasma. The absolute concentrations of three selected SAP-peptide pairs among 290 Asian individuals were measured by selected reaction monitoring (SRM) approach, and their associations with both obesity and diabetes were further analyzed. This work revealed that heterozygotes and homozygotes with various SAPs in a population could have different associations with particular traits. In addition, the SRM approach allows us for the first time to separately measure the absolute concentration of each SAP peptide in the heterozygotes, which also shows different associations with particular traits.

Keywords: single amino acid polymorphism (SAP), proteomics, selected reaction monitoring (SRM), diabetes, obesity

Received June 16, 2011. Revised July 21, 2011. Accepted August 5, 2011.
* The Author (2011). Published by Oxford University Press on behalf of Journal of Molecular Cell Biology, IBCB, SIBS, CAS. All rights reserved.
amino acids of proteins can be detected both qualitatively and quantitatively at the proteome level (Figure 1A). Here, we apply a term of ‘SAP’ for single amino acid polymorphism for simplicity. Although Mendelian point mutations of non-synonymous amino acids that cause functional change of proteins such as hemoglobin of anemia (Kutlar, 2007) or β-amyloid precursor protein of Alzheimer’s disease (di Fede et al., 2009) are extensively studied, little attention has been paid to detect SAPs at population level so far. In our work, a SAP database of human plasma proteome was constructed based on high-confidence Swiss-Prot annotation and proteins that were systematically analyzed by shotgun proteomics strategy from 33 selected Asian plasma samples (Figure 1B). To further reveal the properties of SAPs in population level, we applied a targeted proteomic approach, selected reaction monitoring (SRM) (Addona et al., 2009; Picotti et al., 2009), to both qualitatively and quantitatively analyze three pairs of SAP peptides in total 290 Asian plasma samples, and their associations with both obesity and diabetes were further analyzed.

Results

Discovery and validation of SAP peptides

For human plasma, it is essential to establish a reliable plasma-related SAP database. Based on the shotgun 2D-LC-MS/MS results of proteome from 33 human plasma samples, 717 high-confidence plasma proteins were identified after searching against a human International Protein Index (IPI) database. Then, the proteins whose identified sequence coverage was more than 40% and complement system-related proteins as well as apolipoproteins (Li et al., 2008) were reserved, while the antibody-related proteins were excluded. According to Swiss-Prot high-confidence SNP annotation, we established a reliable human plasma-specific SAP database including 439 SAPs, which matched with 29 plasma proteins. Using the same search engine, SEQUEST, all MS/MS spectra were searched against this human plasma-specific SAP database. As a result, 2154 SAP peptides corresponding to 46 sequence-unique SAP peptides were identified. This first step database searching was considered as a ‘Discovery’ phase (Supplementary Figure S1A).

Using the in-house software BuildSummary, all the 2154 spectra were extracted again, and searched with SEQUEST algorithm against another ‘total’ database, which combined the human IPI database (version 3.07) with the SAP database mentioned above. Finally, by matching 35 sequence-unique SAP peptides, 2029 SAP peptides were reserved; out of which, 24 sequence-unique SAP peptides corresponding to ‘wild-type’ peptides were also identified (Supplementary Table S1). This second step of database searching was considered as a ‘Validation’ phase (Supplementary Figure S1A).

In the ‘Validation’ phase, there were four cases as follows (Supplementary Figure S1B). Case 1: the identified peptides against the SAP database were exactly the same as that against the ‘total’ database, and these SAP peptides were reserved for further analysis. Case 2: the identified peptide sequences against the SAP database were shared with the peptide sequences of different proteins in ‘total’ database. These SAP peptides were discarded. Case 3: the identified peptides against the ‘total’ database were different from those against the SAP database. It is because the identified peptides against the ‘total’ database were scored higher and deltaCn was no less than 0.1. In this case, these SAP peptides were discarded. Case 4: the identified peptides against the SAP database were not identified from the ‘total’ database. It is because these peptides did not meet the filter conditions, such as deltaCn <0.1. These SAP peptides were discarded as well.

Qualitative distribution of SAP peptides in subpopulation

To further reveal the properties of SAPs in population level, we applied SRM approach for total 290 Asian plasma samples to analyze three pairs of SAP peptides selected from above 24 SAP-peptide pairs both qualitatively and quantitatively, all of which belong to the complement system in the serum (Table 1 and Supplementary Table S2). According to the amino acid sequences of these SAP-peptide pairs, six artificial peptides were synthesized with stable-isotope-labeled amino acids. The similarities of MS/MS spectra as well as the relative pattern of transitions derived from the endogenous peptides and

<table>
<thead>
<tr>
<th>Protein Description of selected SAP-peptide annotation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Complement component C7</td>
</tr>
<tr>
<td>Complement factor H</td>
</tr>
<tr>
<td>Complement component C5</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
synthesized peptides indicated the accurate identification SAP peptides in plasma (Supplementary Figure S2A–F).

All the 290 Asians could be divided into four phenotypes: diabetes-normal weight (D, n = 36), normal plasma glucose-overweight/orbidity (O, n = 109), diabetes-overweight/orbidity (D&O, n = 109), and normal subjects with both normal plasma glucose and normal weight (N, n = 36) (Supplementary Table S3). We analyzed the variable distribution of homozygote and heterozygote of SAP peptides in the Asian individuals (Figure 2A). For complement component C5, the percentages of the heterozygous SAP peptide in diabetic and overweight/obese subjects were significantly higher than those in the normal subjects (Figure 2A). On the other hand, in the cases of complement component C7 and complement factor H (CFH), there is no significant difference in the percentages of the heterozygous SAP peptides in these four sub-populations (Figure 2A).

Furthermore, we analyzed each homozygote distribution of the SAP-peptide pairs in sub-populations. The results indicated that both homozygous C7_587T and CFH_62V were dominant in this selected Asian population, and the percentage of the C7_587P homozygote in the normal subjects was significantly higher than that in the other three metabolic disorder subjects (Figure 2B). Interestingly, in the case of C5 protein, there was no C5_966Y homozygote found in any individual (Figure 2B), which was in agreement with a previous observation of widespread monoallelic expressions in human autonomies (Gimelbrant et al., 2007).

Quantitative distribution of SAP peptides in homozygous individuals

Taking advantage of SRM approach, we could detect each SAP peptide in the plasma of these 290 Asians in terms of the absolute concentration of peptides (Supplementary Table S3). Figure 3 demonstrates the quantitative distribution of SAP peptides in homozygous individuals. For C7_587T homozygous subjects, the concentration of C7_587T peptide was significantly over-represented in the plasma of the diabetic-overweight/obese individuals, whereas the C7_587P peptide level did not present any significant difference within four sub-populations in C7_587P homozygous. These results suggest that each homozygous form of a SAP-peptide pair might have independent association with a particular physiological or pathological trait. In the case of CFH, both of the SAP-peptide pair, CFH_62V and CFH_62I, were significantly higher in overweight/obese and diabetic-overweight/obese sub-populations than that of normal sub-populations. In addition, there was no C5_966Y homozygote detected in any individuals, and the other SAP form, C5_966D, also did not show any significant variance in the four sub-populations (Figure 3).

The lack of significance for some SAP peptides (C7_587P and CFH_62I in D sub-population) was inconclusive due to the small number of homozygous with certain mutation. Therefore, a larger population would be needed to assess its significance.

Quantitative distribution of SAP peptides in heterozygous individuals

Making use of SRM results, we also analyzed the quantitative distribution of SAP peptides in heterozygous sub-populations (Figure 4). In the case of C7 SAP-peptide pair in heterozygous subjects, the concentration of C7_587T peptide, rather than C7_587P peptide, was significantly over-represented in diabetic-overweight/obese individuals, which was similar with that shown in homozygous individuals (Figure 3). However, C5 SAP-peptide pair in heterozygous subjects did not show any significant difference. On the other hand, for CFH heterozygous SAP-peptide pair detected in total 139 individuals, the concentrations of both CFH_62V and CFH_62I peptides were significantly higher in overweight/obese and diabetic-overweight/obese sub-populations than that in normal ones (Figure 4), which indicated the association of CFH expression with obesity.

Interestingly, the sum of both SAP-peptide concentrations of the CFH in heterozygous individuals showed similarly significant variances among the four sub-populations as either single SAP-peptide expression did (Figure 4), whereas the sum of both SAP-peptide concentrations of the C7 heterozygote showed weaker significant variances among the four sub-populations than that the C7_587T expression alone showed (Figure 4).

In addition, we also calculated the allelic ratio of heterozygous SAP-peptide expression, in which SAP peptides (C7_587T, CFH_62V, and C5_966D) with higher ‘penetrance’ were considered as numerators. There was no significant difference of average allelic ratios of C7 and C5 SAP-peptide expression (Figure 4). Importantly, the average allelic ratio of CFH in normal sub-population was significantly higher than that in overweight/obese and diabetic-overweight/obese sub-populations.

Figure 2 Qualitative distribution of SAP peptides in sub-populations.
(A) The composition of heterozygote/homozygote with selected SAP peptides for C7, CFH, and C5 in each sub-population, including N (Normal individuals), D (Diabetic individuals), O (Overweight/Obese individuals), and D&O (Diabetic-Overweight/Obese individuals).
(B) Distribution of SAP-peptide pairs of C7, CFH, and C5 in homozygous population. In A and B, these sub-populations were compared with χ² test, respectively, and all the P-values were adjusted via Holm correction algorithm.
Discussion

The SNP analysis at the genome level has been extensively studied to reveal their associations with particular traits or diseases (Kubo et al., 2007; Sun et al., 2008; di Fede et al., 2009), which is indirect and non-quantitative. On the other hand, the SAPs play more important roles in affecting functions of genes and present the direct association with phenotypes. Furthermore, the quantitative traits should be ultimately determined by the expressed proteins. For example, di Fede et al. (2009) found a β-amyloid precursor protein mutation (APP_A673V) that caused familial Alzheimer's disease only in the homozygous state, while heterozygous carriers were unaffected since the wild-type peptide could interact with the mutant peptide. In the present work, we combined non-targeted and targeted proteomics strategy, both qualitatively and quantitatively detected SAP peptides in population level.

In the present work, these three selected pairs of peptide variants represented different features of SAPs. For C7, the qualitative analysis already indicated the role of C7_587P against obesity and diabetes, and quantitative data both in homozygote and heterozygote also demonstrated the lack of association of C7_587P with obesity and diabetes. On the other hand, C7_587T showed significant association with obesity and diabetes, which displayed inconsistent distribution with C7_587P. Therefore, the...
association of this protein with particular traits of a population could not be accurately measured by classical protein-level assays such as ELISA or western blotting, when the concentrations of SAP-peptide pair presented discrepancy in distribution significance. Regarding to CFH, both of the SAP-peptide pair showed positive correlation with obesity either in homozygote or in heterozygote. However, the ratio of these two SAP peptides showed differential distributions associated with obese in heterozygous subjects. These results suggested that not only the respective concentration of each SAP might be related to phenotype, but the relative ratio between SAP variants could be another dimension to consider the linking of SAP with phenotype. As for C5, though no quantitative results showed the significance of either peptide associated with pathological traits possibly due to the small sample size or the low frequency of C5_966Y, the qualitative analysis surprisingly indicates the protecting role of C5_966D to obesity and diabetes, since there were more homozygous subjects (C5_966D) in the normal sub-population.

The substitution of amino acids at protein level, especially at certain important domains, may alter the normal structures of proteins, and eventually result in specific physiological or pathological changes, which may arise from either alteration of post-translational modifications or abundances of functional proteins. In the present study, we selected three pairs of SAP variants to measure their absolute concentrations in Asian plasmas. In the case of C7, the non-polar proline of 587th site was substituted for the polare threonine. This dramatic change might alter the structure of C7, and further affected the normal function of C7 in plasma. In our SRM data, this pair of SAP variants was tightly associated with obesity and diabetes (Figures 2–4). For CFH, we found a substitution lied in 62th site, where valine was changed to isoleucine. The structures of these two amino acids were similar, and this alteration between them may not affect the structure of whole protein dramatically. However, it has been reported that this mutation, V62I, is strongly associated with polypoidal choroidal vasculopathy (Kondo et al., 2009) and some other diseases (Hageman et al., 2005). In the present work, we found that both CFH_62V and CFH_62I as well as the ratio of their concentrations were significantly associated with obesity.

In conclusion, we deliver a new concept of SAP representing non-synonymous SNP at the proteome level (Figure 1A), which defines a new direction to directly analyze polymorphisms of proteins in population. Our qualitative analysis indicates that the heterozygous and homozygous SAPs in a population could have various significant associations with particular traits, and therefore should be separately analyzed. Furthermore, the SRM approach allows us for the first time to separately measure the absolute concentration of each SAP peptide in heterozygote, which also shows various significant associations with particular traits. Importantly, the plasma-SAP analysis based on the SRM approach might provide a new kind of biomarkers for detecting physiological or pathological traits.

Materials and methods

Sample collection

The human plasma samples of this study were derived from a population-based study including 1059 Chinese men and women aged 35–54 years (Sun et al., 2010), which was approved by the Institutional Review Board of the Institute for Nutritional Sciences. The recruitment, anthropometric measurements, and blood sample collection were described previously (Sun et al., 2010).

Overweight/obesity was defined as body mass index ≥24.0 kg/m². Type 2 diabetes was defined as fasting plasma glucose (FPG) ≥7.0 mM or 2 h post-load plasma glucose ≥11.1 mM during an oral glucose tolerance test. Impaired fasting glucose (IFG) was defined as 5.6 mM ≤ FPG < 7.0 mM. Sample preparation and digestion

Thirty-three human plasma samples for shotgun proteomic analysis were obtained from the following six groups: five healthy donors with both normal plasma glucose and normal weight (Group 1), five IFG-normal weight donors (Group 2), five diabetic-normal weight donors (Group 3), six normal plasma glucose-overweight/obese donors (Group 4), six IFG-overweight/obese donors (Group 5), and six diabetic-overweight/obese donors (Group 6). In each group, equal volumes of plasma samples were mixed together. Using the method of ethanol precipitation developed by previous work (Colantonio et al., 2005), a majority of albumin was removed. Each 50 μl pooled plasma was diluted with 250 μl buffer (10 mM NaCl, 100 mM HEPES, pH 7.4) at 4°C before it was filtered by 0.22 μm film (Agilent Technologies). To reach a final concentration of 42% of ethanol buffer, 180 μl pre-chilled ethanol was added into 260 μl filtrate, and incubated at 4°C for 1 h. The plasma proteins in this ethanol buffer were centrifuged at 16000 g for 45 min at 4°C. The pellet component was collected as the albumin-depleted fraction. The pellet fraction was then lyophilized and re-suspended in 200 μl lysis buffer (8 M urea, 4% CHAPS, 40 mM Tris-base, 65 mM DTT). The disulfide bonds were reduced by DTT at 37°C for 2.5 h, then the mixture was added with 10 μl of 1 M IAA and incubated for an additional 40 min at room temperature in darkness. After that, the protein mixtures were subjected to five volumes of pre-chilled precipitation solution containing 50% acetone, 50% ethanol, and 0.1% acetic acid, and incubated at −20°C for 20 h. Then the mixture was centrifuged at 15000 g for 1 h, and the precipitate was resolved in 300 μl of 50 mM NH₄HCO₃. Trypsin (1:25, enzyme to protein) was added and incubated with the mixture at 37°C for 24 h. After ultrafiltration and lyophilization, the tryptic peptide mixtures were collected.

For SRM assay, 145 pairs of age-, gender-, and obese status-matched diabetic cases and controls were obtained from 36 healthy donors with both of normal plasma glucose and normal weight (N), 36 diabetic-normal weight donors (D), 109 normal plasma glucose-overweight/obese donors (O), and 109 diabetic-overweight/obese donors (D&O). Each 3 μl plasma was diluted with 97 μl lysis buffer (8 M urea, 40 mM Tris-base, 20 mM DTT), and the disulfide bonds were reduced by DTT at 37°C for 2.5 h. Then the mixture was added with 5 μl of 1 M IAA and incubated at room temperature in darkness for 40 min. After that, the protein mixtures were transferred to a 10 k filter (Pall Corporation) and centrifuged at 10000 g for 1 h. The filter was washed by adding 200 μl of 50 mM NH₄HCO₃ again. Finally, 100 μl of 50 mM NH₄HCO₃ and trypsin (1:25, enzyme to protein) was added and incubated with the mixture at 37°C for
20 h. The tryptic peptide mixtures were collected for further analysis.

The online 2D-LC-MS/MS system

The peptide fractionation and mass spectrometric identification were performed on the automated online continuous pH and reverse-phase gradient SCX–RP–LC (SCX, strong cation exchange; RP, reverse phase; LC, liquid chromatography) system coupled with a mass spectrometer (MS) (Zhou et al., 2007). The SCX solvents were pH 2.5 buffer (A) and pH 8.5 buffer (B), and the RP solvents were 0.1% formic acid in either water (A) or acetonitrile (B). Briefly, ~10 μl of plasma-derived protein digests was loaded onto the SCX column (320 μm × 100 mm; Column Technology Inc.) in SCX buffer A and eluted by a continuous pH gradient (from pH 2.5 to 8.5). Ten fractions were eluted to two C18 RP trap columns (300 μm × 5 mm; C18, 5 μm; Agilent Technologies) alternatively, followed by one C18 RP analytical column (75 μm × 150 mm; C18, 5 μm; Column Technology Inc.). A linear ion trap MS (Thermo Electron) was operated in a data-dependent mode. The MS was set that each full MS scan was followed by 10 MS/MS scans on the 10 most intense ions from the full MS spectrum with the following Dynamic Exclusion settings: repeat counts; 2; repeat duration, 30 sec, exclusion duration, 120 sec. Two duplicates of each plasma digestion were performed.

Database searching

All MS/MS spectra were searched with SEQUEST algorithm against the human IPI database (version 3.51), in which each genuine protein sequence was followed by a reversed amino acid sequence. Carboxymethylation (57.0215 Da) was searched as a fixed modification on cysteine, and oxidation (15.9949 Da) was set as a variable modification on methionine. Only one missing cleavage site was allowed. Using in-house software BuildSummary, the output results were combined together and filtered. The peptide and protein false discovery rate was fixed at ≤0.01 and 0.05, respectively, and deltaCn was fixed at ≥0.1.

Plasma SAP database searching

Using the same search engine, SEQUEST, all MS/MS spectra were searched against the human plasma-specific SAP database. The searching parameter was the same as mentioned above. Finally, all the accepted results must meet the following criterion: Xcorr ≥ 1.9, 2.2, and 3.75 for charge +1, +2, and +3 of precursor ion, respectively; deltaCn ≥ 0.1.

Internal standard SAP peptides preparation

Peptides DGFVQDEGPMFPVGGK(13C5,15N2-labeled), DGFVQDEGTMFVGGK(13C5,15N2-labeled), IPLYLV(13C5,15N-labeled)PK, IPLDLV(13C5,15N-labeled)PK, SL(13C6,15N-labeled)GNIVMVCN, and SL(13C6,15N-labeled)GNMVICNR were synthesized with stable-isotope labeling as internal standards (IS) by GL Biochem Ltd. After purification and lyophilization, certain milligrams of peptides were obtained by accurate weighting. The absolute amounts of IS peptides were adjusted by their purities, respectively, which were from 95.18% to 97.80%. The reduction and alkylation process of IS peptides was the same as those mentioned above. After that, IS peptides were desalted by C18 cartridges (C18, 55 μm; Phenomenex) and lyophilized. As a result, an aliquot of 1 ng DGFVQDEGPMFPVGGK, 1 ng DGFVQDEGTMFVGGK, 1 ng IPLYLVPK, 1 ng IPLDLVPK, 25 ng SLGNNMVCR, and 25 ng SLGNNVMVCN was mixed and spiked into 0.05 μl plasma-derived tryptic peptides for SRM assay.

SRM assay

RP-HPLC fractionation was performed on a PicoRfrit column (75 μm × 100 mm; C18, 5 μm; tip, 15 μm; New Objective) using a gradient B (0.1% formic acid in acetonitrile) from 2% to 35% in 30 min at a flow rate of ~300 ml/min after split. A triple quadrupole MS (TSQ Vantage, ThermoFisher Scientific) was used to acquire SRM data. For ionization, 1800 V of spray voltage and 200°C of capillary temperature were set. The selectivity of Q3 and Q2 was set at 0.2 and 0.7 Da (FWHM), respectively, and the collision gas pressure of Q2 was set at 1.2 mTorr argon. Using the following formulas, the collision energy (CE) was calculated.

\[ CE = 0.034 \times \frac{m}{z_{precursor \ ion}} + 3.314, \text{ for charge} +2 \text{ precursor ion} \]

and

\[ CE = 0.044 \times \frac{m}{z_{precursor \ ion}} + 3.314, \text{ for charge} +3 \text{ precursor ion} \]

For each peptide, five SRM transitions were monitored, and a scan time of 15 ms was used. The best transition was used to quantify the corresponding SAP peptides. It was considered as NA if the signal-to-noise ratio (S/N) was < 3 (Keshishian et al., 2007).

Determining limit of detection and limit of quantification

To determine limit of detection (LOD) and limit of quantification (LOQ) of each SAP peptide, a dilution series of IS peptides were spiked into 1 μl 20-fold diluted plasma-derived tryptic peptides, respectively. The amount of IS peptides (X-axis) to the heavy/light peak area ratio (Y-axis) was plotted to acquire a calibration curve. LOD was considered as the lowest amount of the calibration curve (Kuzyk et al., 2009), and LOQ was considered as one-third of LOQ (Addona et al., 2009).

Supplementary material

Supplementary material is available at Journal of Molecular Cell Biology online.

Funding

This work is supported by the grants from the Ministry of Science and Technology (2011CB910200, 2011CB910601), a grant from the National Natural Science Foundation of China (30821065), and the grants from the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX1-YW-02, KJCX2-YW-M15).

Conflict of interest: none declared.

References


Detecting single amino acid polymorphisms


