Comparison of Global Chemical Modifications of the Human Plasma Proteome Across Two Different Age Groups

Liu Yongtao¹, Zhao Mindi², Pan Xuanzhen¹, Gao Youhe^{1*}

2020-02-11

ChinaRxiv

Abstract

Protein chemical modification refers to covalent group reactions involving its amino acid residues or chain termini, thereby altering the molecular structure and its functions in regulation and information transmission. This study employed LC-MS/MS technology combined with an unrestricted modification identification algorithm to compare differences in the overall chemical modification levels of human plasma proteomes between two age groups. The study found that succinylation and phosphorylation modifications of cysteine, as well as the substitution of lysine with threonine, were significantly higher in the older group compared to the younger group, whereas carbamylation modification of lysine was lower in the older group. The thiol group in cysteine residues is an important moiety for disulfide bond formation and maintenance of protein structure, and the aforementioned cysteine-related modifications involve participation of the thiol group. Lysine is a basic amino acid, and modifications on its amino group alter the acid-base properties of proteins, potentially affecting protein structure and function. In summary, this study identified four types of protein molecular modifications and substitutions with significant differences in the plasma proteome across different age groups. It is speculated that the increase of certain modified proteins in the blood of elderly individuals may reflect an accumulation of proteins that have lost normal function, which may be one of the reasons why the risk of age-related diseases such as metabolic disorders and tumors is higher in the elderly than in young people.

Full Text

Comparison of Global Chemical Modifications of the Human Plasma Proteome Across Two Different Age Groups

Liu Yongtao1, Zhao Mindi2, Pan Xuanzhen1, Gao Youhe1*

1 (Department of Biochemistry and Molecular Biology, Beijing Key Laboratory of Gene Engineering Drugs and Biotechnology, College of Life Sciences, Beijing Normal University, Beijing 100875)

2 (Department of Laboratory Medicine, National Geriatrics Center, Beijing Hospital, Beijing 100730)

Abstract:
Objective: Chemical modifications of proteins refer to covalent group reactions involving their amino acid residues or chain termini, thereby altering the molecular structure and the functions they perform in regulation and information transmission. The human plasma proteome contains numerous types of chemical modifications, such as phosphorylation, methylation, and acetylation.
Methods: This study employed LC-MS/MS technology combined with an un-

restricted modification identification algorithm to compare differences in the overall levels of chemical modifications in the plasma proteomes of two groups at different age stages. **Results:** The study found that, in the older group, succinylation and phosphorylation of homocysteine, as well as the modification in which lysine is replaced by ornithine, were significantly higher than in the younger group, whereas the carbamylation modification of lysine was lower than in the younger group. The sulfhydryl group in homocysteine residues is an important group for forming disulfide bonds and maintaining protein structure, and the above modifications related to homocysteine involve participation of the sulfhydryl group. Lysine is an alkaline amino acid, and modifications on its amino group alter the acid-base properties of proteins, potentially affecting protein structure and function. **Conclusion:** This study identified four types of protein molecular modifications and substitutions with significant differences in the plasma proteomes of people at different age stages. It is inferred that older individuals have an increase in certain modified proteins in the blood, which may reflect an increase in proteins that have lost normal function in the blood. This may be one of the reasons why older individuals have a higher probability than younger people of developing age-related diseases such as metabolic diseases, cardiovascular and cerebrovascular diseases, and tumors.

Keywords: plasma; proteome; chemical modification; tandem mass spectrometry; unrestricted modification identification

Classification No.: Q516 R34

Comparison of global chemical modifications of human plasma proteome at two different age groups

Yongtao Liu1, Mindi Zhao2, Xuanzhen Pan1, Youhe Gao1*

1 (Department of Biochemistry and Molecular Biology, Beijing Key Laboratory of Gene Engineering Drug and Biotechnology, College of Life Science, Beijing Normal University, Beijing 100875, China)

2 (Department of Laboratory Medicine, National Geriatrics Center, Beijing Hospital, Beijing 100730, China)

Abstract:
Objective The chemical modifications of proteins refer to the covalent group reaction involved in its amino acid residues or chain ends, which in turn change

---

Gao Youhe (1964.06–), male, Ph.D., professor, doctoral supervisor, recipient of the National Science Fund for Distinguished Young Scholars; currently Director of the Beijing Key Laboratory of Gene Engineering Drugs and Biotechnology,

and Editor-in-Chief of *Urine*. Email: gaoyouhe@bnu.edu.cn. This work is one of the research outcomes of the National Key R&D Program projects "Development and Application Demonstration of Key Technologies for Proteomics Data Quality Control" and "Development of New Technologies for Early Diagnosis of Alzheimer' s Disease" (Project Nos. 2018YFC0910202 and 2016YFC1306300).

the molecular structure and function. There are many types of molecular modifications in the human plasma proteome, such as phosphorylation, methylation, acetylation, etc.

Methods In this study, two groups of human plasma proteome samples from different age groups were used to compare global chemical modifications by LC-MS/MS combined with non-limiting modification identification algorithms.

Results A total of 4 molecular modifications were found to have statistical differences. The succinylation and phosphorylation modifications of cysteine (Cys, C), and the modification of lysine (Lys, K) with threonine (Thr, T), were significantly higher in the old group than in the young group, while the carbamylation of lysine was lower in the young group. The sulfhydryl group in the cysteine residue is an important group for forming disulfide bonds and maintaining the structure of the protein. All differential cysteine-related modifications may cause structural and functional changes. Lysine is a basic amino acid, and modification of its amino group will change the charge state of the protein, which may affect the structure and function of the protein.

Conclusions In summary, four types of protein chemical modifications and substitutions were found to be differential in the plasma proteome across different age groups. We speculate that certain modified proteins increase in the blood of elderly people, which in turn changes the functions of these proteins. This may be one of the reasons why elderly people are more likely than young people to be at risk for age-related diseases such as metabolic disease, cerebral and cardiovascular diseases, and cancer.

Keywords Plasma, Proteome, Chemical modifications, Tandem mass spectrometry, Non-limiting modification identification

Gao Youhe (1964.06–), male, Ph.D., professor, doctoral supervisor, recipient of the National Science Fund for Distinguished Young Scholars, currently director of the Beijing Key Laboratory of Genetic Engineering Drugs and Biotechnology, and editor-in-chief of *Urine*. E-mail: gaoyouhe@bnu.edu.cn. This work is one of the research outcomes of the National Key Research and Development Program projects "Development and Application Demonstration of Key Technologies for Quality Control of Proteomics Data" and "Development of New Technologies for the Early Diagnosis of Alzheimer' s Disease" (project Nos. 2018YFC0910202 and 2016YFC1306300).

1 Introduction

Biological processes in living organisms can operate simultaneously and in coordination with one another, and they are inseparable from the synthesis, catalysis, and regulatory reactions in which proteins participate. Proteins are a class of structurally complex biological macromolecules. Differences in their higher-order structures determine their distinct biochemical activities, and, through mutual combinations among proteins, they form higher-level networks to execute specific functions[^1]. Chemical modification of proteins refers to covalent reactions involving amino acid residues or chain termini. Usually, a small number of changes in chemical structure do not affect the biological activity of a protein and are therefore termed modifications of nonessential parts of the protein. In the vast majority of cases, however, changes in molecular structure significantly alter the physicochemical properties of proteins, change their conformation, cause alterations in protein activity, and thereby change the functions they perform[^2]. Therefore, even if the protein abundance level does not change, small changes in its level of chemical modification can also cause marked changes in protein function. This means that chemical modification of proteins enriches, in another dimension, the different functions that proteins can exert. The effects of chemical modification on protein function are mainly manifested in the following three aspects: 1. even the occurrence of a single modification on a protein can affect its function; 2. for the same protein, the effects of the same or different modifications occurring on different amino acids on its function also differ; 3. the same protein may undergo multiple types of chemical modifications, thereby making the biological processes in which it participates more variable and complex[^3]. The chemical modifications associated with proteins mainly include the following types: 1. post-translational modification (Post-translational modification, PTM)[^4], which refers to chemical modifications carried out on proteins after translation. Precursors of post-translationally modified proteins often have no biological activity and usually must undergo post-translational modification processing by specific modifying enzymes before they can become mature proteins with functional activity and exert their specific biological functions[^5,^6]. 2. Chemical-derived modification of proteins (Chemical derived modification) refers to a class of modifications in which new groups are introduced into protein side chains or original groups are removed; it usually includes spontaneous non-enzymatic modifications, as well as modifications introduced by crosslinkers and artificial reagents. 3. Amino acid substitution (Amino acid substitution) refers to changes in protein properties and functions caused by the substitution of an original amino acid in a protein side chain by another type of amino acid. All of these are types of modifications that can significantly affect protein function.

Mass spectrometric analysis can not only enable large-scale collection and in-depth mining of omics data, but also accurately determine targeted modifications of specific proteins. With the continuous development of instrumental science, ultra-high-resolution and tandem mass spectrometry (LC–MS/MS) has

provided richer information and data for proteomics and studies of its modifications, and has also offered powerful assistance for accurately identifying chemical modification sites in protein chains. Usually, when performing proteomic data analysis, species-specific proteome database searching and comparison are required. High-resolution tandem mass spectrometry is the main way to obtain large amounts of protein modification information. When using search engines for database retrieval, it is usually necessary to manually set known types of protein chemical modifications. This type of search method is called a restricted search; when the types of protein modifications in a sample are unknown and one wishes to discover new modification types, it is difficult to achieve this[^7]. Therefore, comprehensive, unrestricted modification searching is of critical importance for understanding all chemical modification information contained in a sample proteome. The Open-pFind algorithm is an open sequence-database search algorithm that integrates information from the UniMod database and, by using an open-search approach, analyzes and processes the collected mass spectrometry data to obtain the overall chemical modification information of the proteome[^8-^10].

Plasma is an important component of the internal environment of the body and plays roles in transporting substances needed to sustain life activities and wastes produced in the body. Plasma contains abundant types and quantities of proteins, and its composition is readily affected by the body' s metabolic, physiological, and pathological processes. Metabolites and wastes in cells and tissues are transported and exchanged through the blood. In theory, the study of plasma proteomics can reflect the physiological state of the body at a specific stage. Another important research aspect of plasma proteomics is its level of chemical modifications. A comprehensive comparison of changes in the chemical modification levels of the plasma proteome will also provide richer information for studying physiological changes in the body[^6]. At present, there are more than 1,500 types of protein chemical modifications recorded in the UniMod[^11], PSI-MOD, and RESID databases, and the human plasma proteome contains numerous types of chemical modifications, such as

acetylation at the protein N-terminus, phosphorylation, methylation, glycosylation, and ubiquitination of side chains, as well as disulfide-bond modifications within and between protein chains. The plasma proteome can reflect subtle conditions associated with age and aging[^12]. Urine, another body fluid similarly rich in protein information, has been found through studies to contain a urinary proteome that also has the ability to reflect information on organismal aging[^13]; even common physiological processes such as hunger can be traced in the urinary proteome[^14]. Comparisons of the overall chemical modifications of proteins in plasma and urine show differences in modifications among different types of samples[^15]. Based on the importance of chemical modifications in the human plasma proteome, this study used high-resolution tandem mass spectrometry combined with an unrestricted modification search (Open-pFind) to compare the differences in the overall chemical modification levels of the plasma proteome between two groups of people in different age ranges.

2 Materials and Methods

2.1 Sample collection

Plasma samples from 20 individuals were collected from the clinical laboratory department of a Beijing hospital; all were discarded clinical laboratory samples. The samples were divided by age into two groups: a young group and an older group. Samples were randomly selected from the laboratory specimens, and there were no restrictions or requirements regarding the subjects' diet, medications, or other factors. Only age and sex information was extracted from the discarded samples. Detailed sample information is shown in Table 1.

Table 1. Statistical results for the obtained plasma sample information

Table 1. Statistical data of plasma samples

Sample type	Young-group plasma	Older-group urine
Number of samples (n)	10	10
Sex:		
Male	2	3
Female	8	7
Age:		
Mean	31.8 ± 4.19	59.3 ± 3.46
Range	26–37	56–66

2.2 Protein sample preparation and tryptic digestion

Plasma was obtained by centrifugation after whole-blood anticoagulation treatment. Plasma samples ($n = 20$) were diluted 40-fold with Milli-Q water, and then 100 μL was taken for subsequent experiments. Dithiothreitol (DTT) at 20 mmol/L was used to react with the samples at 37°C for 1 h to denature the disulfide bonds in the protein structure. Subsequently, 55 mmol/L iodoacetamide (IAA) was added, and the reaction was carried out in the dark for 30 min to alkylate the disulfide-bond binding sites. At −20°C, the supernatant was precipitated with three volumes of prechilled acetone for 2 h, and then centrifuged at 4°C and 12,000 × g for 30 min to obtain the protein precipitate. The precipitate was then resuspended in an appropriate amount of protein dissolution buffer (8 mol/L urea, 2 mol/L thiourea, 25 mmol/L DTT, and 50 mmol/L Tris). The concentration of the protein extract was measured using the Bradford assay. Using the filter-aided sample preparation (FASP) method, 100 μg of protein from each sample was digested with trypsin at a ratio of 50:1 (Trypsin Gold, Mass Spec Grade, Promega, Fitchburg, WI, USA). The digested peptides were dried using a vacuum centrifugal concentrator, and the dried peptides were sealed and stored at −80°C.

2.3 Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis and database searching

Before analysis, the dried peptide samples were dissolved in 0.1% formic acid solution, and the final concentration was controlled at 0.1

μg/μL, and each sample was analyzed using 1 μg of peptide mass. The Thermo EASY-nLC1200 chromatographic system was loaded with the precolumn and analytical column. Proteomic data were acquired using a Thermo Orbitrap Fusion Lumos mass spectrometry system (Thermo Fisher Scientific, Bremen, Germany). The liquid chromatography method was as follows: precolumn: 75 μm × 2 cm, nanoViper C18, 2 μm, 100 Å; analytical column: 50 μm × 15 cm, nanoViper C18, 2 μm, 100 Å; injection volume: 10 μL; flow rate: 250 nL/min; mobile phase: phase A: 100% mass-spectrometry-grade water (Fisher Scientific, Spain)/1‰ formic acid (Fisher Scientific), phase B: 80% acetonitrile (Fisher Scientific, USA)/20% water/1‰ formic acid; 120 min gradient elution: 0 min, 3% phase B; 0 min–3 min, 8% phase B; 3 min–93 min, 22% phase B; 93 min–113 min, 35% phase B; 113 min–120 min, 90% phase B. The mass spectrometry method was as follows: ion source: nanoESI; spray voltage: 2.0 kV; capillary temperature: 320 °C; S-lens RF Level: 30; resolution settings: MS1 (Orbitrap) 120,000 @m/z 200, MS2 30,000 (Orbitrap) @m/z 200; precursor-ion scan range: m/z 350–1350; fragment-ion scan range: start from m/z 110; MS1 AGC: 4e5; charge range: 2–7; ion injection time: 50 ms; MS2 AGC: 1e5; ion injection time: 50 ms; ion selection window: 2.0 m/z; fragmentation mode: HCD; energy NCE 32; Data-dependent MS/MS: Top 20; dynamic exclusion time: 15 s; internal calibration mass: 445.12003.

pFind Studio software (version 3.1.3, Institute of Computing Technology, Chinese Academy of Sciences) was used for label-free quantitative analysis of the LC-MS/MS data. The target search database was the Homo Sapiens database downloaded from UniProt (updated to October 2018). During the search, the instrument type was HCD-FTMS, enzyme specificity was trypsin, and up to two missed cleavage sites were allowed. "Open search" was selected. Filtering criteria: FDR at the peptide level was less than 1%, and the Q value at the protein level was less than 1%. The identification data were analyzed using a forward and reverse database search strategy simultaneously.

3 Results

3.1 Identification of the total protein using Bottom-Up proteomics technology

In label-free proteomic analysis, 20 (10/10) samples were analyzed by LC-MS/MS. After searching the data (.raw) with open-pFind software, the analysis results could be browsed and exported in pBuild. After organizing and statistically summarizing the sample results, through 120-min liquid-phase gradient analysis, 628–781 proteins (mean value 724) were identified in plasma samples without high-abundance-protein depletion, and the number of unique

Venn diagram showing 89 modifications unique to the older group, 1080 shared, and 74 unique to the younger group. The left circle is labeled "older group," and the right circle is labeled "younger group."

Figure 1: Venn diagram showing 89 modifications unique to the older group, 1080 shared, and 74 unique to the younger group. The left circle is labeled "older group," and the right circle is labeled "younger group."

peptide identifications was 7,526–11,464 (mean value 10,238).

3.2 Comparison of modifications after translation of the plasma proteome between the young group and the elderly group

In the plasma samples from the elderly group, a total of 1,169 types of modifications were identified (including low-abundance modification types, i.e., each type of modification was identified at least once in the samples), whereas in the plasma samples from the young group, a total of 1,154 types of modifications were identified (including low-abundance modification types). Among them, 88% of the modification types were shared, and the remaining 12% of differential modifications were related (Figure 1).

Figure 1 Statistics of the overall plasma proteome chemical modifications in the old and young groups

Among the 163 unique protein chemical modifications accumulated across the two groups, 158 had within-group reproducibility below 50%; only 5 met the criterion of reproducibility higher than 50%. These were, respectively, four modifications appearing in the older group: Arg->Phe[R], Arg->Trp[R], Xlink_DMP[K], and DTT_ST[S], and one modification appearing in the younger group: bisANS-sulfonates[T]. Overall, the unique chemical modification types were molecular modification types with small numbers of identifications and poor within-group reproducibility.

Chemical modifications of non-low abundance—that is, modifications identified 10 or more times in the samples—were analyzed statistically, and the identification coverage within each group was required to be greater than 50%. Among the 1080 chemical modifications, 120 met these criteria. These were subjected to unsupervised clustering analysis, which could roughly distinguish the younger-group and older-group samples; however, 40% of the younger-group samples and older-group samples were clustered into the same group and were not assigned to categories together with other within-group samples. Figure 2 shows the unsupervised clustering results for the specific samples. The criterion for screening differential chemical modifications was a between-group p value less than 0.05. The mean number of chemical modifications in each sample within each group was calculated by normalizing against the total number of identified chemical modification spectra. Four shared modifications with between-group fold changes greater than 1.5 were obtained: 2-succinyl[C], Lys->Thr[K], Phos-

pho[C], and Carbamyl[K]. Detailed information is shown in Table 2.

Table 2 Information on the molecular modifications of the plasma proteome between the older and younger groups

No.	ID*	Name	Modification type	*p* value	Older-group mean	Younger-group mean	Older group / younger group
1	957	2-succinyl[C]	Chemical derivative modification	0.001	0.00091	0.00046	1.98 ↑
2	594	Lys->Thr[K]	Amino-acid substitution	0.003	0.00107	0.00065	1.65 ↑
3	21	Phospho[C]	Post-translational modification	0.013	0.00087	0.000548	1.59 ↑
4	1175	Pro->Trp[P]	Amino-acid substitution	0.042	0.00091	0.00061	1.49 ↑
5	21	Phospho[S]	Post-translational modification	0.027	0.00105	0.00077	1.36 ↑
6	64	Succinyl[Any N-term]	Post-translational modification	0.032	0.00414	0.00317	1.31 ↑
7	23	Dehydrated[S]	Post-translational modification	0.022	0.0041	0.00446	1.09 ↓

Figure 2. Unsupervised cluster analysis of overall plasma proteome molecular modifications in the young and old groups.

Figure 2: Figure 2. Unsupervised cluster analysis of overall plasma proteome molecular modifications in the young and old groups.

No.	ID*	Name	Modification type	p value	Older-group mean	Younger-group mean	Older group / younger group
8	562	Glu->Asp[E]	Amino-acid sub-stitu-tion	0.029	0.0015	0.00184	1.22 ↓
9	5	Carbamyl[K]	Polymerization	0.049	0.01858	0.03081	1.67 ↓

* The ID is the accession number in the Unimod database; for details, see: http://www.unimod.org/

Figure 2 Unsupervised cluster analysis of overall plasma proteome molecular modifications in the young and old groups. The red mark (Y) indicates samples from the young group, and the light-blue mark (O) indicates samples from the older group. The abscissa shows the unsupervised clustering status and specific sample information, and the ordinate shows the names of the specific molecular modifications.

Figure 2 Unsupervised cluster analysis of overall plasma proteome molecular modifications in the young and old groups. The red mark (Y) is the sample of the young group, and the blue sign (O) is the sample of the old group. The abscissa is the unsupervised clustering and sample specific information, and the ordinate is the specific molecular modification name.

3.3 Random grouping test

The four differential modifications that met the differential screening criteria (p value less than 0.05 and fold change greater than 1.5) were each subjected independently to random grouping tests to verify the false-positive probability of each differential modification. The data from 10 young-group cases and 10 old-group cases, for a total of 20 datasets, were randomly divided into two groups, with 10 samples in each group. There were 92,378 different combinations $\left(\frac{1}{2}C_{20}^{10}\right)$. For the combinations in each case, statistical analysis was performed using the same differential screening criteria. After exhaustive calculation and statistics, the random combination results for the four differential modifications were obtained; see Table 3.

Table 3 Random grouping calculation results of four differential chemical modifications in plasma proteins

No.	Name	Total number of occurrences in random groupings	Frequency of occurrence	Confidence (%)
1	2-succinyl [C]	4857	0.0525	94.74
2	Lys->Thr [K]	1135	0.0123	98.77
3	Phospho [C]	2126	0.0230	97.70
4	Carbamyl [K]	4664	0.0505	94.95

Through random grouping tests, it was found that the randomness of the four differential modifications was approximately within the range of 5%, and the confidence levels were greater than 90%, indicating that the possibility that the differences in these four modifications between different age groups were generated at random is relatively low.

4 Discussion

From translational assembly to ultimate degradation, proteins experience many harmful environments throughout the entire process. They may be modified, thereby changing the structure and function of protein molecules; this process is also referred to as molecular aging[⁀18]. These reactions are mainly caused by non-enzymatic binding of reactive small molecules to functional groups on proteins. Modifications are classified as reversible or irreversible. Through an unrestricted modification search, several protein chemical modifications with statistically significant differences were found in plasma from two groups of people of different ages. These included succinylamide modification of hemisuccinic acid, substitution of lysine residues by threonine residues, phosphorylation of hemisuccinic acid, substitution of carbamyl residues by selenocysteine residues, phosphorylation of cysteine, N-terminal modification of succinylamide, dehydration of cysteine, substitution of valine residues by aspartic acid residues, and methylcarbamylation of lysine. Among them, succinylation of hemisuccinic acid, phosphorylation modification, and the replacement of lysine by threonine were significantly higher in the older group than in the younger group, whereas methylcarbamylation of lysine was higher in the younger group than in the older group.

Hemisuccinic acid is one of the least abundant amino acids in proteins. Residue statistics for protein composition indicate that the average frequency of

hemisuccinic acid in eukaryotes is approximately 2%; meanwhile, 70% of proteins derived from the reducing proteome are present in the intracellular environment[^19]. Studies have shown that the abundance of hemisuccinic acid in proteins is affected by protein function. The high activity of thiol groups makes them readily participate in many chemical modifications; conversely, it also determines the distribution and topological properties of hemisuccinic acid residues in protein structures. Thiol groups are also important groups for forming disulfide bonds and maintaining protein structure[^20].

The succinylation and phosphorylation of hemisuccinic acid found in this study both involve the participation of thiol groups and may therefore affect the formation of disulfide bonds. Through a Michael addition reaction, fumaric acid is added to the free thiol sites of certain Cys residues in proteins, thereby forming S-(2-succinic acid) hemisuccinic acid[^21]. This modification was first detected in plasma proteins (including albumin) and is formed through an irreversible reaction. This modification has been reported in association with diabetes, obesity, and fumarate hydratase–related diseases and in a model of Reed syndrome. At the same time, increased levels of succinated proteins have also been found in mouse 3T3-L1 adipocytes cultured in high-glucose medium (30 mM, whereas the physiological level is 5 mM), as well as in rat tissues treated with streptozotocin[^16,^22,^23]. Reports in the literature indicate that nutrient (sugar) excess leads to increases in ATP:ADP, NADH:NAD^+ and the mitochondrial membrane potential, while the increased NADH:NAD^+ inhibits oxidative phosphorylation, thereby leading to sustained accumulation of mitochondrial intermediates (including fumaric acid), which in turn causes increased protein succination. The accumulation of succinated proteins may even be caused by reduced fumarate hydratase activity. Fumarate hydratase catalyzes the conversion of fumaric acid to malic acid in the tricarboxylic acid cycle. Importantly, loss of function and mutations in the known fumarate hydratase gene make affected individuals susceptible to multiple skin diseases and uterine leiomyoma, as well as

hereditary leiomyomatosis and renal cell carcinoma (HLRCC)[19]. Although the exact role of related proteins such as carbamylated half-amides modified by succinic acid remains to be fully elucidated[24–26], it is associated with the cancer response related to fumarate hydratase[27–29]. Increased protein succinylation has also been described in the brainstem of *Ndufs4* knockout mice (a model of Leigh syndrome)[30], indicating that this type of protein chemical modification has a potential role in the pathogenesis of this mitochondrial disease. By comparing with annotations in the UniProt database, Park et al. found that among the protein succinylation sites detected, 16 succinylation sites were located in cofactor-binding regions or catalytic regions, and 74 succinylation sites were present around enzyme active sites[31]. Baynes et al. found that succinylation of hemicarbamic acid in human skin collagen increased with age[23,32,33].

Cysteine phosphorylation has recently been discovered in prokaryotic and eukaryotic systems and is considered to play a key role in signal transduction and

regulation of cellular responses. Because of the low chemical stability of thiophosphate esters in peptide molecules, there have been few studies of *in vivo* phosphorylation of Cys side chains[34,35]. Cysteine phosphorylation is an important function of cysteine-dependent protein phosphatases (CDPs). This enzyme belongs to the protein tyrosine phosphatase (PTP) subfamily and catalyzes hydrolysis of the phosphoester bond through formation of a phospho-cysteine intermediate. Studies have shown that this reversible post-translational regulation (PTM) is critically important in regulating the expression of virulence determinants and bacterial resistance to antibiotics. In addition, studies have also indicated that cysteine phosphorylation, previously regarded as a rare form, may be more widespread in nature and may play an important role in biological regulation in various organisms.

Replacement of lysine by glutamine alters the acidity and alkalinity of proteins, thereby affecting protein activity and function. It was discovered long ago that substitution of lysine residues in hemoglobin by glutamine residues reduces its oxygen affinity and tension[36]. Sun WY et al. found that mutation of exon 14 of nucleophosmin 20040, in which glutamine replaced Lys at amino acid 556, reduced the procoagulant activity of prothrombin by 50%[37]. The decrease in oxygen affinity also reflects a slowing of the organism' s metabolism, and this, together with changes in a series of physiological states such as coagulation activity, may be an early manifestation of aging. In addition, it has been found that human and mouse embryonic stem cells require specific amino acids for proliferation. mES cells require glutamine metabolism to complete apparent epigenetic protein modification. Glutamine is converted into glutamate and acetyl-coenzyme A, and glutamate metabolism specifically regulates trimethylation of lysine (Lys) residues in histone H3 (H3K4me3)[38]. In addition, we also found that the older group showed a decrease in lysine carbamylation modification compared with the younger group. Carbamylation is an irreversible, nonenzymatic modification process in which the decomposition products of urea react with the N-terminus of proteins or with the side chains of lysine residues; it has previously been reported to be associated with protein aging[39]. Lysine carbamylation can promote the coordination of metal ions with specific enzymatic activities. Studies have pointed out that the amount of carbamylation in the plasma of patients with elevated urea levels (for example, patients with kidney disease) is significantly increased[40].

Aging is an inevitable, spontaneous process in organisms over time. It is a complex natural phenomenon, manifested as degenerative changes in structure and decline in function, with decreased adaptability and resistance. Aging is one of the greatest known risk factors for most human diseases: among the approximately 150,000 deaths worldwide each day, about two thirds are due to age-related causes. The causes of aging are currently unclear. The prevailing theory at present is the damage theory, in which DNA damage is regarded as the common basis of cancer and aging; some believe that the intrinsic causes of DNA damage are the most important driving force of aging. The waste-accumulation theory holds that the accumulation of waste in cells may interfere

with metabolism. For example, one type of waste called lipofuscin is formed through complex reactions in which fats bind to proteins in cells. These wastes accumulate in cells in the form of small granules, and their size increases with age. At the same time, the accumulation of biological macromolecules whose structures have been damaged or even inactivated may lead to the progressive failure of organisms and systems; this is also considered a concept of aging.

In addition to being associated with physiological aging, molecular aging of proteins is also accelerated in many chronic diseases, such as diabetes and chronic renal insufficiency, and participates in the occurrence and development of disease-specific complications.

Through the above study, we found several types of protein chemical modifications and amino-acid substitutions with significant differences in the plasma proteomes of populations at different ages. These chemical modifications can alter the structure and properties of proteins.

thereby affecting the functions performed by proteins. It is speculated that the gradual accumulation of certain harmful and irreversible chemical modifications in the plasma proteins of the elderly may reflect the process of organismal aging, and may also be one of the reasons why the incidence of aging-related diseases in older individuals—such as metabolic diseases, cardiovascular and cerebrovascular diseases, and tumors—is higher than in younger people.

References:

[1] Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature, 2016, 537: 347-355

[2] Sakamoto S, Hamachi I. Recent progress in chemical modification of proteins. Anal Sci, 2019, 35: 5-27

[3] Wang C, Zou P, Yang C, *et al.* Dynamic modifications of biomacromolecules: Mechanism and chemical interventions. Sci China Life Sci, 2019, 62: 1459-1471

[4] Minguez P, Parca L, Diella F, *et al.* Deciphering a global network of functionally associated post-translational modifications. Mol Syst Biol, 2012, 8:

[5] Pieroni L, Iavarone F, Olianas A, *et al.* Enrichments of post-translational modifications in proteomic studies. J Sep Sci, 2020, 43: 313-336

[6] Goto T, Kudo Y, Lee SH, *et al.* Efficient strategy for screening chemical modifications on human serum albumin: Use of lc/ms/ms and differential analysis software. Bunseki Kagaku, 2015, 64: 653-659

[7] An Z, Zhai L, Ying W, *et al.* Ptminer: Localization and quality control of protein modifications detected in an open search and its application to comprehensive post-translational modification characterization in human proteome. Mol Cell Proteomics, 2019, 18: 391-405

[8] Li D, Fu Y, Sun R, *et al.* Pfind: A novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry.

Bioinformatics, 2005, 21: 3049-3050

[9] Chi H, He K, Yang B, *et al.* Pfind-alioth: A novel unrestricted database search algorithm to improve the interpretation of high-resolution ms/ms data. J Proteomics, 2015, 125: 89-97

[10] Wang LH, Li DQ, Fu Y, *et al.* Pfind 2.0: A software package for peptide and protein identification via tandem mass spectrometry. Rapid Commun Mass Spectrom, 2007, 21: 2985-2991

[11] Creasy DM, Cottrell JS. Unimod: Protein modifications for mass spectrometry. Proteomics, 2004, 4: 1534-1536

[12] Lehallier B, Gate D, Schaum N, *et al.* Undulating changes in human plasma proteome profiles across the lifespan. Nat Med, 2019, 25: 1843-+

[13] Li X, Gao Y. Potential urinary aging markers of 20-month-old rats. PeerJ, 2016, 4: e2058

[14] Yuan Y, Z. F, N. Y, *et al.* Urinary protein changes in a rat starvation model. MOJ Proteomics & Bioinformatics, 2017, 6: 00202

[15] L. Y, D. M, P. X, *et al.* Why are there proteins in the urine of healthy people? Sciencepaper Online, 2019, 201911-28:

[16] Blatnik M, Thorpe SR, Baynes JW. Succination of proteins by fumarate: Mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes. Ann N Y Acad Sci, 2008, 1126: 272-275

[17] Van Driessche G, Vandenberghe I, Jacquemotte F, *et al.* Mass spectrometric identification of in vivo carbamylation of the amino terminus of ectothiorhodospira mobilis high-potential iron-sulfur protein, isozyme 1. J Mass Spectrom, 2002, 37: 858-866

[18] Gillery P, Jaisson S. Usefulness of non-enzymatic post-translational modification derived products (ptmdps) as biomarkers of chronic diseases. J Proteomics, 2013, 92: 228-238

[19] Pe' er I, Felder CE, Man O, *et al.* Proteomic signatures: Amino acid and oligopeptide compositions differentiate among phyla. Proteins, 2004, 54: 20-40

[20] Gunnoo SB, Madder A. Chemical protein modification through cysteine. Chembiochem, 2016, 17: 529-553

[21] Alderson NL, Wang YP, Blatnik M, *et al.* S-(2-succinyl)cysteine: A novel chemical modification of tissue proteins by a krebs cycle intermediate. Arch Biochem Biophys, 2006, 450: 1-8

[22] Thomas SA, Storey KB, Baynes JW, *et al.* Tissue distribution of S-(2-succino)cysteine (2sc), a biomarker of mitochondrial stress in obesity and diabetes. Obesity, 2012, 20: 263-269

[23] Frizzell N, Rajesh M, Jepson MJ, *et al.* Succination of thiol groups in adipose tissue proteins in diabetes: Succination inhibits polymerization and secretion of adiponectin. J Biol Chem, 2009, 284: 25772-25781

[24] Zheng L, Cardaci S, Jerby L, *et al.* Fumarate induces redox-dependent senescence by modifying glutathione metabolism. Nat Commun, 2015, 6: 6001

[25] Yang M, Soga T, Pollard PJ, *et al.* The emerging role of fumarate as an oncometabolite. Front Oncol, 2012, 2: 85

[26] Frezza C, Pollard PJ, Gottlieb E. Inborn and acquired metabolic defects in cancer. J Mol Med, 2011, 89: 213-220

[27] Ooi A, Wong JC, Petillo D, *et al.* An antioxidant response phenotype shared between hereditary and sporadic type 2 papillary renal cell carcinoma. Cancer Cell, 2011, 20: 511-523

[28] Bardella C, El-Bahrawy M, Frizzell N, *et al.* Aberrant succination of proteins in fumarate hydratase-deficient mice and hlrcc patients is a robust biomarker of mutation status. J Pathol, 2011, 225: 4-11

[29] Adam J, Hatipoglu E, O' Flaherty L, *et al.* Renal cyst formation in fh1-deficient mice is independent of the hif/phd pathway: Roles for fumarate in keap1 succination and nrf2 signaling. Cancer Cell, 2011, 20: 524-537

[30] Piroli GG, Manuel AM, Clapper AC, *et al.* Succination is increased on select proteins in the brainstem of the nadh dehydrogenase (ubiquinone) fe-s protein 4 (ndufs4) knockout mouse, a model of leigh syndrome. Mol Cell Proteomics, 2016, 15: 445-461

[31] Piroli GG, Manuel AM, Walla MD, *et al.* Identification of protein succination as a novel modification of tubulin. Biochem J, 2014, 462: 231-245

[32] Biatnik M, Thorpe SR, Baynes JW. Succination of proteins by fumarate — mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes. Ann Ny Acad Sci, 2008, 1126: 272-275

[33] Frizzell N, Lima M, Baynes JW. Succination of proteins in diabetes. Free Radic Res, 2011, 45: 101-109

[34] Zu XL, Besant PG, Imhof A, *et al.* Mass spectrometric analysis of protein histidine phosphorylation. Amino Acids, 2007, 32: 347-357

[35] Attwood PV, Piggott MJ, Zu XL, *et al.* Focus on phosphohistidine. Amino Acids, 2007, 32: 145-156

[36] Shih DT, Jones RT, Shih MF, *et al.* Hemoglobin chico [beta 66(e10)lys--thr]: A new variant with decreased oxygen affinity. Hemoglobin, 1987, 11: 453-464

[37] Sun WY, Smirnow D, Jenkins ML, *et al.* Prothrombin scranton: Substitution of an amino acid residue involved in the binding of na+ (lys-556 to thr) leads to dysprothrombinemia. Thromb Haemost,

2001, 85: 651-654

[38] Van Winkle LJ, Ryznar R. One-carbon metabolism regulates embryonic stem cell fate through epigenetic DNA and histone modifications: Implications for transgenerational metabolic disorders in adults. Front Cell Dev Biol, 2019, 7: 300

[39] Badar A, Arif Z, Alam K. Role of carbamylated biomolecules in human diseases. IUBMB Life, 2018, 70: 267-275

[40] Dalle-Donne I, Aldini G, Carini M, *et al.* Protein carbonylation, cellular dysfunction, and disease progression. J Cell Mol Med, 2006, 10: 389-406

Note: Figure translations are in progress. See original paper for figures.