


RESEARCH ARTICLE

Open Access



Non-human peptides revealed in blood reflect the composition of intestinal microbiota

Georgij P. Arapidi^{1,2*} , Anatoly S. Urban^{1,2}, Maria S. Osetrova¹, Victoria O. Shender^{1,2}, Ivan O. Butenko^{2,4}, Olga N. Bukato², Alexandr A. Kuznetsov², Tatjana M. Saveleva¹, Grigorii A. Nos¹, Olga M. Ivanova^{1,3}, Leonid V. Lopukhov⁵, Alexander V. Laikov⁵, Nina I. Sharova⁶, Margarita F. Nikonova⁶, Alexander N. Mitin⁶, Alexander I. Martinov⁶, Tatiana V. Grigorieva⁵, Elena N. Ilina^{2,4}, Vadim T. Ivanov¹ and Vadim M. Govorun^{1,2,4}

Abstract

Background The previously underestimated effects of commensal gut microbiota on the human body are increasingly being investigated using omics. The discovery of active molecules of interaction between the microbiota and the host may be an important step towards elucidating the mechanisms of symbiosis.

Results Here, we show that in the bloodstream of healthy people, there are over 900 peptides that are fragments of proteins from microorganisms which naturally inhabit human biotopes, including the intestinal microbiota. Absolute quantitation by multiple reaction monitoring has confirmed the presence of bacterial peptides in the blood plasma and serum in the range of approximately 0.1 nM to 1 μM. The abundance of microbiota peptides reaches its maximum about 5 h after a meal. Most of the peptides correlate with the bacterial composition of the small intestine and are likely obtained by hydrolysis of membrane proteins with trypsin, chymotrypsin and pepsin – the main proteases of the gastrointestinal tract. The peptides have physicochemical properties that likely allow them to selectively pass the intestinal mucosal barrier and resist fibrinolysis.

Conclusions The proposed approach to the identification of microbiota peptides in the blood, after additional validation, may be useful for determining the microbiota composition of hard-to-reach intestinal areas and monitoring the permeability of the intestinal mucosal barrier.

Keywords Blood plasma and serum, Healthy donors, Immunomodulatory peptide properties, Microbiota of the small intestine, Selective permeability of mucus barriers

*Correspondence:

Georgij P. Arapidi
arapidi@gmail.com

¹ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117997, Russian Federation

² Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Malaya Pirogovskaya 1a, Moscow 119435, Russian Federation

³ Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Malaya Pirogovskaya 1a, Moscow 119435, Russian Federation

⁴ Research Institute for Systems Biology and Medicine, Nauchny Proezd 18, Moscow 117246, Russian Federation

⁵ Kazan Federal University, Kremlyovskaya Str. 18, Kazan 420008, Russian Federation

⁶ National Research Center—Institute of Immunology Federal Medical-Biological Agency of Russia, Kashirskoe Highway 24, Moscow 115522, Russian Federation



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

The human microflora, or microbiota, is a complex of microorganisms found in tissues and biological fluids of the body [1]. The National Institutes of Health (NIH) Human Microbiome Project (HMP) [2, 3] advises to consider different types of microbiota: gastrointestinal, urogenital, oral, skin, airways, eye, ear, wound, nose, lymph nodes, blood, bone, heart, and liver. Currently, it is known that the number of intestinal microbiota organisms is comparable to the number of cells in the human body [4]. The diversity of metabolic functions of commensal bacteria is enormous because the total number of genes of these symbiotic microorganisms exceeds the number of human genes by almost 150 times [5]. The relationship of the intestinal microbiota in various diseases is shown, particularly for conditions related to the human digestive system, such as obesity [6, 7], inflammatory bowel disease [8], and colon cancer [9]. However, the effects of the microbiota are not restricted to this “niche”. The connection between the intestinal microflora and diabetes mellitus [10], metabolic syndrome [11], symptomatic atherosclerosis [12], liver cirrhosis [13], Alzheimer’s and Parkinson’s diseases [14] has been shown. In some diseases related to the microbiota, fecal microbiota transplantation from a healthy donor is used in therapeutic practice to correct the microbiota composition of the patient [15]. Despite the achievements aimed at a better understanding of the intra-intestinal crosstalk between the microbiota and the immune system of the intestine [16], the extraintestinal mechanisms affecting the human body are poorly understood [17]. We consider that an important step towards elucidating the mechanisms of symbiosis between the human body and its microbiota is the discovery of active molecules—agents of interaction between the microbiota and the human body. To the best of our knowledge, currently, there are only studies on single interaction agents [18, 19]. For example, the group of Sergueï Fetissov studies signaling via the bacterial homolog of a human neuropeptide, α -melanocyte-stimulating hormone, as a regulator of the gut–brain axis [19]. Gabanyi et al. found that bacterial peptidoglycan-derived muropeptides reach the brain, pass through the blood–brain barrier, and influence the activity of a subset of brain neurons in mice [20]. Studies on neuroregulatory capabilities of the intestinal microbiota suggest that peptides produced by the intestinal microbiota are able to penetrate not only the intestine but also the blood–brain barrier. The intestinal mucosal barrier refers to the selectivity of the intestinal mucosa that prevents the unwanted intestinal contents from penetrating the

intestinal epithelium, while the barrier maintains nutrient permeability [21]. The permeability of the intestinal epithelial layer in normal and pathological conditions was studied by Spadoni et al., who demonstrated an unconstrained penetration of a 4-kDa fluorescent isothiocyanate dextran through the intestinal epithelium of mice [22]. Larger molecules (70 kDa) can penetrate the epithelial layer only if the intercellular contacts of the epithelium are disturbed due to intestinal infection.

Blood is a very important connective tissue, and it can theoretically contain products of all processes occurring in different parts of the body [23]. The blood plasma proteome was studied quite well [24–28]. Naturally, the next step was to investigate the blood peptidome [23, 29–31]. When examining the blood plasma and serum, we observe peptide fragments (up to 5 kDa) of proteins from solid tissues, immunoglobulins, long-range and local ligands, and other proteins [32–35]. According to N.L. Anderson & N.G. Anderson, “alien” proteins appear in the blood only upon infection or in the presence of parasites [36]. Can bacterial peptides and proteins be found in the bloodstream of healthy people? Christmann et al. detected antibodies to microbiota peptides in the serum of healthy adults and children [37]. They analyzed the serum by protein microarray and identified an IgG response to a panel of microbiota antigens. The authors attributed a similar relationship between adaptive immunity and the microbiota to “autoantibodies”. After isolation of CD137+ regulatory T cells from the donor blood, the group of Alexander Schefold was able to identify clones specific to some symbiotic bacteria [38]. Many studies show that proteins are degraded during digestion and enter the bloodstream mainly in the form of amino acids, as well as di- and tripeptides [39]. However, there are reports about detection of longer peptides in the bloodstream after ingestion of dairy and plant products [40–42]. A study on bioactive soybean peptides demonstrated the bloodstream penetration of a 43-amino acid peptide, lunasin, which has anti-inflammatory properties [42]. Thus, we can assume that the human intestinal microbiota, with its significant metabolic capabilities, can produce a pool of various peptides which enter the bloodstream and can perform various biological functions—also acting as the agents of interaction between the microbiota and the body.

In this work, we show, for the first time, that a significant number of peptides of non-human origin is present in the plasma and serum of healthy people. These peptides are fragments of proteins from microorganisms which naturally inhabit human biotopes—especially the intestinal microbiota.

Results

Mass spectrometry analysis of the blood plasma and serum samples from healthy donors reveals a high representation of peptides that are not fragments of known human proteins

Previously, we studied in detail blood plasma and serum peptides of ten healthy males and ten healthy females using an integrated approach developed by us based on the use of several types of sorbents (cation exchange Toyopearl CM-650M, CM Bio-Gel A, SP Sephadex C-25 and anion exchange QAE Sephadex A-25; Additional file 1: Sheet 1) [34]. Our approach allowed us to identify 15,530 unique peptides that belong to 2127 protein groups of human origin (Additional file 1: Sheets 2–3). However, many uninterpreted spectra remain. A single analysis using TripleTOF 5600+ produces around 40,000 tandem mass spectra within a 1.5-h-long gradient (Additional file 1: Sheet 1). Our results showed that we were able to identify only around 16.67% of all measured spectra (Additional file 1: Sheet 1). The mass spectra quality analysis based on QualScore [43] showed that, among the spectra of good quality and with a high chance of identification (on average 8.11% of the total number of mass spectra), more than half of them were not interpreted as fragments of known human proteins (Additional file 1: Sheet 1). This finding raises a question about the origin of these non-human peptides.

De novo identification of blood plasma and serum peptides reveals a significant presence of protein fragments from the human microbiota

Blood is a complex tissue, and it contains a large number of metabolic products produced by various organs [23]. To identify potential sources of “alien” (non-human) plasma peptides, we performed a de novo analysis of mass spectrometry data. De novo analysis is a less efficient method of identification than search against databases [44], since the accuracy and resolution of modern

mass spectrometers such as QTOF or Orbitrap remains not high enough. However, we used de novo analysis to identify organisms that included the most abundant components of the complex peptide mixture. That procedure allowed us to develop a hypothesis and then perform a standard search against a database of proteins of identified organisms.

The complete list of identified peptides, according to the results of the de novo analysis, as well as the list of organisms that contain these sequences in their genomes, are given in Additional file 2. Figure 1A shows the variety of sources of peptides in the blood plasma and serum; 62.2% of the identified spectra were assigned to animal proteins (Additional file 2: Sheet 4), 57.7% were interpreted precisely as fragments of mammalian proteins (Additional file 2: Sheet 6), including human, and 22.1% of the interpreted spectra were homologous for different kingdoms and superkingdoms (“any creature”); for 8.1% of the identified sequences, no taxonomic information was found at the superkingdom or kingdom level (Additional file 2: Sheet 4). A significant number of identified peptides (7.0%) belongs to bacterial components in the plasma and serum peptidome of healthy donors (Fig. 1A). At the same time, 96.6% of the superkingdom *Bacteria* identifications belong to various phyla of human microbiota organisms (Additional file 2: Sheet 5). *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, according to the de novo analysis, are the most represented phyla of superkingdom *Bacteria* and account for 6.1% of the total number of identified spectra. Moreover, genus *Helicobacter* (class *Epsilonproteobacteria*, phylum *Proteobacteria*) accounts for 3.5% of the total number of identifications (Additional file 2: Sheet 9), and species *Peptostreptococcus bacterium VA2* (phylum *Firmicutes*) accounts for 0.8% of the total number of identifications (Additional file 2: Sheet 10). This microorganism, according to the NIH HMP [3], is typical of the oral microbiota. Kingdoms and superkingdoms such as *Viridiplantae*, *Fungi*,

(See figure on next page.)

Fig. 1 Identification and validation of peptide fragments of microbiota proteins in human blood plasma (17 samples) and serum (55 samples). See also Additional file 2, Additional file 3, Additional file 4, and Additional file 5. **A** Distribution of the identified spectra by different taxonomic groups based on the results of the de novo analysis of plasma and serum samples from healthy donors. **B** Scheme of microbiota peptides identification in the mass spectra of human blood samples. Joint database (DB) allows to validate the identifications against the database of annotated genomes of microorganisms (defined by the NIH Human Metagenome Project), using identifications according to the UniProt Knowledgebase (UniProtKB). For validation, we used (1) machine learning algorithms embedded in Scaffold 4 software; (2) filtering out sequences of bacterial peptides homologous to known human proteins, taking into account possible single amino acid polymorphism (SAP); (3) comparing identified spectra with the spectra of synthetic peptides. **C** Comparison between the spectrum of the peptide identified in plasma/serum (bottom) and synthetic peptide WSDGSYLEDDQDMWR (top). b- and y-fragment peaks are labeled. Common peaks are shown in a slightly darker color. The estimates of Mascot Ion Score identification reliability and correlation between the spectra are given. Similar pairs of spectra of 30 synthesized peptides are shown in Additional file 4. **D** Comparison of our results of quantitative analysis of the bacterial peptide content of the blood of healthy persons, according to the MRM (multiple reaction monitoring) data, with the data on the permissible content of hormones in the blood courtesy of MedlinePlus from the National Library of Medicine

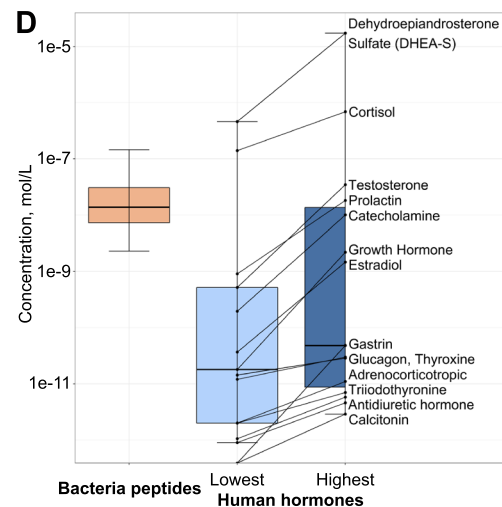
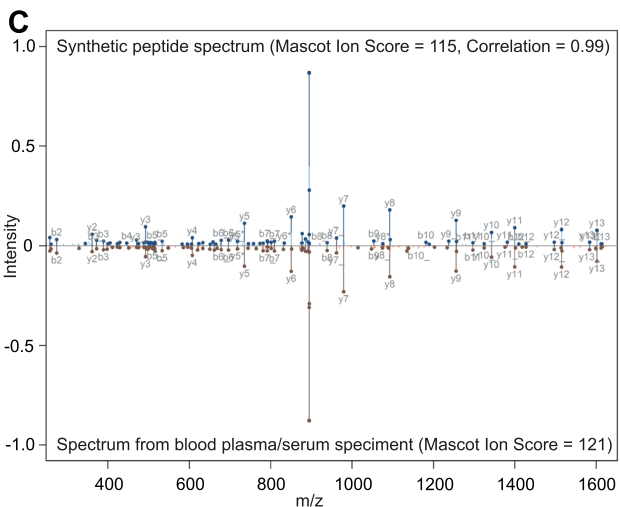
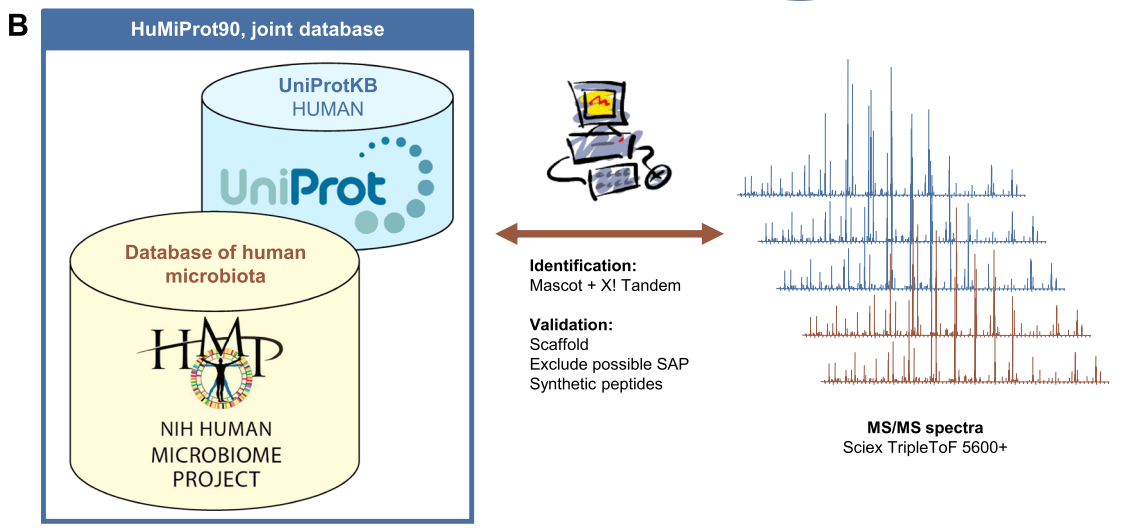
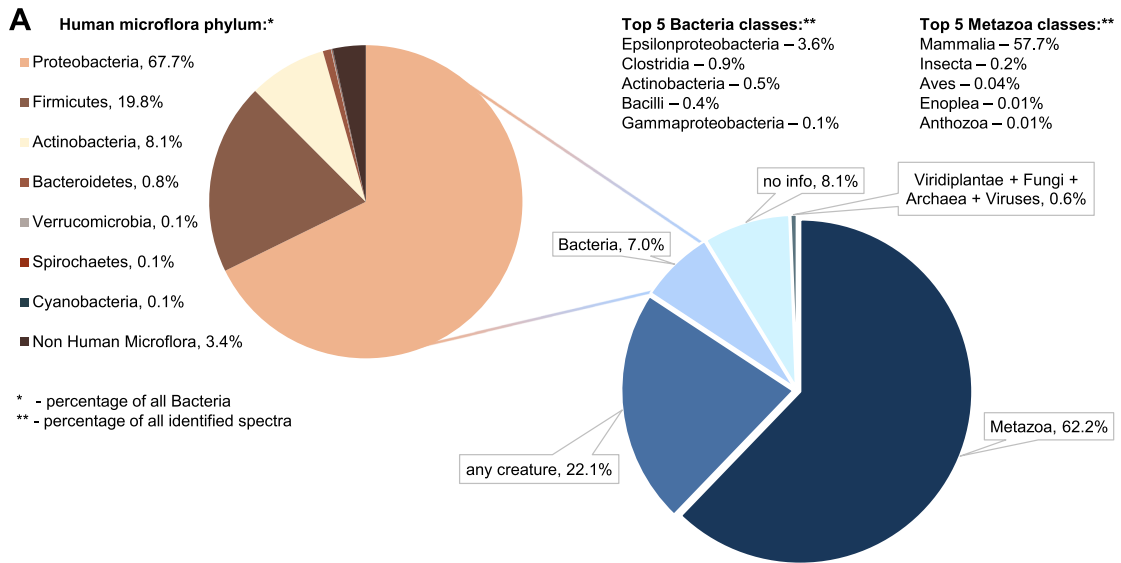


Fig. 1 (See legend on previous page.)

Archaea, and *Viruses*, as well as *Metazoa* classes such as *Insecta*, according to the result of the de novo analysis, contributed less than 1% to the total number of identified peptides, which can be considered insignificant (Fig. 1A). Our analysis did not reveal mass spectra attributed to the species of phylum *Firmicutes*, which are characteristic of soil, dust, or dirt microflora (*Bacillus subtilis*, *Bacillus mycoides*, *Bacillus mesentericus*, *Bacillus megaterium*, *Clostridium perfringens*, *Clostridium oedomaticus*, *Clostridium histolyticus*, *Clostridium botulinum*, *Clostridium chauvoeij*). This suggests that the detected peptides appeared in the samples not in the process of sample preparation or due to violations of the conditions for obtaining or storing the biomaterial.

LC-MS/MS analysis identified a significant number of peptides related to the human microbiota in the blood plasma and serum samples from healthy donors

The results of the de novo analysis indicate that the blood contains a considerable number of peptides that are fragments of proteins from the human microbiota. We tried to identify this pool of exogenous peptides according to the classical scheme (Fig. 1B). We combined the UniProt Knowledgebase (taxon human) and the database of annotated genomes of microorganisms identified by the NIH Human Metagenome Project (Additional file 3: Sheet 1). The database of annotated genomes of microorganisms was compressed by 90% homology using CD-HIT [45]. We named the resulting database HuMiProt90 (human and microbiota proteins). We identified a total of 12,713 unique peptides related to human proteins and 912 unique peptides related to human microbiota proteins (Table 1). The number of microbiota peptides in the total pool of identified peptides was about 7%. All identified amino acid sequences are listed in Additional file 3: Sheets 2–3.

The peptides mentioned in Table 1 and shown in Additional file 3 were validated both in silico and by comparison with the spectra of synthetic peptides (Fig. 1B). The in silico approach included machine learning used

by Scaffold 4 software to validate identifications based on the target-decoy approach (FDR 0.56% by PSM; false discovery rate by peptide to spectrum match), as well as filtering homologous sequences among known human proteins, taking into account possible single amino acid polymorphisms. The key stage of validation was the production of 30 synthetic peptides identified as fragments of proteins from the human microbiota. Figure 1C demonstrates an example of spectra comparison for a sequence identified in the plasma/serum and a synthetic peptide. Similar pairs of spectra for all 30 peptides are shown in Additional file 4. The correlation between the mass spectra of blood plasma/serum samples and the spectra of synthetic peptides is not less than 0.7 (for 23 peptides, the correlation is more than 0.8). The successful validation of the identified peptides, especially when compared to synthetic samples, suggests that these microbiota peptides are actually present in the blood of healthy donors.

Absolute quantitation confirmed the presence of bacterial peptides in the blood plasma and serum at physiologically significant concentrations

We quantified the abundance of the identified peptides in the serum and plasma by multiple reaction monitoring (MRM) using 30 synthetic bacterial peptides as external standards (Additional file 5: Sheet 1). Of these 30 peptides, 15 were quantified reliably. Quantitative analysis by MRM confirmed the identification of bacterial peptides in the blood and showed that these peptides were present at concentrations ranging from 1.01 to 4.28×10^{-4} $\mu\text{g/ml}$ (Fig. 1D, Additional file 5: Sheet 2). Using the medical blood test information courtesy of MedlinePlus from the National Library of Medicine (Additional file 5: Sheet 3), we compared the amounts of bacterial peptides in the blood (measured by mass spectrometry) with the publicly available data on acceptable blood hormone concentrations (Fig. 1D). It turned out that the identified bacterial peptides were present in the blood at concentrations comparable to or greater than the physiologically

Table 1 Total number of peptides (human or human microbiota) and percentage of those peptides that belong to the microbiota. See also Additional file 1 and Additional file 3

| Sample group | Number of peptides identified as fragments of human proteins | Number of peptides identified as fragments of microbiota proteins | Microbiota/total |
|---------------------------------|--|---|------------------|
| Plasma, Toyopearl, 17 samples | 4109 | 305 | 6.91% |
| Serum, Toyopearl, 16 samples | 6046 | 334 | 5.24% |
| Serum, Bio-Gel, 15 samples | 4941 | 123 | 2.43% |
| Serum, SP-Sephadex, 9 samples | 1675 | 95 | 5.37% |
| Serum, QAE-Sephadex, 15 samples | 2916 | 186 | 6.00% |
| Total number, 72 samples | 12,713 | 912 | 6.69% |

significant concentrations of human hormones in normal conditions.

Most of the identified bacterial peptides originate from the digestive system microflora as a result of hydrolysis of membrane-associated proteins by the main proteases of the gastrointestinal tract

We estimated the abundance of various phyla of microorganisms according to the blood peptidome data, based on the number of PSMs (peptide to spectrum matches) for the identified peptides (Fig. 2A, Additional file 3: Sheets 5–6), as well as the annotation of microorganisms by localization “niches” (Fig. 2B, Additional file 3: Sheets 12–13). *Proteobacteria*, *Firmicutes*, and *Actinobacteria* were the most abundant bacterial phyla (Fig. 2A). Among the identified bacterial peptides, 40.62% are fragments of proteins from the intestinal microbiota and 16.61% are fragments of proteins from oral microorganisms (Fig. 2B).

We performed an enrichment analysis for the precursor proteins of the identified peptides against orthologous groups from the eggNOG database (Table 2). Six out of nine significantly represented gene ontologies correspond to membrane or extracellular components.

Next we analyzed proteolysis sites in the peptides identified in the serum and plasma. For 36 proteolysis patterns (mostly corresponding to different proteases) described in R-package cleaver (part of Bioconductor), we compared the number of peptides of bacterial and human origin with less than two hydrolysis sites within an amino acid sequence (Additional file 6). Fisher’s exact test was used for comparison. The enzymes that are actively involved in digestion in the small intestine [47] have high odds ratios: chymotrypsin (3.00, 95% conf. interval 1.51–5.54, p -value 1.15×10^{-3}), pepsin (pH 1.3) (1.80, 95% conf. interval 1.11–3.11, p -value 4.64×10^{-2}), and trypsin (1.47, 95% conf. interval 1.24–1.73, p -value 1.11×10^{-5}). We also analyzed the proteolysis patterns of several bacterial proteases (Asp-N endopeptidase, LysC,

Arg-C proteinase or clostripain, staphylococcal peptidase I and thermolysin) and patterns of several eucaryotic proteases (caspases from caspase1 to caspase10, enterokinase, factor Xa, glutamyl endopeptidase, granzyme B, Lys-N, neutrophil elastase, proline endopeptidase, proteinase K, and thrombin). We did not get statistically significant results according to Fisher’s exact test for any of those proteases. Consequently, the bacterial peptides which circulate in the blood are statistically proven to differ significantly from the surrounding human peptides in that the bacterial peptides are likely to have been hydrolyzed in the digestive system.

The identified bacterial peptides have physicochemical properties likely allowing them more selectively pass the intestinal mucosal barrier and resist fibrinolysis compared to human peptides in the blood

We also analyzed a number of physicochemical properties of the identified peptides—of bacterial and human origin (Additional file 7). The comparison of the obtained distributions using the Wilcoxon rank sum test with continuity correction showed that the bacterial peptides were much more hydrophobic than the human peptides (Fig. 2C), consisted of amino acids with less branched side groups, as well as aliphatic (Fig. 2D) and non-polar amino acids. The analysis of the amino acid composition of the identified peptides demonstrated the prevalence of Ala, Cys, Ile, and Leu in bacterial peptides, while human peptides contained more His, Arg, Lys, Ser, and Gln (Fig. 2E). Lower representation of peptides with positively charged amino acids (Arg and Lys) could be due to the higher binding of the positively charged molecules to the polyanionic mucin polymers and the lower the diffusion of such molecules through the mucin [48, 49]. The identification of peptide fragments from human microbiota proteins with similar physicochemical properties in both plasma and serum (Fig. 2E) suggests that this pool of bloodstream peptides probably has properties

(See figure on next page.)

Fig. 2 Various properties of the identified bacterial peptides in the blood (72 samples). See also Additional file 3, Additional file 6, Additional file 7, and Additional file 8. **A** Evaluation of the abundance of various phyla of microorganisms, according to the analysis of the plasma (17 samples) and serum (55 samples) peptidome from healthy persons against the human microbiota protein database HuMiProt90. **B** Annotation of the microbiota isolation body sites within the human body from where the peptide fragments may have originated. The information about the isolation body sites was taken from the NIH Human Metagenome Project (HMP) database. The “Average” column shows average values for all 72 plasma/serum samples. **C** Analysis of the hydrophobicity of the peptides identified in the blood, of bacterial and human origin. **D** Analysis of the aliphaticity of the peptides identified in the blood, of bacterial and human origin. **E** Comparison of the amino acid compositions of the identified peptides of bacterial and human origin in plasma (17 samples) and serum (55 samples). Asterisks indicate amino acids for which there is a significant difference according to the Wilcoxon rank sum test with continuity correction between the peptides of different origin. **F** Left side: regional gastrointestinal transit time [46]. Right side: number of bacterial peptides identified in the serum before and after a meal (4 individuals, 19 samples) and assessment of the abundance of various phyla of microorganisms, according to the mass spectrometry analysis of peptides at these time points

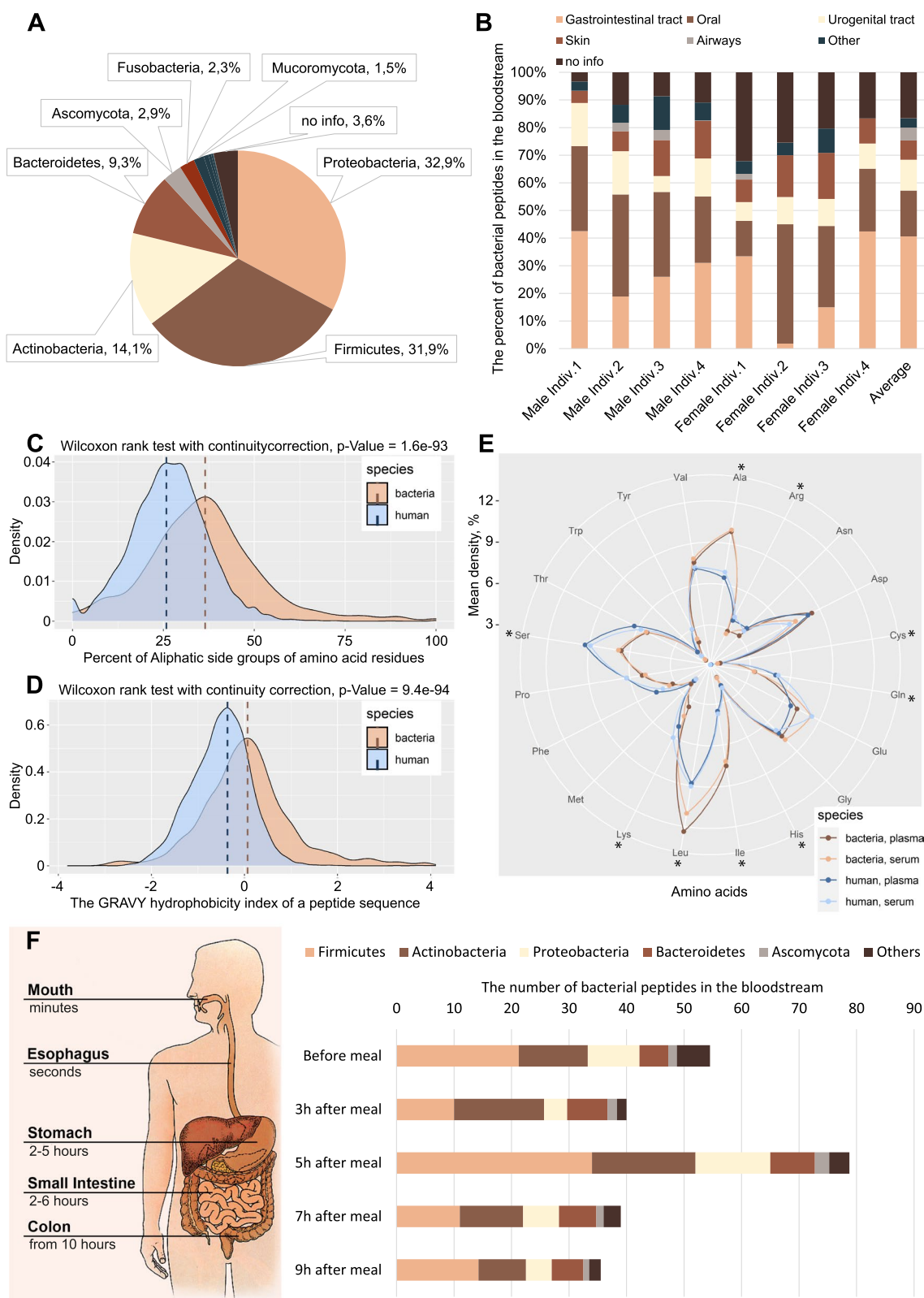


Fig. 2 (See legend on previous page.)

Table 2 Gene ontology (GO) orthologous groups from the eggNOG database enriched in precursor proteins of the identified bacterial peptides (for all 72 plasma/serum samples). Annotations for all proteins of microorganisms, according to the NIH Human Microbiome Project, were used as a background. The Fisher *p*-values were adjusted for multiple testing

| GO ID | Term | Annotated | Significant | Expected | Fisher <i>p</i> -value |
|------------|---|-----------|-------------|----------|------------------------|
| GO:1990220 | GroEL-GroES complex | 215 | 22 | 0.5 | 2.30×10^{-29} |
| GO:0009341 | Beta-galactosidase complex | 22 | 9 | 0.05 | 8.90×10^{-19} |
| GO:0009275 | Gram-positive-bacterium-type cell wall | 7 | 5 | 0.02 | 1.40×10^{-12} |
| GO:0010339 | External side of cell wall | 7 | 5 | 0.02 | 1.40×10^{-12} |
| GO:0005576 | Extracellular region | 6263 | 43 | 14.63 | 3.10×10^{-10} |
| GO:0045262 | Plasma membrane proton-transporting ATP synthase complex, catalytic core F(1) | 69 | 7 | 0.16 | 3.40×10^{-10} |
| GO:0045203 | Integral component of cell outer membrane | 142 | 7 | 0.33 | 5.40×10^{-8} |
| GO:0005737 | Cytoplasm | 74,850 | 223 | 174.85 | 2.70×10^{-6} |
| GO:0009986 | Cell surface | 1686 | 15 | 3.94 | 0.006 |

of resistance to fibrinolysis. The pool of peptides is preserved after passing through the digestive tract.

The change in the representation of peptides in the blood after a meal indicates the probable penetration of bacterial peptides into the bloodstream from the small intestine

According to our peptidome data, *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Bacteroidetes* are the most abundant bacterial phyla with about 33%, 32%, 14%, and 9% of the corresponding identified bacterial peptides, respectively (Fig. 2A). According to the published metagenomic data, the luminal microbiota of the large intestine of a healthy adult person contains about 50–55% *Firmicutes*, 25–30% *Bacteroides*, and less than 25% account for other types of microorganisms, including *Proteobacteria* and *Actinobacteria* [50–54]. On the other hand, according to Vuik et al., the mucosa-associated upper gastrointestinal tract microbiota consists mainly of *Proteobacteria* (mean abundance $40 \pm 2.1\%$) and *Firmicutes* ($38 \pm 2.3\%$) with low levels of *Actinobacteria* and *Bacteroidetes* ($8 \pm 1.6\%$) [55]. In order to find out the microbiota of which part of the intestine most likely makes the greatest contribution to the observed peptidome of the bloodstream, we took serum samples from four healthy donors before a meal and 3, 5, 7, and 9 h after a meal (Additional file 8: Sheet 1). The analysis of the serum peptidome against the HuMiProt90 database showed that the amount of non-human peptides (potentially bacterial peptides) in the bloodstream significantly increased 5 h after a meal (Fig. 2F, Additional file 8: Sheet 3). The comparison of the data with the transit time of food in different parts of the gastrointestinal tract suggests that these peptides probably enter the bloodstream predominantly in the small intestine. At the same time, phyla *Proteobacteria*, *Firmicutes*, and *Actinobacteria* are the main contributors to the increase in (potentially)

bacterial peptides 5 h after a meal (Fig. 2F, Additional file 8: Sheets 5–6).

Some of the identified bacterial peptides are able to activate the immune system by increasing the secretion of pro-inflammatory cytokines and colony-stimulating factors by immunocompetent cells of the peripheral blood

Since the bacterial peptides identified in the blood were present at concentrations comparable to or greater than the physiologically significant levels of human hormones in normal conditions (Fig. 1D), we assumed that these peptides could be the agents of interaction between the microbiota and the human body. Because a significant part of the identified bacterial peptides most likely originates from the intestine, and the intestine is an immune-rich organ, we analyzed the potential immunomodulatory properties of the synthesized peptides, fragments, of microbiota proteins. To suppose the immunological role of the peptides, we incubated isolated peripheral blood mononuclear cells for 3 days with each of the 28 synthetic peptides (at concentration 100 $\mu\text{g}/\text{ml}$) and analyzed the culture medium for the presence of secreted 17 human cytokines by Bio-Plex 200 (Bio-Rad, USA). The profiling revealed that some peptides increased the secretion of pro-inflammatory cytokines IL-1, IL-2, IL-6, IL-12, IL-17, and TNF-alpha, as well as colony-stimulating factors IL-5, IL-7, and GM-CSF—thus, they are potentially able to activate the immune system (Additional file 9, Fig. 3A–E). The secretion of IL-7, IL-13, and monocyte chemoattractant protein 1 (MCP-1) underwent a particularly significant and similar change for many bacterial peptides (Fig. 3C–E). IL-7 is a lymphopoietic growth factor that plays an important role in the maturation and multiplication of lymphoid cells [56], while IL-13 is the central regulator of IgE synthesis and mucin hypersecretion [57]. MCP-1, whose secretion was decreased in the presence

of most of the synthetic peptides, is a factor that plays a role in the chemotaxis of monocytes, memory T cells and dendritic cells to the foci of inflammation [58].

Discussion

An approach was developed to identify peptide fragments of microbiota proteins in complex biological fluids as human blood plasma and serum

The peptidome of blood derivatives has very complex composition, and the analysis is hindered due to the presence of highly represented plasma proteins and other contaminating components, as well as a high range of concentrations of major and minor components [23, 29, 59]. The diversity of the peptides in the blood can be compared with the diversity of the protein composition of the blood plasma. The latter is estimated to have more than 10 million different proteins, with concentrations differing by almost 12 orders of magnitude [36]. We previously published a number of papers on the isolation of peptides from various complex biological fluids, including the human blood serum and plasma [32–35]. In our approach for isolation of bloodstream peptides, we took into account the necessary stages: desorption of low-abundance peptides from the surface of major blood plasma proteins, separation of low molecular weight peptides from proteins, and removal of salts and other components that reduce the efficiency of mass spectrometry analysis of non-tryptic peptides.

The result of the identification, based on a search against a database of amino acid sequences, depends on which database was used [60]. That is why one of the limitations of our study is the inability to increase the number of the analyzed organisms for all food products (animal and plant origin) consumed by donors, as well as to add to the analysis the whole variety of possible biological modifications of human proteins and transcripts. The probability of misinterpretation of data (the number of false positive identifications) increases with the number of protein sequences in the database [61, 62]. In order to assure quality control of this process, different

approaches are used [61], for example, the target-decoy approach [63], which was used in this study. These approaches directly or indirectly increase the threshold score, which, on the one hand, not only reduces the number of false positive identifications but also the number of true identifications. For example, by comparing the results described in this research work, we can see a ten-fold increase in the number of protein sequences (210,438 entries in the database of human proteins UniProtKB-HS vs. 2,128,106 entries in the combined database of human and microbiota proteins HuMiProt90) which results in an 18% decrease in the identifications related to human proteins (15,530 and 12,713 unique peptides, respectively; Additional file 1 and Table 1, Additional file 3). Thus, by uncontrollably increasing the number of proteins in the database used to interpret mass spectrometric data, we will not obtain additional identification results but will lose the existing ones.

It is necessary to perform a search against a database containing only the most likely proteins. We identified such a set of most probable proteins using de novo analysis. De novo identification reveals a significant presence of human and microbiota protein fragments (Fig. 1A, Additional file 2). For example, the contribution of plant proteins (vegetables, fruits, and cereals were included in the diet of donors) was 0.2%, while the contribution of microbial proteins was 7%. Therefore, we purposefully limited the database used. It is known that the total microbiota genome is estimated to exceed the human genome by two orders of magnitude [2], so the identification of low-abundance peptides related to microbiota proteins against the background of more abundant human protein fragments—is a complex computational task, which can lead to a large number of false positive interpretations. To reduce the complexity of the database, CD-HIT [45] was used to collapse the proteins of the human microbiota with more than 90% homology. This procedure reduced the size of the microbiota protein database by almost four times—from 7,202,633 to 1,917,668 protein entries.

(See figure on next page.)

Fig. 3 Analysis of the potential immunomodulatory properties of the bacterial peptides from the blood of healthy donors (three individuals in two replicates). See also Additional file 9. **A** Changes in the secretion level of 17 cytokines by peripheral blood mononuclear cells in response to incubation with Peptide_03 (FISYDNTKTI). “Control” corresponds to the cells to which no peptides were added. The dots show emissions in measured values. Asterisks indicate statistically significant deviations in secreted cytokines compared to the control, according to the Wilcoxon rank sum test, with corrections for multiple testing. **B** Changes in the secretion level of 17 cytokines by peripheral blood mononuclear cells in response to incubation with Peptide_25 (ALEGDAEWEAK). **C** Changes in the secretion level of interleukin 7 (IL-7). “Control” corresponds to the cells to which no peptides were added. p01–p28 are order numbers of the studied bacterial peptides (see peptide sequences in Additional file 9). The upper, middle, and lower dashed lines correspond to 25%, 50%, and 75% quartiles for the values measured for cells without added peptides, respectively. Asterisks indicate statistically significant deviations in secreted cytokines compared to the control, according to the Wilcoxon rank sum test, with corrections for multiple testing. **D** Changes in the secretion level of interleukin 13 (IL-13). **E** Changes in the secretion level of monocyte chemoattractant protein 1 (MCP-1)

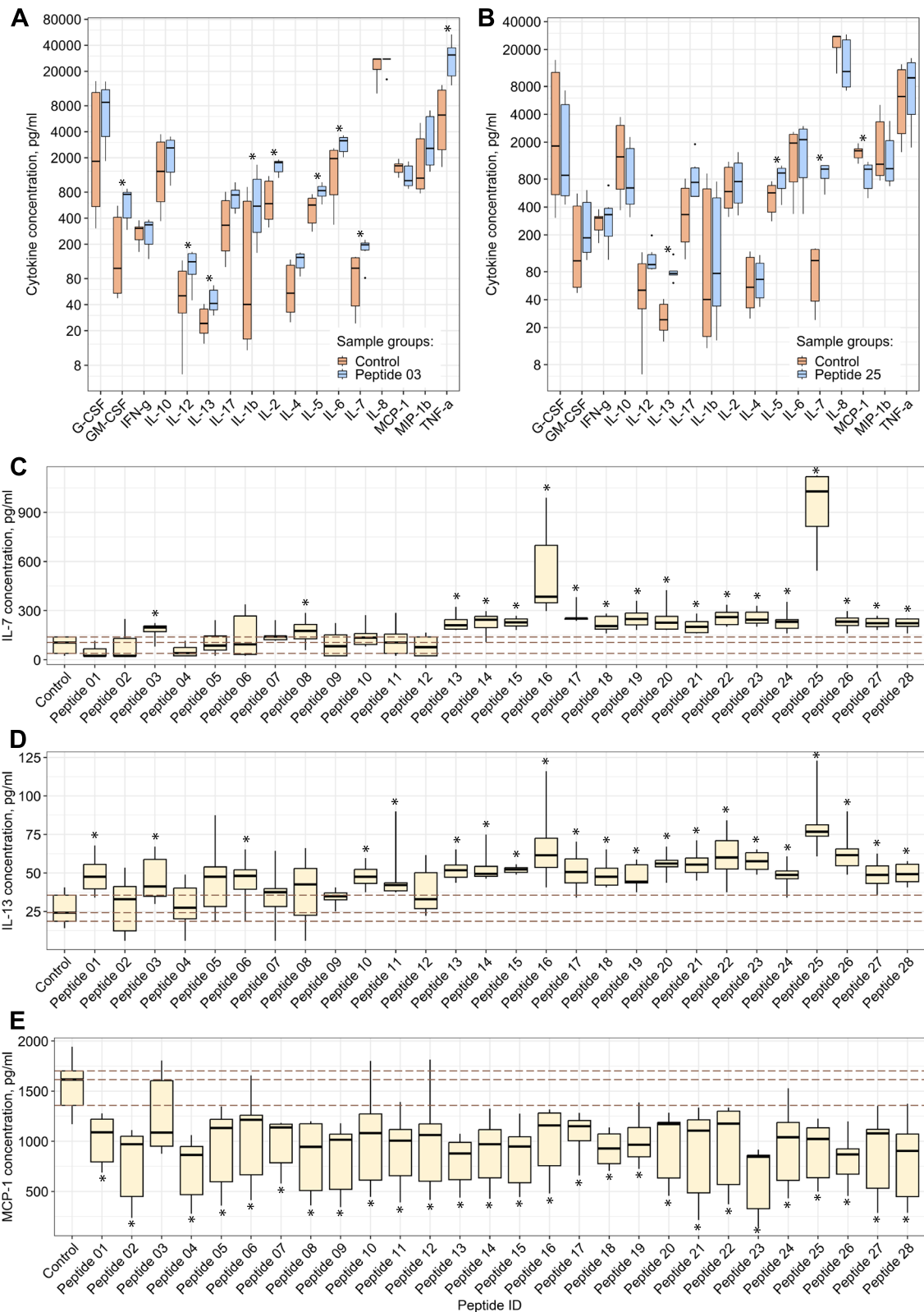


Fig. 3 (See legend on previous page.)

The predominant presence of peptide fragments of human proteins in the blood (Fig. 1A, Additional file 2) allows to use the identification results against the human protein database as a reference to validate the identification results against the protein database for microorganisms. Thus, the scoring algorithm in Scaffold, that includes Naïve Bayes classifier, uses the most represented and well-fragmented human peptides for calibration. For this purpose, the database of proteins of the human microbiota with the database of human proteins was combined. In addition, the possibility of identifying a peptide that can be of microbiota or human origin at the same time was checked. If a mass spectrum could be interpreted as either human or microbiota-derived peptide (possibly because of leucine–isoleucine substitutions since they are not distinguishable in our mass spectrometry conditions), we considered it to be (more likely) of human origin. It was also investigated whether the sequences identified as microbiota peptides could in fact be fragments of human proteins—considering single amino acid substitutions (due to individual genome point mutations), in addition to possible leucine–isoleucine substitutions. Such matches were removed from the analysis altogether, which reduced the number of false positive identifications of microbiota peptides. As an alternative approach to an *in silico* validation, the fragmentation spectra for 30 plasma/serum peptides identified as microbiota-derived were compared with the spectra of synthetic peptides (Fig. 1C, Additional file 4). Identification, *de novo* or by searching against databases, cannot serve as the ultimate stage of investigation because it contains some deliberately incorrect identifications. The successful validation of 30 peptides indicates that we have been able to develop an approach and identify microbiota peptides in the blood of healthy people.

Our approach allowed us to identify 12,713 and 912 unique peptides that belong to human and bacteria proteins, respectively (Table 1, Additional file 3: Sheets 2–3). However, our results showed that with this approach, we were able to identify only around 14.63% of all measured spectra, so the majority of spectra still remain of unknown origin. Many of these spectra are of good quality and with a high chance of identification. We hope that these spectra can lead to new and intriguing discoveries.

We anticipate that our assay could be a starting point for a more sophisticated analysis of the composition of the mucosa-associated bacteria in the small intestine and for monitoring the permeability of the intestinal mucosal barrier based on the blood serum/plasma peptidome

According to the published metagenomic data, the intestinal microbiota of an adult person contains 50–55%

Firmicutes and 25–30% *Bacteroides*, while other types of microorganisms account for less than 25% of the microbiome [50–54]. Our results (Fig. 2A, Additional file 3: Sheets 5–6) show that the peptides from *Proteobacteria* (32.9%), *Firmicutes* (31.9%), and *Actinobacteria* (14.1%) phyla are the most abundant microbiota-derived peptides in the blood. This significant discrepancy requires careful discussion. The gut microbiota is highly heterogeneous throughout the length of the intestine [64]. In addition, it is necessary to distinguish between the luminal and mucosal microbiota [65]. The results of metagenomic analyses, for the most part, reflect the abundance of the luminal microbiota of the large intestine, which has little to do with the composition of the mucosa-associated bacteria of the small intestine. Meanwhile, selective permeability of the small intestine is of particular interest, including the effect of many pharmaceuticals. According to Vuik et al., the mucosal biopsies from 9 different sites, taken with antegrade and subsequent retrograde double-balloon enteroscopy from 14 donors, indicate that the mucosa-associated upper gastrointestinal tract microbiota consists mainly of *Proteobacteria* (mean abundance $40 \pm 2.1\%$) and *Firmicutes* ($38 \pm 2.3\%$) [55]. In the lower gastrointestinal tract, the percentage of *Proteobacteria* is much lower ($5.3 \pm 0.4\%$). On the contrary, *Firmicutes* are predominant species in the large intestine ($64 \pm 7\%$). The *Bacteroidetes* are more represented in the lower gastrointestinal tract ($28 \pm 1.6\%$) than in the small intestine ($8 \pm 1.6\%$). Similar results were obtained by Zhang et al. for the swine intestinal tract [66]. According to this data, the percentages of *Bacteroidetes* and *Proteobacteria* are 9.08% and 30.13%, respectively, in the ileum; 35.36% and 15.45%, respectively, in the cecum; and 27.96% and 18.04%, respectively, in the colon. The *Firmicutes* remain almost equally represented in these three sections of the gut—around 50%. The description of the human ileal bacterial microbiota by Villmones et al. shows that *Firmicutes* and *Actinobacteria* are the most represented bacterial phyla, while *Proteobacteria* are not abundant at all [67]. The analysis of the samples from esophagogastroduodenoscopy and antegrade double-balloon enteroscopy in the jejunum by Leite et al. points at three most represented phyla: *Firmicutes*, *Proteobacteria*, and (much less abundant) *Actinobacteria* [68].

The analysis of peptides found in the blood serum before and after a meal indicates that non-human peptides (potentially bacterial peptides) are likely to enter the bloodstream mainly in the small intestine (Fig. 2F, Additional file 8). At the same time, we understand that we may have taken as “bacterial” part of peptides, food degradation products, which probably also enter the bloodstream. Nonetheless, this experiment, together with the publications cited above, confirms that the bloodstream

pattern of peptide fragments of microbiota proteins can be explained by the peculiarity of the composition of the mucosa-associated bacteria in the small intestine. Intestinal microorganisms shape the mucosal barrier. The mucus of mice whose microbiota was enriched with *Proteobacteria* and TM7 bacteria was more penetrable by bacteria or bacterium-sized beads [69]. The analysis of the physicochemical properties of the identified bacterial peptides shows that such peptides more often contain Ala, Cys, Ile, and Leu, while His, Arg, Lys, Ser, and Gln are not represented as much (Fig. 2E). Lower representation of positively charged amino acids (Arg and Lys) contributes to the peptide's easier permeability of the mucosal barrier [48, 49]. We assume that our approach to the identification of peptide fragments of microbiota proteins in the human plasma and serum may be useful for determining the microbiota composition of hard-to-reach intestinal areas, such as the small intestine, and for monitoring the permeability of the intestinal mucosal barrier.

It should be noted that despite all the arguments presented in this work that the microbiota of the small intestine is a potential source of the bacterial peptides we detected in the bloodstream, we did not receive definitive proof of this thesis. One of the counterarguments is that the representation of the microbiota of the gastrointestinal tract is significantly heterogeneous, from $10e+3/ml$ to $10e+12/ml$, and it is in the large intestine that it reaches its maximum. It can be said that the presence of microbiota in the small intestine is not sufficient to produce peptides in concentrations of 1–100 μM . In addition, the growth of the microbiota may not occur due to the direct entry of food into the corresponding part of the intestine, but due to the release of nutrients by intestinal epithelial cells in response to changes in the level of digestive system hormones [70]. In the studies of the bloodstream peptidome, it was shown that albumin and other highly abundant blood proteins sorb on their surface and transport various low-molecular blood components, including small proteins and peptides [32, 71–73]. According to theoretical estimates, these low-molecular components of the blood proteome are almost completely associated with the surface of transporters [72, 73]. The complex with a large protein is able to protect peptides from being removed from the bloodstream in the kidneys and promotes their accumulation in the blood in quantities suitable for analysis. In addition, the high permeability of the small intestine to various metabolites, including dietary peptides, must be taken into account. Chabance et al. showed that the level of detectable cow caseinoglycopeptide in the bloodstream of volunteers who consumed 500 mL of milk or yogurt was significantly higher in 20 min compared to those who consumed water and

it reached a maximum after 40–60 min [41]. In another study by Vermont et al., 50 g of soy protein were given in the diet of five volunteers for 5 days [42]. The concentration of lunasin, a soybean peptide, in the plasma varies from 50.2 to 110.6 ng/mL for the blood taken at 30 min and from 33.5 to 122.7 ng/mL for the blood withdrawn 1 h after food intake on day 5.

We anticipate that the microbiota peptide metabolites we have identified are involved in the host inflammatory response through modulation cytokine–microbiota or microbiota–cytokine interactions

The interaction of the human organism with microflora occurs at birth and is important for host homeostasis, not only as a direct defense against pathogens but also as a source of constant stimulation and regulation of the immune system [74]. Microbiota is involved in the modulation of inflammatory pathways through molecular interactions, which leads to the formation of the correct immune response by the host organism in response to pathogens. For example, germ-free mice were found to have difficulty producing pro-inflammatory cytokines and recruiting immune cells, resulting in inflammatory hyperreactivity to bacterial infection of the lungs [75]. Microbiota-produced metabolites, especially short chain fatty acids (SCFAs), are essential for a proper host inflammatory response [76]. Our cytokine profiling revealed that some of the identified bacterial peptides increased the secretion of pro-inflammatory cytokines (IL-1, IL-2, IL-6, IL-12, IL-17, and TNF- α), as well as colony-stimulating factors (IL-5, IL-7, and GM-CSF)—thus, they are potentially essential for maintaining host homeostasis (Additional file 9, Fig. 3A–E).

Elucidation of the mechanisms of how individual members of microbiota or the metabolites they produce influence the production of cytokines is of particular importance for maintaining the homeostasis of the host organism [74]. More research is needed to identify pro-inflammatory components of the gut microbiota in various diseases. The findings demonstrate the importance of microbiota peptide metabolites and targeting this mechanism may improve current strategies to modulate the microbiota through fecal transplantation, prebiotics, probiotics, or in other ways for the treatment of diseases through changes in the level of cytokines and inflammatory response.

Conclusions

We identified 912 peptides related to the human microbiota in the blood plasma and serum of healthy persons, using a search against the protein sequence databases that we specially prepared. The identified peptides were validated by a number of bioinformatics approaches, as

well as by comparison with the spectra of synthetic peptides and by quantitative label-free approach, multiple reaction monitoring (MRM). In this work, we have for the first time proven the presence of a large number of peptide fragments originating from the human microbiota proteins in the blood of healthy persons.

The microbiota peptides account for 6.69% of the total peptide content of the blood. The identified bacterial peptides are present at (potentially) physiologically significant concentrations, between 1.01 and 4.28×10^{-4} $\mu\text{g/ml}$. Moreover, the majority of the identified bacterial peptides belong to phyla *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, and they most likely originate from the intestinal microflora as a result of hydrolysis of membrane-associated proteins with trypsin, chymotrypsin, and pepsin—the main proteases of the gastrointestinal tract. Some of the identified bacterial peptides are able to activate the immune system; they increase the secretion of pro-inflammatory cytokines and colony-stimulating factors by immunocompetent cells of the peripheral blood.

The bacterial peptides identified in the blood are likely to originate from the mucosa-associated bacteria of the small intestine, according to what is known about the microbiota of this organ. Such peptides—identified both in the plasma and the serum—are likely to have properties of resistance to fibrinolysis, since the pool of peptides is preserved after passing through the digestive tract. The physicochemical properties of the identified bacterial peptides are similar to those required for selective permeability of mucosal barriers. Our approach to the identification of microbiota peptides in the blood serum and plasma, after additional validation, may be useful for determining the microbiota composition of hard-to-reach intestinal areas, such as the small intestine, and for monitoring the permeability of the intestinal mucosal barrier. The microbiota peptides that we were able to identify in the plasma and serum of healthy persons can be used for further research on host–pathogen interactions.

Methods

Previously, we published a dataset of human blood plasma and serum samples of 10 healthy males and 10 healthy females, fractionated on a set of sorbents (cation exchange Toyopearl CM-650M, CM Bio-Gel A, SP Sephadex C-25 and anion exchange QAE Sephadex A-25) and analyzed by LC-MS/MS individually and pooled in equal amounts (72 samples in total; Additional file 1: Sheet 1) [34]. The mass spectrometry peptidomics data was deposited to the ProteomeXchange Consortium via the PRIDE [77] partner repository (dataset identifier 10.6019/PXD008141) [78]. We analyzed this dataset

again within this work. Additionally, we analyzed a new dataset of blood serum samples from 4 healthy male donors before a meal and 3, 5, 7, and 9 h after a meal (20 samples in total; Additional file 8: Sheet 1), and this dataset is published here for the first time.

Clinical characteristics of healthy donors

Fasting blood plasma and serum samples of ten healthy male (average age 32 years) and ten healthy female (average age 26 years) donors were obtained at the Lopukhin Federal Research and Clinical Center of Physico-Chemical Medicine of Federal Medical and Biological Agency, as well as serum samples from four healthy male donors (average age 28), taken before a meal and 3, 5, 7, and 9 h after a meal. The diet of these four donors was very varied. The volunteers were considered healthy based on their anamneses and follow-up checks. All volunteers included in the study were examined by a therapist before taking the blood, recording vital indices, and monitoring potential infectious sites (oropharynx). If any health problems were detected in a volunteer before or within a month after taking the material, samples from such a volunteer were excluded from the study. The study was approved by the Ethics Committee of the Lopukhin Federal Research and Clinical Center of Physico-Chemical Medicine of the Federal Medical and Biological Agency, and each donor gave a written informed consent to participate in the study.

Sample collection

To obtain plasma or serum, the blood was collected from the cubital vein into blood collection tubes (REF 456023 or REF 456092, respectively; Vacuette tube, Austria). The plasma or serum was obtained from blood samples within 15 min after collection or after coagulation within 1 h at room temperature, respectively. Collection tubes were centrifuged for 15 min at 700 g at room temperature. Blood derivative was separated, aliquoted, and stored at -80 °C until analysis.

Chemicals

Formic acid (FA), trifluoroacetic acid (TFA), acetic acid (AcOH), ammonium hydroxide solution (NH₄OH), and guanidine hydrochloride (GuHCl) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile hypergrade for LC-MS LiChrosolv (AcN), methanol gradient grade for liquid chromatography LiChrosolv (MeOH), and HPLC (high-performance liquid chromatography) grade water were acquired from Merck (Darmstadt, Germany).

Fractionation and peptide extraction

Fasting blood plasma and serum samples were fractionated on Toyopearl CM-650 M (Tosoh Bioscience LLC, USA) weak cation exchange particles. Fasting blood serum samples were also fractionated on CM Bio-Gel A (Bio-Rad Laboratories, Inc., USA) cation exchange gel and SP Sephadex C-25 (GE Healthcare Bio-Sciences AB, Sweden) strong cation exchange particles. Preliminarily, 80 μ l of sorbent was washed two times with 400 μ l WS1 (20 mM AcOH, pH 3.5). To precipitate, the sorbent was centrifuged at 500 g for 10 s. Then 200 μ l of plasma/serum was diluted with 400 μ l WS1 and added to the sorbent. After 30 min of incubation at room temperature with constant stirring, the sorbent was precipitated and the plasma/serum-buffer solution was removed. Then the sorbent was washed three times with 700 μ l WS1. The sorbent was incubated for 15 min with 800 μ l of 0.1% NH₄OH (pH 11) and precipitated and the eluate was collected. Finally, 9 μ l of FA was added to eluate to adjust pH to approximately 3.

Fasting blood serum samples were also fractionated on QAE Sephadex A-25 (GE Healthcare Bio-Sciences AB, Sweden) strong anion exchange particles. Preliminarily, 160 μ l of sorbent was washed two times with 400 μ l WS2 (20 mM Tris, pH 8.26). To precipitate, the sorbent was centrifuged at 500 g for 10 s. Then 200 μ l of serum was diluted with 400 μ l WS2 and added to the sorbent. After 30 min of incubation at room temperature with constant stirring, the sorbent was precipitated and the serum-buffer solution was removed. Then the sorbent was washed three times with 700 μ l WS2. The sorbent was incubated for 15 min with 800 μ l of 0.5% TFA and precipitated and the eluate was collected.

To desorb peptides from the surface of abundant blood proteins, the eluate after peptide extraction was incubated at 98 °C water for 15 min [32]. After heating, additional fractionation was carried out on Discovery DSC-18 SPE Tube reverse phase columns (100 mg, Supelco, USA). The column was conditioned with 500 μ l of MeOH and equilibrated three times with 500 μ l of WS3 (3% AcN in 0.1% TFA). Then the eluate from the previous step was loaded onto the column. The column was washed three times with 300 μ l of WS3. Elution was performed with 1 ml 80% AcN in 0.1%TFA. The eluate was evaporated till 5 μ l on a centrifugal vacuum evaporator, and it was diluted with 10 μ l of WS3 only before the mass spectrometry analysis.

Isolation of peptides from the serum of four healthy male donors (before and after a meal) was performed according to the following protocol; 500 μ l of serum were incubated with 100 μ l of 8 M GuHCl for 30 min at room temperature with constant stirring. The mixture of peptides was separated from large proteins by centrifugation

on Vivaspin 500 concentrators (Sartorius, Germany) with a PES membrane resistant to high concentrations of salts and detergents (10 kDa molecular weight cut-off). Vivaspin 500 concentrators were pre-washed twice with 500 μ l of WS4 (1.33 M GuHCl). Each sample was centrifuged for approximately 2 h at 13,000 g until a clot (approximately 200 μ l) was formed. Then, an additional 500 μ l of WS4 was passed through the concentrator. The mixture of peptides was purified from salts (including GuHCl) on Discovery DSC-18 SPE Tube reverse phase columns (100 mg, Supelco, USA). The column was conditioned with 1 ml of MeOH and equilibrated with 2 ml of WS5 (1% MeOH in 0.1% TFA). Then, each sample was mixed with 100 μ l of 10% MeOH in 1% TFA and loaded onto the column. The column was washed with 2 ml of WS5. Elution was performed with 400 μ l of 50% MeOH in 0.1% TFA. The eluate was evaporated till dry on a centrifugal vacuum evaporator, and it was dissolved in 15 μ l of WS5 only before the mass spectrometry analysis.

LC-MS/MS analysis

The analysis of eluate of fasting blood plasma and serum samples as well as synthetic peptides was performed on a TripleTOF 5600+ mass spectrometer (Sciex, Canada) with a NanoSpray III ion source coupled with a NanoLC Ultra 2D+ nano-HPLC system (Eksigent, USA). The HPLC system was configured in the trap-elute mode. Sample loading solvent and solvent A: 98.9% water, 1% MeOH, 0.1% FA (v/v); solvent B: 99.9% AcN, 0.1% FA (v/v). Samples were loaded on a trap column (Chrom XP C18 (3 μ m 120 Å) 350 μ m \times 0.5 mm; Eksigent) at a flow rate of 3 μ l/min within 10 min and eluted through the separation column (3C18-CL-120 (3 μ m 120 Å) 75 μ m \times 150 mm; Eksigent) at a flow rate of 300 nl/min. For the analyses of plasma and serum samples, the gradient of solvent B was from 5 to 40% (in 90 min) followed by 10 min at 95% and 20 min of re-equilibration with 5%. For the analyses of synthetic peptides, the gradient of solvent B was from 0 to 10% (in 5 min) and from 10 to 50% (in 5 min). Between different samples, two blank 45-min runs consisting of 5–8 min waves of solvent B (5-95-95-5%) were required to wash the system and to prevent carryover. The mass spectra were acquired in a positive ion mode. The information-dependent mass-spectrometer experiment included one survey MS (mass-spectrometry) 1 scan, followed by 50 dependent MS2 scans. The MS1 acquisition parameters were as follows: the mass range for MS2 analysis and subsequent ion selection for MS2 was set to 300–1250 m/z, signal accumulation time 250 ms. Ions for MS2 analysis were selected on the basis of intensity with a threshold of 200 counts per second and the charge state from 2 to 5. The MS2 acquisition parameters were as follows: resolution of the quadrupole

was set to 0.7 Da (UNIT), the measurement mass range was set to 200–1800 m/z , optimization of ion beam focus to obtain maximal sensitivity, signal accumulation time was 50 ms for each parent ion. Collision-activated dissociation was performed with nitrogen gas with the collision energy ramping from 25 to 55 V within 50 ms signal accumulation time. Analyzed parent ions were sent to the dynamic exclusion list for 15 s in order to get the next MS2 spectra of the same compound around its chromatographic peak apex (minimum peak width throughout the gradient was about 30 s). β -Galactosidase tryptic solution (20 fmol) was run with a 15-min gradient (5–25% of solvent B) every two samples and between sample sets to calibrate the mass spectrometer and to control the overall system performance, stability, and reproducibility.

The analysis of eluate of the serum before and after a meal was performed on a Q Exactive HF-X mass spectrometer (ThermoFisher Scientific, USA) with a Nano-Spray Flex ion source coupled to an EASY-nLC 1200 nano-HPLC system. The HPLC system was configured in a trap-elute mode. Solvent A was water with 0.1% FA (v/v), solvent B was 80% AcN, 0.1% FA (v/v). The samples were loaded on a custom-made trap column (100 $\mu\text{m} \times 10 \text{ mm}$ C18 1.8 μm 100 Å Aeris Peptide) at a flow rate of 2.5 $\mu\text{l}/\text{min}$ within 10 min and eluted through a custom-made separation column (75 $\mu\text{m} \times 150 \text{ mm}$ C18 1.8 μm 100 Å Aeris Peptide) at a flow rate of 300 nl/min . The gradient of solvent B was 0–6% (in 5 min) and 6–50% (in 60 min). The mass spectra were acquired in a positive ion mode. The spray voltage was set to 2.2 kV, funnel RF level at 40, and heated capillary at 250 °C. The instrument was configured for data-dependent acquisition (DDA) using the full MS/DD-MS/MS setup with 20 dependent fragment spectra. The full MS resolution was set to 120,000 at m/z 200, and full MS AGC target was 3E6 with an IT of 32 ms. The mass range was set to 400–1600 m/z . The AGC target value for fragment spectra was set at 2E5, and the intensity threshold was kept at 2.5E5. The isolation width was set at 1.4 m/z . The normalized collision energy was set at 27%. The charges for selection were set from 2 to 7, and the peptide match was set to preferred. Isotope exclusion was on and dynamic exclusion was set to 45 s. All the data were acquired in a profile mode.

Peptide identification

To identify the peptides related to the proteome of the human body, we downloaded the UniProt Knowledgebase (UniProtKB), taxon human (with 210,438 entries, hereinafter UniProtKB-HS), on April 1, 2020. This database contains, among other things, the translated

products of short reading frames and transcripts previously annotated as human lncRNA as confirmed at the time of downloading. To identify the peptides belonging to one of the known human microbiota, a database of all known proteins of the human microbiota was created (with 6,387,796 entries). The data on microorganisms belonging to a particular microbiota were taken from the NIH Human Microbiome Project (HMP) [3]. The amino acid sequences of the annotated genes of microorganisms were downloaded from the NCBI nr database. The microorganisms of the human microbiota, for which genomes were sequenced and annotated within the HMP, are listed in Additional file 3: Sheet 1.

To eliminate highly homologous protein sequences from closely related microorganisms, we used the CD-HIT [45] tool which compressed the human microbiota database by 90% homology. The parameters used previously to create the UniRef90 [79] database were applied. The compressed database contained 1,917,668 entries. To identify the microbiota peptides against the background of more represented human peptides, we created a combined protein database consisting of human proteins (UniProtKB-HS) and the CD-HIT compressed database of known proteins of the human microbiota. We named the resulting database HuMiProt90 (Human and Microbiota Proteins, with 2,128,106 entries).

Raw LC-MS/MS data from the Q Exactive HF-X mass spectrometer were converted to.mgf peak lists with MSConvert (ProteoWizard Software Foundation). For this procedure, we used the following parameters: "--mgf --filter peakPicking true [1,2]". Raw LC-MS/MS data from the TripleTOF 5600+ mass spectrometer were converted to.mgf peak lists with ProteinPilot (version 4.5, Sciex, Canada). For this procedure, we used the following parameters: no specific digestion, TripleTOF 5600 instrument, thorough ID search with detected protein threshold 95.0%. For a thorough protein identification, the generated peak lists were searched with MASCOT (version 2.5.1, Matrix Science Ltd., UK) and X! Tandem (ALANINE, 2017.02.01, The Global Proteome Machine Organization) against the UniProtKB-HS or HuMiProt90 database with the concatenated reverse decoy dataset. The precursor and fragment mass tolerance were set at 20 ppm and 50 ppm, respectively. The database search parameters included no enzyme or possible post-translational modification (PTM). For X! Tandem, we also selected parameters that allowed us to quickly check for protein N-terminal acetylation, peptide N-terminal glutamine ammonia loss or peptide N-terminal glutamic acid water loss. The resulting files were submitted to Scaffold 4 (version 4.2.1, Proteome Software Inc., USA) for validation and further analysis.

De novo analysis of mass spectrometry data

The de novo analysis was performed with PEAKS Studio 8.0 (Bioinformatics Solutions Inc., Canada) using the following parameters: charge states from 1 to 4; error tolerance for precursor mass 10.0 ppm, for fragment ion 0.01 Da; no enzyme or possible PTM; and report up to 30 candidates per spectrum. The identified spectra were further analyzed if the ALC (average local confidence) estimated by the software was not lower than 80% for at least one interpretation. Additionally, peptide interpretations from selected spectra were used only if their ALC was not lower than 50%.

The search for precursor proteins of the identified peptides, as well as the search for organisms that contain these proteins in their genomes, was carried out by finding all exact matches of the identified sequences among the annotated genomes at the RefSeq non-redundant database NCBI nr. In this case, possible leucine–isoleucine substitutions were taken into account, since they are indistinguishable in our mass spectrometry conditions. The mass spectrometry driven BLAST analysis involved assignment of mass spectra identified as protein fragments of different organisms to the taxonomic level at which these different organisms converge. That is, if as a result of the de novo analysis a particular mass spectrum was interpreted as a peptide fragment at the same belonging to a protein from, for example, *Escherichia coli* (class *Gammaproteobacteria*) and *Helicobacter pylori* (class *Epsilonproteobacteria*), then this mass spectrum was attributed to the taxonomic level “phylum_Proteobacteria”. If a particular mass spectrum was interpreted as a peptide fragment at the same belonging to a protein from, for example, *Helicobacter pylori* (superkingdom *Bacteria*) and *Homo sapiens* (superkingdom *Eukaryota*), then we could not attribute this mass spectrum to any taxonomic level and marked it as “any creature”.

Spectral counting analysis

Spectral counting (interpreted as a certain peptide) is a straightforward label-free quantitation strategy [80, 81] which we employed to analyze the representation of certain groups of organisms. We took one interpreted spectrum as a unit of measurement. After identification against the human and microbial protein database (HuMiProt90) and validation of the results, we found all occurrences of the identified peptides in various proteins of different microorganisms in the database (the microbial protein database was used before homology compression). Spectral counting of the identified peptides was divided in equal proportions among all possible precursor proteins. Spectral counting was then summarized between proteins

belonging to organisms of the same taxonomic category to estimate the representation of this category in the total pool of bacterial peptides. The mass spectra identified as protein fragments of different organisms were assigned to the taxonomic level at which these different organisms converge. That is, if a particular mass spectrum was interpreted as a peptide fragment at the same belonging to a protein from, for example, *Escherichia coli* (class *Gammaproteobacteria*) and *Helicobacter pylori* (class *Epsilonproteobacteria*), then this mass spectrum was attributed to the taxonomic level “phylum_Proteobacteria”.

In the de novo analysis, each spectrum could be interpreted by several amino acid sequences. Some of these sequences may have multiple entries in the NCBI nr database, and some may not have a single entry. In this case, one spectral counting was divided equally among all occurrences in NCBI nr of possible interpretations. If a possible interpretation had no occurrences in NCBI nr, it was not taken into account. Spectral counting was then summarized between peptides belonging to organisms of the same taxonomic category to estimate the possible contribution of organisms of this category to the identified complex peptide mixture.

In silico validation

The identified peptides were validated in several stages. First, we performed in silico validation based on machine learning, as well as the search for homologous sequences in the UniProtKB-HS database. At the second stage, we validated the MS/MS spectra selected at the first stage by comparing the identified peptides with the synthetic ones.

First, we used the LFDR (local false discovery rate) scoring algorithm in Scaffold with standard experiment wide protein grouping, which includes Naïve Bayes classifier on top score and decoy sets of identifications. For the evaluation of peptide hits, the LFDR of 5% was selected for peptides only. False-positive identifications were based on reverse DB analysis. This led to a global FDR of 0.56%.

We also compared the sequences of the identified bacterial peptides with peptide fragments of human proteins, taking into account possible single amino acid polymorphism. For this validation, we checked possible coincidences of sequences identified as microbiota peptides with protein fragments from the UniProtKB-HS database, paying attention to leucine–isoleucine substitutions (since they are indistinguishable in our mass spectrometry conditions), as well as any single amino acid substitutions. If such coincidences were found, then these identifications were removed from the analysis.

Validation via synthetic peptides

To check the correctness of identification by mass spectrometry, we synthesized 30 amino acid sequences identical to those identified as fragments of microbiota proteins and passed all stages of the *in silico* validation. For this synthesis, we selected sequences with different reproducibility, according to the identification data both in the serum and in the plasma; this way, we considered all peptides, not only those identified with high reliability. The synthesis was carried out by Shanghai Ruifu Peptide Co., Ltd. (China). The quality of the synthesis and the purity of the product (over 90%) were verified by chromatography and mass spectrometry.

To calculate the correlation of the spectra identified in plasma/serum samples and the spectra of synthetic peptides, as well as to visually compare the two spectra, we used R/Bioconductor package “MSnbase” [82].

Validation and quantitation by MRM

Peptide quantitation was performed by multiple reaction monitoring (MRM) using a mass spectrometer QTRAP 6500 (AB Sciex, USA) combined with a liquid chromatograph Infinity1290 (Agilent, USA). Separation was done on a UHPLC column Titan C18, 1.9 μm , 10 cm \times 2.1 mm (Supelco, USA) using the following scheme: 0–0.3 min—5% phase B, 0.3–17 min—gradient from 5 to 50% phase B, 17–18 min—gradient from 50 to 95% phase B, 18–21.5 min—95% phase B, 21.5–23 min—gradient from 95 to 5% phase B, and 23–25 min—5% phase B. Composition of phase A: water 94.9%, AcN 5%, FA 0.1%; composition of phase B: AcN 94.9%, water 5%, FA 0.1%. Other parameters: flow 0.4 ml/min, temperature 40°C. Detection was performed in positive ionization mode by electrospray ionization (Turbo Spray IonDrive) with source temperature 500°C, capillary voltage 5200 V, curtain gas pressure 35 psi, ion source gas 1 pressure 60 psi, and ion source gas 2 pressure 60 psi.

The representation of peptides in the sample was assessed by Skyline 20.1.0.31 (MacCoss Lab, USA) using the area of the peaks of MRM transitions specific for each investigated peptide (Additional file 5: Sheet 1). The results of computational analysis were checked manually. The external calibration consisted of 15 independent measurements of mixtures of 30 synthetic peptides at concentrations between 4.89×10^{-11} M and 8.00×10^{-7} M each. From these measurements, the detection and quantification thresholds were set for each peptide, and a calibration curve was constructed in the region above the quantification threshold. The concentrations of measured bacterial peptides in the plasma and serum samples are shown in Additional file 5.

Cytokine profiling

To study the possible immunological role of the identified bacterial peptides, we incubated peripheral blood mononuclear cells (PBMCs) from 3 healthy donors in 2 replicates for 3 days individually with each of 28 synthesized bacterial peptides at a concentration of 100 $\mu\text{g}/\text{ml}$. To obtain PBMCs, we collected blood samples from the cubital vein into blood collection tubes (REF 456092, Vacuette tube, Austria). The PBMC fraction was separated by centrifugation in Ficoll gradient (1.077 g/ml). PBMCs were seeded on 96-well plates (100,000 per well) in 100 μl of complete medium. Each of the 28 synthesized peptides was dissolved separately in complete nutrient medium at 200 $\mu\text{g}/\text{ml}$, and 100 μl were added per well. Complete culture medium without synthetic peptides was used as a control. The culture medium after incubation was analyzed for the presence of secreted human cytokines using Bio-Plex Pro Human Cytokine 17-plex Assay (Bio-Rad, USA) and Bio-Plex 200 Systems (Bio-Rad, USA). To determine statistically significant deviations in secreted cytokines compared to the control, we used the Wilcoxon rank sum test, with corrections for multiple testing.

Analysis of proteolysis sites in peptides

We analyzed proteolysis sites in the peptides identified in the blood serum and plasma. For 36 proteolysis patterns (mostly corresponding to different proteases) described in R/Bioconductor package “cleaver”, we compared the number of peptides of bacterial and human origin with less than two hydrolysis sites within an amino acid sequence. Fisher’s exact test was used for comparison. If we could find more than one hydrolysis site of a particular protease within the amino acid sequence, we considered this enzyme unlikely to be involved in the production of this particular peptide. Since peptides circulating in the bloodstream are hydrolyzed by exopeptidases, we did not analyze the terminal amino acids.

Analysis of peptide physicochemical properties

We used R/Bioconductor package “Peptides” [83] to analyze the following physicochemical properties of the identified peptides (both bacterial and human): size of the side groups of amino acid residues, aliphaticity, aromaticity, polarity, charge, basic acid properties, length, pI, hydrophobicity, and Boman index (potential protein interaction). The Wilcoxon rank sum test with continuity correction was used to compare the obtained distributions of bacterial and human peptides.

Enrichment analysis of bacterial proteins

Enrichment analysis was performed using eggNOG mapper v2 [84]. The eggNOG database was downloaded

using `download_eggnog_data.py`. Next, annotation was carried out according to orthologous groups from the eggNOG database using `emapper.py` separately for precursor proteins of the identified peptides and for all proteins of microorganisms described in the NIH Human Microbiome Project (HMP) [3]. Enrichment analysis for the precursor proteins of the identified bloodstream peptides was performed with R/Bioconductor package “topGO”. The annotation for all proteins of microorganisms described in the HMP was used as a background. Fisher’s exact test was used for comparison. The *p*-value by Fisher’s test was less than 0.01 and the number of precursor proteins (“Significant” parameter) was more than 2 as threshold values.

Abbreviations

| | |
|--------------------|--|
| NIH | National Institutes of Health |
| HMP | Human microbiome project |
| FDR | False discovery rate |
| PSM | Peptide to spectrum match |
| MRM | Multiple reaction monitoring |
| FA | Formic acid |
| TFA | Trifluoroacetic acid |
| AcOH | Acetic acid |
| NH ₄ OH | Ammonium hydroxide solution |
| GuHCl | Guanidine hydrochloride |
| AcN | Acetonitrile |
| MeOH | Methanol |
| HPLC | High-performance liquid chromatography |
| MS | Mass spectrometry |
| PTM | Post-translational modification |
| ALC | Average local confidence |
| DDA | Data-dependent acquisition |
| UniProtKB-HS | UniProt Knowledgebase, taxon human |
| HuMiProt90 | Human and microbial protein database |
| LFDR | Local false discovery rate |
| PBMCs | Peripheral blood mononuclear cells |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01975-1>.

Additional file 1. Related to Table 1. Summary table of the amino acid sequences identified against the database of human proteins. Sheet 1 – description of all 72 samples, total number of identified peptides and spectra, number of precursor proteins for each analyzed sample. Sheet 2 – list of identified sequences related to human. Sheet 3 – list of protein groups of possible precursor proteins for identified sequences related to human.

Additional file 2. Related to Fig. 1. Summary table of the amino acid sequences identified using the de novo approach, as well as a list of organisms that contain these sequences in their genomes. Sheet 1 – list of identified peptides. Sheets 2, 3, 4, 5, 6, 7, 8, 9, and 10 – lists of organisms that probably contain the identified sequences in their genomes.

Additional file 3. Related to Figs. 1 and 2, and Table 1. Summary table of the amino acid sequences identified against the database of human and microbiota proteins HuMiProt90. Sheet 1 – microorganisms of human microbiota; annotations of genomes taken to compile the database of human and microbiota proteins HuMiProt90. Sheet 2 – list of identified sequences related to human. Sheet 3 – list of identified sequences related to the human microbiota. Sheets 4, 5, 6, 7, 8, 9, 10, and 11 – lists of organisms that probably contain the identified sequences in their genomes. Sheets 12 and 13 – isolation body sites where the microbiota with the

identified peptides is probably located, according to the NIH Human Microbiome Project.

Additional file 4. Related to Fig. 1. Comparison of 30 spectra identified in the blood plasma/serum samples and the spectra of synthetic peptides. b- and y-fragment peaks are labeled. Common peaks are shown in a slightly darker color. The estimates of Mascot Ion Score identification reliability are given.

Additional file 5. Related to Fig. 1. Results of the absolute quantitative MRManalysis of bacterial peptides in the blood plasma and serum. Sheet 1 – MRM transitions specific for each studied peptide. Sheet 2 – results of absolute quantitative MRM analysis. Sheet 3 – low and high amounts of human hormone concentrations in the blood of healthy people.

Additional file 6. Related to Fig. 2. Analysis of proteolysis sites in the peptides identified in the blood serum and plasma. For 36 proteolysis patterns described in cleaver, we compared the number of peptides of bacterial and human origin with less than two hydrolysis sites within an amino acid sequence. Fisher’s exact test and the Chi-square test were used for comparison.

Additional file 7. Related to Fig. 2. Physicochemical properties of the peptides identified in the blood, of bacterial and human origin. The following physicochemical properties are shown: size of the side groups of amino acid residues; aliphaticity; aromaticity; polarity; charge; basic acid properties; length; pI; hydrophobicity; Boman index. The Wilcoxon rank sum test with continuity correction was used to compare the obtained distributions.

Additional file 8. Related to Fig. 2. Summary table of the amino acid sequences identified against the database of human and microbiota proteins HuMiProt90 in the blood serum samples of four healthy donors before a meal and 3, 5, 7, and 9 h after a meal. Sheet 1 – description of samples, total number of identified peptides and spectra, number of precursor proteins for each analyzed sample. Sheet 2 – list of identified sequences related to human. Sheet 3 – list of identified sequences related to the human microbiota. Sheets 4, 5, 6, 7, 8, 9, 10, and 11 – lists of organisms that probably contain the identified sequences in their genomes. Sheets 12 and 13 – isolation body sites where the microbiota with the identified peptides is probably located, according to the NIH Human Microbiome Project.

Additional file 9. Related to Fig. 3. Cytokine response of immunocompetent cells isolated from the bloodstream to incubation with each of the 28 synthetic bacterial peptides identified in the blood plasma/serum of healthy persons.

Acknowledgements

The research was performed using the equipment of Interdisciplinary center for shared use of Kazan Federal University. We thank Dr Roman Trikin for critical reading and editing of the manuscript.

Authors’ contributions

Conceptualization: VMG, GPA, VTI, ENI, and AIM. Data curation: GPA. Formal analysis: GPA, ASU, MSO, TMS, and GAN. Funding acquisition: VMG, VTI, and GPA. Investigation: GPA, ASU, MSO, VOS, IOB, ONB, TMS, OMI, LVL, AVL, NIS, MFN, and ANM. Methodology: GPA, TMS, MSO, and OMI. Project administration: GPA, VMI, ANM, and TVG. Resources: VMG, TVG, and AIM. Software: GPA, ASU, MSO, TMS, and GAN. Supervision: VMG, GPA, VTI, AIM, and TVG. Validation: GPA, LVL, and AVL. Visualization: GPA. Writing—original draft: GPA, MSO, and AAK. Writing—review and editing: GPA, VMG, VTI, ASU, OMI, and LVL. All authors read and approved the final manuscript.

Funding

Part of this work (collection and mass spectrometry analysis of the blood plasma and serum samples from healthy persons) was supported by the Russian Foundation for Basic Research projects no. 17-00-00461 (GPA, ASU, VTI). Part of this work (collection and mass spectrometry analysis of the blood serum samples taken from four healthy donors before a meal and 3, 5, 7, and 9 h after a meal) was supported by the Russian Science Foundation project no. 20-15-00400 (GPA, AAK).

Part of this work (bioinformatics analysis of mass spectrometry data of the blood plasma and serum samples from healthy persons) was supported by grant 075-15-2019-1669 from the Ministry of Science and Higher Education of the Russian Federation (OMI).

Part of this work (peptide extraction and purification) was carried out with state funding for the "Immunopeptidome" project, state registration number of R&D 124031200004-7 (GPA).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The datasets generated and analyzed during the current study have been deposited to the ProteomeXchange Consortium via the PRIDE [77] partner repository. This paper also analyzes existing, publicly available data. The PRIDE partner repository identifiers are 10.6019/PXD027552, PXD027701, 10.6019/PXD027587, 10.6019/PXD027675, PXD027793, and 10.6019/PXD027618 [85–90]. This paper does not report original code.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Lopukhin Federal Research and Clinical Center of Physico-Chemical Medicine of the Federal Medical and Biological Agency (protocol No. 2018/05, dated 08.05.2018), and each donor gave a written informed consent to participate in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 25 July 2023 Accepted: 7 August 2024

Published online: 26 August 2024

References

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015;3:31.
2. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–14.
3. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012;486:215–21.
4. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell*. 2016;164:337–40.
5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464:59–65.
6. Ley R, Turnbaugh P, Klein S, Gordon J. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006;444:1022–3.
7. Turnbaugh PJ, Hamady M, Yatsunenkov T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;457:480–4.
8. Lepage P, Häslér R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology*. 2011;141:227–36.
9. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst*. 2013;105:1907–11.
10. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490:55–60.
11. Ussar S, Griffin NW, Bezy O, Bry L, Gordon JI, Kahn CR, et al. Interactions between gut microbiota, host genetics and diet modulate the predisposition to obesity and article interactions between gut microbiota, host genetics and diet modulate the predisposition to obesity and metabolic syndrome. *Cell Metab*. 2015;22:1–15.
12. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun*. 2012;3:1245.
13. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature*. 2014;513:59–64.
14. Rutsch A, Kantsjö JB, Ronchi F. The gut-brain axis: how microbiota and host inflammasome influence brain physiology and pathology. *Front Immunol*. 2020;11:604179.
15. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science*. 2016;352:586–9.
16. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. *Nature*. 2016;535:65–74.
17. Million M, Tomas J, Wagner C, Lelouard H, Raouf D, Gorvel J-P. New insights in gut microbiota and mucosal immunity of the small intestine. *Hum Microbiome J*. 2018;7:8:23–32.
18. Mäkiyuokko H, Lahtinen SJ, Wacklin P, Tuovinen E, Tenkanen H, Nikkilä J, et al. Association between the ABO blood group and the human intestinal microbiota composition. *BMC Microbiol*. 2012;12:94.
19. Breton J, Tennoune N, Lucas N, Francois M, Legrand R, Jacquemot J, et al. Gut commensal *E. coli* proteins activate host satiety pathways following nutrient-induced bacterial growth. *Cell Metab*. 2016;23:1–11.
20. Gabanyi I, Lepousez G, Wheeler R, Vieites-Prado A, Nissant A, Chevalier G, et al. Bacterial sensing via neuronal Nod2 regulates appetite and body temperature. *Science*. 2022;376:eabj3986.
21. Turner JR. Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol*. 2009;9:799–809.
22. Spadoni I, Zagato E, Bertocchi A, Paolinelli R, Hot E, Di Sabatino A, et al. A gut-vascular barrier controls the systemic dissemination of bacteria. *Science*. 2015;350:830–4.
23. Petricoin EF, Belluco C, Araujo RP, Liotta LA. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat Rev Cancer*. 2006;6:961–7.
24. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*. 2002;359:572–7.
25. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. *Nature*. 2003;425:905.
26. Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics*. 2004;3:311–26.
27. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly available database. *Proteomics*. 2005;5:3226–45.
28. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics*. 2011;10:M110.006353.
29. Schulz-Knappe P, Schrader M, Zucht H-D. The peptidomics concept. *Comb Chem High Throughput Screen*. 2005;8:697–704.
30. Schrader M, Schulz-Knappe P, Fricker LD. Historical perspective of peptidomics. *EuPA Open Proteom*. 2014;3:171–82.
31. Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, et al. Current peptidomics: applications, purification, identification, quantification, and functional analysis. *Proteomics*. 2015;15:1026–38.
32. Ziganshin R, Arapidi G, Azarkin I, Zaryadieva E, Alexeev D, Govorun V, et al. New method for peptide desorption from abundant blood proteins for plasma/serum peptidome analyses by mass spectrometry. *J Proteomics*. 2011;74:595–606.
33. Ziganshin RH, Ivanova OM, Lomakin YA, Belogurov AA Jr, Kovalchuk SI, Azarkin IV, et al. The Pathogenesis of the demyelinating form of Guillain-Barre Syndrome (GBS): proteo-peptidomic and immunological profiling of physiological fluids. *Mol Cell Proteomics*. 2016;15:2366–78.
34. Arapidi G, Osetrova M, Ivanova O, Butenko I, Saveleva T, Pavlovich P, et al. Peptidomics dataset: blood plasma and serum samples of healthy donors fractionated on a set of chromatography sorbents. *Data Brief*. 2018;18:1204–11.

35. Shender V, Arapidi G, Butenko I, Anikanov N, Ivanova O, Govorun V. Peptidome profiling dataset of ovarian cancer and non-cancer proximal fluids: ascites and blood sera. *Data Brief*. 2019;22:557–62.
36. Anderson NL, Anderson NG. The human plasma proteome history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1:845–67.
37. Christmann BS, Abrahamsson TR, Bernstein CN, Duck LW, Mannon PJ, Berg G, et al. Human seroreactivity to gut microbiota antigens. *J Allergy Clin Immunol*. 2015;136:1378–86.e1–5.
38. Bacher P, Heinrich F, Stervbo U, Nienen M, Vahldieck M, Iwert C, et al. Regulatory T cell specificity directs tolerance versus allergy against aeroantigens in humans. *Cell*. 2016;167:1067–78.e16.
39. Iwai K, Hasegawa T, Taguchi Y, Morimatsu F, Sato K, Nakamura Y, et al. Identification of food-derived collagen peptides in human blood after oral ingestion of gelatin hydrolysates. *J Agric Food Chem*. 2005;53:6531–6.
40. Chabance B, Jollès P, Izquierdo C, Mazoyer E, Francoual C, Drouet L, et al. Characterization of an antithrombotic peptide from α -casein in newborn plasma after milk ingestion. *Br J Nutr*. 1995;73:583–90.
41. Chabance B, Marteau P, Rambaud JC, Migliore-Samour D, Boynard M, Perrotin P, et al. Casein peptide release and passage to the blood in humans during digestion of milk or yogurt. *Biochimie*. 1998;80:155–65.
42. Dia VP, Torres S, De Lumen BO, Erdman JW, De Mejia EG. Presence of lunasin in plasma of men after soy protein consumption. *J Agric Food Chem*. 2009;57:1260–6.
43. Nesvizhskii AI, Roos FF, Grossmann J. Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol Cell*. 2006;5:652–70.
44. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics*. 2012;11:M111.010587.
45. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
46. Lee YY, Erdogan A, Rao SSC. How to assess regional and whole gut transit time with wireless motility capsule. *J Neurogastroenterol Motil*. 2014;20:265–70.
47. Hur SJ, Lim BO, Decker EA, Julian MD. In vitro human digestion models for food applications. *Food Chem*. 2011;125:1–12.
48. Witten J, Samad T, Ribbeck K. Selective permeability of mucus barriers. *Curr Opin Biotechnol*. 2018;52:124–33.
49. Huang JX, Blaskovich MAT, Pelingon R, Ramu S, Kavanagh A, Elliott AG, et al. Mucin binding reduces colistin antimicrobial activity. *Antimicrob Agents Chemother*. 2015;59:5925–31.
50. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao J-Z, et al. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol*. 2016;16:90.
51. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486:222–7.
52. Reid G, Younes JA, Van der Mei HC, Gloor GB, Knight R, Busscher HJ. Microbiota restoration: natural and supplemented recovery of human microbial communities. *Nat Rev Microbiol*. 2011;9:27–38.
53. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. 2011;108(Suppl 1):4578–85.
54. Kostic AD, Howitt MR, Garrett WS. Exploring host-microbiota interactions in animal models and humans. *Genes Dev*. 2013;27:701–18.
55. Vuik F, Dicksved J, Lam SY, Fuhler GM, van der Laan L, van de Winkel A, et al. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United European Gastroenterol J*. 2019;7:897–907.
56. Huang H-Y, Luther SA. Expression and function of interleukin-7 in secondary and tertiary lymphoid organs. *Semin Immunol*. 2012;24:175–89.
57. Rael EL, Lockey RF. Interleukin-13 signaling and its role in asthma. *World Allergy Organ J*. 2011;4:54–64.
58. Carr MW, Roth SJ, Luther E, Rose SS, Springer TA. Monocyte chemoattractant protein 1 acts as a T-lymphocyte chemoattractant. *Proc Natl Acad Sci U S A*. 1994;91:3652–6.
59. Mahboob S, Mohamedali A, Ahn SB, Schulz-Knappe P, Nice E, Baker MS. Is isolation of comprehensive human plasma peptidomes an achievable quest? *J Proteomics*. 2015;127 Pt B:300–9.
60. Gorshkov V, Hotta SYK, Verano-Braga T, Kjeldsen F. Peptide de novo sequencing of mixture tandem mass spectra. *Proteomics*. 2016;16:2470–9.
61. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods*. 2007;4:787–97.
62. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002;74:5383–92.
63. Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res*. 2008;7:29–34.
64. Dieterich W, Schink M, Zopf Y. Microbiota in the gastrointestinal tract. *Med Sci (Basel)*. 2018;6(4):116.
65. Crespo-Piazuelo D, Estellé J, Revilla M, Criado-Mesas L, Ramayo-Caldas Y, Óvilo C, et al. Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. *Sci Rep*. 2018;8:12727.
66. Zhang L, Wu W, Lee Y-K, Xie J, Zhang H. Spatial heterogeneity and co-occurrence of mucosal and luminal microbiome across swine intestinal tract. *Front Microbiol*. 2018;9:48.
67. Villmones HC, Haug ES, Ulvestad E, Grude N, Stenstad T, Halland A, et al. Species level description of the human ileal bacterial microbiota. *Sci Rep*. 2018;8:4736.
68. Leite GGS, Weitsman S, Parodi G, Celly S, Sedighi R, Sanchez M, et al. Mapping the segmental microbiomes in the human small bowel in comparison with stool: a REIMAGINE Study. *Dig Dis Sci*. 2020;65:2595–604.
69. Jakobsson HE, Rodríguez-Piñero AM, Schütte A, Ermund A, Boysen P, Bemark M, et al. The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep*. 2015;16:164–77.
70. Power ML, Schulkin J. Anticipatory physiological regulation in feeding biology: cephalic phase responses. *Appetite*. 2008;50:194–206.
71. Gundry RL, Fu Q, Jelinek CA, Van Eyk JE, Cotter RJ. Investigation of an albumin-enriched fraction of human serum and its albuminome. *Proteomics Clin Appl*. 2007;1:73–88.
72. Geho DH, Liotta LA, Petricoin EF, Zhao W, Araujo RP. The amplified peptidome: the new treasure chest of candidate biomarkers. *Curr Opin Chem Biol*. 2006;10:50–5.
73. Lowenthal MS, Mehta AI, Frogale K, Bandle RW, Araujo RP, Hood BL, et al. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem*. 2005;51:1933–45.
74. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell*. 2014;157:121–41.
75. Fagundes CT, Amaral FA, Vieira AT, Soares AC, Pinho V, Nicoli JR, et al. Transient TLR activation restores inflammatory response and ability to control pulmonary bacterial infection in germfree mice. *J Immunol*. 2012;188:1411–20.
76. Vieira AT, Macia L, Galvão I, Martins FS, Canesso MCC, Amaral FA, et al. A role for gut microbiota and the metabolite-sensing receptor GPR43 in a murine model of gout. *Arthritis Rheumatol*. 2015;67:1646–56.
77. Vizzaino JA, Csordas A, Del-Toro N, Dianas JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. 2016;44:11033.
78. ProteomeXchange dataset. Blood plasma and serum samples of healthy donors fractionated on different chromatography sorbents. <https://www.ebi.ac.uk/pride/archive/projects/PXD008141>. The PRIDE partner repository identifier is 10.6019/PXD008141.
79. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. 2015;31:926–32.
80. Geis-Asteggiane L, Ostrand-Rosenberg S, Fenselau C, Edwards NJ. Evaluation of spectral counting for relative quantitation of proteoforms in top-down proteomics. *Anal Chem*. 2016;88:10900–7.
81. Dowle AA, Wilson J, Thomas JR. Comparing the diagnostic classification accuracy of iTRAQ, peak-area, spectral-counting, and emPAI methods for relative quantification in expression proteomics. *J Proteome Res*. 2016;15:3550–62.
82. Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*. 2012;28:288–9.

83. Osorio D, Rondón-Villarreal P, Torres R. Peptides: a package for data mining of antimicrobial peptides. *R J.* 2015;7:4.
84. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47:D309-14.
85. ProteomeXchange dataset. Mass spectrometry analysis of the blood plasma and serum samples from healthy donors. <https://www.ebi.ac.uk/pride/archive/projects/PXD027552>. The PRIDE partner repository identifier is 10.6019/PXD027552.
86. ProteomeXchange dataset. *De novo* identification of blood plasma and serum peptides from samples of healthy donors. <https://www.ebi.ac.uk/pride/archive/projects/PXD027701>. The PRIDE partner repository identifier is PXD027701.
87. ProteomeXchange dataset. Reanalysis of the blood plasma and serum samples from healthy donors. <https://www.ebi.ac.uk/pride/archive/projects/PXD027587>. The PRIDE partner repository identifier is 10.6019/PXD027587.
88. ProteomeXchange dataset. 30 synthetic peptide fragments of proteins from the human microbiota. <https://www.ebi.ac.uk/pride/archive/projects/PXD027675>. The PRIDE partner repository identifier is 10.6019/PXD027675.
89. ProteomeXchange dataset. MRM (Multiple Reaction Monitoring) quantitation of the blood plasma and serum samples from healthy donors. <https://www.ebi.ac.uk/pride/archive/projects/PXD027793>. The PRIDE partner repository identifier is PXD027793.
90. ProteomeXchange dataset. Mass spectrometry analysis of the blood plasma and serum samples from healthy donors before and after a meal. <https://www.ebi.ac.uk/pride/archive/projects/PXD027618>. The PRIDE partner repository identifier is 10.6019/PXD027618.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.