The intestinal proteome of diabetic and control children is enriched with different microbial and host proteins

Elsa Pinto,1 Marisol Anselmo,2 Manuela Calha,2 Andrew Bottrill,3 Isabel Duarte,1 Peter W. Andrew4 and Maria L. Faleiro1*  

Abstract

In this study, the intestinal microbial proteome of children with established type 1 diabetes (T1D) was compared with the proteome of healthy children (Control) with the aim to identify differences in the activity of the intestinal microbiota that not only will contribute to a deeper knowledge of the functionality of the gut in these children but also may provide new approaches to improve the control of the disease. Faecal protein extracts collected from three T1D children (aged 9.3±0.6 years) and three Control children (aged 9.3±1.5 years) were analysed using a combination of 2D gel electrophoresis and spectral counting. The results evidenced markedly differences between the intestinal proteome of T1D children and the Control. The T1D microbial intestinal proteome was enriched with proteins of clostridial cluster XVa and cluster IV and Bacteroides. In contrast, the Control proteome was enriched with bifidobacterial proteins. In both groups, proteins with moonlight function were observed. Human proteins also distinguished the two groups with T1D children depleted in exocrine pancreatic enzymes.

INTRODUCTION

The human gut is colonized by an extraordinary microbial consortium, the activity of which is crucial for human nutrition and physiology. Even though comprehension of the impact of these consortia is still in its early stages, a significant number of diseases related with disturbed intestinal microbiota have been reported [1, 2]. Several studies have shown disturbances in the microbiome signatures during the development of autoimmunity for type 1 diabetes (T1D) [3–6]. In turn, the composition of the gut microbiota is influenced by geographical factors and lifestyle [7–10]. Recently, how the microbiome of infants at high risk of T1D is modulated by geographical location was investigated [11]. In that study, pronounced microbiome diversity across locations was found. Namely, the abundance of Bifidobacterium, Veillonella, Faecalibacterium, Streptococcus and Akkermansia species was significantly different over the sampling locations, evidencing that the microbiome diversity is shaped by geographical region despite the presence of a comparable HLA class II population [11]. Knowledge about the composition of the microbiome of children with established diabetes is very limited [12, 13]. Using a quantitative PCR (qPCR) approach, Murri et al. [12] showed that, in Spanish T1D children, the population of Bifidobacterium, Lactobacillus, Blautia cocoides/Eubacterium rectale group and Prevotella was diminished whereas the proportion of Clostridium, Bacteroides and Veillonella increased in comparison to healthy children. In Turkish T1D children, using a plate counting approach, a decrease also was found in Bifidobacterium population but along with an increase in the population of Candida albicans and Enterobacteriaceae other than Escherichia coli [13]. The shifts in the intestinal microbial population in T1D individuals are key to comprehending the events in the progression of the disease [6]. Although there are significant amounts of data about the microbiota in individuals at risk of autoimmunity, the bacterial functions that are implicated in the disease are so far limited. Since identified proteins can be assigned to taxa and functions, the metaproteomic approach may help us to identify the associated microbiota functions in T1D. The high complexity of the proteome within the human gut can, in part, explain the so far limited number of metaproteomic studies.

Received 23 June 2016; Accepted 12 December 2016  
Author affiliations: 1Faculdade de Ciências e Tecnologia, Centro de Investigação em Biomedicina, Universidade do Algarve Campus de Gambelas, 8005-139 Faro, Portugal; 2Núcleo de Diabetologia, Nutrição e Doenças Metabólicas (NDNDM) do Hospital de Faro, 8000-386 Faro, Portugal; 3Protein and Nucleic Acid Chemistry Laboratory (PNACL), University of Leicester, Leicester LE1 7RH, UK; 4Department of Infection, Immunity and Inflammation, University of Leicester, Leicester LE1 7RH, UK.  
*Correspondence: Maria L. Faleiro, mfaleiro@ualg.pt  
Keywords: type 1 diabetes; Metaproteome; microbiota.  
Abbreviations: 2-DE, 2D gel electrophoresis; COG, cluster of orthologous group; HMRG, human microbial isolate reference genome; IAP, intestinal alkaline phosphatase; Plg, plasminogen; qPCR, quantitative PCR; T1D, type 1 diabetes.  
Five supplementary figures and three supplementary tables are available with the online Supplementary Material.
although there is a high interest in the impact of the human microbiota in health and disease [14, 15]. Nevertheless, there has been a growing interest in the use of this metaproteomic approach to explore the human intestinal microbiota, following the first metaproteomic study conducted by Klaassens et al. [16] on faecal samples from two infants [17–20]. The metaproteomic approach applied to the study of the intestinal microbial communities in T1D children can probe potential differences of T1D metaproteome. Such metaproteome signatures can help us to understand the interaction of the gut microbiota with diabetes.

The objective of this study was to investigate intestinal proteome differences between children with established T1D and children with no reported disease, or historical disease in the family, by combining 2D gel electrophoresis (2-DE) and spectral counting. The goal of the study was to identify distinct microbial and host protein patterns in the two groups.

**METHODS**

**Faecal samples**

Faecal samples were obtained from three healthy and three T1D children (one male and two females for each group) aged 9.3±1.5 and 9.3±0.6 years, respectively. The glycosylated haemoglobin level (HbA1) in diabetic children was 8.558±0.056 %. Samples from each child were collected, delivered to the laboratory within 2–24 h in chilled conditions and stored at −80 °C.

A questionnaire on the children’s eating habits, lifestyle (sedentary vs sports practice), medical history and medication before and during sampling was conducted among parents (Table S1, available in the online Supplementary Material). The children included in the study were selected by matching age, diet type, sports practice and nearby living. The Control children had no history of T1D in the immediate family. The exclusion criteria included antibiotic treatment, hospitalization and diagnosis of infectious diseases up to 3 months before the beginning of the study.

**Ethical considerations**

The ethics committee of Faro Hospital, Portugal, approved this study and all parents of the participating children gave informed consent for the study.

**Protein extraction**

The extraction of proteins from faeces was adapted from previously published methods [16, 21, 22]. Five grams of faeces was transferred to 20 ml of 0.9 % (w/v) NaCl supplemented with 200 µl of chloramphenicol (100 µg ml⁻¹) to inhibit protein synthesis. Glass beads of 4 mm (approximately the same sample volume of the sample) (Sigma) were added. The samples were homogenized for 5 min in a vortex. Then, each sample was centrifugated at 700 g for 1 min at 4 °C and the supernatant was collected. This procedure was repeated three times and the final centrifugation was done at 2790 g for 5 min at 4 °C. The pellet was then subject to five more washing steps with 400 µl of washing buffer [100 mM Tris/HCl (pH 7), 100 mM EDTA and 0.1 ml 100× protease inhibitor mix, GE Healthcare]. The centrifugation between each washing step was performed at 16 090 g for 5 min at 4 °C. The supernatant was discarded. The pellet was resuspended in 500 µl of lysis buffer [25 mM Tris/HCl (pH 7.0), 50 mM EDTA, 1 % (w/v) DTT and 0.20 ml 100× protease inhibitor mix, GE Healthcare] and, subsequently, the cell lysate was sonicated (Soniprep 150, Sanyo) for 15 min with glass beads (Sigma), added to the samples in a 1 : 1 ratio. To eliminate contamination by nucleic acids, 1 µl of DNase (1 U µl⁻¹, Promega) and 5 µl of RNase A (10 mg ml⁻¹, Promega) were added and further incubated on ice for 30 min. The samples were then centrifuged at 3000 g for 10 min at 4 °C. Proteins were precipitated with ice-cold acetone (Merck) (five times the volume of the supernatant) and 0.2 g of trichloroacetic acid (Merck). The samples were incubated for 1 h at −20 °C. Subsequently, the samples were centrifuged at 18 000 g for 30 min at 4 °C. The supernatant was discarded and the pellet dried briefly. After that, 300 µl of the solubilization buffer [7 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 40 mM DTT and 0.8 % (v/v) Pharmalyte] was added and it was left to dissolve at room temperature for 1 h. Then the mixture was centrifuged at 4020 g for 5 min. Finally, the supernatant was collected, and protein quantification was determined using the Bio-Rad Protein Assay kit II (based on the Bradford method), following the manufacturer’s instructions. Protein samples were stored at −80 °C and, prior to electrophoresis, resuspended in rehydration buffer (GE Healthcare).

**Two-dimensional gel electrophoresis**

The protein pattern of faecal microbiota of the two groups of children was evaluated by 2-DE. About 125 µg of protein was used. The protein samples were separated in the first dimension using a linear 18 cm pH 4–7 Immobiline Dry Strip (GE Healthcare). The pH value of the strips was selected based on previous reports of microbial proteins found in the intestinal tract [18–20]. Rehydration and isoelectric focusing was done using an IPGPhor apparatus (Amersham Pharmacia Biotech). The Immobiline strips were hydrated for 1 h at room temperature and rehydrated for 11 h at 20 °C, 30 V and 50 µA per strip. After 12 h of hydration, isoelectric focusing of each strip was done at 20 °C, 50 µA per strip, according to the following programme: 100 V, 1 h, step-n-hold; 500 V, 1 h, step-n-hold; 8000 V, 1 h, gradient, up to 8000 V, 60 000 V per hour, step-n-hold. Afterwards, each strip was equilibrated for 20 min in equilibration buffer [6 M urea, 75 mM Tris/HCl (pH 8.8), 2 % (w/v) SDS, 30 % (v/v) glycerol and traces of bromphenol blue] with 30 mM DTT (Sigma) and following using the same fresh buffer but containing 5 % (w/v) iodoacetamide (GE Healthcare).

The second dimension electrophoresis was performed in 12.5 % (w/v) SDS-PAGE gels, using an Ettan Dalt six apparatus (GE Healthcare) in TGS buffer [250 mM Tris, 1.92 M glycine and 1 % (w/v) SDS]. Protein spots were visualized...
with silver nitrate staining according to the EMBL – Proteomics Core Facility protocol (http://www.embl.de/proteomics/proteomics_services/protocols/index.html).

The determination of the protein patterns was done using an Image Scanner II (GE Healthcare), in combination with computational image analysis done by using IMAGE MASTER 2D Platinum software, version 6 (GE Healthcare). Duplicate 2D gels for each child were analysed (N=6).

**In-gel protein digestion**

The 2-DE maps showed a very complex protein pattern, and the majority of the proteins spots were observed in the region of pI 5 and 6 (Fig. S1a, b). For further analysis of this gel area, the region between the 100 and 20 kDa bands of a prestained marker (Biorad) was selected.

This region was partitioned into 1 cm squares and the fractions were manually excised from the stained 2-DE gels. Then, these fractions were subjected to in-gel protein digestion. The pieces of each gel were destained and proteins were reduced, alkylated and digested with trypsin. In-gel digestion was carried out as before [23] by adding a freshly prepared 1:1 (v/v) mixture of 30 mM potassium ferri-cyanide and 100 mM sodium thiocyanate to the gel slices. Destaining was done for 5 min. The destaining solution was removed and the gel pieces were washed with water, following which the gel pieces were immersed for 20 min in a solution of 400 mM ammonium bicarbonate/100 % acetonitrile 1:1 (v/v) to equilibrate the pH to pH 8. For removal of the aqueous solution, they were washed with acetonitrile. The acetonitrile was removed by air drying for 10 min.

The reduction and alkylation of the proteins was done by adding a freshly prepared solution of 10 mM DTT (Calbio-chem) in 50 mM ammonium bicarbonate for 30 min at 60 °C in a heating block. After removal of excess liquid, a freshly prepared solution of 100 mM iodoacetamide (Sigma) in 50 mM ammonium bicarbonate was added for 30 min at room temperature in the dark. The iodoacetamide solution was removed and the gel slices were transferred to 400 mM ammonium bicarbonate/100 % acetonitrile 1:1 (v/v) for 15 min. This process was repeated twice. The mixture was removed and the gel particles were covered with acetonitrile and left at room temperature for about 15 min. Acetonitrile was removed and air dried for 10 min.

The enzymatic digestion of proteins was done using a trypsin solution [2 µg (Promega, modified porcine trypsin) in 50 mM ammonium bicarbonate], and the mixture was incubated overnight at 37 °C. After digestion, each sample was acidified using 0.1 % (v/v) formic acid and was analysed by LC-MS/MS.

LC-MS/MS was carried out using an RSLCnano HPLC system (Dionex, UK) and an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). Samples were loaded at high flow rate onto a reverse-phase trap column (0.3 mm i.d. × 1 mm), containing 5 µm C18 300 Å Acclaim PepMap media (Dionex) maintained at a temperature of 37 °C. The loading buffer was 0.1 % (v/v) formic acid/0.05 % (v/v) trifluoroacetic acid/2 % (v/v) acetonitrile.

Peptides were eluted from the trap column at a flow rate of 0.3 µl min⁻¹ and through a reverse-phase PicoFrit capillary column (75 µm i.d. × 400 mm) containing Symmetry C18 100 Å medium (Waters, UK) that was packed in-house using a high-pressure device (Proxeon Biosystems, Denmark). Peptides were eluted over a period of 1 h, with the output of the column sprayed directly into the nanospray ion source of an LTQ Orbitrap Velos mass spectrometer.

The LTQ Orbitrap Velos mass spectrometer was set to acquire a 1 microscan FTMS scan event at 60 000 resolution over the m/z range 300–1800 Da in positive ion mode. The maximum injection time for MS was 500 ms and the AGC target setting was 1e⁶. Accurate calibration of the FTMS scan was achieved using a background ion lock mass for C₆H₁₂O₁₄S₂ (401.922718 Da). Subsequently, up to 10 data-dependent HCD MS/MS were triggered from the FTMS scan. The isolation width was 2.0 Da, and normalized collision energy was 42.5. Dynamic exclusion was enabled. The maximum injection time for MS/MS was 250 ms and the AGC target setting was 5e⁴.

**Protein databases**

Tandem MS/MS spectra were searched against two databases, namely, the metadb [17] and the Human Microbial Isolate Reference Genome (HMRG) database [19]. The database metadb [17] is combined with the database db1 [24]. The database db1 includes the metagenomes of two human subjects, a human database, as well as frequent contaminants, such as human keratins and trypsin among others. Since db1 showed insufficient Bacteroides sequences, which is largely recognized to be a quite abundant genus in the human gut, genomic sequences of Bacteroides were added to the metadb database jointly with other bacterial sequences usually found on the gut microbiota (e.g. Clostridia, Lactobacillus and Bifidobacteria). Such sequences were retrieved from the Joint Genome Institute IMG database (https://img.jgi.doe.gov/). The database also includes ‘distracters’ that are sequences unlikely to be found in the healthy gut (supplementary information for this database can be found following the link: http://compbio.orl.gov/human_gut_microbial_metaproteome/databases/).

The HMRG database is composed of 51 human-derived reference isolate genomes, that are known to be common gut inhabitants, from the Joint Genome Institute IMG human microbiome project in a FASTA-formatted protein sequence database and concatenated with the estimated human protein database (National Center for Biotechnology Information). Common contaminants, such as trypsin and keratin, were also included [19]. The construction of the HMRG database based on 51 human-derived reference isolate genomes was done by selecting the genera that have been found from the 16S data from a previous study [25], including also usual gut genome sequences, while preventing redundancy between analogous species and strains.
This database was used to support and complement the metadb database since the HMRG can provide definitive species/protein identifications.

To access this database, the following link was used: http://combio.ornl.gov/crohns_disease_metagenomics_metaproteomics/databases/.

**Protein identification and quantification**

The data were searched using Mascot (version 2.2.04, Matrix Science) and upon import to Scaffold with XiTandem (version 2006.9.15.4, The Global Proteome Machine Organization). Scaffold software (version 3.0, PROTEOME Software) was used to validate and quantify MS/MS-based peptide and protein identifications from the two databases. For protein identification, at least three peptide matches were required (each having a minimum peptide and protein identification probability of 95 %). A protein false discovery rate of \( \leq 0.01\% \) was calculated by the Scaffold software. To improve protein inference, peptides can be credited uniquely to the protein with the highest probability. This approach is followed by the Scaffold software [26]. Protein clusters are formed based on shared peptide evidence. Scaffold’s Protein Cluster Analysis is similar to Mascot’s hierarchical family clustering but is more demanding to consider two proteins in a unique cluster [26]. Protein clusters can be viewed as a representative for a single identification [26–29].

The total spectral count was used to quantify the relative abundance of proteins in the intestinal proteome of the two groups of children. Statistically significant differences in protein spectral counts between the two groups were calculated using Fisher’s exact test and the fold change was analysed. Proteins with \( P < 0.05 \) were considered as differential proteins. In addition, proteins that showed a fold change higher than 2 and below 0.5 were sorted. All Scaffold files, which comprise the search results, are available on request.

**Functional assignment**

For the annotation of the identified proteins, the Cluster of Orthologous Group (COG) database using RPS-BLAST through the Conserved Domain Database website (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used.

The COG assigned to the query protein was the one associated with the best BLAST hit (<1E-10 cutoff). The identified COGs were also mapped on the Kyoto Encyclopedia of Genes and Genomes pathway database (www.genome.jp/kegg/pathway.html).

**Molecular interaction network analysis**

The network analysis of the COGs identified by the proteomics study for the Control (Fig. S4) and T1D (Fig. S5) samples was conducted using the STRING online database version 10 [30]. The search was performed in COG mode, using default parameter values, displaying all interaction data sources except for the textmining (which was not included in the network). After collecting the molecular interaction data, the network visualization was generated using Cytoscape version 3.3.0 [31], where the node size is proportional to its degree (number of connections), and they are coloured according to their COG functional category, namely: (i) green gradient for COGs related to Metabolism (categories C, E, F, G, H and I); (ii) pink colours for Information Storage and Processing (categories J and K) and (iii) orange colours for Cellular Processes and Signalling (categories M and O).

**Phylogenetic trees and presence–absence dot-plots**

In order to visualize the pattern of COGs present in each sample, we manually constructed a dot-plot displaying the presence–absence of the identified COGs in each micro-organism, maintaining the colour code used on the molecular interaction networks (Figs S4 and S5). The phylogenetic relationships between the identified micro-organisms were retrieved from the National Center for Biotechnology Information’s Taxonomy database using its ‘Taxonomy Common Tree’ tool (www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi).

**Bacterial counts by real-time quantitative PCR**

The population of *Bifidobacterium* spp. and *Bacteroides* spp. was estimated by real-time qPCR. Extraction of DNA from stool samples was done using the kit QIAamp DNA Stool Mini Kit, QIAGEN, according to the manufacturer’s instructions. The primers for targeting *Bifidobacterium* spp. were reported by Penders et al. [32] and, for *Bacteroides* spp., the primers were described by Liu et al. [33]. The quantifications by qPCR were performed in a final volume of 25 µl containing the following: 12.5 µl of IQ Supermix (Bio-Rad), 1 µl of each primer (10 pmol µl\(^{-1}\)), 1 µl of probe (5 pmol µl\(^{-1}\)), 5 µl of purified target DNA and nuclease DNA-free water. Amplification and detection were done according to the following conditions: 2 min at 50 °C, 5 min at 95 °C and 42 cycles of 15 s at 95 °C and 1 min at 60 °C. Fluorescence was collected during the extension step. The values of \( \log_{10} \) total cells per gram wet weight faeces of *Bifidobacterium* and *Bacteroides* spp. were calculated from the threshold cycle (CT) values using the constructed standard curves.

**RESULTS AND DISCUSSION**

**Data generation and searched databases**

It was not possible to match the 2-DE maps of faecal samples of T1D children with those of the Control. The majority of the proteins spots were observed in the region of pI 5 and 6 (Fig. S1a, b). Therefore, this area of the gel was selected to identify and quantify the microbial and human proteins being differently represented in the faecal samples of T1D and Control children.

The two databases used, metadb and HRMG, allowed the identification of several proteins in these databases. Moreover, they allowed us to compare the identification of proteins from each database (Table 1). Using the HRMG
were observed in all biological samples and which showed a
The proteome analysis proceeded with the proteins that
allowing more definitive species/protein identification.
linking 51 human-derived reference isolate genomes from
databases; in particular, metadb is a metagenome-derived
(Figs S2 and S3). The difference in the overlap between the
databases superimposing a higher number of proteins but
of proteins between the two groups was observed when the
metadb database was used, namely, 499 proteins overlap
(Figs S2 and S3). The difference in the overlap between the
two databases, likely, is related to the assembly of the two
databases; in particular, metadb is a metagenome-derived
database superimposing a higher number of proteins but
clearly is a more taxonomy-restricted database in compar-
ison to the HMRG database, which was mainly generated by
linking 51 human-derived reference isolate genomes from the
Joint Genome Institute IMG human microbiome project
allowing more definitive species/protein identification.
The proteome analysis proceeded with the proteins that
were observed in all biological samples and which showed a
fold change >2.0 and P<0.05.
It should be noted that, despite the great potential of a meta-
proteomics approach for understanding the impact of the
intestinal microbiota on host health and disease, this meth-
ology still poses some challenges; namely, the different
protein extraction methods (differential centrifugation,
ultracentrifugation with density gradient or direct extraction
from stool samples) may have a meaningful impact on pro-
tein/peptide identification, particularly the differential
centrifugation, which may result in diminished host and
food-associated proteins along with microbial cell surface
proteins [34]. Additionally, technical variability arises
mostly from the protein extraction method, with the diges-
tion step and instrument-related variables (such as instru-
ment’s variance and stability) also contributing to it [35].
However, despite the significant impact of the sample type
and proteomic methodology on technical variability, studies
with microbial cultures are expected to show lower technical
variability when compared to other types of biological sam-
ple, such as faecal samples, mostly due to their homogeni-
zation being easier to accomplish [36]. Our coefficients of
variance in spectral counts (Table S2 and S3) may result from
all these factors.
An additional demanding step is the selection of the appropri-
ate databases for the identification of mass spectra, which is
very challenging by virtue of the tremendous diversity and
individual variations of the gut microbiota [29]. In human
metaproteomic studies, a recently recommendation for data-
base purposes has been the combination of metagenomic
sequencing with metaproteomic analysis, and the use of a
database comprising all metagenomic sequences from all the
samples seems to be more suitable in comparison to a sample-
matched database [37–39].

The majority of the identified proteins, either in the Control
or in the T1D samples, were of microbial origin. Human (or
contaminants) proteins represented 22 % of the Control and
21 % of the T1D samples (Table S2 and S3). In the Control
samples, bacterial proteins from Bifidobacterium adolescen-
tis, Bifidobacterium longum subsp. infantis, Ruminococcus,
Collinsella aerofaciens, Coprococcus comes and Clostridium
sp. were the most abundant (Table 1, Fig. 1a). In contrast,
the bacterial origin of the proteins identified in T1D sam-
dles was different from Control samples, namely, including
proteins produced by Eubacterium rectale, Faecalibacterium
prausnitzii Bacteroides dorei and Bacteroides uniformis
(Table S3).

The 21 most abundant bacterial proteins observed in the
Control group were distributed among seven functional cat-
egories (Table 1), with the highest number of proteins being
associated with carbohydrate transport and metabolism
(seven proteins). The molecular interaction network for
Control samples (Fig. S4) suggests the over-representation
of metabolism-related COGs (depicted in green), particu-
larly the aforementioned functional category G, which is
not only the most frequent class (6 distinct nodes out of a
total of 14) but also the most densely connected (bigger
nodes imply more connections since the size of the node is
proportional to its degree). As expected, metabolic COGs
connect mostly to each other, given their close functional
interaction within metabolic pathways.
In contrast, in the T1D proteome, the 26 most abundant
bacterial proteins were distributed among 11 functional cat-
egories, some of which were not observed in the Control
group, namely, amino acid transport and metabolism, co-
enzyme transport and metabolism, post-translational modi-
fication, protein turnover and chaperones and transcription
(Table 1, Fig. 1b).

The T1D interaction network (Fig. S5) clearly shows that the
metaproteome of T1D samples is functionally enriched
in core energy metabolism proteins, with a strong emphasis
on sugar transport and processing. This network is largely
composed of metabolic functional categories, three of which
are absent from the Control network (the aforementioned
categories E, H and K). This network is tightly packed, with
many nodes displaying high connectivity to a diverse set of
COGs, mirroring the intricate nature of the cellular pro-
ceses taking place in the gut microbiota. Remarkably, this
network displays two dominant categories, each presenting
five distinct COGs: carbohydrate transport and metabolism
(G) and energy production and conversion, which are two
Table 1. Proteins identified using the HMRG database that were more abundant or only observed in the intestinal proteome of T1D or Control children

Differences in the identification of proteins using the metadb are marked. Using the metadb:
a, this protein was identified as a NifU homologue involved in Fe-S cluster formation; b, was not identified; c, was identified only in two control and two diabetic children; d, the origin of this protein was Bacteroides fragilis NCTC 9343; e, was not identified in the three diabetic children; f, was not quantitatively significant; g, the origin of this protein was Lactococcus lactis lactis II1403; h, the origin of this protein was Bacteroides thetaiotaomicron VPI-5482; i, also identified in control children; j, also identified in control children and significantly more abundant.

<table>
<thead>
<tr>
<th>COG functional category</th>
<th>COG</th>
<th>Origin</th>
<th>Protein name</th>
<th>Molecular mass (kDa)</th>
<th>pI</th>
<th>Spectral count</th>
<th>Fisher’s test (P value)</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteins more abundant in Control children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td></td>
<td></td>
<td>Elastase 3A, pancreatic preprotein</td>
<td>29</td>
<td>6.43</td>
<td>&lt;0.0001</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elastase 3B preprotein</td>
<td>29</td>
<td>5.85</td>
<td>0.00037</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Energy production and conversion</td>
<td>COG026</td>
<td>Bifidobacterium longum infantis ATCC 15697</td>
<td>Acetate kinase</td>
<td>44</td>
<td>5.54</td>
<td>0.0055</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0822</td>
<td>Collinsella aerofaciens ATCC 25986</td>
<td>Hypothetical proteina</td>
<td>26</td>
<td>5.03</td>
<td>0.029</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG1454</td>
<td>Collinsella aerofaciens ATCC 25986</td>
<td>Alcohol dehydrogenase, class IV</td>
<td>45</td>
<td>4.93</td>
<td>0.0085</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Nucleotide transport and metabolism</td>
<td>COG2779</td>
<td>Clostridium baratti DSM 16793</td>
<td>Formate-tetrahydrofolate ligase</td>
<td>60</td>
<td>5.61</td>
<td>&lt;0.0001</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Carbohydrate transport and metabolism</td>
<td>COG3957</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>Phosphoketolase</td>
<td>93</td>
<td>5.09</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG2407</td>
<td>Ruminococcus obenum ATCC 29174</td>
<td>1-Fucose isomerase</td>
<td>44</td>
<td>5.01</td>
<td>&lt;0.0001</td>
<td>11.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0537</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>Glyceraldehyde-3-phosphate dehydrogenaseb</td>
<td>38</td>
<td>5.16</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0021</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>Transketolase</td>
<td>76</td>
<td>4.99</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proteins only observed in Control children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COG176</td>
<td>Ruminococcus obenum ATCC 29174</td>
<td>Transaldolase, putative, TaC familyc</td>
<td>24</td>
<td>5.29</td>
<td>0.00043</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clostridium sp. SS2/1</td>
<td></td>
<td>24</td>
<td>4.96</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid transport and metabolism</td>
<td>COG1024</td>
<td>Clostridium sp. SS2/1</td>
<td>Enoyl-CoA hydratase/carnithine racemase</td>
<td>28</td>
<td>5.12</td>
<td>0.0028</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG183</td>
<td>Clostridium sp. SS2/1</td>
<td>Acetyl-CoA acetyltransferases</td>
<td>41</td>
<td>5.41</td>
<td>&lt;0.0001</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation, ribosomal structure and biogenesis</td>
<td>COG0264</td>
<td>Bifidobacterium longum infantis ATCC 15697</td>
<td>Translation elongation factor Ts</td>
<td>30</td>
<td>5.07</td>
<td>0.034</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG050</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>Translation elongation factor 1A (EF-1A/EFTu)</td>
<td>44</td>
<td>4.94</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>Cell wall/membrane/envelope biogenesis</td>
<td>COG1078</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>UDP-galactose 4-epimerase</td>
<td>39</td>
<td>5.84</td>
<td>&lt;0.0001</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proteins only observed in T1D children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COG0888</td>
<td>Bifidobacterium adolescentis L2-32</td>
<td>Chymotrypsin C preproprotein</td>
<td>29</td>
<td>7.54</td>
<td>&lt;0.0001</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proteins more abundant in T1D children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COG functional category</td>
<td>COG</td>
<td>Gene name</td>
<td>Protein name</td>
<td>Molecular mass (kDa)</td>
<td>pI</td>
<td>Spectral count</td>
<td>Fold change</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----</td>
<td>-----------</td>
<td>--------------</td>
<td>---------------------</td>
<td>----</td>
<td>---------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>Energy production and conversion</strong></td>
<td>COG1274</td>
<td>Faecalibacterium prausnitzii A2-165</td>
<td>Phosphoenolpyruvate carboxykinase</td>
<td>59</td>
<td>5.21</td>
<td>5.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG2086</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Electron transfer flavoprotein, beta subunit</td>
<td>28</td>
<td>5.05</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0597</td>
<td>Roseburia intestinalis L1-82</td>
<td>Alkylamine monooxygenase</td>
<td>49</td>
<td>5.65</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0341</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>3-Hydroxyacyl-CoA dehydrogenase</td>
<td>17</td>
<td>5.85</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amino acid transport and metabolism</strong></td>
<td>COG0334</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Glutamate dehydrogenase [NADP(+)]</td>
<td>49</td>
<td>5.43</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0597</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Fumarate hydratase</td>
<td>128</td>
<td>5.64</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0115</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Branched-chain amino acid transaminase (ilvE)</td>
<td>41</td>
<td>5.48</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0366</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Pyruvate:ferredoxin (flavodoxin) oxidoreductase</td>
<td>49</td>
<td>5.43</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0179</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>NAD-dependent glyceraldehyde-3-phosphate dehydrogenase</td>
<td>39</td>
<td>5.10</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Carbohydrate transport and metabolism</strong></td>
<td>COG0166</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Glucose-6-phosphate isomerase</td>
<td>59</td>
<td>5.21</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0126</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Phosphoglycerate kinase</td>
<td>43</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0138</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>5-Aminoimidazole-4-carboxamide ribonucleotide transformylase</td>
<td>58</td>
<td>6.48</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lipid transport and metabolism</strong></td>
<td>COG0148</td>
<td>Bacteroides uniformis ATCC 8492</td>
<td>Enolase</td>
<td>46</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0188</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>3-Hydroxyacyl-CoA dehydrogenase</td>
<td>28</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0419</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>3-Hydroxyacyl-CoA dehydrogenase</td>
<td>28</td>
<td>4.97</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Translation, ribosomal structure and biogenesis</strong></td>
<td>COG0480</td>
<td>Claribacter leptum DSM 753</td>
<td>Translation elongation factor 2 (EF-Tu)</td>
<td>48</td>
<td>5.43</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0498</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Securin G2</td>
<td>58</td>
<td>5.12</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0631</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Securin G2</td>
<td>58</td>
<td>5.20</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post-translational modification, protein turnover, and chaperones</strong></td>
<td>COG0191</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>Securin G2</td>
<td>58</td>
<td>4.97</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proteins only observed in T1D children</strong></td>
<td>COG0166</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Glutamate dehydrogenase [NADP(+)]</td>
<td>49</td>
<td>5.43</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0126</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>Phosphoglycerate kinase</td>
<td>43</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0138</td>
<td>Eubacterium rectale ATCC 33656</td>
<td>5-Aminoimidazole-4-carboxamide ribonucleotide transformylase</td>
<td>58</td>
<td>6.48</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0148</td>
<td>Bacteroides uniformis ATCC 8492</td>
<td>Enolase</td>
<td>46</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0188</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>3-Hydroxyacyl-CoA dehydrogenase</td>
<td>28</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COG0419</td>
<td>Faecalibacterium prausnitzii M21/2</td>
<td>3-Hydroxyacyl-CoA dehydrogenase</td>
<td>28</td>
<td>4.97</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pinto et al., Microbiology 2017;163:161–174
strongly interconnected classes, i.e. the carbohydrate transport and metabolism represents a particularly well studied branch of the cell’s core energy metabolism.

A more detailed description is given below.

The proteome of healthy children is enriched with bifidobacteria proteins

Among the 21 more abundant proteins in the intestinal proteome of healthy children, 8 proteins were from bifidobacteria, 3 of which were associated with carbohydrate transport and metabolism (Table 1, Fig. 1a). These are phosphoketolase, glyceraldehyde-3-phosphate dehydrogenase and transketolase (Bifidobacterium adolescentis and Bifidobacterium longum subsp. infantis). Interestingly, from the proteins only found in the intestinal Control proteome, only phosphoglycerate mutase (BADO_0743, EC 5.4.2.12) produced by Bifidobacterium adolescentis, was observed (Table 1). Transketolase (BADO_0882, EC 2.2.1.1) and phosphoketolase (BADO_0732, EC 4.1.2.9/4.1.2.22) are two enzymes of the pentose phosphate pathway that participate in glucose breakdown after the cleavage to fructose-6-phosphate. Another protein, acetate kinase (Blong_1731), which converts acetyl phosphate into acetate with ATP production, was also abundant. Interestingly, in the study of Kim et al. [40], the acetate kinase (Blong_1731) of Bifidobacterium longum infantis ATCC 15697 was constitutively produced in Mueller–Hinton medium supplemented with different substrates (e.g. lactose, mucin, inulin, FOS and GOS). Another interesting observation was reported in the study of Erickson et al. [19], which also used a metaproteomic approach to study the human host–microbiota signatures of Crohn’s disease. In their study, the abundance of acetate kinase originated by Bifidobacterium adolescentis L2-32 and Faecalibacterium prausnitzii strain A2-165 and strain M21/1 was lower in patients with Crohn’s disease in the ileum but its abundance was similar between healthy and patients with Crohn’s disease in the colon. The abundance of this enzyme in the intestinal proteome of the Control children suggests the existence of an unidentified substrate or environmental condition that induces the synthesis of this enzyme. The abundance of proteins of the pentose and glycolytic pathways in response to the intestinal environment has been reported in Bifidobacterium longum and also in the pathogenic Listeria monocytogenes [41, 42].

Besides the known role of phosphoglycerate mutase in glycolysis, this enzyme can be involved with the plasminogen (Plg)—plasmin system of the human host [43]. The presence of human Plg on the bifidobacterial surface gives the bacteria a cell surface proteolytic activity aiding their movement within the gastrointestinal system and helping with the acquisition of essential nutrients during colonization of the host [43]. Another abundant protein from Bifidobacterium adolescentis, with reported Plg-binding proprieties, is the elongation factor 1A (EF-1A/EFTu) (EF-Tu) [44, 45]. An elongation factor from Bifidobacterium longum subsp. infantis also was identified, namely, EF-Ts.
Our data show that the gastrointestinal environment of the Control children is particularly enriched with *Bifidobacterium* proteins, particularly with Plg-binding properties, in contrast to the T1D that is enriched with *Eubacterium rectale*, *Faecalibacterium prausnitzii* and *Bacteroides dorei* proteins. The reasons why the bifidobacteria Plg-binding proteins were under-represented in the samples of T1D children are not obvious.

In the Control proteome, formate-tetrahydrofolate ligase from *Clostridium bartletti* also was prominent. Perhaps, it is relevant that the synthesis of this enzyme has been reported in *Lactobacillus reuteri* cells exposed to bile salts [46].

The counts of *Bifidobacterium* spp. between Control and T1D children were not significantly different (*P*=0.084). For Control children, the *Bifidobacterium* spp. population achieved...
10.50±0.50 log_{10} cells g^{-1} faeces, whereas in T1D children, this population reached 9.30±0.41 log_{10} cells g^{-1} faeces. This finding suggests that the population of *Bifidobacterium* is not disturbed by the diabetic intestinal milieu but its protein pattern is modified.

**T1D proteome is enriched with proteins of some species of clostridial clusters XIVa and IV and from Bacteroides**

In the T1D faecal samples, it was observed that eight proteins from *Faecalibacterium prausnitzii*, that belong to the clostridial cluster IV (also known as *Clostridium leptum*), were more abundant [47]. Namely, phosphoenolpyruvate carboxykinase; electron transfer flavoprotein, beta subunit; a hypothetical protein; pyruvate-ferredoxin (flavodoxin) oxidoreductase; glutamate dehydrogenase [NADP(+)]; glyceraldehyde-3-phosphate dehydrogenase, type I; phosphoglycerate kinase; and enoyl-CoA hydratase/carnithine racemase. Other gut bacteria also showed higher abundance of some of these proteins; e.g. phosphoenolpyruvate carboxykinase was associated with two *Faecalibacterium prausnitzii* strains, strains A2-165 and M21/2. The high abundance of phosphoenolpyruvate carboxykinase in the metaproteome of human faecal samples was previously reported [20, 48]. Electron transfer flavoprotein, beta subunit, and phosphoglycerate kinase from *Eubacterium rectale* were also more abundant. The higher abundance of glutamate dehydrogenase [NADP(+)] was seen from both *Eubacterium rectale* and *Akkermansia muciniphila*, and the greatest abundance of glyceraldehyde-3-phosphate dehydrogenase, type I, was observed simultaneously from *Eubacterium rectale* and *Bacteroides dorei* (Table 1, Fig. 1b). The anaerobe *Faecalibacterium prausnitzii* is a major member of the intestinal microbiota and a key producer of butyrate, consuming acetate in the process [47, 49–51]. In most anaerobic micro-organisms, the oxidative decarboxylation of pyruvate to produce acetyl-CoA is carried out by pyruvate-ferredoxin (flavodoxin) oxidoreductase [52], which in *Faecalibacterium prausnitzii* plays a central role in butyrate production [53]. It is possible that, in the intestinal milieu of diabetic children, *Faecalibacterium prausnitzii* is experiencing a deficit of acetate supplied by other sources, such as those provided by *Bifidobacterium* spp. [54–56], thus recycling its own source of acetate to proceed with the production of butyrate [53]. Phosphoenolpyruvate carboxykinase (FP2_13680) is a gluconeogenic enzyme, and the fact that the abundance of this enzyme could be assigned to two *Faecalibacterium prausnitzii* strains in the T1D proteome suggests that faecalibacteria are experiencing low levels of glycolytic substrates. Such low levels of glycolytic sources may occur due to competition [57] resulting from the diabetic diet that is restricted in carbohydrates (of the recommended total daily energy intake, carbohydrate is only 50–55%) [58]. As mentioned above, the presence of glutamate dehydrogenase [NADP(+)] was linked to *Eubacterium rectale* and *Akkermansia muciniphila*. Glutamate dehydrogenase [NADP(+)] can play different physiological roles, namely, it is involved in the reductive amination of α-ketoglutarate, osmotic equilibrium and also as an electron sink [20, 59, 60]. This latter role is protective by resulting in a reduced level of free electrons in strict anaerobes [59]. In addition to the overall enrichment of the proteins mentioned above, *Eubacterium rectale*, that is included in the cluster XIVa (also known as *Clostridium cocoides* cluster) [47], showed three abundant proteins: 4-hydroxybutyrate CoA transferase (EUBREC_3075), branched-chain amino acid transaminase (iLvE) (EUBREC_2380) and 5-aminoimidazole-4-carboxamide ribonucleotide transformylase (EUBREC_1567). *Eubacterium rectale* is another key butyrate producer, consuming acetate in the human gut [61, 62]. The 4-hydroxybutyrate CoA transferase is the terminal enzyme of the aminobutyrate/succinate pathway involved in butyrate synthesis [62]. Its increased abundance seems to be consistent with the hypothesis of low acetate available in the T1D gut mentioned above, forcing *Eubacterium rectale* to produce butyrate through the alternative pathway of aminobutyrate/succinate. From the group of bacterial proteins only found in T1D children, four belong to *Eubacterium rectale* [l,l-diaminopimelate aminotransferase (EUBREC_2541), orotine 5-phosphate decarboxylase (EUBREC_2481), 3-hydroxybutyryl-CoA dehydrogenase (EUBREC_1017, EC4.2.1.17) and UDP-glucose-epimerase (EUBREC_3245, EC5.1.3.2)]. The 3-hydroxybutyryl-CoA dehydrogenase is involved in butyrate metabolism and increased synthesis has been reported in germ-free mice colonized with *Eubacterium rectale* and *Bacteroides thetaiotaomycron* [63] suggesting that, in the diabetic intestinal milieu, *Eubacterium rectale* encounters other gastrointestinal members that induce over-production of this enzyme. UDP-glucose-epimerase is required for galactose metabolism to interconvert UDP-galactose and UDP-glucose and it has been associated with the osmotic stress response by sustaining the cohesion of lipoteichoic acid layers of Gram-positive bacteria or the lipopolysaccharide of Gram-negative bacteria, thus making cells more resistant to osmotic stress [64]. Because its production was assigned only to *Eubacterium rectale*, we can suggest that UDP-glucose-epimerase helps this bacterium to overcome the osmotic challenges in the gastrointestinal tract of T1D children. Interestingly, the highest number of proteins included in the functional group of amino acid transport and metabolism (this functional group was absent in the Controls) were originated from *Eubacterium rectale* (Table 1, Fig. 1b), namely the branched-chain amino acid transaminase (iLvE) (EUBREC_2380, involved in valine, leucine and isoleucine degradation), the glutamate dehydrogenase [NADP(+)] (EUBREC_3138) and the l,l-diaminopimelate aminotransferase (EUBREC_2541, involved in lysine synthesis). It is known that the main lysine fermentation products obtained by the growth of intestinal bacteria are butyrate and acetate [65], whereas valine, leucine and isoleucine fermentation products are the short-chain fatty acids isobutyrate from valine, isovalerate and ammonia from leucine and the 2-methylbutyrate from isoleucine [65]. These facts reinforce the hypothesis that short-chain fatty acids production by *Eubacterium rectale* under T1D intestinal environment require amino acid metabolism. Two intriguing observations from the T1D intestinal proteome were noticed: (i) the abundance of glyceraldehyde-3-phosphate dehydrogenase (EL88_10785, EC 1.2.1.12), glutamate dehydrogenase (NAD⁺) and GroEL by *Bacteroides dorei*
and (ii) DnaK produced by *Bacteroides dorei* and *Bacteroides uniformis* was only observed in the T1D intestinal proteome. The higher incidence of *Bacteroides dorei* in the gut of Finnish children at risk or at the age of the first T1D autoantibody was recently reported [66]. The reason why Finnish T1D children are colonized with five times more *Bacteroides dorei* in comparison to Control children so far is not clear [66, 67]. GroEL and DnaK are chaperone proteins that allow the correct protein folding. These chaperones are constitutively expressed under balanced growth conditions, but when the bacterial cells are exposed to environmental stresses (e.g. high temperature, oxidative stress, bile salts), their production is upregulated. Surface location of GroEL and DnaK in intestinal bacteria and pathogenic bacteria has been reported [41, 68–71]. Moreover, their ability to bind to epithelial cells and mucins (moonlighting nature) was reported for the intestinal bacteria *Lactobacillus johnsonii, Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium animalis* [69–72]. However, as suggested by Davis-Richardson et al. [66] and Davis-Richardson and triplet [67], if *Bacteroides dorei* is implicated in the progress of the autoimmunity state, it will be through the disturbance of epithelial cells and/or interference with the development of the immune system. To what extent the higher abundance and the moonlighting character of *Bacteroides dorei* proteins contribute to the development and/or maintenance of T1D state needs further investigation.

The *Bacteroides* population in the two groups of children was estimated by real-time qPCR and no significant differences were observed (*P*=0.916). For Control children, the *Bacteroides* population reached 8.01±0.43 log_{10} cells g^{-1} faeces, and for T1D children, this population attained 8.07±0.36 log_{10} cells g^{-1} faeces. Such finding leads us to suggest that the intestinal milieu of T1D children is not impairing the *Bacteroides* population but rather is shaping its protein profile, in particular, *Bacteroides dorei* and *Bacteroides uniformis*.

In contrast to our study, Murri et al. [12] found higher numbers of *Bacteroides* in faecal samples of Spanish T1D children. Such differences may be associated with geographical location factors as has been previously reported [66, 67, 73, 74].

**Human proteins**

The procedure of protein extraction was targeted to the microbial protein; even so, it is not unexpected that some host proteins could also be observed, especially those that will be the most abundant [34]. In the intestinal proteome of Control children, the most abundant human proteins were elastase 3A, pancreatic preprotease, elastase 3B preprotein and the protein chymotrypsin C, which was found exclusively in the Control intestinal proteome (Table 1). In contrast, in T1D children, the most abundant human proteins were mucin 2 (MUC2) precursor, meprin A subunit alpha and the intestinal alkaline phosphatase (IAP) precursor (Table 1). Elastases 3A and 3B are members of the chymotrypsin-like elastase family (EC 3.4.21.70). These enzymes have serine protease activity and their content in faeces has been used as a biomarker for exocrine pancreatic insufficiency, including in diabetic children [75, 76] and adults [77–79]. Our findings using a proteomic approach showed diminished exocrine pancreatic enzymes in T1D children, which is in agreement with the reports of the lower exocrine pancreatic function in T1D children following the determination of faecal elastase 1 [75, 76].

MUC2 is the major component of the intestinal mucus layers that show a polymeric net-like structure being attached to the epithelial cells. The secreted mucus layer is a crucial barrier to infection avoiding the translocation of pathogens or commensal bacteria across the gut epithelium [80–82]. When inflammation occurs, one of the host responses is to increase the mucus production in order to eliminate the damage of the epithelial cells or beat the infection. Moreover, several bioactive factors can activate the secretion of MUC2, namely, microbes and their toxins and metabolic products, inflammatory cytokines, reactive oxygen and nitrogen species and neuropeptides [83–88]. Which factors are contributing to the higher production of MUC2 in T1D children still needs to be elucidated.

Meprins belong to the group of zinc metalloproteinases that are made up of two subunits; meprin α (MEP 1A) or meprin β (MEP 1B), which can be found in higher levels in the proximal colon [18, 89]. Interesting meprins’ abundance in the proximal colon was observed through a metaproteomic approach to evaluate the human microbial ecosystems at the mucosal luminal interface [18]. Both MEP 1A and 1B have proteolytic activity against a wide range of protein substrates. However, each subunit displays different protein and peptide bond specificities [90–92]. Low levels of mRNA transcripts of the *MEPIA* gene have been observed in inflammatory bowel disease patients, plus *Mepla* knockout mice display greater colon damage and inflammation than wild-type mice [93]. An interesting finding was observed in the study of Debyser et al. [48] in which the faecal metaproteome of 15 children suffering from cystic fibrosis and with pancreatic insufficiency was investigated. The authors found a significantly higher abundance of MUC2 precursor and meprin A subunit alpha in patients in comparison to their siblings without the disease. These findings suggest that MUC2 precursor and meprin A subunit alpha are associated with intestinal dysbiosis. IAP is known to be critical for intestinal homeostasis and health by virtue of its multiple interactions with intestinal commensal and pathogenic microbiota, diet and mucosal interface [94, 95]. An increase in IAP activity was reported in a diabetic mouse model [96]. It is possible that, in the T1D children’s intestinal milieu, different factors lead to the over-production of IAP, for example, the carbohydrate-restricted diet. Thus, the increased abundance of MUC2, MEP 1A and IAP in the intestinal environment of T1D children suggests a response of the gut cells designed to improve the health of the gastrointestinal tract. The abundance of these proteins still requires further investigation.
The contribution of the identified bacterial proteins to such changes in the human proteins in T1D children needs further investigation.

**Conclusions**

The analysis of the intestinal proteome of T1D and Control children allowed us to gain insights into how the T1D metaproteome differs in functionality from Controls. The possibility that *Bacteroides dorei* proteins are involved in the development or maintenance of the T1D state is particularly noteworthy. Strategies to improve the composition and functionality of the *Bifidobacterium* population in T1D children must be explored and their impact on the course of the disease must be evaluated. To the best of our knowledge, this is the first study that reports both the low and high changes in abundance of human intestinal proteins in T1D children that play essential roles in gut homeostasis and health. Such changes in human proteins in diabetic children require additional experiments. This study has been conducted on a small number of T1D children and further studies, integrating metagenomics and metaproteomics using a larger cohort, are crucial to gain a better insight regarding the state of intestinal dysbiosis in children with established diabetes.

Funding information

E.P. is thankful to Fundação para a Ciência e Tecnologia for the grant SFRH/BD/66954/2009. This work was partially funded through the project PEst-OE/EEB/LA0023/2011 and PEst-OE/EEB/LA0023/2013.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References


82. Johansson ME, Larsson JM, Hansson GC. The two mucus layers of colon are organized by the MUc2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci USA 2011;108:4659–4665.


Edited by: K. M. Dobos and S. V. Gordon