

Master's degree in computer science

Gut Protein Interacting with Human Body Protein Report

Foundation

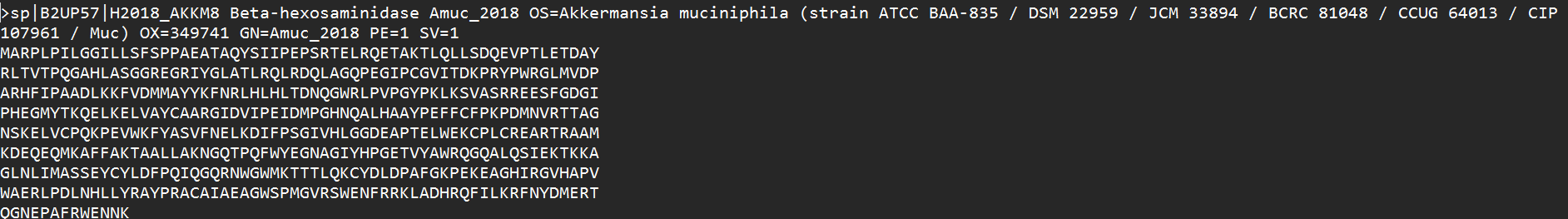
This work consisted on infering if there could be, possibly, an interaction between a protein that would be relevant for a disease (in this case Type 2 Diabetes) that interacted with a Human Body protein, thus linking these PPI with that disease. When I asked ChatGPT “what are the main organisms associated with Type 2 diabetes, preferably from the intestine”, I got the following answer:

* Microorganism considered beneficial:
  + **Akkermansia muciniphila** - Associated with the integrity of the intestinal barrier and regulation of inflammation. Reduced levels are often observed in patients with Type 2 Diabetes.
  + **Faecalibacterium prausnitzii** - A butyrate producer with anti-inflammatory properties. Low levels are commonly found in individuals with Type 2 Diabetes.
  + **Bifidobacterium spp.** - Contributes to metabolic health, fiber fermentation, and inflammation regulation.
* Microorganism potentially detrimental:
  + **Ruminococcus spp.** - Some species of this genus are associated with mucin degradation, which can compromise the intestinal barrier.
  + **Prevotella spp.** - Although some species have probiotic properties, others are linked to inflammation in populations with Type 2 Diabetes.
  + **Desulfovibrio spp.** - Sulfate-reducing bacteria associated with the production of toxic compounds like hydrogen sulfide, which can harm the intestinal mucosa.
  + **Proteobacteria (Ex.: Escherichia coli patogênica)** - Represents a pro-inflammatory group that increases under conditions of intestinal dysbiosis in Type 2 Diabetes.
* Other Microorganism of interest:
  + **Clostridium spp. (Certain groups)** - Some subgroups of Clostridium are butyrate producers and have positive effects, while others are associated with inflammation and insulin resistance.

Micro-organisms Proteome acquisition

The proteome from each of the following microorganisms was obtained, mainly from UniProt. As an example I leave the link for the Proteome acquisition of Akkermansia muciniphila <https://www.uniprot.org/proteomes/UP000001031>.

The data was downloaded in a fasta format (the predefined format) and the obtained file was comprised of the following information. It is important to note that this information is, of course, microorganism deppendant, but the first protein in the proteome of the Akkermansia muciniphila can be found bellow.



Dataset Breakdown

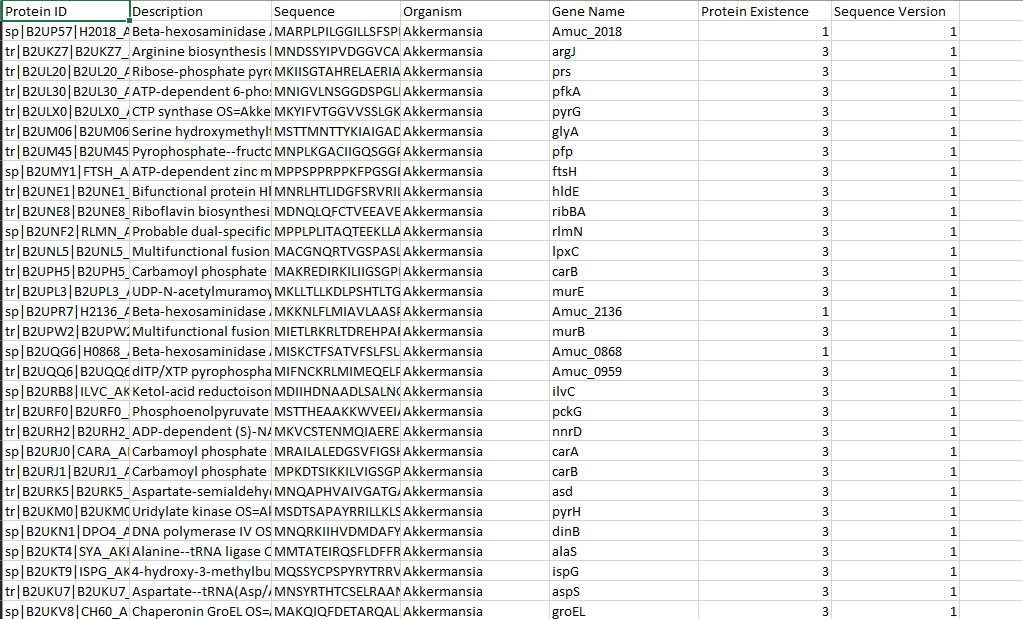
* **sp|B2UP57**: This indicates a UniProt protein entry (sp stands for Swiss-Prot), after that the Protein ID can be found.
* **H2018\_AKKM8 Beta-hexosaminidase**: This is the name of the protein, in this case, **Beta-hexosaminidase**.
* **Amuc\_2018**: This is the gene name associated with the protein in this organism.
* **OS=Akkermansia muciniphila**: The organism’s name.
* **OX=349741**: The NCBI taxonomy identifier for Akkermansia muciniphila.
* **PE=1**: Protein existence evidence level, with **1** indicating strong evidence (e.g., experimental).
* **SV=1**: Sequence version, showing the version number of this protein sequence.
* **MARPPLPILGGILLSSFSPPAEATAQYSIIPPSRTELRQETAKT...:** Amino Acid sequence of the protein

Dataset Filtering

Knowing the objective of the work, an Interactome was used. Interactomes contain a database of PPI thus enabling the comparison. Interactome 3D from <https://interactome3d.irbbarcelona.org/> was used.

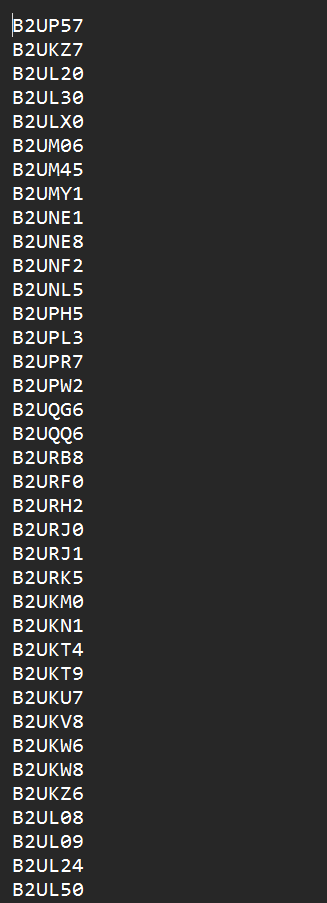
For this, all the info needed to be filtered and processed. The data for this step was mostly kept intact but some of the entries were excluded. The entries, proteins, that had PE, the protein existence higher than 3 were removed (PE = 4, PE = 5). A protein level than 3 would be higher than 3 is purely prediction or blatantly uncertain thus not matching the needed criteria.

|  |  |  |
| --- | --- | --- |
| PE Level | Description | Evidence Type |
| PE=1 | Experimental evidence at the protein level | Protein detected by direct experiments |
| PE=2 | Experimental evidence at the transcript level | mRNA detected, but no direct protein evidence |
| PE=3 | Inferred from homology | Protein predicted based on related proteins |
| PE=4 | Predicted | Computational prediction only |
| PE=5 | Uncertain | The protein is a dubious prediction |

The resulting proteins were then converted into EXCEL (for easier analysis and display) and the result was the following.

Finally, the data was retrieved from the EXCEL, removed all excess data and just kept the Protein IDS (only thing necessary to input in the Interactome). All the protein IDS from all proteomes were stored in a single doc file called “LISTA”. The main reason for this choice is simply for ease of use, to have all the protein IDS condensed into a single place.

Some of the entries for the Lista file are the following:

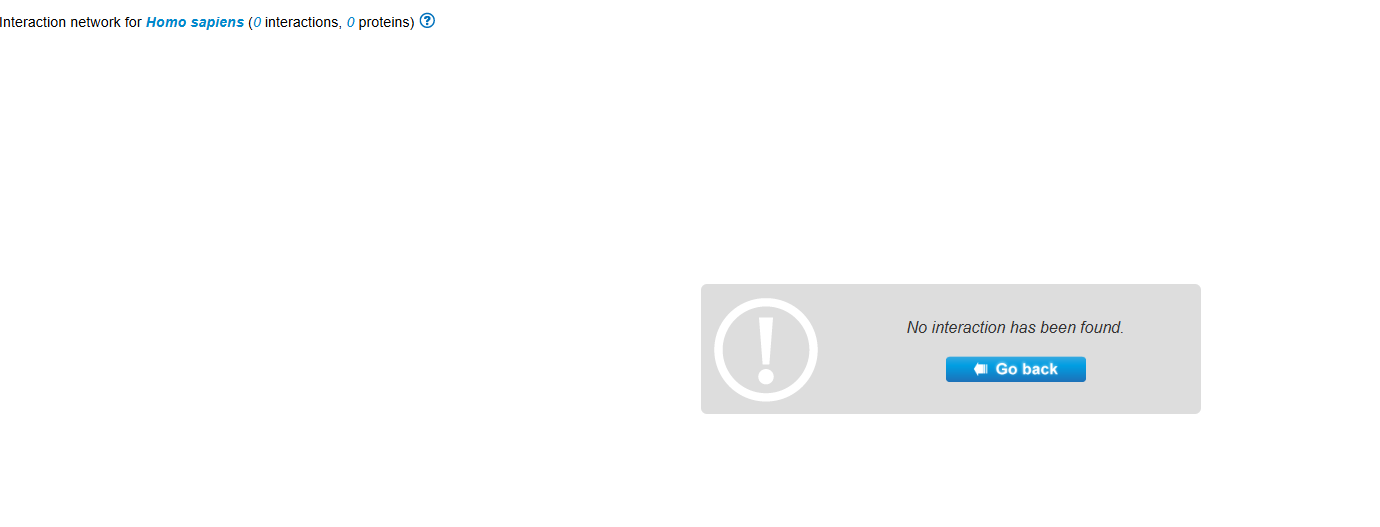


The LISTA file, at the end had over 6500 protein IDS (6676 to be exact) that would be inputted in the referred website.

Results

All the protein IDS were inserted into the proteome. At first, an error was received because of a mal format that was inserted into the Ruminococcus spp. Protein IDs. The website couldn’t retrieve the protein IDS or they simply weren’t a match to any existing protein ID’s. After careful consideration and after taking a look at the actual dataset, it was determined that the protein IDS from Ruminococcus spp were actually valid and that, perhaps, there was simply just not a match at all and thus the data from that microorganism was simply removed.

All the other data was then inputted into the website and the result was:



In the end, there were no interactions between all the protein Ids and the human body.

Result Analysis

There could have been many errors that happened in each of the following steps:

* From the website side (not filtering/having) all the interactions necessary;
* From the proteome info (if the info was not reliable/incomplete) we could be missing vital data and possibly some interactions;

In the end, the only conclusion that can be drawn is that, for the introduced data in this specific website, there were NO RECORDED interactions between those protein IDS and the Human Body.