



## Database Tool

# MAHMI database: a comprehensive MetaHit-based resource for the study of the mechanism of action of the human microbiota

Aitor Blanco-Míguez<sup>1,2</sup>, Alberto Gutiérrez-Jácome<sup>1</sup>,  
Florentino Fdez-Riverola<sup>1</sup>, Anália Lourenço<sup>1,3,\*</sup> and Borja Sánchez<sup>2\*</sup>

<sup>1</sup>ESEI – Department of Computer Science, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n 32004, Ourense, Spain, <sup>2</sup>Department of Microbiology and Biochemistry of Dairy Products, Instituto de Productos Lácteos de Asturias (IPLA), Consejo Superior de Investigaciones Científicas (CSIC), Villaviciosa, Asturias, Spain and <sup>3</sup>CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

\*Corresponding author: Email: [analia@uvigo.es](mailto:analia@uvigo.es); Correspondence may also be addressed to Borja Sánchez. Email: [borja.sanchez@csic.es](mailto:borja.sanchez@csic.es)

Citation details: Blanco-Míguez,A., Gutiérrez-Jácome,A. Fdez-Riverola,F. *et al.* Large-scale inference of gene function through phylogenetic annotation of gene ontology terms: case study of the apoptosis and autophagy cellular processes. *Database* (2016) Vol. 2016: article ID baw157; doi:10.1093/database/baw157.

Received 4 May 2016; Revised 4 November 2016; Accepted 5 November 2016

## Abstract

The Mechanism of Action of the Human Microbiome (MAHMI) database is a unique resource that provides comprehensive information about the sequence of potential immunomodulatory and antiproliferative peptides encrypted in the proteins produced by the human gut microbiota. Currently, MAHMI database contains over 300 hundred million peptide entries, with detailed information about peptide sequence, sources and potential bioactivity. The reference peptide data section is curated manually by domain experts. The *in silico* peptide data section is populated automatically through the systematic processing of publicly available exoproteomes of the human microbiome. Bioactivity prediction is based on the global alignment of the automatically processed peptides with experimentally validated immunomodulatory and antiproliferative peptides, in the reference section. MAHMI provides researchers with a comparative tool for inspecting the potential immunomodulatory or antiproliferative bioactivity of new amino acid sequences and identifying promising peptides to be further investigated. Moreover, researchers are welcome to submit new experimental evidence on peptide bioactivity, namely, empiric and structural data, as a proactive, expert means to keep the database updated and improve the implemented bioactivity prediction method. Bioactive peptides identified by MAHMI have a huge biotechnological potential, including the manipulation of aberrant immune responses and the design of new functional ingredients/foods based on the genetic sequences of the human microbiome. Hopefully, the resources provided by MAHMI will be useful to those researching gastrointestinal disorders of autoimmune and inflammatory

nature, such as Inflammatory Bowel Diseases. MAHMI database is routinely updated and is available free of charge.

**Database URL:** <http://mahmi.org/>

## Introduction

Human beings live in a homeostatic symbiosis with their gastrointestinal microbes (1). The host provides microbiota with nutrients and a friendly environment, whereas the microbiota helps in the development of the gut mucosa, including the immune system, and provides the host with nutritional contributions (2). Different functions have been assigned to human gut microbiota, including maintenance of the epithelial barrier, direct or indirect inhibition of pathogen adhesion to intestinal surfaces, modulation and proper maturation of the immune system, degradation of otherwise non-digestible carbon sources (such as plant polysaccharides), and production of different metabolites (such as vitamins and short-chain fatty acids), among which butyric acid is of paramount importance for colonocyte physiology (3). Despite the growing number of solid scientific evidences showing that our intestinal microbiota is implicated in maintaining the gut physiology, we are far from reaching a clear understanding about the molecular mechanisms underlying this mutualism and how alterations in the relative amounts of microbiota populations are related to human health (4).

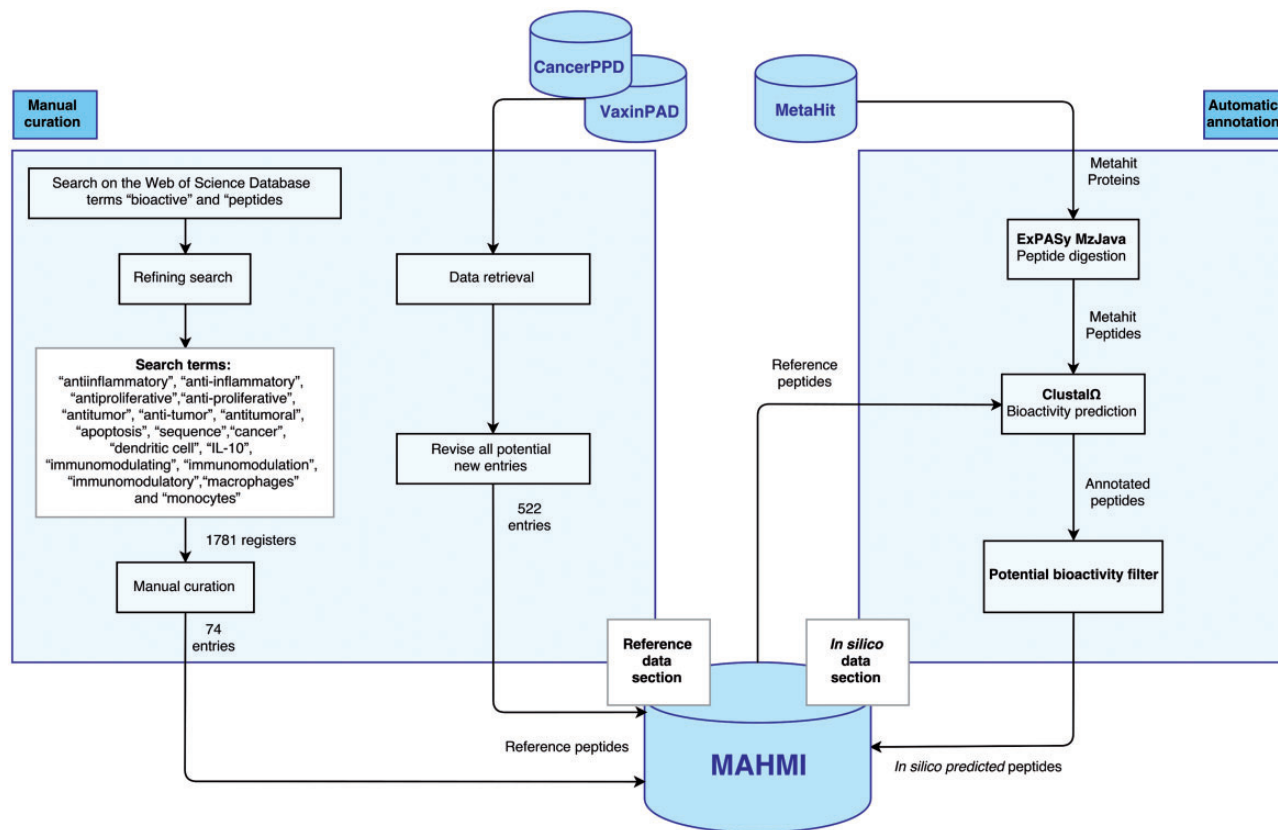
In order to maintain intestinal homeostasis, host bacteria and human immune system are constantly engaged in sensing and reacting to one another by means of specific molecules such as extracellular proteins (5). The human gastrointestinal tract (GIT), and more concisely the gut-associated lymphoid tissue (GALT), is perhaps the main point of contact between gut microorganisms and the human host. Peyer's patches and lymphoid/plasma cells are important GALT structures in the lower parts of the GIT, and whereas the first have precise localizations, the second are diffusely distributed throughout the gut mucosa (6). Interaction between microbiota and host, also referred to as molecular cross-talking, is elicited by a series of relatively well-conserved molecules known as microbial-associated molecular patterns (MAMPs). MAMPs interact with specific receptors on the surface of host cells named pattern-recognition receptors (PRRs), driving the normal interaction occurring in a homeostatic situation and also some of the beneficial effects attributed to probiotics (7). The interaction of MAMPs with PRRs also affects the maturation of antigen-presentation cells, which will

subsequently define the type of T-cell immune response (mainly Th1, Th2, Th17 or Treg) (8).

Several non-infectious human diseases, including autoimmune, inflammatory diseases and certain types of cancer, are related to both an immunological imbalance and a dysbiosis of the gut microbiota (9–12). Possibly, the paradigm of the link between gut microbiota and immunity are the Inflammatory Bowel Diseases (IBD), in which an aberrant immune response to the gut microbiota is related to development, persistence and relapse of the disease (13, 14). Indeed, faecal microbiota transplantation leads to complete remission in about 40% of the cases, even though there is a relative high heterogeneity among patients (15). In this regard, some species have been found to be underrepresented in the faecal microbiota of patients suffering from IBD, such as the butyrate producing bacteria *Faecalibacterium prausnitzii*, which has been postulated as a biomarker or even as a potential therapeutic agent (16, 17). However, it should be taken into account that several studies point to the involvement of genetic predisposition due to IBD-promoting gene mutations (18, 19).

This work introduces the Mechanism of Action of the Human Microbiome (MAHMI) database, a novel and comprehensive resource containing peptides extracted from the human microbiome complemented with potential immunomodulatory or antiproliferative bioactivity. MAHMI has two data sections: the reference data section that contains peptides whose bioactivity has been experimentally tested and the *in silico* data section, which keeps peptides encrypted in larger bacterial proteins that exhibit sequence similarity with the peptides contained in the reference data section, and which might be potentially bioactive. Our bioactivity prediction method encompasses *in silico* protein digestion using intestinal proteases from protein microbiome complements and global sequence similarity search against the reference peptides.

The overall aim of MAHMI is to facilitate the work of those scientific teams working in the field of the molecular interaction between the human host and the intestinal microbiota. Most notably, MAHMI Web-based interface enables the comparison of user-submitted amino acid sequences against database contents and thus, it can be ideally used to identify peptides that may be used to manipulate altered immune responses. MAHMI has a clear



**Figure 1.** Data sections and curation workflow. The curation workflow is divided into two processes: manual curation and automatic annotation.

application in the framework of IBD, but also in other immune-associated diseases where the gut microbiota plays a part, including Irritable Bowel Syndrome (20, 21), Celiac Disease (22, 23) and Type-1 Diabetes (24, 25).

## Database construction and update

### Database overview

The system integrates a Web-based front-end, a Java application as processing core (version 1.8.0\_66) and a MySQL data management system (version 14.14). The front-end is a PHP application (version 5.5.9) with a Bootstrap interface. The MAHMI site is best viewed by Google Chrome 46.0, Firefox 45.0.1, Opera 36.0 and Safari browsers.

### Data sources: reference and *in silico* sections

The MAHMI database consists of two data sections (Figure 1). The *reference* section (left side) is manually annotated and reviewed, containing information extracted from the scientific literature and other public, manually curated databases about experimentally validated immunomodulatory peptides. The *in silico* data section (right side) is automatically generated and annotated based on the computational analysis of the exoproteomes of the

human microbiome and the identification of peptides with immunomodulatory potential.

The main sources of the reference section are scientific literature, the VaxinPAD server of peptide-based vaccine adjuvants (26) and the CancerPPD database on antiproliferative peptides and proteins (27). Although retrieval and annotation of some of these data is automated at some extent, this curation workflow is primarily manual. Most notably, scientific literature is retrieved from the Web of Science Database (23/05/2015) and manually curated by experts. Search starts by using the terms 'bioactive' and 'peptides', and is further refined using terms such as 'anti-inflammatory', 'anti-inflammatory', 'antiproliferative', 'anti-proliferative', 'antitumor', 'anti-tumor', 'antitumoral', 'apoptosis', 'sequence', 'cancer', 'dendritic cell', 'IL-10', 'immunomodulating', 'immunomodulation', 'immunomodulatory', 'macrophages' and 'monocytes'. The first literature retrieval rendered a total of 1781 records. Works reporting the amino acid sequence of an antiproliferative or immunomodulatory peptide were kept (74 records). The contents of VaxinPAD and CancerPPD databases are automatically retrieved, but our curators revise all new entries before insertion into MAHMI database. Moreover, all peptide entries in MAMHI maintain cross-links to the original sources (i.e. database identifiers and PubMed identifiers, in the case of scientific literature).

Complementarily, the MetaHIT project (<http://www.metahit.eu/>) is the primary data source of the *in silico* data section (28). Specifically, protein data downloadable via the GigaDB FTP site ([ftp://climb.genomics.cn/pub/10.5524/100001\\_101000/100064/1.GeneCatalogs/](ftp://climb.genomics.cn/pub/10.5524/100001_101000/100064/1.GeneCatalogs/)). These proteins are processed and the peptides that show immunomodulatory or anti-inflammatory potential are inserted in MAHMI database. Noteworthy, this data compilation and screening process runs in continuous mode and MAHMI database is routinely updated on a weekly basis.

### *In silico* bioactivity prediction

This module is implemented as a Java 8 application that uses a REST API to communicate with the rest of the system. Our REST API ensures the modular design of the system, such that each component performs a specific/limited scope functionality and hence, individual components may be interchanged as needed without affecting the rest of the system. Notably, communication is based on the interchange of XML and JSON objects. The data workflow is as follows.

#### Protein digestion

This module is responsible for the digestion of the proteins downloaded from MetaHIT. *In silico* digestion is performed using the ExPASy mzJava (<http://mzjava.expasy.org>) library (29) and considering the major human intestinal endoproteases, i.e. trypsin, chymotrypsin and pepsin (pH < 2). The resulting collections of peptides, denominated peptidomes, are stored in the MySQL data repository.

#### Bioactivity prediction

All peptides generated by enzyme digestion are processed to predict any possible bioactivity. The MAHMI module responsible for this task is built on top of ClustalΩ, the latest and more scalable version of the well-known Clustal multiple sequence alignment tool (30). Peptides are aligned against the peptides of the reference data section (described in the *Data Sources* section).

For the first version of the database and in an attempt to give a first bioactivity index, a peptide was considered as potentially bioactive if its amino acid similarity with experimentally validated peptides was >60% as returned by the ClustalΩ algorithm.

#### Query capabilities

The graphical user interface (GUI) of MAHMI database is available at <http://mahmi.org>. This is a PHP application with a dynamic and user-friendly Bootstrap interface, which expedites the use of HTML5, CSS3 and JQuery technologies in interface development (<http://getbootstrap.com/>).

#### Quick search

By selecting this option, the user can compare amino acidic sequences against the two data sections of MAHMI. The list of results is produced based on the similarity scores calculated by the NCBI BLAST+ tool BLASTP alignment (31).

#### Advanced search

The configuration of advanced search options (Figure 2) allows the user to select a specific data section and use search facets, such as the bioactivity potential threshold (i.e. predicted peptide should have a similarity equal or greater than the specified threshold) or the bioactivity type (i.e. immunomodulatory, antiproliferative or both). These filters are applied over BLASTP alignment results.

#### Results page

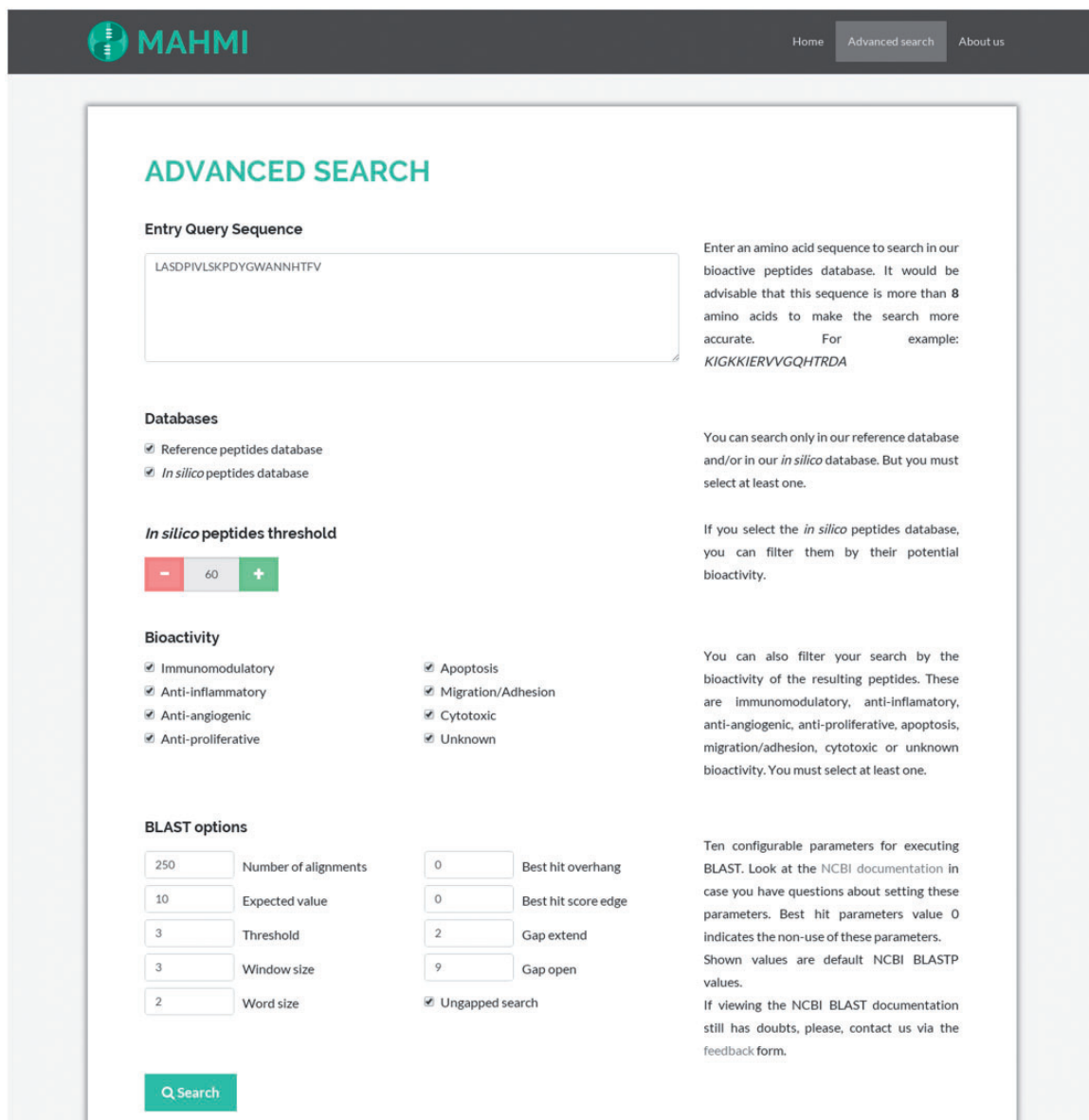
MAHMI Web interface enables the graphical interpretation of all alignments of the user sequence to database records. This interpretation is enabled by two visual and interactive components: (1) a HTML object that presents the distribution of BLAST hits on the query sequence, indicating the extent and region of the alignments; (2) a table that describes the main high-scoring pair attributes of all the hits (32). These two visual components have been built using BlasterJS, an in-house developed and publicly available JavaScript library for displaying BLAST results (<http://sing.ei.uvigo.es/blasterjs>).

As illustrated in Figure 3, the hits distribution component summarises all search results. The visual appearance and interpretation of this component is similar to that of NCBI BLAST service, which is quite familiar to most domain users (<http://blast.ncbi.nlm.nih.gov>). Exception being that in our visual component hits coming from the reference data section will be presented first. Hits are coloured by score (full-colour and grey-scale views are available).

By clicking on one hit, the user views the corresponding hit in the alignment table. As illustrated in Figure 4, the alignment hits table gathers together all BLAST hits and their attributes: bit score, expected value (*e*-value), identity, positives and gaps (32). Record ordering is two-fold: first, entries are sorted by data section provenance, i.e. reference entries show first than *in silico* predicted entries; and then, results are sorted by BLAST similarity score.

#### Hit alignment page

By clicking on an alignment description on the alignment hits table (Figure 4), the system shows an in-depth visual representation (amino acid by amino acid) of the alignment (Figure 5). Moreover, the user has access to additional information about the matching peptide, such as sequence length, bioactivity potential (potential or experimentally validated), molecular weight and isoelectric point. In



**Figure 2.** Snapshot of the advanced search. The query sequence is 'LASDPIVLSKPDYGWANNHTFV' and both data sections are searched, with the following custom BLAST options: threshold set to 3, window size set to 3, word size set to 2, gap extend set to 2, gap open set to 9 and ungapped search selected.

addition to these data, MAHMI records are compliant with the Minimum information about any (x) sequence (MIxS) checklist for metagenomes (34). Whenever available, source proteins are cross-linked to external database records, namely, UniProt (33), CancerPPD (27), VaxinPAD (26) and Metahit metagenome records.

### Results download

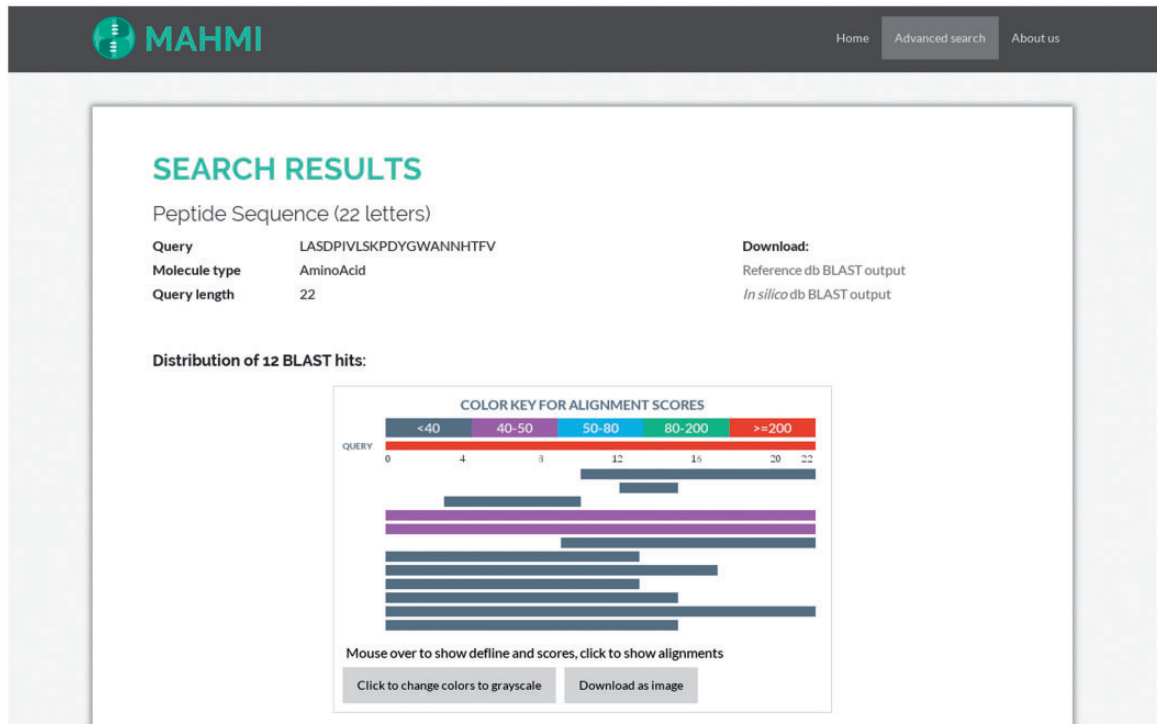
All search results, namely the BLASTP alignments, the peptide sequences and all figures can be downloaded. Specifically, the user can obtain the BLASTP output file

(default format) as well as the hits distribution component as image (PNG format). The alignment table can be exported as image (PNG format) or plain text file (CSV format). On the alignment information page, the detailed alignment figure can be downloaded as PNG image. The user can also download additional peptide information, namely, sequence and possible sources, whenever available.

### Public REST API

In addition to the graphical user interface, MAHMI provides a public REST API to download all data relating to





**Figure 3.** Snapshot of the hit distribution of results for a query search. The query sequence is 'LASDPVLSKPDYGWANNHTFV' and both data sections are searched, with the following custom BLAST options: threshold set to 3, window size set to 3, word size set to 2, gap extend set to 2, gap open set to 9 and ungapped search selected. Returned hits are presented in the same order as they appear in the BLAST file. Also, hits are coloured based on the assigned bit score.

**Sequences producing significant alignments:**

Description	Score	E value	Identities	Positives	Gaps
MHM_REF451 reference IMM LZ-8, immunomodulatory	15.0	1.6	33%	67%	0%
MHM_REF282 reference ANT Anti-inflammatory	13.8	2.2	100%	100%	0%
MHM_REF435 reference UNK TNF-related apoptosis-inducing ligand (TRAIL), unknown, antitumor	14.6	2.3	71%	86%	0%
MHM_POS23370840 possible ANT 75%MHM_REF119	53.2	7.0E-11	100%	100%	0%
MHM_POS22797112 possible IMM 66.666667%MHM_REF590	51.2	4.0E-10	91%	100%	0%
MHM_POS29544994 possible IMM 66.666667%MHM_REF590	33.1	0.004	85%	92%	0%
MHM_POS23374042 possible ANT 60%MHM_REF83	30.7	0.03	100%	100%	0%
MHM_POS22574314 possible IMM 66.666667%MHM_REF590	28.7	0.16	53%	82%	0%
MHM_POS22803732 possible ANT 60%MHM_REF83	28.7	0.17	85%	100%	0%
MHM_POS30291642 possible IMM 66.666667%MHM_REF590	23.8	8.3	47%	87%	0%
MHM_POS23472559 possible ANT 66.666667%MHM_REF148	23.8	8.3	45%	59%	0%
MHM_POS21188442 possible IMM 80%MHM_REF544	23.8	8.6	47%	67%	0%

Download as csv    Download as image

**Figure 4.** Snapshot of the table detailing the hits of a query search. This table gathers together all BLAST hits and their attributes (bit score, expected value, identities, positives and gaps). Clicking on one hit, it redirects the user to the alignment information of this hit.


**MAHMI**  
Last update: 21/09/2016

[Home](#)
[Explore peptides](#)
[Data submission](#)
[Advanced search](#)
[Contact us](#)

## RESULTING ALIGNMENT

Peptide Sequence (34 letters)

**Description:**  
 MHM\_POS23370840 possible ANT 75%MHM\_REF119 [Download peptide sequence](#)

**Properties:**

Score	E value	Identities	Positives	Gaps	Length	Potential bioactivity	Molecular Weight	Isoelectric point
53.2	7.0E-11	100%	100%	0%	34	Anti-inflammatory	3747.13	4.13

**Potential bioactivity:**  
 Potential *anti-inflammatory*. 75% similarity with reference peptide MHM\_REF119.

**Source proteins:** [Download source proteins fasta](#)

Protein *MH0409\_GL0141390*

- Uniprot entry [P82594](#)
  - Organism: *Streptomyces chartreusis*
  - Gene: *N/A*
  - Protein: *Extracellular exo-alpha-(1->5)-L-arabinofuranosidase*
- Metagenome *MH0409* MixS data:
  - Investigation type: *metagenome*
  - Project name: *MetaHIT-Richness*
  - Collection date: *2008-2010*
  - Environmental package: *human-gut*
  - Geographic location: *Denmark*
  - Environment biome: *human gut*
  - Environment feature: *human-associated habitat*
  - Environment material: *faeces*
  - Sequencing method: *Illumina*
  - Geographic location (latitude): *55.676097 DD*
  - Geographic location (longitude): *12.568337 DD*
- Additional metagenome information:
  - Gene length: *1800*
  - Gene completeness: *Complete*
  - Gene origin: *EUR*
  - Species annotation phylum: *unknown*
  - Species annotation genus: *unknown*
  - Kegg annotation phylum: *K01209*
  - Sample occurrence frequency: *0.0015785319652723*
  - Individual occurrence frequency: *0.00186915887850467*
  - Kegg functional category: *Carbohydrate Metabolism*

**Alignment:**

Query	1	<span style="background-color: #90EE90; padding: 2px;">L</span> <span style="background-color: #FFB6C1; padding: 2px;">A</span> <span style="background-color: #90EE90; padding: 2px;">S</span> <span style="background-color: #FFB6C1; padding: 2px;">D</span> <span style="background-color: #90EE90; padding: 2px;">P</span> <span style="background-color: #FFB6C1; padding: 2px;">I</span> <span style="background-color: #90EE90; padding: 2px;">V</span> <span style="background-color: #FFB6C1; padding: 2px;">L</span> <span style="background-color: #90EE90; padding: 2px;">L</span> <span style="background-color: #FFB6C1; padding: 2px;">S</span> <span style="background-color: #90EE90; padding: 2px;">K</span> <span style="background-color: #FFB6C1; padding: 2px;">P</span> <span style="background-color: #90EE90; padding: 2px;">D</span> <span style="background-color: #FFB6C1; padding: 2px;">Y</span> <span style="background-color: #90EE90; padding: 2px;">G</span> <span style="background-color: #FFB6C1; padding: 2px;">W</span> <span style="background-color: #90EE90; padding: 2px;">A</span> <span style="background-color: #FFB6C1; padding: 2px;">N</span> <span style="background-color: #90EE90; padding: 2px;">N</span> <span style="background-color: #FFB6C1; padding: 2px;">H</span> <span style="background-color: #90EE90; padding: 2px;">T</span> <span style="background-color: #FFB6C1; padding: 2px;">F</span> <span style="background-color: #90EE90; padding: 2px;">V</span>	22
Subject	1	<span style="background-color: #90EE90; padding: 2px;">L</span> <span style="background-color: #FFB6C1; padding: 2px;">A</span> <span style="background-color: #90EE90; padding: 2px;">S</span> <span style="background-color: #FFB6C1; padding: 2px;">D</span> <span style="background-color: #90EE90; padding: 2px;">P</span> <span style="background-color: #FFB6C1; padding: 2px;">I</span> <span style="background-color: #90EE90; padding: 2px;">V</span> <span style="background-color: #FFB6C1; padding: 2px;">L</span> <span style="background-color: #90EE90; padding: 2px;">L</span> <span style="background-color: #FFB6C1; padding: 2px;">S</span> <span style="background-color: #90EE90; padding: 2px;">K</span> <span style="background-color: #FFB6C1; padding: 2px;">P</span> <span style="background-color: #90EE90; padding: 2px;">D</span> <span style="background-color: #FFB6C1; padding: 2px;">Y</span> <span style="background-color: #90EE90; padding: 2px;">G</span> <span style="background-color: #FFB6C1; padding: 2px;">W</span> <span style="background-color: #90EE90; padding: 2px;">A</span> <span style="background-color: #FFB6C1; padding: 2px;">N</span> <span style="background-color: #90EE90; padding: 2px;">N</span> <span style="background-color: #FFB6C1; padding: 2px;">H</span> <span style="background-color: #90EE90; padding: 2px;">T</span> <span style="background-color: #FFB6C1; padding: 2px;">F</span> <span style="background-color: #90EE90; padding: 2px;">V</span>	22

[Download as image](#)  
Back

**Figure 5.** Snapshot of the hit alignment information. Properties table includes the high-scoring pair attributes of the hit, bioactivity (potential bioactivity, in the case of *in silico* predictions), molecular weight and isoelectric point. Source proteins are cross-linked to UniProt and Metahit metagenome entries. The alignment figure compares in more depth (amino acid by amino acid) the alignment of the proposed hit with the query sequence.

proteins and peptides in the database. The main methods provided by this API are listed below:

- GET/peptides/reference, which returns the reference peptides;
- GET/peptides that returns the bioactive peptides;
- POST/peptides/search + search XML object (see details in [Supplementary material 1](#)), which enables batch searches, similar to those that can be performed on the web;
- GET/peptides/sourceProteins/\$peptide\_id, which returns information about a given peptide, namely the list of source proteins (cross-referenced to Uniprot) and meta-genome information, namely MIXS compliant data;
- GET/proteins/\$peptide\_id, which returns the proteins that produce a specific peptide.

This API is publicly available at <http://mahmi.org/api>.

### Database utility

To the best of our knowledge, and with the exception of VaxinPAD and CancerPPD, no effort has been conducted to extensively collect data on peptides encrypted in the protein complement of the human microbiome with immunomodulatory and antiproliferative potential bioactivities. This prompted us to create MAHMI, an online and accessible resource containing massive information on potential bioactive peptides. Sequences contained within MAHMI database may help the basic researcher working in the field of bacteria-human host molecular cross-talk, as this kind of peptides are relevant from a physiological and immune point of view.

Proteins produced by our intestinal microbiota are important molecular targets that can allow the study of the mechanism of action of our gut microbes (35). Notably, surface proteins are key extracellular components in the mediation of certain interactions, such as immunomodulation or mucosal colonization (5). A subset of these proteins, the secreted proteins, may even be able to interact directly with mucosal cells once released by the microorganism to their surrounding environment (36). Extracellular proteins from the gut microbiota are either cleaved by GIT proteases or processed by professional antigen presenting cells such as dendritic cells (37). This interaction mediated by peptides produced by the digestion of our gut microbe proteins ends in physiological changes of the cell (38), including signalling pathway regulation, changes in the secretion of chemokins, cytokines and anti-bactericidal peptides (defensins), mucus secretion, production of pseudopods by dendritic cells (DCs),

rearrangement of the tight-junctions in epithelial cells, and modulation of the immune function and the response of the gut-associated lymphoid tissue (GALT) cells (39).

This is a very promising field and, for instance, recent results have revealed that a peptide encrypted in a *Lactobacillus plantarum* extracellular protein, more concisely an aggregation factor, is able to modulate the immune status of both intestinal and monocyte-differentiated DCs, including cytokine secretion and toll-like receptors (TLR) expression (40, 41). By querying the content of MAHMI database, the researcher will easily have insights about the potential bioactivity of the user-submitted amino acid sequence.

### Conclusions

MAHMI database offers a novel and comprehensive repository about peptides with immunomodulatory and anti-proliferative potential. MAHMI has the potential to aid in the study of how to balance altered immune responses and, ideally, in the innovative development of new functional ingredients or drugs based on bioactive peptides obtained from the genetic sequences of the human microbiome. These may be directed to the prevention of several gastrointestinal disorders of autoimmune and inflammatory nature as well as the minimisation of the effects of the disease in patients.

### Availability and requirement

The Mechanism of Action of the Human Microbiome (MAHMI) database is freely accessible for research purposes for non-profit and academic organisations at <http://mahmi.org/>.

### Supplementary data

Supplementary data are available at *Database Online*.

### Acknowledgements

Borja Sánchez was recipient of a Ramón y Cajal postdoctoral contract from the Spanish Ministry of Economy and Competitiveness.

### Funding

This work was funded by the Grant AGL2013-44039-R from the Spanish “Plan Estatal de I+D+I”, and the Grant EM2014/046 from the “Plan Galego de investigación, innovación e crecemento 2011-2015”. Borja Sánchez was recipient of a Ramón y Cajal



postdoctoral contract from the Spanish Ministry of Economy and Competitiveness.

This work was also partially funded by the [14VI05] Contract-Programme from the University of Vigo and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273) and the European Union's Seventh Framework Programme FP7/REGPOT-2012-2013.1 under grant agreement n° 316265, BIOCAPS. This document reflects only the author's views and the European Union is not liable for any use that may be made of the information contained herein.

*Conflict of interest.* None declared.

## References

- Hooper,L.V. and Gordon,J.I. (2001) Commensal host-bacterial relationships in the gut. *Science*, 292, 1115–1118.
- Eckburg,P.B., Bik,E.M., Bernstein,C.N. *et al.* (2005) Diversity of the human intestinal microbial flora. *Science*, 308, 1635–1638.
- Gevers,D., Kugathasan,S., Denson,L.A. *et al.* (2014) The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe*, 15, 382–392.
- Sommer,F., and Bäckhed,F. (2013) The gut microbiota—masters of host development and physiology. *Nat. Rev. Microbiol.*, 11, 227–238.
- Sánchez,B., Urdaci,M.C., and Margolles,A. (2010) Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa-bacteria interactions. *Microbiology*, 156, 3232–3242.
- Kamada,N., Seo,S.U., Chen,G.Y. and Núñez,G. (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nat. Rev. Immunol.*, 13, 321–335.
- Bouskra,D., Brézillon,C., Bérard,M. *et al.* (2008) Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature*, 456, 507–510.
- Maloy,K.J. and Powrie,F. (2011) Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature*, 474, 298–306.
- Bull,M.J. and Plummer,N.T. (2014) Part 1: The human gut microbiome in health and disease. *Integr. Med. (Encinitas)*, 13, 17–22.
- Singh,B., Qin,N. and Reid,G. (2016) Microbiome regulation of autoimmune, gut and liver associated diseases. *Inflamm. Allergy Drug Targets*, 14, 84–93.
- Hunt,R.H., Camilleri,M., Crowe,S.E. *et al.* (2015) The stomach in health and disease. *Gut*, 64, 1650–1668.
- Wu,H., Tremaroli,V. and Bäckhed,F. (2015) Linking microbiota to human diseases: a systems biology perspective. *Trends Endocrinol. Metab.*, 26, 758–770.
- Buttó,L.F. and Haller,D. (2016) Dysbiosis in intestinal inflammation: cause or consequence. *Int. J. Med. Microbiol.*, 306, 302–309.
- Kataoka,K. (2016) The intestinal microbiota and its role in human health and disease. *J. Med. Invest.*, 63, 27–37.
- Colman,R.J. and Rubin,D.T. (2014) Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. *J. Crohns. Colitis*, 8, 1569–1581.
- Willing,B., Halfvarson,J., Dicksved,J. *et al.* (2009) Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm. Bowel Dis.*, 15, 653–660.
- Maslowski,K.M., Vieira,A.T., Ng,A. *et al.* (2009) Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature*, 461, 1282–1286.
- Basso,P.J., Fonseca,M.T.C., Bonfá,G. *et al.* (2014) Association among genetic predisposition, gut microbiota, and host immune response in the etiopathogenesis of inflammatory bowel disease. *Brazilian J. Med. Biol. Res.*, 47, 727–737.
- Xu,X.M. and Zhang,H.J. (2016) miRNAs as new molecular insights into inflammatory bowel disease: crucial regulators in autoimmunity and inflammation. *World J. Gastroenterol.*, 22, 2206–2218.
- Distrutti,E., Monaldi,L., Ricci,P., and Fiorucci,S. (2016) Gut microbiota role in irritable bowel syndrome: new therapeutic strategies. *World J. Gastroenterol.*, 22, 2219–2241.
- Hornby,P.J. (2015) Drug discovery approaches to irritable bowel syndrome. *Expert Opin. Drug Discov.*, 10, 809–824.
- Kyriacou,A., McLaughlin,J. and Syed,A.A. (2015) Thyroid disorders and gastrointestinal and liver dysfunction: a state of the art review. *Eur. J. Intern. Med.*, 26, 563–571.
- Korponay-Szabó,I.R., Troncone,R. and Discepolo,V. (2015) Adaptive diagnosis of coeliac disease. *Best Pract. Res. Clin. Gastroenterol.*, 29, 381–398.
- Paun,A., Yau,C. and Danska,J.S. (2016) Immune recognition and response to the intestinal microbiome in type 1 diabetes. *J. Autoimmun.*, 71, 10–18.
- Knip,M. and Siljander,H. (2016) The role of the intestinal microbiota in type 1 diabetes mellitus. *Nat. Rev. Endocrinol.*, 12, 154–167.
- Nagpal,G., Gupta,S., Chaudhary,K. *et al.* (2015) VaccineDA: Prediction, design and genome-wide screening of oligodeoxynucleotide-based vaccine adjuvants. *Sci. Rep.*, 5, 12478.
- Tyagi,A., Tuknait,A., Anand,P. *et al.* (2015) CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Res.*, 43, D837–D843.
- Qin,J., Li,R., Raes,J. *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464, 59–65.
- Horlacher,O., Nikitin,F., Alocci,D. *et al.* (2015) MzJava: an open source library for mass spectrometry data processing. *J. Proteomics*, 129, 63–67.
- Sievers,F., Wilm,A., Dineen,D. *et al.* (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7, 539–539.
- Camacho,C., Coulouris,G., Avagyan,V. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421.
- Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.
- Bairoch,A. (2004) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 33, D154–D159.
- Yilmaz,P., Kottmann,R., Field,D. *et al.* (2011) Minimum information about a marker gene sequence (MIMARKS) and

- minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.*, 29, 415–420.
35. Sánchez,B., González-Tejedo,C., Ruas-Madiedo,P. *et al.* (2011) *Lactobacillus plantarum* extracellular chitin-binding protein and its role in the interaction between chitin, caco-2 cells, and mucin. *Appl. Environ. Microbiol.*, 77, 1123–1126.
36. Ruiz,L., Hevia,A., Bernardo,D. *et al.* (2014) Extracellular molecular effectors mediating probiotic attributes. *FEMS Microbiol. Lett.*, 359, 1–11.
37. Swiecki,M. and Colonna,M. (2015) The multifaceted biology of plasmacytoid dendritic cells. *Nat. Rev. Immunol.*, 15, 471–485.
38. Lebeer,S., Vanderleyden,J. and De Keersmaecker,S.C.J. (2010) Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nat. Rev. Microbiol.*, 8, 171–184.
39. Hevia,A., Delgado,S., Sánchez,B. and Margolles,A. (2015) Molecular players involved in the interaction between beneficial bacteria and the immune system. *Front. Microbiol.*, 6, 1285.
40. Al-Hassi,H.O., Mann,E.R., Sanchez,B. *et al.* (2014) Altered human gut dendritic cell properties in ulcerative colitis are reversed by *Lactobacillus plantarum* extracellular encrypted peptide STp. *Mol. Nutr. Food Res.*, 58, 1132–1143.
41. Bernardo,D., Sánchez,B., Al-Hassi,H.O. *et al.* (2012) Microbiota/host crosstalk biomarkers: regulatory response of human intestinal dendritic cells exposed to *Lactobacillus* extracellular encrypted peptide. *PLoS One*, 7, 1–8.