Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: In–silico study

sapna negi

Abstract

Citation: sapna negi Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: In–silico study. **protocols.io**
dx.doi.org/10.17504/protocols.io.ibccaiw
Published: 06 Jul 2017

Protocol

Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: In–silico study

Step 1.

- **Candidate peptides identification and characterization**

1. **Identification of candidate peptides:** Sequence identity search was conducted between all gut bacterial species/genus and complete human proteome**.** Here we took gut bacterial species/genus information from Human Microbiome Project database (www.http://hmpdacc.org/) and used all available annotated protein sequences of these bacterial species/genus from Uniprot database (https://www.ebi.ac.uk/uniprot). Wherever, gut bacterial species information was mentioned in HMP database, the related species were looked for sequence homology with human proteins. In 319 bacterial strain (from a total of 823 in HMP) found in gut did not have species information but had genus information. Protein sequences of all the species of these genera were searched against human protein database (Uniprot database). Sequence similarity search using pBLAST (Basic Local Alignment Search Tool for protein sequences) was carried out, to get peptide similarity between gut bacteria and human expressed proteins. Data on length and sequence of homologous regions, homologous peptide sequence and protein/gene IDs and names were recorded. Peptides having homologous regions of ≥9 aa were included in the study.

All the gut bacterial proteins as well as human proteins corresponding to homologous peptides were subjected to various characterization in order to further understand their biological relevance to disease etiology.

1. **Candidate peptide antigenic potential:** The homologous peptides were subjected to HLA class II binding profile using *in-silico* The peptide candidates (as seen in Supplementary table 1**)** were analyzed for HLA class II using ProPred software [21, 22]. Propred is a matrix-based method that allows prediction of MHC binders for various alleles based on experimental binding profiles. Binding scores were generated for 50 HLA class II (HLA-II) alleles. Affinity values were generated for candidate as well as random peptides from the same protein. The threshold binding values for all peptides were subjected for significant association using non-parametric Mann-Whiteny test. A significant difference in binding threshold values defines a difference in binding to candidate peptide than to others from same protein with respect to presence of specific HLA-II allele in the host. However, all the peptides (irrespective of their significant binding to HLA) were considered for further analysis for their role in human diseases. The candidate peptides having association of corresponding human protein with common diseases are displayed in heatmap to show their binding affinity with various HLA-II alleles.

Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: In–silico study

Step 2.

- **Bacterial characterization**

Bacterial proteins were studied for location of peptides in bacterial cell and for their antigenic potential. This may help us in predicting effect of antigen on our immune system. Predominant gut bacteria phyla possessing these auto-immune peptides were then estimated in the dataset.

1. **Bacterial protein location:** The set of unique candidate peptides were subjected to PSLpred [23] server for identification of bacterial candidate protein's cellular location, that is, extracellular, periplasmic, outer membrane, inner membrane or cytoplasmic**.**

2. **Predominant gut bacterial phyla in dataset:** The bacterial species encoding candidate peptides were subjected for phylum level classification (microwiki.com) to know predominance of each bacterial phylum in our dataset of autoimmune candidate peptides. Difference in number of species in each phylum was compared with number of autoimmune candidates encoded by them. P-value and 95% CI for a significant difference was calculated using pearson's correlation using SISA statistical tool (www.quantitativeskills.com/sisa/).

Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: In–silico study

Step 3.

- **Human protein characterization**

The human proteins corresponding to candidate peptides (candidate human proteins) were characterized using systems biology approach to know which biological process, KEGG pathway, and tissue may be affected and the type of diseases associated, with raising of antibody to the candidate peptide.

1. **Disease association:** 575 human proteins were annotated under disease subgroups using the Genetic Association Database (GAD). Genes were classified using GAD module of DAVID functional annotation tool (https://david-d.ncifcrf.gov/summary.jsp). To know diseases with significant number of associated candidate human proteins matching our list, p-value was generated using boneferroni multiple correction method. A distance matrix to disease was created based on number of gut bacterial species having the candidate peptide with autoimmunity potential. Closer the peptide to the disease in the cytoscape higher is the number of bacterial species harboring the specific candidate peptide.
2. **Tissue specificity:** The details on tissue specific expression of candidate human proteins was generated using Uniprot tissue database (UP_Tissue database) and p-values were calculated through DAVID functional annotation tool. Boneferroni multiple correction was used to get adjusted p-values. A total of 667 candidate human proteins could be annotated for tissue expression.
3. **Gene Ontology and Functional annotation:** Candidate human proteins were studied for their involvement in biological processes using DAVID functional annotation tool. These genes were also studied for their involvement in molecular pathways using KEGG database and significant pathway associations were recorded. All the p-values were checked for multiple corrections using boneferroni. Total 670 and 470 candidate human proteins were annotated by biological processes and KEGG pathways respectively.