Analysis of the relationship between proteins of microorganisms and MS disease by bioinformatics methods

Ali Bigdeli
Tarbiat Modares University Faculty of Biological Sciences
Seyed Shahriar Arab (sh.arab@modares.ac.ir)
Tarbiat Modares University Faculty of Biological Sciences https://orcid.org/0000-0002-2739-1617
Mehrdad Behmanesh
Tarbiat Modares University Faculty of Biological Sciences

Research Article

Keywords: MS, Bioinformatics, Myelin basic protein, Myelin oligodendrocytes glycoprotein, Microorganisms

Posted Date: April 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1510914/v1

Abstract

Multiple Sclerosis (MS) is an autoimmune disease that is caused by the invasion and destruction of the myelin sheath in the human nervous system. Several studies have shown that there is a powerful link between infectious agents such as bacteria and viruses, with the development or severity of MS disease. With regards to the normal duty of the immune system, we hypothesize that this attack can be caused by sensitizing the immune system to an antigen/s containing resemble myelin sheath components. This study is tried to use bioinformatics and its related tools as a new approach to finding the possible resemble protein amino acid sequences or 3D structures in different microorganisms with human myelin basic protein as a main protein component of myelin sheath human nerve system. Based on the limitations of the proteins 3D structures prediction, the study was categorized in two ways: 1. Simultaneous study and analysis of protein structure and sequence 2. Study exclusively through Sequence similarities between proteins. The similarity of the sequence in several analysis suggests that the presence of these microorganisms or their entry into the body increases the risk of MS. The results indicated the presence of species of fungi is highly capable of infecting immune cells associated with MS. The sequence similarity in the bacterial structures, specifically in the intestinal flora, indicates the class of bacteria that intercede the metabolism of the human body and promote natural activities. Their imbalance increases the possibility of the immune system activating against the body itself.

Highlights

- Two target proteins in the nervous system associated with MS have been studied with pathogenic and non-pathogenic microorganisms
- A different approach to MS-related structural proteins is proposed.
- This study is based on the preserved similarities of amino acids and the tertiary structure of proteins.
- The possibility of error in the immune system has been investigated by examining antigenic epitopes in sequences.

1. Introduction

Multiple sclerosis (MS) is an autoimmune disease that causes inflammation. MS is a central nervous system (CNS) disease that produces inflammatory plaques in the brain, spinal cord, and optic nerve, as well as nerve cell degeneration[1,2]. Several proteins are involved in the development of the nervous system's myelin structure. The myelin basic protein (MBP) is one of these proteins that is required for the formation of myelin[3]. The second of numerous proteins involved in this structure is myelin oligodendrocyte glycoprotein (MOG). This protein is only found on the outside of the CNS myelin and oligodendrocytes membranes[4].

The immune system is activated by microorganisms. The immune system is reacted to infectious agents in the body. The immune system recognizes specific antigens for each of them and then secretes antibodies to overcome them. However, structural similarities between viral and bacterial antigens and their own antigens promote immune system malfunction, and autoimmune disease is caused by an attack on the host body itself. The process through which antigenic similarities generate autoimmune diseases is referred to as molecular mimicry[5,6]. Researchers have conducted considerable studies on these diseases as a result of the link between microbes and autoimmune diseases, and research is still ongoing[7–9]. MS exacerbation has also been linked to infections of the upper respiratory tract, gastrointestinal tract, and genitourinary system. Microbial pathogens may impact the neurological system of immunocompromised people, notwithstanding the absence of clear evidence for an infectious etiology of MS in humans. Viruses including Epstein-Barr virus (EBV), Human Herpesvirus 6 (HHV-6), and the Human endogenous retroviruses (HERV) family, as well as bacteria like Helicobacter pylori, Chlamydia pneumonia, and Staphylococcus aureus enterotoxins, have all been associated with the development or exacerbation of MS in various studies[10].

1.1 Relation of Parvoviridae family with MS

The Parvoviridae family of microorganisms belongs to the virus category. The link of Parvovirus B19 species of this family with most autoimmune diseases has been researched. However, the status of pathogenic proteins of this virus in relation to MS has not been thoroughly determined. Through molecular mimicry of the processes reported in Human parvovirus B19 (HPV-B19) associated autoimmune proteins, IgG antibodies is interacted with recognized human autoantigens, such as MBP[11]. In relation to this virus, only the tertiary structure of the capsid protein is submitted in the RCSB database. The possibility of similarity of this protein with MBP was investigated in this study.

Adeno associated viruses (AAV), which are utilized as vectors in genetic engineering transmissions, are the second member of this family[12]. Serotype 4 of this virus has a three-dimensional (3D) structure that is highly similar to HPV-B19 and its association with MS can be investigated.

1.2 Relation a member of Herpesviridae with MS

The following virus is a member of the Herpesviridae family that is linked to autoimmune disorders[13]. Using this virus family, serotype 6 has been linked to the autoimmune disorder MS in the literature[14]. In this system, the immune response and function of CD4+ T cells were studied and verified using the MDRPRTPPPSYSE sequence of U24 protein from HHV-6 and the IVTPRTPPPSQGK sequence of MBP[15]. According to previous data, the impact of trypsin on the amino acids Arginine and Threonine leads to the hypothesis of autoantigens and MS[16].

1.3 Association of gut microbiota and MOG

Bacteria will be the subject of the following report. The synergistic impact of Lactobacillus reuteri and OTU0002 Erysipelotrichaceae in the gut generated T cell responses in the intestine and particular response to MOG in the nervous system that produces autoimmune encephalitis. This action is according to information collected from intestinal microorganisms. This mechanism relationship is unknown, although it is being investigated. In this case, scientists discovered two cases of bacteria and MOG-dependent CD4+ cells. This result identified their proteins and their similarity in the target species due to the induction of autoimmune encephalitis in mice by this disruption[17].

1.4 Concurrent association of virus and bacterium in MS

In the next case, according to a research, portion of the peptides of two proteins from bacteria and the EBV were found in the cerebrospinal fluid and serum of MS patients, causing an immune response. The study looked at the cross-reaction between the peptides of this protein from viruses and bacteria, as well as the cross-reaction in the MBP and Interferon regulatory factor 5 protein, (IRF-5) and came up with the aforementioned finding[18]. IRF-5 regulatory factor epitope matching to EBV virus and Mycobacterium paratuberculosis (MAP) produces a particular humoral immune response in MS patients. Peptides of these two bacteria have been discovered using myelin-based anti-protein antibodies[19,20].

1.5 Epsilon toxin from Clostridium perfringens and MS

The other type is the link between epsilon toxin and Clostridium perfringens, which has been examined in MS patients, and clinical findings reveal antibodies to this toxin. This toxin causes brain damage in animals and humans by affecting their neurological systems. Observations by Western blotting confirmed the presence of antibodies to epsilon toxin in some patients with multiple sclerosis in the United States[21]. In another study, three groups of MS patients were tested in a control group of 125. Two groups demonstrated a stronger response (14 percent) in the serum of patients than the control group using epsilon toxin. Peptide 170-184 with the TGVSLTTSYSFANTN sequence is the Epsilon toxin of the bacterium Clostridium perfringens. This peptide is obtained from the Western blot results of a group of patients that was not present in the other experimental groups[22].

1.6 SARS-CoV-2019

With the outspread of the COVID-19 pandemic in 2019, researchers have had many questions about the SARS-CoV-2 virus. In an investigation, the potential link between this virus and autoimmune diseases in various population groups. During the pandemic, numerous reports and evidence of inflammation and autoimmunity have been published in scientific journals. The study of autoantibodies in clinical trials, genetic diversity in populations, and the possibility of autoimmunity after remission of the pandemic in the world are investigated. As well as reports of cases such as Guillain-Barré syndrome, Myasthenia gravis, Psoriasis, Inflammatory arthritis in studies, have highlighted the importance of this disease research. Equally important is the presence of autoantibodies in COVID-19 patients, which can include Anti-nuclear antibodies (ANA), Antiprothrombin IgM, Anti-CCP antibodies, etc[23–25].

1.7 Gut microbiota balance

The next two examples are distinct from the aforesaid example and contribute to the overall balance of immune system. The immune system is disrupted and autoimmune disorders are caused by an imbalance in the gut microbiome. Butyric acid, a short-lived fatty acid that influences gut bacteria metabolism, is decreased in MS patients[26]. Bile salt hydrolase (BSH) is involved in the metabolism of butyric acid and is found in many gut bacteria. They play an important part in its synthesis, working in tandem with human enzymes[26]. Polysaccharide A, found in Bacteroides fragilis, is the last example. This polysaccharide has the potential to protect nerve cells in the human body against demyelination. The synthesis of this polysaccharide necessitates the use of many enzymes[27]. The immune system reacts against nerve cells when the equilibrium of this bacteria in the gut microbiome is disrupted. The findings of this study reveal that a part of the MOG and MBP autoantigen sequences are similar to some enzyme sequences involved in butyric acid and polysaccharide A metabolism.

Advances in numerous sectors of life sciences, as well as the discovery of various methods for disease treatment are noticeable. Bioinformatics and its numerous technologies, such as drug design, molecular dynamics and simulation, proteomics investigations, and structural proteins, play a critical role in developing clinical research in the treatment of diseases and improving community health[28]. In this regard, bioinformatics research has been conducted on several autoimmune disorders[29,30]. The goal of this study is to look at viral pathogenic proteins, which are bacteria that have a comparable structure to myelin-based proteins in the nervous system in terms of protein sequence and 3D structure. Using a variety of bioinformatics techniques, researchers have solved this problem and provided a novel strategy to the treatment and prevention of autoimmune disorders, especially MS, which is the focus of this project. In this study, structural relationships and protein sequence similarity with respect to the microorganisms mentioned in the introduction section are investigated in order to evaluate the possibility of association with MOG and MBP.

2. Materials And Methods

2.1 Data extraction

In general, the information used in this study includes the amino acid sequences of MBP and MOG from humans, as well as several groups of viruses, bacteria, and molds that have been linked to MS. This information from patients has been obtained clinically. Experiments with mice have also been studied, specifically in this project and in relation to their structural proteins.

2.2 Extraction of data related to MBP

The amino acid sequence of protein was obtained from the Uniprot database under the accession number P02686. The sequence was checked in the RCSB database for the presence of a tertiary structure. The MBP sequence was checked using the Basic Local Alignment Search Tool (BLAST) online software available in the National Center for Biotechnology Information (NCBI) database and the Protein Data Bank (PDB) database as the sequence comparison reference. The information obtained from the NCBI database in the MBP analysis with the PDB reference database revealed that among the 304 amino acids, there are only a few known 3D structures. (Fig. 1).

2.3 The tertiary structures of the MBP

To retrieve the structure of MBP, the AlphaFold database, created by the DeepMind team using an artificial intelligence methodology, is utilized. DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) have created an online web server to provide free 3D prediction structures to the scientific community (http://alphafold.ebi.ac.uk). AlphaFold predicts 3D protein structures with scoring more than 90 in CASP14 (Critical Assessment of protein Structure Prediction) challenging. A deep learning algorithm is used for prediction, which makes use of physical, biological, and bioinformatics information about protein structure, as well as multiple sequence alignments[31,32].

2.4 MBP sequence alignment with HPV-B19 and AAV-4 viral capsids proteins

The HPV-B19 and AAV-4 viral capsids sequences separately were aligned with MBP. The PRALINE[33,34] server was used to these alignment, which has the ability to consider the predicted secondary structure, when the amino acid sequences are aligned. The results were utilized for further investigation in the tertiary structures. Also, to evaluate the obtained scores and classify the amino acid alignments information, their position in the 3D structure was investigated. The tertiary structure was investigated using the YASARA[35] software. As well as the structural overlap feature was used to analyze the similarities between the two 3D structures of viruses using the pairwise alignment of software (MUSTANG algorithm).

2.5 Data related to Oligodendrocyte glycoprotein

Data on oligodendrocyte glycoproteins are retrieved from the UniProt database using the human and mouse codes Q16653 and Q61885, respectively. The YRSPFSRVV motif is part of a protein sequence in mice that is linked to the development of autoimmune encephalitis of the nervous system and, as a result, MS[17]. This MOG motif in mice has a specified tertiary structure that was discovered using X-ray crystallography. YRPPFSRVV is a human sequence that is similar to this motif, and the tertiary structure of this domain is also included in the PDB database.

2.6 List of microbe motifs related to MS disease

The information used as input in the alignments and adaptations of this study to find connections with accurate information that has been thoroughly examined (Table 1). This table is obtained from the search for the relationship between microorganisms and MBP and MOG.

Table1. Information on Previously Investigated Sequences, Alignments, and Adaptations.

Protein	Sequence	Organism	References
U24 protein	MDRPRTPPPSYSE	Human herpesvirus 6B (strain Z29) (HHV-6 variant B) (Human B lymphotropic virus)	[15]
Elongation factor G	GRLTFLRVY	Lactobacillus reuteri (strain DSM 20016)	
UvrABC system protein A	KREGFVRVQ	Lactobacillus reuteri	
Aminopeptidase (AP)	ERDGFTRVL	Erysipelotrichaceae bacterium OTU0002	
Amino acid ABC transporter substrate-binding protein	KRIAFSRVY	Lactobacillus reuteri	[17]
Tyr recombinase domain-containing protein	PGRRPFTRKELQ	Mycolicibacterium paratuberculosis (strain ATCC BAA-968 / K-10) (Mycobacterium paratuberculosis)	
Epstein-Barr nuclear antigen 1	PGRRPFFHPVGEAD	Epstein-Barr virus (strain B95-8) (HHV-4) (Human herpesvirus 4)	
Inner tegument protein	AVPVLAFDAARLRLLE	Epstein-Barr virus (strain B95-8) (HHV-4) (Human herpesvirus 4)	
Uncharacterized protein (MAP_4027)	AVVPVLAYAAARLLL	Mycolicibacterium paratuberculosis (strain ATCC BAA-968 / K-10) (Mycobacterium paratuberculosis)	[18,19]
Epsilon-toxin type B	TGVSLTTSYSFANTN	Clostridium perfringens	[22]
Conjugated bile acid hydrolase	All of Sequence	Clostridium perfringens (strain 13 / Type A)	
Conjugated bile acid hydrolase	All of Sequence	Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1)	[26]
Capsular polysaccharide biosynthesis protein	All of Sequence	Bacteroides fragilis (strain ATCC 25285 / DSM 2151 / JCM 11019 / NCTC 9343)	[27]

3. Results

3.1 Similarity in the sequences and structural alignments

3.1.1 HPV-B19

Code Q9PZT0 in the UniProt database represents the capsid sequence of the HPV-B19. The tertiary structure of the HPV-B19 capsid protein in the RSCB database with the 1S58 code is obtained through the X-ray diffraction method. Using this information, the structure and sequence alignment of the aforementioned protein with MBP was performed. It should be noted that there are multiple references that associate the HPV-B19 virus and MS. However, research into the location of pathogenic proteins of this virus in relation to MS has not been adequately identified[36]. Sequence alignment of two proteins displays in the Fig. 2a.

3.1.2 AAV-4

The Adeno-Associated Virus serotype 4 (AAV-4) capsid sequence is represented by the UniProt database code O41855. AAV-4 capsid with 2G8G code obtained by X-ray diffraction method was selected from the RCSB database. The tertiary structure of HPV-B19 and AAV-4 capsids have Root-mean-square deviation (RMSD) 9.64 for the Cα atoms. According to this score, similarity between them is high, therefore alignment of structure and sequence for AAV-4 with MBP had done. The alignment result of protein sequences is shown in Fig. 3a.

The alignment of the tertiary structure between the HPV-B19 and AAV-4 of capsid proteins and the MBP has been done by MUSTANG algorithm (Fig. 4a, b).

Aligned amino acids with a high score of 8 in Figs. 2 and 3 were chosen for the 3D position of protein in the two viral capsid structures (Fig. 5a, b, c). The similarity in the aligned positions between the two virus structures is very noticeable on the left side of the images of the two 3D structures.

3.2 Similarity of sequences

3.2.1 Human Herpes Virus and MBP

The BLAST was done on the PRTPPPS sequence that is described in the introduction section. The information, which is related to three fungi and one amoebae species, was obtained (Table 2).

Table 2. BLAST Peptide of MBP sequence 230 to 236 provided the following results

Protein sequence (as a study input) and its position				
Human	230	PRTPPPS	236	MBP
Protein sequences found using BLAST and their positions				
Aspergillus clavatus NRRL1	1124	PSTPPPS	1130	Protein transport protein sec31
Aspergillus fischeri NRRL 181	1110	PSTPPPS	1116	Protein transport protein sec31
Aspergillus fumigatus Af293	1110	PSTPPPS	1116	Protein transport protein sec31
Dictyostelium discoideum	95	PPTPPPS	101	Homeobox protein 12

The obtained results are analyzed in the Immune Epitope Database (IEDB) to assess the immunogenicity of the aforementioned sequences and their effect on CD4+ in immune system T cells. Because sequences in the IEDB in this section must be at least 15 amino acids long, the length of the sequences is greater than the length shown in the table above. Given the previously demonstrated association between the human leukocyte antigens (HLA)-DRB1:15:01 allele and MS, it is important to note the scores obtained from this allele by the IEDB[37]. The obtained results are shown in Table 3.

Table 3. CD4+ T cell immunogenicity prediction scores. All sequences were analyzed using the CD4+ T cell prediction tool in the IEDB. HLA-DRB1:15:01 allele .scores are bolded in the table and underlined because of their importance in MS association

Protein Name	Organism, Amino acids position	Peptide	Combined Score	Immuno-genicity Score	Peptide core	Median Percentile Rank (7-allele)	HLA-DRB 1:03:01	HLA-DRB 1:07:01	HLA-DRB 1:15:01	H D 3:
Protein transport protein sec31	Aspergillus clavatus 1122-1136	YAPSTPPPSQLPMQQ	87.84488	99.6122	YAPSTPPPS	80	93	73	**95**	7
	Aspergillus fischeri, Aspergilllus fumigatus 1108-1122	YAPSTPPPSQLPMQQ	87.84488	99.6122	YAPSTPPPS	80	93	73	**95**	7
Homeobox protein 12	Dictyostelium discoideum (Slime mold) 92-106	QATPPTPPPSSSSLL	99.99964	99.9991	PTPPPSSSS	100	100	63	**100**	9
U24 protein	Human herpesvirus 6A 1-15	MDRPRTPPPSYSEVL	96.95668	99.8917	DRPRTPPPS	95	97	91	**92**	8
MBP	Human 227-241	IVTPRTPPPSQGKGR	97.43512	99.5878	VTPRTPPPS	96	88	98	**96**	9

The method used in the results of Table 3 is a combination of prediction methods major histocompatibility complex (MHC) binding (using 7-allele method)[38] and immunogenicity.

The Dictyostelium discoideum has a higher immunogenicity score than other sequences. Also, the score of HLA-DRB:15:01 is higher than other cases in this sequence, indicating the importance of the information obtained due to the association of this allele with MS (Table 3).

3.2.2 Relation of the mouse gut flora amino acid sequences and the MOG

Miyauchi et al. discovered similarities in the protein sequences of Lactobacillus, Erysipelotrichaceae, and MOG. The sequences of these proteins are displayed in Table 4[17].

Table 4. shows the similarity information of proteins sequence and conserved amino acids in the study by Miyauchi et al.

Organism	Sequence information			Protein name
Lactobacillus reuteri	172	KREGFVRVQ	180	UvrABC system protein A
Lactobacillus reuteri (strain DSM 20016)	324	GRLTFLRVY	332	Elongation factor G
Lactobacillus reuteri (strain DSM 20016)	122	KRIAFSRVY	130	Amino acid ABC transporter substrate-binding protein
OTU0002 Erysipelotrichaceae		ERDGFTRVL		Xaa-Pro Aminopeptidase
Mus musculus	68	YRSPFSRVV	76	MOG (MOG_MOUSE)

Then these sequences were then analyzed using the BLAST database, and the results were imported into Table 5.

The findings indicate the presence of highly similar proteins in various bacteria. Repetition of Lactobacillus, which is part of the gut flora, has been observed in several cases[39]. The resulting proteins are involved in the binding of a specific type of RNA, the DNA damage and catalysis system, and the translation of the protein by the elongation factor G. Lactobacillus sakei subspecies are used in the food industry, and studies show that this bacterium plays a protective role in the sinus cavities, preventing sinusitis[40,41].

Clostridium and its elongating factor G, were found in the results. In the next phase of the research, other bacterial species that secrete epsilon toxin and are linked to MS will be discussed further.

Leptotrix, a bacterium that oxidizes iron and manganese, is abundant in running water containing these elements[42]. A ribosomal protein with cation-dependent endonuclease activity is involved in this similarity.

The lactic acid bacterium Oenococcus oeni is used in a variety of industries, for example, its high content in butter causes more aroma and flavor[43].

In another case, an Aspergillus species capable of secreting patulin toxin and causing neurological damage can enter the respiratory system and cause pathogenicity[44]. The 6-methylsalicylic acid synthase is one of the enzymes involved in patulin biosynthesis[45]. Methanococcus is a thermophilic archaebacterium that performs methanogenic metabolism and participates in biochemical cycles[46].

Two of the Penicillium species in Table 5 can also produce the extremely dangerous patulin toxin. In addition, one of these two species is known as blue mold, and it has been linked to apple contamination[47,48]. In both cases of penicillium, the elongation factor G are the proteins with structural similarities and associated sequences with the studied bacteria.

Table 5. BLAST results of bacterial proteins from Table 4.

Organism	Sequence information			Protein name
Aspergillus clavatus (strain ATCC 1007 / CBS 513.65 / DSM 816 / NCTC 3887 / NRRL 1)	1468	TRAGFTRVL	1475	6-methylsalicylic acid synthase
Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) (Methanococcus jannaschii)	375	KRDGFIRVL	383	Adenine deaminase
Lactobacillus sakei subsp. sakei (strain 23K)	50	LRDGFIRVR	58	SsrA-binding protein
Penicillium expansum (Blue mold rot fungus)	1523	TRDAFNRVL	1531	6-methylsalicylic acid synthase
Penicillium patulum (Penicillium griseofulvum)	1523	TRDAFNRVL	1531	6-methylsalicylic acid synthase
Lactobacillus plantarum	172	KREGFVRVR	180	UvrABC system protein A (UVRA_LACPL)
Oenococcus oeni (strain ATCC BAA-331 / PSU-1)	332	GRLTFLRVY	340	Elongation factor G
Lactobacillus fermentum (strain NBRC 3956 / LMG 18251)	324	GRLTFLRVY	332	Elongation factor G
Clostridium perfringens (strain ATCC 13124 / DSM 756 / JCM 1290 / NCIMB 6125 / NCTC 8237 / Type A)	321	GRLAFTRVY	329	Elongation factor G
Leptothrix cholodnii (strain ATCC 51168 / LMG 8142 / SP-6) (Leptothrix discophora (strain SP-6))	50	LRIAFDRVE	58	30S ribosomal protein S16

3.2.3 Relation of MAP, EBV, and IRF5 proteins with MBP

According to the information provided in the introduction section, which is related to MAP and EBV, the results of the study by Mameli et al. are shown in Table 6.

Table 6. shows the sequences related to the two microorganisms mentioned, as well as the results of the study conducted by Mameli et al.

Organism	Sequence information			Protein name	reference
Mycolicibacterium paratuberculosis (strain ATCC BAA-968 / K-10)	121	PGRRPFTRKELQ	132	Tyr recombinase domain-containing protein (MAP_0106)	[18]
Epstein-Barr virus (strain B95-8) (HHV-4)	400	PGRRPFFHPVGEAD	413	Epstein-Barr nuclear antigen 1 (EBNA1)	
Human) Homo sapiens (	85	PGSRPHLIRLFSRD	98	MBP (MBP_HUMAN)	
Organism	**Sequence information**			**Protein name**	**reference**
Mycolicibacterium paratuberculosis (strain ATCC BAA-968 / K-10)	18	AVVPVLAYAAARLLL	32	Uncharacterized protein (MAP_4027)	[19]
Epstein-Barr virus (strain B95-8) (HHV-4)	305	AVPVLAFDAARLRLLE	320	Inner tegument protein (BOLF1)	
Homo sapiens (Human)	424	RLLLEMFSGEL	434	Interferon regulatory factor 5 (IRF5_HUMAN)	

The obtained results from the Table 6 are aligned in with each other (Figs. 6-7).

The sequences and conserved sites of each of these proteins (1.MAP 0106, EBNA1, and MBP 2.MAP 4027, BOLF1, and IRF5) were performed in the BLAST database. The information on each sequence, function, and position in the microorganism are provided by Table 7. Additionally, multiple alignments of each of the protein classes have done (Figs. 8-9).

The BLAST results show that different pathogens have sequences that are similar to the two groups of MAP and EBV peptide sequences, which will be introduced. These findings are summarized in Table 7.

A species of Aspergillus fungus that causes lung infection and, due to the conditions growing of fungus and resistance in the environment, can cause a destructive infection in the lung and result in the death of patient[49].

A similar sequence was found in a fungal protein that is responsible for editing pre-mRNA. Klebsiella pneumonia is the second case in the results of Table 7. This bacterium is found in the flora of the mouth, skin, and intestines and, if inhaled, can cause lung alveoli infections and pneumonia. This type of pneumonia is more common in people with weakened immune systems and in nosocomial infections[50]. In bacteria, this sequence is part of a protein that synthesized pyrroloquinoline quinone.

Schizosaccharomyces pombe, is the next item in Table 7. It is a yeast that use in basic molecular biology studies, and an extension of these studies can be seen in biotechnology. This yeast is also one of two yeasts with more than 200 human-like genes, 23 of which are involved in cancer[51]. This sequence is part of a protein that is responsible for histone nucleosome binding and is a prerequisite for chromatin regeneration in bacteria.

The two following findings are related to the bacterium Borkhdelriaia. Because of the similarity in the immune response. The first case (B. thailandensis) is used to model the infection of other species of this bacterium family because it is easier to work with and does not require high-level biological facilities. The second strain of this bacterium (B. cenocepacia) causes lung infections in cystic fibrosis patients and is one of the most dangerous pathogens for these patients[52,53]. The first protein from the bacterium Borkhdeldriaia is involved in controlling the function of the s70 ribosome and the maturation of srRNA16 in its final stages. The role of the second protein of this bacterium in the biosynthesis pathway of converting GTP to 7,8-dihydroneuroprotein triphosphate.

Bordetella avium, the final bacterium in the first group, causes a respiratory infection in turkeys, causing disease by attaching to and spreading to respiratory epithelial cells. It should be noted that this disease is known as Bordetellosis in birds. The presence of this bacterium has been reported in studies performed on patients with cystic fibrosis and is considered a risk factor. The presence of this bacterium has been reported in studies performed on patients with cystic fibrosis and is considered an opportunistic pathogen in humans[54,55]. In this study, a portion of the enzyme glucose 6-phosphate isomerase was obtained from this bacterium (this enzyme is involved in the glycolysis cycle).

The first bacterium in the second category of BLAST results is MAP. According to the researches, this bacterium causes tuberculosis and respiratory infections[56]. Trehalose Phosphate Phosphatase is one of this proteins of the bacterium, which is similar to the second group of sequences, functions as a hydrolase, separating phosphate from trehalose phosphate.

The second bacterium is a freshwater single-celled cyanobacterium. The photosynthesis and mechanisms involved in cyanobacteria are used as examples. Synechocystis sp. PCC6803 can grow in the absence of light; thus, it is one of the most widely used cyanobacteria[57]. This sequence is part of the NAD(P)H-quinone oxidoreductase enzyme belonging to the mentioned bacteria, which is shown in the results of Table 7.

The following bacterium is a Streptomyces species that is involved in the production of antibiotics and vitamin B12[58]. ElmGT protein of Streptomyces bacteria is involved in the biosynthesis of alaromycin in glucose transport.

The final bacterium obtained from the results is a cyanobacterium with autotrophic and rapid growth against yeast. Its ability to grow rapidly in the presence of sunlight is the reason why it is used in biotechnology as a photosynthetic bioreactor model[59]. The obtained sequence, like that of other cyanobacteria, is related to the NAD(P)H-quinone oxidoreductase enzyme in Thermosynechococcus elongatus.

In addition to the above information, the probability percentage of each sequence as an epitope of an antigen in CD4+ T cell activation is showed in the Table 7. It should be noted that this score is only related to the leukocyte antigen allele HLA-DRB1:15:01 and is related to MS, as determined by the IEDB[60].

Table 7. BLAST results in two categories of protein sequences EBV, MAP, IRF5, MBP.

Results obtained from the first group (EBNA1, MAP_0106, MBP)					HLA-DRB 1:15:01 Score
Sequence	**Sequence information**			**Protein name**	
Aspergillus fumigatus (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100)	818	PGRRPRGDSLSRSPS	832	Pre-mRNA-splicing factor cwc22	97
Klebsiella pneumoniae	803	PGRRPEAILQAILRQ	817	Coenzyme PQQ synthesis protein F	35
pombe (strain 972 / ATCC 24843)Schizosaccharomyces (Fission yeast)	346	PGRRPRSMTAPKGGA	360	Bromodomain-containing protein C631.02	82
Burkholderia thailandensis (strain ATCC 700388 / DSM 13276 / CIP 106301 / E264)	128	PGRRPTSTPNRRAPA	142	Endoribonuclease YbeY	93
Burkholderia cenocepacia (strain MC0-3)	25	PGRRPLEWVGMQGID	39	GTP cyclohydrolase FolE2 2	60
Bordetella avium (strain 197N)	436	PGRRPSTLIVLPRMT	450	Glucose-6-phosphate isomerase	43
Results obtained from the second group (BOLF1, IRF5, MAP4027)					
Sequence	**Sequence information**			**Protein name**	
Mycobacterium paratuberculosis (strain ATCC BAA-968 / K-10)	226	AIPVLAQAAGRLRDE	240	Trehalose-phosphate phosphatase	13
Synechocystis sp. (strain PCC 6803 / Kazusa)	19	LVPVLALTASKLLRP	33	-quinoneNAD(P)H oxidoreductase subunit 3	3.6
Streptomyces olivaceus	87	LAPVLAYLAGRLLEP	101	Elloramycin glycosyltransferase ElmGT	0.41
(strain BP-1)Thermosynechococcus elongatus	31	LVPVLALAASALLRP	45	NAD(P)H-quinone oxidoreductase subunit 3	4.3

The multiple alignments results were shown based on the protein data recorded in two different categories (Figs. 8-9).

These sequences were compared in terms of the possibility of eliciting an immune system response in B cell antibodies. (Fig. 10 and Table 8). The amino acids are listed separately, along with their scores as part of the antigen epitope that induces activity of immune system and exclusively B cells (Fig. 10). This prediction was performed using the Bepipred Linear Epitope Prediction 2.0 method, in which the BepiPred-2.0 server uses protein sequences to predict B cell epitopes[61].

The Table of amino acid is shown scores for each sequence separately (Fig. 10). The horizontal axis in this diagram represents the position of amino acids in each sequence, and the vertical axis shows the points related to each amino acid.

Table 8. Obtained sequences as the epitope structure of an antigen. each amino acid from sequences has a special score in epitope structure and the scores of each amino acid have been shown in Figure 10. Bolded amino acids have scored higher than the threshold (0.5) in this prediction tool.

Aspergillus fumigatus	Klebsiella pneumoniae	Schizosaccharomyces pombe	Burkholderia thailandensis	Burkholderia cenocepacia	Bordetella avium	Mycobacterium paratuberculosis	Synechocystis sp	Streptomyces olivaceus
P	P	P	P	P	P	A	L	L
G	G	G	G	G	G	I	V	A
R	R	R	R	R	R	P	P	P
R	R	R	R	R	R	V	V	V
P	P	**P**	**P**	P	P	L	L	L
R	E	**R**	**T**	**L**	S	A	A	A
G	A	**S**	**S**	**E**	T	Q	L	Y
D	I	**M**	**T**	**W**	L	A	T	L
S	L	**T**	**P**	**V**	I	A	A	A
L	Q	**A**	**N**	**G**	V	G	S	G
S	A	**P**	**R**	**M**	L	**R**	K	R
R	I	K	**R**	Q	P	L	L	L
S	L	G	A	G	R	R	L	L
P	R	G	P	I	M	D	R	E
S	Q	A	A	D	T	E	P	P

3.2.4 Correlation of Epsilon toxin from Clostridium perfringens with MS

The amino acid sequence TGVSLTTSYSFANTN (170–184) of epsilon protein, according to previous studies of researchers and the introduction information is linked to MS disease. A similar portion was obtained using the BLAST database, is belongs to Noravirus. The two sequences aligned each other (Fig. 11). This similar sequence is hosted by the Drosophila, Uncharacterized ORF3 protein, according to the Uniprot database. This virus family can infect the intestines of all Drosophila species that is a member of the Picorna virus family.

According to the IEDB, the Bepipred Linear Epitope Prediction 2.0 method shows the epitope score of antigen in the part of Noravirus sequence that is higher than the bacterial epsilon sequence. The amino acids individually investigated, along with their scores as part of the antigen epitope (Fig 12). The diagram in this Figure shows the position of amino acids in each sequence on the horizontal axis and the score of each amino acid on the vertical axis. Also, the amino acids are shown separately, in which score of them are higher than a threshold and they were bolded (Table 9).

Table 9. Norovirus and Clostridium perfringens related sequences as antigen epitopes.

Position	Noravirus	Clostridium perfringens
1	T	T
2	G	I
3	V	T
4	S	S
5	L	L
6	T	**T**
7	T	**T**
8	**S**	**S**
9	**Y**	**Y**
10	**S**	**S**
11	**F**	**L**
12	A	A
13	N	N
14	T	V
15	N	P

3.2.5 Spike and protease of SARS-CoV-2

Because of the global COVID-19 pandemic, researchers are currently interested in researching and studying various aspects of this disease. The similarities of two SARS-CoV-2 proteins with parts of two proteins, MOG and MBP, were investigated in this study. First, the sequence similarity of MOG and MBP with the spike glycoprotein of virus (UniProt ID: P0DTC2) was investigated. MBP and MOG with glycoprotein spike sequences have aligned (Figs 13-14).

In the next part, the MOG and MBP sequences were aligned with the SARS-CoV-2 main protease sequence. The results of these alignments are shown in Figs 15-16.

3.3 Gut microbiota balance

Considering the role of butyric acid and polysaccharide A in the protection of nervous system cells, bacterial enzymes that play a role in the metabolism of these two metabolites is taken into account. The investigation was divided into two parts: BSH and related enzyme to Polysaccharide A production. firstly, possibility of similarity between a portion of the MOG and MBP sequences and BSH enzyme is investigated as follow.

3.3.1 BSH enzyme involved in butyric acid metabolism

Fig. 17 demonstrates the bacteria involved in butyric acid metabolism, which, along with human enzymes, play an important role in its formation[26].

In this segment, sequences 246 to 256 of the MBP (inducing autoimmune encephalitis) and 68 to 76 of the MOG (with four protected amino acids in a study of mouse gut flora by Miyauchi et al.) were aligned with the BSH enzyme in several species of bacteria in the gut flora to investigate the possible association between a microbial imbalance in the gut. The results obtained with the active site of the enzyme in a protected part of the aligned sequence are shown below. The alignment of the MBP sequence (246 to 256) and the Cl. perfringens BSH enzyme sequence has done (Fig. 18).

Also, the multiple alignment of MBP sequences 246–256 and bacterial sequences in their common region was investigated (Fig. 19).

The presence of Lactobacillus was considered in the results, so in addition to the above alignment, the entire sequence of this protein from the bacterium with the results of Table 5 and the MOG of mouse and human was aligned (Fig. 20). Three amino acids from the four protected amino acids in the obtained results are the same (according to list of Table 5). Following the determination of the precise location of the MOG and Cl. perfringens alignment, the next alignment with the other bacteria (Bi. longum, La. plantarum) was performed (Fig. 21).

The 3D structure of two proteins of Bi. Longum (PDB ID: 2HF0) and Cl. Perfringens (PDB ID: 2RF8) has been determined by the X-ray diffraction method. As a result, these two structures were matched to each other (Fig. 22). The common motifs obtained from the alignment with the MOG sequence, Bi. longum, and Cl. perfringens are marked in red and green, purple, and yellow, respectively. It should be noted that the sequence similarity between these two proteins of the bacteria is 37%. However, according to the image, the alignment of the tertiary structure of these two proteins shows a high overlap. Additionally, the RMSD between two chains of molecules is 3.53, as calculated for Cα atoms.

3.3.2 Polysaccharide A in Bacteroides fragilis

Several enzymes are involved in the production of polysaccharide A[62], which is shown in Fig. 23.

According to the results of the studies, a portion of the MOG sequence and a portion of the two proteins of the polysaccharide A metabolic pathway have similar amino acid sequence. The first similarity is the MOG association with Wzy. This enzyme of bacterium acts in the periplasmic space as a polymerase in the binding of polysaccharide subunits. MOG alignment with this enzyme was done (Fig. 24).

In the second case, a sequence was performed between glycosyltransferase (WcfQ) and MOG (Fig. 25).

4. Conclusion

The importance of recognizing the structural relationship of MBP with MS disease is not obscured. The association between microorganisms with this disease and myelin structural protein has been reported in many life-science studies. In this study, an attempt was made to analyze the data of laboratory studies that showed the relationship of a microorganism and its specific metabolite with patients or animal models of MS, with the possibility of structural similarity with other microorganisms and even the same microorganism. Also, bioinformatics analysis and more comprehensive information on possible mechanisms in the immune system with some of microorganisms and MS had been reported.

Figure 5c depicts an examination of the alignment regions with a normalized score greater than 8 between the virus capsid sequences HPV-B19 and AAV-4 with MBP. This result demonstrates that, in addition to sequence similarity, three different parts of the tertiary structure of capsids overlap with MBP. Due to the existence of these motifs on the surface of capsid proteins, the similarity of the sequence in the alignment performed in these motifs, as well as the possibility of immune system error in the face of it, additional laboratory research is warranted.

Several transporter proteins from Aspergillus species and one species from Dictyostelium have been obtained with the similarity of a sequence of HHV-6. This information due to effect of the virus sequence on CD4+ cells of the immune system is noteworthy. The results for analysis in relation to the mentioned cases were evaluated by the IEDB (Table 3). Results of Table 3 include Aspergillus fischeri, whose genome is very similar to Aspergillus fumigatus, a fumigatus

species that causes Aspergillosis[63]. Aspergillus clavatus, which is used in the production of enzymes[64]. Also included are the soil amoebae Dictyostelium discoideum, which is used as a model in immune system studies[65]. Three mentioned organisms have a direct relationship with humans that are notable as pathogens, as well as for research and industrial application. The high score of Dictyostelium compared to the HHV-6 virus sequence and MBP, in the effect with CD4+ cell activation and HLA-DRB:15:01 involvement in the immune system, are noteworthy in these results. The results also show that all cases had a high score when compared to the HHV-6 virus sequence.

The notable results obtained in BLAST with a portion of the MOG sequence. Four amino acids Arginine, Phenylalanine, Arginine, and Valine protected in the result of Table 5 respectively. Two Penicillium species and one Aspergillus species are present in the results. Their proteins are 6-methylsalicylic acid synthase. 6-methylsalicylic acid synthase is involved in patulin biosynthesis, a toxin that causes various reactions in the immune, gastrointestinal, and nervous systems. The Elongation factor G in three cases of the results of is obtained (One of these cases is Cl. perfringens bacterium, whose epsilon toxin is associated with MS in the results). Also, the UvrABC system protein in Lactobacillus plantarum is present in the results. All of the mentioned cases are significant due to their high similarity with the Miyauchi et al study results.

In Table 7, three of the obtained sequences are related to Aspergillus fumigatus, Schizosaccharomyces pombe, and Burkholderia thailandensis. These sequence had a score above 80 in predicting the CD4+ antigen epitope and specifically in relation to the HLA-DRB1:15:01 allele, according to the IEDB.

In the same database, the above cases were predicted as an antigen epitope in the face of immune system B cells (Fig. 10). All of this information (Table 7 and Fig. 10) suggests that these three sequences have the potential to cause immune system errors, and various body conflicts. It should be noted that, as previously stated, the fungus Aspergillus fumigatus secretes patulin toxin.

The ORF3 protein sequence of NoraVirus in Drosophila is very similar to the Epsilon toxin sequence in Cl.perfringens. This sequence has a higher score than the epsilon toxin sequence in the antigen epitope position, according to data predicted from the IDEB database. Given the prevalence of this virus in Drosophila populations and the use of this organism in laboratory studies, there is an urgent need for additional research on this issue.

Extensive research on Murine Coronavirus has been conducted prior to the COVID-19 pandemic on the effect of this coronavirus on the nervous system of mice and the induction of encephalomyelitis by this virus. it is worth mentioning that during these studies, Demyelination in neurons also had similar structural to human MS disease[66–68]. Murine and Sars-CoV-2 viruses belong to the same genus, and new research shows that Sars-CoV-2 remains have been found in the human brain[69]. According to the results obtained in the sequences alignments between the part of MOG and MBP with Spike and protease proteins of the virus, the amino acid similarity is particularly high in MOG (Figs 14 and 16). These motifs are located on the surface of these virus proteins in their 3D structure, so when the immune system is exposed to these proteins, the possibility of error increases, and as a result, the individual develops autoimmunity. This is a case that needs further investigation.

BSH enzyme involved in butyric acid synthesis and enzymes involved in polysaccharide A synthesis, the study approach of which differs from all cases studied in this research. Because the human intestine coexists with microbial flora, two factors that contribute to the production of human metabolites were investigated. The goal of these cases is to look into the possibility of dysbiosis intestinal flora composition in humans. The results of the aforementioned cases show that there is sequence similarity between the MBP and MOG and the BSH, WcfQ, and Wzy enzymes. In the alignment study of MBP, the amino acid Arginine in the second position of its sequence is identical to the active site amino acid in the BSH enzyme. Another observation is the alignment of a portion of MOG with the WcfQ and Wzy enzymes, the protected amino acids Phenylalanine and Arginine. The results show that if there is an imbalance in the human microbial flora, the production of metabolites is impaired (as reported in MS patients, as mentioned in the introduction). It is possible that these cases will be misidentified by the immune system as pathogenic factors. Because of the similarity in protein sequence, this disorder may cause autoimmune reactions in the human body.

Abbreviations

3D	three-dimensional	HPV-B19	Human parvovirus B19
AAV-4	Adeno-associated virus serotype 4	IEDB	Immune Epitope Database
BLAST	Basic Local Alignment Search Tool	IRF-5	Interferon regulatory factor 5
BSH	bile salt hydrolase	MAP	Mycobacterium paratuberculosis
CD4	cluster of differentiation 4	MBP	myelin basic protein
CNS	central nervous system	MOG	myelin oligodendrocytes glycoprotein
EBV	Epstein-Barr virus	MS	Multiple Sclerosis
HERV	Human endogenous retroviruses	NCBI	National Center for Biotechnology Information
HHV-6	Human herpesvirus serotype 6	PDB	Protein Data Bank
HLA	Human leukocyte antigens	RMSD	Root-mean-square deviation

Declarations

Author information

Affiliations

Department of Biophysics, School of Biological Sciences, Tarbiat Modares University, Tehran, Iran.

Ali Bigdeli, Seyed Shahriar Arab.

Department of Molecular Genetics, School of Biological Sciences, Tarbiat Modares University, Tehran, Iran.

Mehrdad Behmanesh.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ali Bigdeli. Seyed Shahriar Arab had supervised to the continuing of procedure. Mehrdad Behmanesh was advisor of project. The first draft of the manuscript was written by Ali Bigdeli and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Corresponding author

Seyed Shahriar Arab, Ph.D.

Bioinformatics Lab, Faculty of Biological Sciences

Tarbiat Modares University, Jalal AleAhmad St

Tehran, Iran.

Email: sh.arab@modares.ac.ir

Tel: +98-21-82883494

Ethics approval

Not applicable

Availability of data and materials

Data are available from the authors upon request.

Declaration of interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

Acknowledgements

The authors would like to acknowledge the Faculty of Biological Science, Tarbiat Modares University, and the Bioinformatics laboratory for their support and contribution to this study.

Funding

This work was supported by Tarbiat Modares University. Co-Responder has received research support from this university.

Consent to Participate

Not applicable.

Consent for Publication

All authors whose names appear on the submission agreed to the version to be published.

Competing Interests

The authors declare no competing interests.

References

[1] A. Compston, A. Coles, Multiple sclerosis, Lancet. 372 (2008) 1502–1517. https://doi.org/10.1016/S0140-6736(08)61620-7.

[2] M.M. Goldenberg, Multiple Sclerosis Review, Pharm. Ther. 37 (2012) 175. /pmc/articles/PMC3351877/ (accessed September 27, 2021).

[3] M.A. Moscarello, Myelin Basic Protein, the “Executive” Molecule of the Myelin Membrane, in: Cell Biol. Pathol. Myelin, 1997. https://doi.org/10.1007/978-1-4615-5949-8_2.

[4] P. Peschl, M. Bradl, R. Höftberger, T. Berger, M. Reindl, Myelin oligodendrocyte glycoprotein: Deciphering a target in inflammatory demyelinating diseases, Front. Immunol. 8 (2017) 1. https://doi.org/10.3389/fimmu.2017.00529.

[5] R. M, R.-J. P, M. DM, P. Y, A.-A. Y, R.-S. C, L. PSC, A. AA, G. ME, A. JM, Molecular mimicry and autoimmunity, J. Autoimmun. 95 (2018) 100–123. https://doi.org/10.1016/J.JAUT.2018.10.012.

[6] M.F. Cusick, J.E. Libbey, R.S. Fujinami, Molecular Mimicry as a Mechanism of Autoimmune Disease, Clin. Rev. Allergy Immunol. 42 (2012) 102. https://doi.org/10.1007/S12016-011-8294-7.

[7] H. Xu, M. Liu, J. Cao, X. Li, D. Fan, Y. Xia, X. Lu, J. Li, D. Ju, H. Zhao, Q. Guan, The Dynamic Interplay between the Gut Microbiota and Autoimmune Diseases, J. Immunol. Res. 2019 (2019). https://doi.org/10.1155/2019/7546047.

[8] S. Trott, I.L. King, An introduction to the microbiome and MS, Mult. Scler. (2018). https://doi.org/10.1177/1352458517737391.

[9] M.K. Smatti, F.S. Cyprian, G.K. Nasrallah, A.A. Al Thani, R.O. Almishal, H.M. Yassine, Viruses and autoimmunity: A review on the potential interaction and molecular mechanisms, Viruses. 11 (2019). https://doi.org/10.3390/v11080762.

[10] M. M, A. L, F. MF, C. J, The role of infections in multiple sclerosis, Mult. Scler. 25 (2019) 891–901. https://doi.org/10.1177/1352458518823940.

[11] J.R. Kerr, The role of parvovirus B19 in the pathogenesis of autoimmunity and autoimmune disease, J. Clin. Pathol. (2016). https://doi.org/10.1136/jclinpath-2015-203455.

[12] D. Wang, P.W.L. Tai, G. Gao, Adeno-associated virus vector as a platform for gene therapy delivery, Nat. Rev. Drug Discov. (2019). https://doi.org/10.1038/s41573-019-0012-9.

[13] D.N. Posnett, Herpesviruses and autoimmunity, Curr. Opin. Investig. Drugs. 9 (2008) 505–514. https://europepmc.org/article/med/18465661 (accessed December 2, 2020).

[14] J. Gutiérrez, M.J. Vergara, M. Guerrero, O. Fernández, G. Piédrola, P. Morales, M. Del, Multiple sclerosis and human herpesvirus 6, Infection. (2002). https://doi.org/10.1007/s15010-002-2056-7.

[15] M. V. Tejada-Simon, Y.C.Q. Zang, J. Hong, V.M. Rivera, J.Z. Zhang, Cross-reactivity with myelin basic protein and human herpesvirus-6 in multiple sclerosis, Ann. Neurol. (2003). https://doi.org/10.1002/ana.10425.

[16] P. Medveczky, J. Antal, A. Patthy, K. Kékesi, G. Juhász, L. Szilágyi, L. Gráf, Myelin basic protein, an autoantigen in multiple sclerosis, is selectively processed by human trypsin 4, FEBS Lett. 580 (2006) 545–552. https://doi.org/10.1016/j.febslet.2005.12.067.

[17] E. Miyauchi, S.W. Kim, W. Suda, M. Kawasumi, S. Onawa, N. Taguchi-Atarashi, H. Morita, T.D. Taylor, M. Hattori, H. Ohno, Gut microorganisms act together to exacerbate inflammation in spinal cords, Nature. (2020). https://doi.org/10.1038/s41586-020-2634-9.

[18] G. Mameli, E. Cocco, J. Frau, M.G. Marrosu, L.A. Sechi, Epstein Barr Virus and Mycobacterium avium subsp. paratuberculosis peptides are recognized in sera and cerebrospinal fluid of MS patients, Sci. Rep. 6 (2016). https://doi.org/10.1038/srep22401.

[19] C. D, M. G, G. G, C. E, M. S, F. J, M. MG, M. R, S. LA, Human interferon regulatory factor 5 homologous epitopes of Epstein-Barr virus and Mycobacterium avium subsp. paratuberculosis induce a specific humoral and cellular immune response in multiple sclerosis patients, Mult. Scler. 21 (2015) 984–995. https://doi.org/10.1177/1352458514557304.

[20] M. G, C. D, C. E, M. S, F. J, M. MG, S. LA, Epstein-Barr virus and Mycobacterium avium subsp. paratuberculosis peptides are cross recognized by anti-myelin basic protein antibodies in multiple sclerosis patients, J. Neuroimmunol. 270 (2014) 51–55. https://doi.org/10.1016/J.JNEUROIM.2014.02.013.

[21] K.R. Rumah, J. Linden, V.A. Fischetti, T. Vartanian, Isolation of Clostridium perfringens type B in an individual at first clinical presentation of multiple sclerosis provides clues for environmental triggers of the disease, PLoS One. 8 (2013). https://doi.org/10.1371/JOURNAL.PONE.0076359.

[22] S. Wagley, M. Bokori-Brown, H. Morcrette, A. Malaspina, C. D'Arcy, S. Gnanapavan, N. Lewis, M.R. Popoff, D. Raciborska, R. Nicholas, B. Turner, R.W. Titball, Evidence of Clostridium perfringens epsilon toxin associated with multiple sclerosis, Mult. Scler. J. 25 (2019) 653–660. https://doi.org/10.1177/1352458518767327.

[23] M. Mateu-Salat, E. Urgell, A. Chico, SARS-COV-2 as a trigger for autoimmune disease: report of two cases of Graves' disease after COVID-19, J. Endocrinol. Investig. 2020 4310. 43 (2020) 1527–1528. https://doi.org/10.1007/S40618-020-01366-7.

[24] G. Halpert, Y. Shoenfeld, SARS-CoV-2, the autoimmune virus, Autoimmun. Rev. 19 (2020) 102695. https://doi.org/10.1016/J.AUTREV.2020.102695.

[25] M.C. Sacchi, S. Tamiazzo, P. Stobbione, L. Agatea, P. De Gaspari, A. Stecca, E.C. Lauritano, A. Roveta, R. Tozzoli, R. Guaschino, R. Bonometti, SARS-CoV-2 infection as a trigger of autoimmune response, Clin. Transl. Sci. 14 (2021) 898–907. https://doi.org/10.1111/CTS.12953.

[26] R. JM, K. DJ, H. PB, Bile salt biotransformations by human intestinal bacteria, J. Lipid Res. 47 (2006) 241–259. https://doi.org/10.1194/JLR.R500013-JLR200.

[27] J. Ochoa-Repáraz, D.W. Mielcarz, L.E. Ditrio, A.R. Burroughs, S. Begum-Haque, S. Dasgupta, D.L. Kasper, L.H. Kasper, Central Nervous System Demyelinating Disease Protection by the Human Commensal Bacteroides fragilis Depends on Polysaccharide A Expression , J. Immunol. 185 (2010) 4101–4108. https://doi.org/10.4049/jimmunol.1001443.

[28] S. Gill, A. Christopher, V. Gupta, P. Bansal, Emerging role of bioinformatics tools and software in evolution of clinical research, Perspect. Clin. Res. (2016). https://doi.org/10.4103/2229-3485.184782.

[29] N. Petrovsky, V. Brusic, Bioinformatics for study of autoimmunity, Autoimmunity. 39 (2006) 635–643. https://doi.org/10.1080/08916930601062437.

[30] M. Trela, P.N. Nelson, P.B. Rylance, The role of molecular mimicry and other factors in the association of Human Endogenous Retroviruses and autoimmunity, APMIS. 124 (2016) 88–104. https://doi.org/10.1111/APM.12487.

[31] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Applying and improving AlphaFold at CASP14, Proteins Struct. Funct. Bioinforma. 89 (2021) 1711–1721. https://doi.org/10.1002/PROT.26257.

[32] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nat. 2021 5967873. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[33] V.A. Simossis, J. Heringa, PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information, Nucleic Acids Res. (2005). https://doi.org/10.1093/nar/gki390.

[34] P. Bawono, J. Heringa, PRALINE: A versatile multiple sequence alignment toolkit, Methods Mol. Biol. 1079 (2014) 245–262. https://doi.org/10.1007/978-1-62703-646-7_16.

[35] H. Land, M.S. Humble, YASARA: A tool to obtain structural guidance in biocatalytic investigations, in: Methods Mol. Biol., Humana Press Inc., 2018: pp. 43–67. https://doi.org/10.1007/978-1-4939-7366-8_4.

[36] J.R. Kerr, The role of parvovirus B19 in the pathogenesis of autoimmunity and autoimmune disease, J. Clin. Pathol. 69 (2016) 279–291. https://doi.org/10.1136/JCLINPATH-2015-203455.

[37] L. Kular, Y. Liu, S. Ruhrmann, G. Zheleznyakova, F. Marabita, D. Gomez-Cabrero, T. James, E. Ewing, M. Lindén, B. Górnikiewicz, S. Aeinehband, P. Stridh, J. Link, T.F.M. Andlauer, C. Gasperi, H. Wiendl, F. Zipp, R. Gold, B. Tackenberg, F. Weber, B. Hemmer, K. Strauch, S. Heilmann-Heimbach, R. Rawal, U. Schminke, C.O. Schmidt, T. Kacprowski, A. Franke, M. Laudes, A.T. Dilthey, E.G. Celius, H.B. Søndergaard, J. Tegnér, H.F. Harbo, A.B. Oturai, S. Olafsson, H.P. Eggertsson, B. V. Halldorsson, H. Hjaltason, E. Olafsson, I. Jonsdottir, K. Stefansson, T. Olsson, F. Piehl, T.J. Ekström, I. Kockum, A.P. Feinberg, M. Jagodic, DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis, Nat. Commun. 2018 91. 9 (2018) 1–15. https://doi.org/10.1038/s41467-018-04732-5.

[38] S. Paul, C.S. Lindestam Arlehamn, T.J. Scriba, M.B.C. Dillon, C. Oseroff, D. Hinz, D.M. McKinney, S. Carrasco Pro, J. Sidney, B. Peters, A. Sette, Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes, J. Immunol. Methods. 422 (2015) 28–34. https://doi.org/10.1016/J.JIM.2015.03.022.

[39] T.R. KLAENHAMMER, C. FREMAUX, C. AHN, K. MILTON, Molecular Biology of Bacteriocins Produced by Lactobacillus, in: Bacteriocins Lact. Acid Bact., Elsevier, 1993: pp. 151–180. https://doi.org/10.1016/b978-0-12-355510-6.50015-4.

[40] N.A. Abreu, N.A. Nagalingam, Y. Song, F.C. Roediger, S.D. Pletcher, A.N. Goldberg, S. V. Lynch, Sinus microbiome diversity depletion and Corynebacterium tuberculostearicum enrichment mediates rhinosinusitis, Sci. Transl. Med. 4 (2012) 151ra124. https://doi.org/10.1126/scitranslmed.3003783.

[41] C. Montanari, F. Barbieri, M. Magnani, L. Grazia, F. Gardini, G. Tabanelli, Phenotypic diversity of Lactobacillus sakei strains, Front. Microbiol. 9 (2018) 2003. https://doi.org/10.3389/fmicb.2018.02003.

[42] M. Takeda, H. Makita, K. Ohno, Y. Nakahara, J.I. Koizumi, Structural analysis of the sheath of a sheathed bacterium, Leptothrix cholodnii, Int. J. Biol. Macromol. 37 (2005) 92–98. https://doi.org/10.1016/j.ijbiomac.2005.09.002.

[43] E. Mani-López, E. Palou, A. López-Malo, Biopreservatives as Agents to Prevent Food Spoilage, in: Microb. Contam. Food Degrad., Elsevier, 2018: pp. 235–270. https://doi.org/10.1016/b978-0-12-811515-2.00008-1.

[44] A.D. Hocking, Aspergillus and related teleomorphs, in: Food Spoilage Microorg., Elsevier Ltd, 2006: pp. 451–487. https://doi.org/10.1533/9781845691417.4.451.

[45] O. Puel, P. Galtier, I.P. Oswald, Biosynthesis and toxicological effects of patulin, Toxins (Basel). 2 (2010) 613–631. https://doi.org/10.3390/TOXINS2040613.

[46] S. DasSarma, J.A. Coker, P. DasSarma, Archaea (overview), in: Encycl. Microbiol., Elsevier Inc., 2009: pp. 1–23. https://doi.org/10.1016/B978-012373944-5.00108-5.

[47] W.K. Nip, F.S. Chu, Production of [14C]patulin by Penicillium patulum, Appl. Environ. Microbiol. 33 (1977) 814–816. https://doi.org/10.1128/aem.33.4.814-816.1977.

[48] D. Errampalli, Penicillium expansum (Blue Mold), in: Postharvest Decay Control Strateg., Elsevier Inc., 2014: pp. 189–231. https://doi.org/10.1016/B978-0-12-411552-1.00006-5.

[49] J.A. Sugui, K.J. Kwon-Chung, P.R. Juvvadi, J.-P. Latgé, W.J. Steinbach, Aspergillus fumigatus and Related Species, Cold Spring Harb. Perspect. Med. 5 (2015). https://doi.org/10.1101/CSHPERSPECT.A019786.

[50] S. Suster, C.A. Moran, *Klebsiella* Pneumonia, Diagnostic Pathol. Thorac. (2017) 530–533. https://doi.org/10.1016/B978-0-323-37715-7.50113-4.

[51] E.A. Johnson, C. Echavarri-Erasun, Yeast Biotechnology, The Yeasts. 1 (2011) 21–44. https://doi.org/10.1016/B978-0-444-52149-1.00003-3.

[52] D. P, M. E, Burkholderia cenocepacia in cystic fibrosis: epidemiology and molecular mechanisms of virulence, Clin. Microbiol. Infect. 16 (2010) 821–830. https://doi.org/10.1111/J.1469-0691.2010.03237.X.

[53] A. Haraga, T.E. West, M.J. Brittnacher, S.J. Skerrett, S.I. Miller, Burkholderia thailandensis as a model system for the study of the virulence-associated type III secretion system of burkholderia pseudomallei, Infect. Immun. 76 (2008) 5402–5411. https://doi.org/10.1128/IAI.00626-08.

[54] A.T. Harrington, J.A. Castellanos, T.M. Ziedalski, J.E. Clarridge, III, B.T. Cookson, Isolation of Bordetella avium and Novel Bordetella Strain from Patients with Respiratory Disease, Emerg. Infect. Dis. 15 (2009) 72. https://doi.org/10.3201/EID1501.071677.

[55] A.J. Chaves Hernández, Poultry and Avian Diseases, Encycl. Agric. Food Syst. (2014) 504–520. https://doi.org/10.1016/B978-0-444-52512-3.00183-2.

[56] C.T. Dow, Mycobacterium avium Subspecies Paratuberculosis and Human Disease: Bridging Infection and Autoimmunity, Infect. Autoimmun. (2015) 569–581. https://doi.org/10.1016/B978-0-444-63269-2.00037-4.

[57] Y. Yu, L. You, D. Liu, W. Hollinshead, Y.J. Tang, F. Zhang, Development of Synechocystis sp. PCC 6803 as a Phototrophic Cell Factory, Mar. Drugs. 11 (2013) 2894. https://doi.org/10.3390/MD11082894.

[58] D. Hu, S.M.-Y. Lee, K. Li, K.M. Mok, Secondary Metabolite Production Potential of Mangrove-Derived Streptomyces olivaceus, Mar. Drugs 2021, Vol. 19, Page 332. 19 (2021) 332. https://doi.org/10.3390/MD19060332.

[59] R. Azevedo, J.L. Lopes, M.M. de Souza, B.F. Quirino, L.J. Cançado, L.F. Marins, Synechococcus elongatus as a model of photosynthetic bioreactor for expression of recombinant β-glucosidases, Biotechnol. Biofuels 2019 121. 12 (2019) 1–10. https://doi.org/10.1186/S13068-019-1505-9.

[60] L. Moutsianas, L. Jostins, A.H. Beecham, A.T. Dilthey, D.K. Xifara, M. Ban, T.S. Shah, N.A. Patsopoulos, L. Alfredsson, C.A. Anderson, K.E. Attfield, S.E. Baranzini, J. Barrett, T.M.C. Binder, D. Booth, D. Buck, E.G. Celius, C. Cotsapas, S. D'Alfonso, C.A. Dendrou, P. Donnelly, B. Dubois, B. Fontaine, L. Fugger, A. Goris, P.-A. Gourraud, C. Graetz, B. Hemmer, J. Hillert, I.I.G.C. (IIBDGC), I. Kockum, S. Leslie, C.M. Lill, F. Martinelli-Boneschi, J.R. Oksenberg, T. Olsson, A. Oturai, J. Saarela, H.B. Søndergaard, A. Spurkland, B. Taylor, J. Winkelmann, F. Zipp, J.L. Haines, M.A. Pericak-Vance, C.C.A. Spencer, G. Stewart, D.A. Hafler, A.J. Ivinson, H.F. Harbo, S.L. Hauser, P.L. De Jager, A. Compston, J.L. McCauley, S. Sawcer, G. McVean, Class II HLA interactions modulate genetic risk for multiple sclerosis, Nat. Genet. 47 (2015) 1107. https://doi.org/10.1038/NG.3395.

[61] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (2017) W24–W29. https://doi.org/10.1093/NAR/GKX346.

[62] H. FF, S. L, S. C, B. M, S. C, Pyruvate Substitutions on Glycoconjugates, Int. J. Mol. Sci. 20 (2019). https://doi.org/10.3390/IJMS20194929.

[63] M.E. Mead, S.L. Knowles, H.A. Raja, S.R. Beattie, C.H. Kowalski, J.L. Steenwyk, L.P. Silva, J. Chiaratto, L.N.A. Ries, G.H. Goldman, R.A. Cramer, N.H. Oberlies, A. Rokas, Characterizing the Pathogenic, Genomic, and Chemical Traits of Aspergillus fischeri , a Close Relative of the Major Human Fungal Pathogen Aspergillus fumigatus , MSphere. 4 (2019). https://doi.org/10.1128/MSPHERE.00018-19/SUPPL_FILE/MSPHERE.00018-19-SF007.PDF.

[64] B.R. Shruthi, R.N.H. Achur, T. Nayaka Boramuthi, Optimized Solid-State Fermentation Medium Enhances the Multienzymes Production from Penicillium citrinum and Aspergillus clavatus, Curr. Microbiol. 2020 779. 77 (2020) 2192–2206. https://doi.org/10.1007/S00284-020-02036-W.

[65] T. Jin, X. Xu, J. Fang, N. Isik, J. Yan, J.A. Brzostowski, D. Hereld, How human leukocytes track down and destroy pathogens: lessons learned from the model organism Dictyostelium discoideum, Immunol. Res. 2008 431. 43 (2008) 118–127. https://doi.org/10.1007/S12026-008-8056-7.

[66] M.P. Hosking, T.E. Lane, The pathogenesis of murine coronavirus infection of the central nervous system, Crit. Rev. Immunol. 30 (2010) 119–130. https://doi.org/10.1615/CRITREVIMMUNOL.V30.I2.20.

[67] S.P. Templeton, S. Perlman, Pathogenesis of acute and chronic central nervous system infection with variants of mouse hepatitis virus, strain JHM, Immunol. Res. 39 (2007) 160–172. https://doi.org/10.1007/s12026-007-0079-y.

[68] J.J. Phillips, M.M. Chua, S.H. Seo, S.R. Weiss, Multiple regions of the murine coronavirus spike glycoprotein influence neurovirulence, J. Neurovirol. 7 (2001) 421. https://doi.org/10.1080/135502801753170273.

[69] SARS-CoV-2 infection and persistence throughout the human body and brain, (2021). https://doi.org/10.21203/RS.3.RS-1139035/V1.

Figures

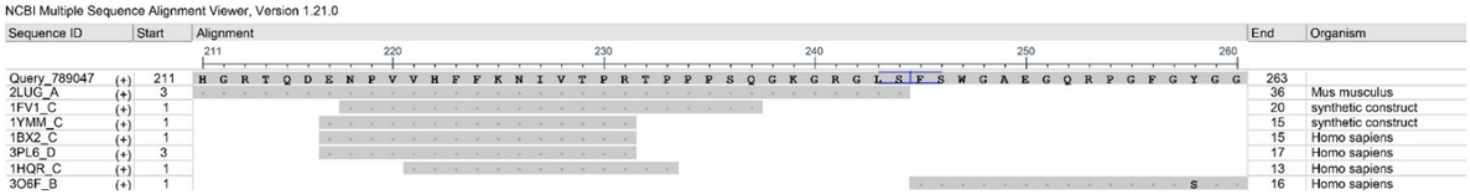


Figure 1

Current PDB structures related to a portion of the MBP sequence. From 304 amino acids of MBP (isoform 1 of MBP, Golli-MBP1), three-dimensional structure of this protein is incomplete in PDB database, in other words, only a few amino acids of this

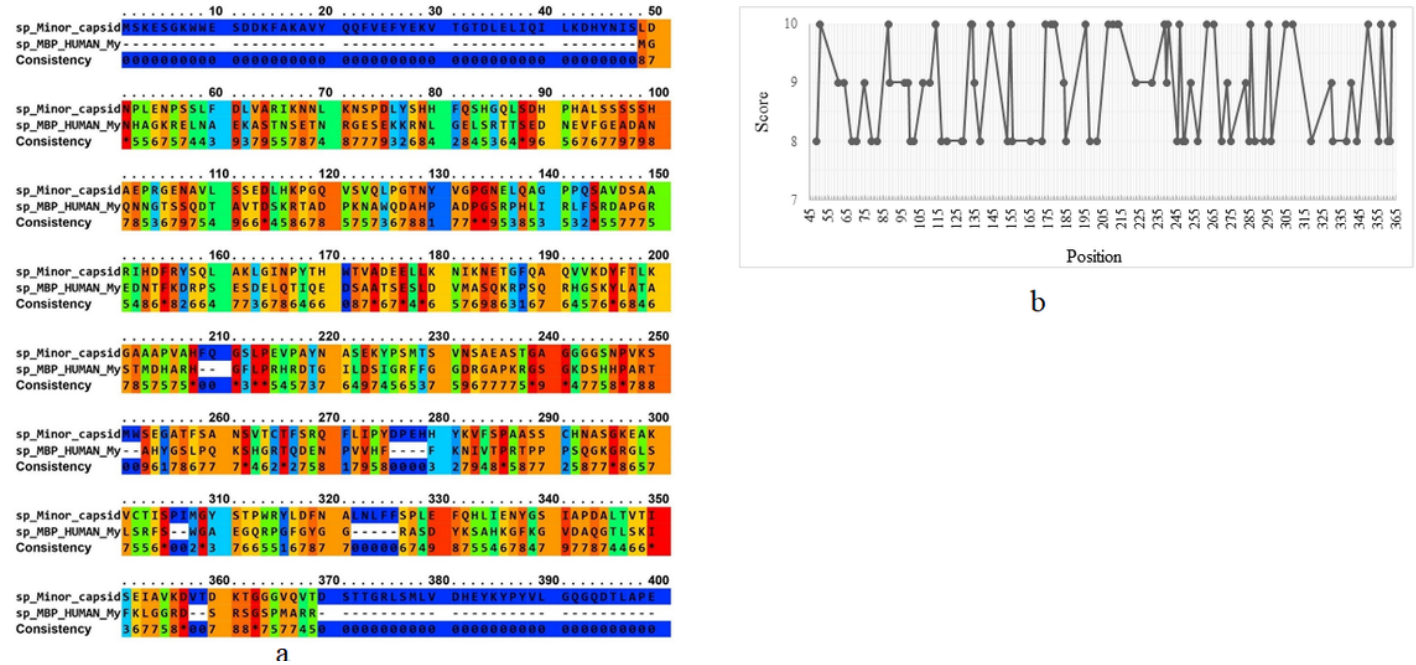


Figure 2

a) Alignment of MBP with the HPV-B19 viral capsid. There are three lines in image, and each line contains 50 characters. The viral capsid amino acid sequence is on the first line, the MBP amino acid sequence is on the second line, and the scores associated with the amino acid alignment between the two

sequences are on the third line. In amino acid alignment, the colors blue and red are used to represent the gap and identity mode, respectively. **b)** The graph of Alignment. According to a section of this fig, the horizontal axis shows alignment position and the vertical axis is related to scoring above 8 on alignment.

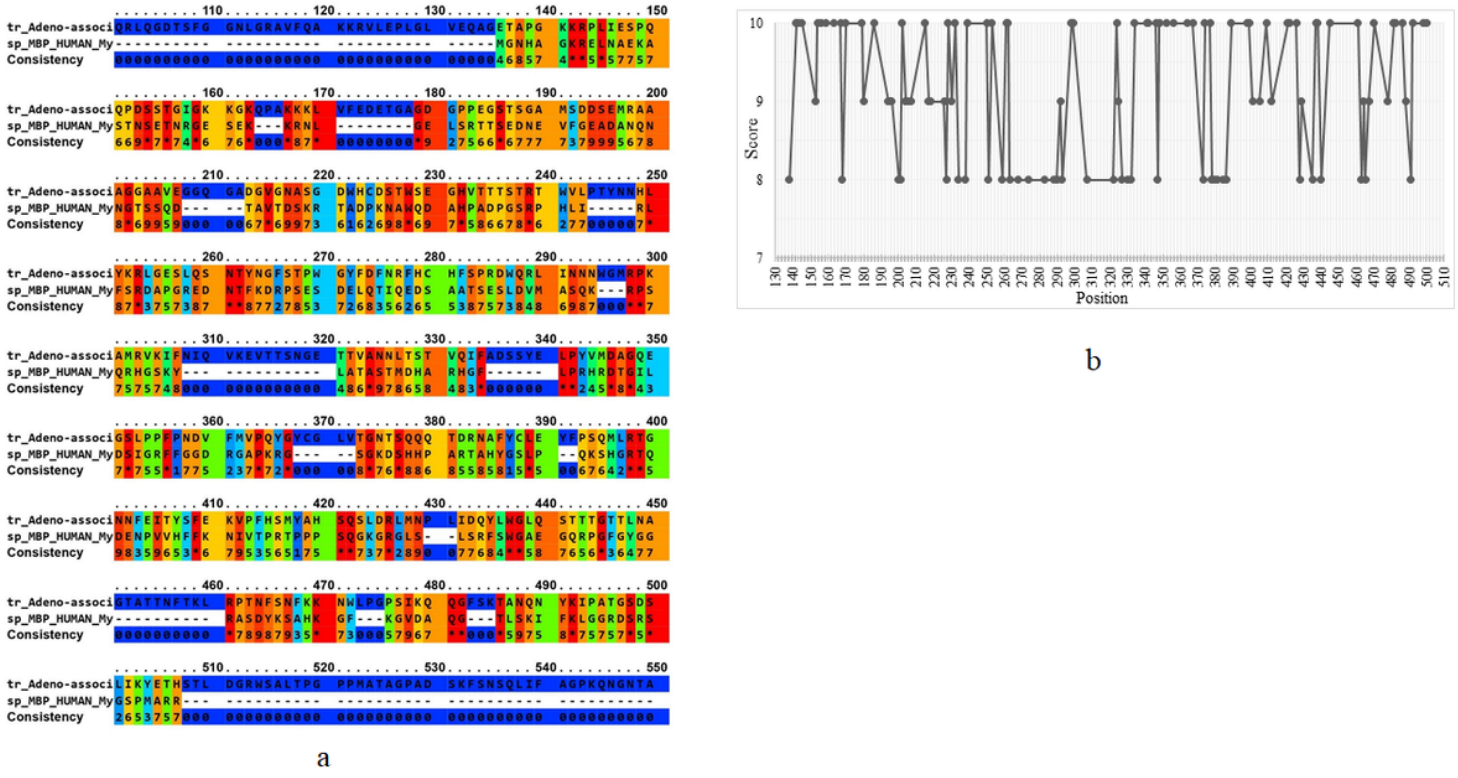


Figure 3

a) AAV-4 viral capsid coat alignment with MBP sequences**.** The sequences below are listed in three lines with 50 characters in each: The first line contains the amino acid sequence of the viral capsid coat, whereas the second line contains the amino acid sequence of the MBP. The third line depicts the consistency of each amino acid alignment from 0 to 10 (star). The similarity and identity of two amino acids are represented by points, with 0 representing the gap mode and a star representing the identity of two amino acids. **b**) Diagram of alignment. In this part, the alignment diagram of section A is drawn, which depicts the horizontal and vertical axes of position and score greater than 8 between sequences, respectively.

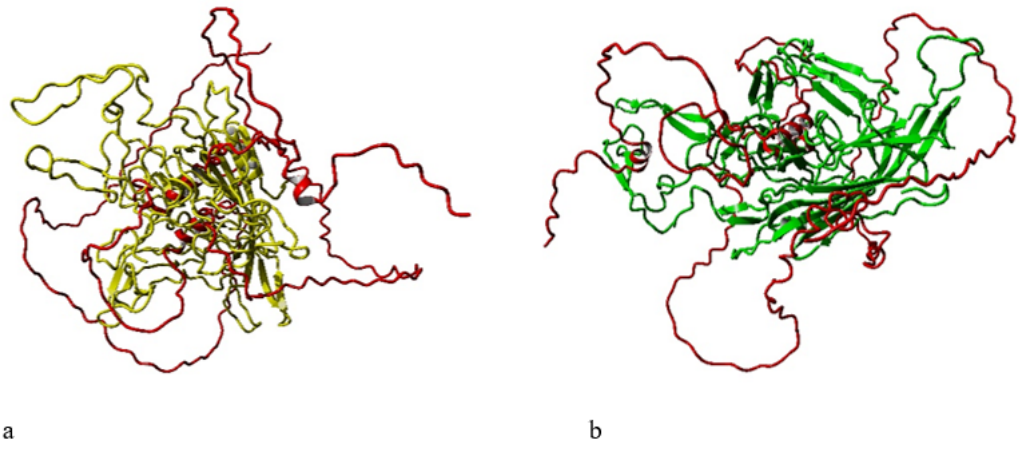


Figure 4

a) Alignment of tertiary structure between HPV-B19 capsid and MBP. **b)** Alignment of tertiary structure between AAV-4 capsid and MBP.

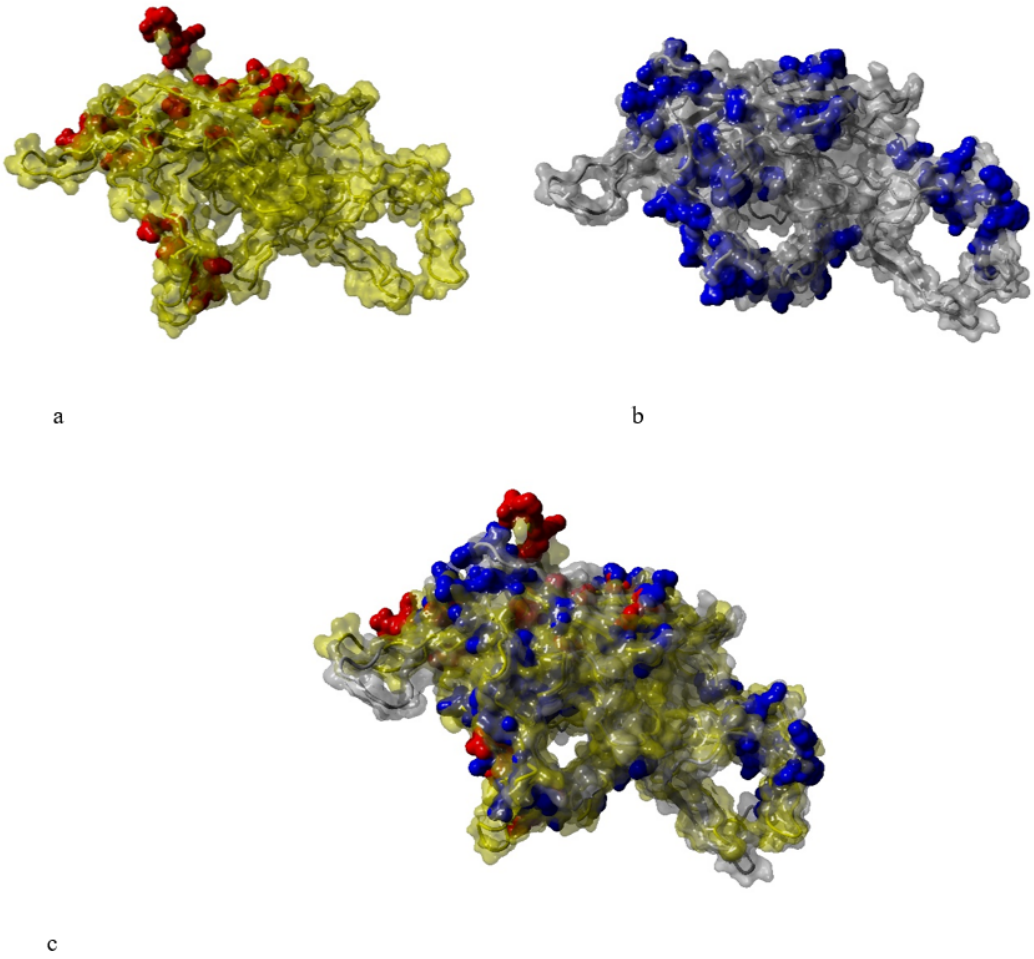


Figure 5

depicts the 3D structure of virus capsid (with emphasis on aligned positions). **a)** The 3D structure of the HPV-B19 capsid that red states indicate aligned amino acids with a score greater than 8. **b)** The blue states in the 3D structure of the AAV-4 virus capsid represent amino acids in this protein's sequence that are aligned with the MBP sequence. **c)** Alignment of the two viruses' capsid structures for overlap in the blue and red areas.

Mycolicibacterium paratuberculosis	121	P	G	S	R	P	H	L	I	R	L	F	S	R	D	132
Epstein-Barr virus	400	P	G	R	R	P	F	F	H	P	V	G	E	A	D	413
MBP	85	P	G	R	R	P	F	T	R	K	E	L	Q	_	_	98
Consistency		*	*	7	*	*	5	4	5	6	5	3	7	1	3	

Figure 6

Multiple alignment of MAP_0106, EBNA1, and MBP.

Mycolicibacterium paratuberculosis	18	A	V	V	P	V	L	A	Y	A	A	A	R	L	L	L	_	_	32
Epstein-Barr virus	305	_	A	V	P	V	L	A	F	D	A	A	R	L	R	L	L	E	320
IRF5	424	_	_	R	L	L	L	E	M	F	S	G	E	L	_	_	_	_	434
Consistency		0	1	4	3	7	*	5	4	1	7	6	6	*	0	3	0	0	

Figure 7

Multiple alignment of MAP_4027⁄BOLF1 and IRF5 proteins.

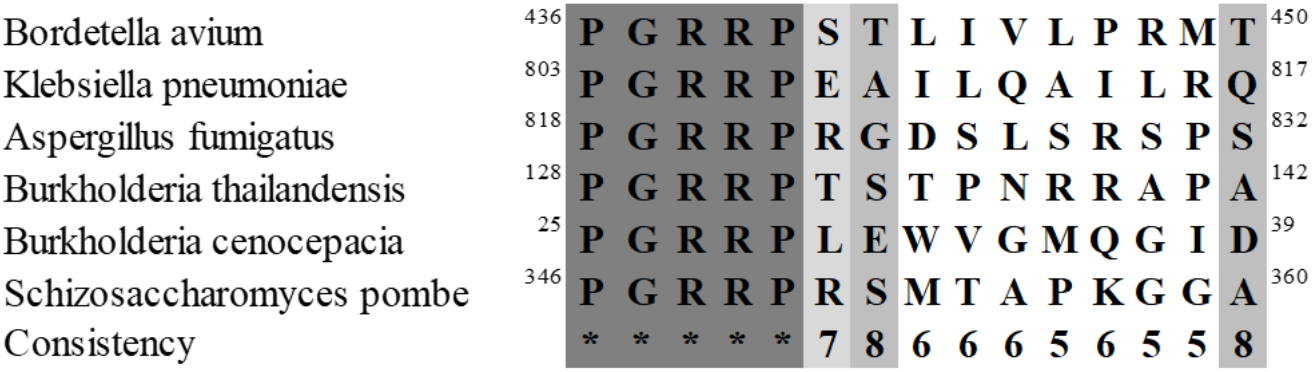


Figure 8

Multiple alignment results of the first group of BLAST protein sequences (EBNA1, MAP 0106, MBP). The amino acids Proline, Glycine, two consecutive Arginine, and Proline can be seen as protected in the sequences in this alignment.

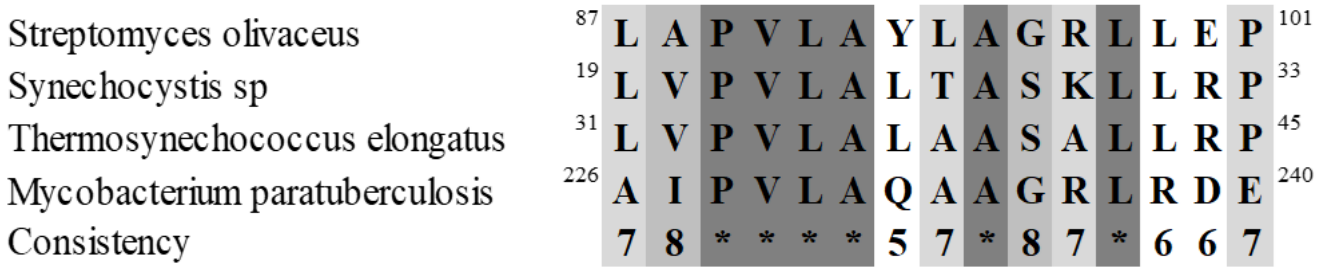
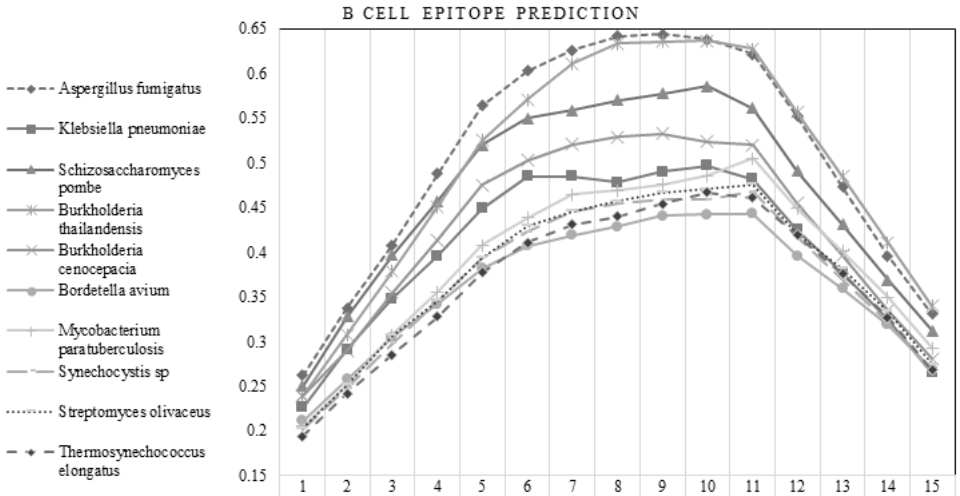


Figure 9

Multiple alignment results of the second group of BLAST protein sequences (BOLF1, IRF5, MAP4027). The amino acids Proline, Valine, Leucine, two Alanine, and Leucine can be seen as protected in the sequences in this alignment, respectively.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Aspergillus fumigatus	0.262	0.338	0.408	0.488	0.564	0.603	0.626	0.642	0.644	0.638	0.622	0.552	0.473	0.396	0.331
Klebsiella pneumoniae	0.226	0.291	0.348	0.396	0.449	0.485	0.485	0.478	0.49	0.497	0.482	0.425	0.377	0.324	0.265
Schizosaccharomyces pombe	0.25	0.328	0.397	0.457	0.52	0.55	0.559	0.57	0.578	0.586	0.561	0.491	0.431	0.369	0.312
Burkholderia thailandensis	0.24	0.308	0.379	0.451	0.525	0.571	0.611	0.634	0.636	0.637	0.628	0.557	0.485	0.411	0.34
Burkholderia cenocepacia	0.239	0.29	0.355	0.414	0.475	0.503	0.521	0.529	0.533	0.524	0.52	0.455	0.397	0.335	0.281
Bordetella avium	0.211	0.258	0.305	0.342	0.382	0.407	0.419	0.429	0.441	0.442	0.443	0.396	0.359	0.319	0.269
Mycobacterium paratuberculosis	0.205	0.252	0.308	0.356	0.408	0.439	0.465	0.469	0.476	0.486	0.505	0.449	0.401	0.35	0.293
Synechocystis sp	0.202	0.248	0.298	0.346	0.394	0.423	0.446	0.455	0.459	0.459	0.468	0.418	0.37	0.322	0.266
Streptomyces olivaceus	0.203	0.252	0.307	0.346	0.394	0.43	0.446	0.457	0.467	0.471	0.476	0.421	0.382	0.334	0.275
Thermosynechococcus elongatus	0.194	0.242	0.285	0.328	0.377	0.411	0.431	0.44	0.454	0.467	0.461	0.419	0.376	0.327	0.268

Figure 10

Diagram of the sequences obtained from Table 8 as epitopes of an antigen.

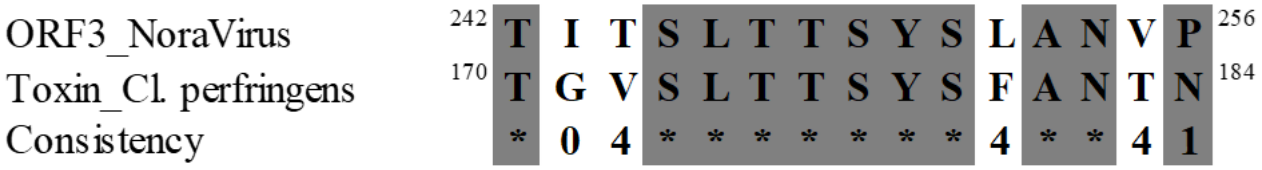


Figure 11

Alignment of two proteins of Clostridium perfringens and Noravirus.

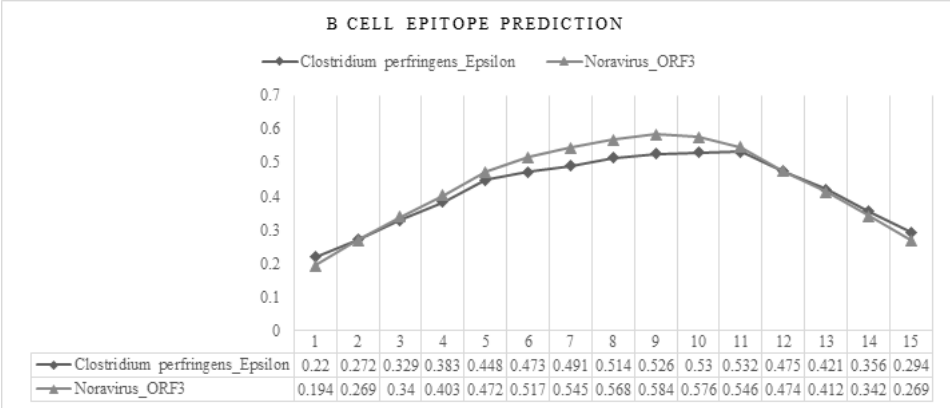


Figure 12

Sequence analysis of epsilon and ORF3 proteins as an antigen epitope.

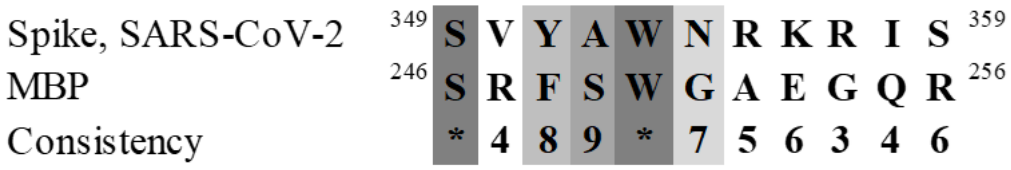


Figure 13

Sequence alignment of MBP 246–256 with spike glycoprotein 349–359. The spike glycoprotein sequence 349-359 is located in its RBD domain.

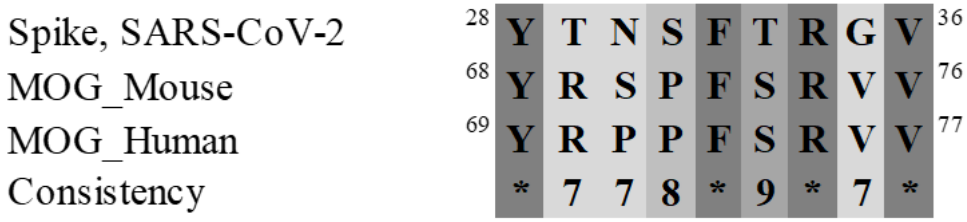


Figure 14

Sequence alignment of 68 to 76 MOG with 28 to 36 spike glycoprotein. Two of the four amino acids (in the Miyauchi et al. study) are conserved, and the amino acids at the beginning and end of this sequences are the same. This sequence is located in the N-terminal of the spike of the virus.

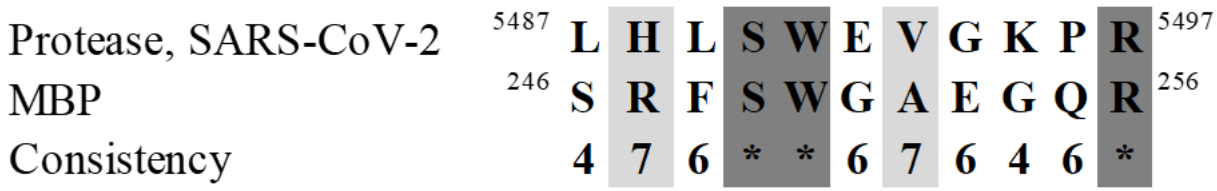


Figure 15

Sequence alignment of MBP 246–256 with main protease 5487–5497. This sequence is found in the Helicase nsp13 chain of the protease.

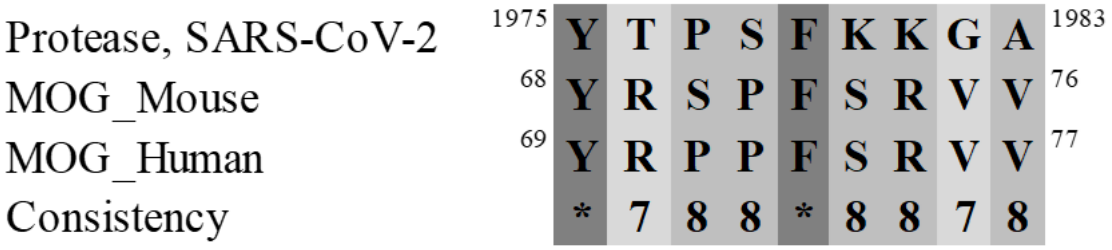


Figure 16

Sequence alignment of 68 to 76 MOG with 1975 to 1983 main protease. All of the amino acids scores are higher than 7. One of the four amino acids (F) in the Miyauchi et al. study is conserved, and the amino acid at the beginning of this sequences are the same. The virus protease sequence from 1975 to 1983 is in the chain of papain-like protease nsp3.

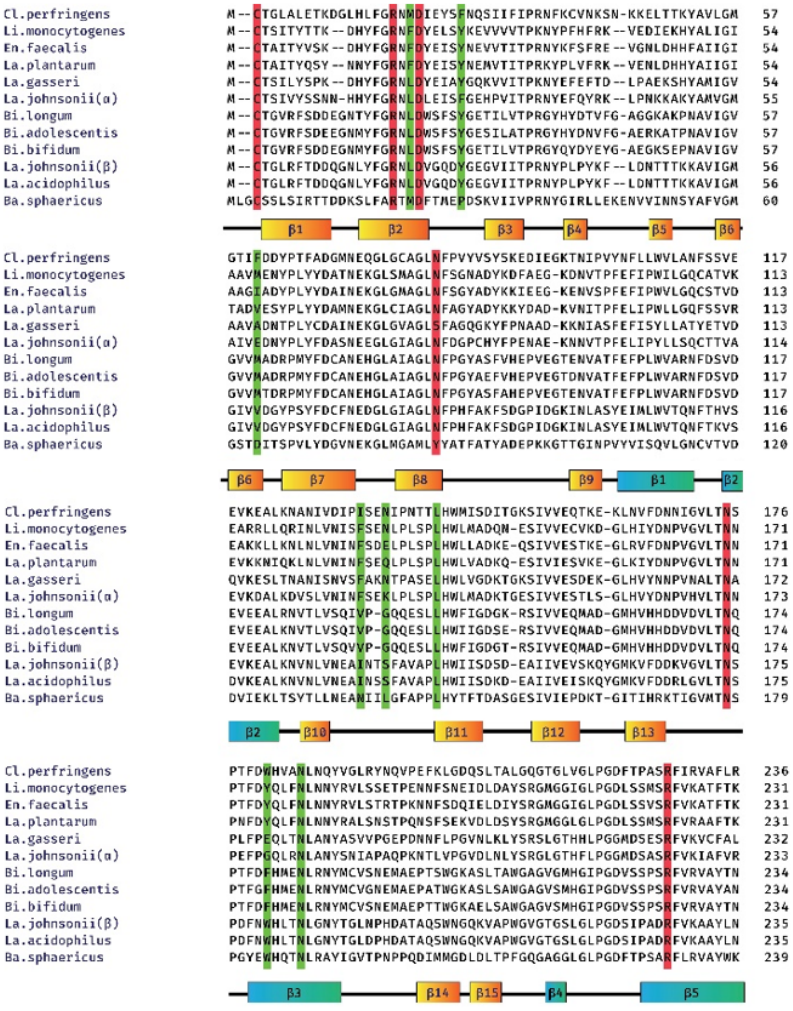


Figure 17

Bacteria and enzymes involved in the butyric acid metabolic pathway (red areas of active sites based on the structure obtained from Cl. perfringens).

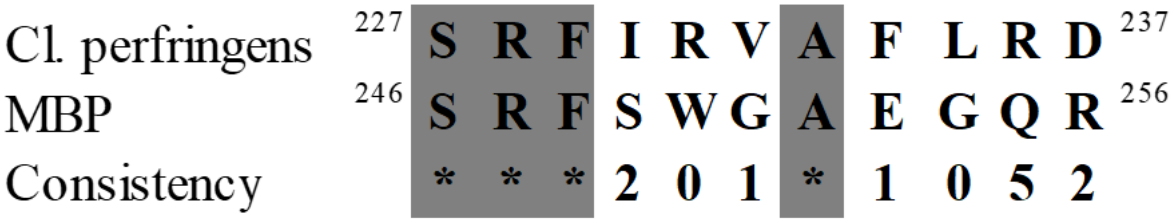


Figure 18

Alignment of Cl.perfringens MBP (from 246 to 256) and BSH enzyme sequences (227-238). According to the alignment of this Figure, the arginine amino acid at position 228 of the bacterial protein sequence is the active site of the enzyme, which can also be seen in red at the fourth position of domain B5 in Figure 17.

Figure 19

Protected points in the alignment three bacterial sequences of Cl. perfringens, Bi. longum, La. plantarum, and MBP. The amino acids in the sequence are arranged as follows: 227–237 for Cl. perfringens, 225–235 for Bi. longum, 222–232 for La. plantarum, and 246–256 for MBP. Furthermore, the amino acid arginine, which is present in the active site of the bacterial enzyme, shares a protected site with the MBP.

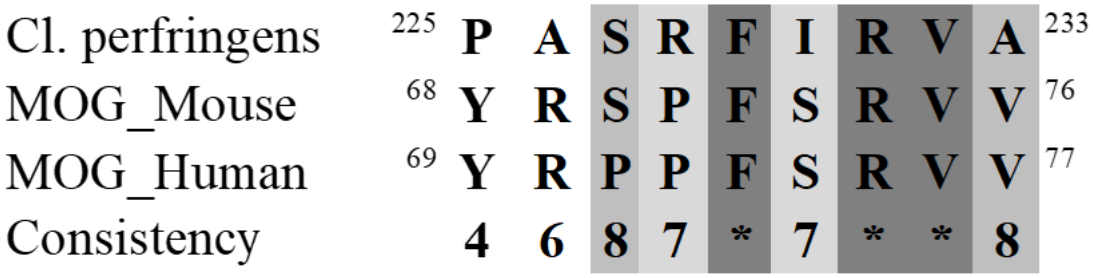


Figure 20

Alignment of human and mouse MOG with BSH protein from Cl.perfringens. Seven amino acids out of nine amino acids that are aligned, have a score above seven, and three amino acids out of four amino acids (in the study of Miyauchi et al.) are conserved (F, R, V).

Cl. perfringens	225	P	A	S	R	F	I	R	V	A	233
Bi. longum	223	S	P	S	R	F	V	R	V	A	231
La. plantarum	220	S	M	S	R	F	V	R	A	A	228
MOG_Mouse	68	Y	R	S	P	F	S	R	V	V	76
MOG_Human	69	Y	R	P	P	F	S	R	V	V	77
Consistency		3	3		5	*	5	*	8	7	

Figure 21

Multiple alignment of MOG with Cl. perfringens, Bi. longum, La. plantarum. Two of the four default amino acids are protected amino acids. The bacterial enzyme's active site is the amino acid Arginine, which is similar to the amino acid proline in human and mouse proteins.

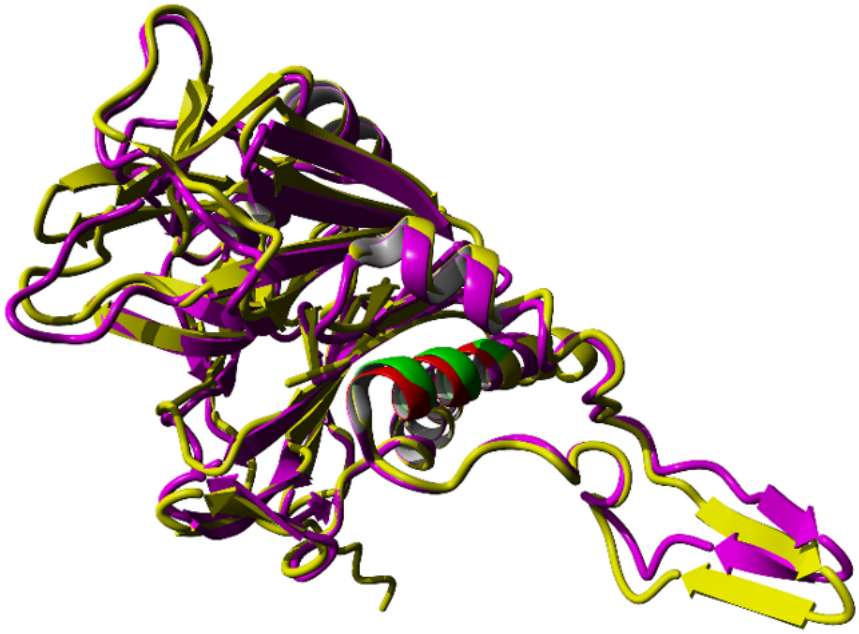


Figure 22

Structure alignment of Bi.longum (purple) and Cl.perfringens (yellow).

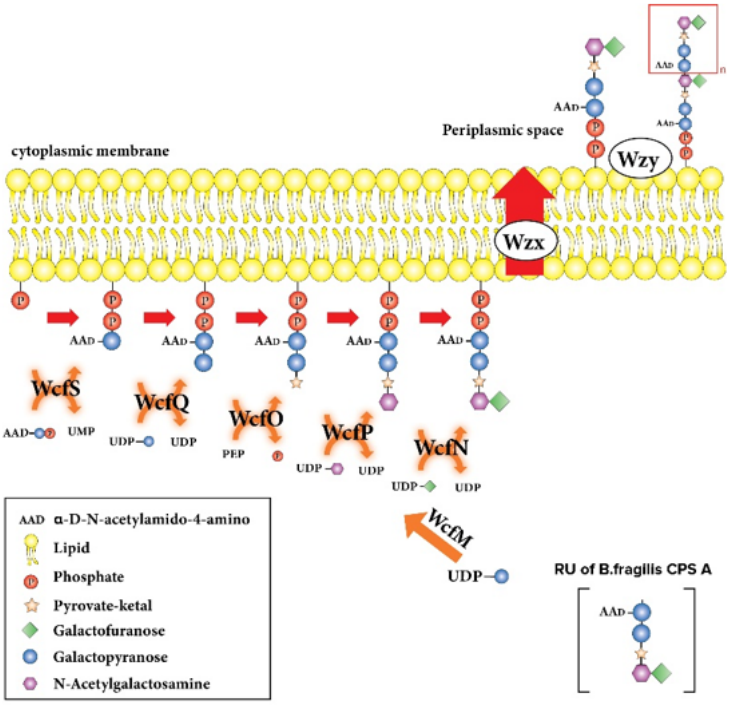


Figure 23

Metabolism of the related enzyme to Polysaccharide A production.

Wzy	311	A	I	E	A	F	S	R	I	M	319
MOG_Mouse	68	Y	R	S	P	F	S	R	V	V	76
MOG_Human	69	Y	R	P	P	F	S	R	V	V	77
Consistency		5	3	3	5	*	*	*	9	6	

Figure 24

Relationship between MOG and the Wzy enzyme. Four amino acids in this alignment have scored above 7. Phenylalanine, Serine, and Arginine are conserved. The protected position in the Phenylalanine and Arginine amino acids, as well as the high similarity of Valine and Isoleucine, can be seen in this alignment, that these amino acids had been conserved in Miyauchi et al study.

WcfQ	163	R	R	S	Y	F	E	K	A	G	171
MOG_Mouse	68	Y	R	S	P	F	S	R	V	V	76
MOG_Human	69	Y	R	P	P	F	S	R	V	V	77
Consistency		5	*	5	4	*	5	7	6	3	

Figure 25

MOG (Mouse and Human) alignment with the bacterial enzyme WcfQ. Three out of the nine amino acids have scored higher than 7 and are highlighted in Figure. Also, two amino acids out of four (according to Miyauchi et al study) are conserved (R, F).