The complex relationship between infections and the development of autoimmune diseases

Arturo Bayo Martín

Master's degree in Bioinformatics and Biostatistics

Drug Design and Structural Biology

Luis Franco Serrano
Nuria Pérez Álvarez

Universitat Oberta de Catalunya

UNIVERSITAT DE BARCELONA

January 15th, 2023

FICHA DEL TRABAJO FINAL (SPANISH)

Título del trabajo:	***The complex relationship between infections and the development of autoimmune diseases***
Nombre del autor:	*Arturo Bayo Martín*
Nombre del consultor/a:	*Luis Franco Serrano*
Nombre del PRA:	*Nuria Pérez Álvarez*
Fecha de entrega (mm/aaaa):	*01/2023*
Titulación o programa:	*Master en Bioinformática y Bioestadística*
Área del Trabajo Final:	*Drug Design and Structural Biology*
Idioma del trabajo:	*Inglés*
Keywords	*Autoimmune diseases development* *Autoantigens epitopes homology* *Infections related with autoimmune diseases*

Resumen del Trabajo

Las enfermedades autoinmunes (EA) ocurren cuando el sistema inmune ataca las células sanas del cuerpo. Algunas de las EAs más comunes son: artritis reumatoide, lupus, celiaquía, esclerosis múltiple, diabetes tipo 1, y psoriasis. Estas enfermedades aparecen de hecho más comúnmente de lo que la gente piensa: las estimaciones indican que alrededor de un 7.5% de la población de Estados Unidos y un 4% de la población global sufren de al menos una enfermedad autoinmune. La prevalencia de las EAs parece estar aumentando también, cada año la incidencia y prevalencia mundial de enfermedades autoinmunes se incrementa en hasta un 19.1%, siendo más comunes en mujeres que en hombres (2-3 veces mayor propensión en mujeres). La causa exacta de las EAs es desconocida, pero se piensa que una combinación de factores genéticos, ambientales, y del sistema inmunitario pueden contribuir a su desarrollo. Los tratamientos actuales para las EAs normalmente incluyen medicamentos que reprimen el sistema inmunitario, pero ya que estos pueden tener efectos secundarios significativos, los investigadores están trabajando en desarrollar tratamientos más dirigidos y efectivos, que puedan dirigirse específicamente a las células involucradas en las EAs sin suprimir el sistema inmunitario por completo. Algunos estudios sugieren que las infecciones con ciertos tipos de bacterias o virus pueden estar asociadas a un mayor riesgo de desarrollar trastornos autoinmunes. La posibilidad de confirmar esa relación puede llevar al desarrollo de tratamientos más avanzados y específicos para algunos trastornos autoinmunes. Este trabajo explora esa relación para las EAs más comunes, analizando la homología de los epítopos conocidos contra el proteoma completo de todos los organismos procariotas.

Abstract

Autoimmune diseases (ADs) occur when the immune system attacks the body's own healthy cells. Some of the most common ADs are: rheumatoid arthritis, lupus, celiac disease, multiple sclerosis, type 1 diabetes, and psoriasis. These diseases are indeed more common than most people realize, with estimates indicating that around 7.5% of the population in the United States and 4% of the global population suffer from at least one autoimmune disease. The prevalence of ADs appears to be increasing too, with yearly increases in the overall worldwide incidence and prevalence of autoimmune diseases of up to 19.1%, being more common in women than men (2-3 times higher propensity in women). The exact cause of ADs is not fully understood, but it is thought that a combination of genetic, environmental, and immune system factors may contribute to their development. Current treatments for ADs often involve medications that suppress the immune system, but since these can have significant side effects, researchers are working to develop more targeted and effective treatments that can specifically target the cells or pathways involved in ADs without suppressing the immune system as a whole. Some studies suggest that infections with certain types of bacteria or viruses may be associated with an increased risk of developing ADs. Confirming that relationship may lead to the development of more advanced, targeted treatments for certain ADs. This work explores that relationship for the most common ADs by analyzing their known epitopes against the complete proteome of all prokaryotic organisms.

Table of contents

1. Introduction

Autoimmune diseases (AD) occur when the immune system attacks the body's own healthy cells. These diseases are more common than most people think: A study from the U.S. Department of Health and Human Services (Office on Women's Health) in 2010 [1] indicated that 23.5 million people in the United States (approximately 7.5% of the population in the year when the study was published) suffer autoimmune diseases. A more recent article published by Nature in 2021 [2] estimates that 4% of the world population (313 million people in 2021) have at least one autoimmune disease. For reasons yet unknown, the prevalence of autoimmune diseases is increasing. Some examples of autoimmune diseases are rheumatoid arthritis, lupus, celiac disease, multiple sclerosis, type 1 diabetes, alopecia areata, and vasculitis. There are epidemiological studies for most of these diseases that help to understand the frequency, distribution, main determinants, and events related with the disease that allow the causes to be delimited to a certain extent, and figure out if the influence of genetic, epigenetic, environmental factors, viruses, or infections are the main cause of those autoimmune diseases.

Out of the affected population, autoimmune diseases in women have between two and three times more propensity than men [3]. According to a study from 2000 [4], 80% of the population affected by autoimmune diseases in the United States were women, although more recent studies have shown that figure really depends on the disease.

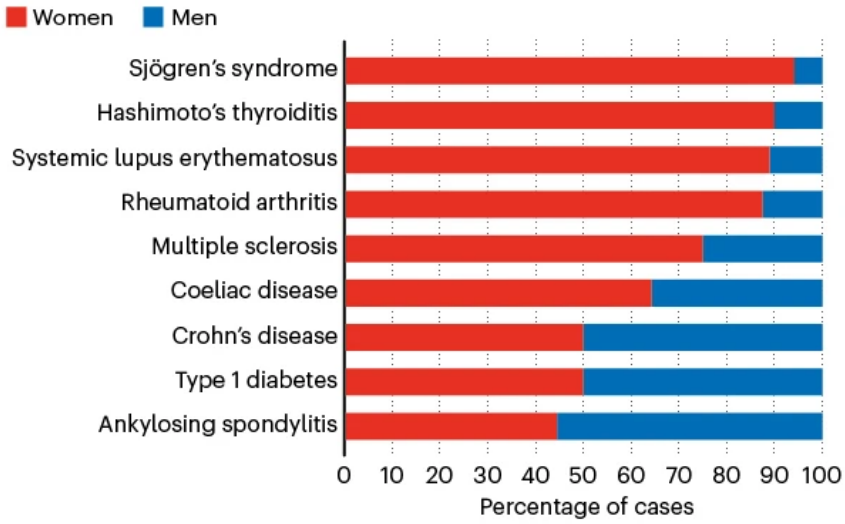


Figure 1: Study of sex differences in common autoimmune diseases

Investigators haven't yet figured out what exactly cause autoimmune diseases. The current agreement is that a combination of genetic susceptibility and environmental influence such as infections, diet, and exposure to chemicals may play an important role on the development of autoimmune diseases.

1.1. Context and work rationale

There's significant evidence [5] that suggests that different classes of pathogens (bacteria, viruses, and parasites) are involved in the development or propagation of autoreactive

immune responses. However, the evidence of an actual link between autoimmunity and a previous infection is stronger on certain diseases than others.

Bacteria and viruses trigger a similar immune response, suggesting that the antibodies produced as a response to certain infections may also attack some healthy cells (autoantibody), because they resemble to those bacteria or viruses that cause the infection and bind to specific molecules, called autoantigens, to initiate the immune response. Some investigators suggest that infections can actually damage the autoimmune system, leading to developing autoimmune disorders.

These autoantigens may be various molecules, including proteins, enzymes, and hormones. They are typically found on the surface of the cells, in the blood, or in other fluids. The specific autoantigens targeted in each autoimmune disease may vary. For instance, in rheumatoid arthritis, the immune system targets proteins in the joints, causing inflammation and joint damage. In multiple sclerosis, the immune system attacks proteins in the myelin-based protective covering on nerve fibers.

The specific site on an antigen (and autoantigen) molecule that is recognized and targeted by the immune system is known as epitope. This region is usually composed of a small number of aminoacids and is typically located on the surface of the molecule. Epitopes are usually highly conserved, meaning that they don’t change much over time, allowing the immune system to recognize and target them effectively.

The prevalence of autoimmune diseases varies depending on the specific disease and the population being studied. Despite anecdotal and clinical implications that microbial pathogens act as triggers for autoimmune diseases, little is known about the mechanisms by which infectious agents actually cause these diseases.

1.2. Work goals

This work will explore the relationship between the known epitopes of the most common autoimmune diseases and the proteome of known prokaryote organisms. There are X main goals that this work will aim to achieve:

1. Study the prevalence of autoimmune diseases in different populations, including the demographics and geographic distribution of these diseases.

2. Investigate the role of previous bacterial and viral infections in the development of these autoimmune diseases.

3. Explore the mechanisms by which autoimmune diseases may develop, including the role of autoantigens and autoantibodies in the immune system's response to these substances.

4. Identify and characterize the specific autoantigens and epitopes targeted in different autoimmune diseases.

1.3. Impact on sustainability, ethical-sociability, and diversity

Conducting research on autoimmune diseases has a number of potential impacts on sustainability, ethical-sociability, and diversity:

In terms of sustainability, advances in the understanding and treatment of autoimmune diseases would lead to more effective and efficient use of healthcare resources, potentially reducing the overall burden on the healthcare system. Also, the development of new treatments for autoimmune diseases may improve the quality of life of individuals affected by these diseases, potentially reducing the need for long-term care and supporting more independent living.

From an ethical-sociability perspective, research on autoimmune diseases could help to improve the lives of individuals with these conditions and reduce the societal stigma often associated with these diseases.

In terms of diversity, it is important to ensure that the results of the research are disseminated and applied in a way that benefits all members of society and not just a particular group. The results of this work apply equally to millions of people worldwide, independently of the origin, gender, and social backgrounds of the individuals affected by these diseases.

1.4. Approach and methodology

There are many different approaches that could be taken to carry out research on autoimmune diseases. This work will be developed following the principles of the scientific method, based on research, observation, measurement, analysis, and formulation of hypotheses to answer the questions raised and write a report with the findings in this document.

Following these steps guarantees the incremental construction of the necessary solid foundations in order to tackle each subsequent step, that will allow for the evaluation and formulation of more probable hypotheses.

1.5. Work plan

Sprint	Start Date	End Date	Objective	Milestones
1	**Oct 24**	**Nov 7**	**Research and collect information on autoimmune diseases**	**Completed the Context and State of the art sections in the memory**
1- Collect information on academic search engines, articles, scientific studies, and bibliographies. 2- Organize the information and describe each of the identified autoimmune diseases 3- Construct a comparative table between the studied diseases that contains at least the frequency, distribution, main determinants, and relationship with previous infections. 4- Search for information about the relationship of each of these diseases with viral or bacterial infections 5- Write the introduction, context, and state of the art in the memory				
2	**Nov 7**	**Nov 21**	**Narrow list of autoimmune diseases, identify object of study, and study relationship with viral or bacterial infections**	**Completed the first version of Materials and methods, and Results sections.**

1- Select a disease based on existing clues about viral or bacterial origin 2- Search databases for the epitopes of its pathogens, for example ViPR 3- Identify the amino acid sequence of the epitopes 4- Translate the amino acid sequence to nucleotides 5- Search for alignment with the nucleotide sequence in databases of microorganisms, using tools such as Clustal and BLAST 6- Analyze the alignment results 7- Start the study of the possible relationship with known viruses, bacteria, and parasites 8- Review progress of work with mentor				
3	**Nov 21**	**Dec 5**	**Reflect on results based on selected disease and continue with study**	**Completed the second version of the Materials and methods, and Results sections.**
1- Decide if the selected disease is a candidate to provide significant results for the work 2- Consider the selection of another disease if sufficient evidence is not found to perform a study that provides significant results 3- Continue or start with the study of the selected autoimmune disease, iterating on the study of the located epitopes and their alignment with the proteome of microorganisms 4- Review and refine the sections drafted in the previous sprint				
4	**Dec 5**	**Dec 19**	**Narrow autoimmune diseases related to viral or bacterial infections and write final report and conclusion**	**Completed the final version of the report and conclusion sections.**
1- Reflect on the results obtained so far and identify the autoimmune diseases that may be related to viral or bacterial infections 2- Write the final report and conclusion 3- Review and refine the sections drafted in the previous sprints 4- Prepare the presentation				
5	**Dec 19**	**Dec 23**	**Review work with tutor to focus on closing the report**	**Draft final report reviewed and areas for improvement identified**
1- Review progress of work and memory 2- Identify potential areas for improvement 3- Plan implementation of improvements in the home stretch				
6 *(Final)*	**Dec 27**	**Jan 15**	**Finalize and close the report**	**Polished final version of the report**
1- Implement improvements agreed upon with tutor in review 2- Proofread and format check 3- Review and complete bibliographic references used throughout the work 4- Review the content of the memory to ensure it follows a consistent argumentative thread 5- Final review with tutor prior to submission 6- Wrap up the final version of the document				

1.6. Brief summary of product output

An analysis of the relationship between the most impactful autoimmune diseases among the most common, and previous infections of virus and bacteria.

2. State of the art

A healthy immune system defends the body against disease and infection, by responding to invading microorganisms, such as bacteria and viruses. But if the immune system malfunctions, it may mistakenly attack healthy cells, including organs and tissues. An autoimmune disease (AD) occurs when the immune system mistakenly attacks and destroys the body's own healthy cells, weakening body functions and even turning life-threatening.

According to the list available on Global Autoimmune Institute's site [6], there are 157 known autoimmune diseases in 2022. Some of them are well known, such as type 1 diabetes, celiac disease, and multiple sclerosis. The scientific community does not yet fully understand the immune system and what causes the body to produce an immune response to itself, and the diagnosis of these diseases is often difficult to determine.

However, it is known that there are several triggers and risk factors which play a role in developing an autoimmune disease, for example infections, exposure to toxins, genetic heritage, and certain side effects from some medications.

Antibodies are a key part of the immune response, and they also play a central role in the development of autoimmune diseases. In particular, antibodies that specifically react with self-antigens (called autoantibodies) are generated as a result of the loss of tolerance response against self-antigens and can be pathogenic. Autoantibodies are often used as diagnostic markers for these conditions. The presence of high levels of specific autoantibodies in the blood can be a strong indication that a person has an autoimmune disease, and many autoimmune diseases are associated with characteristic patterns of autoantibody production. For example, people with rheumatoid arthritis often have high levels of autoantibodies directed against the body's own joint tissue, while people with multiple sclerosis often have autoantibodies directed against the central nervous system.

In recent years, interest has shifted to uncover the target epitopes of these autoantibodies. Understanding the specific epitopes that are targeted by autoantibodies can provide valuable insights into the mechanisms of autoimmune diseases and the development of new treatments.

3. Materials and methods

4.1. Exploration of the most common autoimmune diseases

The consensus on which ADs are most common varies. It is likely based on patient-reported information, physician experience in-clinic, hospital data, and research studies. Based on current information [7], some of the most common autoimmune diseases include:

1. Type 1 diabetes (T1D)
2. Rheumatoid arthritis
3. Psoriasis/psoriatic arthritis
4. Multiple sclerosis
5. Systemic lupus erythematosus (SLE)
6. Inflammatory bowel disease (Crohn's Disease)
7. Celiac disease
8. Scleroderma
9. Grave's disease
10. Sjögren's syndrome

3.1.1 Type 1 diabetes (T1D)

Type 1 Diabetes (T1D) is a chronic autoimmune disease that inhibits the pancreas to produce enough insulin, the main anabolic hormone of the body that regulates the metabolism of carbohydrates, fats, and protein by promoting the absorption of glucose from the blood stream into liver, muscle cells, and adipose tissue. As a result, the body would suffer an autoimmune reaction that destroys the cells in the pancreas that produces the insulin, called beta cells. This leads to a very low level of insulin that results in high blood glucose level and low glucose mobilization to organs and tissues. Utilization of glucose is hampered, and even if there is glucose in the blood stream, the body can't utilize it, leading to an excess amount of glucose in the blood which might have severe consequences (such as increased risk of heart disease and stroke, kidney disease, vision problems, and nerve problems).

According to a recent study published on The Lancet [8] in 2022, there are about 8.4 (95% uncertainty interval, so ranging between 8.1 and 8.8) million individuals worldwide with type 1 diabetes. Out of these, 1.5 million (18%) were younger than 20 years, 5.4 million (64%) were aged 20–59 years, and 1.6 million (19%) were aged 60 years or older.

There are several studies [9] (2008) that suggest a relation between previous infection and type 1 diabetes development. Enteroviruses such as Coxsackievirus B (CVB) (10), but also rotavirus [11][12], mumps virus [13], and cytomegalovirus [14] have been associated with type 1 diabetes.

3.1.2 Rheumatoid arthritis

Rheumatoid arthritis (RA) [15] is a chronic autoimmune disease which produces joint inflammation, pain, deformity, and movement difficulty. On individuals affected by RA,

the immune system mistakenly sends antibodies to the lining of joints, where they attack the tissue surrounding the joint, causing the thin layer of cells (*synovium*) covering the joints to become sore and inflamed, releasing chemicals that damage nearby bones, cartilage, tendons and ligaments. The most frequently inflamed joints are the wrists, fingers and toes, elbows, shoulders, hips, knees, and ankles. According to the United States' Center for Disease Control and Prevention (CDC) [16], the specific causes of RA are unknown.

RA has an incidence of about 0.25% worldwide, according to the Global Burden of Disease study in 2010 [17], and 0.5 to 1% on adults in the developed world [18]. Similar to other autoimmune diseases, RA is also twice as common in women than in men.

This autoimmune disease is difficult to diagnose, there's no diagnostic criteria specific to RA and nor pathognomonic finding for RA either.

Some studies [19] [20] have suggested that certain viruses, such as the Epstein-Barr virus or the human parvovirus, may trigger the immune system to attack the joints and lead to the development of rheumatoid arthritis.

3.1.3 Psoriasis and psoriatic arthritis

Psoriasis is an autoimmune disease that cause inflammation of the body, typically with signs of inflammation such as raised plaques and scales on the skin. An overactive immune system speeds up skin cell growth,

It affects around 125 million people worldwide, around 2% to 3% of the total population [21][22]. One in three people with psoriasis may also develop psoriatic arthritis.

According to a study published by NatureDirect in 2007 [23], there's strong evidence of psoriasis induction by a preceding infection, in particular Streptococcus pyogenes infection, which appears to involve initial superantigenic T cell activation by streptococcal toxins, followed by an antigen-specific T cell response. Exacerbation, rather than induction, of psoriasis has been linked to another Gram-positive bacteria, Streptococcus aureus, which probably also exerts its effects via the production of superantigenic toxins and 2 species of fungi, Malassezia and Candida, and viruses such as HPV5. Another virus that has been linked [24] with psoriasis is the human herpes virus 8 (HHV-8), also known as the Kaposi's sarcoma-associated herpes virus (KSHV). This virus is associated with a type of skin cancer called Kaposi's sarcoma, and some studies have suggested that it may also play a role in the development of psoriasis. Finally, it is believed [25] that the human immunodeficiency virus (HIV), which causes the immune deficiency syndrome AIDS, may increase the risk of developing psoriasis.

3.1.4 Multiple sclerosis

Multiple sclerosis (MS) [26] is an inflammatory and neurodegenerative autoimmune disorder of the brain and spinal cord in which focal lymphocytic infiltration leads to damage of myelin (the fatty substance that coats and protects nerve fibers in the brain and spinal cord) and axons. This damage disrupts the ability of parts of the nervous system to

transmit signals, resulting in a range of signs and symptoms, including physical, mental, and sometimes psychiatric problems.

It affects around 2.8 million people in the world [27]. A study [28] that aggregated data from another 65 studies on multiple sclerosis across 24 countries covering ~3% of the world population indicated that MS prevalence has increased over time.

Although the causes of MS are still unknown, there's a common agreement that the combination of genetic heritage and environmental factors seems to be responsible. However, there are a few studies such as *Cusick MF, Libbey JE, Fujinami RS. Multiple sclerosis: autoimmunity and viruses 2013* [29] that explore the involvement of viruses in multiple sclerosis pathogenesis, and concludes that although it is unlikely that any one microbe will be determined to be the causative agent of MS, complex interactions between multiple viral infections leading to a variety of viral specificities may result in the activation of T cells that recognize self and induce multiple sclerosis. In particular, the study calls out interactions between human endogenous retroviruses and autoreactive $CD8^+$ T cells with dual T-cell receptors as potential triggering mechanisms of the disease.

Another virus that has been linked [30] with multiple sclerosis is the Epstein-Barr virus (EBV), which is a type of herpes virus. In the same way, the human herpes virus 6 (HHV-6), a common cause of childhood infections, was linked in a study on February 2022 [31] to playing a role in developing multiple sclerosis.

3.1.5 Systemic lupus erythematosus (SLE)

SLE is the most common types of lupus. It is an autoimmune disease that occurs when the immune system attacks its own healthy tissues and organs, causing widespread inflammation and tissue damage. The inflammation typically impacts joints and skin but it may also affect the brain, kidneys, blood cells, heart, and lungs.

The incidence of SLE [32] is estimated to be between 2 and 7 per 10000 people in the world, that is, it affects between 1.5 to 5.4 million people worldwide. It is 10 times more common in women than in men.

It has been proven that some virus infections play a role in inducing the disease [33]. In particular, Epstein–Barr virus (EBV), parvovirus B19 (B19V), Human herpesviruses (HHVs), and human endogenous retroviruses (HERVs) are involved in SLE pathogenesis.

3.1.6 Crohn's disease

Chron's disease (CD) is an inflammatory bowel disease that causes parts of the digestive system to become inflamed. It can happen anywhere in the guts. Commonly it tends to affect the middle part of the bowel, and tends to cause the bowel to become thickened, swollen and inflamed.

In 2015, a study [34, table 3] estimated that Crohn's disease affects around 11.2 million people worldwide.

There's a recent study from 2021 [35] that reviews the relationship between viral infections and the development of inflammatory bowel disease. In particular, a study referred there [36] calls out specifically murine norovirus (MNV) to trigger CD on patients with a specific gene mutation (Atg16L1).

3.1.7 Celiac disease

It is an autoimmune disease that damages the small intestine and alters the absorption of vitamin, minerals, and other nutrients. This disorder comes mainly from intolerance to a specific protein called gluten, present in cereals, such as wheat, oats, barley, and rye. It produces symptoms such as digestive problems.

Celiac disease is present in ~1% of the world population, that is about 78 million people worldwide.

Some researchers [37] suggest that a higher number of infections at a young age and certain infections of the digestive tract may increase the risk of contracting celiac disease. Other researchers believe that changes in the microbiome (bacteria in the digestive tract that help with digestion) could influence the development of celiac disease. There is evidence that certain infections may increase the risk of developing celiac disease, according to some studies that have found an association between celiac disease and infections with certain viruses, such as rotavirus or norovirus, and certain types of bacteria, such as Helicobacter pylori and Salmonella.

3.1.8 CREST syndrome (Scleroderma)

CREST syndrome, also known CREST scleroderma, is a subtype of systemic scleroderma that is characterized by certain features. The acronym CREST stands for the following:

- C: Calcinosis, the accumulation of calcium deposits in the skin or internal organs
- R: Raynaud's phenomenon, a condition that causes the blood vessels in the fingers and toes to constrict in response to cold or stress
- E: Esophageal dysfunction, referring to problems with the function of the esophagus
- S: Sclerodactyly, which is the hardening and thickening of the skin on the fingers and toes
- T: Telangiectasia, the formation of small blood vessels that are visible through the skin

Crest syndrome is a rare form of scleroderma that affects about 10-20% of people with systemic scleroderma, about 1.5 million people worldwide. It is more common in women and tends to be less severe than other forms of systemic scleroderma. The course of the disease and the prognosis can vary widely, and treatment options may include medications, physical therapy, and skin care.

Some studies suggest that CREST syndrome is related with bacterial infections, such as Helicobacter pylori [38] and Salmonella; viral infections, such as rotavirus and norovirus; fungal infections, such as Candida; and parasitic infections, such as toxoplasmosis.

3.1.9 Graves' disease

It is the most common form of autoimmune hyperthyroidism. The hormones thyroxine (T4) and triiodothyronine (T3) produced in the thyroid gland increase metabolism, stimulate growth, and have a significant effect on the heart, muscles, and nervous system, as well as on the body's protein, sugar, and ion metabolism. Adequate intake of iodine is essential for the production of thyroid hormones. When they are overproduced, the metabolism becomes excessively accelerated, resulting in the symptoms of Graves' disease, such as tremors in limbs, sleep disturbance, weight loss, light avoidance, muscle weakness, double vision, loss of libido, enlarged thyroid gland, etc.

Graves' disease can appear at any age and in both genders, but it most commonly affects individuals between the ages of 20 and 40. Women are five times more affected than men. It impacts females more frequently than males by a ratio of 5-10 to 1. The condition usually develops during middle age with a peak incidence of 40-60, but it can also affect children, adolescents, and the elderly. It occurs in almost any part of the world and is estimated to affect 2%-3% of the general population.

Graves' disease is an autoimmune disease, but other factors may contribute to its development, including genetic, environmental, or other factors. People with this disease often have a family history of thyroid or autoimmune problems. Some family members might have hyperthyroidism, underactive thyroid, or premature graying of the hair. Similarly, there may be a family history of related immune problems, including juvenile diabetes, pernicious anemia, or vitiligo.

Some research [39] has suggested that certain viruses, such as the Epstein-Barr virus and the cytomegalovirus, may play a role in the development of the disease,

3.1.10 Sjögren's syndrome

The disease is triggered by an autoimmune process in which autoantibodies are produced against the tear and salivary glands, causing them to maintain a constant state of inflammation. As a result of the inflammation, the cells that make up the glands gradually die and are no longer able to do their job, and the symptoms of the disease appear, such as dry eyes, dry mouth, reduced saliva volume, etc.

Sjögren's syndrome is a relatively common disease, affecting one in 100 people, 1% of the population [40]. More than 90% of patients are women, with an average age of 50-60 years, but it is usually more severe in men. Between 400,000 and 3.1 million adults have Sjögren's syndrome. This disease can affect individuals of any age, but symptoms usually appear between the ages of 45 and 55. It has been reported worldwide, and there appears to be no racial or geographic bias in incidence. The female: male ratio is approximately 9:1.

There is no certainty why people develop Sjogren's syndrome. Specific genes put people at higher risk, but infection with a particular virus or bacteria is also necessary to trigger the disease. People with Sjogren's syndrome are expected to also have a rheumatic disease — such as rheumatoid arthritis or lupus. Laboratory analyses implicate viruses, including Epstein-Barr virus (EBV), as participants in disease pathogenesis.

4.2. Selection of autoimmune diseases in scope of the study

The previous section presents data that can be used to create a comparison table for the top 10 most common autoimmune diseases. This table will enable the selection of the most impactful diseases to study, based on frequency, distribution, known epitopes, and any known relationship with viral or bacterial infections according to current research.

Apart from the data gathered from the different studies referenced, the epitopes data will be enriched with entries from the Immune Epitope Database (IEDB), that is referenced in some studies [41] as a valuable data source.

The following table summarizes the autoimmune diseases being considered for this study based on these criteria:

Disease name	**Frequency worldwide**	**Distribution (men/women)**	**Known autoantigens (related human peptides/proteins)**	**Related infections according to studies**
Type 1 diabetes (T1D)	8.4 million	More common under 30	GAD65, IA-2, Insulin, Insulin isoform 2	Enteroviruses, CVB, Rotavirus, and Mumps virus.
Rheumatoid arthritis	19 million (0.25% world population)	2x more common in women than in men	Cardiolipin, Alpha-enolase, Keratin type I cytoskeletal 18	Epstein-Barr virus or the human parvovirus,
Psoriasis/psoriatic arthritis	125 million (2% - 3% world population)	People between age of 15 and 35	Keratin 16 and 17, cathelicidin LL-37, melanocytic ADAMTSL5, lipid antigen PLA2G4D and keratin 17	Streptococcus pyogenes, and HHV-8
Multiple sclerosis	2.8 million	Women twice as likely than men, individuals between 20 and 50	Myelin basic protein (MBP), myelin oligodendrocyte glycoprotein (MOG)	Retroviruses, Epstein-Barr, and HHV-6
Systemic lupus erythematosus (SLE)	1.5 to 5.4 million people worldwide	0 times more common in women than in men	Cardiolipin antigen	Epstein–Barr virus (EBV), parvovirus B19 (B19V), HHV
Inflammatory bowel disease (Crohn's Disease)	11.2 million	Similar in men and women	Flagellin protein, HSP60, and TNF receptor	murine norovirus
Celiac disease	78 million (1% world population)	1.5 to 2 times more common in women than in men	Tissue transglutaminase 2 (tTG 2), Endomysium	Rotavirus, norovirus, Helicobacter pylori and Salmonella
Scleroderma (Crest syndrome)	30 thousand to 2.7 million	Individuals between 30 and 50 years	Topoisomerase I enzyme, U1-70kD ribonucleoprotein (RNP) complex, and centromere	HHV-8, and Borrelia burgdorferi

			protein C (CENP-C)	
Graves' disease	125 million (2% - 3% world population)	Individuals between 20-40 years. 5x times more frequent in women	Thyroid-stimulating hormone receptor (TSHR) and the thyroid peroxidase (TPO) enzyme	Epstein-Barr virus and the cytomegalovirus
Sjögren's syndrome	79 million (1% world population)	90% (9:1 women:men) women between 50 and 60 years	Ro/SSA and La/SSB ribonucleoprotein (RNP) complexes, salivary gland protein 1 (SG1), and carbonic anhydrase 6 (CA6) enzyme	Epstein-Barr virus (EBV), Parvovirus B19

4.3. Methodology

In order to study the relationship between the selected autoimmune diseases and previous infections, researchers typically use information about their epitopes, specific regions on the surface of proteins that can be recognized and bound by the immune system (antibodies).

Therefore, in order to study a potential relationship between developing those autoimmune diseases and having previous bacterial or viral infections, this work will analyze the relationship between the epitopes of those autoimmune diseases and the complete proteome of known bacteria and viruses, using protein sequence alignments, adhering to the following process:

First, a list of autoimmune diseases and their corresponding epitopes needs to be curated. The IEDB is a good source of information to identify the most common epitopes for the list of diseases, and in some cases, it even provides the aminoacid sequence of the epitopes. When IEDB doesn't provide the sequence directly, other data sources such as EMBL-EBI [42] or UniProt [43] will be used to find the sequence for those proteins.

The second step is to compute sequence alignments across multiple bacteria and viruses databases using BLAST [44]. In particular, this work uses Position-Specific Iterated BLAST (PSI-BLAST), selecting the following databases: All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF, that contain 512512737 sequences as of 2022/10/30, as shown in the following picture:

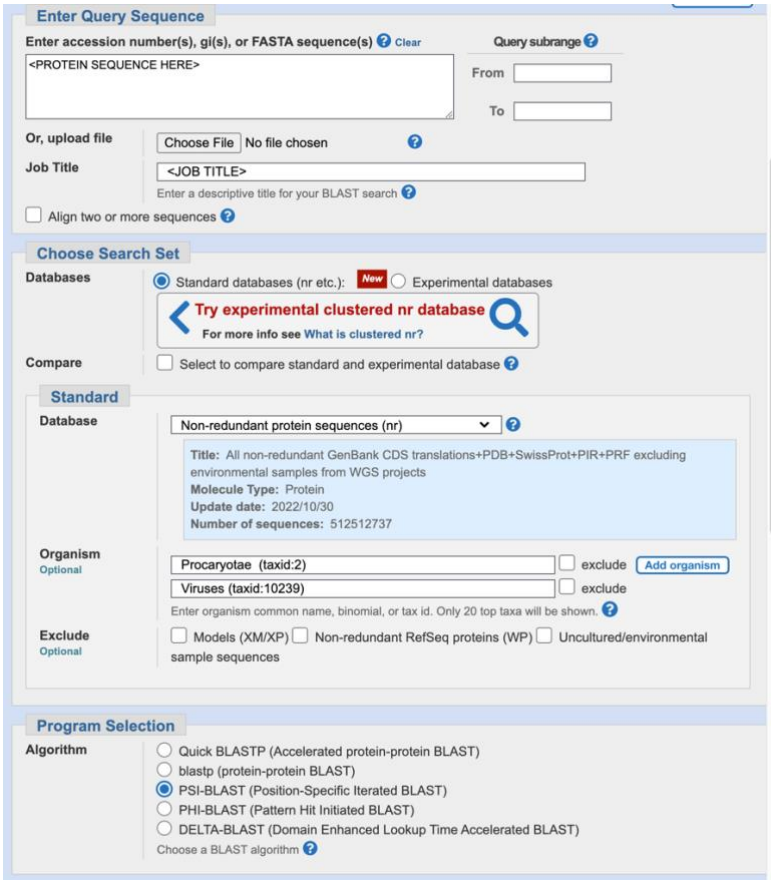


Figure X: NCBI's BLASTP web user interface

To ensure the reliability of the results, a psi-blast threshold of 0.005 is recommended, in order to filter out low-scoring alignments and reducing the likelihood of false positive hits.

Third, once the job execution finishes, it will show a list of sequences producing significant alignments, as shown in the picture below. These alignments are ordered by E-value, which indicates the number of hits expected to be seen by chance. The closer to 0, the better. This work will only consider sequences with E-value better than the threshold 0.005. Other attributes of interest are:

- **Total score**: Sum of the alignment scores of all of the segments from the sequence. The higher the score, the better the alignment.
- **Query coverage**: the % of the contig length that aligns with the NCBI hit. A small query coverage % means only a tiny portion of the contig is aligning. If there is an alignment with 100% identity and a 5% query coverage, the sequence is probably not that taxon.
- **Percent identity**: the % of bases that are identical to the reference genome. A query sequence can have a low % identity, but still be a real hit. It is essential to take the e value into account and look for homology between conserved regions- this will be evident at the protein level.

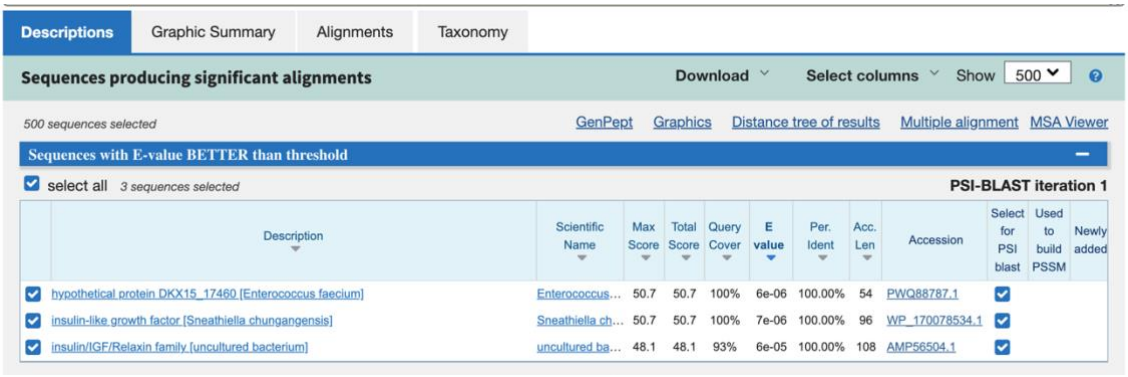


Figure X: NCBI's BLASTP sequence alignment results interface

Fourth, the link "Distance tree of results" presents a dendrogram or tree display that clusters sequences according to their distance from the query sequence. This display is useful for identifying aberrant or unusual sequences, as well as potentially natural groupings of related sequences such as members of gene families or homologs from other species in the BLAST output.

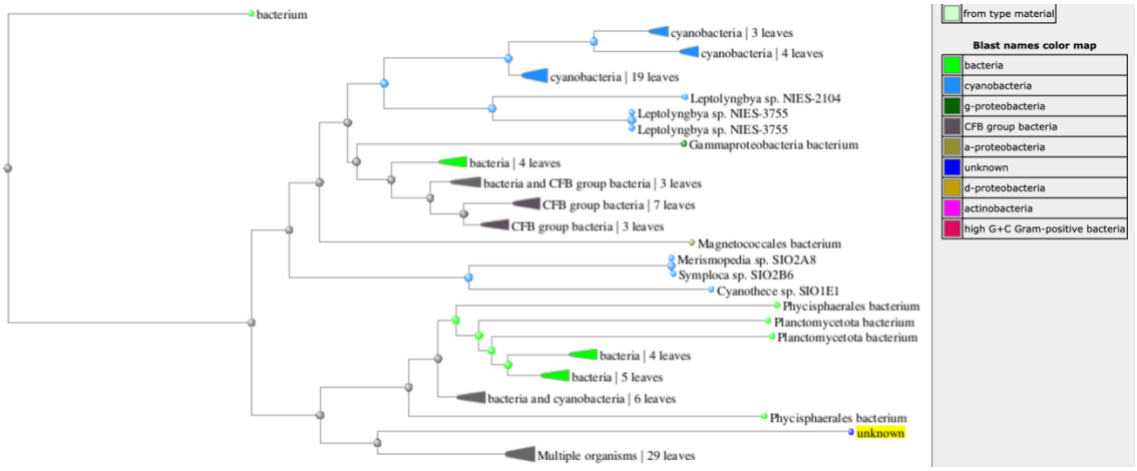


Figure X: NCBI's BLASTP distance tree diagram

Finally, the tab "Taxonomy" provides a report that summarizes everything that BLAST has classified about the relationships between all of the organisms found in the BLAST hitlist.

Once a potential relationship is found, the specific protein of the matching organism will be described using Uniprot database (https://www.uniprot.org), in particular checking its function and subcellular location to figure out if it actually may be related with the development of the autoimmune disease.

4. Results

Using the methodology described in the previous chapter, and leveraging advanced sequence alignment techniques from BLASTP (psi-BLAST in particular), it's possible to analyze sequence alignment between the epitopes of those autoimmune diseases in question and the proteome data of all prokaryote organisms across GenBank CDS, PDB, SwissProt, PIR, and PRF databases, in order to find insights on potential relationships between these diseases and previous infections of these organisms.

The following sections describe the results obtained for each individual analysis of the autoimmune diseases with the highest prevalence worldwide:

4.1. Type 1 diabetes (T1D)

According to IEDB, the most common epitopes specific to Type 1 Diabetes for the known human proteins that act as autoantigens are GAD65, IA-2, Insulin, and Insulin isoform 2. Using psi-BLAST to search for sequence alignments with these epitopes across prokaryote proteome databases throws the following results:

Autoantigen: **GAD65**
Epitope sequences: VMNILLQYVV (1) and NFFRMVISNPAAT (2)

The first sequence (1) doesn't trigger any significant matches better than the 0.005 threshold, and the non-significant matches are alignments with long sequences (400+ residues) mostly of DnaD domain proteins from firmicutes such as *Mammaliicoccus lentus* and *Macrococcus canis*.

The second sequence (2) doesn't produce significant matches either, and the non-significant ones are partial matches of 9 out of 13 aminoacids with high e-values.

As a result, the epitope sequences of human GAD65 don't seem to have homologous proteins within known proteomes of bacteria and viruses, since the e-value of those non-significant matches tells there's a high probability that they can occur simply by chance. On the other hand, there are studies [45] that highlight notable similarities between human and bacterial GAD, proposing a physiological hypothesis in which changes in the gut microbiome in T1D patients result in a release of bacterial GAD, thus causing miseducation of the host immune system, where deputized immune cells may then target human beta cells leading to the development of T1D.

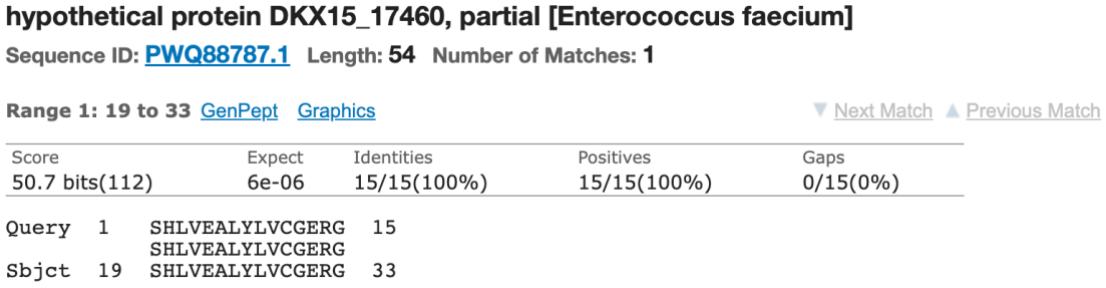
Autoantigen: **IA-2**
Epitope sequence: KLQVFLIVL
This epitope doesn't trigger any match that may be considered significant, since the length of the protein sequences (434+) is much longer than the query sequence (9 residues), and e-values of 12+ indicate these matches have a high probability of occurring simply by chance.

Autoantigen: **Insulin**
Epitope sequences: ALWGPDPAAA (1) and HLVEALYLV (2)

Output: None of the two epitopes (1) and (2) produce significant alignments with the proteome of prokaryote organisms.

Autoantigen: **Insulin isoform 2**
Epitope sequence: SHLVEALYLVCGERG

This sequence produces 3 significant hits within 3 Bacteria organisms: bacillota *Enterococcus faecium* (DKX15_17460) with 50.7 score, expect value of 6e-06, 100% query coverage, and 100% identity; the high G+C Gram-positive bacteria *Streptomyces sp. Lzd4kr* (WP_269646015.1) with 50.7 score, expect value of 7e-06, 100% coverage, and 100% identity; and a-proteobacteria *Sneathiella chungangensis* (WP_170078534), with 50.7 score, expect value of 7e-06, 100% coverage, and 100% identity. Does not produce significant alignments with viruses.



hypothetical protein DKX15_17460, partial [Enterococcus faecium]

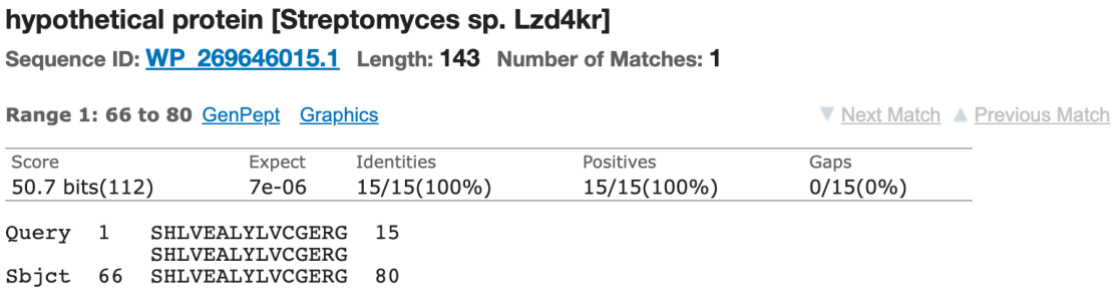
Sequence ID: PWQ88787.1 **Length:** 54 **Number of Matches:** 1

Range 1: 19 to 33 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
50.7 bits(112)	6e-06	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   SHLVEALYLVCGERG  15
           SHLVEALYLVCGERG
Sbjct  19  SHLVEALYLVCGERG  33
```

Figure X: Insulin isoform 2 alignment (1)



hypothetical protein [Streptomyces sp. Lzd4kr]

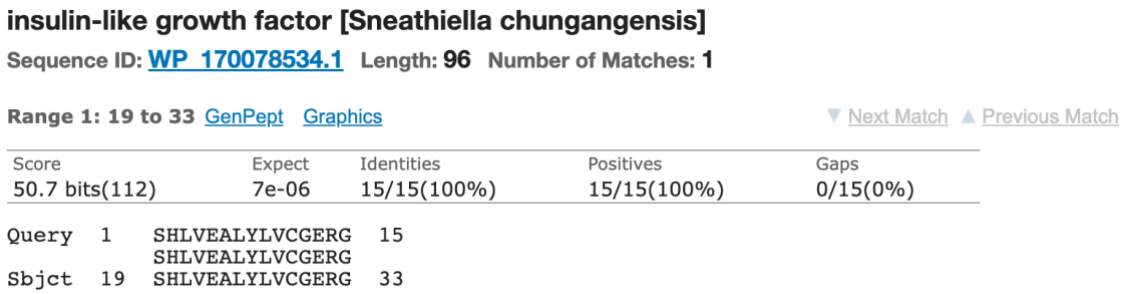
Sequence ID: WP_269646015.1 **Length:** 143 **Number of Matches:** 1

Range 1: 66 to 80 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
50.7 bits(112)	7e-06	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   SHLVEALYLVCGERG  15
           SHLVEALYLVCGERG
Sbjct  66  SHLVEALYLVCGERG  80
```

Figure X: Insulin isoform 2 alignment (2)



insulin-like growth factor [Sneathiella chungangensis]

Sequence ID: WP_170078534.1 **Length:** 96 **Number of Matches:** 1

Range 1: 19 to 33 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
50.7 bits(112)	7e-06	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   SHLVEALYLVCGERG  15
           SHLVEALYLVCGERG
Sbjct  19  SHLVEALYLVCGERG  33
```

Figure X: Insulin isoform 2 alignment (3)

The following diagram shows the distance between the epitope and the matches:

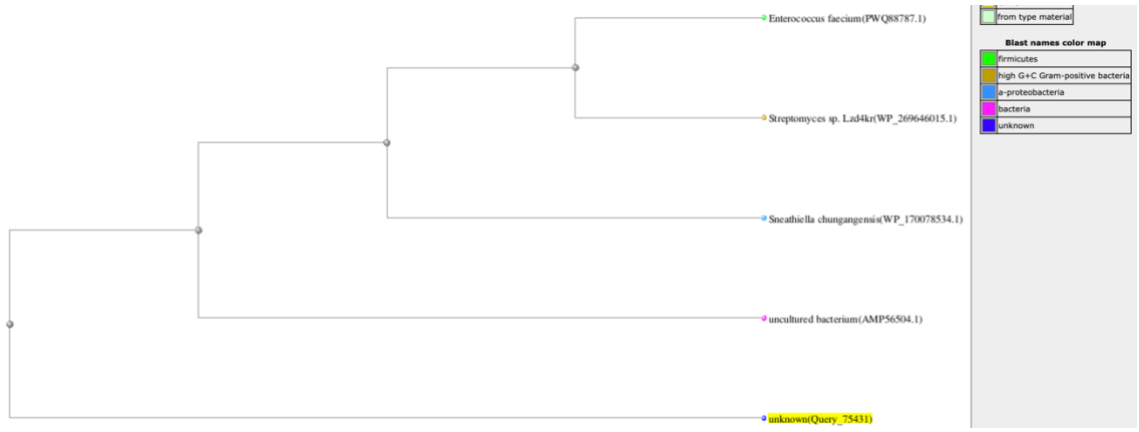


Figure X: Distance tree diagram for Insulin isoform 2 alignments

Based on the significant alignments identified, it seems that an infection of these bacteria may contribute to subsequent triggering of an autoimmune response against Insulin isoform 2, leading to the development of type 1 diabetes.

4.2. Rheumatoid Arthritis

According to IEDB, the most common epitope for Rheumatoid Arthritis is Cardiolipin. Using psi-blast to search for sequence alignments with this epitope across prokaryote proteome databases throws the following results:

Autontigen: **Cardiolipin**
Epitope sequence:
MLALRVARGSWGALRGAAWAPGTRPSKRRACWALLPPVPCCLGCLAERWRLRPAAL
GLRLPGIGQRNHCSGAGKAAPRPAAGAGAAAEAPGGQWGPASTPSLYENPWTIPNMLS
MTRIGLAPVLGYLIIEEDFNIALGVFALAGLTDLLDGFIARNWANQRSALGSALDPLAD
KILISILYVSLTYADLIPVPLTYMIISRDVMLIAAVFYVRYRTLPTPRTLAKYFNPCYATA
RLKPTFISKVNTAVQLILVAASLAAPVFNYADSIYLQILWCFTAFTTAASAYSYYHYGR
KTVQVIKD

The psi-BLAST tool produces 731 significant hits for this sequence within 285 Bacteria organisms, mostly delta/epsilon proteobacteria and a-proteobacteria, such as a-proteobacteria *Magnetococcales bacterium* (CDP-alcohol phosphatidyltransferase family protein) with 117 score, 60% query coverage and 39% identity; d-proteobacteria *Desulfatitalea sp. M08but* (CDP-alcohol phosphatidyltransferase family protein) with 114 score, 64% query coverage and 38% identity; and g-proteobacteria *Gammaproteobacteria bacterium* (CDP-alcohol phosphatidyltransferase) with 104 score, 59% query coverage, and 33% identity. Does not produce any significant alignments with viruses.

CDP-alcohol phosphatidyltransferase family protein [Magnetococcales bacterium]

Sequence ID: MBF0444013.1 **Length:** 187 **Number of Matches:** 1

Range 1: 11 to 174 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
108 bits(249)	2e-20	Compositional matrix adjust.	71/181(39%)	82/181(45%)	17/181(9%)

```
Query  108  TIPNMLSMTRIGLAPVLGYLIIEEDFNIALGVFALAGLTDLLDGFIARNWANQRSALGSA  167
            IPN LS+ R   AP    L I   FN AL  F  A LTD  DG+IA  W  Q   LG
Sbjct  11   NIPNSLSLIRLLCAPLFCWLLISNHFNFALVLFLVASLTDAADGYIAKRW-KQETELGRY  69

Query  168  LDPLADKILISILYVSLTYADLIPVPLTYMIISRDVMLIAAVFYVRYRTLPTPRTLAKYF  227
            LDPLADK+L  I ++ L    L PV LT  I+SRD+ LI  V             L K
Sbjct  70   LDPLADKVLLMIGFITLSWIQLLPVWLTLIIVSRDIVLIGGVI-----------LSKMV  117

Query  228  NPCYATARLKPTFISKVNTAVQLILVAASLAAPVFNYADSIYLQILWCFTAFTTAASAYS  287
              Y     P  ISKVNT  Q+ L+   L    F   +S+     W  TA  T AS
Sbjct  118  AKVYVV---QPLIISKVNTFLQMLLIMLVLLNSQFTELNSVVTMHIW-LTAISTVASLWA  173

Query  288  Y  288
            Y
Sbjct  174  Y  174
```

Figure X: Cardiolipin alignment (1)

CDP-alcohol phosphatidyltransferase family protein [Desulfatitalea sp. M08but]

Sequence ID: WP_246902604.1 **Length:** 192 **Number of Matches:** 1

See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 5 to 180 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
106 bits(243)	1e-19	Compositional matrix adjust.	76/193(39%)	83/193(43%)	17/193(8%)

```
Query  105  NPWTIPNMLSMTRIGLAPVLGYLIIEEDFNIALGVFALAGLTDLLDGFIARNWANQRSAL  164
             P  IPN L + RI L PV   L     +  AL VF  AG  D LDGFIAR   NQR  L
Sbjct  5    TPISIPNILTLVRILLTPVFIILLLRHLYWQALAVFVVAGVSDALDGFIARYF-NQRTTL  63

Query  165  GSALDPLADKILISILYVSLTYADLIPVPLTYMIISRDVMLIAAVFYVRYRTLPTPRTLA  224
            G  LDP ADK L    Y+SL   +  P   T  +I RDV     +      TL T  T
Sbjct  64   GAYLDPVADKLLLASAYISLAVLEVLPAWVTVIVIARDVIISLGIAIL---TL-TQKT--  117

Query  225  KYFNPCYATARLKPTFISKVNTAVQLILVAASLAAPVFNYADSIYLQILWCFTAFTTAAS  284
                  Y      PT ISK  TA QL+LV   L  P        YL  LW  TA  T AS
Sbjct  118  ------YEV---RPTIISKCTTAAQLVLVIVALCDPARIHLTILYLPLLW-LTAGLTTAS  167

Query  285  AYSYYHYGRKTVQ  297
             Y   G    Q
Sbjct  168  GLHYIYIGMNILQ  180
```

Figure X: Cardiolipin alignment (2)

CDP-alcohol phosphatidyltransferase [Gammaproteobacteria bacterium]

Sequence ID: MAQ20199.1 **Length:** 184 **Number of Matches:** 1

Range 1: 7 to 169 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
100 bits(230)	6e-18	Compositional matrix adjust.	64/179(36%)	77/179(43%)	16/179(8%)

```
Query  110  PNMLSMTRIGLAPVLGYLIIEEDFNIALGVFALAGLTDLLDGFIARNWANQRSALGSALD  169
            PN     R  L          E++ +AL    LAG  D LDG  ARN  N R  +G  LD
Sbjct  7    PNAICIARAVLVAPIVVALLGENYTVALALIGLAGFSDALDGMLARNF-NWRTRMGGVLD  65

Query  170  PLADKILISILYVSLTYADLIPVPLTYMIISRDVMLIAAVFYVRYRTLPTPRTLAKYFNP  229
            P ADK ++   ++SLTY +L+PV L  ++ISRD  LI  V    Y  L           P
Sbjct  66   PAADKLMVFSVFISLTYLELVPVFLAVLVISRD--LIIIVGAMTYQFL---------IEP  114

Query  230  CYATARLKPTFISKVNTAVQLILVAASLAAPVFNYADSIYLQILWCFTAFTTAASAYSY  288
                  P FISK NT VQL      L    F    S +L IL     FT   S   Y
Sbjct  115  VEG----EPAFISKLNTGVQLLFMMVTLMSAAFHWPLSSFLLILGAIVIFTSITSGLNY  169
```

Figure X: Cardiolipin alignment (3)

The distance tree diagram below shows the alignment of the epitope sequence and these organisms:

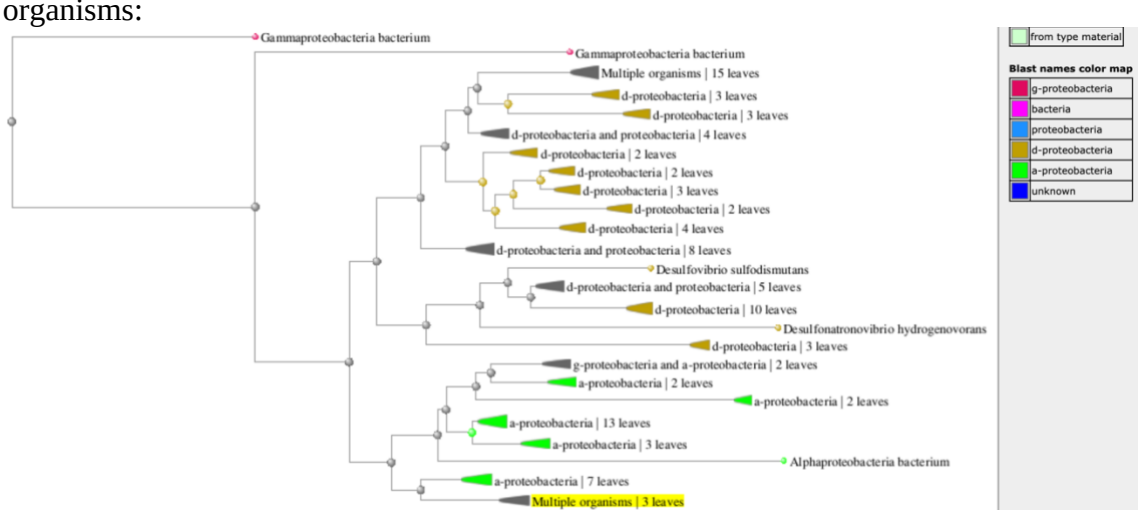


Figure X: Distance tree diagram for Cardiolipin alignments

Most of the significant alignments (482 out of the top 500 results) are with proteins from the CDP-alcohol phosphatidyl-transferase family, that participate in phospholipid biosynthetic process in the cellular membrane. Since bacterial antigens occur on the surface of the cell, seems there may be a potential relationship between developing autoimmunity to cardiolipin and previous infections with some of the bacteria found.

Autoantigen: **Alpha-enolase**
Epitope sequence: KIHAREIFDSRGNPTVE

This epitope from alpha-enolase produces 997 significant hits within 332 organisms, out of which 206 are FCB group bacteria. Most of the alignments produced (481 out of top 500) are with the phosphopyruvate hydratase protein of these bacteria, such as *Klebsiella pneumoniae*, *Enterococcus faecalis*, and *Staphylococcus aureus*.

phosphopyruvate hydratase [Klebsiella pneumoniae]

Sequence ID: MCP6508410.1 **Length:** 58 **Number of Matches:** 1

Range 1: 5 to 21 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
57.9 bits(129)	2e-08	17/17(100%)	17/17(100%)	0/17(0%)

```
Query  1   KIHAREIFDSRGNPTVE  17
           KIHAREIFDSRGNPTVE
Sbjct  5   KIHAREIFDSRGNPTVE  21
```

Figure X: Alpha-enolase alignment (1)

phosphopyruvate hydratase [Enterococcus faecalis]

Sequence ID: MBW3682561.1 Length: 59 Number of Matches: 1

Range 1: 5 to 21 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
57.9 bits(129)	2e-08	17/17(100%)	17/17(100%)	0/17(0%)

```
Query  1   KIHAREIFDSRGNPTVE  17
           KIHAREIFDSRGNPTVE
Sbjct  5   KIHAREIFDSRGNPTVE  21
```

Figure X: Alpha-enolase alignment (2)

hypothetical protein [Staphylococcus aureus]

Sequence ID: MBO8666304.1 Length: 98 Number of Matches: 1

Range 1: 5 to 21 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
57.9 bits(129)	3e-08	17/17(100%)	17/17(100%)	0/17(0%)

```
Query  1   KIHAREIFDSRGNPTVE  17
           KIHAREIFDSRGNPTVE
Sbjct  5   KIHAREIFDSRGNPTVE  21
```

Figure X: Alpha-enolase alignment (3)

The majority of the alignments have 100% identity, no gaps, a score above 55 and significant e-value (below 0.005), which indicates there's a potential relationship between this epitope from human alpha-enolase, and phosphopyruvate hydratase (enolase) bacterial protein. Since this protein is involved in cell wall formation and RNA turnover and as a plasminogen receptor, there's a possible relationship between autoimmune reaction to human alpha-enolase, and a previous bacterial infection.

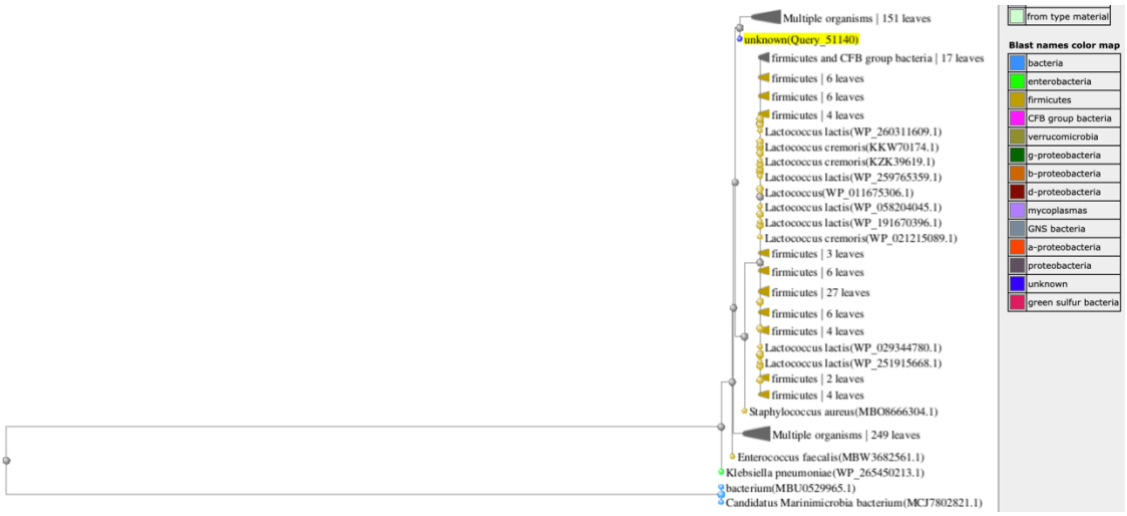
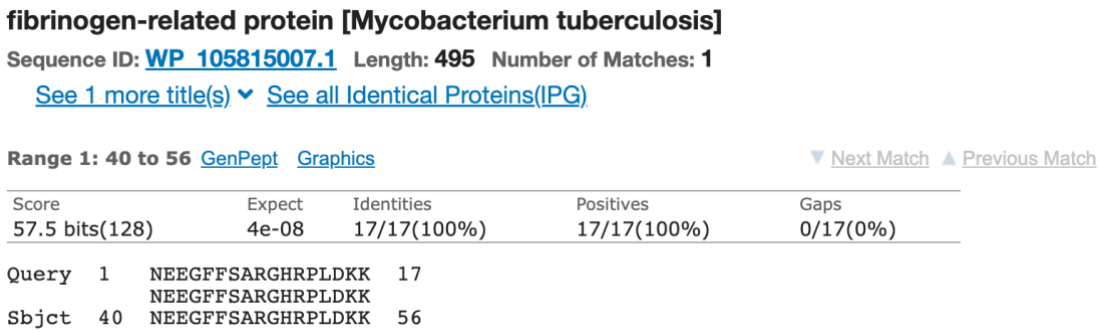


Figure X: Distance tree diagram for Alpha-enolase alignments

Autoantigen: **Fibrinogen beta chain**

Epitope sequences: NEEGFFSARGHRPLDKK and RPAPPPISGGGYRAR

The first sequence (1) generates one significant alignment with a fibrinogen-related protein in *Mycobacterium tuberculosis*:



fibrinogen-related protein [Mycobacterium tuberculosis]

Sequence ID: WP_105815007.1 **Length:** 495 **Number of Matches:** 1

See 1 more title(s) See all Identical Proteins(IPG)

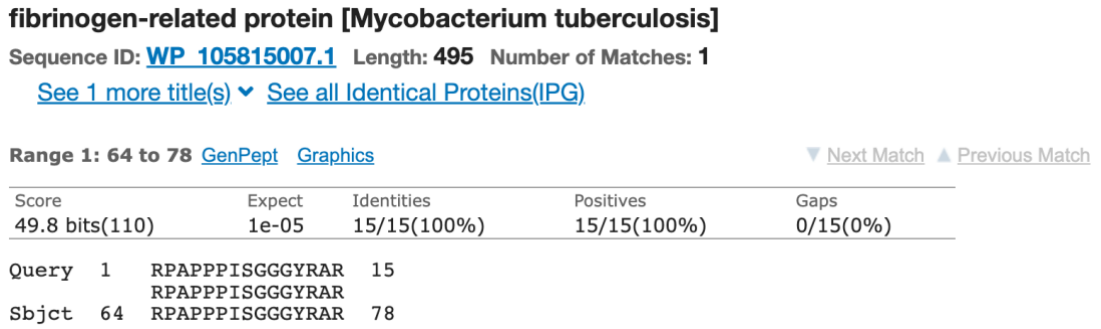
Range 1: 40 to 56 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
57.5 bits(128)	4e-08	17/17(100%)	17/17(100%)	0/17(0%)

```
Query  1   NEEGFFSARGHRPLDKK  17
           NEEGFFSARGHRPLDKK
Sbjct  40  NEEGFFSARGHRPLDKK  56
```

Figure X: Fibrinogen beta chain alignment (1)

The second sequence (2) produces two significant alignments, also with fibrinogen-related proteins from *Mycobacterium tuberculosis* and *Salmonella enterica*:



fibrinogen-related protein [Mycobacterium tuberculosis]

Sequence ID: WP_105815007.1 **Length:** 495 **Number of Matches:** 1

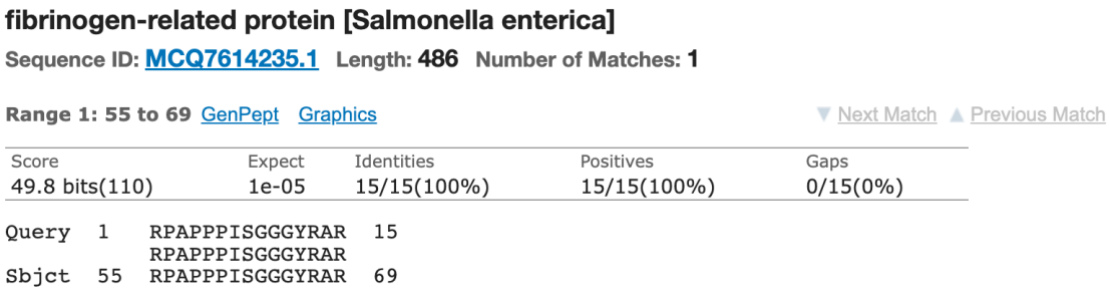
See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 64 to 78 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
49.8 bits(110)	1e-05	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   RPAPPPISGGGYRAR  15
           RPAPPPISGGGYRAR
Sbjct  64  RPAPPPISGGGYRAR  78
```

Figure X: Fibrinogen beta chain alignment (2)



fibrinogen-related protein [Salmonella enterica]

Sequence ID: MCQ7614235.1 **Length:** 486 **Number of Matches:** 1

Range 1: 55 to 69 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
49.8 bits(110)	1e-05	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   RPAPPPISGGGYRAR  15
           RPAPPPISGGGYRAR
Sbjct  55  RPAPPPISGGGYRAR  69
```

Figure X: Fibrinogen beta chain alignment (3)

Due to the complexity of both epitope sequences, even though the length of the protein sequences found is long (~500 aminoacids), these alignments are considered significant.

In the light of the results obtained, seems that there's a possible relationship between developing autoimmune response to human fibrinogen beta chain proteins and a previous infection of *Mycobacterium tuberculosis* or *Salmonella enterica*, which can led to the development of Rheumatoid Arthritis.

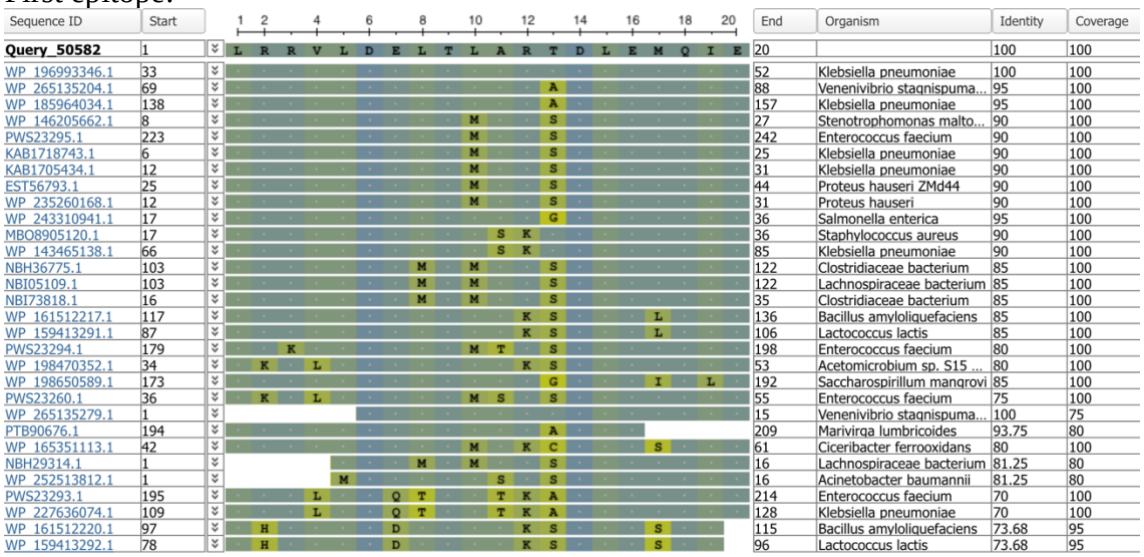
4.3. Psoriasis

According to IEDB, the most common epitopes in Psoriasis are present in Keratin, type I cytoskeletal 16, Keratin, type I cytoskeletal 18, and ADAMTS-like protein 5. Using psi-blast to search for sequence alignments with this epitope across prokaryote proteome databases throws the following results:

Autoantigen: **Keratin, type I cytoskeletal 16**
Epitope sequences: LRRVLDELTLARTDLEMQIE, ANILLQIDNARLAAD, ENRYCVQLSQIQGLI, VRALEEANTELEVKI

These four epitope sequences produce the following significant alignments:

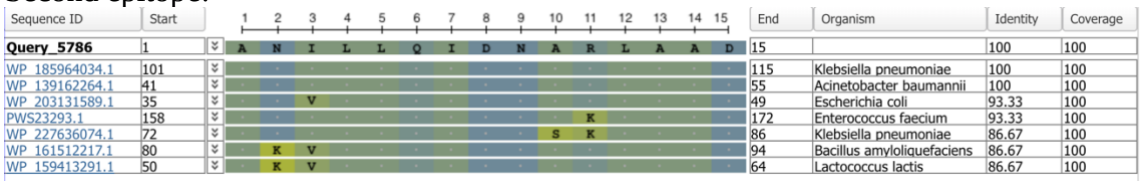
First epitope:



Sequence ID	Start	End	Organism	Identity	Coverage
Query_50582	1	20		100	100
WP_196993346.1	33	52	Klebsiella pneumoniae	100	100
WP_265135204.1	69	88	Venenivibrio stagnispuma...	95	100
WP_185964034.1	138	157	Klebsiella pneumoniae	95	100
WP_146205662.1	8	27	Stenotrophomonas malto...	90	100
PWS23295.1	223	242	Enterococcus faecium	90	100
KAB1718743.1	6	25	Klebsiella pneumoniae	90	100
KAB1705434.1	12	31	Klebsiella pneumoniae	90	100
EST56793.1	25	44	Proteus hauseri ZMd44	90	100
WP_235260168.1	12	31	Proteus hauseri	90	100
WP_243310941.1	17	36	Salmonella enterica	95	100
MBQ8905120.1	17	36	Staphylococcus aureus	90	100
WP_143465138.1	66	85	Klebsiella pneumoniae	90	100
NBH36775.1	103	122	Clostridiaceae bacterium	85	100
NBI05109.1	103	122	Lachnospiraceae bacterium	85	100
NBI73818.1	16	35	Clostridiaceae bacterium	85	100
WP_161512217.1	117	136	Bacillus amyloliquefaciens	85	100
WP_159413291.1	87	106	Lactococcus lactis	85	100
PWS23294.1	179	198	Enterococcus faecium	80	100
WP_198470352.1	34	53	Acetomicrobium sp. S15 ...	80	100
WP_198650589.1	173	192	Saccharospirillum mangrovi	85	100
PWS23260.1	36	55	Enterococcus faecium	75	100
WP_265135279.1	1	15	Venenivibrio stagnispuma...	100	75
PTB90676.1	194	209	Marivirga lumbricoides	93.75	80
WP_165351113.1	42	61	Ciceribacter ferrooxidans	80	100
NBH29314.1	1	16	Lachnospiraceae bacterium	81.25	80
WP_252513812.1	1	16	Acinetobacter baumannii	81.25	80
PWS23293.1	195	214	Enterococcus faecium	70	100
WP_227636074.1	109	128	Klebsiella pneumoniae	70	100
WP_161512220.1	97	115	Bacillus amyloliquefaciens	73.68	95
WP_159413292.1	78	96	Lactococcus lactis	73.68	95

Figure X: Multiple sequence alignment viewer for Keratin, type I cytoskeletal 16 (1)

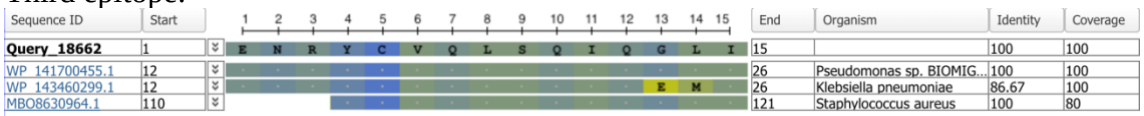
Second epitope:



Sequence ID	Start	End	Organism	Identity	Coverage
Query_5786	1	15		100	100
WP_185964034.1	101	115	Klebsiella pneumoniae	100	100
WP_139162264.1	41	55	Acinetobacter baumannii	100	100
WP_203131589.1	35	49	Escherichia coli	93.33	100
PWS23293.1	158	172	Enterococcus faecium	93.33	100
WP_227636074.1	72	86	Klebsiella pneumoniae	86.67	100
WP_161512217.1	80	94	Bacillus amyloliquefaciens	86.67	100
WP_159413291.1	50	64	Lactococcus lactis	86.67	100

Figure X: Multiple sequence alignment for Keratin, type I cytoskeletal 16 (2)

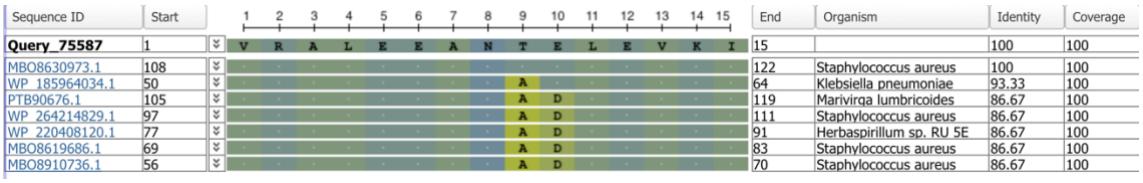
Third epitope:



Sequence ID	Start	End	Organism	Identity	Coverage
Query_18662	1	15		100	100
WP_141700455.1	12	26	Pseudomonas sp. BIOMIG...	100	100
WP_143460299.1	12	26	Klebsiella pneumoniae	86.67	100
MBQ8630964.1	110	121	Staphylococcus aureus	100	80

Figure X: Multiple sequence alignment for Keratin, type I cytoskeletal 16 (3)

Fourth epitope:



Sequence ID	Start	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	End	Organism	Identity	Coverage
Query_75587	1	V	R	A	L	E	E	A	N	T	E	L	E	V	K	I	15		100	100
MBO8630973.1	108	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	122	Staphylococcus aureus	100	100
WP_185964034.1	50	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	64	Klebsiella pneumoniae	93.33	100
PTB90676.1	105	·	·	·	·	·	·	·	·	A	D	·	·	·	·	·	119	Marivirga lumbricoides	86.67	100
WP_264214829.1	97	·	·	·	·	·	·	·	·	A	D	·	·	·	·	·	111	Staphylococcus aureus	86.67	100
WP_220408120.1	77	·	·	·	·	·	·	·	·	A	D	·	·	·	·	·	91	Herbaspirillum sp. RU 5E	86.67	100
MBO8619686.1	69	·	·	·	·	·	·	·	·	A	D	·	·	·	·	·	83	Staphylococcus aureus	86.67	100
MBO8910736.1	56	·	·	·	·	·	·	·	·	A	D	·	·	·	·	·	70	Staphylococcus aureus	86.67	100

Figure X: Multiple sequence alignment for Keratin, type I cytoskeletal 16 (4)

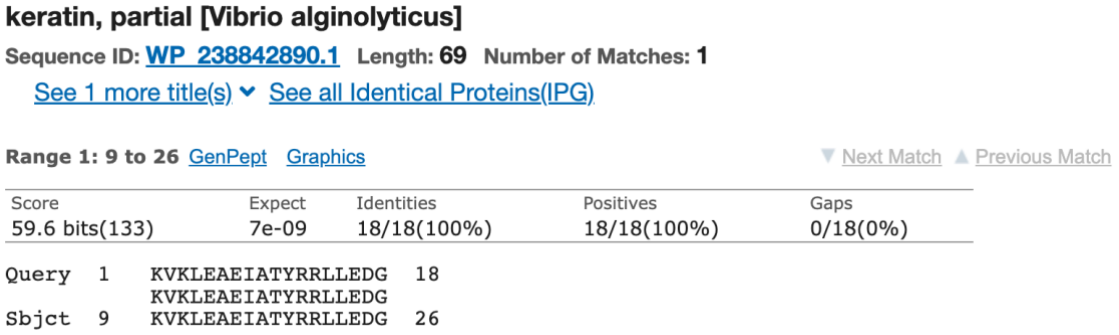
Half of the alignments produced (22 out of 47) are bacterial Keratin, and the other half hypothetical proteins from these bacteria related with filament region (intermediate filament protein), such as and WP_185964034 and DKP78_13785, both of them with similar structural functions.

Although the structural function of these proteins plays a role in allowing bacteria to adhere to mammalian host cells, Keratin is typically not considered an antigen in any infection, however the immune system may respond to damaged or altered keratin triggering the immune response against this protein, such as in psoriasis patients.

Some capsular polysaccharides are key to the pathogenesis of bacterial infections, such as K6, K16, and K17 in *klebsiella pneumoniae* infection [46]. The alignment between these epitopes in human autoantigens and bacterial keratin may indicate a potential relationship between developing psoriasis and a previous infection of such bacteria.

Autoantigen: **Keratin, type I cytoskeletal 18**
Epitope sequence: KVKLEAEIATYRRLLEDG

Alignment with this epitope sequence produces similar results than with Keratin 16, psi-BLAST returned 2 high quality results from 2 organisms: *Vibrio alginolyticus*, and *Klebsiella pneumoniae*:



keratin, partial [Vibrio alginolyticus]

Sequence ID: WP_238842890.1 **Length:** 69 **Number of Matches:** 1

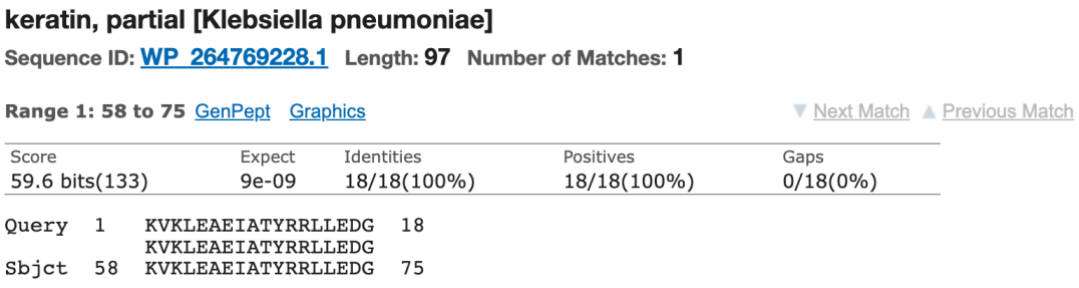
See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 9 to 26 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
59.6 bits(133)	7e-09	18/18(100%)	18/18(100%)	0/18(0%)

```
Query  1   KVKLEAEIATYRRLLEDG  18
           KVKLEAEIATYRRLLEDG
Sbjct  9   KVKLEAEIATYRRLLEDG  26
```

Figure X: Keratin, type I cytoskeletal 18 alignment (1)



keratin, partial [Klebsiella pneumoniae]

Sequence ID: WP_264769228.1 **Length:** 97 **Number of Matches:** 1

Range 1: 58 to 75 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
59.6 bits(133)	9e-09	18/18(100%)	18/18(100%)	0/18(0%)

```
Query  1   KVKLEAEIATYRRLLEDG  18
           KVKLEAEIATYRRLLEDG
Sbjct  58  KVKLEAEIATYRRLLEDG  75
```

Figure X: Keratin, type I cytoskeletal 18 alignment (2)

These alignments have 100% identity and high significance, which indicates homology with specific regions of bacterial keratin, and signaling up a potential relationship between developing an autoimmune response to structural proteins such as Keratin and a previous infection by these organisms, especially if these regions assemble peptides related with known bacterial antigens.

Autoantigen: **ADAMTS-like protein 5**
Epitope sequence: VRSRRCLRL
Output: No significant alignment detected within virus and bacteria databases.

4.4. Celiac disease

According to IEDB, the most common epitopes for Psoriasis are located in antigens Tissue transglutaminase and Calreticulin. Using psi-BLAST to search for sequence alignments with these epitopes across prokaryote proteome databases throws the following results:

Autoantigen: **Tissue transglutaminase**
Epitope sequence: LEPFSGKALCSWSIC
Output: No significant alignment detected within virus and bacteria databases.

Autoantigen: **Calreticulin**
Epitope sequences: EDKKRKEEEE, EEEEAEDKED, EQRLKEEEED
Output: No significant alignment detected within virus and bacteria databases for any of the three epitope sequences.

In view of these results, seems that these known autoantigens in humans for celiac disease aren't related with previous bacterial or viral infections. However, since the main antigen for this disease is the well-known Gluten, or alpha-beta gliadin, with epitopes such as Tri a 21, it's worth exploring the relationship between this molecule and known antigens from both prokaryote and human proteomes:

Tri a 21 epitope sequence: QLQPFPQPQLPY

This epitope produces an alignment with HLA-DQ2-glia-alpha1, but doesn't generate any alignments with the proteome of bacteria and viruses.

4.5. Graves' disease

The most common epitopes for Graves' disease are located in Thyrotropin receptor (TSHR), and Thyroid peroxidase (TPO). Using psi-blast to search for sequence alignments with this epitope across prokaryote proteome databases throws the following results:

Autoantigen: **Thyrotropin receptor (TSHR)**

Epitope sequence: PDLTKVYSTDIFFILEITDN, PSTQTLKLIETHLRTIPSHA
Output: No significant alignment detected within virus and bacteria databases for any of the two epitope sequences.

Autoantigen: **Thyroid peroxidase (TPO)**
Epitope sequence: GLPRLETPADLSTAIASRS
Output: No significant alignment detected within virus and bacteria databases.

With these data points, it seems that these specific epitopes aren't related with prokaryote proteome.

4.6. Multiple sclerosis

According to IEDB, the most common epitopes for Multiple Sclerosis are located in Myelin basic protein (MBP), and Myelin-oligodendrocyte glycoprotein (MOG). Using psi-BLAST to search for sequence alignments with these epitopes across prokaryote proteome databases provides the following results:

Autoantigen: **Myelin basic protein (MBP)**
Epitope sequence: ENPVVHFFKNIVTPR, DENPVVHFFKNIVTPRTPP

The first sequence (1) generates one alignment with a hypothetical protein from Gammaproteobacteria *Xylella fastidiosa subsp. multiplex*

hypothetical protein [Xylella fastidiosa subsp. multiplex]
Sequence ID: MRT95218.1 Length: 93 Number of Matches: 1

Range 1: 9 to 23 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
53.2 bits(118)	9e-07	15/15(100%)	15/15(100%)	0/15(0%)

```
Query  1   ENPVVHFFKNIVTPR  15
           ENPVVHFFKNIVTPR
Sbjct  9   ENPVVHFFKNIVTPR  23
```

Figure X: Myelin basic protein (MBP) alignment (1)

The sequence alignment has 100% identity and high significance. The second epitope sequence (2) is aligned with the same hypothetical protein from *Xylella fastidiosa subsp. multiplex* identified:

hypothetical protein [Xylella fastidiosa subsp. multiplex]
Sequence ID: MRT95218.1 Length: 93 Number of Matches: 1

Range 1: 8 to 26 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
66.4 bits(149)	4e-11	19/19(100%)	19/19(100%)	0/19(0%)

```
Query  1   DENPVVHFFKNIVTPRTPP  19
           DENPVVHFFKNIVTPRTPP
Sbjct  8   DENPVVHFFKNIVTPRTPP  26
```

Figure X: Myelin basic protein (MBP) alignment (2)

In order to confirm potential homology, let's run sequence alignment between the full sequence of that hypothetical protein and human proteome:

> hypothetical protein [Xylella fastidiosa subsp. Multiplex]
QKSQRTQDENPVVHFFKNIVTPRTPPPSQGKGRGLSLSRFSWGAEGQKPGFGYGGRAS
DYKSAHKGFKGAYDARARFPKSLSWEGETAAMDTP

The tool psi-BLAST produces 35 significant alignments with human proteins, 20 of which are myelin basic proteins:

myelin basic protein [Homo sapiens]

Sequence ID: KAI2587336.1 **Length:** 149 **Number of Matches:** 1

See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 53 to 145 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
139 bits(349)	1e-42	Compositional matrix adjust.	72/94(77%)	76/94(80%)	2/94(2%)

```
Query  1    QKSQ-RTQDENPVVHFFKNIVTPRTPPPSQGKGRGLSLSRFSWGAEGQKPGFGYGGRASD  59
            QKS  RTQDENPVVHFFKNIVTPRTPPPSQGKGRGLSLSRFSWGAEGQ+PGFGYGGRASD
Sbjct  53   QKSHGRTQDENPVVHFFKNIVTPRTPPPSQGKGRGLSLSRFSWGAEGQRPGFGYGGRASD  112

Query  60   YKSAHKGFKGAYDARARFPKSLSWEGETAAMDTP  93
            YKSAHKGFKG  DA+    K     G  +   +P
Sbjct  113  YKSAHKGFKGV-DAQGTLSKIFKLGGRDSRSGSP  145
```

Figure X: Hypothetical protein alignment with human myelin basic protein

This specific alignment has high significance, 77% of identity and 2% gaps (2 out of 94 residues). This indicates that there's certain homology between the hypothetical protein from *Xylella fastidiosa subsp. multiplex* and human myelin basic protein.

In light of these results, seems that a previous infection of this bacteria might trigger immune response to Myelin basic protein (MBP), one of the autoantigens linked with the development of Multiple Sclerosis.

Autoantigen: **Myelin-oligodendrocyte glycoprotein (MOG)**
Epitope sequence: MEVGWYRPPFSRVVHLYRNGK
Output: No significant alignment detected within virus and bacteria databases.

5. Conclusion and future work

As referenced throughout the document, some existing studies [47] suggest that there are certain bacteria and viruses that might play a role in the development of autoimmune diseases. However, the interpretation of existing results is limited because the number of studies and patients studied is relatively small.

This work confirms those findings with evidence extracted from the analysis of homology between the epitopes localized in the known autoantigens for the most common autoimmune diseases and the proteome of prokaryote organisms such as bacteria and viruses.

In particular, based on the analysis reported in the previous sections, there is evidence that suggest that infections with certain types of bacteria, such as *Enterococcus faecium*, *Klebsiella pneumoniae*, *Salmonella enterica*, and *Xylella fastidiosa subsp. Multiplex*, play a role in the development of certain autoimmune diseases, such as type 1 diabetes, rheumatoid arthritis, psoriasis, and sclerosis multiple.

The following table summarizes the results reported in the previous section:

Autoimmune disease	Found relationships
Type 1 diabetes	*Enterococcus faecium* *Streptomyces sp. Lzd4kr* *Sneathiella chungangensis*
Rheumatoid Arthritis	*Magnetococcales bacterium* *Desulfatitalea sp. M08but* *Klebsiella pneumoniae* *Mycobacterium tuberculosis* *Salmonella enterica*
Psoriasis	*klebsiella pneumoniae*
Celiac disease	N/A
Graves' disease	N/A
Multiple sclerosis	*Xylella fastidiosa subsp. Multiplex*

It is important to note that the relationship between infections and autoimmune diseases is complex and not fully understood. While some infections may be related with these diseases, and increase the risk of developing autoimmune disorders, they are not necessarily the cause of those conditions.

In conclusion, while there is evidence that suggests that viral or bacterial infections may be associated with an increased risk of developing autoimmune diseases, the exact relationship between these infections and autoimmune disorders is not fully understood yet. Further research is needed in order to determine the exact role of infections in the development of autoimmune diseases, and to identify potential prevention or treatment strategies.

The future of autoimmune disease research is promising, and there is hope that ongoing combination of efforts to develop new treatments, understand the causes of these diseases,

and prevent their occurrence will lead to a better understanding of these conditions and the development of more effective treatments.

Regarding future work, the analysis carried out in this work could be automated to provide new tools that check alignments between epitopes and prokaryote proteomes automatically. The tools used in this work are publicly available research tools that, while they are a gold mine of information, the user interfaces and documentation seem outdated. For example, there's a huge opportunity for automation at IEDB, that currently doesn't don't have APIs to access the information, providing only an option to export the information in batch [48]. Also, BLAST alignment process is very manual, as most of the information provided by the user interface is not available programmatically. There are some promising initiatives, such as Biocontainers [49], but other initiatives such as Blast in the cloud [50] used by the Bio.Blast Python package [51] are now deprecated, and need an update.

6. Glosary

1. **Autoantibody:** Antibodies produced by the immune system that mistakenly attack healthy cells.
2. **Autoantigens:** Specific molecules targeted by autoantibodies, which may vary depending on the autoimmune disease.
3. **Autoimmune diseases (AD):** A type of disease where the immune system attacks the body's own healthy cells.
4. **Beta cells:** The cells in the pancreas that produce insulin.
5. **Center for Disease Control and Prevention (CDC):** A national public health institute in the United States.
6. **CVB:** Coxsackie virus B, a type of enterovirus.
7. **Epitope:** The specific site on an antigen (and autoantigen) molecule that is recognized and targeted by the immune system.
8. **Environmental influence:** Factors in the environment that may contribute to the development of autoimmune diseases.
9. **Epstein-Barr virus:** A virus that can cause mononucleosis and is associated with certain types of cancer.
10. **Enteroviruses:** A group of viruses that can cause a wide range of illnesses.
11. **Global Burden of Disease study:** A study that examines the incidence, prevalence, and disability-adjusted life years (DALYs) of diseases and injuries worldwide.
12. **Human parvovirus:** A virus that can cause a wide range of illnesses.
13. **Mumps virus:** A virus that can cause mumps, a disease that primarily affects the parotid glands.
14. **Pathogens:** Microorganisms that can cause diseases.
15. **Prevalence:** The proportion of people affected by a certain disease or condition in a specific population.
16. **Rotavirus:** A virus that can cause diarrhea, especially in young children.
17. **Synovium:** The thin layer of cells covering the joints.

7. References

[1] U.S. Department of Health and Human Services, Office on Women's Health; 2010; Autoimmune diseases: Overview; Accessed Oct 2022
https://web.archive.org/web/20161005193812/https:/www.womenshealth.gov/files/assets/docs/fact-sheets/autoimmune-diseases.pdf

[2] Emily Sohn; 2021; Why autoimmunity is most common in women; Accessed Oct 2022
https://www.nature.com/articles/d41586-021-01836-9

[3] National Stem Cell Foundation; Autoimmune disease; Accessed Oct 2022
https://nationalstemcellfoundation.org/glossary/autoimmune-disease/

[4] Michael McCarthy; 2000; The "gender gap" in autoimmune disease; TheLancet
https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(05)74535-9/fulltext

[5] Nese Sinmaz, Tina Nguyen, Fiona Tea, Russell C. Dale, and Fabienne Brilot; 2016; Mapping autoantigen epitopes: molecular insights into autoantibody-associated disorders of the nervous system; Accessed Oct 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5006540/

[6] Global Autoimmune Institute; Accessed Nov 2022
https://www.autoimmuneinstitute.org/

[7] Stephanie Watson; 2019/2022; Autoimmune Diseases: Types, Symptoms, Causes, and More; Accessed Nov 2022
https://www.healthline.com/health/autoimmune-disorders

[8] Gabriel A Gregory, Thomas I G Robinson, Sarah E Linklater, Fei Wang, Stephen Colagiuri, Carine de Beaufort, et al.; 2022; Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study; Accessed Nov 2022
https://www.thelancet.com/journals/landia/article/PIIS2213-8587(22)00218-2/fulltext

[9] Christophe M. Filippi and Matthias G. von Herrath; 2008; Viral Trigger for Type 1 Diabetes; Accessed Nov 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570378/

[10] Hyoty H, Taylor KW; The role of viruses in human diabetes; 1353–1361; 2002;
https://pubmed.ncbi.nlm.nih.gov/12378375

[11] Honeyman MC, Stone NL, Harrison LC; T-cell epitopes in type 1 diabetes autoantigen tyrosine phosphatase IA-2: potential for mimicry with rotavirus and other environmental agents; 231-239; 1998
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2230363/

[12] Honeyman MC, Coulson BS, Stone NL, Gellert SA, Goldwater PN, Steele CE, Couper JJ, Tait BD, Colman PG, Harrison LC; Association between rotavirus infection

and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes; 1319–1324; 2000
https://pubmed.ncbi.nlm.nih.gov/10923632

[13] Hyoty H, Leinikki P, Reunanen A, Ilonen J, Surcel HM, Rilva A, Kaar ML, Huupponen T, Hakulinen A, Makela AL, et al.; Mumps infections in the etiology of type 1 (insulin-dependent) diabetes; 111–116; 1988
https://pubmed.ncbi.nlm.nih.gov/3243043

[14] Pak CY, Eun HM, McArthur RG, Yoon JW; Association of cytomegalovirus infection with autoimmune type 1 diabetes; 1–4; 1988
https://pubmed.ncbi.nlm.nih.gov/2898620

[15] Smolen, J. S., Aletaha, D., and McInnes, I. B; 2016; Rheumatoid Arthritis; Accessed Nov 2022
http://eprints.gla.ac.uk/131249/1/131249.pdf

[16] Centers for Disease Control and Prevention, Division of Population Health; 2020; Rheumatoid Arthritis (RA); Accessed Nov 2022
https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html

[17] Marita Cross, Emma Smith, Damian Hoy, Loreto Carmona, Frederick Wolfe, Theo Vos, Benjamin Williams, Sherine Gabriel, Marissa Lassere, Nicole Johns, Rachelle Buchbinder, Anthony Woolf, Lyn March; 2010; The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study; Accessed Nov 2022
https://pubmed.ncbi.nlm.nih.gov/24550173/

[18] Josef S. Smolen; 2016; Rheumatoid Arthritis; Accessed Nov 2022
http://eprints.gla.ac.uk/131249/1/131249.pdf

[19] Karen H Costenbader and Elizabeth W Karlson; 2006; Epstein–Barr virus and rheumatoid arthritis: is there a link?; Accessed Nov 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1526553/

[20] Mark J. Soloski, Eleanor S. Metcalf; 2007; Salmonella as an Inducer of Autoimmunity; Accessed Nov 2022
https://journals.asm.org/doi/10.1128/ecosalplus.8.8.13

[21] Rosa Parisi 1, Deborah P M Symmons, Christopher E M Griffiths, Darren M Ashcroft; 2013; Global epidemiology of psoriasis: a systematic review of incidence and prevalence; Accessed Nov 2022
https://pubmed.ncbi.nlm.nih.gov/23014338/

[22] National Psoriasis Foundation; Psoriasis Statistics; Accessed Nov 2022
https://www.psoriasis.org/psoriasis-statistics/

[23] Lionel Fry, Barbara S.Baker; 2007; Triggering psoriasis: the role of infections and medications; Accessed Nov 2022
https://www.sciencedirect.com/science/article/pii/S0738081X07001599?via%3Dihub

[24] Lucia Brambilla, Giovanni Genovese, Athanasia Tourlaki, Silvia Della Bella; 2018; Coexistence of Kaposi's sarcoma and psoriasis: is there a hidden relationship?; Accessed Nov 2022
https://pubmed.ncbi.nlm.nih.gov/30105990/

[25] Nilesh Morar, Saffron A Willis-Owen, Toby Maurer, Christopher B Bunker; 2010; HIV-associated psoriasis: pathogenesis, clinical features, and management; Accessed Nov 2022
https://pubmed.ncbi.nlm.nih.gov/20610329/

[26] Alastair Compston, Alasdair Coles; 2008; Multiple Sclerosis; Accessed Nov 2022
https://pubmed.ncbi.nlm.nih.gov/18970977/

[27] Jo Lane, Huah Shin Ng, Carmel Poyser, Robyn M, Lucas Helen Tremlett; 2022; Multiple sclerosis incidence: A systematic review of change over time by geographical region; Accessed Nov 2022
https://www.msard-journal.com/article/S2211-0348(22)00443-6/fulltext#bib0001

[28] Jo Lane, Huah Shin Ng, Carmel Poyser, Robyn M, Lucas Helen Tremlett; 2022; Multiple sclerosis incidence: A systematic review of change over time by geographical region; “References”; Accessed Nov 2022
https://www.msard-journal.com/article/S2211-0348(22)00443-6/fulltext#bib0001

[29] Matthew F. Cusick, Jane E. Libbey, and Robert S. Fujinami; 2022; Multiple Sclerosis: Autoimmunity and viruses; Accessed Nov 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4406397/

[30] National Institutes of Health, U.S. Department of Health and Human Services; 2022; Study suggests Epstein-Barr virus may cause multiple sclerosis; Accessed Nov 2022
https://www.nih.gov/news-events/nih-research-matters/study-suggests-epstein-barr-virus-may-cause-multiple-sclerosis

[31] K. I. Voumvourakis, P.C. Fragkou, D. K. Kitsos, K. Foska, M. Chondrogianni, and S. Tsiodras; 2022; Human herpesvirus 6 infection as a trigger of multiple sclerosis: an update of recent literature; Accessed Nov 2022
https://bmcneurol.biomedcentral.com/articles/10.1186/s12883-022-02568-7

[32] N Danchenko, J A Satia, and M S Anthony; 2016; Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden; Accessed Nov 2022
https://journals.sagepub.com/doi/10.1191/0961203306lu2305xx

[33] Marco Quaglia, Guido Merlotti, Marco De Andrea, Cinzia Borgogna, and Vincenzo Cantaluppi; 2021; Viral Infections and Systemic Lupus Erythematosus: New Players in an Old Story; Accessed Nov 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7916951/

[34] Prof Theo Vos; 2015; Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015; Accessed Nov 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5055577/

[35] Klaudia Farkas, Daniella Pigniczki, and Tamás Molnár; 2021; The complex relationship between viruses and inflammatory bowel disease; Accessed Nov 2022 https://journals.sagepub.com/doi/full/10.1177/1756284820988198

[36] Ken Cadwell, Khushbu K. Patel, Nicole S. Maloney, Ramnik Xavier, Thaddeus S. Stappenbeck, Herbert W. Virgin; 2010; Virus-Plus-Susceptibility Gene Interaction Determines Crohn's Disease Gene Atg16L1 Phenotypes in Intestine; Accessed Nov 2022 https://www.cell.com/cell/fulltext/S0092-8674(10)00545-3

[37] Giacomo Caio, Umberto Volta, Anna Sapone, Daniel A. Leffler, Roberto De Giorgio, Carlo Catassi, and Alessio Fasano; 2019; Celiac disease: a comprehensive current review; Accessed Nov 2022 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6647104/

[38] Daniel S Smyk, Andreas L Koutsoumpas, Maria G Mytilinaiou, Eirini I Rigopoulou, Lazaros I Sakkas, and Dimitrios P Bogdanos; 2014; Helicobacter pylori and autoimmune disease: Cause or bystander; Accessed Nov 2022 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3921471/

[39] Aleksandra Pyzik, Ewelina Grywalska, Beata Matyjaszek-Matuszek, Jarosław Ludian, Ewa Kiszczak-Bochyńska, Agata Smoleń, Jacek Roliński, and Dawid Pyzik; 2019; Does the Epstein–Barr Virus Play a Role in the Pathogenesis of Graves' Disease?; Accessed Nov 2022 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6650880/

[40] Steven E. Carsons, and Bhupendra C. Patel; 2022; Sjogren Syndrome; Accessed Nov 2022 https://www.ncbi.nlm.nih.gov/books/NBK431049/

[41] Kerrie Vaughan, Bjoern Peters, Kevin C. O'Connor, Roland Martin, and Alessandro Sette; 2013; A molecular view of multiple sclerosis and experimental autoimmune encephalitis: What can we learn from the epitope data?; Accessed Dec 2022 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4784960/

[42] EMBL's European Bioinformatics Institute: https://www.ebi.ac.uk/; Accessed Dec 2022

[43] UniProt: https://www.uniprot.org/; Accessed Dec 2022

[44] National Library of Medicine, BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi; Accessed Dec 2022

[45] Suhana Bedi, Tiffany M Richardson, Baofeng Jia, Hadeel Saab, Fiona S L Brinkman, Monica Westley; 2022; Similarities between bacterial GAD and human GAD65: Implications in gut mediated autoimmune type 1 diabetes; Accessed Dec 2022 https://pubmed.ncbi.nlm.nih.gov/35196314/

[46] Yi-Jiun Pan, Tzu-Lung Lin, Yen-Hua Chen, Chun-Ru Hsu, Pei-Fang Hsieh, Meng-Chuan Wu, and Jin-Town Wang; 2013; Capsular Types of Klebsiella pneumoniae Revisited by wzc Sequencing; Accessed Dec 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3857182/

[47] Daniel S Smyk, Andreas L Koutsoumpas, Maria G Mytilinaiou, Eirini I Rigopoulou, Lazaros I Sakkas, and Dimitrios P Bogdanos; 2014; Helicobacter pylori and autoimmune disease: Cause or bystander; Accessed Dec 2022
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3921471/

[48] Immune Epitope Database and Analysis Resource:
https://www.iedb.org/database_export_v3.php; Accessed Dec 2022

[49] Biocontainers: https://biocontainers.pro/; Accessed Dec 2022

[50] NCBI BLAST Cloud Documentation: http://ncbi.github.io/blast-cloud/; Accessed Dec 2022

[51] Bio.Blast.NCBIWWW Python module documentation:
https://biopython.org/docs/1.75/api/Bio.Blast.NCBIWWW.html; Accessed Dec 2022