# VaxiDB: A Database for Potential Antigens of Different Pathogens to Develop Novel Vaccines and Diagnostics

## HELP AND GUIDELINES

**CONTENTS**

## 1. OVERVIEW

**VaxiDB** is an innovative online database and tool that facilitates the identification and analysis of potential vaccine protein candidates while also enabling the design of customized vaccine formulations through its unique epitope reshuffling feature, providing users with all possible combinations of the input epitopes.

## 2. BACKGROUND

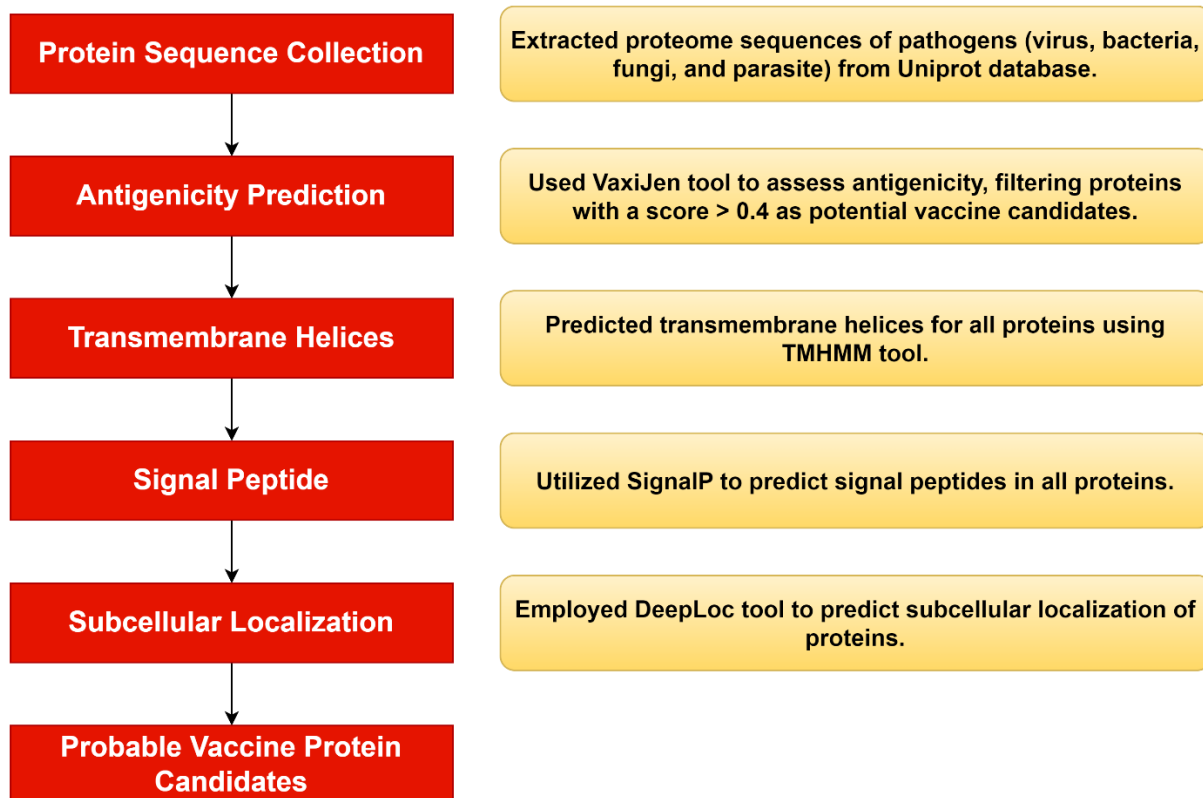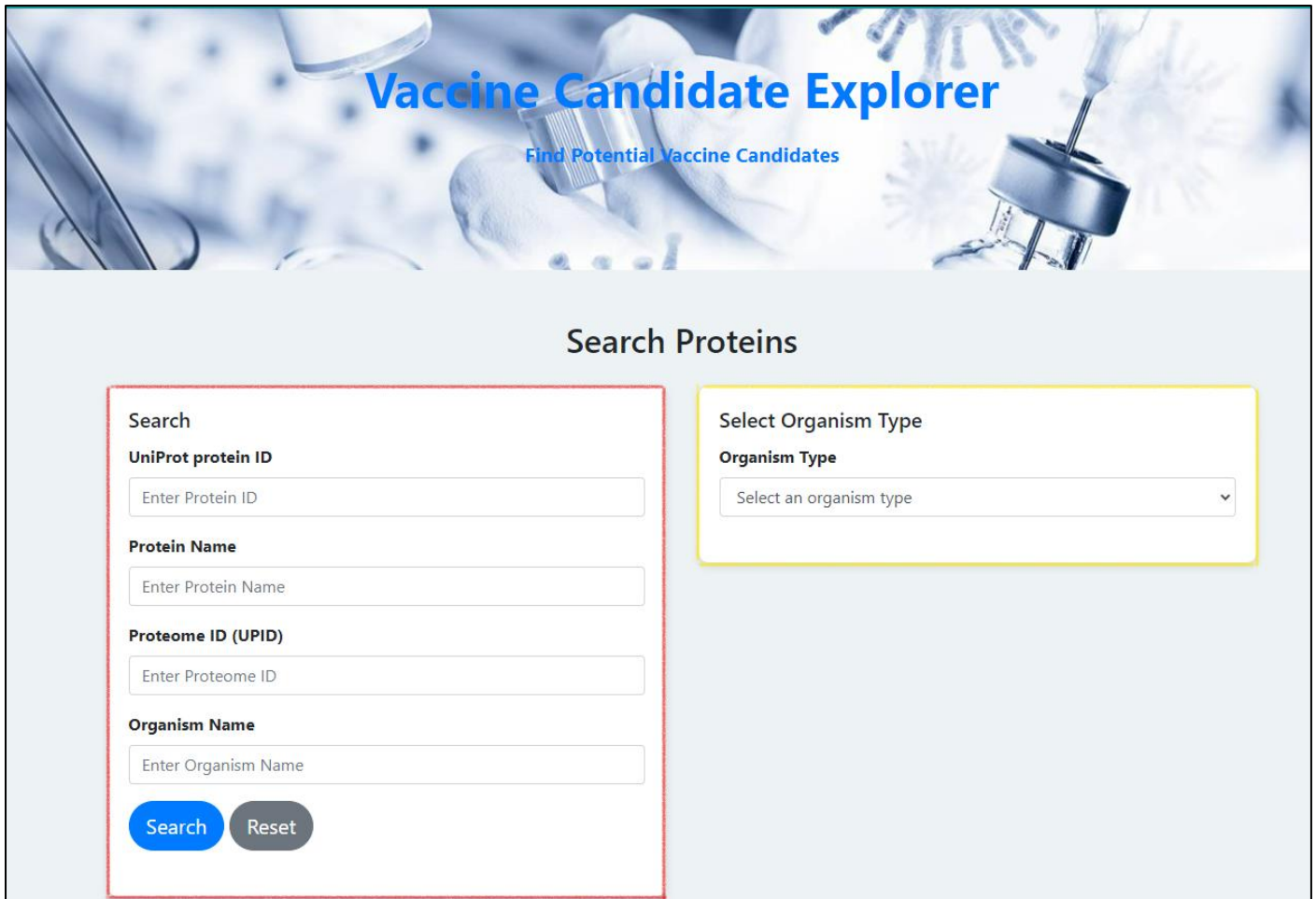Methodology used to train our model:



**Figure 1**: Flow chart of development of the database.

## 3. VACCINE CANDIDATE EXPLORER

As its name suggests, the Vaccine Candidate Explorer helps identify potential vaccine candidates through two approaches:

1. Users can manually input data related to the target protein, including any of the following: UniProt Protein ID, Protein Name, Proteome ID (UPID), or Organism Name.

2. Users can select organism types along with their corresponding Proteome ID and Organism Name from a pre-prepared list within the database.



**Figure 2**: The red box allows for input of either the UniProt Protein ID, Protein Name, Proteome ID (UPID), or Organism Name. The yellow box offers a more streamlined approach, enabling users to choose from predefined options.

**Procedure:**

To demonstrate how the 'Vaccine Candidate Explorer' can be implemented, we will use a classic case from the UniProt database, featuring *Pseudomonas aeruginosa*, the pathogen responsible for lung infections.

**3.1.** Enter the UniProt Protein ID in the box outlined under "**UniProt protein ID**" and hit the highlighted **"Search"** button

The UniProt Protein ID for a given protein is located after the file icon and before the dot its UniProt page as indicated in the figure below.



In this case, "Q9HWS1" has been entered.

**Result displayed:**



You may download the result in .xlsx format by clicking on the "**Download Results**" button.



**3.2.** Enter the name of the protein in the box labeled **'Protein Name.'** A new popup option will appear automatically; **select the organism** you are targeting (as the same protein can be present in different organisms), and then click the highlighted **'Search'** button.

The full name of the protein is displayed below the UniProt Protein ID on the right as indicated in the given figure.



In this example, **'EF-hand domain-containing protein'** has been entered. In the next image, you can see that the organism name or proteome ID option appears automatically.

**The search** is specific to the organism **Pseudomonas aeruginosa** (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1).



**Result Displayed:**

The **search** is specific to the organism Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1), which is why the result shows **only one entry.**



| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Experimentally Validated Antigen | PMID | Submission ID IEDB | DOI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Q9HWS1 | EF-hand domain-containing protein | 1.027 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Mitochondrion | Highly Probable | NA | NA | NA | NA |

| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Experi Valida Antige |
|---|---|---|---|---|---|---|---|---|
| Q9HWS1 | EF-hand domain-containing protein | 1.027 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Mitochondrion | Highly Probable | NA |
| A0A0J5WG44 | EF-hand domain-containing protein | 0.7111 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular | Highly Probable | NA |
| A0A0J5WX59 | EF-hand domain-containing protein | 0.5417 | ANTIGEN | Signal peptide (Sec/SPI) | 0 | Extracellular | Highly Probable | NA |

The result is not specific to the organism ***Pseudomonas aeruginosa*** (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1), it represents various organisms consisting of the protein of our interest i.e. **EF-hand domain-containing protein**, hence the data displayed is larger in size.

However, it is possible to filter it by clicking on the drop-down menu next to "**Vaccine Candidate Probability**" and clicking the "**Download Results**" option.

We'll be selecting the "**Highly Probable**" option for this case. This selection rearranges the Protein IDs and their corresponding information in descending order based on their Predicted Antigenic Score, with all scores above 0.5.



Download the result in .xlsx format by clicking on the "**Download Results"** button

As you can see, the data has been filtered, and coincidently the protein of interest (searched for in the previous step **3.2.**).

**3.3.** Enter the Proteome ID of the protein in the box outlined under "**Proteome ID (UPID)**" and hit the highlighted **"Search"** button

The Proteome ID of a protein can be found below "Proteomes" in the "Names & Taxonomy" section.



In this case, we'll be entering "UP000002438" as the Proteome ID

**Result Obtained:**



| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Experimentally Validated Antigen | PMI |
|---|---|---|---|---|---|---|---|---|---|
| Q9HWS1 | EF-hand domain-containing protein | 1.027 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Mitochondrion | Highly Probable | NA | NA |

**3.4.** Enter the FULL NAME of the protein in the box outlined under "**Organism Name**" and hit the highlighted **"Search"** button

The full name of the organism is displayed below the "Status" information of the protein on the right as indicated in the figure below.



In our case, we enter "*Pseudomonas aeruginosa* (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)".

The result is not specific to the protein **EF-hand domain-containing protein** organism instead it represents various protein in our proteome of interest i.e. *Pseudomonas aeruginosa* (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1), hence the data displayed is larger in size.

However, it is possible to filter it by clicking on the drop-down menu next to "**Vaccine Candidate Probability**" and clicking the "**Download Results**" option.



**Result Displayed:**



| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Exper Valida Antig |
|---|---|---|---|---|---|---|---|---|
| Q9HWS1 | EF-hand domain-containing protein | 1.027 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Mitochondrion | Highly Probable | NA |
| A0A0J5WG44 | EF-hand domain-containing protein | 0.7111 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular | Highly Probable | NA |
| A0A0J5WX59 | EF-hand domain-containing protein | 0.5417 | ANTIGEN | Signal peptide (Sec/SPI) | 0 | Extracellular | Highly Probable | NA |

## 3.6. Selection from Database

### 3.6.1. Exploring the options when selecting the type of organism, we wish to look into



You may select one of the following organisms corresponding to the target organism type of your choice:

- Virus
- Bacteria
- Parasite
- Fungus

Upon selecting each of the organism type, the second drop-down option gets affected - we get different dropdowns depending on the organism type we select.

**Virus:**

**Bacteria:**



**Parasite:**

**Fungus:**



Here, in our case upon selecting "Bacteria" as the organism type and "UP000002438 - Pseudomonas aeruginosa" option in the Proteome ID and Organism Name section.

**Results obtained via Selection from Database – same as in steps 3.1. and 3.3.:**



| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Experimentally Validated Antigen | PMI |
|---|---|---|---|---|---|---|---|---|---|
| Q9HWS1 | EF-hand domain-containing protein | 1.027 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Mitochondrion | Highly Probable | NA | NA |

**Please Note:** We can filter data based on whether the antigen has been experimentally validated.



The database authors perform necessary updates on a weekly basis, incorporating the latest research findings.

| Protein ID | Protein Name | Predicted Antigenic Score | Predicted Antigenic Probability | Predicted Signal Peptide | Number of TMHS | Localizations | Vaccine Candidate Probability | Experimentally Validated Antigen | PMID | Submission ID IEDB | DOI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Q9CC67 | Secreted p60-family protein | 0.5652 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular | Highly Probable | YES | 23283462 | NA | NA |
| P46842 | Alanine and proline-rich secreted protein Apa | 0.4885 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular | Highly Probable | YES | 23283462 | NA | NA |
| Q05868 | MPT51 antigen | 0.4657 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular\|Endoplasmic reticulum | Highly Probable | YES | 16339528 | NA | NA |
| Q9CD80 | Putative phthiocerol dimycocerosate transporter LppX | 0.6513 | ANTIGEN | Signal peptide (Sec/SPI) | 0 | Extracellular | Highly Probable | YES | 7525488 | NA | NA |
| Q49738 | D-ribose-binding protein | 0.5408 | ANTIGEN | Signal peptide (Sec/SPI) | 1 | Extracellular | Highly Probable | YES | 16339528 | NA | NA |
| Q9CBI7 | beta-lactamase | 0.5531 | ANTIGEN | Signal peptide (Sec/SPI) | 0 | Extracellular | Highly Probable | YES | NA | 1000638 | NA |

Users can download only experimentally validated and highly probable data.
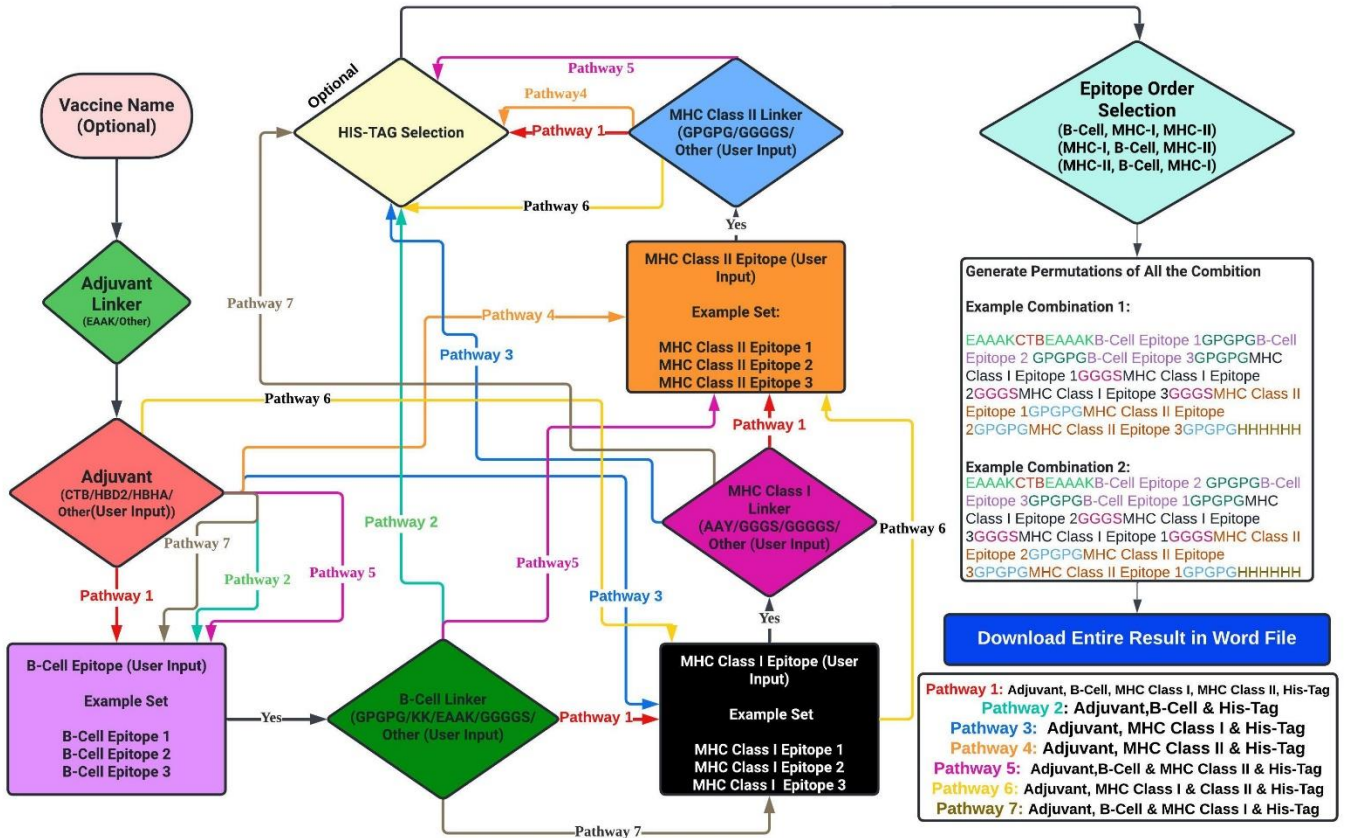
Additionally, they can apply any filters based on their preferences when downloading.

## 4. VACCINE DESIGNING

The vaccine designing tool is a comprehensive selection for customising and subsequently designing a vaccine.

**Guide to Vaccine Design Workflow**



Homepage:

**Procedure:**

## 4.6. Vaccine Name (Optional)

Vaccine Name (Optional)

My Vaccine

## 4.7. Configuring Adjuvant Options

### 4.7.1. Adjuvant Linker

The sequence that connects the adjuvant to the rest of the vaccine construct. Linkers in vaccine design ensure flexibility and proper spacing between the adjuvant and the antigenic epitopes to avoid steric hindrance, which could interfere with epitope presentation and immune recognition. These linkers are typically short peptide sequences. We have provided all the options that are already reported in various research studies. However, we also included an **'Others'** option in all the scenarios, allowing you to input your own if needed.

Select Options

Adjuvant Linker

| Select an option |
| Select an option |
| EAAAK |
| Other |

### 4.7.2. Adjuvant

An adjuvant is a substance used in vaccines to enhance the immune response to an antigen. In a multi-epitope vaccine, the adjuvant can help improve the immunogenicity of the otherwise small epitopes, ensuring a stronger and longer-lasting immune response. Common adjuvants include molecules like toll-like receptor agonists or cytokines. The role of the adjuvant in the construct is to help stimulate the immune system more efficiently.

Adjuvant

| Select an option |
| Select an option |
| Cholera toxin subunit B (CTB) |
| Human β-defensin 2 (HBD2) |
| Heparin-binding hemagglutinin (HBHA) |
| Other |
| NATAFGRAINRRVE |

Select Options

Adjuvant Linker

| EAAAK |

Adjuvant

| Human β-defensin 2 (HBD2) |

## 4.8. B Cell

### 4.8.1. B Cell Epitope

It is the specific part of an antigen that is recognized by B cells and their antibodies. In a vaccine, B cell epitopes are included to elicit a humoral immune response (production of antibodies). The goal is for B cells to recognize these epitopes and produce neutralizing antibodies against the pathogen or its toxin.

Select Options

**B Cell Epitope**

Enter B Cell Epitopes, each on a new line

### 4.8.2. B Cell Linker

It connects individual B cell epitopes or links the B cell epitope to other components of the vaccine (such as MHC epitopes or adjuvants). These linkers ensure that the epitopes are presented in an accessible and flexible manner, maximizing their ability to interact with B cells and elicit an antibody response.

**B Cell Linker**

Select an option

Select an option
GPGPG
KK
EAAAK
GGGGS
Other

Select Options

**B Cell Epitope**

TDAYNQKLSERRAN
NATAEGRAINRRVE

**B Cell Linker**

GGGGS

## 4.9. MHC Class I

### 4.9.1. MHC Class I Epitope

It is a short peptide (usually 8-10 amino acids) that is presented by MHC Class I molecules on the surface of antigen-presenting cells (APCs). These epitopes are recognized by CD8+ cytotoxic T cells, which are responsible for killing infected or cancerous cells. MHC Class I epitopes are crucial for inducing a cellular immune response against intracellular pathogens, like viruses.

Select Options

**MHC Class I Epitope**

Enter MHC Class I Epitopes, each on a new line

### 4.9.2. MHC Class I Linker

It connects multiple MHC Class I epitopes or links an MHC Class I epitope to other elements of the vaccine. The linker ensures that each epitope remains accessible for proper binding to MHC Class I molecules without interference from nearby sequences.

**MHC Class I Linker**

Select an option

| Select an option |
| AAY |
| GGGS |
| GGGGS |
| Other |

**Select Options**

**MHC Class I Epitope**

MAGE-1

**MHC Class I Linker**

AAY

## 4.10. MHC Class II
### 4.10.1. MHC Class II Epitope

A longer peptide (usually 13-18 amino acids) presented by MHC Class II molecules on the surface of APCs. These are recognized by CD4+ helper T cells, which assist in coordinating the overall immune response, including the activation of B cells and cytotoxic T cells. MHC Class II epitopes help stimulate a helper T cell response, promoting a stronger immune system coordination.

**Select Options**

**MHC Class II Epitope**

Enter MHC Class II Epitopes, each on a new line

### 4.10.2. MHC Class II Linker

A sequence used to connect different MHC Class II epitopes or link them to other components in the vaccine construct. These linkers provide flexibility and proper spacing to prevent structural constraints that could interfere with epitope presentation on MHC Class II molecules and recognition by helper T cells.

**MHC Class II Linker**

Select an option

| Select an option |
| GPGPG |
| GGGGS |
| Other |

**Select Options**

**MHC Class II Epitope**

HLA-DOB

**MHC Class II Linker**

GPGPG

## 4.11. HIS-TAG and Epitope Order
## 4.11.1. HIS-TAG

A sequence of histidine residues (usually 6) added to the protein construct for purification purposes. His-tags bind to nickel ions in a purification column, allowing the tagged protein to be easily isolated from other cellular proteins during the production process. His-tags are not typically part of the final vaccine product that is administered to patients, but they are crucial during vaccine development and purification stages.



## 4.11.2. Epitope Order

The sequential arrangement of B cell, MHC Class I, and MHC Class II epitopes within the multi-epitope vaccine construct. The correct ordering of epitopes is critical to ensure proper folding, epitope accessibility, and an effective immune response. The arrangement may affect how well the immune system processes and presents these epitopes, ultimately determining the strength and breadth of the immune response.

## 4.12. Result:

You may now download the results of the permutations of the combinations of your vaccine design and view it in doc file.