In Silico Design of a Tubulin Beta-2A Chain Protein Targeted Multi-Epitope Vaccine Against Colorectal Cancer

By

Md. Yousuf Wasif

Student ID:20146063

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (B. Pharm.)

School of Pharmacy

BRAC University

November 2024

Declaration

It is hereby declared that.

1. The thesis submitted is my own original work while completing a degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all the main sources of help.

Student's Full Name & Signature:

Md. Yousuf Wasif
Student ID: 20146063

Approval

The thesis titled “In Silico Design of a Tubulin Beta-2A Chain Protein Targeted Multi-Epitope Vaccine Against Colorectal Cancer” submitted by Md. Yousuf Wasif 20146063, of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy on …./12/2024.

Supervised By:

_____

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
BRAC University

Approved By:

_____

Dean:

A.F.M. Yusuf Haider, PhD
Acting Dean, School of Pharmacy
Professor, Department of Mathematics and Natural Sciences
BRAC University

Ethics Statement

All ethical standards were maintained in the completion of this thesis. Conducting the research with integrity and following the ethical principles was given utmost importance.

Abstract

Colorectal cancer (CRC) is a predominant cause of cancer-related mortality worldwide, requiring novel strategies for prevention and therapy. This paper details the in-silico design of a multi-epitope vaccine aimed at the tubulin beta-2A chain protein, exhibiting favorable antigenic characteristics for colorectal cancer immunotherapy. The vaccine design procedure utilized computational techniques to find cytotoxic T lymphocytes (CTL), helper T lymphocytes (HTL), and B-cell epitopes. Each epitope underwent evaluation for antigenicity, allergenicity, and toxicity, resulting in the identification of ideal candidates. Linkers were utilized to assemble the ultimate multi-epitope vaccination sequence, augmenting immunogenicity. Structural and biochemical stability were confirmed using homology modeling and immunoinformatic methods. Molecular docking demonstrated robust binding interactions between the vaccine and Toll-like receptor 3 (TLR3), corroborating its effectiveness in eliciting an immunological response. Immune simulations forecasted substantial development of antibodies and memory cells. The results suggest that the proposed vaccination has potential for addressing CRC, pending additional validation both in vitro and in vivo.

Keywords

Colorectal cancer (CRC), Multi-epitope vaccine, Tubulin beta-2A chain protein, In-silico design, Molecular docking.

Dedication

I would like to dedicate this to my wonderful parents, whose love and support have shaped me into who I am today. Forever grateful for your unwavering guidance and endless sacrifices.

Acknowledgement

I would like to thank Almighty Allah, who helped me achieve everything I have today.

I would like to express my love and gratitude to my parents who are truly deserved, And I appreciate the person I became today.

I would like to show my utmost gratitude to my supervisor Senior Lecturer Mohammad Kawsar Sharif Siam for guiding me in every step of this work. Without his support and motivation, I could not have passed this phase of my educational journey. In addition, I would like to show my sincere gratitude to all the faculty members of the School of Pharmacy, without whom I could not have accomplished my university journey.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CRC- Colorectal Cancer

CTL- Cytotoxic T-Lymphocyte

BA – Binding Affinity

HTL- Helper T-lymphocyte

APCs- Antigen Presenting Cell

GRAVY- Grand Average of Hydropathy

T3DB- The Toxin and Toxin-Target Database

Chapter 1

Introduction

Colorectal cancer (CRC) ranks as the second foremost cause of cancer-related mortality among both genders. But more than half of all occurrences and fatalities can be avoided because of modifiable risk factors like smoking, eating poorly, drinking too much alcohol, not exercising, and being overweight (Siegel et al., 2020b). Colorectal cancer (CRC) originates in the colon or the rectum. These malignancies may be referred to as colon cancer or rectal cancer, dependent upon their origin. Colon cancer and rectal cancer are frequently categorized together due to their numerous shared characteristics. Colorectal cancer typically progresses gradually over several years, initiating as benign growths known as polyps that may ultimately transform into malignant tumors if not addressed.

Colon cancer is the second leading cause of cancer-related deaths worldwide. In 2020, it was projected that over 1.9 million new cases of colorectal cancer and more than 930,000 fatalities from colorectal cancer occurred worldwide (Xi & Xu, 2021). Marked geographical variations in incidence and fatality rates were observed. The incidence rates were highest in Europe, Australia, and New Zealand, whereas the mortality rates were greatest in Eastern Europe (Ilic & Ilic, 2022). By 2040, the incidence of colorectal cancer is anticipated to reach 3.2 million new cases year (a 63% increase) and 1.6 million deaths per year (a 73% increase) (World Health Organization: WHO & World Health Organization: WHO, 2023).

Colorectal cancer in its initial phases may exhibit little or absent symptoms, complicating detection without routine screening. This highlights the significance of early detection, since the outcome for persons diagnosed in the initial stages of the disease is markedly superior to that of those whose cancer has metastasized. Typical signs of colorectal cancer (CRC) may encompass alterations in bowel habits, including chronic diarrhea or constipation, hematochezia, unexplained weight reduction, lethargy, and stomach discomfort. By McCulloch et al. (2019) Nonetheless, these symptoms may also signify other illnesses, underscoring the importance of regular screening, especially for individuals over 50 or with a familial history of colorectal cancer.

Multiple variables increase the risk of acquiring colorectal cancer. Factors include age, predominantly affecting individuals over 50, with lifestyle elements such as a diet rich in red and processed meats, insufficient physical activity, tobacco use, and excessive alcohol intake. Genetic predispositions and a personal history of inflammatory bowel illnesses, such as Crohn's disease or ulcerative colitis, elevate the risk of getting colorectal cancer (CRC) (Lewandowska et al., 2022).

Screening technologies, such as colonoscopies, fecal occult blood tests, and flexible sigmoidoscopy, have considerably improved early detection, leading to greater survival rates (Holme et al., 2013). Treatment generally encompasses surgery, chemotherapy, radiation therapy, and targeted therapies, based upon the cancer's stage and location. Progress in personalized

medicine is providing more customized strategies, enhancing patient outcomes (Krzyszczyk et al., 2018). Colorectal cancer persists as a significant public health concern; however, heightened awareness, routine screening, and advancements in therapy are enhancing survival rates and effective disease management.

Chapter 2

Method

This in-silico vaccine design for colorectal cancer followed a set of sequential procedures.

Figure 1 illustrates the process of constructing a vaccine against colorectal cancer.

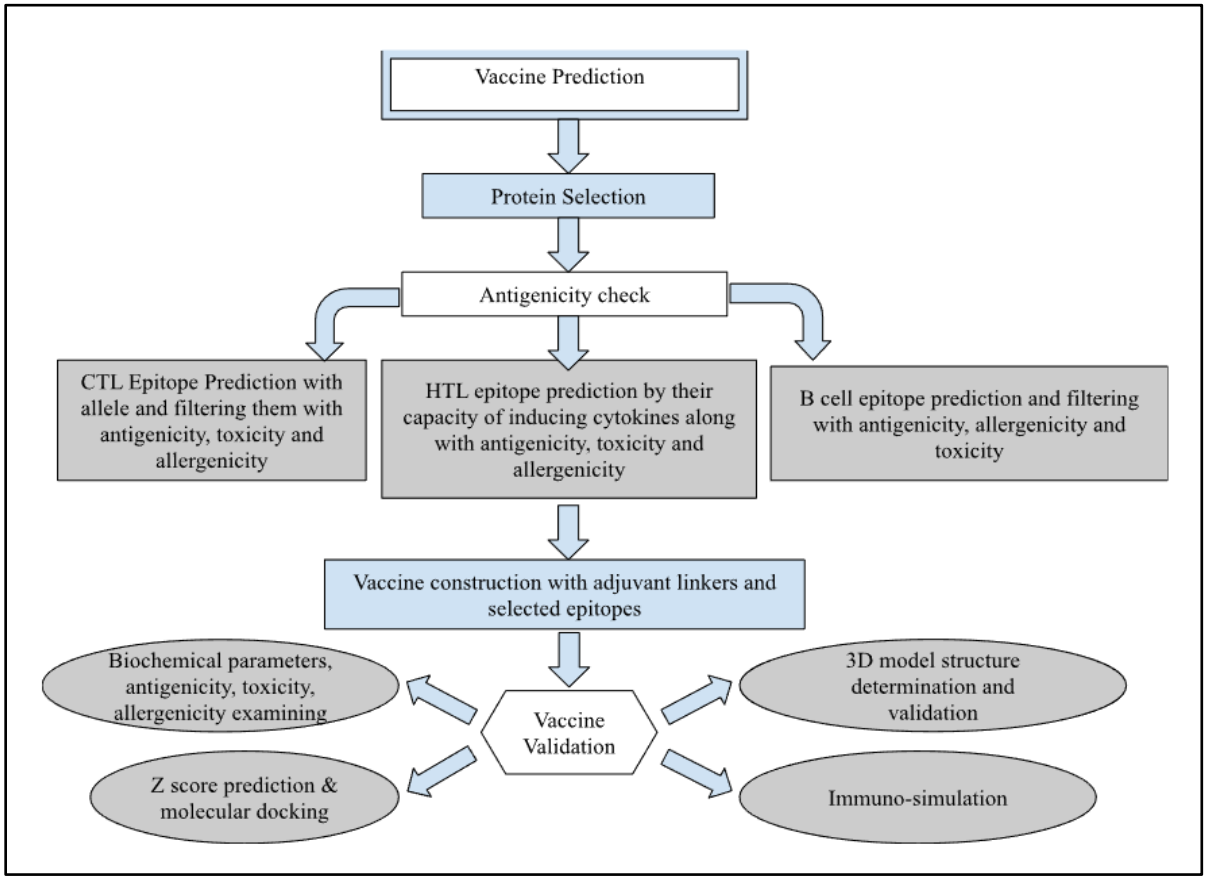


Figure 1: Sequential process used in in-silico vaccine design for Colorectal Cancer (CRC)

.

2.1 Literature Review

To achieve a comprehensive understanding of current research trends and findings in this field, search and review are being done on the Elsevier journals , Google Scholar, Scopus, PubMed, ResearchGate and MDPI. For searching relevant information certain keywords such as ‘colorectal cancer, antigenicity, allergenicity testing software, vaccine design, toxicity testing, Tubulin Beta-2 Chain, HTL, CTL, B cell, molecular docking, immunogenicity, MHC, in-silico process was used.

2.2 Choosing and Assembling of Envelope Protein Sequence of Colorectal Cancer

The right protein with antigenicity was chosen using the Uniport database. Tubulin beta-2A chain, the chosen target protein, was obtained from the Uniport database in FASTA format. The VaxiJen v2.0 server was utilized to evaluate antigenicity. Bacteria were selected as the target organism, and the threshold was set at 0.5 (Doytchinova & Flower, 2007). The protein was chosen for vaccine development due to its favorable amino acid length and quality.

2.3 The epitope of cytotoxic T lymphocytes (CTLs) and allele selection

M. V. Larsen et al. (2005) assert that CTL differentiates between healthy and infected cells. CTLs eradicate infections or malignant cells by releasing various cytokines, therefore influencing disease progression (Ito & Seishima, 2010). Cytotoxic T lymphocytes (CTLs) can recognize MHC class I molecules and trigger an immunological response due to the binding of CTL epitopes (M. V. Larsen et al., 2005). The Net CTL 1.2 server was utilized for CTL epitope prediction due to its

superior predictive efficacy. M. V. Larsen et al. (2007) discovered a CTL epitope with Tap transport capability for peptide translocation into the endoplasmic reticulum. All parameters were configured to their normal settings for server-based prediction, except for the sorting score, which was modified to the combined score. Protein sequences were entered in FASTA format in the designated field.

High-affinity alleles for MHC class I were identified using (https://services.healthtech.dtu.dk/services/NetMHCIIpan-4.1/) to assess CTL epitopes (Reynisson et al., 2020). Alleles were chosen for BA prediction after all CTL epitopes were entered in the designated field using the FASTA format and a 9-mer peptide length selection.

Furthermore, the toxicity, allergenicity, and antigenicity of the alleles were evaluated. The antigenicity of bacterial species was evaluated using the VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with a threshold set at 0.5 (Doytchinova & Flower, 2007). Individual CTL epitopes were employed to evaluate allergenicity, as per Dimitrov et al. (2013) (https://www.ddg-pharmfac.net/AllerTOP/). Batch submission was employed, and servers were utilized for toxicity evaluation (Gupta et al., 2013).

2.4 Selection of epitopes on HTL (Helper T-cell Lymphocytes)

HTL stimulates the generation of CTL and antibodies, inducing responses from both the humoral and cellular immune systems. They also secrete several lymphokines that target viruses, bacteria, parasites, and other pathogens. It is feasible to find antigens that bind to MHC class II (Alexander et al., 1998). A server (https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0) was

utilized to detect HTL epitopes (MontesGrajales & Olivero-Verbe, 2021). Alleles were chosen utilizing BA prediction, the protein sequence was inputted in the designated field, and the peptide length was configured at 9 Mer for server utilization. The methodology employed in CTL selection was utilized to assess the antigenicity, allergenicity, and toxicity of HTL epitopes.

2.5 HTL Epitopes Detection and Sorting

Specifically, HTL cells generate IFN-γ, which can regulate bacterial infections (Bao et al., 2014). Furthermore, they engage with antigen-presenting cells (APCs), stimulate macrophages, and influence the cell-mediated immune response (Walker et al., 2021). The IFN-γ-inducing capabilities assessment server was utilized (Kalkal et al., 2022). Interleukin-4 is crucial for the survival and proliferation of B lymphocytes. They also synthesize IgG1 and IgE from immunoglobulin. None et al. (2021) utilized the server (https://webs.iiitd.edu.in/raghava/il4pred/design.php) to assess the IL-4 epitope-inducing capabilities, maintaining all configurations at their default parameters. Helper T cells synthesize IL-10, which is crucial for averting autoimmune and inflammatory disorders, as well as for maintaining homeostasis after acute infections and tissue injury (Iyer & Cheng, 2012). In the HTL capability assessment, nine servers were utilized, with all modes maintained at their default configurations (Kalkal et al., 2022).

2.6 Forecasting B-cell Epitopes

B cells generate prolonged immunological protection via antibody synthesis. Linear epitopes of B lymphocytes comprise uninterrupted sequences of antigenic protein residues. The selection of linear B cell epitopes utilizes the server provided at http://tools.iedb.org/bcell/. Protein sequences were submitted in a straightforward format with the BepiPred linear epitope prediction approach (Jespersen et al., 2017). Antigenicity, allergenicity, and toxicity of B-cell epitopes were evaluated using the same methods as that used for CTL and HTL epitopes.

2.7 Vaccine Construction with Linkers

Linkers are condensed forms of amino acids. The lack of linkers will result in inadequate immunization outcomes, marked by low yields and possible misfolding (Gong et al., 2022). In favor of developing a multiepitope vaccine, various linkers were employed to join distinct epitopes and adjuvants. We connected the linker to the adjuvant and epitopes, associated the EAAAK linker with CTL, the GPGPG linker with HTL, and used the KK linker to unite B-cells (Ayyagari et al., 2022). These inflexible linkers establish appropriate spacing between proteins and various epitopes, so they minimize interactions and preserve their biological activity (Gong et al., 2022).

2.8 Validating the New Vaccine's Allergenicity and Toxicity

The antigenicity of the finished vaccine was reassessed using the designated server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), configured for bacterial organisms with a threshold of 0.5 (Doytchinova & Flower, 2007).

To evaluate the vaccine's toxicity, servers utilizing the sequence search feature were employed (Gong et al., 2022). The vaccination sequence was entered in FASTA format, with all parameters set to default. The allergenicity of the vaccine sequence was evaluated using servers that do sequence searches in FASTA format (Sircar et al., 2014) at http://www.allergenonline.org/.

2.9 Reviewing Designed Vaccine's Biochemical functionality

The service at (https://web.expasy.org/protparam/) was utilized to forecast physicochemical characteristics, including GRAVY, molecular weight, therapeutic pI, aliphatic index, and instability index. and expected half-life (Ayyagari et al., 2022). The study of N-terminal amino acids in protein sequences yields the half-life estimation results for humans, E. coli, and yeast. These results show how long it takes for the amount of protein in the cell to drop by 50% during synthesis.

When the rate is less than 40, the instability index shows stability in the protein or vaccine test tube. The cumulative hydropathy value of all amino acids divided by the total number of residues is referred to as GRAVY (Grand Average of Hydropathy) (Gasteiger et al., n.d.). Hydrophilicity is implied by a negative GRAVY score, while hydrophobicity is by a positive grade. The volume of the aliphatic side chain is referred to as the aliphatic index (H. Wang et al., 2021). The light absorption of proteins at a specific wavelength is related to the extinction coefficient (Gasteiger et al., n.d.).

2.10 Construction of a Vaccine Using 3D Models

The Phyr2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was employed in intensive mode for three-dimensional model generation from a single sequence, providing models via a synthesis of simplified ab initio folding simulations and multiple template modeling (Kelley et al., 2015).

2.11 Validation Analysis of Three-Dimensional Models

To validate three-dimensional models (https://swissmodel.expasy.org/), servers were employed that provide data on homology models, encompassing global and local quality assessments, ligands, and the Ramachandran plot, among other attributes. The Ramachandran plot predominantly depicts the amino acid residues of protein or vaccine structures (Waterhouse et al., 2018).

2.12 Z-Score Three-Dimensional Model Quality Analysis

The ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php) is applied to examine the quality of 3D models via z-score analysis and error verification. The server examines the overall quality score and presents it graphically, representing the score of the examined protein chain in the Protein Data Bank (Wiederstein & Sippl, 2007).

The z-score is calculated from server results, indicating the model's overall quality and quantifying the departure of the structure's total energy from the actual results of a random confirmation distribution (Wiederstein & Sippl, 2007).

2.13 Molecular Docking Assessment

Molecular docking is the process of evaluating binding affinity and interactions between

Molecular docking is the method used to evaluate the binding affinity and interactions between a developed vaccination and the human toll-like receptor. The human toll-like receptor 3 (TLR3) was obtained from the RCSB Protein Data Bank in PDB format (PDB ID: 3ILS). Molecular docking was conducted using the ClusPro server, as noted by Israel T. Desta (https://cluspro.org/home.php). The server implements the fast Fourier transform (FFT) and is denoted as PIPER, which sets one protein on a stationary grid and another on a dynamic grid.

The server enabled the modification of the distance between the receptor and ligand to replicate the binding process (Desta et al., 2020).

2.14 Immunogenicity Reflection

In the last step, immune response prediction was performed utilizing the C-immsim server (https://kraken.iac.rm.cnr.it/C-IMMSIM/) to evaluate the vaccine's capacity to induce immunogenicity in the organism (Rapin et al., 2010). The server depicts several immune responses of the humoral and cellular immune systems, encompassing B lymphocytes, T lymphocytes, and innate immune cells, in graphical representations (Rueckert & Guzmán, 2012). Three injection doses were provided to the server at time intervals of 1, 84, and 168 steps according to the immunization sequence. One step equates to an 8-hour duration, but the interval between two dosages is 4 weeks (Nain et al., 2020).

2.15 Material and method assertion

The design, prediction and verification of the colorectal cancer vaccine were conducted using in-silico methods. Various internet server tools were crucial for the development of vaccines, as extensively documented in diverse literature. Furthermore, the objective was to forecast an effective therapeutic vaccination for humanity; yet, uncertainties may arise, necessitating additional investigation.

Chapter 3

Results

3.1 Antigenicity & protein selection

Table 1: Selected protein information using UniProtKB

NAME OF THE PROTEIN	Tubulin beta-2A chain
GENE	TUBB2A
AMINO ACID	445
ORGANISM	Homo sapiens (Human)
STATUS	UniProtKB reviewed (Swiss-Prot)

Full sequence:
MREIVHIQAGQCGNQIGAKFWEVISDEHGID
PTGSYHGDSDLQLERINVYYNEAAGNKYVPR
AILVDLEPGTMDSVRSGPFGQIFRPDNFVFG
QSGAGNNWAKGHYTEGAELVDSVLDVVRKES
ESCDCLQGFQLTHSLGGGTGSGMGTLLISKI
REEYPDRIMNTFSVMPSPKVSDTVVEPYNAT
LSVHQLVENTDETYSIDNEALYDICFRTLKL
TTPTYGDLNHLVSATMSGVTTCLRFPGQLNA
DLRKLAVNMVPFPRLHFFMPGFAPLTSRGSQ
QYRALTVPELTQQMFDSKNMMAACDPRHGRY
LTVAAIFRGRMSMKEVDEQMLNVQNKNSSYF
VEWIPNNVKTAVCDIPPRGLKMSATFIGNST
AIQELFKRISEQFTAMFRRKAFLHWYTGEGM
DEMEFTEAESNMNDLVSEYQQYQDATADEQG
EFEEEEGEDEA

Additionally, antigenicity was assessed using the VaxiJen v2.0 program, yielding a result of 0.6009.

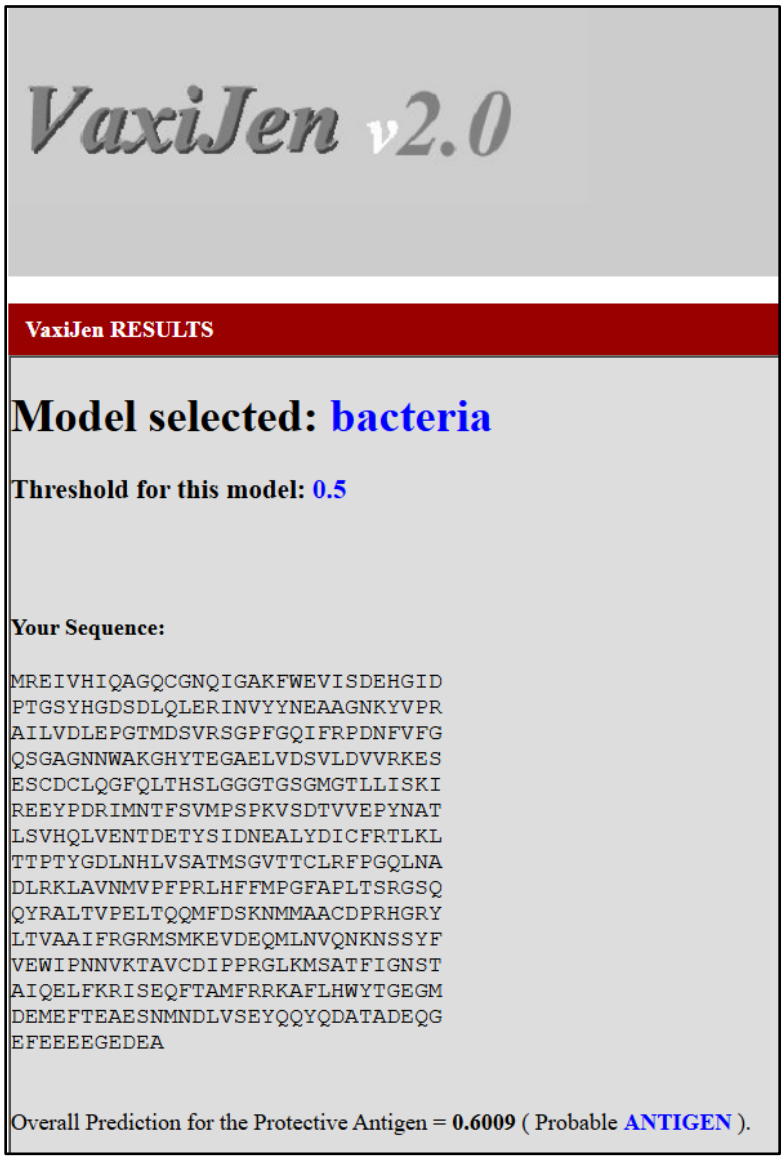
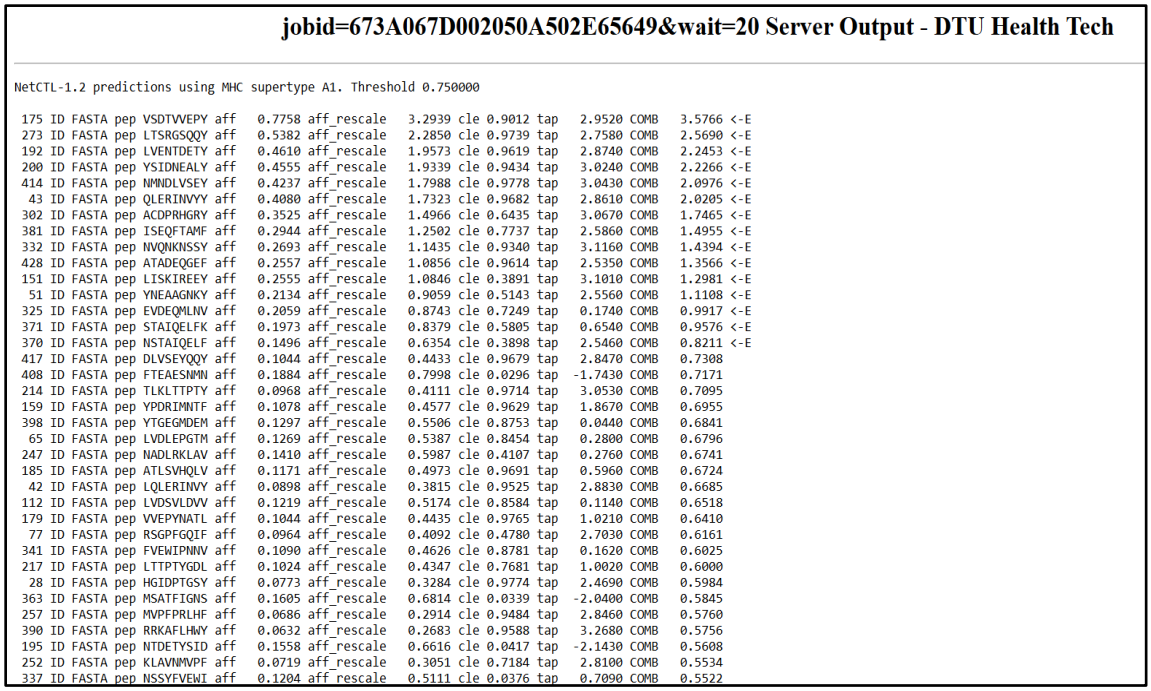


Figure 2: Antigenicity score of protein in VaxiJen v2.0 server (Doytchinova & Flower, 2007)

3.2 CTL and Allele Screening

The NetCTL1.2 software facilitated the identification of CTL epitopes using a threshold of 0.75, an A1 supertype, 0.05 tap transit efficiency, and a composite score (M. V. Larsen et al., 2007). The algorithm identified nine predicted CTL epitopes (Figure 3). Epitopes featuring <-E were chosen for the preliminary step.



jobid=673A067D002050A502E65649&wait=20 Server Output - DTU Health Tech

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

175	ID	FASTA	pep	VSDTVVEPY	aff	0.7758	aff_rescale	3.2939	cle 0.9012	tap	2.9520	COMB	3.5766 <-E
273	ID	FASTA	pep	LTSRGSQQY	aff	0.5382	aff_rescale	2.2850	cle 0.9739	tap	2.7580	COMB	2.5690 <-E
192	ID	FASTA	pep	LVENTDETY	aff	0.4610	aff_rescale	1.9573	cle 0.9619	tap	2.8740	COMB	2.2453 <-E
200	ID	FASTA	pep	YSIDNEALY	aff	0.4555	aff_rescale	1.9339	cle 0.9434	tap	3.0240	COMB	2.2266 <-E
414	ID	FASTA	pep	NMNDLVSEY	aff	0.4237	aff_rescale	1.7988	cle 0.9778	tap	3.0430	COMB	2.0976 <-E
43	ID	FASTA	pep	QLERINVYY	aff	0.4080	aff_rescale	1.7323	cle 0.9682	tap	2.8610	COMB	2.0205 <-E
302	ID	FASTA	pep	ACDPRHGRY	aff	0.3525	aff_rescale	1.4966	cle 0.6435	tap	3.0670	COMB	1.7465 <-E
381	ID	FASTA	pep	ISEQFTAMF	aff	0.2944	aff_rescale	1.2502	cle 0.7737	tap	2.5860	COMB	1.4955 <-E
332	ID	FASTA	pep	NVQNKNSSY	aff	0.2693	aff_rescale	1.1435	cle 0.9340	tap	3.1160	COMB	1.4394 <-E
428	ID	FASTA	pep	ATADEQGEF	aff	0.2557	aff_rescale	1.0856	cle 0.9614	tap	2.5350	COMB	1.3566 <-E
151	ID	FASTA	pep	LISKIREEY	aff	0.2555	aff_rescale	1.0846	cle 0.3891	tap	3.1010	COMB	1.2981 <-E
51	ID	FASTA	pep	YNEAAGNKY	aff	0.2134	aff_rescale	0.9059	cle 0.5143	tap	2.5560	COMB	1.1108 <-E
325	ID	FASTA	pep	EVDEQMLNV	aff	0.2059	aff_rescale	0.8743	cle 0.7249	tap	0.1740	COMB	0.9917 <-E
371	ID	FASTA	pep	STAIQELFK	aff	0.1973	aff_rescale	0.8379	cle 0.5805	tap	0.6540	COMB	0.9576 <-E
370	ID	FASTA	pep	NSTAIQELF	aff	0.1496	aff_rescale	0.6354	cle 0.3898	tap	2.5460	COMB	0.8211 <-E
417	ID	FASTA	pep	DLVSEYQQY	aff	0.1044	aff_rescale	0.4433	cle 0.9679	tap	2.8470	COMB	0.7308
408	ID	FASTA	pep	FTEAESNMN	aff	0.1884	aff_rescale	0.7998	cle 0.0296	tap	-1.7430	COMB	0.7171
214	ID	FASTA	pep	TLKLTTPTY	aff	0.0968	aff_rescale	0.4111	cle 0.9714	tap	3.0530	COMB	0.7095
159	ID	FASTA	pep	YPDRIMNTF	aff	0.1078	aff_rescale	0.4577	cle 0.9629	tap	1.8670	COMB	0.6955
398	ID	FASTA	pep	YTGEGMDEM	aff	0.1297	aff_rescale	0.5506	cle 0.8753	tap	0.0440	COMB	0.6841
65	ID	FASTA	pep	LVDLEPGTM	aff	0.1269	aff_rescale	0.5387	cle 0.8454	tap	0.2800	COMB	0.6796
247	ID	FASTA	pep	NADLRKLAV	aff	0.1410	aff_rescale	0.5987	cle 0.4107	tap	0.2760	COMB	0.6741
185	ID	FASTA	pep	ATLSVHQLV	aff	0.1171	aff_rescale	0.4973	cle 0.9691	tap	0.5960	COMB	0.6724
42	ID	FASTA	pep	LQLERINVY	aff	0.0898	aff_rescale	0.3815	cle 0.9525	tap	2.8830	COMB	0.6685
112	ID	FASTA	pep	LVDSVLDVV	aff	0.1219	aff_rescale	0.5174	cle 0.8584	tap	0.1140	COMB	0.6518
179	ID	FASTA	pep	VVEPYNATL	aff	0.1044	aff_rescale	0.4435	cle 0.9765	tap	1.0210	COMB	0.6410
77	ID	FASTA	pep	RSGPFGQIF	aff	0.0964	aff_rescale	0.4092	cle 0.4780	tap	2.7030	COMB	0.6161
341	ID	FASTA	pep	FVEWIPNNV	aff	0.1090	aff_rescale	0.4626	cle 0.8781	tap	0.1620	COMB	0.6025
217	ID	FASTA	pep	LTTPTYGDL	aff	0.1024	aff_rescale	0.4347	cle 0.7681	tap	1.0020	COMB	0.6000
28	ID	FASTA	pep	HGIDPTGSY	aff	0.0773	aff_rescale	0.3284	cle 0.9774	tap	2.4690	COMB	0.5984
363	ID	FASTA	pep	MSATFIGNS	aff	0.1605	aff_rescale	0.6814	cle 0.0339	tap	-2.0400	COMB	0.5845
257	ID	FASTA	pep	MVPFPRLHF	aff	0.0686	aff_rescale	0.2914	cle 0.9484	tap	2.8460	COMB	0.5760
390	ID	FASTA	pep	RRKAFLHWY	aff	0.0632	aff_rescale	0.2683	cle 0.9588	tap	3.2680	COMB	0.5756
195	ID	FASTA	pep	NTDETYSID	aff	0.1558	aff_rescale	0.6616	cle 0.0417	tap	-2.1430	COMB	0.5608
252	ID	FASTA	pep	KLAVNMVPF	aff	0.0719	aff_rescale	0.3051	cle 0.7184	tap	2.8100	COMB	0.5534
337	ID	FASTA	pep	NSSYFVEWI	aff	0.1204	aff_rescale	0.5111	cle 0.0376	tap	0.7090	COMB	0.5522

Figure 3: CTL prediction score with NetCTL1.2 software (M. V. Larsen et al., 2007)

We further filtered the epitopes derived from Figure 3 based on their combined score exceeding 0.6 and selected them for the second step. Table 2 below presents the evaluated CTL and their aggregated score.

Table 2: CTL epitopes with their combined score

CTL Epitopes	Combined score
VSDTVVEPY	3.5766
LTSRGSQQY	2.5690

LVENTDETY	2.2453
YSIDNEALY	2.2266
NMNDLVSEY	2.0976
QLERINVYY	2.0205
ACDPRHGRY	1.7465
ISEQFTAMF	1.4955
NVQNKNSSY	1.4394
ATADEQGEF	1.3566
LISKIREEY	1.2981
YNEAAGNKY	1.1108
EVDEQMLNV	0.9917
STAIQELFK	0.9576
NSTAIQELF	0.8211

3.3 Allele screening for CTL epitopes

Figure 4 (Reynisson et al., 2020) presents the results of our analysis of selected CTL epitopes for the second stage using the NetMHCPan4.1 service to determine their appropriate binding alleles. Among the six epitopes, five epitopes with binding alleles were identified, as presented in Table 3.

Figure 4:Strong Binding alleles prediction for MHC class-I on NETMHCpan 4.1 server (Reynisson et al., 2020)

Table 3: CTL epitopes with their MHC class-I alleles and ranks

Peptide sequence	Allele	Peptide	%Rank_EL	%Rank_BA	Aff(nM)
01	HLA-A*01:01	VSDTVVEPY	0.030	0.015	16.87
02	HLA-A*01:01	LTSRGSQQY	0.038	0.080	90.60
03	HLA-A*01:01	LVENTDETY	0.082	0.250	411.49
04	HLA-A*01:01	YSIDNEALY	0.054	0.046	48.66
05	HLA-A*01:01	NMNDLVSEY	0.210	0.422	874.90
06	HLA-A*01:01	QLERINVYY	0.103	0.246	401.37
07	HLA-A*01:01	ACDPRHGRY	0.210	0.587	1416.94
08	HLA-A*01:01	NVQNKNSSY	0.498	1.931	6595.28
09	HLA-A*01:01	YNEAAGNKY	0.239	0.530	1214.78
10	HLA-A*03:01	STAIQELFK	0.288	0.281	70.99

11	HLA-A*26:01	ATADEQGEF	0.272	0.446	1338.75
12	HLA-B*58:01	NSTAIQELF	0.329	0.323	56.86
13	HLA-B*15:01	LISKIREEY	0.174	0.771	149.77

3.4 Selection of final CTL epitopes based on toxicity, allergenicity, and antigenicity

Supplementary screening is conducted on epitopes with binding alleles based on their toxicity, allergenicity, and antigenicity. Toxinpred evaluation (Gupta et al., 2013), AllerTOP v2.0 (Dimitrov et al., 2013) for allergenicity, and VaxiJen v2.0 for antigenicity (Doytchinova & Flower, 2007) were utilized. Toxinpred utilized physicochemical criteria like molecular weight, hydrophobicity, hydrophilicity, CTL charge, and hydropathicity to assess toxicity. Figure 5 illustrates the results. The ultimate selection of CTL epitopes was determined by these criteria, leading to the identification of a single suitable CTL epitope (refer to Table 4).

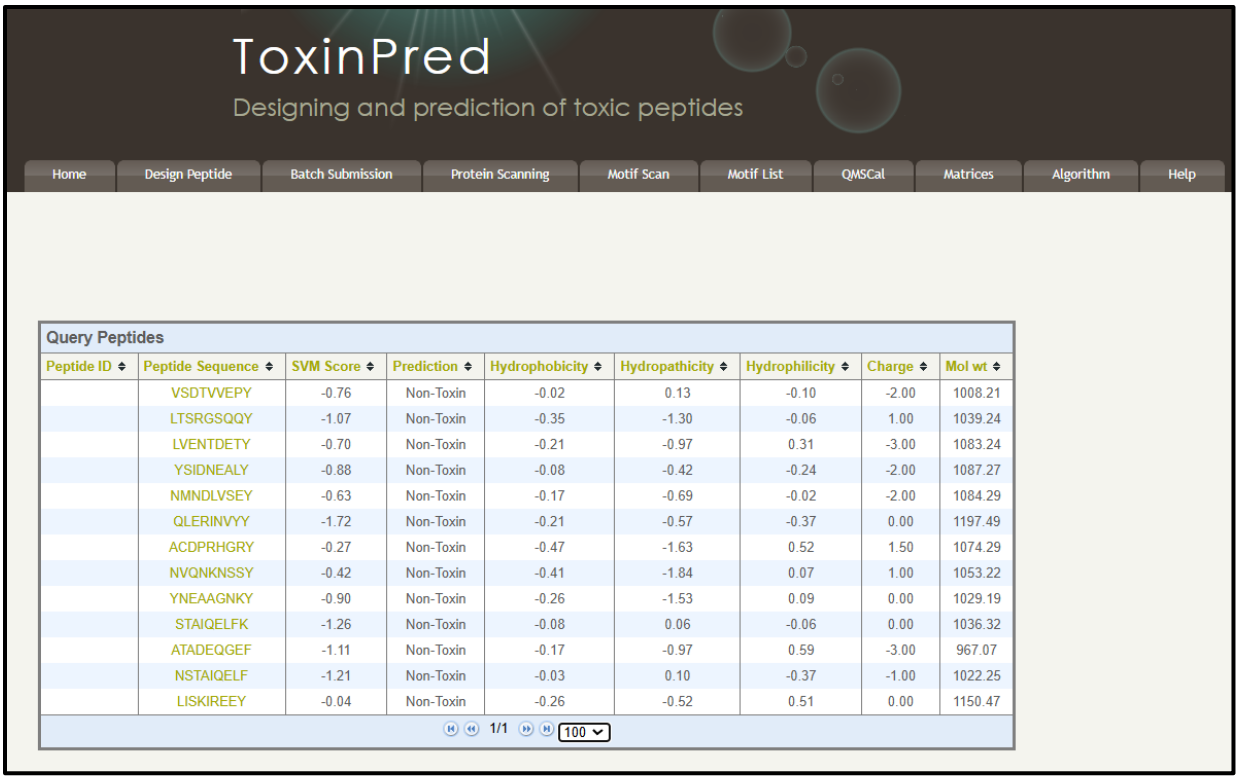


Figure 5:Toxicity prediction of CTL epitopes

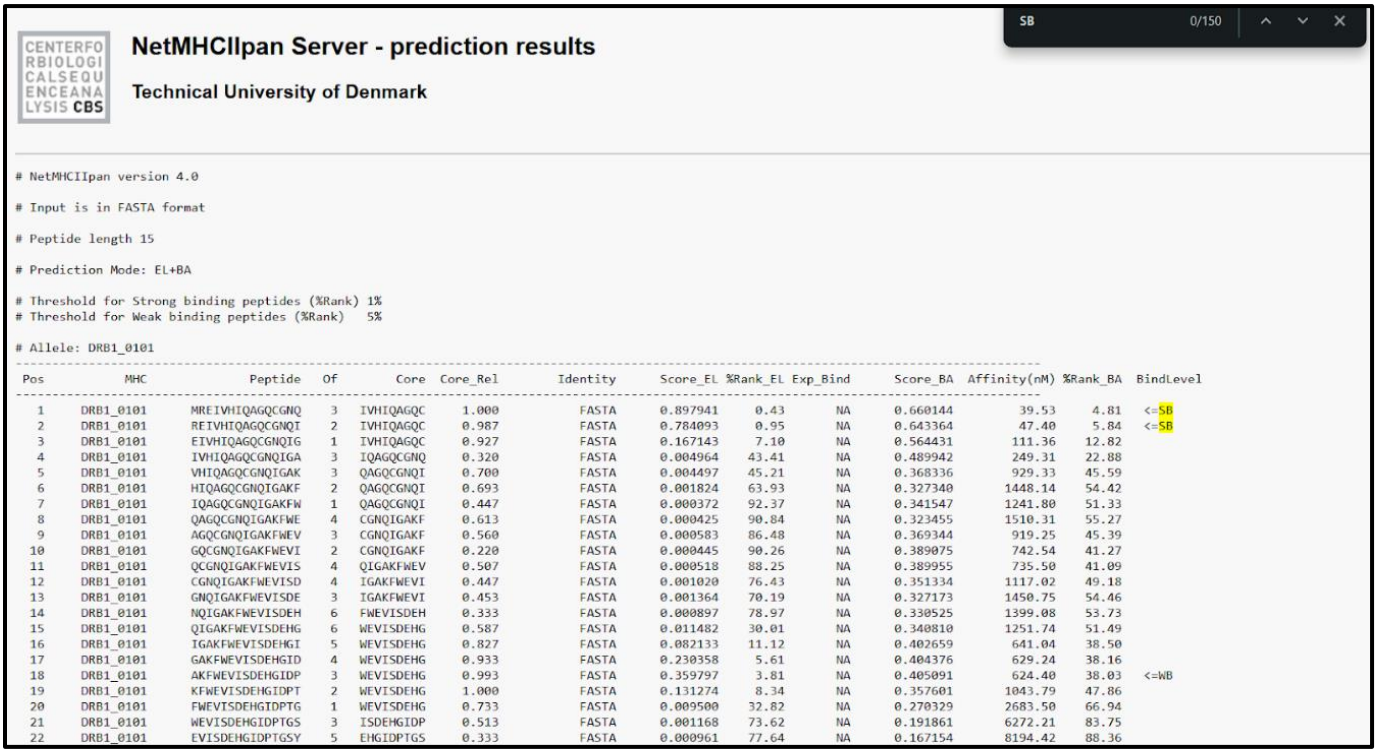
Table 4:Final CTL epitopes selection

Peptide	Antigenicity	Allergenicity	Toxicity
VSDTVVEPY	**Antigen**	**Non allergen**	**Non-toxin**
LTSRGSQQY	**Antigen**	**Non allergen**	**Non-toxin**
LVENTDETY	Non antigen	Allergen	Non-toxin
YSIDNEALY	Non antigen	Allergen	Non-toxin
NMNDLVSEY	Non antigen	Allergen	Non-toxin
QLERINVYY	Non antigen	Allergen	Non-toxin
ACDPRHGRY	Antigen	Allergen	Non-toxin
NVQNKNSSY	**Antigen**	**Non allergen**	**Non-toxin**
YNEAAGNKY	Antigen	Allergen	Non-toxin
STAIQELFK	Non antigen	Allergen	Non-toxin

ATADEQGEF	Antigen	Allergen	Non-toxin
NSTAIQELF	Antigen	Allergen	Non-toxin
LISKIREEY	Non antigen	Non allergen	Non-toxin

3.5 HTL Epitopes Detection and Sorting

To find HTL epitopes and alleles, the NetMHCIIpan 4.0 system server was used. Strong binding allele epitopes were found in Figure 6.



NetMHCIIpan Server - prediction results

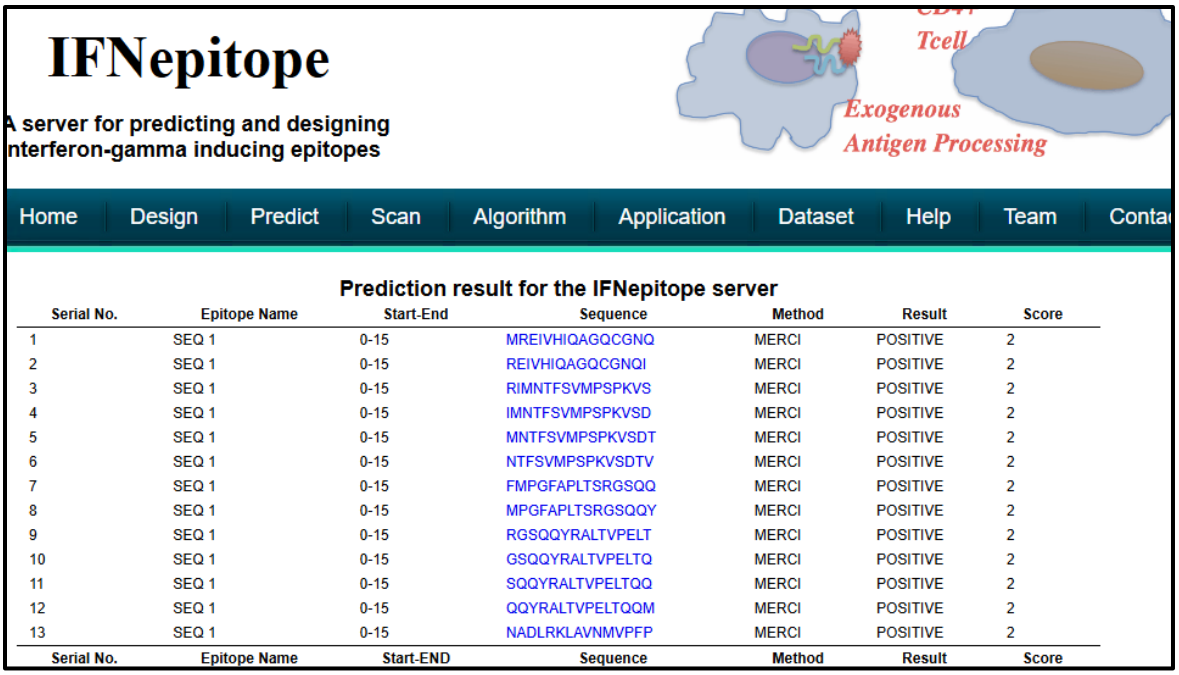
Technical University of Denmark

SB 0/150

```
# NetMHCIIpan version 4.0

# Input is in FASTA format

# Peptide length 15

# Prediction Mode: EL+BA

# Threshold for Strong binding peptides (%Rank) 1%
# Threshold for Weak binding peptides (%Rank)   5%

# Allele: DRB1_0101
```

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
1	DRB1_0101	MREIVHIQAGQCGNQ	3	IVHIQAGQC	1.000	FASTA	0.897941	0.43	NA	0.660144	39.53	4.81	<=SB
2	DRB1_0101	REIVHIQAGQCGNQI	2	IVHIQAGQC	0.987	FASTA	0.784093	0.95	NA	0.643364	47.40	5.84	<=SB
3	DRB1_0101	EIVHIQAGQCGNQIG	1	IVHIQAGQC	0.927	FASTA	0.167143	7.10	NA	0.564431	111.36	12.82	
4	DRB1_0101	IVHIQAGQCGNQIGA	3	IQAGQCGNQ	0.320	FASTA	0.004964	43.41	NA	0.489942	249.31	22.88	
5	DRB1_0101	VHIQAGQCGNQIGAK	3	QAGQCGNQI	0.700	FASTA	0.004497	45.21	NA	0.368336	929.33	45.59	
6	DRB1_0101	HIQAGQCGNQIGAKF	2	QAGQCGNQI	0.693	FASTA	0.001824	63.93	NA	0.327340	1448.14	54.42	
7	DRB1_0101	IQAGQCGNQIGAKFW	1	QAGQCGNQI	0.447	FASTA	0.000372	92.37	NA	0.341547	1241.80	51.33	
8	DRB1_0101	QAGQCGNQIGAKFWE	4	CGNQIGAKF	0.613	FASTA	0.000425	90.84	NA	0.323455	1510.31	55.27	
9	DRB1_0101	AGQCGNQIGAKFWEV	3	CGNQIGAKF	0.560	FASTA	0.000583	86.48	NA	0.369344	919.25	45.39	
10	DRB1_0101	GQCGNQIGAKFWEVI	2	CGNQIGAKF	0.220	FASTA	0.000445	90.26	NA	0.389075	742.54	41.27	
11	DRB1_0101	QCGNQIGAKFWEVIS	4	QIGAKFWEV	0.507	FASTA	0.000518	88.25	NA	0.389955	735.50	41.09	
12	DRB1_0101	CGNQIGAKFWEVISD	4	IGAKFWEVI	0.447	FASTA	0.001020	76.43	NA	0.351334	1117.02	49.18	
13	DRB1_0101	GNQIGAKFWEVISDE	3	IGAKFWEVI	0.453	FASTA	0.001364	70.19	NA	0.327173	1450.75	54.46	
14	DRB1_0101	NQIGAKFWEVISDEH	6	FWEVISDEH	0.333	FASTA	0.000897	78.97	NA	0.330525	1399.08	53.73	
15	DRB1_0101	QIGAKFWEVISDEHG	6	WEVISDEHG	0.587	FASTA	0.011482	30.01	NA	0.340810	1251.74	51.49	
16	DRB1_0101	IGAKFWEVISDEHGI	5	WEVISDEHG	0.827	FASTA	0.082133	11.12	NA	0.402659	641.04	38.50	
17	DRB1_0101	GAKFWEVISDEHGID	4	WEVISDEHG	0.933	FASTA	0.230358	5.61	NA	0.404376	629.24	38.16	
18	DRB1_0101	AKFWEVISDEHGIDP	3	WEVISDEHG	0.993	FASTA	0.359797	3.81	NA	0.405091	624.40	38.03	<=WB
19	DRB1_0101	KFWEVISDEHGIDPT	2	WEVISDEHG	1.000	FASTA	0.131274	8.34	NA	0.357601	1043.79	47.86	
20	DRB1_0101	FWEVISDEHGIDPTG	1	WEVISDEHG	0.733	FASTA	0.009500	32.82	NA	0.270329	2683.50	66.94	
21	DRB1_0101	WEVISDEHGIDPTGS	3	ISDEHGIDP	0.513	FASTA	0.001168	73.62	NA	0.191861	6272.21	83.75	
22	DRB1_0101	EVISDEHGIDPTGSY	5	EHGIDPTGS	0.333	FASTA	0.000961	77.64	NA	0.167154	8194.42	88.36	

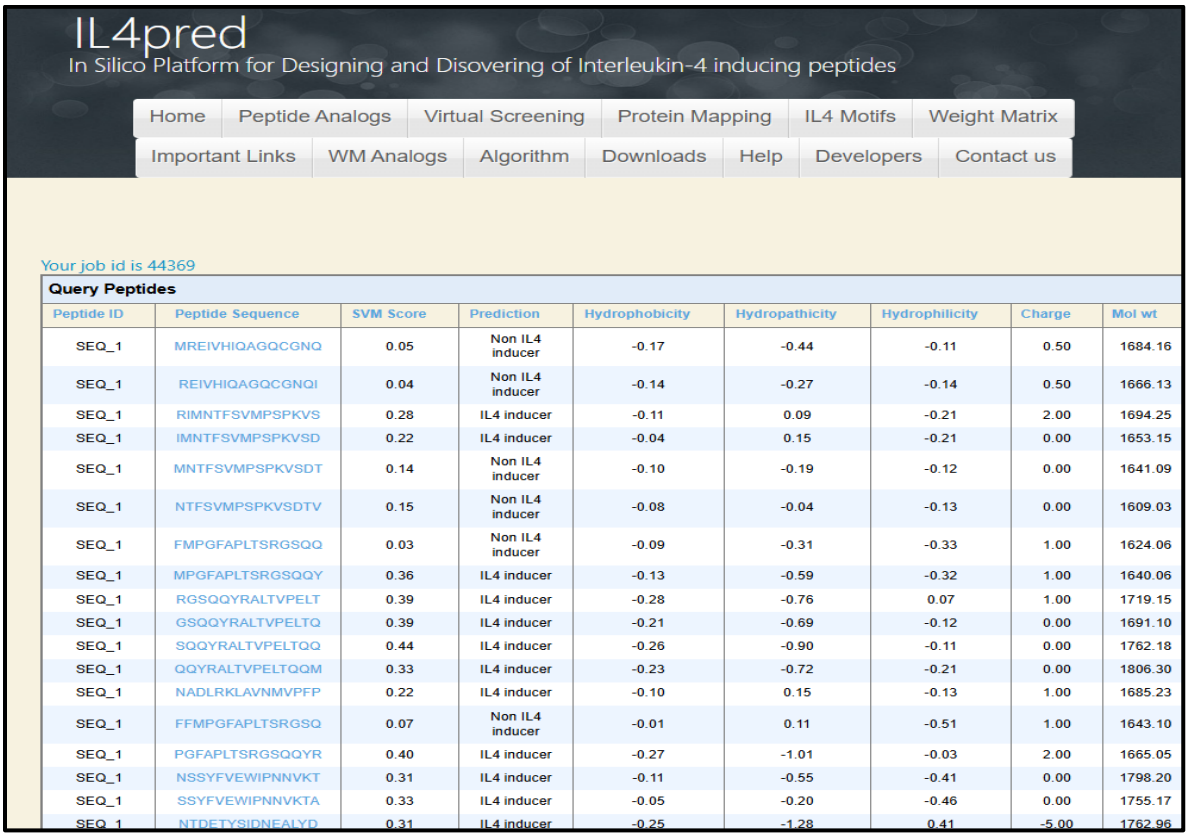
Figure 6:HTL epitopes detection for MHC class-II on NetMHCIIpan4.0 (Reynisson et al., 2020)



Prediction result for the IFNepitope server

Serial No.	Epitope Name	Start-End	Sequence	Method	Result	Score
1	SEQ 1	0-15	MREIVHIQAGQCGNQ	MERCI	POSITIVE	2
2	SEQ 1	0-15	REIVHIQAGQCGNQI	MERCI	POSITIVE	2
3	SEQ 1	0-15	RIMNTFSVMPSPKVS	MERCI	POSITIVE	2
4	SEQ 1	0-15	IMNTFSVMPSPKVSD	MERCI	POSITIVE	2
5	SEQ 1	0-15	MNTFSVMPSPKVSDT	MERCI	POSITIVE	2
6	SEQ 1	0-15	NTFSVMPSPKVSDTV	MERCI	POSITIVE	2
7	SEQ 1	0-15	FMPGFAPLTSRGSQQ	MERCI	POSITIVE	2
8	SEQ 1	0-15	MPGFAPLTSRGSQQY	MERCI	POSITIVE	2
9	SEQ 1	0-15	RGSQQYRALTVPELT	MERCI	POSITIVE	2
10	SEQ 1	0-15	GSQQYRALTVPELTQ	MERCI	POSITIVE	2
11	SEQ 1	0-15	SQQYRALTVPELTQQ	MERCI	POSITIVE	2
12	SEQ 1	0-15	QQYRALTVPELTQQM	MERCI	POSITIVE	2
13	SEQ 1	0-15	NADLRKLAVNMVPFP	MERCI	POSITIVE	2
Serial No.	Epitope Name	Start-END	Sequence	Method	Result	Score

Figure 7:IFNγ inducing abilities inspecting by IFNepitope server (Dhanda, Vir, et al., 2013)



Your job id is 44369

Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
SEQ_1	MREIVHIQAGQCGNQ	0.05	Non IL4 inducer	-0.17	-0.44	-0.11	0.50	1684.16
SEQ_1	REIVHIQAGQCGNQI	0.04	Non IL4 inducer	-0.14	-0.27	-0.14	0.50	1666.13
SEQ_1	RIMNTFSVMPSPKVS	0.28	IL4 inducer	-0.11	0.09	-0.21	2.00	1694.25
SEQ_1	IMNTFSVMPSPKVSD	0.22	IL4 inducer	-0.04	0.15	-0.21	0.00	1653.15
SEQ_1	MNTFSVMPSPKVSDT	0.14	Non IL4 inducer	-0.10	-0.19	-0.12	0.00	1641.09
SEQ_1	NTFSVMPSPKVSDTV	0.15	Non IL4 inducer	-0.08	-0.04	-0.13	0.00	1609.03
SEQ_1	FMPGFAPLTSRGSQQ	0.03	Non IL4 inducer	-0.09	-0.31	-0.33	1.00	1624.06
SEQ_1	MPGFAPLTSRGSQQY	0.36	IL4 inducer	-0.13	-0.59	-0.32	1.00	1640.06
SEQ_1	RGSQQYRALTVPELT	0.39	IL4 inducer	-0.28	-0.76	0.07	1.00	1719.15
SEQ_1	GSQQYRALTVPELTQ	0.39	IL4 inducer	-0.21	-0.69	-0.12	0.00	1691.10
SEQ_1	SQQYRALTVPELTQQ	0.44	IL4 inducer	-0.26	-0.90	-0.11	0.00	1762.18
SEQ_1	QQYRALTVPELTQQM	0.33	IL4 inducer	-0.23	-0.72	-0.21	0.00	1806.30
SEQ_1	NADLRKLAVNMVPFP	0.22	IL4 inducer	-0.10	0.15	-0.13	1.00	1685.23
SEQ_1	FFMPGFAPLTSRGSQ	0.07	Non IL4 inducer	-0.01	0.11	-0.51	1.00	1643.10
SEQ_1	PGFAPLTSRGSQQYR	0.40	IL4 inducer	-0.27	-1.01	-0.03	2.00	1665.05
SEQ_1	NSSYFVEWIPNNVKT	0.31	IL4 inducer	-0.11	-0.55	-0.41	0.00	1798.20
SEQ_1	SSYFVEWIPNNVKTA	0.33	IL4 inducer	-0.05	-0.20	-0.46	0.00	1755.17
SEQ_1	NTDETYSIDNEALYD	0.31	IL4 inducer	-0.25	-1.28	0.41	-5.00	1762.96

Figure 8: IL-4 inducing abilities inspecting on IL4pred server (Dhanda, Gupta, et al., 2013)

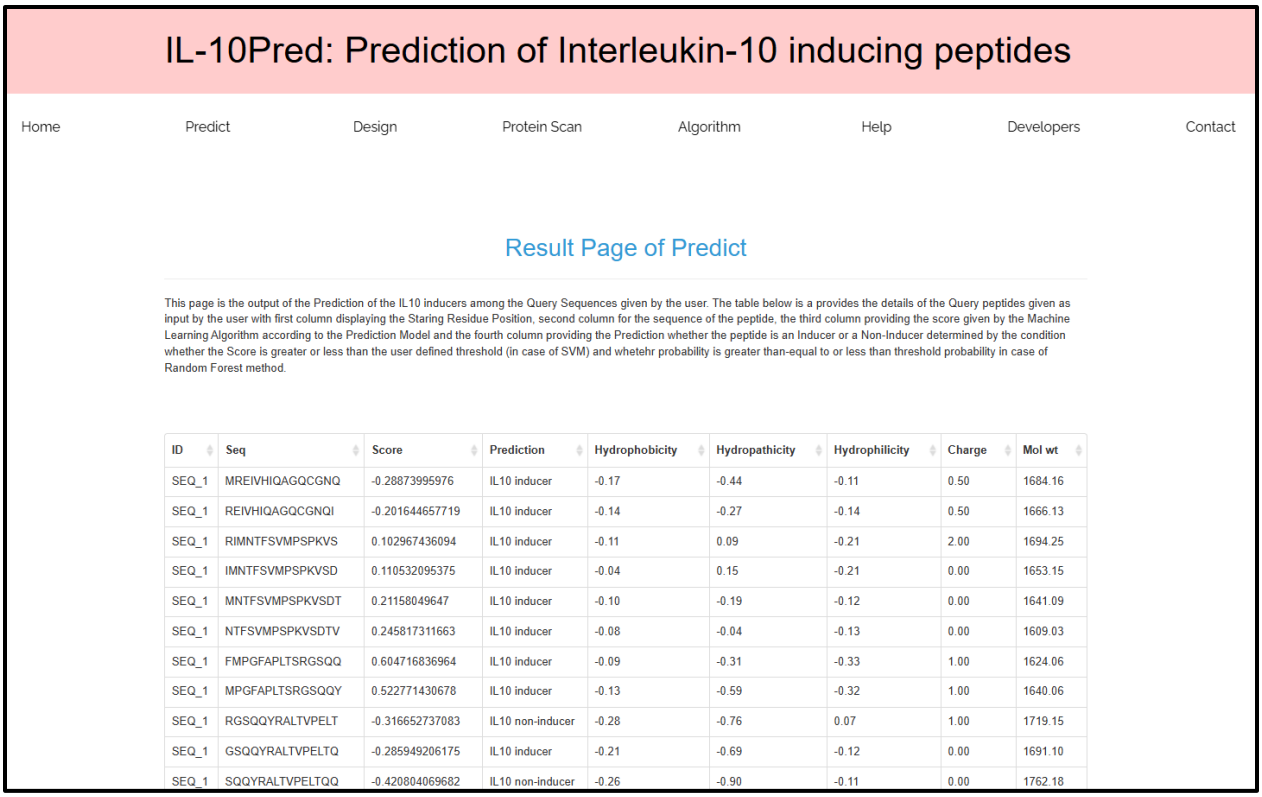


Figure 9: IL-10 inducing abilities inspecting on IL-10pred server (Nagpal et al., 2017)

Table 5:HTL Sorting with different parameters

Htl peptide sequence	**Antigenicity**	**Allergenicity**	**Toxicity**	**IFNγ**	**IL-4**	**IL10**
MREIVHIQAGQCGNQ	Antigen	Allergen	Non-Toxin	Positive	Non-inducer	Inducer
REIVHIQAGQCGNQI	Antigen	Allergen	Non-Toxin	Positive	Non-inducer	Inducer
RIMNTFSVMPSPKVS	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
IMNTFSVMPSPKVSD	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
MNTFSVMPSPKVSDT	Antigen	Non allergen	Non-Toxin	Positive	Non-inducer	Inducer

NTFSVMPSPKVSDTV	Antigen	Non allergen	Non-Toxin	Positive	Non-inducer	Inducer
FMPGFAPLTSRGSQQ	Antigen	Non allergen	Non-Toxin	Positive	Non-inducer	Inducer
MPGFAPLTSRGSQQY	**Antigen**	**Non allergen**	**Non-Toxin**	**Positive**	**Inducer**	**Inducer**
RGSQQYRALTVPELT	Antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
GSQQYRALTVPELTQ	Non antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
SQQYRALTVPELTQQ	Non antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
QQYRALTVPELTQQM	Non antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
NADLRKLAVNMVPFP	**Antigen**	**Non allergen**	**Non-Toxin**	**Positive**	**Inducer**	**Inducer**
FFMPGFAPLTSRGSQ	Antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
PGFAPLTSRGSQQYR	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
NSSYFVEWIPNNVKT	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Non-inducer
SSYFVEWIPNNVKTA	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Non-inducer
NTDETYSIDNEALYD	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
TDETYSIDNEALYDI	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
RFPGQLNADLRKLAV	Non antigen	Allergen	Non-Toxin	Negative	Non-inducer	Inducer
FPGQLNADLRKLAVN	Antigen	Allergen	Non-Toxin	Negative	Non-inducer	Inducer

PGQLNADLRKLAVNM	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
PTGSYHGDSDLQLER	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
TGSYHGDSDLQLERI	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
DETYSIDNEALYDIC	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
ETYSIDNEALYDICF	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
QIGAKFWEVISDEHG	Antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
IGAKFWEVISDEHGI	Antigen	Non allergen	Non-Toxin	Positive	Inducer	Non-inducer
GAKFWEVISDEHGID	Antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
AKFWEVISDEHGIDP	Antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
LQGFQLTHSLGGGTG	Antigen	Allergen	Non-Toxin	Negative	Non-inducer	Inducer
VEPYNATLSVHQLVE	Non antigen	Allergen	Non-Toxin	Negative	Non-inducer	Non-inducer
QLERINVYYNEAAGN	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
ERINVYYNEAAGNKY	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
RINVYYNEAAGNKYV	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
PPRGLKMSATFIGNS	Non antigen	Non allergen	Non-Toxin	Positive	Inducer	Non-inducer
PRGLKMSATFIGNST	Non antigen	Non allergen	Non-Toxin	Positive	Inducer	Non-inducer

KFWEVISDEHGIDPT	Antigen	Allergen	Non-Toxin	Positive	Inducer	Non-inducer
DCLQGFQLTHSLGGGT	Antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
QGFQLTHSLGGGTGS	Antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
VVEPYNATLSVHQLV	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
AAGNKYVPRAILVDL	Antigen	Allergen	Non-Toxin	Negative	Non-inducer	Inducer
AGNKYVPRAILVDLE	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
GNKYVPRAILVDLEP	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
NKYVPRAILVDLEPG	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
INVYYNEAAGNKYVP	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
NVYYNEAAGNKYVPR	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
KRISEQFTAMFRRKA	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
RISEQFTAMFRRKAF	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
ISEQFTAMFRRKAFL	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
SEQFTAMFRRKAFLH	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
LVDSVLDVVRKESES	Non antigen	Allergen	Non-Toxin	Negative	Non-inducer	Inducer
GNSTAIQELFKRISE	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer

NSTAIQELFKRISEQ	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
STAIQELFKRISEQF	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
LNADLRKLAVNMVPF	Antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
MGTLLISKIREEYPD	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
GTLLISKIREEYPDR	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
KLAVNMVPFPRLHFF	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
LAVNMVPFPRLHFFM	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
AVNMVPFPRLHFFMP	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
VNMVPFPRLHFFMPG	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
LERINVYYNEAAGNK	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
DSVLDVVRKESESCD	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
DLQLERINVYYNEAA	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
VSDTVVEPYNATLSV	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Non-inducer
SDTVVEPYNATLSVH	Antigen	Allergen	Non-Toxin	Negative	Non-inducer	Non-inducer
DTVVEPYNATLSVHQ	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Non-inducer
TVVEPYNATLSVHQL	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Non-inducer

RKAFLHWYTGEGMDE	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
KAFLHWYTGEGMDEM	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
RRKAFLHWYTGEGMD	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
VSEYQQYQDATADEQ	Antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
DPTGSYHGDSDLQLE	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
GSYHGDSDLQLERIN	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
ENTDETYSIDNEALY	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
SYFVEWIPNNVKTAV	Non antigen	Non allergen	Non-Toxin	Positive	Inducer	Non-inducer
YFVEWIPNNVKTAVC	Non antigen	Allergen	Non-Toxin	Negative	Non-inducer	Non-inducer
FVEWIPNNVKTAVCD	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Non-inducer
VEWIPNNVKTAVCDI	Non antigen	Allergen	Non-Toxin	Negative	Non-inducer	Non-inducer
KMSATFIGNSTAIQE	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
KMSATFIGNSTAIQE	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
SATFIGNSTAIQELF	Non antigen	Non allergen	Non-Toxin	Negative	Non-inducer	Inducer
ATFIGNSTAIQELFK	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
GAELVDSVLDVVRKE	Non antigen	Allergen	Non-Toxin	Positive	Non-inducer	Inducer

LISKIREEYPDRIMN	Non antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
ISKIREEYPDRIMNT	Non antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
FRGRMSMKEVDEQML	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
RGRMSMKEVDEQMLN	Antigen	Allergen	Non-Toxin	Negative	Inducer	Inducer
MKEVDEQMLNVQNKN	Antigen	Non allergen	Non-Toxin	Negative	Inducer	Inducer
PRHGRYLTVAAIFRG	Non antigen	Allergen	Non-Toxin	Positive	Non-inducer	Inducer
RHGRYLTVAAIFRGR	Non antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer
HGRYLTVAAIFRGRM	Non antigen	Allergen	Non-Toxin	Positive	Inducer	Inducer

3.6 Forecasting B-cell Epitopes

The Bepipred Linear Epitope Prediction 2.0 algorithm employed in the IEDB Analysis Resource produced the following outcome for the fundamental protein sequence. Fourteen peptides were recognized as B cell epitopes, as outlined in Table 6.

Table 6: Predicted B-cells and Sorting

B cell	Start	End	Length	Antigenicity	Allergenicity	Toxicity
TLKLTTP	**214**	**220**	**7**	**Antigen**	**Non allergen**	**Non-toxin**
FPGQLNAD	242	249	8	Non antigen	Allergen	Non-toxin
MSMKEVDE	321	328	8	Antigen	Allergen	Non-toxin
VCDIPPRGL	**353**	**361**	**9**	**Antigen**	**Non allergen**	**Non-toxin**
QNKNSSYFVE	334	343	10	Non antigen	Allergen	Non-toxin
YTGEGMDEMEF	**398**	**408**	**11**	**Antigen**	**Non allergen**	**Non-toxin**
LTSRGSQQYRALTVPE	273	288	16	Non antigen	Non allergen	Non-toxin
PGTMDSVRSGPFGQIFR	70	86	17	Non antigen	Non allergen	Non-toxin
QSGAGNNWAKGHYTEGAELVDS	94	115	22	Antigen	Allergen	Non-toxin
SEYQQYQDATADEQGEFEEEEGE	420	442	23	Antigen	Non allergen	Non-toxin
GIDPTGSYHGDSDLQLERINVYYNEAAG	29	56	28	Antigen	Allergen	Non-toxin

The server provided a plotted graph illustrating the average, minimum, and maximum quantities of B cell epitopes (Figure 10). B-cell epitopes have a minimum value of 0.187 and a maximum value of 0.671, although the server's results yielded an average score of 0.468..

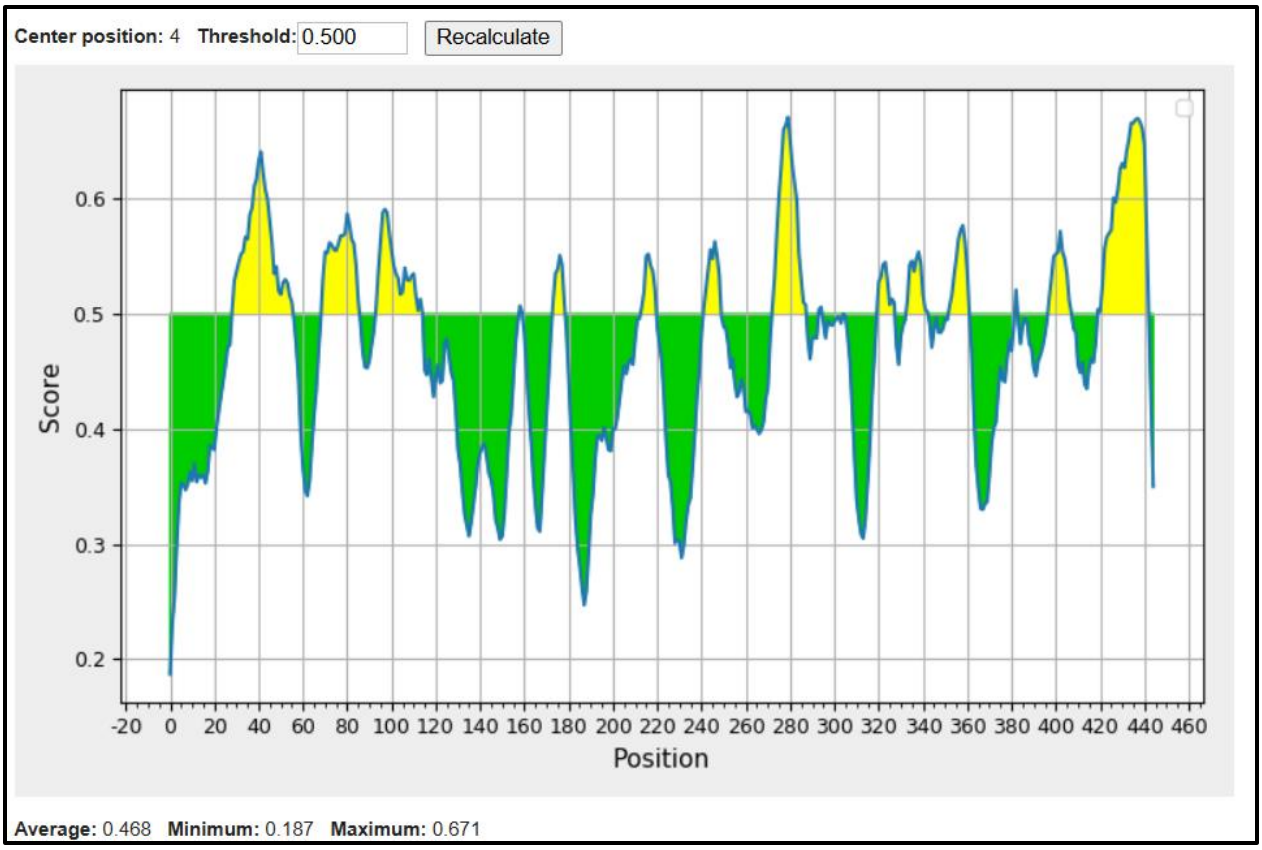


Figure 10: Score Vs position graph of B cell (Jespersen et al., 2017)

3.7 Development of final vaccine

The final vaccine formulation is composed of eligible epitopes categorized using various characteristics. The epitopes were combined with an adjuvant, utilizing the EAAAK, GPGPG, and KK linkers. Two CTL, two HTL, and lastly three B cell epitopes were included in the sequence, utilizing an

EAAAK linker at the N-terminal position to enhance immunogenicity.

The sequence of the constructed vaccine:
MREIVHIQAGQCGNQIGAKFWEVISDEHGID
PTGSYHGDSDLQLERINVYYNEAAGNKYVPR
AILVDLEPGTMDSVRSGPFGQIFRPDNFVFG
QSGAGNNWAKGHYTEGAELVDSVLDVVRKES
ESCDCLQGFQLTHSLGGGTGSGMGTLLISKI
REEYPDRIMNTFSVMPSPKVSDTVVEPYNAT
LSVHQLVENTDETYSIDNEALYDICFRTLKL
TTPTYGDLNHLVSATMSGVTTCLRFPGQLNA

DLRKLAVNMVPFPRLHFFMPGFAPLTSRGSQ
QYRALTVPELTQQMFDSKNMMAACDPRHGRY
LTVAAIFRGRMSMKEVDEQMLNVQNKNSSYF
VEWIPNNVKTAVCDIPPRGLKMSATFIGNST
AIQELFKRISEQFTAMFRRKAFLHWYTGEGM
DEMEFTEAESNMNDLVSEYQQYQDATADEQG
EFEEEEGEDEA**EAAAK**VSDTVVEPY**EAAK**LTS
RGSQQY**EAAAK**NVQNKNSSY**GPGPG**MPGFA
PLTSRGSQQY**GPGPG**NADLRKLAVNMVPFP**K**
KTLKLTTP**KK**VCDIPPRGL**KK**YTGEGMDEME
F

3.8 Antigenicity of Constructed Vaccine

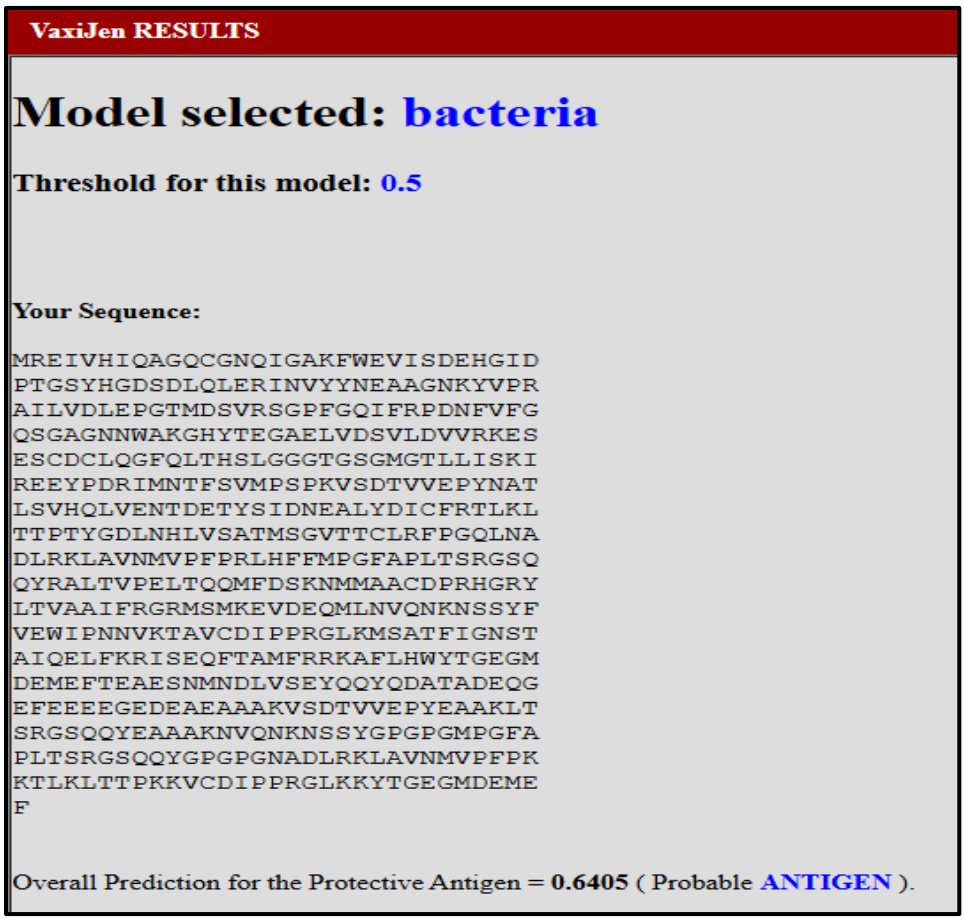


Figure 11: Antigenicity of Designed Vaccine (Doytchinova & Flower, 2007)

3.9 Reviewing Designed vaccine's Biochemical functionality

For the purpose of evaluating vaccination, biochemical features were analyzed using Protparam software. Results of features such as GRAVY, instability index, molecular weight, and molecular formula are provided by the Protparam server.

The therapeutic pI, molecular weight, and amino acid count are shown below (Figure 12). The results show that the person is eligible for the vaccination.

Number of amino acids: 559

Molecular weight: 62078.84
Theoretical pI: 8.06

Amino acid composition: CSV format

Ala (A)	39	7.0%
Arg (R)	26	4.7%
Asn (N)	28	5.0%
Asp (D)	30	5.4%
Cys (C)	8	1.4%
Gln (Q)	27	4.8%
Glu (E)	43	7.7%
Gly (G)	47	8.4%
His (H)	10	1.8%
Ile (I)	20	3.6%
Leu (L)	39	7.0%
Lys (K)	27	4.8%
Met (M)	23	4.1%
Phe (F)	26	4.7%
Pro (P)	32	5.7%
Ser (S)	37	6.6%
Thr (T)	36	6.4%
Trp (W)	4	0.7%
Tyr (Y)	21	3.8%
Val (V)	36	6.4%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 73
Total number of positively charged residues (Arg + Lys): 53

Figure 12: amino acid composition, negative and positive charge residue details by Protparam sever (Gasteiger et al., n.d.)

Atomic composition:

Carbon	C	2726
Hydrogen	H	4223
Nitrogen	N	743
Oxygen	O	855
Sulfur	S	31

Formula: $C_{2726}H_{4223}N_{743}O_{855}S_{31}$
Total number of atoms: 8578

Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 53790
Abs 0.1% (=1 g/l) 0.866, assuming all pairs of Cys residues form cystines

Ext. coefficient 53290
Abs 0.1% (=1 g/l) 0.858, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 36.13
This classifies the protein as stable.

Aliphatic index: 66.82
Grand average of hydropathicity (GRAVY):-0.469

Figure 13: Vaccine's feature evaluation through estimated half-life, instability index, aliphatic index, GRAVY (Gasteiger et al., n.d.)

3.10 Validating the New Vaccine's Antigenicity, Allergenicity and Toxicity

The produced vaccine demonstrated no allergenicity on the AllergenOnline platform, as assessed using an 80-mer sliding window FASTA analysis (Figure 14). The findings demonstrate that the server evaluates 80 possible amino acids from protein sequences in comparison to the server database to attain a minimum recognition rate of 35%.

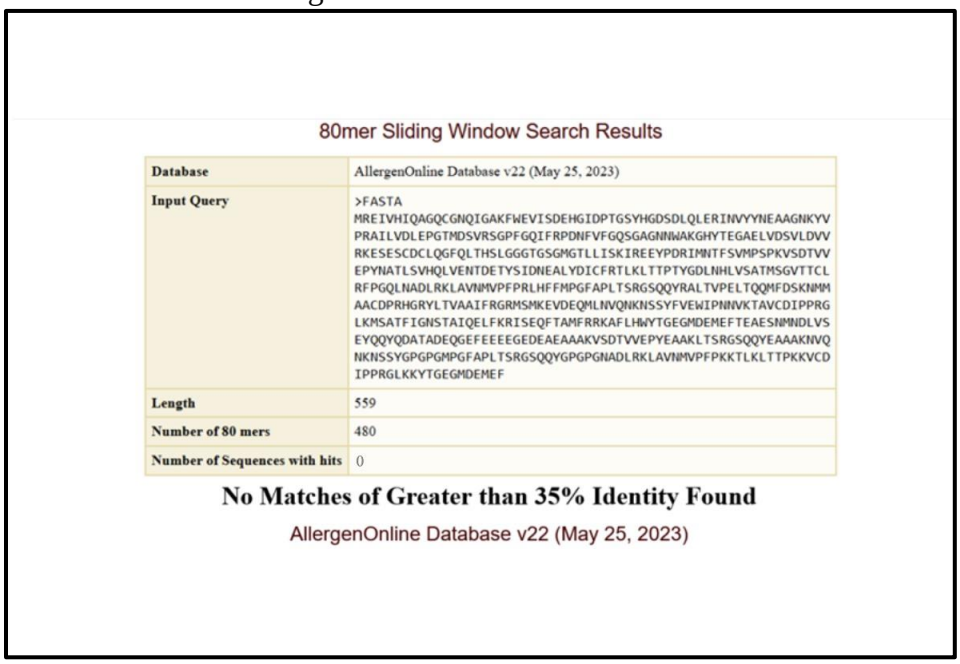


Figure 14: Allergenicity identification on AllergenOnline database of Designed Vaccine (Goodman et al., 2016)

Additionally, toxicity was assessed using the T3DB server to determine the presence of any harmful metabolites or particles, and the results indicated no toxicity (figure 15).

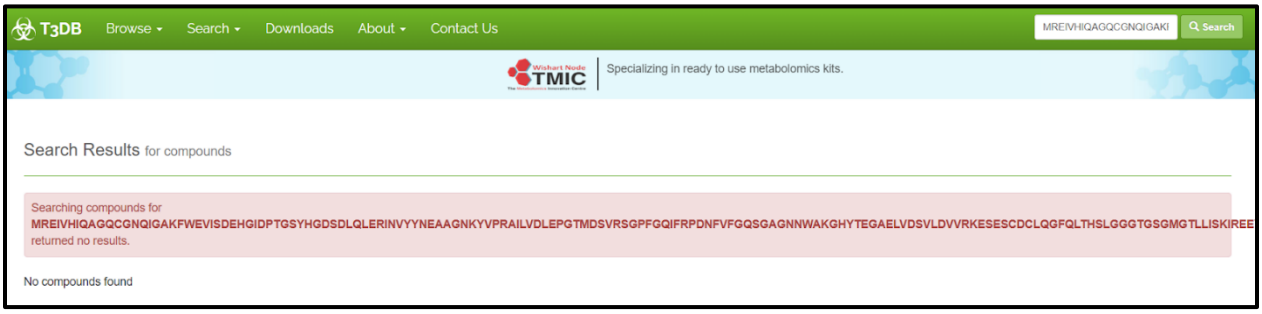


Figure 15: Toxicity identification on T3DB server of designed vaccine (Wishart et al., 2015)

3.11 Construction of a Vaccine Using 3D Models

The Phyre 2 server was used to model the vaccination's homology (figure 16). With 77% coverage and 100% confidence, the server produced a three-dimensional model of the vaccination. In this instance, 433 residues (77% of the sequence) were simulated with complete confidence. A high level of confidence and coverage in % is essential for achieving precise outcomes. . The 3D model of the vaccine enabled the execution of the subsequent biochemical investigation and improved the visual assessment of the vaccine's structure (Kelley et al., 2015).

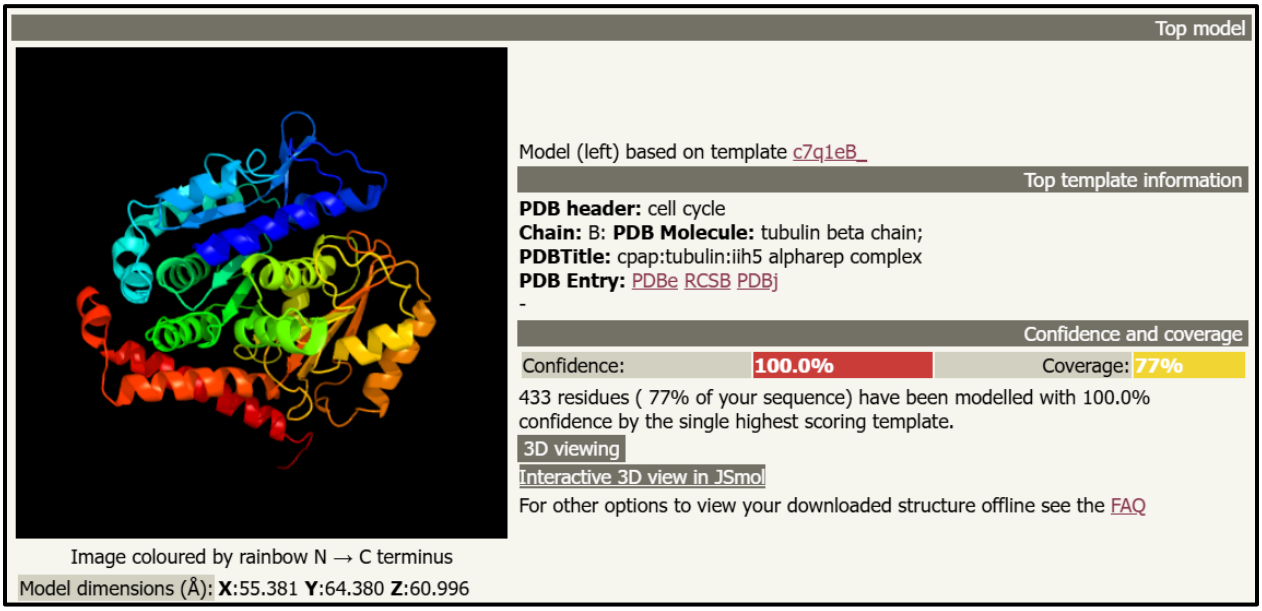


Figure 16: Confidence and coverage score and the constructed 3D model creation on Phyr2 server (Kelley et al., 2015)

3.12 Validation Analysis of Three-Dimensional Models

The Swiss model Expasy utilizes the PDB form file acquired from the Phry2 service. Utilize the Expasy server (Waterhouse et al., 2018) for validation and assess the Ramachandran plot (Figure 17). 95.13 percent. Ramachandran was displayed as the region. The preferred region is where the 34 amino acid levels described are located, and 0.23% of the regions were Ramachandran

outliers (Figure 18). enhanced the visual examination of the vaccine's structure (Kelley et al., 2015).

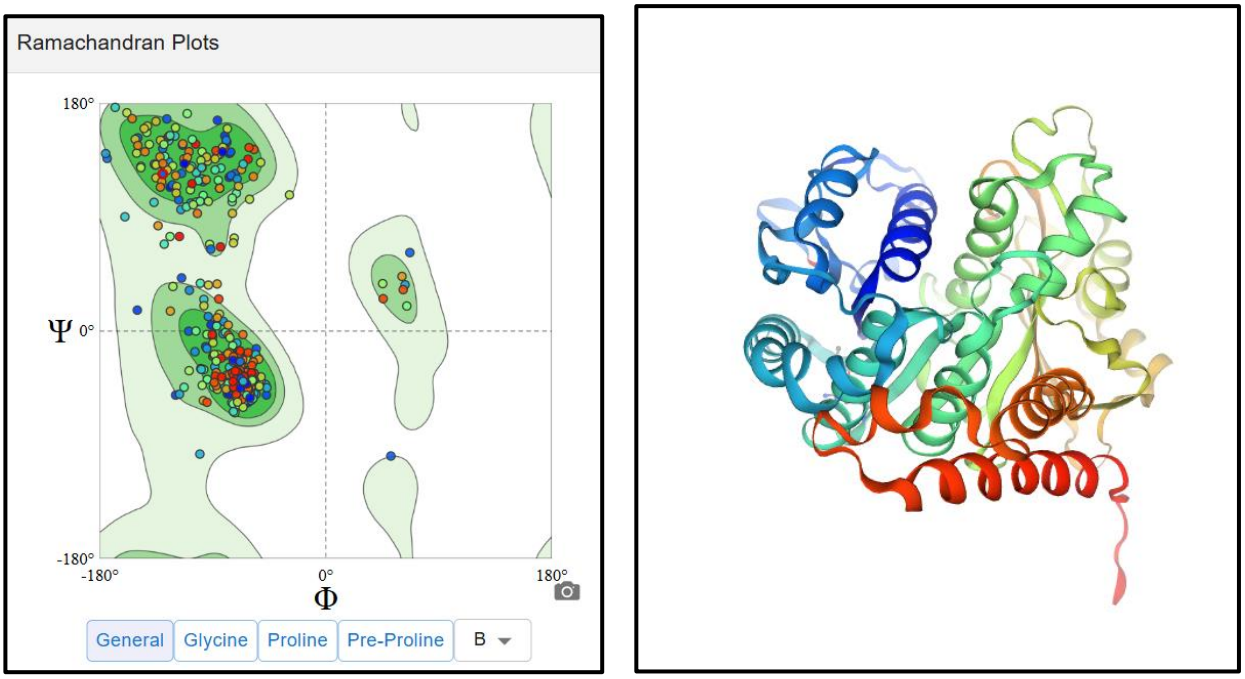


Figure 17: Ramachandran Plot on SWISS PDB plotter (Waterhouse et al., 2018)

MolProbity Results

MolProbity Score	2.37
Clash Score	35.08

(B190 HIS-B414 ASN), (B6 HIS-B50 TYR), (B342 VAL-B344 TRP), (B262 ARG-B421 GLU), (B263 LEU-B422 TYR), (B130 LEU-B162 ARG), (B418 LEU-B422 TYR), (B68 LEU-B112 LEU), (B183 TYR-B388 MET), (B313 VAL-B367 PHE), (B121 ARG-B125 GLU), (B16 ILE-B233 MET), (B279 GLN-B281 TYR), (B305 PRO-B310 TYR), (B257 MET-B266 PHE), (B101 TRP-B184 ASN), (B65 LEU-B90 PHE), (B241 ARG-B242 PHE), (B2 ARG-B46 ARG), (B394 PHE-B397 TRP), (B196 THR-B264 HIS), (B46 ARG-B239 CYS), (B46 ARG-B240 LEU), (B46 ARG-B248 ALA), (B31 ASP-B37 HIS), (B101 TRP-B146 GLY), (B47 ILE-B51 TYR), (B347 ASN-B350 LYS), (B101 TRP-B403 MET), (B291 GLN-B295 ASP)

Ramachandran Favoured	95.13%
Ramachandran Outliers	0.23%

B276 ARG

Rotamer Outliers	0.00%
C-Beta Deviations	0

Figure 18: Ramachandran plot's MolProbity results on SWISS PDB (Waterhouse et al., 2018)

3.13 Z-Score Three-Dimensional Model Quality Analysis

The ProSA-web server employed a knowledge-based energy against sequence location graph (Figure 20) and a z-score vs residue count graph (Figure 19) to depict model quality information. The range of native conformations is represented by a z-score of -9.53. Various colors differentiate structures from different origins (X-ray, NMR). The second graph displays energy distribution, which links energy to amino acid sequence positions, thereby indicating the quality of the local model. Most residues show negative values, except for the slightly positive peaks found near the N-terminal region (Wiederstein & Sippl, 2007).

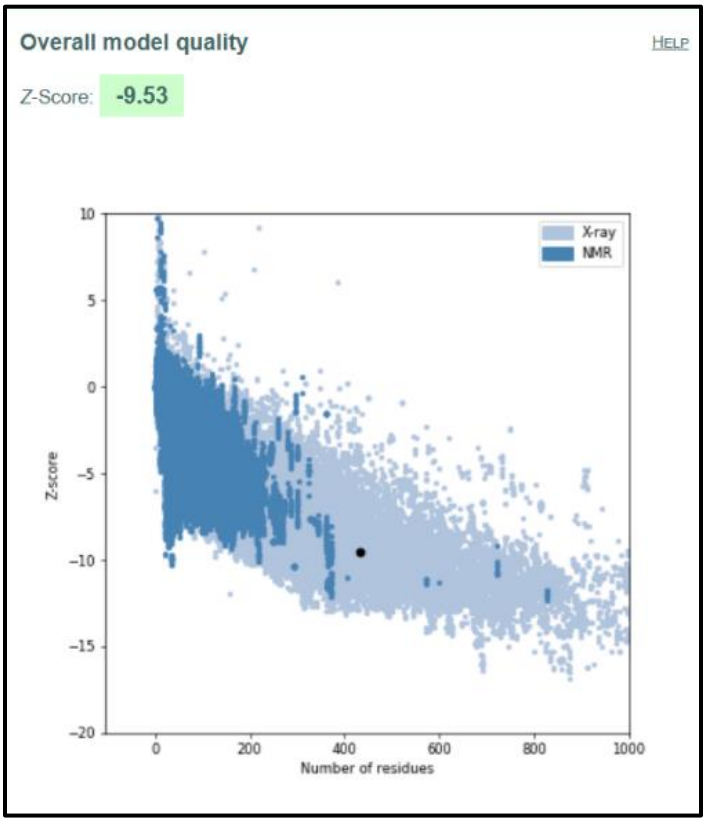


Figure 19: Z score Vs Number of residue graph on ProSA-web server (Wiederstein & Sippl, 2007)

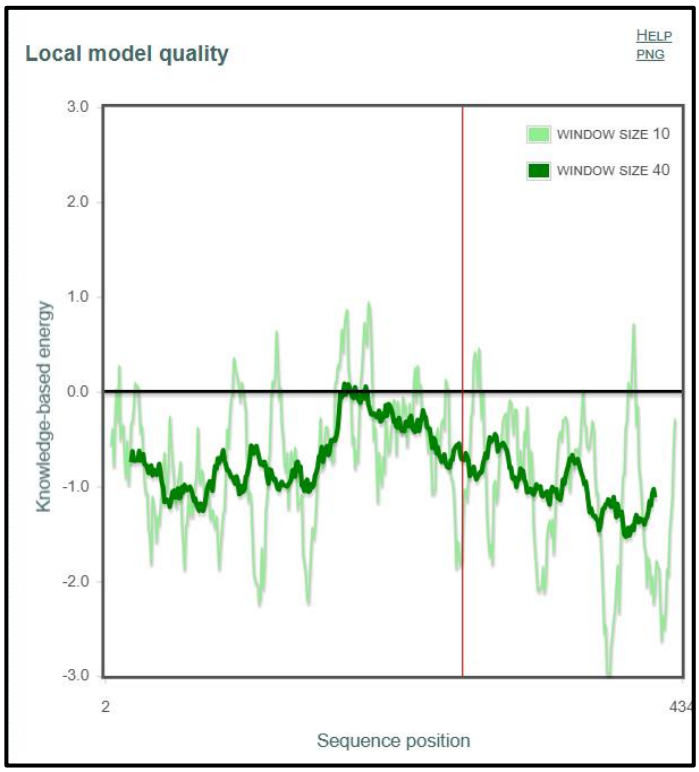


Figure 20: knowledge-based energy Vs sequence position graph on ProSA-web server (Wiederstein & Sippl, 2007)

3.14 Molecular Docking Assessment

The ClusPro service evaluated the binding affinity between Human Toll-like receptor (TLR3) and the designed vaccination to form a ligand-receptor complex. The server produced data indicating 10 clusters among a total of 30 clustered structures. The cluster with the lowest negative score is regarded to be the most significant. The highest-ranking cluster demonstrating the interaction between TLR3 and the engineered vaccination has a ClusPro score of -852.3, the lowest recorded score. The ClusPro score signifies the aggregate of all energies, encompassing van der Waals and electrostatic energies, among others.

The data for the high-ranked model is presented in table 7.

Table 7: Docking complex score of ClusPro server (Kozakov et al., n.d.)

Cluster	Members	Representative	Weighted Score
0	62	Center	-852.3
		Lowest Energy	-852.3
1	44	Center	-903.7
		Lowest Energy	-903.7
2	35	Center	-842.7
		Lowest Energy	-842.7
3	32	Center	-846.6
		Lowest Energy	-846.6
4	28	Center	-752.0
		Lowest Energy	-831.5
5	19	Center	-834.7
		Lowest Energy	-834.7
6	19	Center	-756.3
		Lowest Energy	-846.4
7	17	Center	-820.0
		Lowest Energy	-820.0

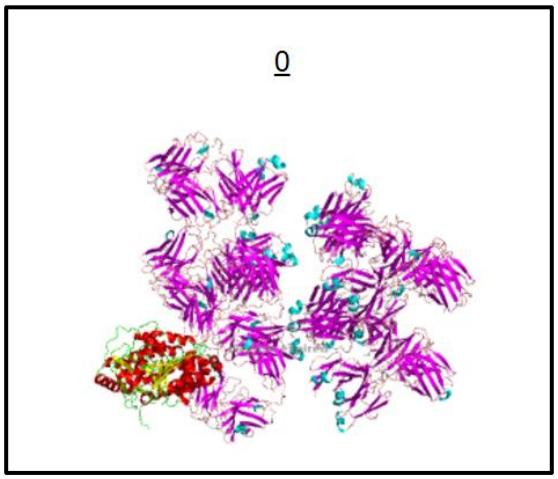


Figure 21: The interactivity of the designed vaccine and TLR3 (Kozakov et al., n.d.)

3.15 Immunogenicity Reflection

Extended doses, such as booster doses are typically required to keep an adequate vaccination level in the body to elicit the desired extended immunological reaction. The immunological response elicited in the body must be evaluated by monitoring the levels of certain antibodies or immunoglobulins in relation to the timing and vaccination schedule for dosage.

Additionally, it is necessary to assess the rise and fall of different immune cells in response to the vaccination doses. The C-IMMSIM server predicted and visually depicted these characteristics in response to the vaccine, as illustrated in Figure 22 (a-j) (Rapin et al., 2010).

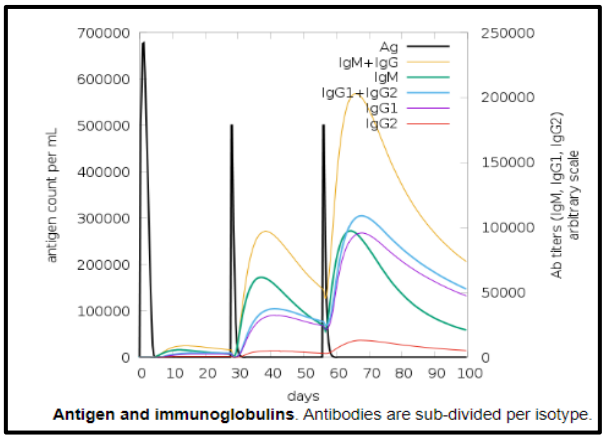


Figure 22: The C-IMMSIM server predicted and visually depicted characteristics in response to the vaccine

(a) Antigen count per mL and antibody titers (Rapin et al., 2010)

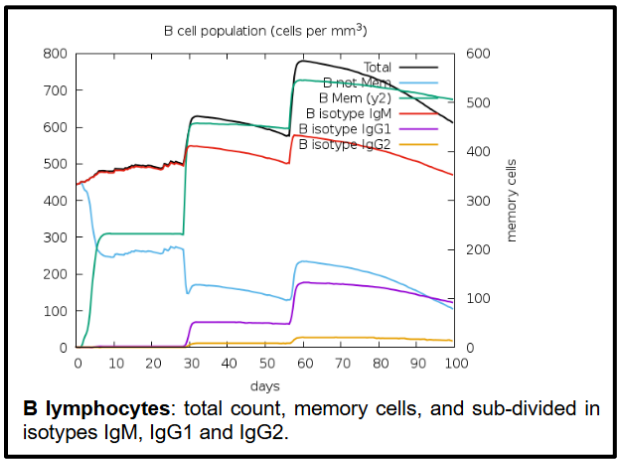


Figure 22:(b) Total count of B lymphocytes and memory cells (Rapin et al., 2010)

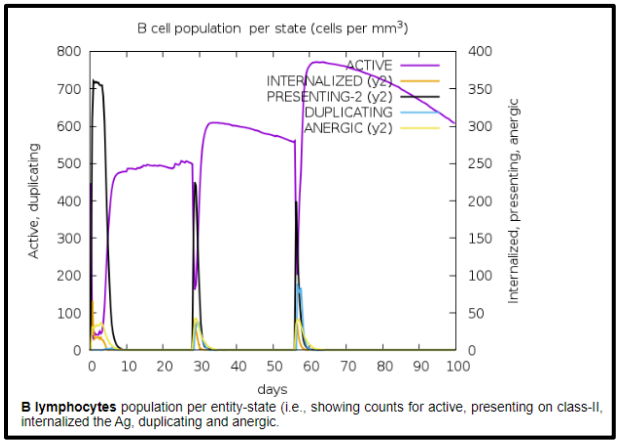


Figure 22:(c) Entity-state of B cell population (Rapin et al., 2010)

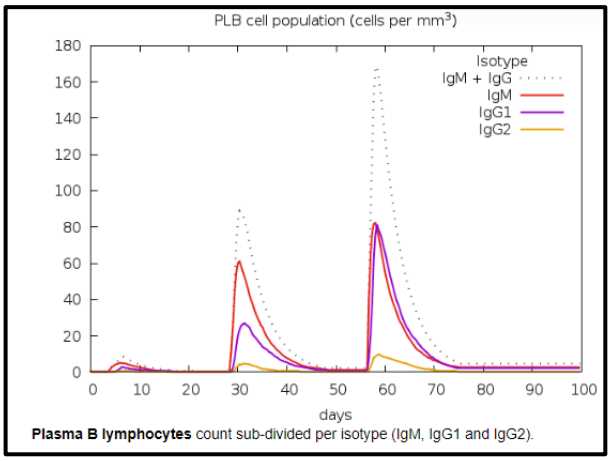


Figure 22:(d) Number of plasma B cells according to their isotypes (Rapin et al., 2010)

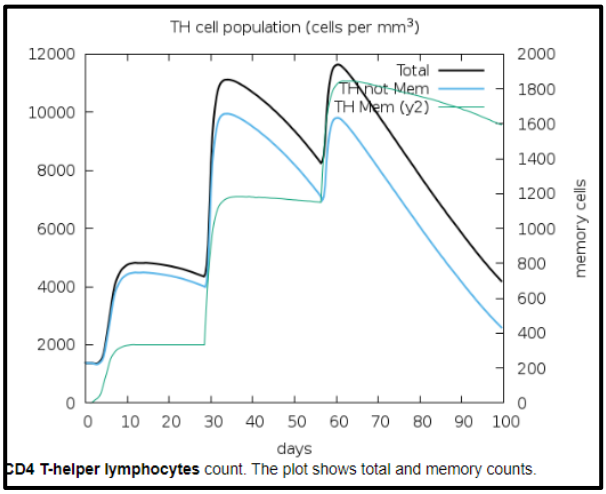


Figure 22:(e) Number of CD4 T helper lymphocytes (Rapin et al., 2010)

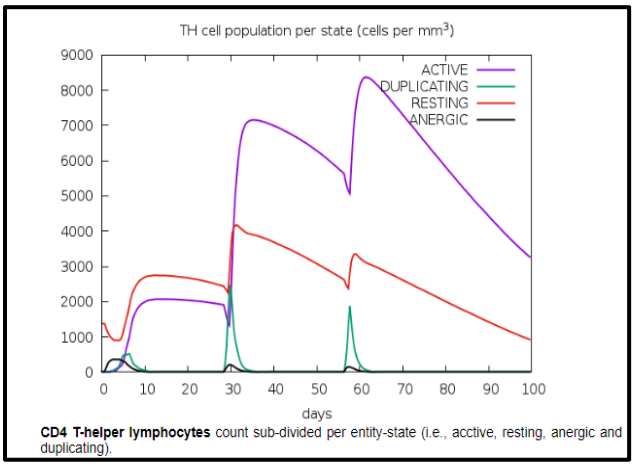


Figure 22:(f) Entity state of CD4 Helper T cells (Rapin et al., 2010)

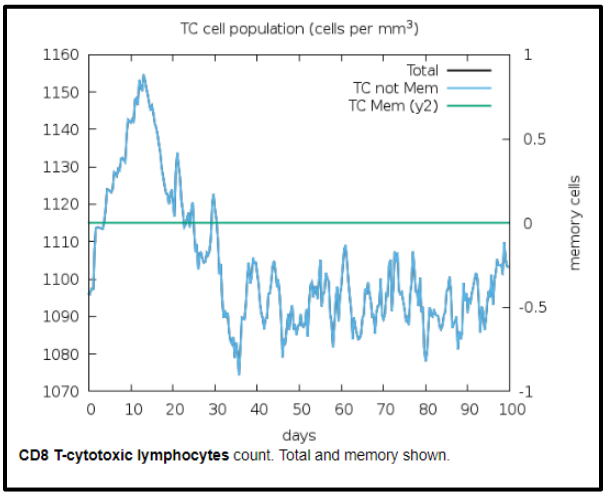


Figure 22:(g) Number of CD8 T cytotoxic (TC) cells (Rapin et al., 2010)

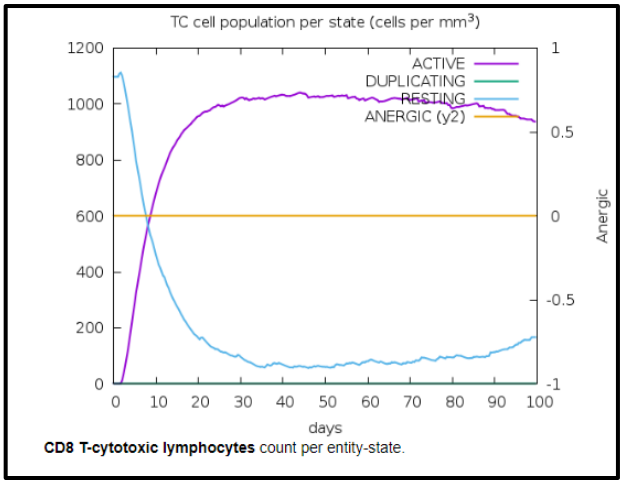


Figure 22:(h) Number of CD8 T cytotoxic (TC) cells per entity (Rapin et al., 2010)

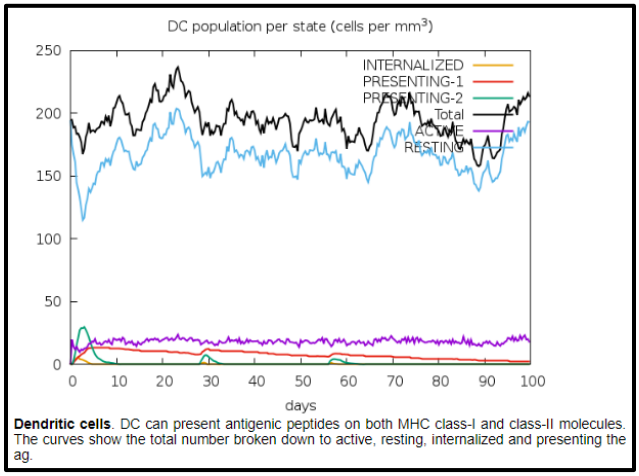


Figure 22:(i) Total number of dendritic cells and their states (Rapin et al., 2010)

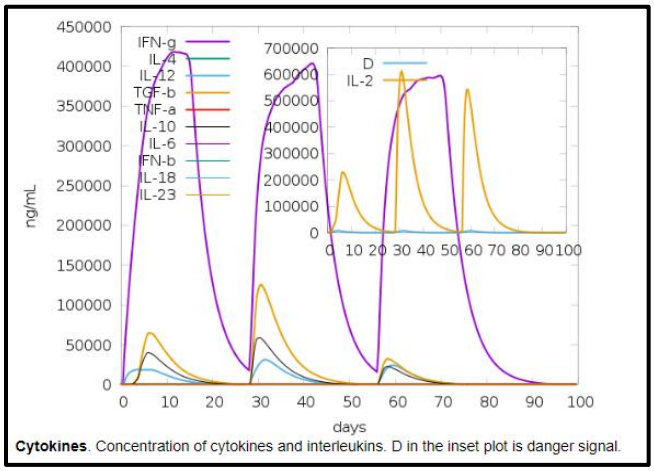


Figure 22:(j) Levels of cytokines including IFN-γ, IL-4, IL-10 (Rapin et al., 2010)

B-lymphocyte (or B cell): immune cells that produce memory cells and antibodies in your body. As shown in Figure (a) after the first immunization, antigen concentration/mL increases. In spite of showing a small increase over the first 28 days, antibodies IgM and IgG showed an impressive rise following approximately two months as vaccine doses had been repeated and completed. The vaccine had induced the expected favorable immunological response and reflected in the production of antibodies B-cell-derived memory cells help the body identify the same infection in the future and speed up the immune response. Thus, effective immunization depends on the

production of competent memory cells. Figure (b) illustrates the total quantity of memory cells and B lymphocytes post-vaccination, categorized by different isotypes. Figure (c) illustrates the variation in representation of B lymphocytes by individual state. It indicated the number of active B cells, the quantity presented on Class II, the count of B cells that received the antigen, and the number of B cells undergoing energy and duplication. Figure (d) illustrates the number of plasma B lymphocytes (PLB) and their respective isotypes, which comprise IgG1, IgG2, IgM, and the combination of IgM + IgG. The number of CD4 T helper cells after each vaccination dose was predicted in Figure (e). The number of Helper T cells that were thought to be active, replicating, at rest, or in an anergic state was replicated by C-IMMSIM in Figure (f).

Total and memory subsets of CD8 T cytotoxic (TC) cells at each of 4 states were modeled (i.e. active, proliferative, resting, anergic), as shown in figures (g) and (h). Figure (i) exhibits active, resting, antigen-presenting and phagocytic dendritic cells. PRESENTING-1 sheds light on dendritic cells presenting antigens in MHC class I, and PRESENTING-2 indicates dendritic cells presenting antigens in MHC class II. The graph in figure (j) has an outline that displays the interleukin-2 levels and danger signal.

Chapter 4

Discussion

This report underscores a significant advancement in formulating an effective vaccine strategy for colorectal cancer. The choice of tubulin beta-2A chain protein as a target was justified by its established antigenicity, rendering it appropriate for immunotherapeutic intervention. Computational techniques were employed for a methodical vaccine design, concentrating on the identification of epitopes with high binding affinities to MHC molecules. Rigorous screening criteria, encompassing toxicity, allergenicity, and antigenicity evaluations, guaranteed the incorporation of safe and effective epitopes in the final vaccine formulation. This study proposes a vaccine using an in-silico approach. An appropriate protein, tubulin beta-2A chain protein, was chosen due to its antigenic properties, which are advantageous for vaccine development (Zarei et al., 2017).

The protein underwent multiple phases using online prediction techniques to identify its CTL, HTL, and B-cell epitopes. Following an investigation of the NetCTL 1.2 server, six CTL epitopes were found, and the alleles were examined using NetMHCpan 4.1, resulting in the selection of the top five CTL epitopes based on their high binding affinity. The epitopes were subsequently assessed based on several parameters (toxicity, antigenicity, allergenicity), and on this occasion, one epitope satisfied the qualifying requirements.

A search utilizing the NetMHCIIpan 4.0 server nearly produced 96 HTL, with cytokine-inducing capabilities employed for categorizing numerous internet servers. Twenty-seven tested positives for IFNγ, thirty-seven were inducers of IL-4, and thirteen for IL-10 Epitopes were once again

classified according to antigenicity, allergenicity, and toxicity, resulting in only two remaining. Utilizing the Bedipred Linear Epitope Prediction 2.0 internet service, eleven B cells were obtained, and subsequent sorting resulted in the selection of three. Sorted candidates utilized in the construction of multi-epitope vaccines to combat H. pylori, incorporating main proteins and linkers. The antigenicity of the vaccine has markedly improved relative to the primary protein. No allergenic components or hazardous entities were identified in the conducted trial. Following vaccine development, a crucial aspect is vaccine stability, which was assessed using the ProtParam server, demonstrating stability for the final vaccine. The chosen molecular weight and negative GRAVY score indicated vaccination suitability. Phry2 homology modeling precisely and comprehensively depicted the three-dimensional structure with full coverage and 77% confidence. Favorable outcomes were noted in the Ramachandran plot using Swiss-Model ExPASy, revealing a favored region of 95.13%, however Z-score analysis performed with ProSA-web produced a score of -9.53 for the vaccination. The binding of the synthesized vaccine to toll-like receptor 5 was shown by a ClusPro server score of -852.3. Nevertheless, the C-immsim server exhibited the expected response of antibodies, characterized by elevated levels of IgG, IgM, and CD4 and CD8 cells following the administration of the dosage.

Chapter 5

Conclusion

To conclude, the research focused on recommending an effective immunization against colorectal cancer since there is no available vaccine on the market till now. The advancement of vaccinations is propelled by antibiotic resistance and several bacterial-related disorders.

The tubulin beta-2A chain protein was targeted to generate a multiepitope peptide vaccine. The identification of essential epitopes and the synthesis of the vaccine were conducted using proven computational approaches, which demonstrated no toxicity or allergenicity. It demonstrated antigenicity at an adequate level, together with cytokine-inducing capabilities and the potential for antibody production. The proposed vaccine has a robust binding affinity for TLR 3 and induces a positive immunological response. Conducting in vitro and in vivo investigations is essential to validate the quality, efficacy, and safety of the developed vaccine.

5.1 Limitations of the study

Despite the in-silico creation of a multi-epitope vaccination targeting colorectal cancer producing encouraging outcomes, this research has notable limitations. The primary concern is that the study relied solely on computer simulations and predictions. In vitro and in vivo investigations are essential to validate the vaccine's antigenicity and confirm its safety and efficacy. Immunogenicity, and binding affinities. Furthermore, it's possible that the immunological simulation predictions don't accurately capture intricate biological reactions in living things. Unresolved issues that need more research include the capacity of vaccine production, possible adverse effects, and genetic diversity-induced variability in patient responses.

5.2 Future Prospect

To assess the vaccine's practical uses, future studies could concentrate on moving this in-silico design closer to preclinical and clinical trials. Targeted administration and vaccination stability may be improved by integration with delivery systems and nanotechnology. Treatment results could be enhanced by extending the strategy to incorporate more proteins unique to colorectal cancer or investigating combinatorial vaccinations. Furthermore, using machine learning and artificial intelligence could improve the accuracy of epitope predictions and expedite the vaccine development process. If these vaccines are developed and validated successfully, they could open the door to customized immunotherapies, which would transform the prevention and treatment of colorectal cancer.

References

Alexander, J., Fikes, J., Hoffman, S., Franke, E., Sacci, J., Appella, E., Chisari, F. V., Guidotti, L. G., Chesnut, R. W., Livingston, B., & Sette, A. (1998b). The optimization of helper T lymphocyte (HTL) function in vaccine development. *Immunologic Research*, *18*(2), 79–92. https://doi.org/10.1007/bf02788751

Bao, Y., Liu, X., Han, C., Xu, S., Xie, B., Zhang, Q., Gu, Y., Hou, J., Qian, L., Qian, C., Han, H., & Cao, X. (2013). Identification of IFN-γ-producing innate B cells. *Cell Research*, *24*(2), 161–176. https://doi.org/10.1038/cr.2013.155

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., & Insana, G. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1). https://doi.org/10.1093/nar/gkac1052

Cirillo, L., Gotta, M., & Meraldi, P. (2017). The Elephant in the room: The role of microtubules in cancer. *Advances in Experimental Medicine and Biology*, 93–124. https://doi.org/10.1007/978-3-319-57127-0_5

Curry, S. J., Byers, T., & Hewitt, M. (2003). *Potential of screening to reduce the burden of cancer*. Fulfilling the Potential of Cancer Prevention and Early Detection - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK223933/

De Groot, A. S., Moise, L., Terry, F., Gutierrez, A. H., Hindocha, P., Richard, G., Hoft, D. F., Ross, T. M., Noe, A. R., Takahashi, Y., Kotraiah, V., Silk, S. E., Nielsen, C. M., Minassian, A. M.,

Ashfield, R., Ardito, M., Draper, S. J., & Martin, W. D. (2020). Better Epitope Discovery, Precision Immune Engineering, and Accelerated Vaccine Design Using Immunoinformatics Tools. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.00442

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013). Prediction of IL4 inducing peptides. *Clinical and Developmental Immunology*, *2013*, 1–9. https://doi.org/10.1155/2013/263952

Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics*, *14*(S6). https://doi.org/10.1186/1471-2105-14-s6-s4

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1). https://doi.org/10.1186/1471-2105-8-4

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., Van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & Taylor, S. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition & Food Research*, *60*(5), 1183–1198. https://doi.org/10.1002/mnfr.201500769

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico approach for predicting toxicity of peptides and proteins. *PLoS ONE*, *8*(9), e73957. https://doi.org/10.1371/journal.pone.0073957

Holme, Ø., Bretthauer, M., Fretheim, A., Odgaard-Jensen, J., & Hoff, G. (2013). Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. *Cochrane Library*, *2014*(3). https://doi.org/10.1002/14651858.cd009259.pub2

Ilic, I., & Ilic, M. (2022). International patterns in incidence and mortality trends of pancreatic cancer in the last three decades: A join point regression analysis. *World Journal of Gastroenterology*, *28*(32), 4698–4715. https://doi.org/10.3748/wjg.v28.i32.4698

Ito, H., & Seishima, M. (2010). Regulation of the induction and function of cytotoxic T lymphocytes by natural killer T cell. *Journal of Biomedicine and Biotechnology*, *2010*, 1–8. https://doi.org/10.1155/2010/641757

Iyer, S. S., & Cheng, G. (2012). Role of Interleukin 10 Transcriptional regulation in inflammation and autoimmune Disease. *Critical Reviews in Immunology*, *32*(1), 23–63. https://doi.org/10.1615/critrevimmunol.v32.i1.30

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29. https://doi.org/10.1093/nar/gkx346

Jin, Y., Zhang, S., & Liu, L. (2020). Circular RNA circ_C16orf62 Suppresses Cell Growth in Gastric Cancer by miR-421/Tubulin beta-2A Chain (TUBB2A) Axis. *Medical Science Monitor*, *26*. https://doi.org/10.12659/msm.924343

Kalkal, M., Kalkal, A., Dhanda, S. K., Das, E., Pande, V., & Das, J. (2022). A comprehensive study of epitopes and immune reactivity among Plasmodium species. *BMC Microbiology*, *22*(1). https://doi.org/10.1186/s12866-022-02480-7

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Krzyszczyk, P., Acevedo, A., Davidoff, E. J., Timmins, L. M., Marrero-Berrios, I., Patel, M., White, C., Lowe, C., Sherba, J. J., Hartmanshenn, C., O'Neill, K. M., Balter, M. L., Fritz, Z. R., Androulakis, I. P., Schloss, R. S., & Yarmush, M. L. (2018). The growing role of precision and personalized medicine for cancer treatment. *Deleted Journal*, *06*(03n04), 79–100. https://doi.org/10.1142/s2339547818300020

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O., & Nielsen, M. (2005). An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology*, *35*(8), 2295–2303. https://doi.org/10.1002/eji.200425811

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, *8*(1). https://doi.org/10.1186/1471-2105-8-424

Lewandowska, A., Rudzki, G., Lewandowski, T., Stryjkowska-Góra, A., & Rudzki, S. (2022). Risk factors for the diagnosis of colorectal cancer. *Cancer Control*, *29*, 107327482110566. https://doi.org/10.1177/10732748211056692

Luo, C., Yao, D., Lim, T. K., Lin, Q., & Liu, Y. (2018). Identification of the altered proteins related to colon carcinogenesis by ITRAQ-based quantitative proteomic analysis. *Current Proteomics*, *16*(4), 297–306. https://doi.org/10.2174/1570164616666181129111542

McCulloch, S. M., Aziz, I., Polster, A. V., Pischel, A., Stålsmeden, H., Shafazand, M., Block, M., Byröd, G., Lindkvist, B., Törnblom, H., Jonefjäll, B., & Simren, M. (2019). The diagnostic value of a change in bowel habit for colorectal cancer within different age groups. *United European Gastroenterology Journal*, *8*(2), 211–219. https://doi.org/10.1177/2050640619888040

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep42851

Protein identification and analysis tools in the Expasy server. (n.d.). https://web.expasy.org/compute_pi/pi_tool-doc.html

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE*, *5*(4), e9862. https://doi.org/10.1371/journal.pone.0009862

Saleem, S., Tariq, S., Aleem, I., Shaheed, N. S., Tahseen, M., Atiq, A., Hassan, S., Bakar, M. A., Khattak, S., Syed, A. A., Ahmad, A. H., Hussain, M., Yusuf, M. A., & Sutton, C. (2019). Proteomics analysis of colon cancer progression. *Clinical Proteomics*, *16*(1). https://doi.org/10.1186/s12014-019-9264-y

Siegel, R. L., Miller, K. D., Sauer, A. G., Fedewa, S. A., Butterly, L. F., Anderson, J. C., Cercek, A., Smith, R. A., & Jemal, A. (2020). Colorectal cancer statistics, 2020. *CA a Cancer Journal for Clinicians*, *70*(3), 145–164. https://doi.org/10.3322/caac.21601

SjöBlom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K. V., Gazdar, A. F., . . . Velculescu, V. E. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, *314*(5797), 268–274. https://doi.org/10.1126/science.1133427

Sircar, G., Sarkar, D., Bhattacharya, S. G., & Saha, S. (2014). Allergen databases. *Methods in Molecular Biology*, 165–181. https://doi.org/10.1007/978-1-4939-1115-8_9

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://academic.oup.com/nar/article/46/W1/W296/5000024

What is colorectal cancer? | How does colorectal cancer start? (n.d.). American Cancer Society. https://www.cancer.org/cancer/types/colon-rectal-cancer/about/what-is-colorectal-cancer.html

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2014). T3DB: the toxic exposome database. *Nucleic Acids Research*, *43*(D1), D928–D934. https://doi.org/10.1093/nar/gku1004

World Health Organization: WHO & World Health Organization: WHO. (2023, July 11). *Colorectal cancer*. https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer

Xi, Y., & Xu, P. (2021). Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology*, *14*(10), 101174. https://doi.org/10.1016/j.tranon.2021.101174

Assignment

ORIGINALITY REPORT

11%	8%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.bracu.ac.bd Internet Source	7%
2	www.frontiersin.org Internet Source	1%
3	Submitted to Universiti Brunei Darussalam Student Paper	1%
4	Yasir Arfat, Imran Zafar, Sheikh Arslan Sehgal, Mazhar Ayaz et al. "In silico designing of multiepitope-based-peptide (MBP) vaccine against MAPK protein express for Alzheimer's disease in Zebrafish", Heliyon, 2023 Publication	<1%
5	Leah Maharaj, Victoria T. Adeleke, Abiodun J. Fatoba, Adebayo A. Adeniyi et al. "Immunoinformatics approach for multi-epitope vaccine design against P. falciparum malaria", Infection, Genetics and Evolution, 2021 Publication	<1%
6	Submitted to Macquarie University Student Paper	<1%

7	Submitted to NHS National School of Healthcare Science Student Paper	<1 %
8	Sarmad Frogh Arshad, Rehana Rehana, Muhammad Asif Saleem, Muhammad Usman et al. "Multi-epitopes vaccine design for surface glycoprotein against SARS-CoV-2 using immunoinformatic approach", Heliyon, 2024 Publication	<1 %
9	kfanhub.com Internet Source	<1 %
10	link.springer.com Internet Source	<1 %
11	Yiwei Zhang, Yujun Zhang, Jingjing Song, Xifu Cheng, Chulin Zhou, Shuo Huang, Wentao Zhao, Zhen Zong, Lingling Yang. "Targeting the “tumor microenvironment”: RNA-binding proteins in the spotlight in colorectal cancer therapy", International Immunopharmacology, 2024 Publication	<1 %
12	assets.researchsquare.com Internet Source	<1 %
13	rcastoragev2.blob.core.windows.net Internet Source	<1 %

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.