

Syrian Arab Republic
Ministry of Higher Education
and Scientifics Research
Syrian Virtual University



الجمهورية العربية السورية
وزارة التعليم العالي والبحث العلمي
الجامعة الافتراضية السورية

In silico analysis of cancer antigens of non-small cell lung cancer (NSCLC) for dendritic cell-based immune-gene therapy application

A Project submitted for the Master's degree of Bioinformatics

Done by student: Hend Zuhair Allaw

Student ID: hend_154309

Supervisors

A.Prof. Majd Aljamali A.Prof. Lama Youssef

2023-1444

Summary

Background: Cancer immunotherapy is the development of efficient therapeutic cancer vaccines. Cancer vaccines are based on tumor antigens expressed in the context of Major Histocompatibility Complex (MHC) molecules able to elicit a strong tumor-specific CTL response, which may result in the killing of tumor cells and cancer regression. We describe here the strategy in the design of a polytope cancer vaccine that has many unique characteristics. Combining different HLA-restricted epitopes from CTAs into one polytope vaccine construct allows the fusion Ag to efficiently enter the ER, then be processed and presented to MHC class I to induce the related CTL responses against all epitopes simultaneously.

Aim of the study: In this project, we aim to design a new structural model containing putative antigenic epitopes. Since immune stimulation is considered one of the most important mechanisms in tumor treatment, tumor cells can escape from the immune system. It may be advantageous to use T cell epitopes of different tumor Antigens simultaneously, the goal of using multiple antigenic epitopes instead of a single antigen is to avoid the specific antigen being lost or mutated. The use of multiple antigenic epitopes in a single structural model would cover a wide range of histocompatibility complex polymorphisms.

Results: we designed a new chimeric construct of CTAs including HLA-restricted epitopes of MAGEA8, SAGE1, and CTA45A2, which contained essential determinants to be recognized by CTLs. Immunogenic epitopes for MAGEA8, SAGE1, and CTA45 A2 proteins were selected along with the hemagglutinin (HA) epitope as a reporter tag, the ER signal peptide, and the ER retention signal.

The Codon Adaptation Index (CAI) of the gene is 0.90. A CAI of 1.0 is considered ideal. Moreover, The GC content of the gene is 62.58%. The ideal percentage range of GC content is between 30% and 70%. The required restriction enzyme sites were added to the ends of the designated gene for future assays. Evaluation of model stability by Ramachandran plot showed that most residues of the chimeric model are in a stable zone. Protein secondary and tertiary structures were predicted by Ab Initio modelling and a software viewer was used for visualization. Epitope binding to MHC and recognition of such

complexes (epitope/MHC) by CTLs is a critical step in inducing a significant immune response. Therefore, a prediction of proteasomal cleavage sites using web-based software had been done. The result showed that the highest-scored cleavage positions are located at the fusion site of each epitope and its adjacent linkers.

Conclusion: We used *in silico* approaches to design our chimeric polytope construct of immune-gene therapy applications. We used several web servers and applications to predict different features of the construct, including GC content, secondary and tertiary structure of the protein, solvent accessibility of the chimeric protein, proteasomal cleavage site, validation of the epitope's prediction, MHC binding affinity, and post-translational modifications. Three epitopes with high immunogenicity scores were included in the study; MAGEA8, SAGE1, and CTA45A2. Both the MAGEA8 epitope and SAGE1 epitope gave a good binding prediction. However, only the SAGE1 epitope showed a strong binding affinity with MCH molecules. For future studies, the CTA45A2 epitope could be substituted with an epitope with a better binding prediction and affinity in order to develop a more effective structural model for cancer immune-gene therapy. Taking all these findings together, this study is promising in the field of the development of multiepitope chimeric vaccines for cancers being rationally designed using immunoinformatics and employing different computational approaches.

Table of Contents

Summary.....	1
Aim of the study:	11
Introduction:	11
Dataset:	11
Methods:	16
I. Sequence analysis.....	16
II. Design of the construct and gene optimization	17
III. In silico structural analysis of chimeric recombinant protein	17
IV. Prediction of the cleavage site.....	17
V. Validation of T-cell epitopes and MHC binding peptides affinity	17
VI. Prediction of post-translational modifications	18
Methods steps summary:	18

Results:	18
I. Design and structure of the chimeric construct	18
Epitope prediction for MAGEA8 protein:.....	19
Epitope prediction for SAGE1 protein:	19
Epitope prediction for CTA 45A2 protein:	19
Signal peptide prediction:	21
Chimeric gene Sequence:	22
Chimeric polytope peptide sequence:.....	22
II. In silico analysis of original chimeric gene	22
III. mRNA structure prediction	24
IV. Prediction of secondary and tertiary structures of chimeric protein	24
V. Evaluation of model stability.....	26
VI. Solvent accessibility prediction.....	27

VII. Prediction of cleavage sites.....	28
VIII. Validation of T-cell epitopes	29
X. MHC binding peptides affinity.....	30
XI. Prediction of post-translational Modifications.....	31
Discussion:	31
Conclusion:.....	34
References:	34

Table of Tables

Table 1. Clinical characteristic of 199 NSCLC cases included in the RNAseq analysis.....	13
Table 2. mRNA expression levels of the three CTAs chosen for further analysis.....	15
Table 3: Epitopes prediction scores via SYFPEITHI.....	20
Table 4: Epitopes immunogenicity scores via IEDB.....	20
Table 5. Prediction of cleavage sites on the construct protein using NetChop server. The threshold is 0.5.....	29
Table 6: Prediction of T-cell epitopes of the construct using different web-based servers.....	30
Table 7: Predictions of MHC-binding peptide affinity for the construct by NetMHC version 4.0. server using ANNs approximation.....	30

Table of Figures

Figure1. The expression of reported cancer testis antigens (CTAs) in non–small-cell lung cancer (NSCLC).....	14
Figure 2. MAGEA8 protein expression in lung cancer comparing to other cancers.....	15
Figure 3. SAGE1 protein expression in lung cancer comparing to other cancers	16
Figure 4. CTA 45 A2 protein expression in lung cancer comparing to other cancers	16
Figure 5. Flowchart of the project steps.	18
Figure 6: Schematic model of the construct. The selected epitopes of MAGEA8, SAGE1 and CTA45A2 are bound together by the linkers for expression in human.....	21
Figure 7. Signal peptide prediction significance results after constructing the polytope chimeric sequence using Phobius web server.....	22
Figure 10: The optimal structure with the lowest ΔG	24

Figure 11: secondary structure prediction Using Jpred 4 for secondary structure prediction the results showed two beta sheets, and one alpha helix comprised of amino acids 1825

Figure 12: Protein 3D structure Ab initio modelling using Robetta web server..., De novo models are built using the Rosetta de novo protocol. The procedure is fully automated25

Figure 13: Secondary structure predicted by AB Initio modelling and the result was viewed by Swiss pdb viewer.....26

Figure 14: Tertiary structure predicted by AB Initio modelling and the result was viewed by Pymol.26

Figure 15: Evaluation of model stability, the structure stability was confirmed based on the Ramachandran plot27

Figure 16: Solvent accessibility prediction. A) Fractional accessible surface area (ASA) analysis, B) Stereochemical/packing quality analysis and C) 3D profile quality analysis of the construct protein28

Table of Abbreviations

Ags	Antigens
ADC	Adenocarcinoma
ANN	Artificial Neural Networks
ASA	Accessible Surface Area
APC	Antigen Presenting Cell
CTAs	Cancer Testis Antigens
CTA 45 A2	Cancer/Testis Antigen Family 45 A2
CAI	Codon Adaption Index
CTL	Cytotoxic T Lymphocytes
DCs	Dendritic Cells
ER	Endoplasmic Reticulum
GC	Guanine, Cytosine
HA	Hemagglutinin Epitope
MAGEA8	Melanoma Associated Antigen 8
MHC	Major Histocompatibility Complex
NSCLC	Non-Small Cell Lung Cancer

PAGE 2

P Antigen Family Member 2

SAGE 1

Sarcoma Antigen 1

SCLC

Small Cell Lung Cancer

SqCC

Squamous Cell Carcinoma

Aim of the study:

In this project, we aim to design a new structural model containing putative antigenic epitopes. Since immune stimulation is considered one of the most important mechanisms in tumor treatment, tumor cells can escape from the immune system. It may be advantageous to use T cell epitopes of different tumor Antigens simultaneously, the goal of using multiple antigenic epitopes instead of a single antigen is to avoid that the specific antigen being lost or mutated. The use of multiple antigenic epitopes in a single structural model would cover a wide range of histocompatibility complex polymorphisms. Some of the most interesting cancer antigens for the development of cancer vaccines are cancer-testis antigens (CTAs). CTAs are aberrantly expressed by different tumor cell types while their normal expression is restricted to a few somatic tissues, including the testis.

Introduction:

The incidence of lung cancer is high, with 2.1 million new cases and 1.8 million casualties estimated worldwide, accounting for 18.4% of all cancer cases. Lung cancer is the third most prevalent cancer type in women globally and the most common cancer type in men [1].

Two major categories are discerned—small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). The first category constitutes approximately 15%, and the second is responsible for approximately 85% of tumors. The two most common entities within the NSCLC category are pulmonary adenocarcinoma (ADC) and pulmonary squamous cell carcinoma (SqCC), which represent about 90% of all NSCLC [2].

In the past, non-small cell lung carcinomas (NSCLC) were grouped without taking into account more specific histological typing (i.e., ADC, SqCC). This was accepted because there was no therapeutic implication to the separation of histological subtypes like ADC and SqCC [3,4].

Based on different escape mechanisms of tumor cells, a variety of novel immunotherapeutic strategies have been designed and optimized. Presentation of specific tumor cell antigens inducing an effective immune response is one of the most important immunotherapeutic mechanisms against tumor cell immune evasion [5]. In vivo immune responses generally begin through unique functions of dendritic cells (DCs), followed by priming of naïve T cells [6]; therefore, to design a DC-based cancer immunotherapy, we can transfect DCs with nucleic acids encoding tumor specific antigens (Ags) or incubate them with tumor-specific molecules, such as proteins, peptides, or lysates [7].

Exclusively expressed tumour Ags, which are recognized by T cells, have been studied in various antitumour immunotherapies [8]. These Ags as potential immunogens are processed to short peptides that bind to MHC class I molecules and present to T-cell receptors of tumor-reactive cytotoxic T lymphocytes (CTLs). Epitope-based vaccines powerfully stimulate immune responses against immunogenic epitopes of different antigens while avoiding unknown properties of using whole gene products [8]. The goal of immunization with these peptide epitopes is to achieve therapeutic benefits; however, to prevent the escape of tumor cells, it may be advantageous to use T cell epitopes of different tumor Ags simultaneously [9].

Some of the most interesting cancer antigens for development of cancer vaccines are CTA. CTAs are aberrantly expressed by different tumor cell types while their normal expression is restricted to a few somatic tissues, including testis. Therefore, CTAs are primary candidates for vaccination in cancer patients [10].

NSCLC, along with melanoma and ovarian cancer, are the most frequently expressed CTAs among the various cancers analyzed [11].

RNA sequencing (RNAseq) data from 199 patients with NSCLC [12] served as a basis for this analysis.

Dataset:

232 CTAs from the CT database were evaluated in a data set of 199 NSCLC cases and 32 normal tissues obtained from 141 individuals. The analysis revealed 96 CTAs that were expressed in NSCLC and showed exclusive expression in testis and placenta among normal tissues. These CTAs were designated as "confirmed CTAs". The information of previous studies regarding mRNA and protein expression for each gene in NSCLC was obtained from CT database, Table 1 shows confirmed CTAs in NSCLC [12,13].

Table 1. Clinical characteristic of 199 NSCLC cases included in the RNAseq analysis [12]

	No.	(%)
All patients	199	(100.0)
Sex		
Female	103	(51.8)
Male	96	(48.2)
Age		
≤70 years	120	(60.3)
>70 years	79	(39.7)
Stage		
IA	70	(35.2)
IB	45	(22.6)
IIA	25	(12.6)
IIB	23	(11.5)
IIIA	33	(16.6)
IV	3	(1.5)
Histology		
Adenocarcinoma	108	(54.3)
Squamous cell carcinoma	67	(33.7)
Not otherwise specified	24	(12.0)
Performance status^A		
0	120	(60.3)
1	77	(38.7)
2	2	(1.0)
Smoking		
Ever	180	(90.5)
Never	19	(9.5)

^AAssessment of a patient's performance status according to WHO score.

mRNA and protein expression profile analysis:

The protein expression of reported CTAs in NSCLC. From 232 genes, 68 were expressed in NSCLC based on the CT database. The CT database (<http://www.cta.lncc.br>) is a systematic data repository for CTAs, currently including 276 genes designated as CTAs, with curated information about gene and protein expression in normal and cancer tissues. Of these, 24 were described based on mRNA and protein levels, while the remaining 44 CTAs were only defined based on mRNA levels (Figure 1) [12,14].

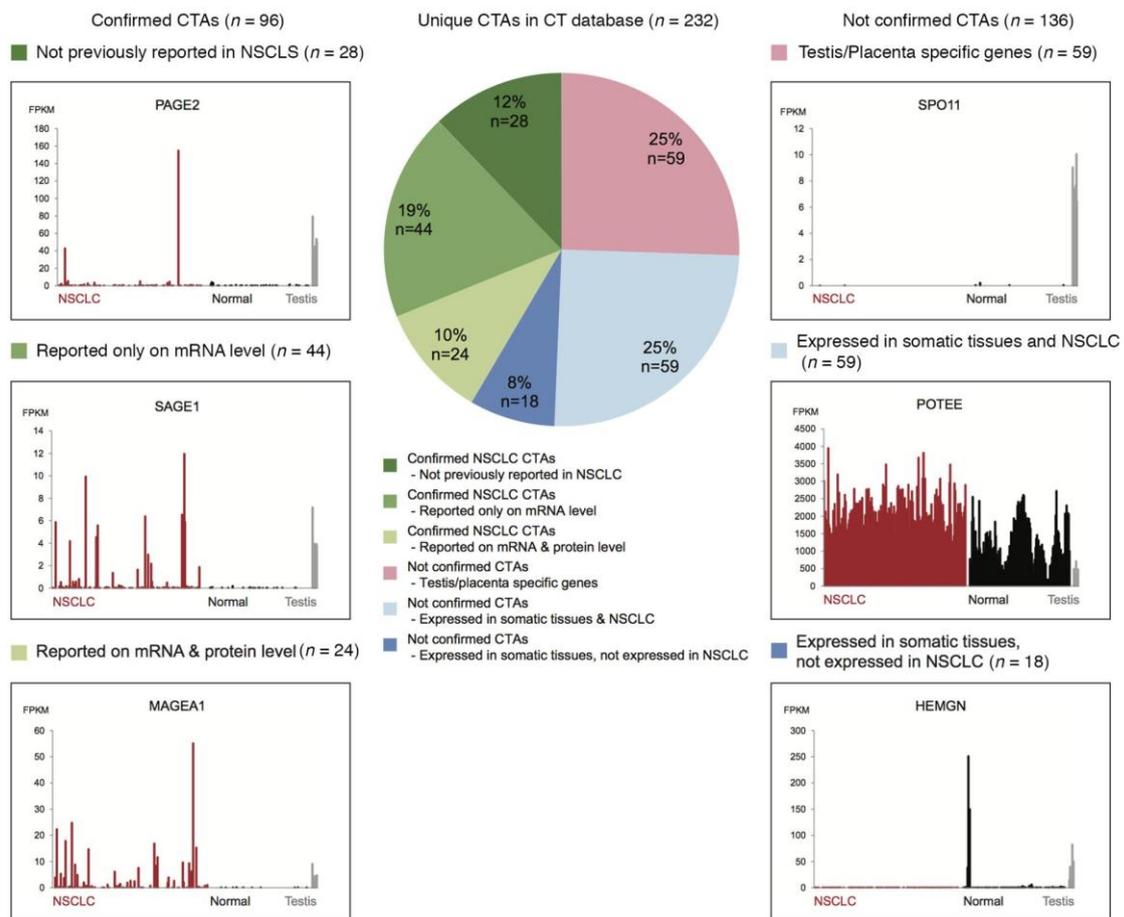


Figure 1. The expression of reported cancer testis antigens (CTAs) in non-small-cell lung cancer (NSCLC). Reported CTAs in the CT database ($n = 232$) were analyzed in 199 NSCLC cases and 142 normal tissues from 32 different organs. Based on the mRNA expression for each CTA in NSCLC and normal tissues, these 232 genes were grouped in either confirmed CTAs (green shades) or not confirmed CTAs and subdivided in testis/placenta-specific genes without expression in NSCLC (pink) or CTAs with expression in somatic tissues (blue shades) [12].

Using the Human Protein Atlas (HPA) image database, we confirmed the protein expression of 8 CTAs in NSCLC (MAGEC2, MAGEB6, PAGE2, PAGE5, PAGE2B, CT45A2, SAGE1 and MAGEA8) [12]. According to the data collected from the HPA and RNA sequence analysis, three CTAs will be chosen for the polytope chimeric syntheses. MAGEA8 (melanoma-associated antigen 8), CT45A2 (Cancer/testis antigen family 45 member A2), and SAGE1 (sarcoma antigen 1) [10].

The mRNA expression levels for the three CTAs were compared in accordance with the CT database. Table 2 illustrates the mRNA expression levels [13].

Table 2. mRNA expression levels of the three CTAs chosen for further analysis [13]

GENE ID	Description	mRNA expression level NSCLC %.
MAGEA8	Melanoma associated antigen 8	74-86
SAGE 1	Sarcoma Antigen 1	69-76
CTA 45 A2	Cancer/testis antigen family 45 A2	25-48

The protein expression for the three CTAs results obtained from HPA [15] are shown in figure 2 (for MAGEA8), figure 3 (for SAGE1), and figure 4 (for CTA 45 A2).

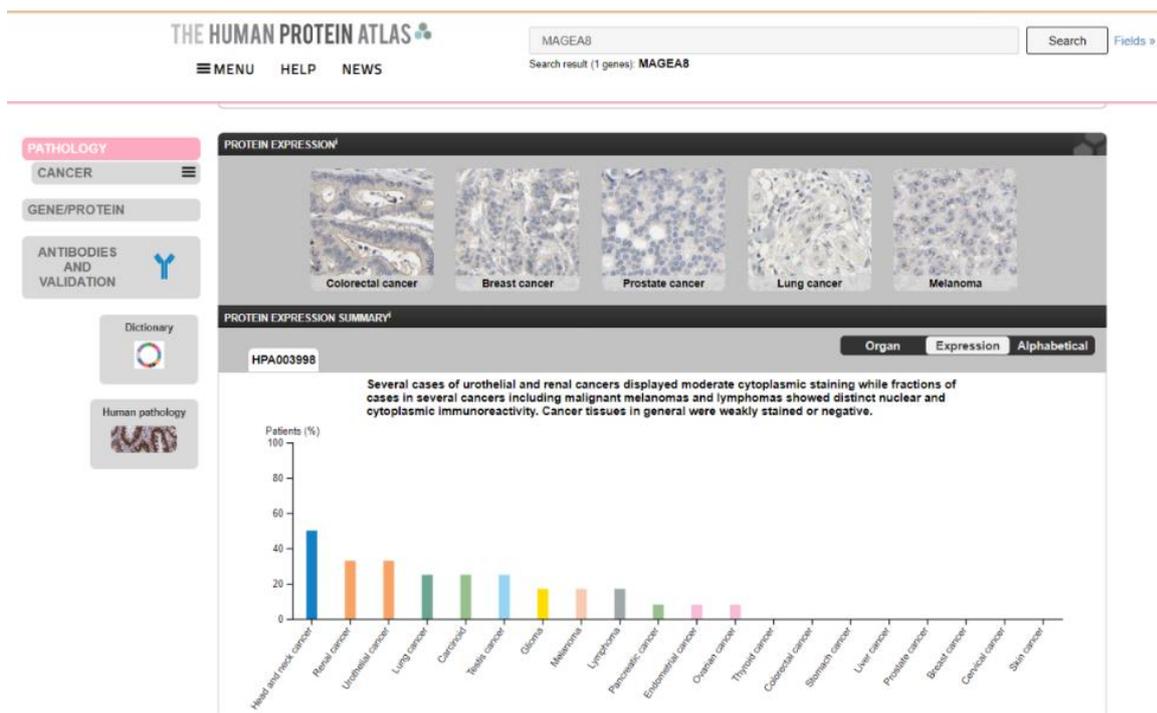


Figure 2. MAGEA8 protein expression in lung cancer comparing to other cancers [15].

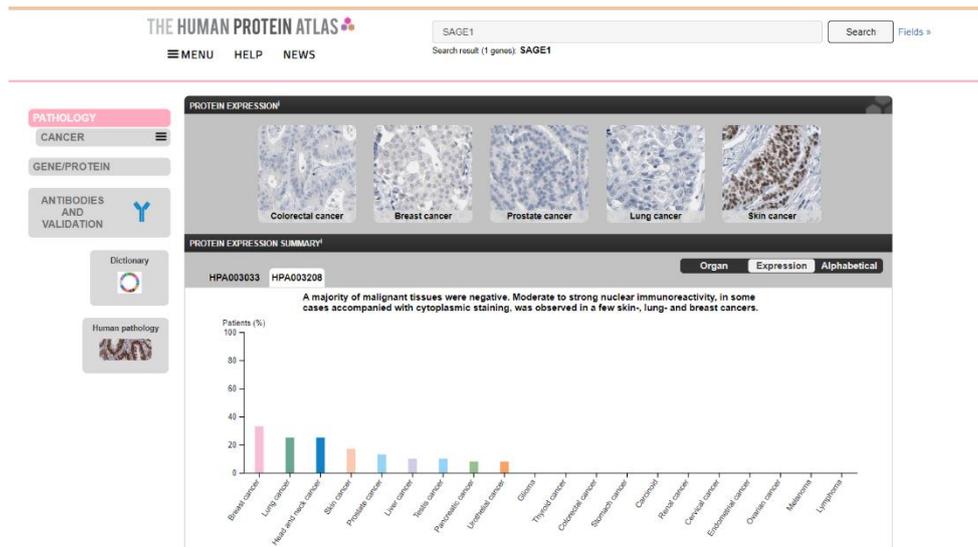


Figure 3. SAGE1 protein expression in lung cancer comparing to other cancers [15].

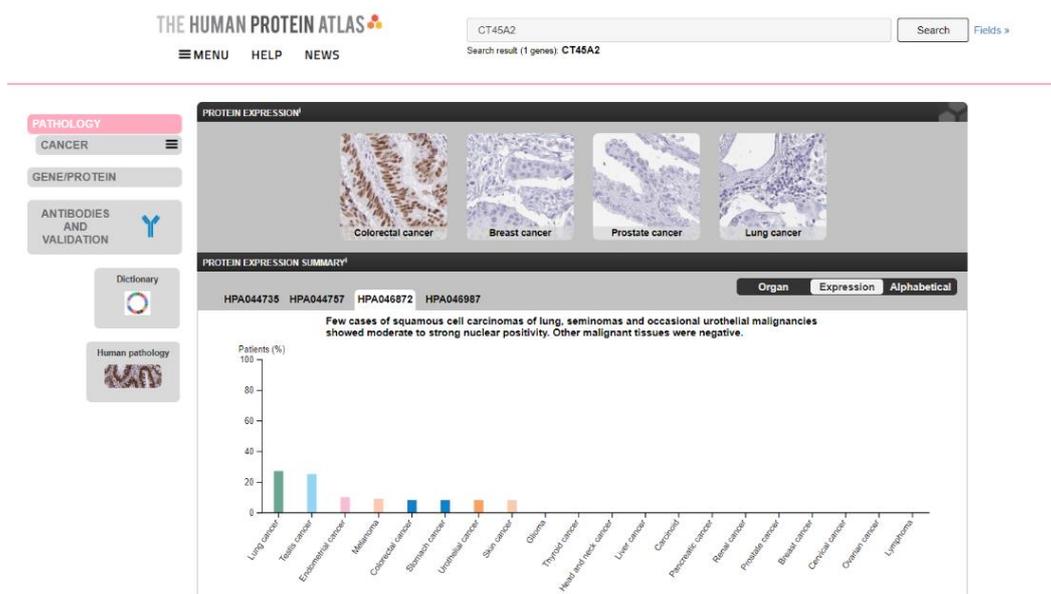


Figure 4. CTA 45 A2 protein expression in lung cancer comparing to other cancers [15].

Methods:

I. Sequence analysis

Immunogenic epitopes for MAGEA8, SAGE1, and CTA 45 A2 proteins were selected along with the hemagglutinin epitope (HA) as a reporter tag, the ER signal peptide, and the ER retention signal. The related nucleotide sequences were obtained from Genbank ([GenBank Overview \(nih.gov\)](http://GenBank Overview (nih.gov))) [16].

II. Design of the construct and gene optimization

A chimeric immunogenic sequence having an HA tag and targeted to the ER was constructed. The selected epitopes were fused using hydrophobic amino acid linkers [8]. To optimize the multiparameter chimeric gene, the in-silico analysis was performed using online data bases such as Emboss translation online tool ([EMBOSS Transeq < Sequence Translation Sites < EMBL-EBI](#)) [17], To identify the epitopes from each Antigen's protein SYFPEITHI a database (<http://www.syfpeithi.de/>), for MHC ligands and peptide motifs were used [18], followed by IEDB The Immune Epitope Database ([IEDB.org: Free epitope database and prediction resource](#)) to assist in the prediction and analysis of epitopes [19]. Following verification of the construct's properties by Gen-Script (NJ, USA) [20].

III. In silico structural analysis of chimeric recombinant protein

The program mfold (<http://www.bioinfo.rpi.edu/applications/mfold>) was used to analyse the secondary structure of the chimeric gene mRNA [21]. The secondary and 3D structures of the recombinant protein were predicted online *ab initio* software Robetta a protein structure prediction service (<https://rosetta.bakerlab.org>) [22]. Energy minimization is determined by analysis of 3D structural stability of the chimeric protein using Swiss-PdbViewer software [23]. Solvent accessibilities of the protein residues were evaluated with the online program VADAR, (<http://redpoll.pharmacy.ualberta.ca/vadar/>) [24].

IV. Prediction of the cleavage site

Cleavage site analysis on the construct protein was performed using NetChop server, an improved neural network training strategy ([NetChop - 3.1 - Services - DTU Health Tech](#)) [25].

This server produces neural network predictions for cleavage sites of the human proteasome using two different methods; C-term 3.0 and 20S 3.0 [8].

V. Validation of T-cell epitopes and MHC binding peptides affinity

The amino acid sequence was analyzed using four web-based T-cell epitope prediction algorithms;

NetCTL (<http://NetCTL - 1.2 - Services - DTU Health Tech>) [26], SYFPEITHI (<http://www.syfpeithi.de/>) [18], CTLPred (<http://www.imtech.res.in/raghava/ctlpred/index.html>) [27], and NetMHC (<http://www.cbs.dtu.dk/services/NetMHC/>) [28]. The NetMHC server produces a neural network prediction of binding affinities for MHC [29].

VI. Prediction of post-translational modifications

To predict post-translational modifications, three web-based servers were used. The NetOglyc 4.0 server produced neural network predictions of mucin-type GalNAc O-glycosylation sites ([NetOglyc - 4.0 - Services - DTU Health Tech](#)) [30]. The NetNglyc server predicted N-glycosylation sites in the construct protein using artificial neural networks (ANNs) that examine the sequence context of Asn-Xaa-Ser/Thr sequences ([NetNGlyc - 1.0 - Services - DTU Health Tech](#)) [31]. The NetPhos 2.0 server predicts serine, threonine, and tyrosine phosphorylation sites ([NetPhos - 3.1 - Services - DTU Health Tech](#)) [32].

Methods steps summary:

Figure 5 shows a flowchart that summarizes the project methods steps.

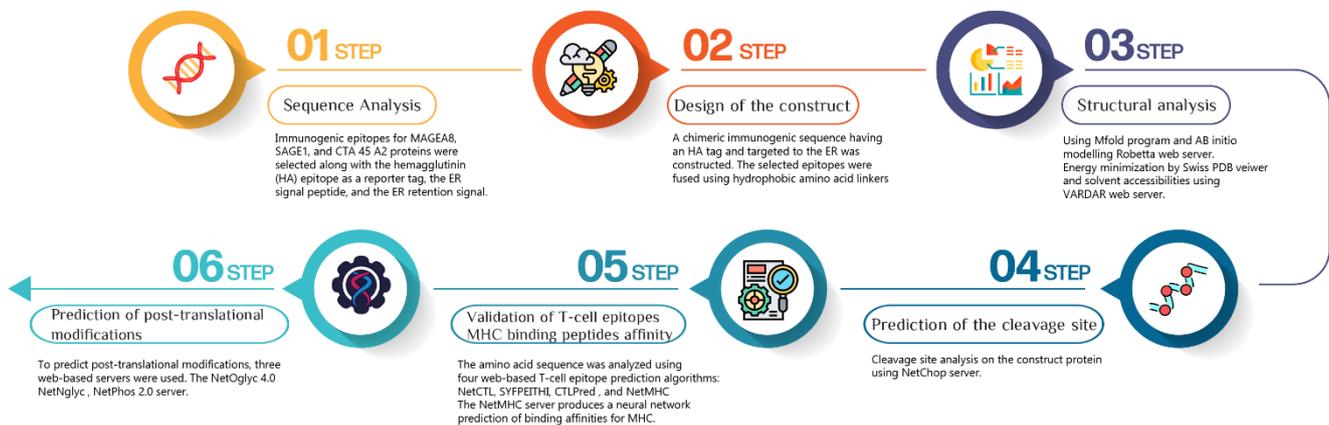


Figure 5. Flowchart of the project steps.

Results:

I. Design and structure of the chimeric construct

After searching for each of the CTA genes in Genebank, we obtained the sequences, and downloaded each sequence from genebank in Fasta file format [16]. Using Emboss to translate the genes for each of the proteins into amino acids sequences for epitopes prediction [17]. Using SYFPEITHI webserver tool which predicts epitopes according to HLA specific selection [18]. The scoring system evaluates every amino acid within a given peptide. high score indicates a strong binder [33].

Epitope prediction for MAGEA8 protein:

The MAGE-A8 protein sequence was screened for HLA-A2.1-binding motifs, six potential peptides were synthesised, and peptides binding to HLA-A2.1 were assured. HLA-A2.1 is the most widespread allele of HLA in the human population, therefore our allele of choice for the search of tumour CTL epitopes [34]. Nonamer had been chosen in the input of SYFPEITHI server. Getting multiple epitopes prediction and using IEDB for immunogenicity score, Epitopes containing amino acids of 115-123 of MAGEA8[18,19]. (Table 3,4)

Epitope prediction for SAGE1 protein:

Sarcoma antigen 1 (SAGE1) is a CTA that was first reported to have a similar expression pattern to the melanoma antigen gene (MAGE) family, which are expressed in bladder carcinoma, lung carcinoma, head and neck carcinoma, and germ-line cells. Subsequent reports have suggested that SAGE1 is also expressed in head and neck squamous cell carcinoma and esophageal squamous cell carcinoma. Due to its tumor-specific expression pattern, SAGE1 is a potential target for cancer immunotherapy [35]. Nonamer had been chosen in the input of SYFPEITHI server. Getting multiple epitopes prediction and using IEDB for immunogenicity score, Epitopes containing amino acids of 841-849 of SAGE1[18,19]. (Table 3,4)

Epitope prediction for CTA 45A2 protein:

CT45 exhibits the typical CT antigen characteristics and shows significant expression in lung cancer, ovarian cancer, and probably other tumor types yet to be tested, we believe that CT45 should remain a strong candidate for cancer vaccine [36]. Getting multiple epitopes prediction and using IEDB for immunogenicity score, Epitopes containing amino acids of 7-15 of CTA45A2 [18,19]. (Table 3,4)

Table 3: Epitopes prediction scores via SYFPEITHI.

Gene	POS	1	2	3	4	5	6	7	8	9	Score
MAGEA8	111	A	L	D	E	K	<u>V</u>	A	E	L	33
	45	L	I	M	G	T	<u>L</u>	E	E	V	29
	204	L	L	I	I	V	<u>L</u>	G	M	I	26
	288	K	V	L	E	H	<u>V</u>	V	R	V	25
	115	K	V	A	E	L	<u>V</u>	R	F	L	24
	179	Y	I	L	V	T	<u>C</u>	L	G	L	24
	240	S	V	Y	W	K	<u>L</u>	R	K	L	24
SAGE1	841	N	Y	E	R	I	F	I	L	L	24
	715	L	Y	A	T	V	I	H	D	I	22
	621	Q	Y	A	A	V	T	H	N	I	21
	597	V	F	S	T	V	P	P	A	F	20
	776	L	Y	K	P	D	S	N	E	F	20
CTA45A2	143	K	I	F	E	M	<u>L</u>	E	G	V	27
	129	Q	L	V	K	E	<u>L</u>	R	C	V	24
	7	K	V	A	V	D	<u>P</u>	E	T	V	19
	34	A	L	L	A	R	<u>K</u>	Q	G	A	19
	50	S	A	M	S	K	<u>E</u>	K	K	L	19

Table 4: Epitopes immunogenicity scores via IEDB.

Gene	Immunogenicity class	Allele	Maske d variables	Peptide	Length	Score
MAGEA8	I	HLA-A0201	[1,2,'cterm']	KVAELVRFL	9	0.25359
				KVLEHVVRV	9	0.23259
				LIMGTLEEV	9	0.14746
				LLIIVLGM I	9	0.13288
				ALDEKVAEL	9	0.0283
				SVYWKLRKL	9	-0.08107
SAGE1	I	HLA-A2402	[2,7,9]	NYERIFILL	9	0.3179
				LYATVIHDI	9	0.2302
				QYAAVTHNI	9	0.12503
				VFSTVPPAF	9	0.03798
				LYKPDSNEF	9	-0.15679
CTA45A2	I	HLA-A0201	[1,2,9]	KVAVDPETV	9	0.17258
				KIFEMLEGV	9	0.06161
				QLVKELRCV	9	-0.10436
				ALLARKQGA	9	-0.19479
				SAMSKEKKL	9	-0.64722

All the peptides are immunogenic and reported to be recognized by CTLs. These three peptides were used to design the chimeric construct. The epitopes were linked by Gly-Pro-Gly-Pro-Gly (GPGPG) repeats. These repeats are expected to prevent formation of junctional epitopes when the protein is cleaved during the presentation process in antigen presenting cells (APCs) [8]. It has been shown that GPGPG spacers eliminate responses against the junctional epitope, allowing the development of a balanced response [37]. To increase the accuracy and efficiency of translation in a human host, the Kozak sequence was added 5' to the start codon. Efficient entrance and accumulation of the recombinant protein in the endoplasmic reticulum (ER) can facilitate processing of epitopes [38]; therefore, an ER signal sequence was added at the 5' end of the construct, and the KDEL sequence was added at the 3' end to make it resident in the ER [38]. The HA epitope tag (YPYDVPDYA) from the human CTL influenza hemagglutinin protein was used to track the gene product in downstream assays. The HA-tag was placed 3' to the ER signal sequence to minimize any potential disruption in tertiary structure, and thus function, of the protein. ER signal sequence was obtained from ncbi query search [8]. The structure of the chimeric gene and arrangements of fragments and linker sites are shown in **Figure 6**.

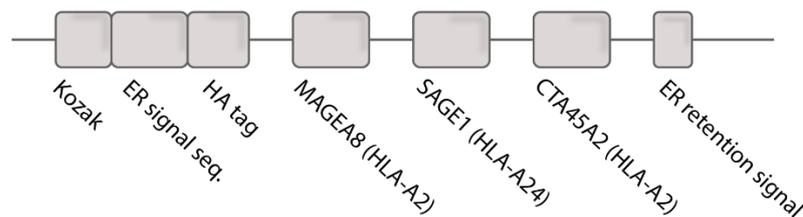


Figure 6: Schematic model of the construct. The selected epitopes of MAGEA8, SAGE1 and CTA45A2 are bound together by the linkers for expression in human. These fragments were selected on the basis of HLA restriction of MHC class I, to be recognized by CTLs.

Signal peptide prediction:

Signal peptide was analysed for the prediction significance after constructing the polytope chimeric sequence. Phobius web server a combined transmembrane topology and signal peptide predictor was used for the signal peptide significance prediction and the result are shown in figure 7 with a significance of 0.999 for all the signal amino acids [39].

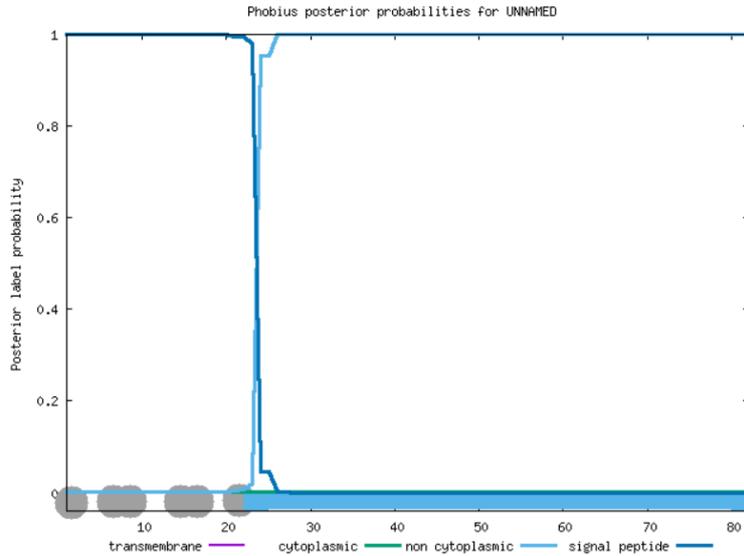


Figure 7. Signal peptide prediction significance results after constructing the polytope chimeric sequence using Phobius web server resulting **in** a significance of 0.999 for all the signal amino acids [39].

Chimeric gene Sequence:

“AAGCTTGCCGCCACCATGGGCATGCAGGTGCAGATCCAGAGCCTGT
 TCCTGCTGCTGCTGTGGGTGCCTGGATCCCGGGGATACCCATATGAC
 GTGCCTGATTACGCTGGACCAGGACCTGGGAAGGTGGCCGAGCTGG
 TGAGGTTCTGGGACCAGGACCTGGGAACTACGAGAGGATCTTCATC
 CTGCTGGGACCAGGACCTGGGAAAGGTGGCCGTGGACCCCGAGACCG
 TGGGACCAGGACCTGGGAAAGATGAACTGTGAGAATTC”

Chimeric polytope peptide sequence:

MGMQVQIQSLFLLLWVPGSRGYDPDYAGPGPGKVAELVRFLGPG
 PGNYERIFILLGPGPGKVAVDPETVGP GPGKDEL

II. In silico analysis of original chimeric gene

Human codon bias was considered to design the chimeric gene. Codon bias and the GC content of the chimeric gene were analyzed. Codon Adaptation Index (CAI) of the gene is 0.90. A CAI of 1.0 is considered ideal while a CAI of >0.8 is rated as good for expression in the desired expression organism. The lower the number, the higher the chance that the gene will be expressed poorly as shown in figure 8 [8,20]. The GC content of the gene is 62.58%. The ideal percentage range of GC content is between 30% and 70%. Any peaks outside of this range will adversely affect transcriptional and translational efficiency as shown in figure 9 [8,20]. No optimization has been conducted. Finally, HindIII

and EcoRI restriction sites were introduced at the 5' and 3' ends of the sequence, respectively.

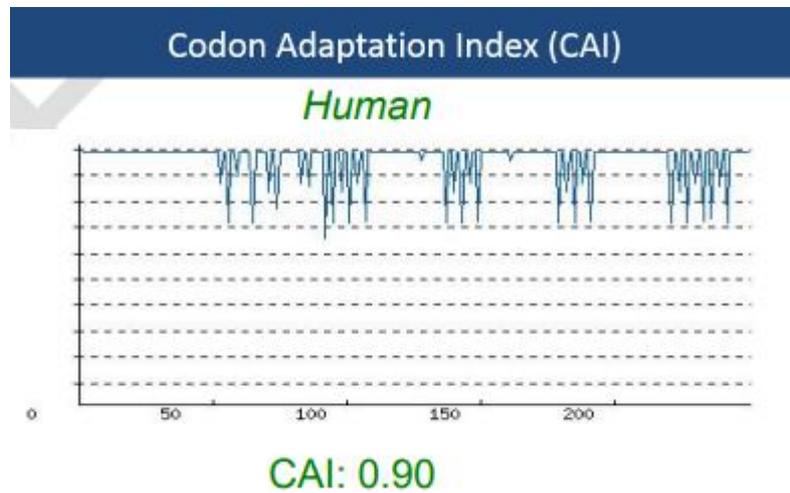


Figure 8. Codon Adaptation Index (CAI) of the gene. The distribution of codon usage frequency along the length of your CDS to be expressed in your target host organism. Possibility of high protein expression level is correlated to the value of CAI- a CAI of 1.0 is considered to be ideal while a CAI of >0.8 is rated as good for expression in the desired expression organism [20].

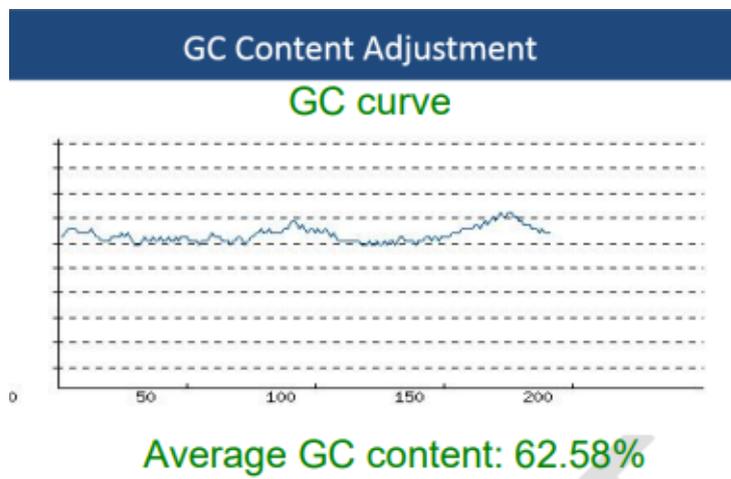


Figure 9. GC content of the gene. The ideal percentage of GC content is between 30% to 70%. Any peaks outside of this range will adversely affect transcriptional and translational efficiency [20].

III. mRNA structure prediction

To verify potential folding of the chimeric mRNA, Mfold webserver for the prediction of peptide secondary structure of single strand nucleic acids was used. The objective of this web server is to provide easy access to RNA and DNA folding and hybridization software to the scientific community at large. Detailed output, in the form of structure plots [21,40]. Ss-count file obtained from the output results. ss-count is the propensity of a base to be single stranded, as measured by the number of times it is single stranded in a group of predicted foldings. The first line of the ss-count file contains the number of expected foldings. 14 folding structures are displayed on the output result page. The optimal structure with the lowest ΔG has been shown in figure 10 [40].

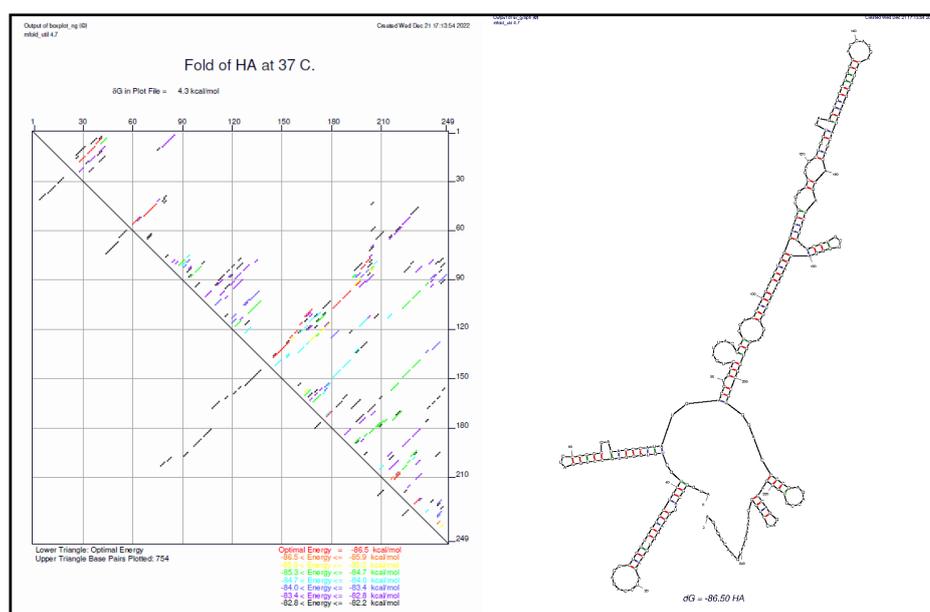


Figure 10: The optimal structure with the lowest ΔG [40]

IV. Prediction of secondary and tertiary structures of chimeric protein

Using online software, different prediction methods were compared to evaluate the secondary structure of the chimeric protein [8]. Using Jpred 4 for secondary structure prediction the results showed two beta sheets, and one alpha helix comprised of amino acids 18 (figure 11) [41]. Chimeric protein 3D models, produced by ab initio modelling using Robetta web server (figure 12) [42] and were uploaded to the Swiss-PdbViewer server to render the tertiary structural illustrations (figure 13) [43]. The final structure of the protein was predicted by Pymol software (figure 14) [44]. One alpha-helix and two beta sheets were identified, supporting the results of the secondary structure analysis.

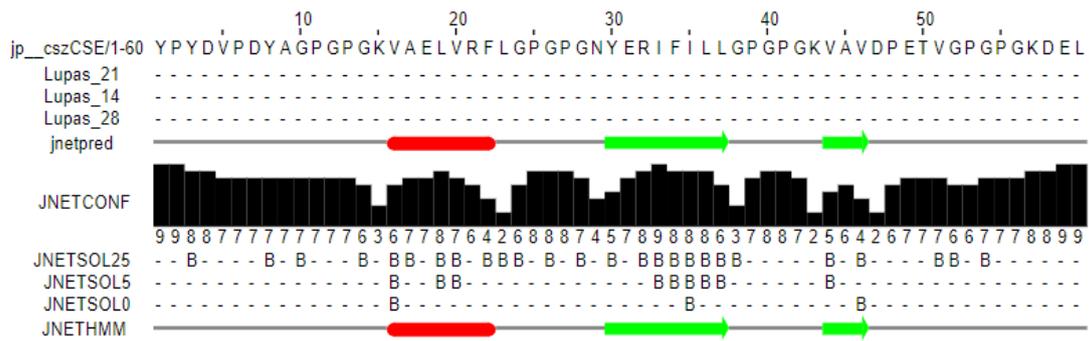


Figure 11: secondary structure prediction Using Jpred 4 for secondary structure prediction the results showed two beta sheets, and one alpha helix comprised of amino acids 18 [41]

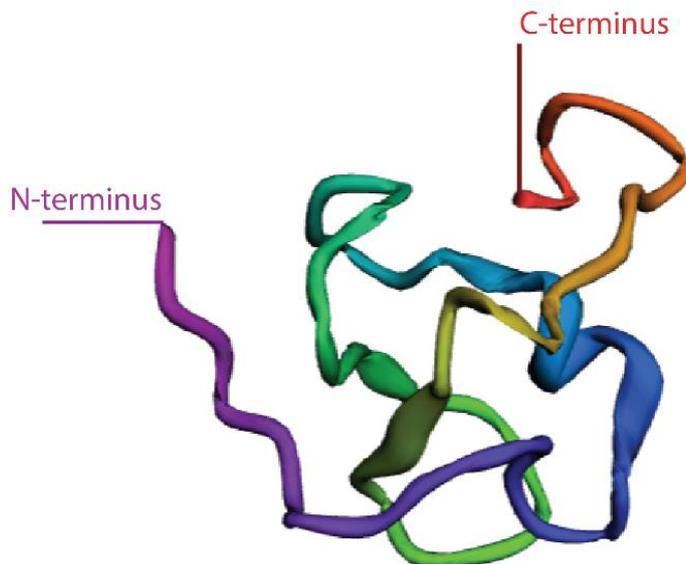


Figure 12: Protein 3D structure Ab initio modelling using Robetta web server De novo models are built using the Rosetta de novo protocol. The procedure is fully automated [42].

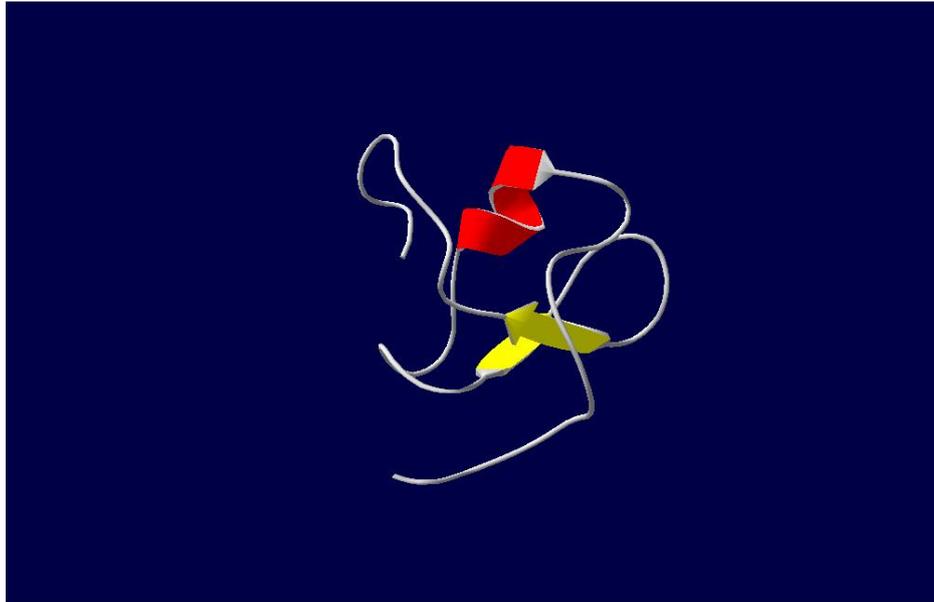


Figure 13: Secondary structure predicted by AB Initio modelling and the result was viewed by Swiss pdb viewer [43].

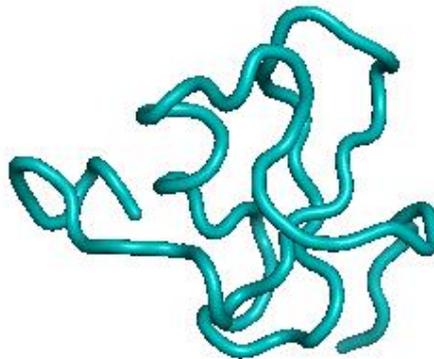


Figure 14: Tertiary structure predicted by AB Initio modelling and the result was viewed by Pymol [44].

V. Evaluation of model stability

By minimizing the energy of a molecule, the stability of the model is confirmed. Energy minimization was determined by analysis of 3D structural stability of the chimeric protein using Swiss-PdbViewer [8]. The energy minimization profile performed by spdbv (Swiss-PdbViewer) and calculated to be -555.381 KJ/mol [43].

This result refers to that the recombinant protein was relatively stable. Also, the structural stability of the chimeric protein was confirmed based on data generated by a Ramachandran plot (figure 15) [24].

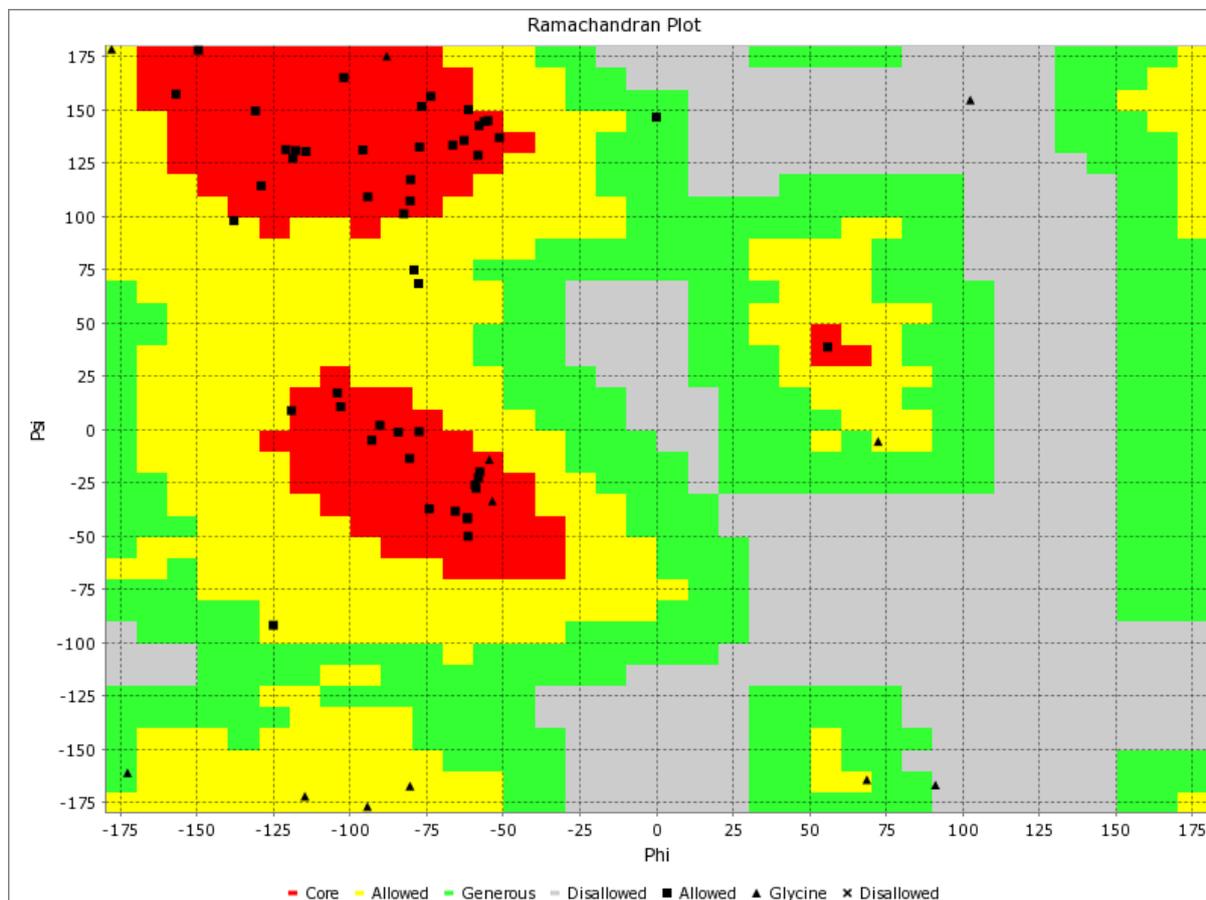


Figure 15: Evaluation of model stability, the structure stability was confirmed based on the Ramachandran plot, the dihedral angles of amino acid residues appear as crosses in the plot.

The red and yellow regions represent the favoured and allowed regions. The red regions correspond to conformations where there are no steric clashes in the model. These favoured regions include the dihedral angles typical of the alpha-helical and beta-sheet conformations. The green areas correspond to conformations where atoms in the protein come closer than the sum of their van der Waals radii. These regions are sterically forbidden for all amino acids with side chains [24].

VI. Solvent accessibility prediction

The analysis of the fractional accessible surface area (ASA) and fractional residue volume showed that all residues have fractional volumes below 1.0. According to VARDAR accessible surface area can be reported in square angstroms or as a fractional ASA (ranging from 0.00 to 1.00). Figure 16-A. Therefore, the protein is efficiently packed with no major packing defects. Stereochemical/packing quality analysis revealed that most residues have good

quality scores near 8 and this showed that the protein has a high-resolution structure. Figure 16-B. 3D profile quality analysis examined local environment, packing, and hydrophobic energy for the protein structure, and the results showed an acceptable quality index. Figure 16-C. (Typically, these threading quality indices range between 5-8. Values that are significantly lower (<5) indicate possible problems with the local structure or local fold) [8,24].

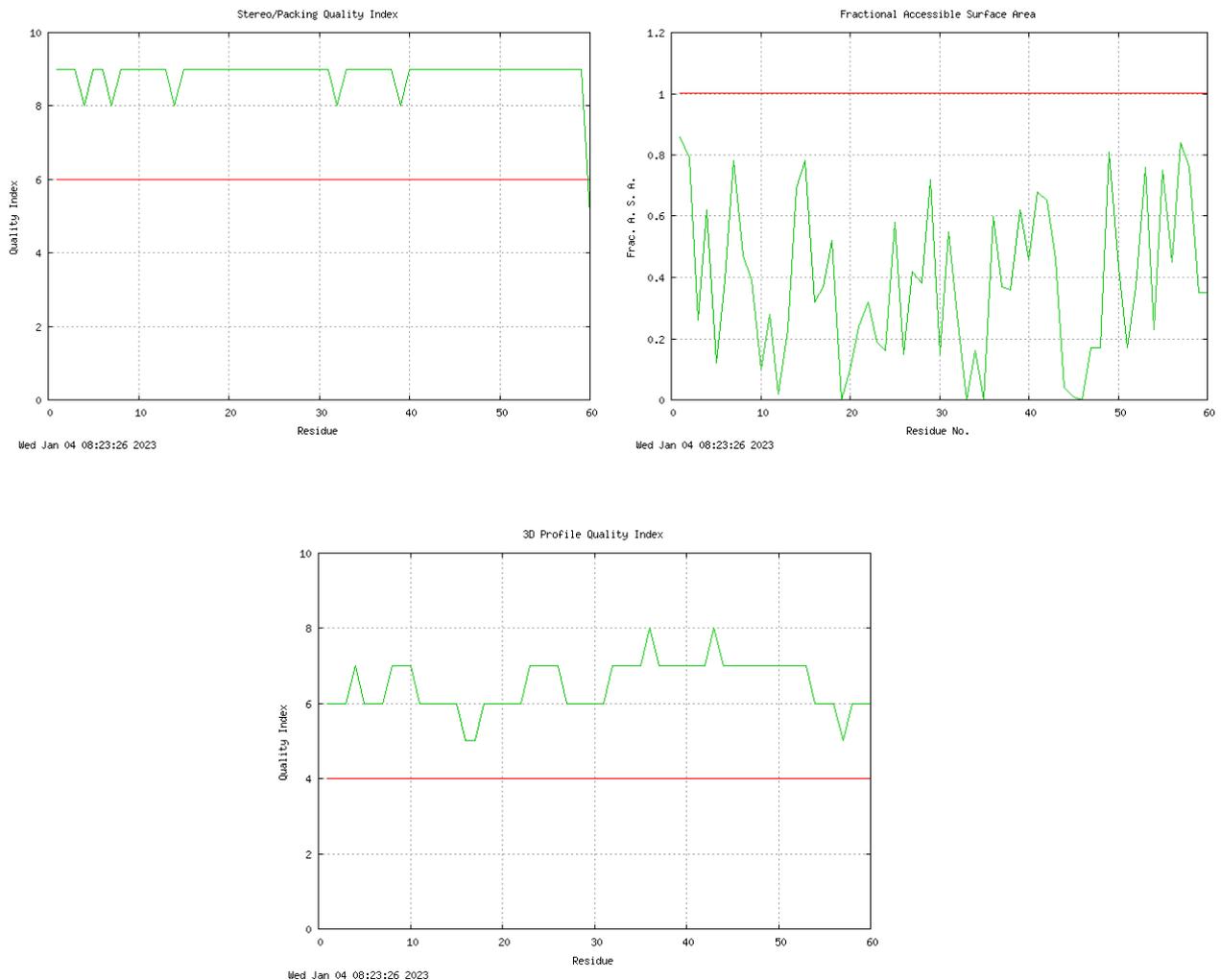


Figure 16: Solvent accessibility prediction. A) Fractional accessible surface area (ASA) analysis, all residues have fractional volumes below 1.0. According to VARDAR accessible surface area can be reported in square angstroms or as a fractional ASA (ranging from 0.00 to 1.00) B) Stereochemical/packing quality analysis revealed that most residues have good quality scores near 8 and this showed that the protein has a high-resolution structure, and C) 3D profile quality analysis of the construct protein, examined local environment, packing, and hydrophobic energy for the protein structure, and the results showed an acceptable quality index [24]

VII. Prediction of cleavage sites

NetChop is a tool to predict cleavage sites of the human proteasome. The NetChop algorithm uses a neural network trained on human proteasome data.

Cleavage sites on the construct protein were analyzed with NetChop. As expected, no cleavage sites were predicted inside the linkers so the production of junctional epitopes was prevented. Also, the cleavage sites with high prediction scores were located at both ends of each selected epitope. The results are summarized in table 5 [24].

Table 5. Prediction of cleavage sites on the constructed protein using NetChop server. Each amino acid in the table is the location of cleavage while no sites are located in the linkers. The threshold is 0.5 [24].

Position	Amino acid	Score
23	Y	0.866598
31	A	0.882795
37	K	0.551773
45	L	0.530573
52	Y	0.894447
59	L	0.964388
65	K	0.945372
73	V	0.826600

VIII. Validation of T-cell epitopes

NetCTL 1.2 server predicted CTL epitopes in the chimeric protein sequence. The server predicted CTL epitopes restricted to 12 MHC class I supertypes using ANNs [45]. The scores from the individual prediction methods were integrated, and thresholds for the integrated scores of each peptide were translated into sensitivity and specificity values (Table 6) [26].

The SYFPEITHI epitope prediction algorithm was also used. To find out the ligation strength to a defined HLA type for a sequence of amino acids. The algorithm used are based on the book "MHC Ligands and Peptide Motifs" by H.G.Rammensee, J.Bachmann and S.Stevanovic. The probability of being

processed and presented is given in order to predict T-cell epitopes. The scoring system of SYFPEITHI evaluated each amino acid in the peptides. The maximum score for HLA*0201 peptides is 36. The scores for epitopes of the chimeric protein are shown in table 6 [18].

CTLPred, is a direct method for prediction of CTL epitopes crucial in subunit vaccine design. The methods are based on elegant machine learning techniques as Artificial Neural Network ANN and Support Vector Machine SVM [46]. The scores of CTLPred predicted epitopes for the chimeric protein are shown in table 6 [27]. The default cut-off score at which the sensitivity and specificity of prediction methods are highly similar was 0.51.

Table 6: Prediction of T-cell epitopes of the construct using different web-based servers.

	SYFPEITHI	CTLPred	NetCTL*
Peptide	Score	score	score
KVAELVRFL	24	0.951	0.8147
NYERIFILL	24	0.931	1.6923
KVAVDPETV	19	0.610	0.7260

*Score>1.25: (sensitivity=0.54, specificity= 0.993), score>1.00: (sensitivity=0.70, specificity= 0.985), score>0.90: (sensitivity=0.74, specificity= 0.980), score>0.75: (sensitivity=0.80, specificity= 0.970), score>0.50: (sensitivity=0.89, specificity= 0.940)

X. MHC binding peptides affinity

NetMHC 4 server Predicted peptide binding to a number of different HLA alleles using artificial neural networks (ANNs) [29]. Rank Threshold for Strong binding peptides is 0.500 and rank threshold for weak binding peptides is 2.000. The results are summarized in table 7 [28].

Table 7: Predictions of MHC-binding peptide affinity for the construct by NetMHC version 4.0. server using ANNs approximation [28].				
Peptide	Log score	Affinity (nM)	Rank	Binding level
KVAELVRFL	0.502	219.23	1.70	WB
NYERIFILL	0.488	253.59	0.50	SB
KVAVDPETV	0.332	1374	5.00	–

XI. Prediction of post-translational Modifications

To predict post-translational modifications, three web-based servers were used. NetOglyc server Find the presence of N-Glycosylation sites in human proteins. Produced neural network predictions of mucin-type GalNAc O-glycosylation sites [30]. The NetNglyc server predicts N-Glycosylation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequences [31]. The NetPhos 3.1 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites [32]. Post-translational modifications, such as glycosylation and phosphorylation, are known to influence protein folding, localization and trafficking, solubility, antigenicity, biological activity, and half-life [47]. The ANNs were trained on the chimeric protein sequence context to predict glycosylation. N-link and O-link are the two main types of glycosylation in mammalian cells.

The sites with scores higher than **0.5** are predicted as glycosylated and marked with the string **POSITIVE** in the comment field. No O-linked glycosylation were predicted. Furthermore, based on the result of the ANN method, which predicts phosphorylation sites with sensitivity in the 69% to 96% range, the construct is potentially phosphorylated at residues Ser-10, Ser-20, and Tyr-30.

Discussion:

One of the principal goals of cancer immunotherapy is the development of efficient therapeutic cancer vaccines. Cancer vaccines are based on tumor antigens expressed in the context of Major Histocompatibility Complex (MHC) molecules able to elicit a strong tumor-specific CTL response, which may result in the killing of tumor cells and cancer regression [48]. Vital questions have arisen during tumor vaccine design. These include choice of the appropriate peptides, formulation, delivery mode, and molecular monitoring of the induced immune responses [49]. Depressed or loss of immunogenic epitopes by tumors and insufficient Ag presentation by APCs are major factors for the failure of the

immune system to establish effective immune responses against these tumor Ags; therefore, an approach that generates both Ag-specific CD4⁺ (Th) and CD8⁺ (CTL) responses may provide optimal immunization against tumors [8,50].

Tumor escape from CTL surveillance, through down regulation of individual tumor Ags and MHC alleles, might be overcome by polytope vaccines, which simultaneously target multiple cancer Antigens [51]. This strategy has advantages over using individual epitopes or intact target antigens, where individual epitopes may be lost or mutated, or where the target antigens may be oncogenic [8]. Antigenic epitopes from diverse antigens can be linked together in a single polytope construct; such insertion of different MHC class I-restricted epitopes allows wide coverage of an MHC polymorphic population [52]. CTAs have been considered promising targets for immunotherapy approaches thanks to their tumor-specificity and strong immunogenicity for the absence of immune tolerance [51]. We describe here the strategy in the design of a polytope cancer vaccine that has many unique characteristics. Combining different HLA-restricted epitopes from CTAs into one polytope vaccine construct allows the fusion Ag to efficiently enter the ER, then be processed and presented to MHC class I to induce the related CTL responses against all epitopes simultaneously [8]. In this study, we designed a new chimeric construct of CTAs including HLA-restricted epitopes of MAGEA8, SAGE1, and CTA45A2, which contained essential determinants to be recognized by CTLs. The DNA fragment encoding these putative antigenic epitopes was designed as a chimeric construct optimally suited for expression in human. Many factors could affect the expression of recombinant genes in a human host, such as mRNA stability, polyadenylation, splicing sites, antiviral motifs, and codon usage preferences [53].

The chimeric gene was designed based on codon usage of highly-expressed nuclear encoded genes in human. Each step of gene expression, from the transcription of DNA into mRNA to the folding and post-translational modification of proteins, is regulated by complex cellular mechanisms. A relationship between mRNA expression and protein solubility can now be predicted [54].

In eukaryotic cell mRNAs, the consensus sequence surrounding the start codon (Kozak seq. 5'GCCACCATGGC) can affect the precision and efficiency of translation. In the chimeric gene, the 5'GCCACC sequence was inserted 5' to the ATG codon. The next codon following the initial methionine ATG codon, GGA, encoding Gly, and the necessary G was provided [8]. Efficient entrance and accumulation of the recombinant protein in the endoplasmic reticulum (ER) can facilitate processing of epitopes. For successful CTL induction, the antigen peptide of interest should be efficiently delivered to the MHC class I-restricted presentation pathway via direct or cross-priming. Various DNA vaccination studies have suggested that cross-priming is more efficient than direct priming

while other studies indicate that direct priming is a very important process for CTL responses. Probably both processes occur following DNA vaccination and the predominant process would be determined by the experimental conditions used, including the type of construct or antigen, and the route of administration. We believe that optimization of the intracellular trafficking of expressed antigen peptide in DCs following direct transfection would be a useful approach for improving the efficacy of MHC class I-restricted presentation and subsequent CTL induction. It has been reported that the direct delivery of antigen peptide to ER improved the efficiency of CTL induction [55]. Codon Adaptation Index (CAI) of the gene is 0.90. A CAI of 1.0 is considered ideal. Moreover, The GC content of the gene is 62.58%. The ideal percentage range of GC content is between 30% and 70% (**Figure 9**). In addition, the required restriction enzyme sites were added to the ends of the designate gene for future assays [8]. Evaluation of model stability by Ramachandran plot showed that most residues of the chimeric model are in a stable zone. CTLs distinguish small peptides eight to ten amino acids long. These epitope peptides are generated by the proteasome system. Protease is responsible for intracellular protein degradation. The proteasome produces the exact C-terminus of CTL epitopes, and the N-terminus with a possible extension [56]. CTL responses could be reduced if the epitopes are destroyed by proteasomes; therefore, prediction of proteasome cleavage sites is valuable for identification of potential immunogenic regions in the chimeric protein. Based on these rules we designed the chimeric protein and then predicted its proteasomal cleavage sites using web-based software. The result showed that the highest-scored cleavage positions are located at the fusion site of each epitope and its adjacent linkers (**Table 5**). The use of GPGPG as a hydrophobic linker restricts the production of junctional epitopes, and this allows efficient downstream processing of the chimeric protein. Furthermore, The NetCTL 1.2 server predicts CTL epitopes in protein sequences. The accuracy of the MHC class I peptide binding affinity is significantly improved compared to the earlier version. Also, the prediction of proteasomal cleavage has been improved and is now identical to the predictions obtained by the NetChop server.

Therefore, based on the prediction results, the selected epitopes of our chimeric construct also showed high-affinity binding to MHC molecules and acceptable sensitivity and specificity to be recognized by CTLs (**Tables 6 and 7, respectively**). Epitope binding to MHC and recognition of such complexes (epitope/MHC) by CTLs is a critical step in inducing a significant immune response.

Conclusion:

We used *in silico* approaches to design our chimeric polytope construct of immune-gene therapy applications. We used several web servers and applications to predict different features of the construct, including GC content, secondary and tertiary structure of the protein, solvent accessibility of the chimeric protein, proteasomal cleavage site, validation of the epitope's prediction, MHC binding affinity, and post-translational modifications. Three epitopes with high immunogenicity scores were included in the study; MAGEA8, SAGE1, and CTA45A2. Both the MAGEA8 epitope and SAGE1 epitope gave a good binding prediction. However, only the SAGE1 epitope showed a strong binding affinity with MCH molecules. For future studies, the CTA45A2 epitope could be substituted with an epitope with a better binding prediction and affinity in order to develop a more effective structural model for cancer immune-gene therapy. Considering all of these results together, this study showed potential for the rational design of multiepitope chimeric cancer vaccines using immunoinformatics and various computational methods. With the ultimate objective of developing therapeutic vaccinations for cancer patients, this study provides the foundation for further refinement and optimization of the fusion gene expression approach.

References:

1. Cronin, K.A. *et al.* (2018) “Annual report to the nation on the status of cancer, part I: National cancer statistics,” *Cancer*, 124(13), pp. 2785–2800. Available at: <https://doi.org/10.1002/cncr.31551>.
2. Chen, Z. *et al.* (2014) “Non-small-cell lung cancers: A heterogeneous set of diseases,” *Nature Reviews Cancer*, 14(8), pp. 535–546. Available at: <https://doi.org/10.1038/nrc3775>.
3. Hirsch, F.R. *et al.* (2008) “The prognostic and predictive role of histology in advanced non-small cell lung cancer: A literature review,” *Journal of Thoracic Oncology*, 3(12), pp. 1468–1481. Available at: <https://doi.org/10.1097/jto.0b013e318189f551>.
4. Paech, D.C. *et al.* (2011) “A systematic review of the interobserver variability for histology in the differentiation between squamous and nonsquamous non-small cell lung cancer,” *Journal of Thoracic Oncology*, 6(1), pp. 55–63. Available at: <https://doi.org/10.1097/jto.0b013e3181fc0878>.
5. Gholamin, M. *et al.* (2010) “Induction of cytotoxic T lymphocytes primed with tumor RNA-loaded dendritic cells in esophageal squamous cell carcinoma: Preliminary step for DC Vaccine Design,” *BMC Cancer*, 10(1). Available at: <https://doi.org/10.1186/1471-2407-10-261>.
6. Landi, A., Babiuk, L.A. and van Drunen Littel-van den Hurk, S. (2007) “High transfection efficiency, gene expression, and viability of monocyte-derived human dendritic cells after nonviral gene transfer,” *Journal of Leukocyte Biology*, 82(4), pp. 849–860. Available at: <https://doi.org/10.1189/jlb.0906561>.
7. Banchereau, J. and Palucka, A.K. (2005) “Dendritic cells as therapeutic vaccines against cancer,” *Nature Reviews Immunology*, 5(4), pp. 296–306. Available at: <https://doi.org/10.1038/nri1592>.
- 8 Forghanifard, M.M. “In silico analysis of chimeric polytope of cancer/testis antigens for dendritic cell-based immune-gene therapy applications,” *Gene Therapy and Molecular Biology*, 14(2012), 87-96.

9. Buteau, C., Markovic, S.N. and Celis, E. (2002) “Challenges in the development of effective peptide vaccines for cancer,” *Mayo Clinic Proceedings*, 77(4), pp. 339–349. Available at: <https://doi.org/10.4065/77.4.339>.
10. Fratta, E. *et al.* (2011) “The biology of cancer testis antigens: Putative function, regulation and therapeutic potential,” *Molecular Oncology*, 5(2), pp. 164–182. Available at: <https://doi.org/10.1016/j.molonc.2011.02.001>.
11. Grunwald, C. *et al.* (2006) “Expression of multiple epigenetically regulated cancer/germline genes in nonsmall cell lung cancer,” *International Journal of Cancer*, 118(10), pp. 2522–2528. Available at: <https://doi.org/10.1002/ijc.21669>.
12. Djureinovic, D. *et al.* (2016) “Profiling cancer testis antigens in non–small-cell lung cancer,” *JCI Insight*, 1(10). Available at: <https://doi.org/10.1172/jci.insight.86837>.
13. *CTDatabase. CTpedia*. Available at: <http://www.cta.lncc.br/>.
14. Almeida, L.G. *et al.* (2009) “CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens,” *Nucleic Acids Research*, 37(Database). Available at: <https://doi.org/10.1093/nar/gkn673>.
15. *The human protein atlas. The Human Protein Atlas*. Available at: <https://www.proteinatlas.org/>.
16. *GenBank Overview. National Center for Biotechnology Information*. U.S. National Library of Medicine. Available at: <https://www.ncbi.nlm.nih.gov/genbank/>.
17. Embl-Ebi. *Emboss Transeq, EBI*. Available at: https://www.ebi.ac.uk/Tools/st/emboss_transeq/.
18. *Syfpeithi. SYFPEITHI*. Available at: <http://www.syfpeithi.de/>.
19. IEDB.org: Free epitope database and prediction resource: Available at: [IEDB.org: Free epitope database and prediction resource](https://www.iedb.org/)

20. Gene synthesis & DNA synthesis – guaranteed delivery time | GenScript. Available at: [GenScript - Make Research Easy - The leader in molecular cloning and gene synthesis, peptide synthesis, protein and antibody engineering.](#)
21. Zuker, M. (2003) “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Research*, 31(13), pp. 3406–3415. Available at: <https://doi.org/10.1093/nar/gkg595>.
22. Robetta. Available at: <https://rosetta.bakerlab.org/>
23. Edwards, Y.J. and Cottage, A. (2003) “Bioinformatics methods to predict protein structure and function: A practical approach,” *Molecular Biotechnology*, 23(2), pp. 139–166. Available at: <https://doi.org/10.1385/mb:23:2:139>.
24. VADAR, Available at: <http://vadar.wishartlab.com/>
25. NetChop-3.1. Available at: <https://services.healthtech.dtu.dk/service.php?NetChop-3.1>
26. NetCTL. Available at: <https://services.healthtech.dtu.dk/service.php?NetChop-3.1>
27. CTLPred. Available at: <http://www.imtech.res.in/raghava/ctlpred/index.html>
28. NetMHC. Available at: <https://services.healthtech.dtu.dk/service.php?NetMHC-4.0>
29. Buus, S. et al. (2003) “Sensitive quantitative predictions of peptide-MHC binding by a ‘query by committee’ Artificial Neural Network Approach,” *Tissue Antigens*, 62(5), pp. 378–384. Available at: <https://doi.org/10.1034/j.1399-0039.2003.00112.x>.
30. NetOglyc 4.0. Available at: <https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0>
31. NetNglyc. Available at: <https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0>
32. NetPhos 3.1. Available at: <https://services.healthtech.dtu.dk/service.php?NetPhos-3.1>

33. Ip, P., Nijman, H. and Daemen, T. (2015) “Epitope prediction assays combined with validation assays strongly narrows down putative cytotoxic T lymphocyte epitopes,” *Vaccines*, 3(2), pp. 203–220. Available at: <https://doi.org/10.3390/vaccines3020203>.
34. Bar-Haim, E. *et al.* (2004) “Mage-A8 overexpression in transitional cell carcinoma of the bladder: Identification of two tumour-associated antigen peptides,” *British Journal of Cancer*, 91(2), pp. 398–407. Available at: <https://doi.org/10.1038/sj.bjc.6601968>.
35. Zhang, Y. *et al.* (2021) “Sage1: A potential target antigen for lung cancer T-cell immunotherapy,” *Molecular Cancer Therapeutics*, 20(11), pp. 2302–2313. Available at: <https://doi.org/10.1158/1535-7163.mct-21-0203>.
36. Chen, Y.-T. *et al.* (2009) “Cancer/testis antigen CT45: Analysis of mrna and protein expression in human cancer,” *International Journal of Cancer*, 124(12), pp. 2893–2898. Available at: <https://doi.org/10.1002/ijc.24296>.
37. Signori, E. *et al.* (2010) “DNA vaccination strategies for anti-tumour effective gene therapy protocols,” *Cancer Immunology, Immunotherapy*, 59(10), pp. 1583–1591. Available at: <https://doi.org/10.1007/s00262-010-0853-x>.
38. Lu, J. *et al.* (2004) “Multiepitope trojan antigen peptide vaccines for the induction of antitumor CTL and th immune responses,” *The Journal of Immunology*, 172(7), pp. 4575–4582. Available at: <https://doi.org/10.4049/jimmunol.172.7.4575>.
39. *Phobius*. Available at: <https://phobius.sbc.su.se/index.html>.
40. *RNA folding form v2.3*. Available at: <http://www.mfold.org/mfold/applications/rna-folding-form-v2.php>
41. *A protein secondary structure prediction server* (no date) *JPred: A Protein Secondary Structure Prediction Server*. Available at: <https://www.compbio.dundee.ac.uk/jpred/>.
42. Robetta web Server https://robetta.bakerlab.org/login.php?next_url=%2Fsubmit.php
43. *Swiss-PdbViewer . Swiss PDB Viewer - Home*. Available at: <https://spdbv.unil.ch/>.

44. *PyMOL is a user-sponsored molecular visualization system on an open-source foundation, maintained and distributed by Schrödinger. We are happy to introduce PyMOL 2.5!* PyMOL. Available at: <https://pymol.org/2/>.
45. Larsen, M.V. *et al.* (2007) “Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction,” *BMC Bioinformatics*, 8(1). Available at: <https://doi.org/10.1186/1471-2105-8-424>.
46. Bhasin, M. and Raghava, G.P.S. (2004) “Prediction of CTL epitopes using QM, SVM and ANN techniques,” *Vaccine*, 22(23-24), pp. 3195–3204. Available at: <https://doi.org/10.1016/j.vaccine.2004.02.005>.
47. Julenius, K. *et al.* (2004) “Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites,” *Glycobiology*, 15(2), pp. 153–164. Available at: <https://doi.org/10.1093/glycob/cwh151>.
48. Buonaguro, L. and Tagliamonte, M. (2020) “Selecting target antigens for cancer vaccine development,” *Vaccines*, 8(4), p. 615. Available at: <https://doi.org/10.3390/vaccines8040615>.
49. Dutoit, V. *et al.* (2002) “Multiepitope CD8+ T cell response to a NY-ESO-1 peptide vaccine results in imprecise tumor targeting,” *Journal of Clinical Investigation*, 110(12), pp. 1813–1822. Available at: <https://doi.org/10.1172/jci16428>.
50. QIN, H. *et al.* (2005) “Specific antitumor immune response induced by a novel DNA vaccine composed of multiple CTL and T helper cell epitopes of prostate cancer associated antigens,” *Immunology Letters*, 99(1), pp. 85–93. Available at: <https://doi.org/10.1016/j.imlet.2005.01.006>.
51. Mateo, L. *et al.* (1999) “An HLA-A2 polyepitope vaccine for melanoma immunotherapy,” *The Journal of Immunology*, 163(7), pp. 4058–4063. Available at: <https://doi.org/10.4049/jimmunol.163.7.4058>.
52. Doan, T. *et al.* (2004) “A polytope DNA vaccine elicits multiple effector and memory CTL responses and protects against human papillomavirus 16 E7-expressing tumour,”

Cancer Immunology, Immunotherapy, 54(2), pp. 157–171. Available at: <https://doi.org/10.1007/s00262-004-0544-6>.

53. Amani, J. *et al.* (2010) “Immunogenic properties of chimeric protein from ESPA, EAE and TIR genes of *Escherichia coli* O157:H7,” *Vaccine*, 28(42), pp. 6923–6929. Available at: <https://doi.org/10.1016/j.vaccine.2010.07.061>.

54. Tartaglia, G.G. *et al.* (2009) “A relationship between mRNA expression levels and protein solubility in *E. coli*,” *Journal of Molecular Biology*, 388(2), pp. 381–389. Available at: <https://doi.org/10.1016/j.jmb.2009.03.002>.

55. Wilson, J.H. and Hunt, T. (2002) *Molecular biology of the cell, 4th edition: A problems approach*. New York: Garland Science.

56. Eggers, M. *et al.* (1995) “The cleavage preference of the proteasome governs the yield of antigenic peptides,” *Journal of Experimental Medicine*, 182(6), pp. 1865–1870. Available at: <https://doi.org/10.1084/jem.182.6.1865>.