

Computational Design of a Multi-Epitope Merkel Cell Polyomavirus Vaccine

Taarun Ganesh

taarun.ganesh@gmail.com

Thomas Jefferson High School for Science and Technology

Research Article

Keywords: immunoinformatics, Merkel cell carcinoma, vaccine, computational

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11516417/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Computational Design of a Multi-Epitope Merkel Cell Polyomavirus Vaccine

Authors: Taarun Ganesh¹

¹Thomas Jefferson High School for Science and Technology, Alexandria, Virginia, USA

Abstract

Merkel cell carcinoma (MCC) is an aggressive neuroendocrine skin cancer caused primarily by Merkel cell polyomavirus (MCV) and typically presents as a painless skin tumor. No vaccine is currently approved for MCC. In this study, I used computational tools to design a multi-epitope vaccine intended to elicit an effective immune response against MCV. B and T cell epitopes were predicted from the key antigenic proteins (VP1, large T antigen, and small T antigen) using the IEDB server, and 22 peptides were selected based on their predicted antigenicity, allergenicity, and toxicity. The selected epitopes were joined with linkers to form a single fusion protein (43.009 kDa) designed for expression in *E. coli*. Molecular docking suggested a plausible interaction between the vaccine candidate and human toll-like receptor 4 (TLR4), and immune simulation predicted antibody and helper T-cell responses. To support efficient expression in the bacterial host, the coding sequence was codon-optimized according to *E. coli* codon preferences. This computational pipeline may enable rapid, early-stage epitope design. However, *in vitro* and *in vivo* studies will be needed to assess expression, stability, and immunogenicity.

Introduction

Merkel cell carcinoma (MCC) is a rare but aggressive skin cancer that predominantly affects elderly and immunocompromised individuals (Silling et al., 2022). It is a highly

aggressive skin cancer with a higher case-fatality rate than melanoma (Becker et al., 2017; Schadendorf et al., 2017). MCC is most frequently found in Caucasian adults with sun exposure, and while its incidence is low overall (0.7 per 100,000 individuals in 2013), from 2000 to 2013 there was a significant increase of approximately 95% (Paulson et al., 2018).

MCC is largely attributable to Merkel cell polyomavirus (MCV), a small circular non-enveloped DNA virus that is classified within the polyomavirus family (Liu et al., 2016; Moens et al., 2017). MCV is detected in approximately 43% to 100% of MCC tumors, making it an important agent in this malignancy (National Comprehensive Cancer Network [NCCN], 2025). The viral genome is organized into an early region, which encodes the oncoproteins large T antigen (LT) and small T antigen (sT), and a late region, which encodes the capsid proteins VP1 and VP2 (Spurgeon & Lambert, 2013; Schowalter & Buck, 2013). VP1 is the major capsid protein that forms the virion surface, whereas VP2 is a minor capsid protein that facilitates a post attachment step of infectious entry in some cell types (Schowalter & Buck, 2013). In MCV-positive MCC cell lines, continued expression of viral T antigens is required for cell proliferation and survival which is consistent with the oncogenic role of the protein (Houben et al., 2010). Finally, MCV DNA is clonally integrated into the genome of 80% of MCV-positive MCC tumors, which provides additional evidence for its contribution to MCC pathogenesis (Feng et al., 2008).

To my knowledge, no vaccine against MCV is currently approved. Previous computational studies have explored epitope-based vaccine design against MCV, identifying the capsid proteins VP1 and VP2 as promising targets (Almansour, 2022). Epitopes are the minimal antigenic regions that can provoke a specific immune response (Shi et al., 2015). Unlike traditional vaccines, which use the whole pathogen, epitope-based vaccines use short antigenic

domains of proteins. They are therefore considered safer than live-attenuated vaccines and easier to produce (Shi et al., 2015; Sette & Fikes, 2003). Such vaccines stimulate a targeted immune response against selected epitopes and have been explored for cancer (Abd-Aziz & Poh, 2022) and viral diseases (Moin et al., 2023). Advances in computational biology now allow B and T cell epitopes to be predicted efficiently (Patronov & Doytchinova, 2013), and the predicted epitopes can be refined through multiple rounds of computational screening. In this study, I screened the entire MCV proteome and identified epitopes from three key antigenic proteins: VP1, and large T antigen and small T antigen, as VP2 and VP3 were not recognized as antigenic proteins. Using various bioinformatics tools, I then designed a multi-epitope vaccine candidate against MCV.

Effective antigen presentation is important for eliciting a strong immune response, and epitope-based vaccines often depend on adjuvants or delivery platforms to achieve it (Skwarczynski & Toth, 2016). First, I designed the multi-epitope fusion protein for expression from a plasmid in *E. coli*.

Methods

Sequence retrieval: The protein sequences of MCV were retrieved from the NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein>) using the RefSeq accession numbers YP_009111420 (major capsid protein VP1), YP_009111421 (large T antigen), and YP_009111422 (small T antigen).

Antigenicity screening: To select proteins suitable for epitope prediction, the retrieved protein sequences were screened for antigenicity using the VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova & Flower, 2007). The virus model

was selected with the default threshold of 0.4, and proteins scoring above this threshold were considered probable antigens and retained for epitope prediction.

Prediction of MHC Class I, Class II, and B cell epitopes: B cell and T cell epitopes were predicted using the Immune Epitope Database (IEDB) analysis resource (<https://www.iedb.org/>) (Vita et al., 2019). T cell receptors recognize peptides presented by major histocompatibility complex (MHC) molecules. MHC class I molecules present peptides to CD8⁺ T cells, whereas MHC class II molecules present peptides to CD4⁺ T cells. I therefore used the IEDB tools for predicting MHC class I and class II binding peptides, as well as B cell epitopes. IEDB also stores and analyzes data on immune epitopes (Vita et al., 2019).

Antigenicity, allergenicity, toxicity, and human homology screening: The antigenicity of the predicted B cell and T cell epitopes was evaluated again using the VaxiJen v2.0 server (Doytchinova & Flower, 2007). Toxicity was predicted using the ToxinPred server (<https://webs.iitd.edu.in/raghava/toxinpred/design.php>), and allergenicity was predicted using the AllerTOP v2.0 server, and epitopes with allergenic potential were filtered out (Gupta et al., 2013; Dimitrov et al., 2014). Finally, human homology was assessed using the BLAST tool in the IEDB Epitope Conservancy Analysis server.

IFN, IL-10 prediction: The IFNepitope server (<http://crdd.osdd.net/raghava/ifnepitope/>) was used to predict epitopes could induce interferon-gamma (IFN- γ). The server classifies epitopes as IFN- γ -inducing or non-inducing using a model built from datasets of inducing and non-inducing peptides (Dhanda et al., 2013). Additionally, the IL-10Pred (<http://crdd.osdd.net/raghava/IL-10pred/>) server was used to predict which epitopes could induce interleukin-10 (IL-10) (Nagpal et al., 2017).

93 **Population coverage:** The IEDB server (<http://tools.iedb.org/population/>) was utilized again to
94 analyze the most promising epitopes across various HLA alleles in different regions of the world
95 in order to determine the population coverage of my epitopes (Bui et al., 2006). Both the IEDB
96 MHC-I and MHC-II binding prediction servers (<http://tools.iedb.org/mhci/>,
97 <http://tools.iedb.org/mhcii/>) predicted the most promising HLA alleles, as the analysis requires
98 MHC-restricted alleles to be computed for each epitope.

99 **Vaccine construct design:** The selected CTL, HTL, and linear B-cell epitopes were joined into a
100 single fusion protein. The PADRE sequence (AKFVAAWTLKAAA) was placed at the N-
101 terminus as a universal helper T-cell epitope (Alexander et al., 1994; del Guercio et al., 1997).
102 Epitopes were connected with class-specific linkers: GGS between CTL epitopes, GP GPG
103 between HTL epitopes, and KK between B-cell epitopes (Livingston et al., 2002; Reddy Chichili
104 et al., 2013; Yano et al., 2005). The antigenicity of the final construct was predicted with
105 VaxiJen v2.0 using the 0.4 threshold (Doytchinova & Flower, 2007).

106 **Physicochemical properties and solubility:** The molecular weight, theoretical isoelectric point
107 (pI), instability index, aliphatic index, grand average of hydropathicity (GRAVY), and estimated
108 half-life of the construct were calculated with the ExPASy ProtParam tool
109 (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005). An instability index above 40
110 classifies a protein as unstable (Guruprasad et al., 1990), the aliphatic index is a measure of
111 thermostability (Ikai, 1980), GRAVY reflects overall hydrophathy (Kyte & Doolittle, 1982), and
112 half-life is estimated from the N-terminal residue using the N-end rule (Bachmair et al., 1986).
113 Intrinsic solubility was predicted with CamSol (Sormanni et al., 2015); scores above 1 indicate a
114 highly soluble protein and scores below -1 a poorly soluble one.

115 **Protein structure prediction and molecular docking:** I modeled the tertiary structure of the
116 construct using AlphaFold2 on the Neurosnap platform, a method that achieved accuracy
117 competitive with experimental structures in the CASP14 benchmark (Jumper et al., 2021). I then
118 performed protein-protein docking of the vaccine candidate with human toll-like receptor 4
119 (TLR4; PDB ID: 4G8A) using the ClusPro v2.0 server (<https://cluspro.bu.edu/login.php>). The
120 server performs rigid-body docking, generates a large number of candidate conformations,
121 clusters the lowest-energy conformations, and reports the size and energy of each cluster; lower
122 energies indicate more favorable predicted binding. ClusPro returned ten ranked models (cluster
123 centers), and I selected the model with the largest cluster size and the lowest energy (Kozakov et
124 al., 2017).

125 **Immune simulation:** Immune simulation was performed using the C-ImmSim server
126 (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>) which uses a position-specific scoring matrix (PSSM)-
127 based method to predict peptide-MHC binding and simulate the immune response to the vaccine
128 (Rapin et al., 2010). All parameters were set to their default values, except that the three
129 injections were administered at time steps 1, 84, and 170, and the number of simulation steps
130 was set to 1050.

131 ***In silico* cloning and vaccine insertion into plasmid:** For codon optimization, the Java Codon
132 Adaptation Tool (JCat) server (<http://www.jcat.de/>) was used to reverse-translate the protein
133 sequence and generate a codon-optimized sequence for expression in *E. coli*, and the
134 corresponding codon adaptation index (CAI) and GC content were calculated (Grote et al.,
135 2005). The ideal CAI value is 1.0, although a score greater than 0.8 is considered adequate, and a
136 GC content between 30% and 70% is considered favorable (Grote et al., 2005). For the *in silico*
137 cloning simulation, the pET-28b(+) vector, which confers kanamycin resistance and supports

high-level expression, was selected, and the optimized vaccine sequence was inserted between the *NheI* and *XhoI* restriction sites using SnapGene (<https://www.snapgene.com/>). The RNAfold server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) was used to predict the secondary structure of the mRNA (Gruber et al., 2008).

Results

Prediction of vaccine candidates

I screened the large T antigen, small T antigen, and VP1 for epitopes. Candidate B-cell and T-cell epitopes from these proteins were then screened for predicted antigenicity, allergenicity, toxicity, and non-homology to the human proteome, and the T-cell epitopes were additionally screened for IFN- γ and IL-10 induction. This screening yielded a final set of 13 MHC class I, 3 MHC class II, and 6 B-cell epitopes (Table 1).

Type	Epitope sequence	Antigenicity Score	Toxicity (SVM) Score
CD8+ T Cell	TSNWYTYTY	0.6999	-0.79
	VSGVNYHMF	0.5914	-1.03
	DLQGLVLDY	0.7124	-1.04
	NPYPVVNLI	1.1456	-0.69
	LTWGECFCY	0.5444	-0.31
	SMFDEVSTK	0.4178	-1.20
	GFPPTWESF	1.0481	-0.61
	LWFGFPPTW	0.7969	-0.86
	HSQSSSSGY	0.7108	-0.58
	RTDGTWEDL	0.4091	-0.62
	LSDYLSHAV	1.2013	-1.14
	TTSDKAIEL	0.5240	-0.80
	AEEASSSQF	0.7112	-0.53
CD4+ T Cell	QVEEVRIYEGSEQLP	0.6444	-1.76
	ASSASFTSTPPKPKK	0.5794	-0.50
	DFPIDLSDYLSHAVY	0.7916	-1.45
B Cell	SNLMPKVSGQPMEGKDNQV	0.5722	-1.48
	QSNQQDKGKAPLKGPKASQKESQT	0.5074	-0.96
	MGVNSPDLPTTSNWYTYTYDLQPKGSSPDQPIKENL	0.6699	-0.87
	RIYEGSEQLPGDPDIVRFLDKFGQEKTVYPKPSVAP	0.9642	-1.18
	QHCSLKTLLKQKNCL	0.5074	-0.04
	QDPLLNIIEEEGPETEETQDSGT	0.9576	-1.41

Table 1. Final epitopes included in the vaccine construct. Antigenicity scores were predicted with VaxiJen v2.0 (threshold 0.4), and toxicity was predicted with ToxinPred (a negative SVM score indicates a predicted non-toxic peptide). The construct contains 13 CD8⁺ T-cell, 3 CD4⁺ T-cell and 6 B-cell epitopes

Population coverage analysis of selected peptides

The selected epitopes were predicted to cover a large share of the global population. MHC class I epitopes covered 85.72% of the world population, and coverage was approximately 80% or higher in most of the regions analyzed. Sweden had the highest class I coverage (96.70%), followed by Germany (92.85%) and England (92.83%). MHC class II epitopes covered 24.18% of the world population, with higher coverage in some regions, reaching about 40% in Sweden and about 48% in Malaysia Oriental (**Fig. 1**). These results suggest that the epitope set could be relevant to a broad range of populations, particularly for MHC class I presentation.

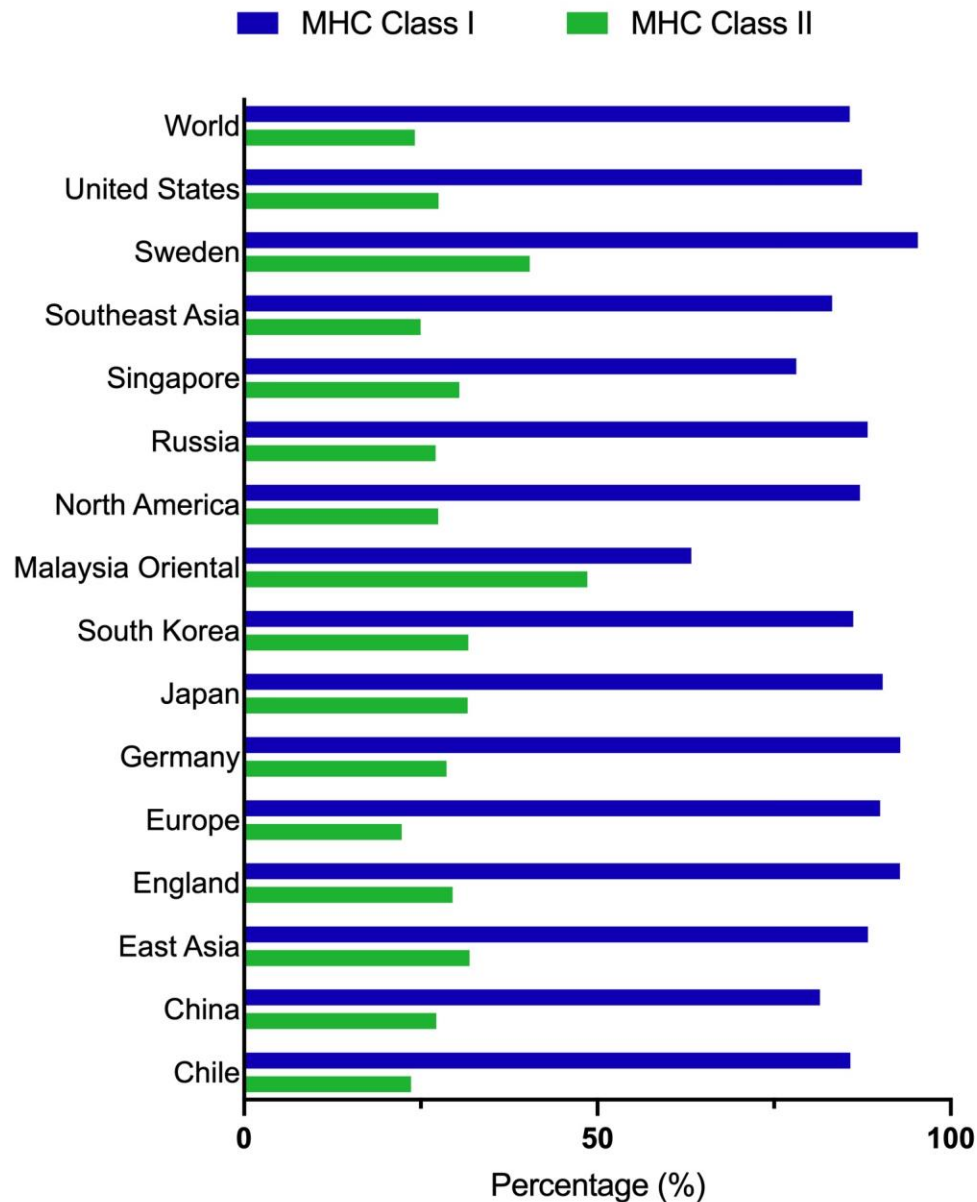


Figure 1. Population coverage of the selected T-cell epitopes. Predicted coverage of MHC class I and class II epitopes across selected regions of the world, calculated with the IEDB population coverage tool from the HLA alleles that bind each epitope.

Creation of a single fusion protein for *E. coli* expression

For expression in *E. coli*, linkers were added to connect all epitopes (Fig. 2), and the entire sequence was inserted into the pET-28b(+) plasmid. The PADRE sequence

(AKFVAAWTLKAAA) was incorporated at the N-terminus of the construct as a universal helper T-cell epitope because it binds multiple HLA-DR alleles (Alexander et al., 1994; del Guercio et al., 1997). A flexible GGGS linker was used between CTL epitopes, whereas a GP GPG linker was used between HTL epitopes, as this linker is widely used in multi-epitope fusion proteins to help preserve epitope immunogenicity and reduce junctional epitopes (Livingston et al., 2002). In addition, a KK linker was added between B-cell epitopes, as this bi-lysine linker is a cleavage site for cathepsin B and is commonly used in epitope-based vaccine designs (Li et al., 2015; Ayyagari et al., 2022).

A. Overall architecture of the multi-epitope vaccine construct



B. Epitope order and linkers

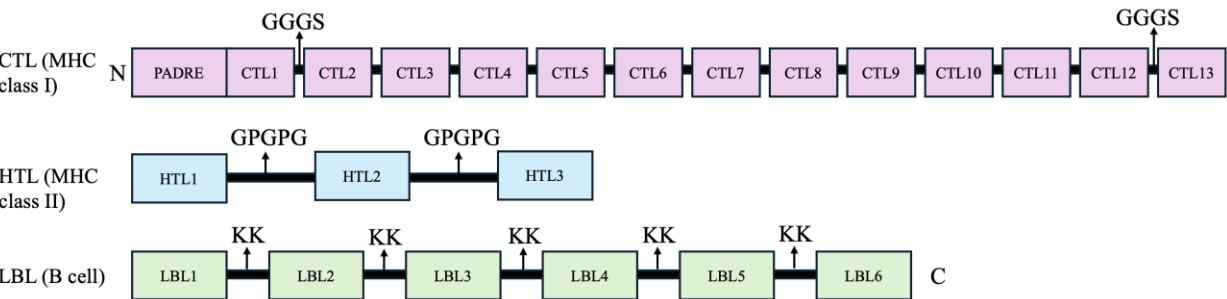


Figure 2. Schematic representation of the multi-epitope vaccine construct with linkers. (A)

Overall architecture from the N-terminus to the C-terminus: PADRE, 13 cytotoxic T lymphocyte (CTL; MHC class I) epitopes, 3 helper T lymphocyte (HTL; MHC class II) epitopes, and 6 linear B lymphocyte (LBL) epitopes. (B) Epitope order and linkers: GGGS between CTL epitopes, GP GPG between HTL epitopes, and KK between LBL epitopes.

Properties of the final vaccine construct

The final vaccine construct had a VaxiJen v2.0 score of 0.6417, above the 0.4

threshold, indicating that it is predicted to be antigenic. Its molecular weight was 43.009 kDa, which is suitable for expression in *E. coli*. The theoretical pI was 5.55 (acidic), and the aliphatic index was 50.29. The estimated half-life was approximately 4.4 hours in mammalian reticulocytes, more than 10 hours in *E. coli*, and more than 20 hours in yeast. Finally, the CamSol intrinsic solubility score was 1.407, indicating that the construct is predicted to be soluble (Table 2).

Final vaccine candidate's properties	
Number of Amino Acids	409 aa
Molecular Weight	43009.37 Da
Antigenicity Score	0.6417
Intrinsic Solubility Score	1.407
Estimated Half-Life (mammalian reticulocytes, <i>in vitro</i>)	4.4 hours
Estimated Half-Life (yeast, <i>in vivo</i>)	> 20 hours
Estimated Half-Life (<i>E. coli</i> , <i>in vivo</i>)	> 10 hours
Aliphatic Index	50.29
Grand Average of Hydropathicity (GRAVY)	-0.74

Table 2. Predicted physicochemical properties of the final vaccine construct.

Properties were calculated with ExPASy ProtParam unless stated otherwise. Antigenicity was predicted with VaxiJen v2.0 (threshold 0.4), intrinsic solubility with CamSol.

Protein structure prediction and molecular docking:

The three-dimensional structure of the vaccine construct was predicted with AlphaFold2 on the Neurosnap platform. To simulate the interaction between the construct and a toll-like receptor, I performed molecular docking of the vaccine construct with TLR4 using the ClusPro v2.0 server, which generated multiple models. I selected the TLR4–vaccine complex model from

the cluster with the largest size and the lowest energy score (−1744.5; Fig. 3). Lower ClusPro energy scores generally indicate a more favorable predicted interaction, although they are not a direct measure of binding affinity.

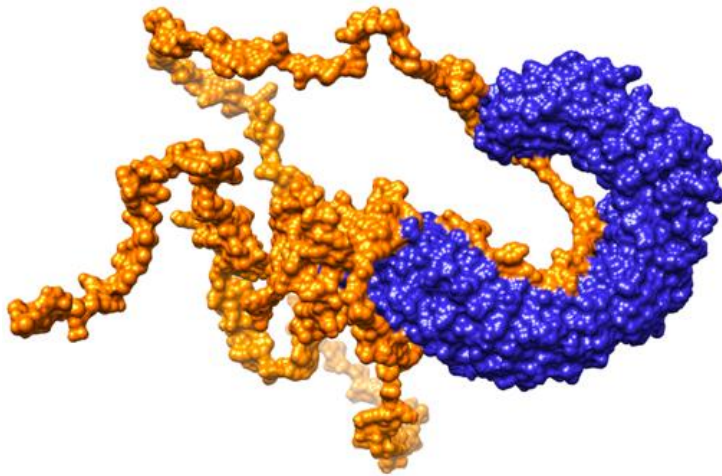


Figure 3. Docked complex of TLR4 and the vaccine construct. The vaccine construct is shown in orange and the TLR4 structure in blue

Immune simulation

The C-ImmSim simulation predicted a vaccine-induced immune response over 350 days after three injections. Cytokine concentrations rose after each dose and returned to baseline within about 100 days. IFN- γ showed the largest peaks, at roughly 400,000 ng/mL after each injection, and IL-2 peaked at about 300,000–500,000 ng/mL, with smaller peaks of other cytokines, including IL-10 (Fig. 4a). Antigen levels fell rapidly after each injection, while antibody titers increased with each dose and were highest after the third. The combined IgM + IgG titer reached roughly 165,000, with IgM and IgG1 as the main components, and titers remained above baseline until the end of the simulation (Fig. 4b). The helper T-cell population

expanded after each injection, peaking at about 10,800 cells after the third dose, and then declined to a lower plateau. Memory helper T cells rose to about 1,700 cells and persisted, decreasing slowly to about 600 cells by day 350 (Fig. 4c). In contrast, the cytotoxic T-cell population fluctuated around its baseline of about 1,100 cells with no sustained expansion after vaccination, and no memory cytotoxic T cells were predicted (Fig. 4d). Overall, the simulation predicted antibody and helper T-cell memory responses to the vaccine construct but no cytotoxic T-cell expansion.

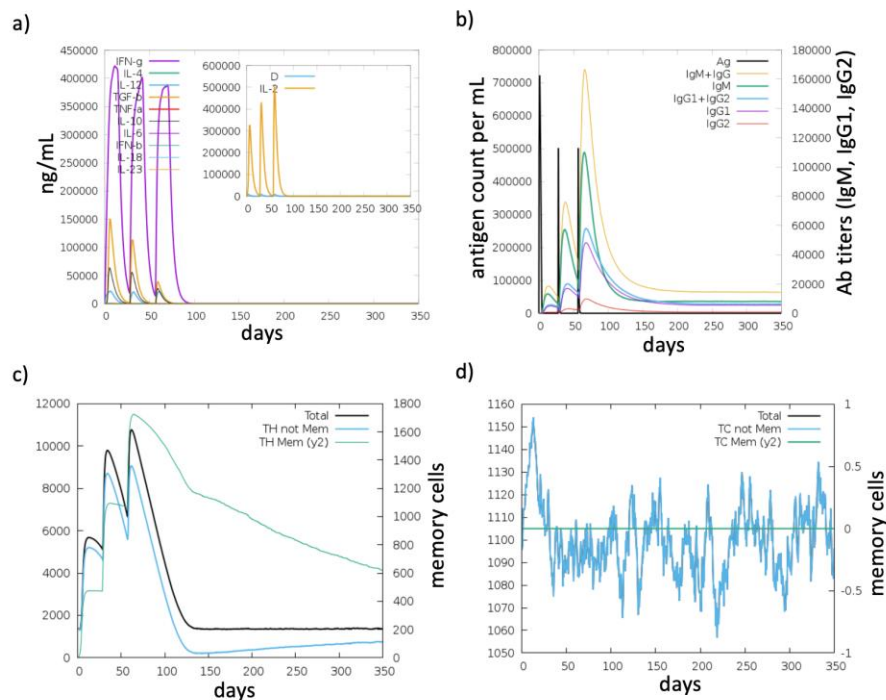


Figure 4. Immune response predicted by the C-ImmSim server. (a) Predicted concentrations of cytokines and interleukins after vaccination; the inset shows IL-2 and the danger signal (D). (b) Antigen count and antibody titers (IgM, IgG1, IgG2, and their combinations) after three vaccine doses. (c) Helper T-cell population by state: total, non-memory, and memory cells. (d) Cytotoxic T-cell population by state: total, non-memory, and memory cells.

Codon optimization and *in silico* cloning

Codon optimization of the vaccine sequence was performed with the JCat server for expression in *E. coli* (strain K12). The optimized sequence had a codon adaptation index (CAI) of 0.956 and a GC content of 69.11%, and both values were within the recommended ranges. The optimized sequence was then inserted into the pET-28b(+) vector between the NheI and XhoI restriction sites (Fig. 5). The mRNA secondary structure was subsequently predicted with the RNAfold server, which gave a minimum free energy (MFE) of -495.96 kcal/mol.

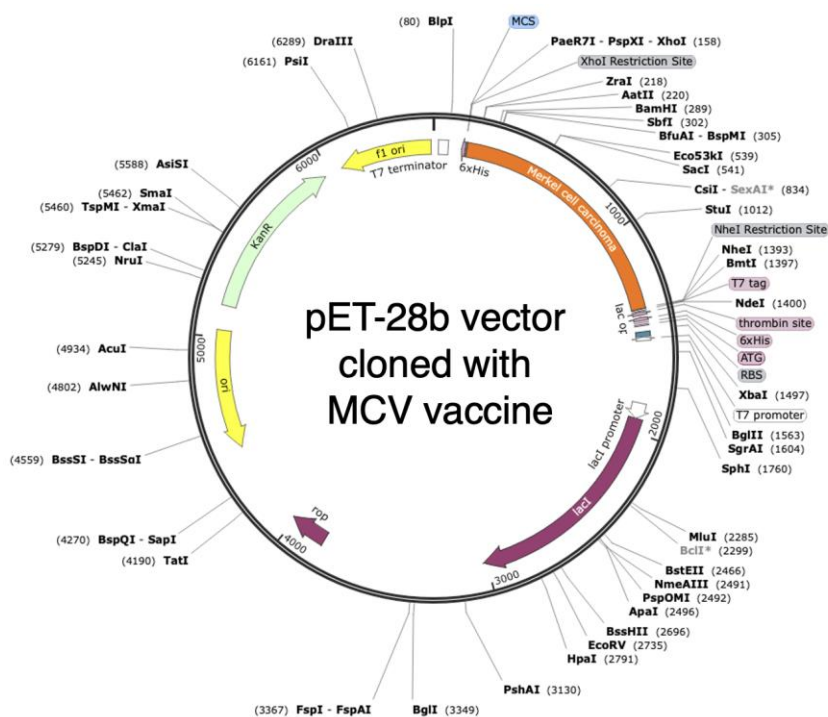


Figure 5. *In silico* cloning of the vaccine construct into the pET-28b(+) expression vector.

The codon-optimized sequence of the multi-epitope vaccine (orange) was inserted between the NheI and XhoI restriction sites. The map shows the vector backbone, including the T7 promoter

and terminator, the N-terminal and C-terminal 6×His tags, the lacI gene, the kanamycin resistance gene (KanR), and the fl origin of replication. The map was generated with SnapGene.

Discussion

Merkel cell carcinoma (MCC) is a rare and aggressive skin cancer, and in most cases in the northern hemisphere it is associated with Merkel cell polyomavirus (MCV). No vaccine against MCV is currently approved. I therefore used an immunoinformatics pipeline to design a multi-epitope vaccine candidate against it. I drew epitopes from the capsid protein VP1 and from the early-region large T and small T antigens. The T antigens are particularly relevant because MCV-positive MCC cell lines depend on T antigen expression for growth (Houben et al., 2010). Targeting proteins that tumors keep expressing could, in principle, allow T-cell responses to act against both infected and transformed cells. Including VP1 added a late-region target, so the construct is not limited to a single stage of the viral life cycle.

Epitope selection prioritized predicted immunogenicity and safety. Candidates were retained only if they were predicted to be antigenic, non-allergenic, and non-toxic, and if they did not match human proteins, which reduces the risk of cross-reactivity and autoimmunity. The final construct contained 13 cytotoxic T lymphocyte (CTL), 3 helper T lymphocyte (HTL), and 6 B-cell epitopes. The imbalance between these groups is a limitation, because only a few HTL epitopes passed the filters, and robust CD8⁺ and antibody responses generally depend on CD4⁺ T-cell help. Some selected epitopes also overlap or share residues, and several B-cell epitopes are long. I kept them because they met my criteria, but these features could affect folding and linker accessibility. Alternative selections could give different results.

The design of the construct follows the strategy used in my earlier work on Rocio virus (Ganesh & Mouroso, 2026). The universal helper T-cell epitope PADRE (AKFVAAWTLKAAA) was placed at the N-terminus to recruit CD4⁺ T-cell help, because it binds a broad range of HLA-DR molecules (Alexander et al., 1994). Epitope classes were joined with class-specific linkers. The flexible, glycine-rich GGGS linker was placed between CTL epitopes to keep adjacent epitopes spatially separate (Reddy Chichili et al., 2013). The GPGPG linker joined the HTL epitopes to help preserve their individual immunogenicity and limit the formation of junctional epitopes (Livingston et al., 2002). The KK linker joined the B-cell epitopes, since it is a cleavage site for the lysosomal protease cathepsin B and may reduce junctional immunogenicity (Yano et al., 2005). Together, these elements are intended to keep each epitope accessible for processing and presentation. The construct could be improved further by incorporating an adjuvant, which would be a logical next step in its development.

The predicted physicochemical properties suggest the construct could be produced in an *E. coli* expression system. The molecular weight is 43.009 kDa, the theoretical pI is 5.55, and the negative GRAVY value of -0.74 indicates a hydrophilic protein. The CamSol intrinsic solubility score of 1.407 suggests good solubility. However, the instability index of 49.20 is above the threshold of 40, which classifies the construct as unstable, and the aliphatic index of 50.29 is low, which points to limited thermostability (Ikai, 1980). Storage and handling conditions would therefore need to be tested, and sequence modifications could be considered. The predicted half-life is short in mammalian reticulocytes (about 4.4 hours) but longer in *E. coli*, which is acceptable for recombinant expression. Codon optimization gave a CAI of 0.956 and a GC content of 69.11%, which is at the upper end of the usual range, and the predicted mRNA structure had a very low minimum free energy (-495.96 kcal/mol). Strong mRNA secondary

structure could reduce translation efficiency, so expression levels remain uncertain. All of these values are computational predictions, and none replaces testing for expression, yield, and folding.

Docking with TLR4 and the immune simulation were used to generate hypotheses. The best ClusPro model had an energy score of -1744.5 . This score ranks models and is not a binding affinity, so the result shows only that a complex is geometrically plausible, not that the construct would signal through TLR4. The C-ImmSim server predicted antibody production and a helper T-cell response with persistent memory, but no clear expansion of cytotoxic T cells, which is a limitation for a vaccine intended to act on tumor cells. The simulator is a simplified, generic model, not including the tumor microenvironment, the immune evasion of MCC, or the real kinetics of an MCV infection.

Here, population coverage is another limitation of my study. The MHC class I epitopes were predicted to cover 85.72% of the world population, but coverage of the HTL epitopes was only 24.18%, which limits how widely the vaccine could work across HLA types. Adding HTL epitopes with broader HLA binding would improve this. This study was based entirely on computational methods. Molecular dynamics or peptide–MHC docking could add some computational support. The main next steps are expression and purification of the construct in *E. coli*, tests of its structure and stability, and tests of antibody and T-cell responses in an animal model.

In conclusion, I designed an in silico multi-epitope vaccine candidate against MCV that targets VP1 and the large and small T antigens. It showed favorable predicted antigenicity and safety, along with clear limitations in stability, cytotoxic T-cell response, and HTL coverage. It is a starting point for experimental work and does not yet show protection.

314

315 **References**

316 Silling, S., Kreuter, A., Gambichler, T., Meyer, T., Stockfleth, E., & Wieland, U. (2022).

317 Epidemiology of Merkel Cell Polyomavirus Infection and Merkel Cell Carcinoma. *Cancers*,

318 *14*(24), 6176. <https://doi.org/10.3390/cancers14246176>

319

320 Paulson, K. G., Park, S. Y., Vandeven, N. A., Lachance, K., Thomas, H., Chapuis, A. G., Harms,

321 K. L., Thompson, J. A., Bhatia, S., Stang, A., & Nghiem, P. (2018). Merkel cell carcinoma:

322 Current US incidence and projected increases based on changing demographics. *Journal of the*

323 *American Academy of Dermatology*, *78*(3), 457-463.e2.

324 <https://doi.org/10.1016/j.jaad.2017.10.028>

325

326 Liu, W., MacDonald, M., & You, J. (2016). Merkel cell polyomavirus infection and Merkel cell

327 carcinoma. *Current opinion in virology*, *20*, 20-27.

328

329 Spurgeon, M. E., & Lambert, P. F. (2013). Merkel cell polyomavirus: a newly discovered human

330 virus with oncogenic potential. *Virology*, *435*(1), 118-130.

331

332 Schowalter, R. M., & Buck, C. B. (2013). The Merkel cell polyomavirus minor capsid protein.

333 PLoS pathogens, *9*(8), e1003558.

334

335 Houben, R., Shuda, M., Weinkam, R., Schrama, D., Feng, H., Chang, Y., ... & Becker, J. C.
 336 (2010). Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of
 337 viral T antigens. *Journal of virology*, 84(14), 7064-7072.
 338
 339 Feng, H., Shuda, M., Chang, Y., & Moore, P. S. (2008). Clonal integration of a polyomavirus in
 340 human Merkel cell carcinoma. *Science*, 319(5866), 1096-1100.
 341
 342 Almansour, N. M. (2022). Immunoinformatics-and bioinformatics-assisted computational
 343 designing of a novel multiepitopes vaccine against cancer-causing Merkel cell
 344 polyomavirus. *Frontiers in Microbiology*, 13, 929669.
 345
 346 Shi, J., Zhang, J., Li, S., Sun, J., Teng, Y., Wu, M., ... & Hu, Y. (2015). Epitope-based vaccine
 347 target screening against highly pathogenic MERS-CoV: an in silico approach applied to
 348 emerging infectious diseases. *PloS one*, 10(12), e0144475.
 349
 350 Sette, A., & Fikes, J. (2003). Epitope-based vaccines: an update on epitope identification,
 351 vaccine design and delivery. *Current opinion in immunology*, 15(4), 461-470.
 352
 353 Abd-Aziz, N., & Poh, C. L. (2022). Development of peptide-based vaccines for cancer. *Journal*
 354 *of Oncology*, 2022.
 355

356 Moin, A. T., Ullah, M. A., Patil, R. B., Faruqui, N. A., Araf, Y., Das, S., ... & Islam, S. (2023). A
 357 computational approach to design a polyvalent vaccine against human respiratory syncytial virus.
 358 Scientific Reports, 13(1), 1-20.

359

360 Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics.
 361 Open biology, 3(1), 120139.

362

363 Doytchinova, I. A. & Flower, D. R. (2007). VaxiJen: A server for prediction of protective
 364 antigens, tumour antigens and subunit vaccines. BMC Bioinform. 8, 4.
 365 <https://doi.org/10.1186/1471-2105-8-4>

366

367 Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... & Peters, B.
 368 (2019). The immune epitope database (IEDB): 2018 update. *Nucleic acids research*, 47(D1),
 369 D339-D343.

370

371 Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v. 2—a server for in
 372 silico prediction of allergens. Journal of molecular modeling, 20(6), 2278.

373

374 Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC
 375 class-II binders. *Biology direct*, 8(1), 1-15.

376

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific reports*, 7(1), 42851.

Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics*, 7(1), 1-5.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., ... & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature protocols*, 12(2), 255-278.

Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., & Sette, A. (2002). A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology*, 168(11), 5499-5506.

Li, X., Guo, L., Kong, M., Su, X., Yang, D., Zou, M., Liu, Y., & Lu, L. (2015). Design and evaluation of a multi-epitope peptide of human metapneumovirus. *Intervirology*, 58(6), 403–412. 10.1159/000445059

400 Ayyagari, V. S., T C, V., K, A. P., & Srirama, K. (2022). Design of a multi-epitope-based
 401 vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of*
 402 *biomolecular structure & dynamics*, 40(7), 2963–2977.
 403 <https://doi.org/10.1080/07391102.2020.1850357>
 404
 405 del Guercio, M. F., Alexander, J., Kubo, R. T., Arrhenius, T., Maewal, A., Appella, E., ... &
 406 Sette, A. (1997). Potent immunogenic short linear peptide constructs composed of B cell
 407 epitopes and Pan DR T helper epitopes (PADRE) for antibody responses in vivo. *Vaccine*, 15(4),
 408 441-448.
 409
 410 Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., ... & Grey, H. M.
 411 (1994). Development of high potency universal DR-restricted helper epitopes by modification of
 412 high affinity DR-blocking peptides. *Immunity*, 1(9), 751-761.
 413
 414 Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of
 415 protein-protein interactions. *Protein science : a publication of the Protein Society*, 22(2), 153–
 416 167. <https://doi.org/10.1002/pro.2206>
 417
 418 Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., & Nisizawa, T. (2005).
 419 An ingenious design for peptide vaccines. *Vaccine*, 23(17-18), 2322-2326.
 420
 421 Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *The Journal of*
 422 *Biochemistry*, 88(6), 1895-1898.

423

424 Ganesh, T., & Mourosi, J. T. (2026, August 21). In silico design of an epitope-based vaccine for

425 Rocio virus using phage display and E. coli expression system. Journal of Emerging

426 Investigators. <https://doi.org/10.59720/24-279>

427

428 Becker, J. C., Stang, A., DeCaprio, J. A., Cerroni, L., Lebbé, C., Veness, M., & Nghiem, P.

429 (2017). Merkel cell carcinoma. Nature reviews Disease primers, 3(1), 17077.

430

431 Schadendorf, D., Lebbé, C., Zur Hausen, A., Avril, M. F., Hariharan, S., Bharmal, M., & Becker,

432 J. C. (2017). Merkel cell carcinoma: epidemiology, prognosis, therapy and unmet medical needs.

433 European journal of cancer, 71, 53-69.

434

435 Moens, U., Calvignac-Spencer, S., Lauber, C., Ramqvist, T., Feltkamp, M. C., Daugherty, M. D.,

436 ... & Ehlers, B. (2017). ICTV virus taxonomy profile: Polyomaviridae. J Gen Virol, 98(6), 1159-

437 1160.

438

439 National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology

440 (NCCN Guidelines®): Merkel Cell Carcinoma. Version 2.2026. October 24, 2025. Accessed

441 October 18, 2026. https://www.nccn.org/professionals/physician_gls/pdf/mcc.pdf

442

443 Skwarczynski, M., & Toth, I. (2016). Peptide-based synthetic vaccines. Chemical Science, 7(2),

444 842-854. <https://doi.org/10.1039/c5sc03892h>

445

446 Bachmair, A., Finley, D., & Varshavsky, A. (1986). In vivo half-life of a protein is a function of
 447 its amino-terminal residue. *Science*, 234(4773), 179-186.
 448
 449 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch,
 450 A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.),
 451 The Proteomics Protocols Handbook (pp. 571-607). Humana Press. [https://doi.org/10.1385/1-](https://doi.org/10.1385/1-59259-890-0:571)
 452 [59259-890-0:571](https://doi.org/10.1385/1-59259-890-0:571)
 453
 454 Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005).
 455 JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic*
 456 *acids research*, 33(suppl_2), W526-W531.
 457
 458 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The vienna
 459 RNA websuite. *Nucleic acids research*, 36(suppl_2), W70-W74.
 460
 461 Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery
 462 Consortium, & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides
 463 and proteins. *PLoS ONE*, 8(9), e73957. <https://doi.org/10.1371/journal.pone.0073957>
 464
 465 Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a
 466 protein and its dipeptide composition: a novel approach for predicting in vivo stability of a
 467 protein from its primary sequence. *Protein Engineering*, 4(2), 155-161.
 468

469 Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a
470 protein. *Journal of Molecular Biology*, 157(1), 105-132.
471
472 Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets
473 bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune
474 system. *PLoS ONE*, 5(4), e9862.
475
476 Sormanni, P., Aprile, F. A., & Vendruscolo, M. (2015). The CamSol method of rational design
477 of protein mutants with enhanced solubility. *Journal of Molecular Biology*, 427(2), 478-490.
478 <https://doi.org/10.1016/j.jmb.2014.09.026>