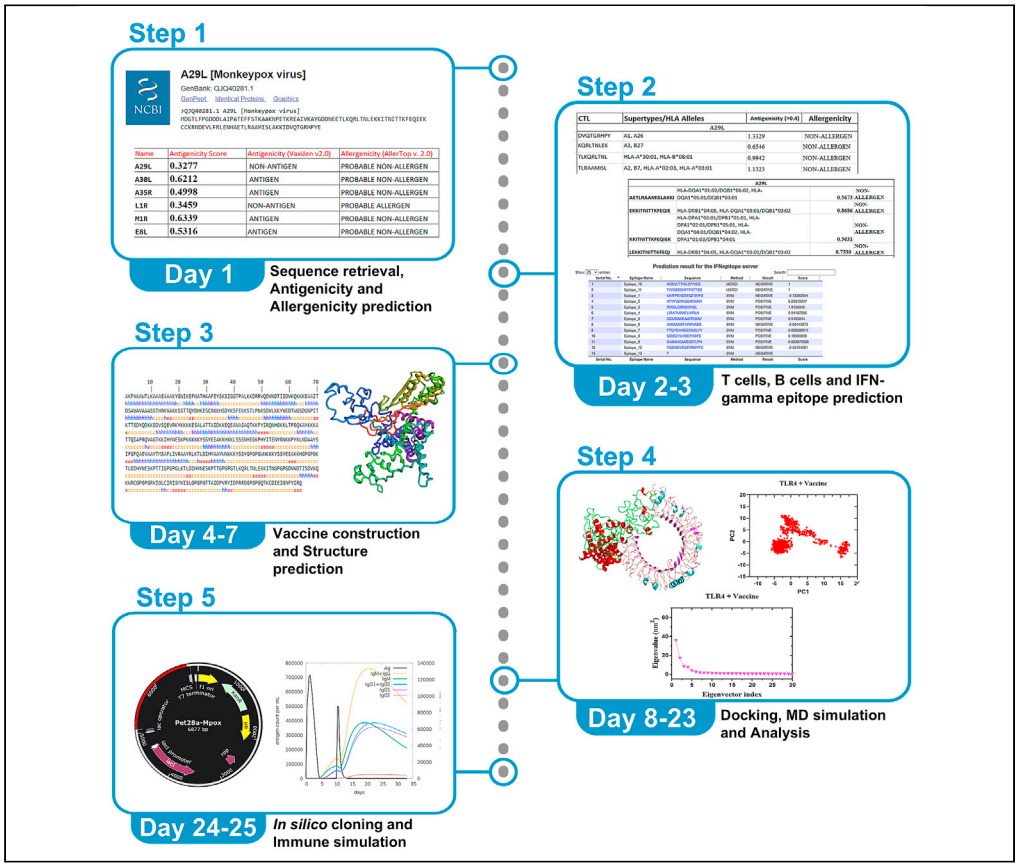
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for designing a peptide-based multi-epitope vaccine targeting monkeypox using reverse vaccine technology



Reverse vaccine technology, supported by advancements in immunoinformatics, facilitates the development of multi-epitope vaccines for rapidly evolving pathogens, thereby strengthening the immune defense. Here, we present a protocol for a peptide-based multi-epitope vaccine targeting monkeypox virus (MPXV) using an open-source approach. We describe steps for evaluating physicochemical properties and allergenicity. We then detail procedures for validating pattern recognition receptor (PRR)-binding affinity and stable major histocompatibility complex (MHC) I/II presentation. Molecular dynamics (MD) simulations confirm immune receptor interactions, enhancing vaccine stability.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Amit Kumar, Garima Nagar, Prithwik Bhowmik, ..., Rakesh Kundu, Indrakant Kumar Singh, Tanmay Majumdar

majumdart@nii.ac.in

Highlights

Steps for predicting specific epitopes for B cells, T cells, and IFN-gamma

Instructions for the construction of 3D vaccine structure and its validation

Steps for molecular docking and MD simulation with PRRs

Guidelines on *in silico* cloning and immune simulation

Kumar et al., STAR Protocols 6, 103671
March 21, 2025
https://doi.org/10.1016/j.xpro.2025.103671

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for designing a peptide-based multi-epitope vaccine targeting monkeypox using reverse vaccine technology

Amit Kumar,[1,5] Garima Nagar,[2,5] Prithwik Bhowmik,[2,5] Geetika Kumari,[1,5] Rasmiranjan Muduli,[1] Mayami Das,[1] Pritha Chakraborty,[3] Anupamjeet Kaur,[1] Kumari Shikha,[1] Suprabhat Mukherjee,[3] Rakesh Kundu,[4] Indrakant Kumar Singh,[2] and Tanmay Majumdar[1,6,7,*]

[1]National Institute of Immunology, New Delhi, India

[2]Department of Zoology, Deshbandhu College, University of Delhi, New Delhi, India

[3]Department of Animal Science, Kazi Nazrul University, Asansol, West Bengal, India

[4]Department of Zoology, Visva-Bharati University, Santiniketan, West Bengal, India

[5]These authors contributed equally

[6]Technical contact

[7]Lead contact

*Correspondence: majumdart@nii.ac.in
https://doi.org/10.1016/j.xpro.2025.103671

SUMMARY

Reverse vaccine technology, supported by advancements in immunoinformatics, facilitates the development of multi-epitope vaccines for rapidly evolving pathogens, thereby strengthening the immune defense. Here, we present a protocol for a peptide-based multi-epitope vaccine targeting monkeypox virus (MPXV) using an open-source approach. We describe steps for evaluating physicochemical properties and allergenicity. We then detail procedures for validating pattern recognition receptor (PRR)-binding affinity and stable major histocompatibility complex (MHC) I/II presentation. Molecular dynamics (MD) simulations confirm immune receptor interactions, enhancing vaccine stability.

For complete details on the use and execution of this protocol, please refer to Kaur et al.[1]

BEFORE YOU BEGIN

Reverse vaccine technology is a bioinformatics-driven approach for vaccine development that identifies immunogenic epitopes directly from a pathogen's genome/proteome. Simultaneous targeting of multiple epitopes triggers an elevated immunogenic response. The present protocol includes mapping of immune response predicting epitopes of MPXV that could foster long-lasting memory immunity against its challenge.

1. Construct the multi-epitope vaccine, its 3D structure and examination of physicochemical properties.
2. Molecular docking and MD simulation to predict the binding affinity of the vaccine with receptors (MHC I, MHC II, and TLRs).
3. Ensure robust immunological response of the multi-epitope vaccine upon booster dosing via *in silico* cloning and immune simulation and assess cellular immune activation and protective immunity through cytokine production and antibody generation.

In this study, we present an insightful approach to improve protection against MPXV infection. We believe the method outlined provides a clear guide for developing a multi-epitope vaccine, and it could also be applied to other infectious disease outbreaks.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Ubuntu 24.04.1 LTS (Linux operating system)	Canonical Ltd.	https://ubuntu.com/tutorials/install-ubuntu-desktop#1-overview
PyMOL	Schrödinger, Inc.	https://PyMOL.org/dokuwiki/doku.php?id=installation
GROMACS	Abraham et al.[2]	https://manual.gromacs.org/current/install-guide/index.html
LigPlot+	Laskowski et al.[3]	https://www.ebi.ac.uk/thornton-srv/software/LigPlus/install.html
GraphPad Prism 10	GraphPad	https://www.graphpad.com/
Snapgene	SnapGene	https://support.snapgene.com/hc/en-us/articles/10304161780628-Download-Install-and-Register-SnapGene
NCBI database	National Library of Medicine (NLM)	https://www.ncbi.nlm.nih.gov/
VaxiJen	Doytchinova et al.[4]	https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
AllerTOP v.2.0	Dimitrov et al.[5]	https://www.ddg-pharmfac.net/AllerTOP/
Expasy ProtParam	Gasteiger et al.[6]	https://web.expasy.org/protparam/
SOPMA	Geourjon et al.[7]	https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
PSIPRED	McGuffin et al.[8]	http://bioinf.cs.ucl.ac.uk/psipred/
NetCTL 1.2	Larsen et al.[9]	https://services.healthtech.dtu.dk/service.php?NetCTL-1.2
IEDB MHC I and II, Antibody Epitope Prediction, Population coverage, Conservancy analysis	Vita et al.[10]	http://tools.iedb.org/mhci/ http://tools.iedb.org/mhcii/, http://tools.iedb.org/bcell/, http://tools.iedb.org/population/ http://tools.iedb.org/conservancy/
NetMHCIIpan - 3.2	Jensen et al.[11]	https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-3.2
IFNepitope server	Dhanda et al.[12]	https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php
SoluProt	Hon et al.[13]	https://loschmidt.chemi.muni.cz/soluprot/
Robetta	Kim et al.[14]	https://robetta.bakerlab.org/queue.php
Galaxy refine	Heo et al.[15]	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
MolProbity	Williams et al.[16]	http://molprobity.biochem.duke.edu/
ClusPro 2.0	Kozakov et al.[17]	https://cluspro.org/help.php
PRODIGY	Xue et al.[18]	https://rascar.science.uu.nl/prodigy/
ToxinPred	Sharma et al.[19]	http://crdd.osdd.net/raghava/toxinpred/
EMBOSS Backtranseq	Rice et al.[20]	https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/
C-IMMSIM	Castiglione et al.[21]	
Other		
System requirements Intel Core i9-10850K 3.60 GHz, Windows 11 Pro, 64 GB RAM, 1 TB ROM, 64-bit OS.	N/A	N/A

STEP-BY-STEP METHOD DETAILS

Part 1: Retrieval of protein sequence

Timing: 30 min

This section enlists the steps for retrieving the protein sequence from the NCBI Database. The NCBI database is an open resource for accessing all genomic and biomedical data information related to advance science and health.

1. Access NCBI for free by using https://www.ncbi.nlm.nih.gov/ link (Figure 1A).
2. Click on the dropdown menu next to the search bar and select the "protein" option.

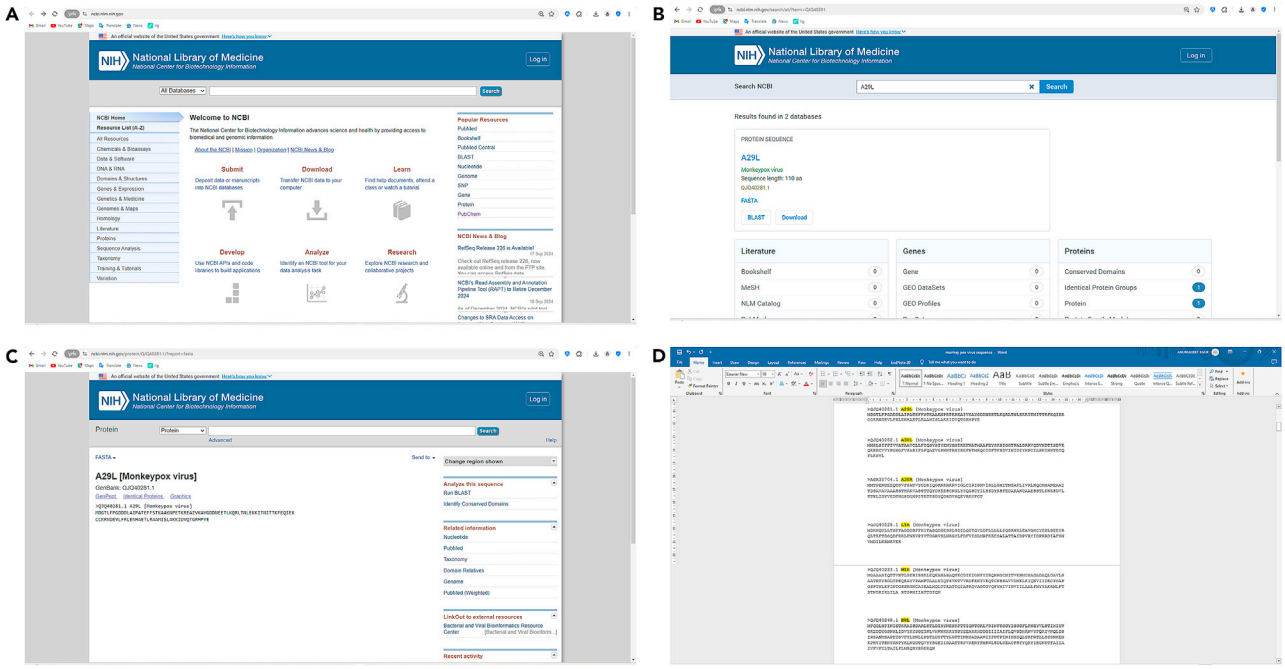


Figure 1. Retrieval of protein sequence using the NCBI database

Retrieved protein FASTA sequences from the NCBI database and compiled a library of target proteins with detailed descriptions.

(A) Home page of NCBI.

(B) Search protein of interest-A29L.

(C) Result of NCBI searched protein.

(D) Word document of protein sequences in FASTA format.

3. In the search, enter the protein name (A29L, A30L, A35R, L1R, M1R, or E8L) (Figure 1B).

   Note: If the desired results are not found, try adding the specific name of the organism next to the protein name to refine the search.

   ⚠ CRITICAL: Please ensure you have a thorough understanding of the proteins you plan to select for candidate vaccine design. In this case, we have chosen the glycoproteins of the mpox virus for consideration in multiepitope vaccine development.

4. Click on FASTA below protein name (Figure 1C).
5. Copy each sequence separately and paste them into a Word document in FASTA format (Figure 1D).

Part 2: Estimation of antigenicity, allergenicity, physiochemical, and secondary structure prediction of selected glycoproteins

Timing: 4 h

This section enlists the procedure and analysis of selected glycoproteins, assessing antigenic properties using VaxiJen v.2.0, allergenic properties with AllerTOP v.2.0, physicochemical characteristics via ProtParam, and secondary structure predictions using SOPMA and PSIPRED 4.0.

6. Open VaxiJen v.2.0 webserver for antigenicity prediction (Figure 2A).
7. Paste the individual protein sequence in the "Enter a PROTEIN sequence here" column.

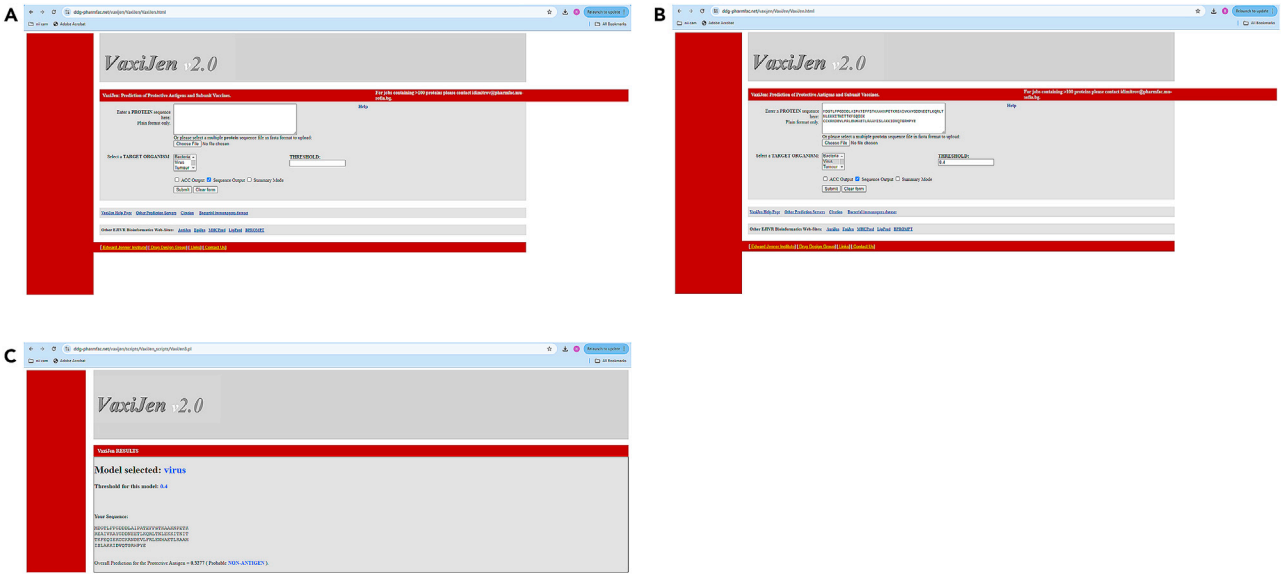


Figure 2. Analysis of antigenicity of selected glycoproteins using VaxiJen v.2.0

(A) Homepage of VaxiJen v2.0 webserver.

(B) Entered protein sequence and selection.

(C) Result page of searched protein.

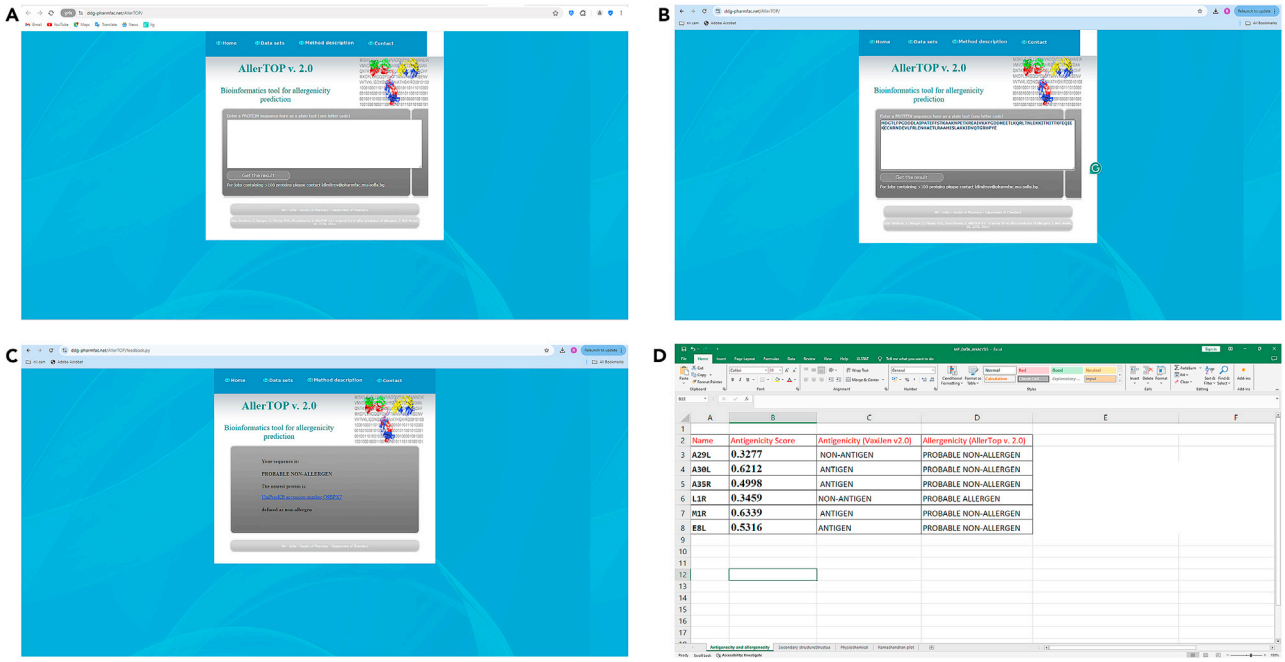
8. Choose the Virus option from the drop down "select a TARGET ORGANISM" section (Figure 2B).
9. Keep the threshold value as default 0.4 (to designate the glycoproteins as antigens with an antigenicity score > 0.4) and click "Submit" to display the result (Figure 2C).

   Note: Results will be displayed if the sequence entered is Probable ANTIGEN or NON-ANTIGEN and an antigenicity score. Score below 0.4 is non-antigenic and above 0.4 is antigenic in nature.

10. Open AllerTOP v.2.0. for allergenicity prediction (Figure 3A).
11. Paste individual protein sequence and click on "Get the results" option (Figure 3B).

    Note: Results will be displayed as "PROBABLE ALLERGEN" or "PROBABLE NON-ALLERGEN" (Figure 3C).

12. Prepare an Excel table as shown in Figure 3D comprising name of protein, antigenicity score, antigenicity and allergenicity status.
13. Examine the physiological parameters using the Expasy ProtParam webserver (Figure 4A).
14. Paste individual protein sequence in "Enter a protein sequence" column and click on "compute parameters" option to display results (Figures 4B and 4C).
15. Prepare an Excel table including size, molecular weight, theoretical pI, instability index, aliphatic index and GRAVY (Figure 4D).
16. Analyze the secondary structure components of the glycoproteins by using the SOPMA prediction model (Figure 5A).
17. Paste the protein sequence in "Paste a protein sequence below" column and click on submit to obtain result option keeping the default parameters (Figures 5B and 5C).
18. Prepare an Excel table compiling different secondary structure parameters generated by this webserver for all the proteins (Figure 5D).
19. Re-validate the protein secondary structure by entering the sequence on PSIPRED 4.0 webserver, click on the submit button (Figures 6A and 6B).



Name	Antigenicity Score	Antigenicity (VaxiJen v2.0)	Allergenicity (AllerTop v. 2.0)
A29L	0.3277	NON-ANTIGEN	PROBABLE NON-ALLERGEN
A30L	0.6212	ANTIGEN	PROBABLE NON-ALLERGEN
A35R	0.4998	ANTIGEN	PROBABLE NON-ALLERGEN
L1R	0.3459	NON-ANTIGEN	PROBABLE ALLERGEN
M1R	0.6339	ANTIGEN	PROBABLE NON-ALLERGEN
E8L	0.5316	ANTIGEN	PROBABLE NON-ALLERGEN

Figure 3. Analysis of allergenicity of selected glycoproteins using AllerTOP v.2.0

(A) Homepage of AllerTOP v.2.0 webserver.

(B) Entered protein sequence and submission.

(C) Result page of searched protein.

(D) Combined excel table containing antigenicity and allergenicity of selected glycoproteins.

20. Download the results generated (as sequence plot) from the right side by clicking the "Get zip file" option (Figure 6C).

Part 3: Prediction of T cell-specific epitopes: MHC I

Timing: 6 h

This section enlists the steps, and focuses on the prediction of T cell-specific epitopes via IEDB MHC I and NetCTL-1.2 servers, which are restricted to MHC class I molecules.

21. Go to MHC I binding predictions tool of Immune Epitope Database (IEDB) (Figure 7A).
22. Paste protein sequence in the dialog box. Click on the "Select HLA allele reference set" check-box (Figure 7B).

Note: The following list of alleles will be automatically selected- HLA-A*01:01, HLA-A*01:01, HLA-A*02:01, HLA-A*02:01, HLA-A*02:03, HLA-A*02:03, HLA-A*02:06, HLA-A*02:06, HLA-A*03:01, HLA-A*03:01, HLA-A*11:01, HLA-A*11:01, HLA-A*23:01, HLA-A*23:01, HLA-A*24:02, HLA-A*24:02, HLA-A*26:01, HLA-A*26:01, HLA-A*30:01, HLA-A*30:01, HLA-A*30:02, HLA-A*30:02, HLA-A*31:01, HLA-A*31:01, HLA-A*32:01, HLA-A*32:01, HLA-A*33:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:01, HLA-A*68:02, HLA-A*68:02, HLA-B*07:02, HLA-B*07:02, HLA-B*08:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:01, HLA-B*35:01, HLA-B*35:01, HLA-B*40:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:02, HLA-B*44:03, HLA-B*44:03, HLA-B*51:01, HLA-B*51:01, HLA-B*53:01, HLA-B*53:01, HLA-B*57:01, HLA-B*57:01, HLA-B*58:01, HLA-B*58:01.

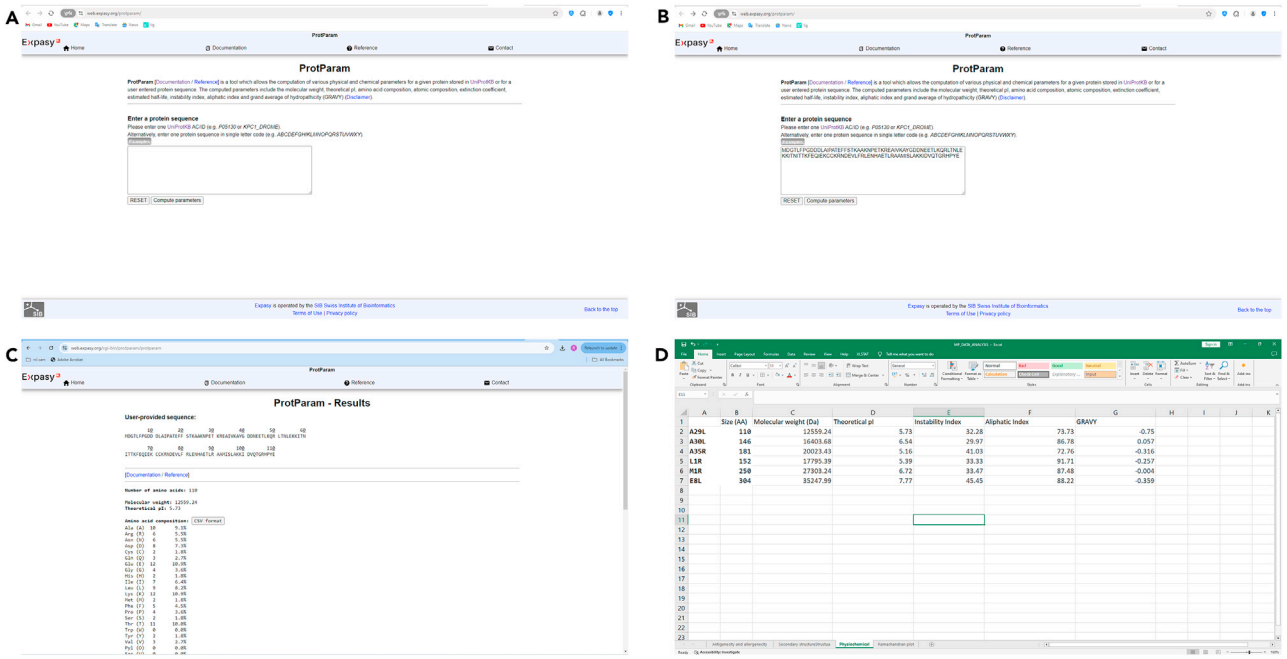


Figure 4. Analysis of physicochemical properties of selected glycoproteins using ProtParam

(A) Homepage of ProtParam webserver.

(B) Entered protein sequence and submission.

(C) Result page of searched protein.

(D) Combined excel table containing physicochemical properties.

23. Sort peptides by "Percentile Rank" and show by "IC50 below(cutoff) nM" and submit (Figure 7C and 7D).
24. Scrutinize and screen the resultant epitopes based on IC50 value $\leq$ 85 and arrange them alphabetically in an Excel file (Figure 8A).
    a. Epitopes showing correlation with multiple alleles are considered as efficacious binders which are selectively sorted.
    b. Club and arrange epitopes with multiple alleles in an Excel sheet. (Figures 8B–8D).
25. Subject the glycoproteins to the NetCTL-1.2 server (DTU Health Tech) for MHC I epitope prediction (Figure 9A).
26. Paste the individual protein sequence in the dialog box, selecting "A1 supertype" and submit the sequence by opting the default parameters (Figure 9B).
27. Select the result ending with (< - E) which signifies epitope being MHC I positive (Figures 9C and 9D).
28. Repeat steps 24–25 for a single protein by opting different supertypes.
29. Paste the results in an Excel sheet and sort the data as step 23 on the basis of peptides with their respective supertype hits (Figures 10A–10C).
30. Highlight and select peptides with multiple and overlapping supertypes.
31. Prepare the combined data from IEDB MHCI and NetCTL.1.2 server, displaying the peptides with their respective supertypes and HLA alleles. (Figure 11A).

Note: Two software tools are used to validate MHC I T-cell epitopes for accurate prediction.

32. Select the multiple HLA alleles binding epitopes on the basis of VaxiJen v.2.0 and AllerTop v 2.0 prediction as done in steps 5–10(Figures 11B and 11C).

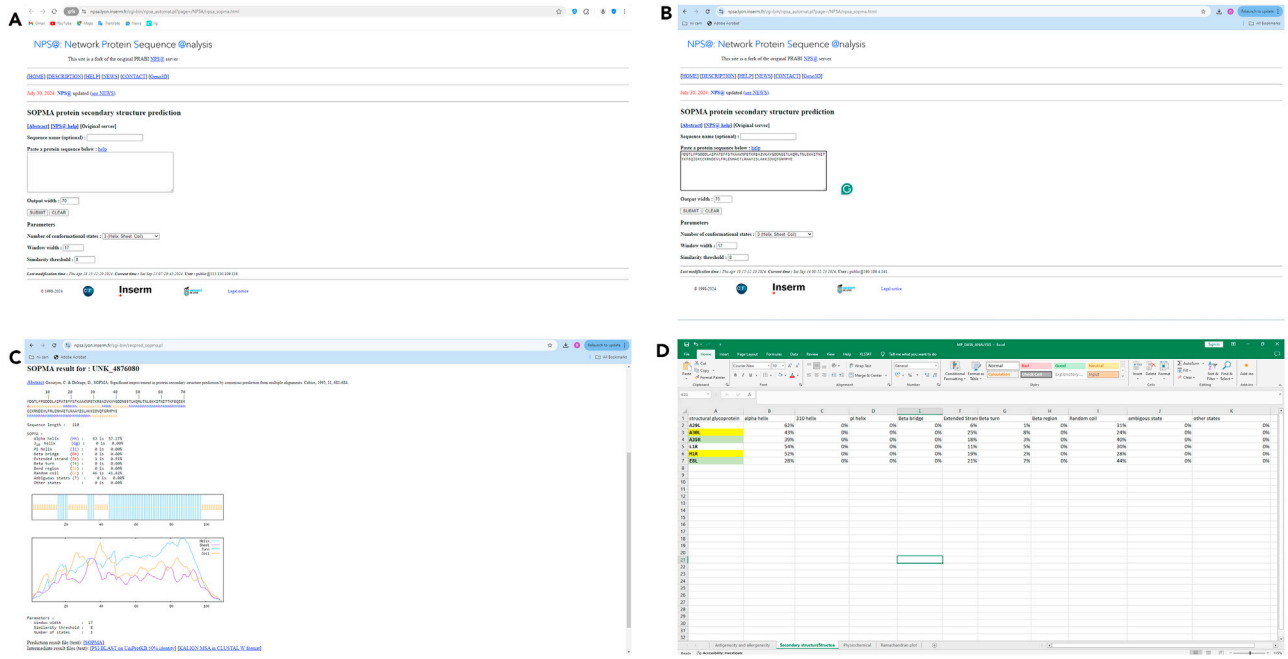


Figure 5. Secondary structure prediction using SOPMA

(A) Homepage of SOPMA.

(B) Entered protein sequence and selection of parameters.

(C) Result page.

(D) Excel sheet containing secondary structure parameters.

33. Combine data of all proteins to generate a final excel file of all selected CTL epitopes with their allergenicity and antigenicity (Figure 11D).

Part 4: Prediction of T cell-specific epitopes: MHC II

Timing: 6 h

This section focuses on steps and analysis of the prediction of T cell-specific epitopes via IEDB MHC II and NetMHCpan-4.0 servers, which are restricted to MHC class II molecules.

34. Open IEDB MHC II Binding Predictions tool (Figure 12A).
    a. Paste the protein sequence in FASTA format.
    b. Select the 7-HLA allele reference set.
    c. Click the "Submit" opting default parameters for rest.
    d. Sort peptides according to percentile rank and multiple alleles against a single epitope (Figures 12B and 12C).

Note: The peptide scoring is based on percentile rank, and epitope sorting is done as step 24 in an Excel datasheet.

35. Use the NetMHCIIpan-4.0 tool for MHC II binding prediction to identify the HLA alleles that recognize epitopes.
    a. Paste the protein in FASTA format.
    b. Opting for peptide length 15 (Figure 13A).
    c. Select loci for 20 alleles at a time.

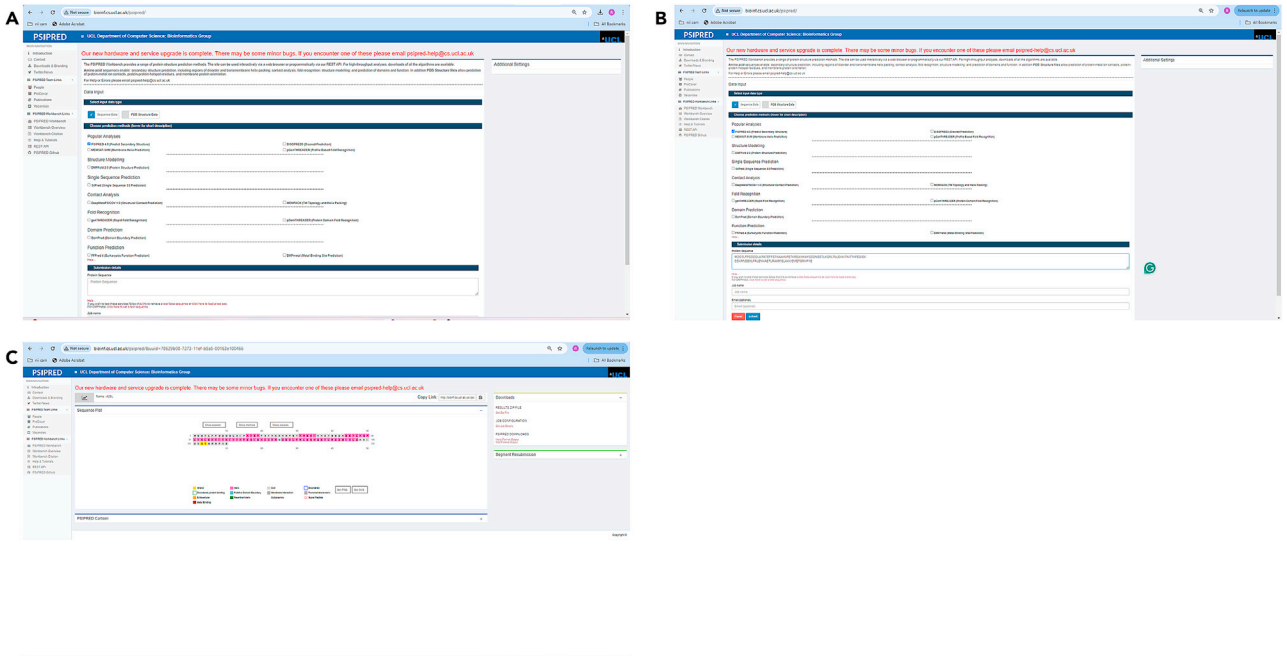


Figure 6. Secondary structure prediction using PSIPRED 4.0
(A) Homepage of PSIPRED.
(B) Entered protein sequence.
(C) Result page.

d. Set the threshold for strong binders to 2%. Repeat this process to cover the remaining alleles (Figure 13B).

36. Select results ending with "(<=SB)," indicating that the epitope is a strong binder with MHC II. Epitope sorting will be same as in MHC I (Figure 13C).
37. Compile epitopes from both the epitope prediction tools in an Excel file comprising peptides (epitopes) with their respective alleles.
38. Arrange the peptides alphabetically, select those epitopes with multiple alleles (Figure 13D).
39. Select the multiple HLA alleles binding epitopes on the basis of VaxiJen v.2.0 and AllerTop v 2.0 prediction as done in steps 5–10 (Figure 14A and 14B).
40. Repeat this process for each protein, create a final Excel spreadsheet, and organize it as described in step 33 (Figures 14C and 14D).

Part 5: Prediction of population coverage of selected T cell epitopes

Timing: 2 h

This section focuses on steps to predict the population coverage of selected T-cell epitopes using the IEDB Population Coverage tool.

41. In an Excel sheet, compile the overlapping CTL and HTL epitopes along with their supertypes and HLA alleles individually for each protein (Figure 15A).

Note: Repeat the steps for all proteins.

42. Transfer the data to a notepad file sequentially, using peptides (epitopes) on the left and supertype alleles on the right and save it for individual protein (Figure 15B).
43. Use the IEDB population coverage tool for population coverage.

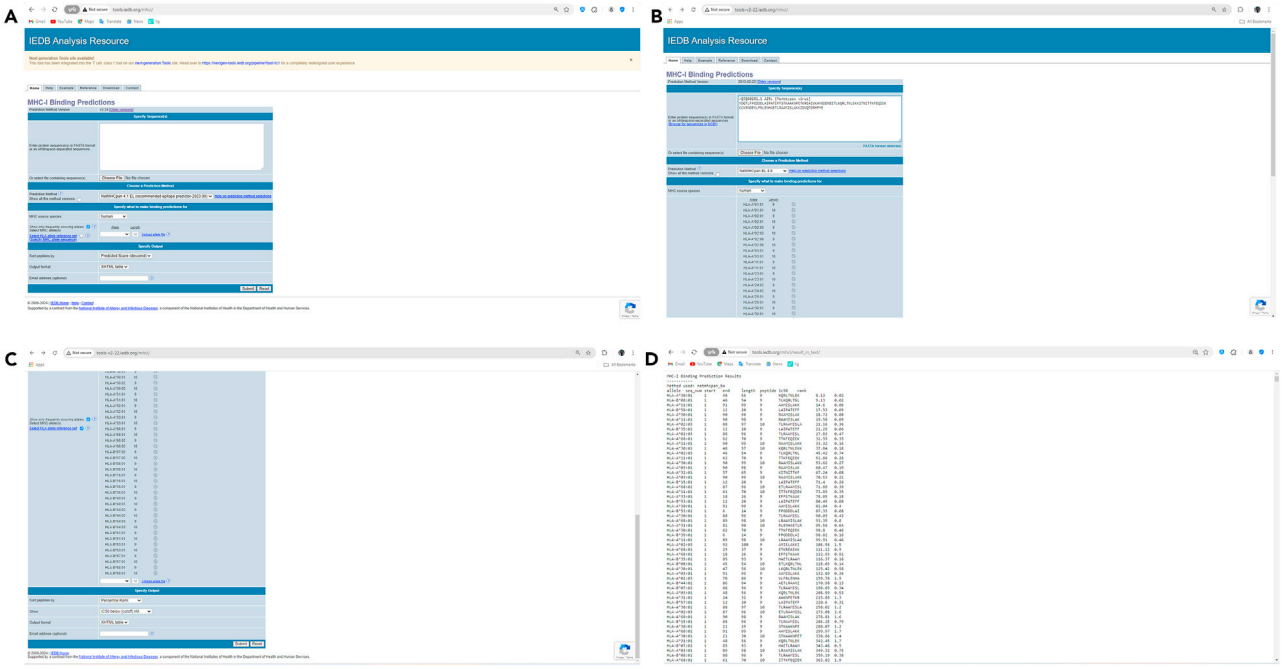


Figure 7. MHC I binding prediction using IEDB MHC I

(A) Homepage of MHC I Binding prediction (B) Entered protein sequence (C) list of all selected HLA alleles reference set. (D) Result page.

44. Select "Class I and II combined" along with "World" option (Figure 16A).
45. Select the saved notepad file in step 42 from "Choose file" option and "Submit" (Figure 16B).
46. Select "View coverage of individual epitope in world" and paste the data in an Excel sheet (Figures 16C and 16D).
47. Merge the epitopes with their respective classes, supertypes, HLA alleles, and the population coverage percentage along with the overall population coverage percentage.
48. Select the epitopes with a cut-off value for CTL>50% and HTL>25%.
49. Create an Excel file of the same including all selected proteins (Figure 17A).
50. Enter the selected epitopes from step 46 in the "epitopes" column and their alleles in "MHC restricted alleles" column of the IEDB population coverage tool manually (Figure 17B).
51. Click on submit to identify the overall population coverage % of selected T cell epitopes (Figure 17C and 17D).

Part 6: Prediction of B cell-specific epitopes

Timing: 2 h

This section enlists the steps and analysis on predicting the B cell specific epitopes using the Antibody epitope prediction tool of IEDB.

52. Use the "Antibody Epitope Prediction" tool of IEDB Analysis Resource to delineate the epitopes for selected proteins.
53. Paste the protein sequence. Select "Bepipred Linear Epitope Prediction 2.0" method and submit (Figures 18A and 18B).
54. Use VaxiJen v.2.0 and AllerTop v 2.0 for antigenicity and allergenicity of predicted epitopes (Figure 18C).
55. Repeat steps 50–52 for all the protein sequences and compile it in an excel file (Figure 18D).

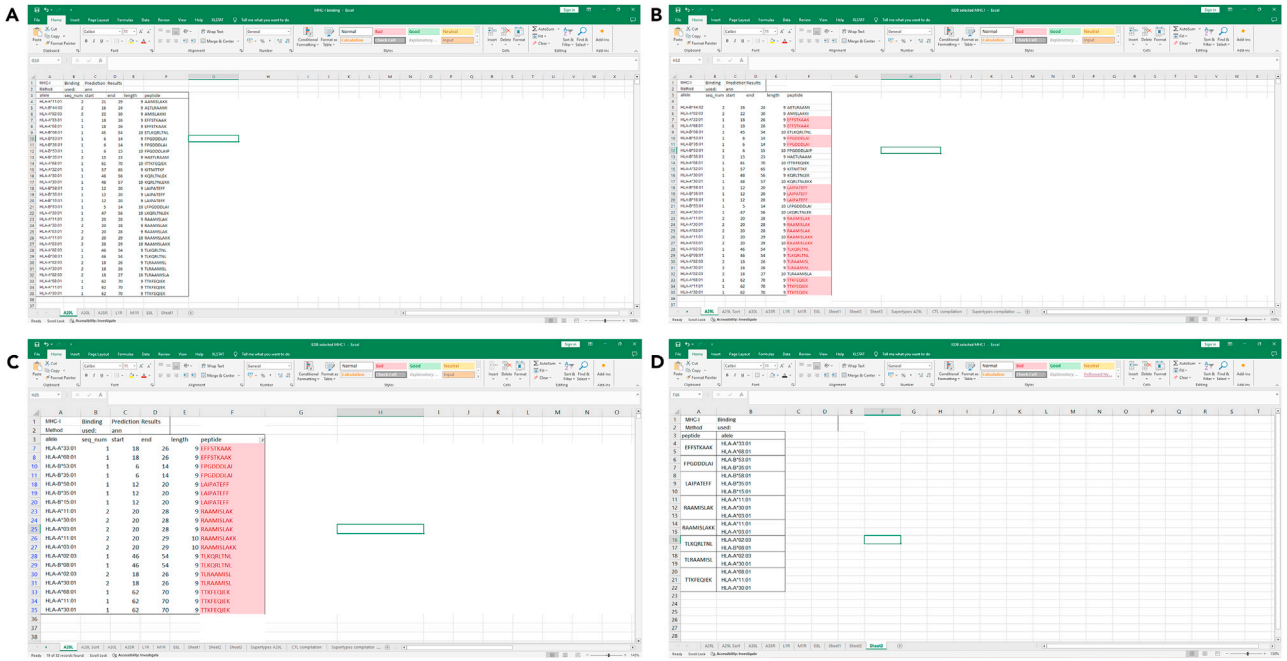


Figure 8. Epitope sorting of MHC I restricted alleles

(A) Excel file of selected epitopes.

(B) Highlighted epitopes.

(C) Filtered highlighted epitopes.

(D) Compiled file of highlighted epitopes against selected alleles.

Part 7: IFN-gamma epitope selection and conservancy analysis of all selected epitopes

Timing: 6 h

This section describes the procedure for prediction of IFN-gamma epitopes via IFNepitope tool, and conservancy analysis of all T cell, B cell and IFN-gamma epitopes via Epitope Conservancy Analysis of IEDB.

56. Utilize the "IFNepitope" tool to predict and design epitopes inducing IFN-gamma for all proteins.
57. Cleave the protein into 15 amino acid chains.
58. Submit by selecting "Motif and SVM hybrid" and "IFN-gamma versus non-IFN-gamma" option (Figures 19A and 19B).
59. Repeat the process for all proteins and create an Excel sheet selecting the epitopes with POSITIVE as result (Figure 19C).
60. Use VaxiJen v.2.0 and AllerTop v 2.0 for antigenicity and allergenicity of predicted epitopes (Figure 19D).

⚠ CRITICAL: Repeat steps for all the proteins and compile the data in an excel file to select one epitope from each protein that shows highest SVM/ MERCI score (the score will be displayed in the last column of your result page (Figure 19E).

Note: Select the available epitopes by prioritizing the "SVM" (Support Vector Machine) algorithm method first. If no epitopes are available from this method, then select from the "MERCI" (Maximum Entropy-based Prediction of Immunogenicity) method.

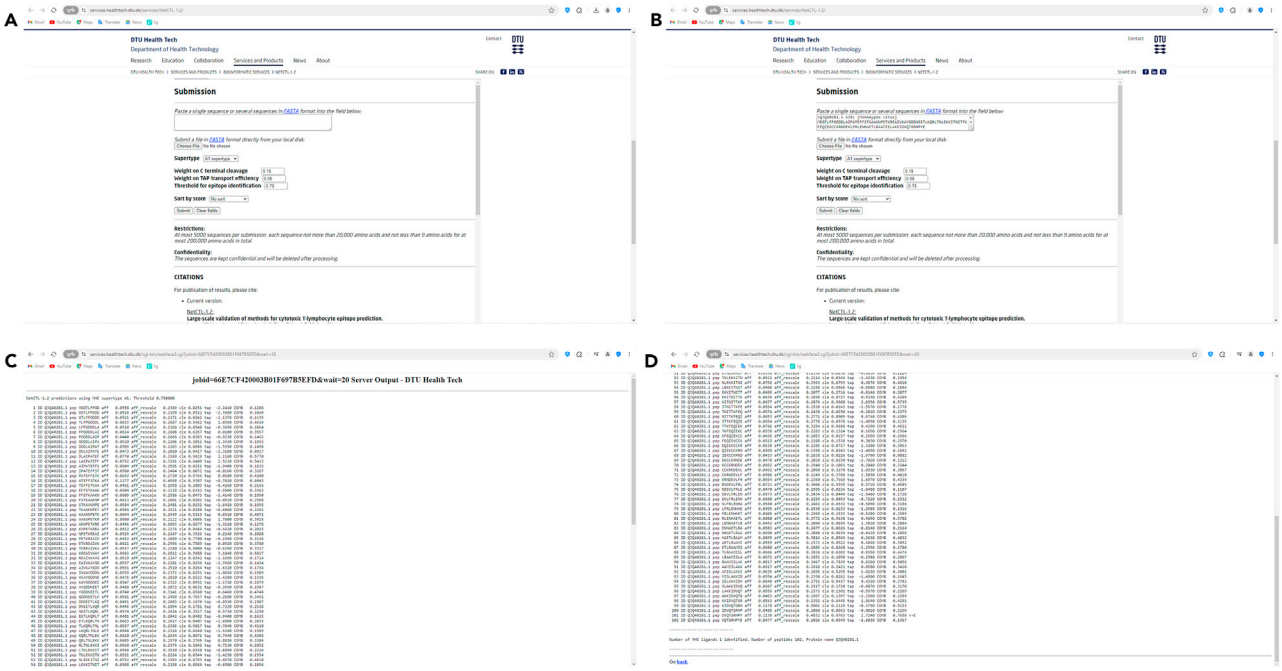


Figure 9. MHC I binding prediction using NetCTL-1.2

(A) Homepage of NetCTL-1.2 (B) Entered protein sequence (C) Result page.
(D) Result page with binding allele shown as <-e (number 101).

61. Perform Conservancy analysis of all selected epitopes, go to the NCBI homepage.
62. Select "Protein" from the dropdown menu in the search bar, and search for the protein of interest (Figures 20A and 20B).

Note: In this case, we have shown the example for B cell epitopes of A35R MPXV protein.

63. Copy the full protein sequence(s) displayed in FASTA format. Paste all sequences into a notepad file and save it (Figure 20C).
64. Go to "IEDB Population coverage analysis" and paste final selected epitope sequences (B cell, T cell and IFN-gamma) in the input box (Figures 21A and 21B).
65. Enter multiple epitopes in FASTA format separately.
66. Select the "Epitope linear sequence conservancy" as the "Analysis type" option to display the amino acid sequence of the specific protein entry with "100%" sequence identity as threshold and submit (Figure 21C).
67. The results will display the conservation percentage for each epitope (Figure 21D).

Note: 90–100% matches indicate highly conserved epitopes ideal for targeting across strains or species.

Part 8: Construction of 3D vaccine structure

Timing: 10 h

This section details the procedure and analyzes construction of 3D vaccine candidate from the selected epitopes, assessing its antigenicity, allergenicity, physicochemical characteristics, secondary and tertiary structure predictions of the designed sequence. This designed candidate is our final vaccine protein which will be further used for docking and simulation.

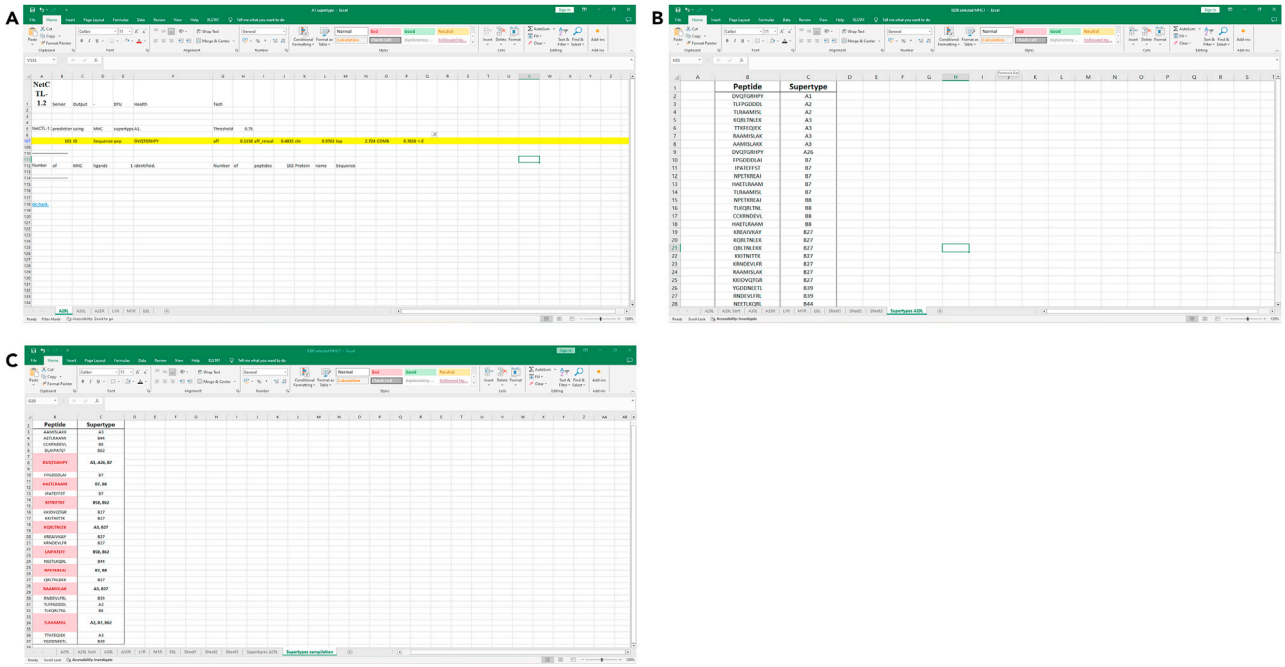


Figure 10. Epitope sorting of MHC I restricted alleles through NetCTL-1.2

(A) Excel file of selected epitopes of single supertype.

(B) Combined excel sheet of epitopes of all supertypes.

(C) Highlighted epitopes with more than one supertypes.

68. Join the selected MHC I, II and IFN-gamma epitope sequences with the "GPGPG" or "AAY" linker and PADRE peptide as adjuvant. For adjuvant, use the "EAAAK" linker to join adjuvant in Excel file (Figure 22A).
69. Paste the formed sequence in a Word file (Figure 22B).
70. Proceed for the antigenicity prediction using VaxiJen v.2.0 (Figures 22C and 22D).
71. Use AllerTop v 2.0 for allergenicity prediction (Figures 23A and 23B).

⚠ CRITICAL: If the sequence does not pass antigenicity and allergenicity use other epitopes and redo the steps.

72. Use the Expasy-ProtParam server to predict the physiological parameters of designed protein sequence as described in step 13–15 (Figures 23C and 23D).
73. Use "Protein Scanning" tool of ToxinPred webserver to find any toxic regions in protein sequence.
74. Paste the designed protein sequence and click on "Run Analysis" keeping the default parameters (Figures 24A and 24B).

⚠ CRITICAL: If the sequence shows toxic epitopes or sequences, reshuffle or change the epitopes and repeat from step 68.

75. Use "SOLUPROT v.1.0" tool to asses solubility of designed protein expression in *Escherichia coli*.
76. Paste the sequence and click on submit to predict the Solubility score (Figures 24C and 24D).

⚠ CRITICAL: If the protein shows no solubility, reshuffle or change the epitopes and repeat from step 68.

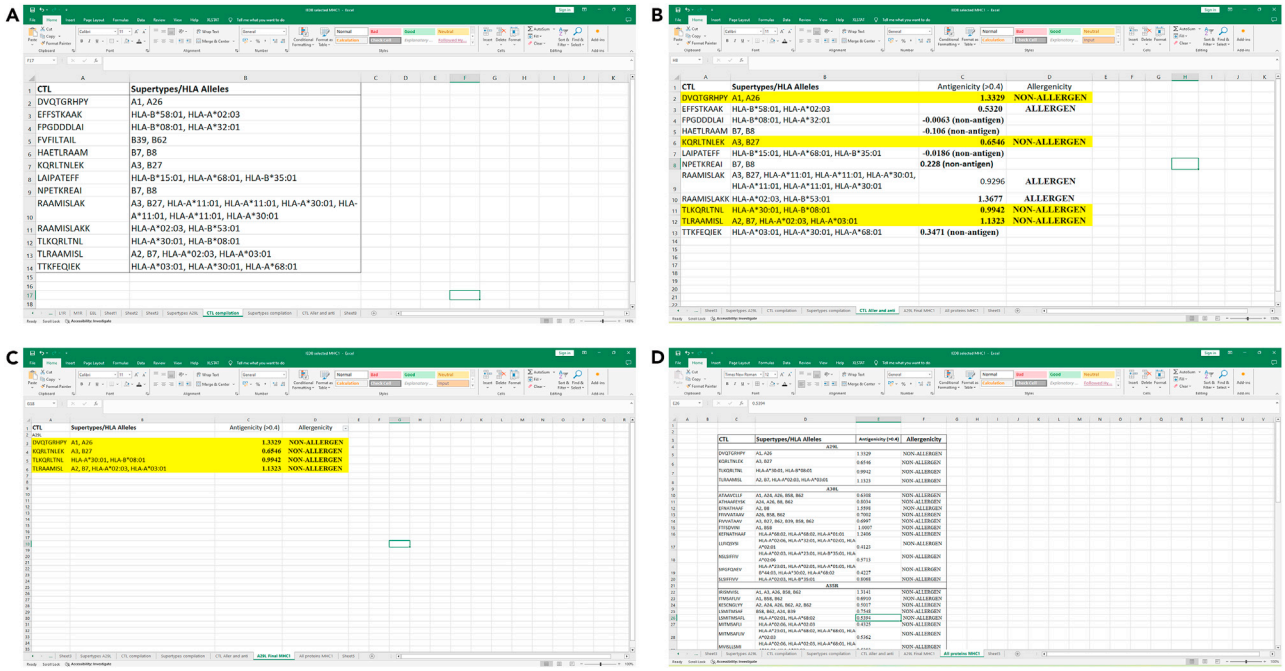


Figure 11. Combined epitopes sorting of MHC I binding prediction and NetCTL-1.2

(A) Combined excel sheet of epitopes of both servers.

(B) Antigenicity and allergenicity of predicted epitopes.

(C) Excel sheet of final selected epitopes of single protein.

(D) Excel sheet of final selected epitopes of all glycoproteins.

77. Use SOPMA for secondary structure prediction of designed protein. Use as described above step 16-18 (Figures 25A and 25B).
78. To predict tertiary structure of protein use "Robetta" tool (Figure 26A).
79. Paste the designed protein sequence in "Protein sequence" column.
80. Enter the "Target Name" and "Submit" the job using default parameters (Figure 26B).

Pause Point: It might take 1–2 days to complete the run.

81. Download the designed protein structure in .pdb format (Figure 26C).
82. Use GalaxyWEB server to refine the predicted protein structure from Robetta (Figure 27A).
83. Upload the downloaded protein structure file (.pdb) in the GalaxyWEB server and submit after adding "Job name" (Figure 27B).

Pause Point: It might take 1–2 days to complete the run.

84. Download the refined structure (Figure 27C).
85. Visualize the 3D structure utilizing MolProbity to check the Ramachandran plot and rotamers.
86. Upload the refined structure. Keep all default parameters and "upload" the job (Figure 28A).
87. Click on continue after new web page appears (Figure 28B).
88. In the next webpage, select "analyze geometry without all atom contacts". Keep parameters default and submit (Figures 28C and 28D).

Note: Verify the structure with the Ramachandran favored percentage (>95% - Valid) (Figure 28E). The verified protein structure will be saved as Final protein.pdb.

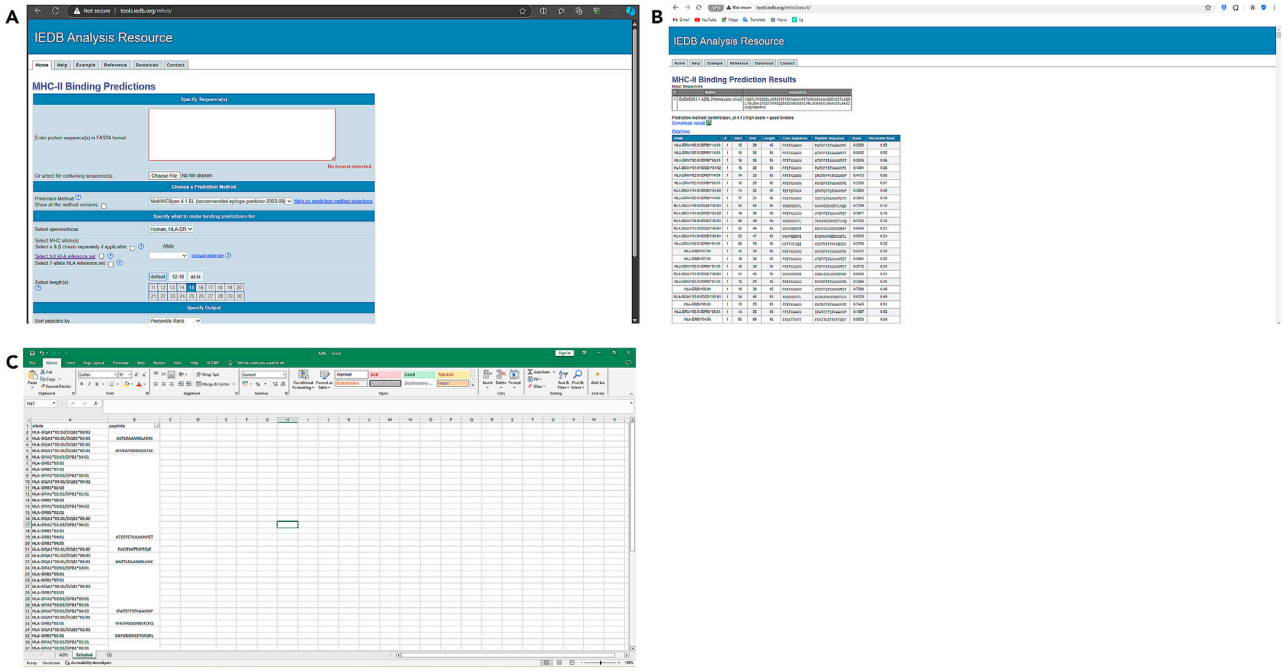


Figure 12. MHC II binding prediction using IEDB MHC II

(A) Homepage of MHC II binding prediction server.

(B) Result page.

(C) Epitope sorting of MHC II binding prediction.

⚠ CRITICAL: Confirm that the 3D epitope-linker combinations meet a Ramachandran plot quality score of at least 95%. If this criterion is not achieved, repeat from step 64 reconfigure and modify the arrangement of linkers and epitopes to create new combinations. Validate the updated 3D structure using the Ramachandran plot to ensure its suitability for further analysis.

Part 9: Molecular docking

Timing: 1 day

This section details the procedure to perform the protein-protein interactions via protein docking platform ClusPro, its validation using PRODIGY webserver and analyze the site of interaction using Ligplot+.

89. Open the PDB website, search "TLR4 structure" and download the structure in PDB format (Figures 29A–29D).

Note: – If ligand is not available download the protein structure (here, PDB: 3FXI) pre- docked with any other molecule.

90. Open the downloaded structure in PyMOL and go to Display – Sequence (Figure 30A).
91. Select and remove the extra unwanted sequences and molecules (if any) (Figure 30B).

⚠ CRITICAL: Verify the presence of any additional molecules and manually remove them; otherwise, the docking will not result in a valid structure.

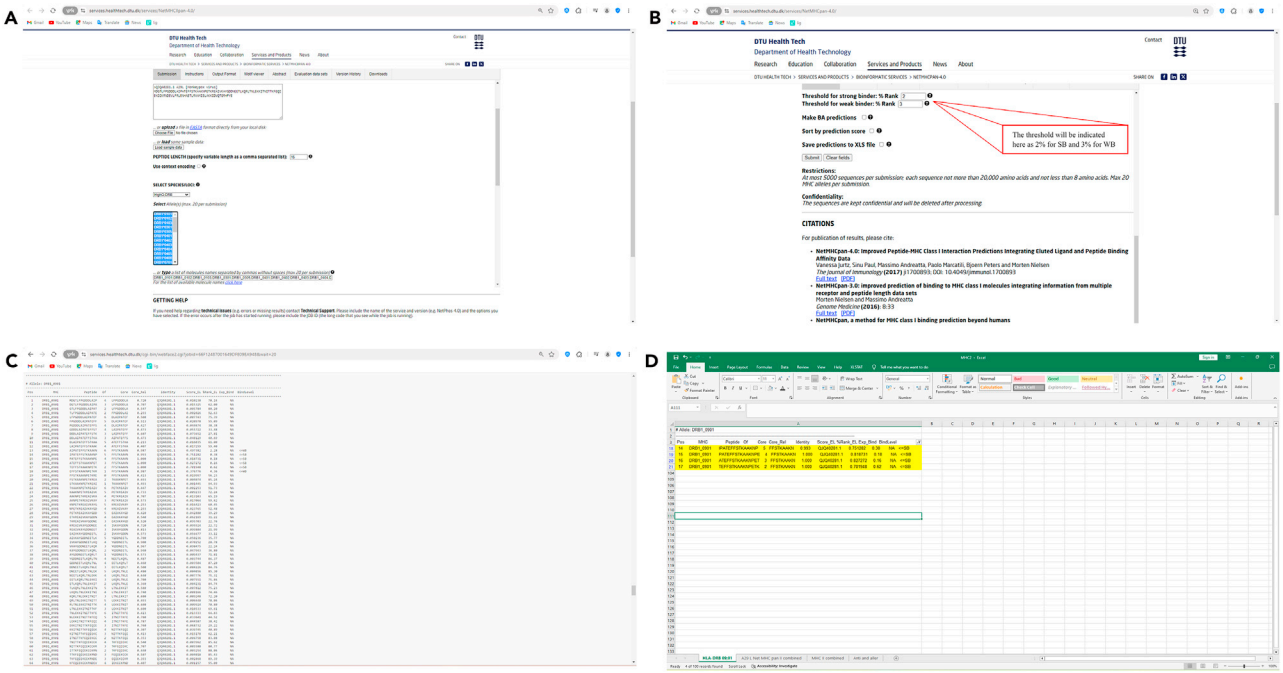


Figure 13. MHC II binding prediction using NetMHCpan 4.0

(A) Homepage with entered protein sequence.

(B) Parameters for selection.

(C) Result page.

(D) Selected epitopes.

92. Delete one of the protein chains to retain a single arm of TLR4.
93. Export it as "PDB file" for molecular docking (Figures 30C and 30D).
94. The final exported file will be saved as TLR4.pdb.

Note: Here we have shown the steps for TLR4. For MHC molecules do not perform step 92.

95. To perform the molecular docking, register an official email ID on the ClusPro server (Figure 31A).
96. Specify the "Job name", select the receptor file (TLR4.pdb) and ligand/protein file (Final protein.pdb).
97. Select the "Dock" option and check the registered email ID for results (Figures 31B and 31C).

Pause Point: It might take 6–8 h to complete the run.

98. Download the PDB of all models (Figure 31D).
99. Run PRODIGY server to validate the protein-protein interaction on the basis of ΔG values of the predicted models.
100. Specify the chain name, temperature (37 degrees) and "Submit Prodigy" by opting "I am not a robot" option to get the results (Figures 32A and 32B).
101. Select the docked structure with the lowest ΔG for further analysis (here: model.000.00.pdb) (Figure 32C).
102. Download LigPlot+ software (Figure 33A).
103. Open selected (model.000.00.pdb) file and opt to run to plot the protein-protein interactions of both chains (Figure 33B).
104. Select "DIMPLOT" and run (Figure 33C).

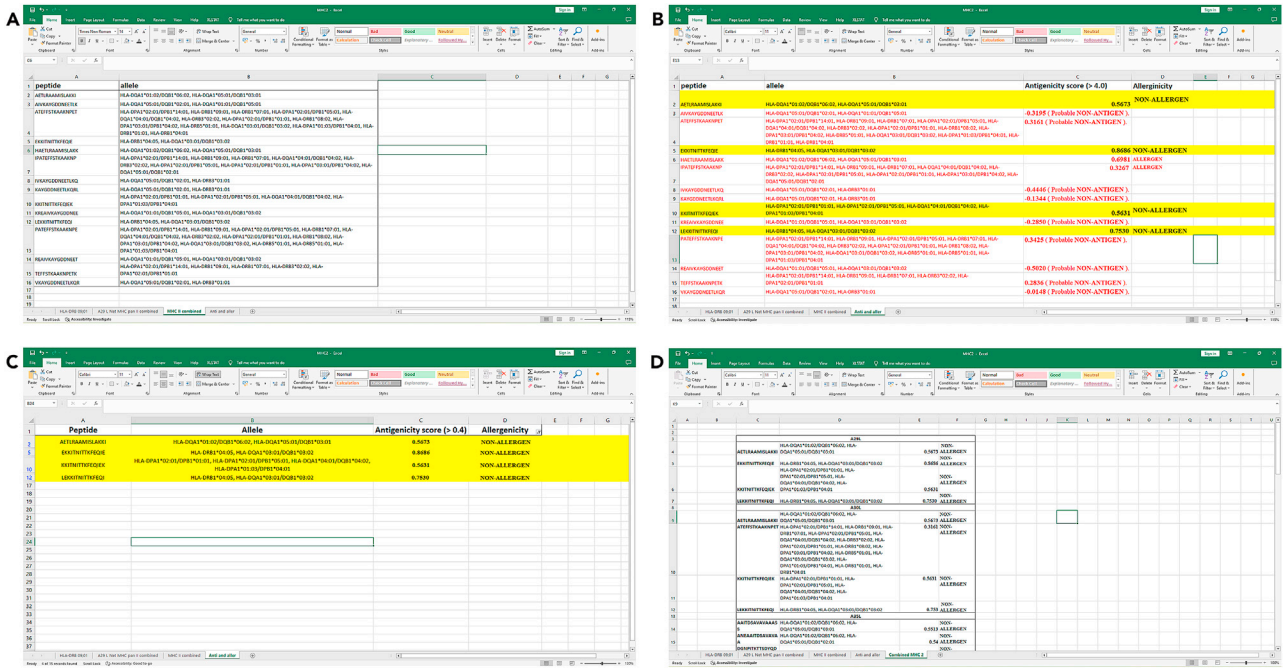


Figure 14. Combined epitopes sorting of MHC II binding prediction and NetMHCpan 4.0

(A) Excel sheet of combined epitopes.

(B) Antigenicity and allergenicity of predicted epitopes.

(C) Excel sheet of final selected epitopes of single protein.

(D) Excel sheet of final selected epitopes of all glycoproteins.

105. The results will be displayed (Figure 33D).

Part 10: Molecular dynamic simulation

Timing: 15 days

This section outlines the procedure to use the GROMACS package for conducting molecular dynamics simulation studies of the designed vaccine-receptor docked complex.

Note: Use GROMACS package for the Molecular dynamic simulation studies of the vaccine receptor docked complex. All the commands are taken from GROMACS tutorial.

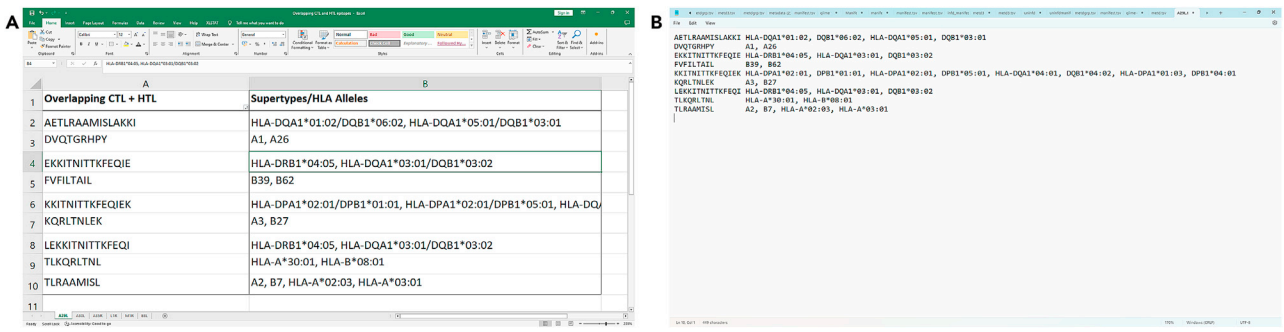


Figure 15. Notepad file for population coverage

(A) excel sheet of combined overlapping epitopes of MHC I and II.

(B) Notepad file of epitopes.

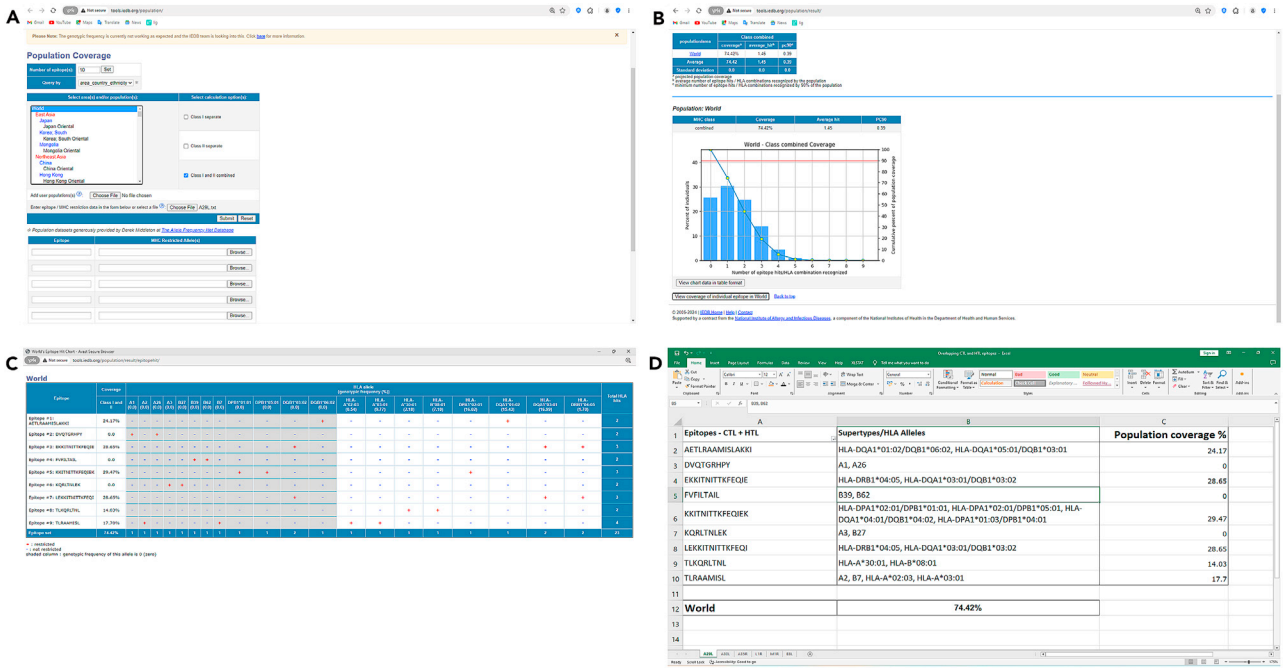


Figure 16. Population coverage of selected T cell epitopes

(A) Homepage of IEDB population coverage tool.

(B) Screenshot of result page.

(C) Result page showing coverage of individual epitope in the world.

(D) Excel sheet of epitopes showing population coverage.

106. Generate topology: Save the selected docked file complex model.000.00.pdb in your respective folder for further steps (here we have saved the docked file in D:\Amit\Docking).

    a. Use the following commands in the Ubuntu terminal for mapping the main directory. `cd /mnt`

```
cd d /Amit /Docking
```

    b. Use the following command to generate a topology file with coordinates.

```
gmx pdb2gmx -f model.000.00.pdb -o complex.gro -water tip3p -ignh
```

Note: model.000.00.pdb is your docked file in .pdb format saved in the docking folder in previous step.

    c. Enter "7" to select the "AMBER99SB -ILDN" force field (Figure 34A).

Note: You can choose force field according to your need and you can also choose your water model, here we have chosen tip3p.

    d. Three files will be generated in the docking folder - topol.top; complex.gro and posre.itp.

107. Define box and solvate: Run the following commands in terminal.

```
gmx editconf -f complex.gro -o newbox.gro -bt dodecahedron -d 1.0

gmx solvate -cp newbox.gro -cs spc216.gro -p topol.top -o solv.gro
```

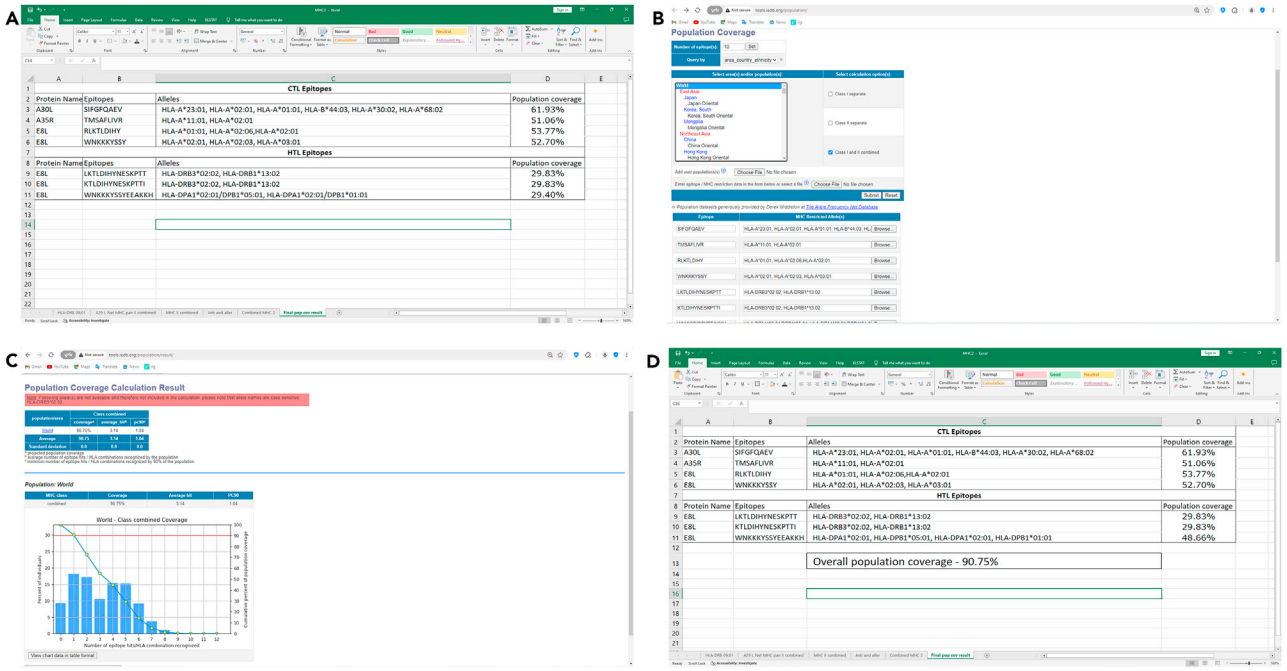


Figure 17. Overall population coverage of selected epitopes

(A) Excel sheet of epitopes that passes the cutoff (CTL>50% and HTL>25%).

(B) Manually entered epitopes and HLA alleles in IEDB population coverage server.

(C) Screenshot of population coverage calculation result.

(D) Excel sheet of overall population coverage.

108. Add ions: Add new ions by copying the text from the given link (http://www.mdtutorials.com/gmx/complex/Files/ions.mdp) and paste it into a newly created document named "ions.mdp" in the docking folder (Create a new notepad file in the docking folder and save it as ions.mdp).
    a. Run the following command in terminal.

```
gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr
```

(a new ions.tpr file will be formed).

    b. Command to pass your .tpr file to genion.

```
gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -neutral -pname NA -nname CL
-neutral -conc 0.15
```

Note: In GROMACS, the .tpr file, or portable binary run input file, contains all the necessary information to perform a molecular dynamics (MD) simulation. This file is essential when using the genion tool in GROMACS to add or replace ions in your system, allowing you to neutralize charge or reach a desired ionic concentration.

    c. Now select the group of solvent molecules (Figure 34B).
    From the obtained list choose the SOL (13)
    In terminal type:

```
13

q
```

109. Energy minimization (EM) of the system: Create the binary input using the tool grompp and this input parameter file:
    a. Copy the text from the given link to perform energy minimization (http://www.mdtutorials.com/gmx/complex/Files/em.mdp) and paste it into a newly created document named "em.mdp" in the docking folder (Create a new notepad file in the docking folder and save it as em.mdp).
    b. Convert .mdp file to .tpr file, run the following command.

```
gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -o em.tpr
```

    c. Run command to carry out Energy minimization (EM) of the system.

```
gmx mdrun -v -deffnm em
```

Note: -deffnm em: Sets "em" as the default prefix for file names. GROMACS will look for an "em.tpr" file (created with 'gmx grompp') and will generate output files such as "em.gro," "em.edr," "em.log," etc., all using "em" as the base name.

110. Equilibration of the protein ligand complex:
    a. Restrain the ligand by running the following command in terminal.

```
gmx genrestr -f ligand.gro -o posre_jz4.itp -fc 1000 1000 1000
```

    b. Include a position restraint file, such as "posre.itp", in your topology file.

⚠ CRITICAL: This allows you to manage when restraints are applied in GROMACS simulations. This approach is useful for selectively restraining certain atoms during specific stages, like energy minimization or equilibration, while letting the system evolve without restraints during the production phase.

To do this go to GROMACS tutorial copy the following text
; Include Position restraint file
#ifdef POSRES
#include "posre.itp"
#endif

    c. Open topol.top file. Paste the above copied text in topol.top file below "; Include ligand topology" and save it (Figure 34C).
    d. Make an index file: Run the following command.

```
gmx make_ndx -f em.gro -o index.ndx
```

From the obtained list choose the protein (1) and ligand (13)
In terminal type:

```
1

13

q
```

Note: This index file defines specific groups of atoms within your system, which you can then use in other GROMACS commands for analysis or for applying restraints, selections,

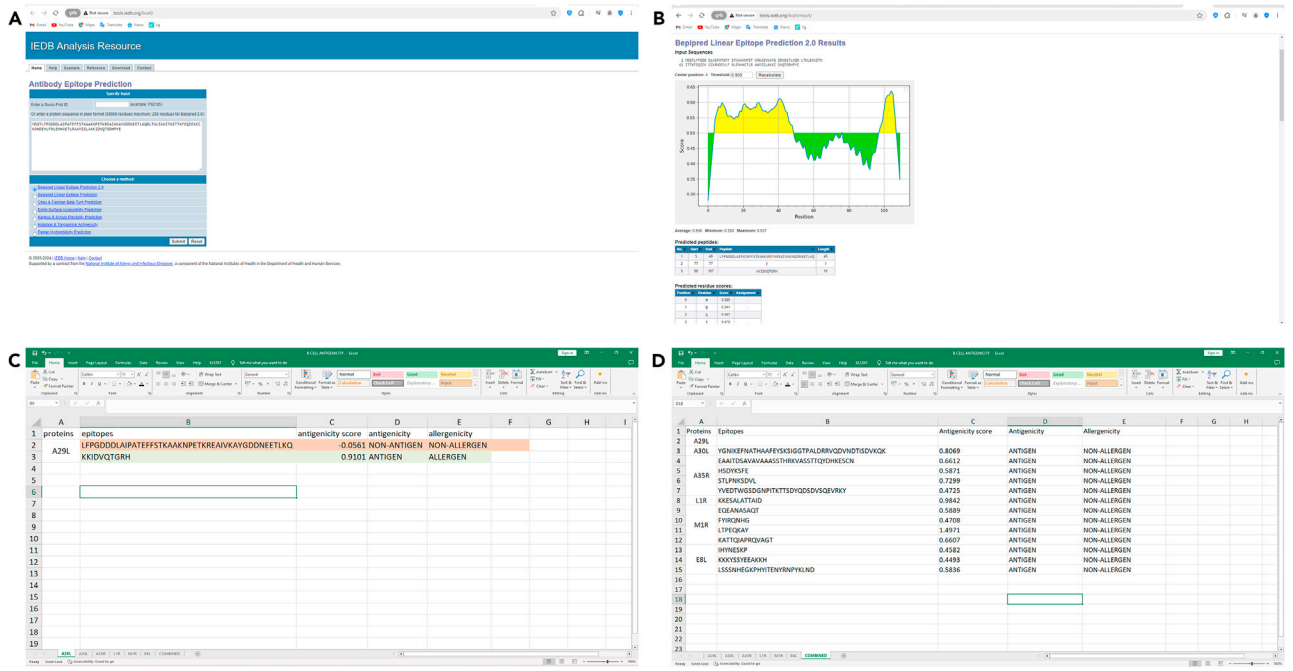


Figure 18. Representation of B cell epitope

Selection of B cells epitopes was performed using "Antibody Epitope Prediction" tool of IEDB Analysis Resource.
(A) Entered protein sequence in dialogue box of homepage.
(B) Bepipred linear epitope prediction result.
(C) Excel sheet containing antigenicity and allergenicity of selected epitopes.
(D) Excel sheet containing final selected epitopes of all proteins.

or calculations. This flexibility is especially helpful for analyses, restraints, or other custom selections that may involve only a subset of the total system.

111. Proceed with NVT equilibration.

Note: number of particles (N), system volume (V), and temperature (T).

a. Copy the text from given link for NVT equilibration (http://www.mdtutorials.com/gmx/complex/Files/nvt.mdp) and paste the text into a newly created document "nvt.mdp" in the docking folder (Create a new notepad file in the docking folder and save it as nvt.mdp).

Note: Change and note down no. of steps and time duration of MD simulation you want to run.

b. Convert .mdp file to .tpr file. Open GROMACS tutorial and run the following command.

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -n index.ndx -o nvt.tpr
```

c. MD run command to carry out NVT equilibration of the system.

```
gmx mdrun -deffnm nvt
```

112. Proceed with NPT equilibration number of particles (N), system pressure (P), and temperature (T).

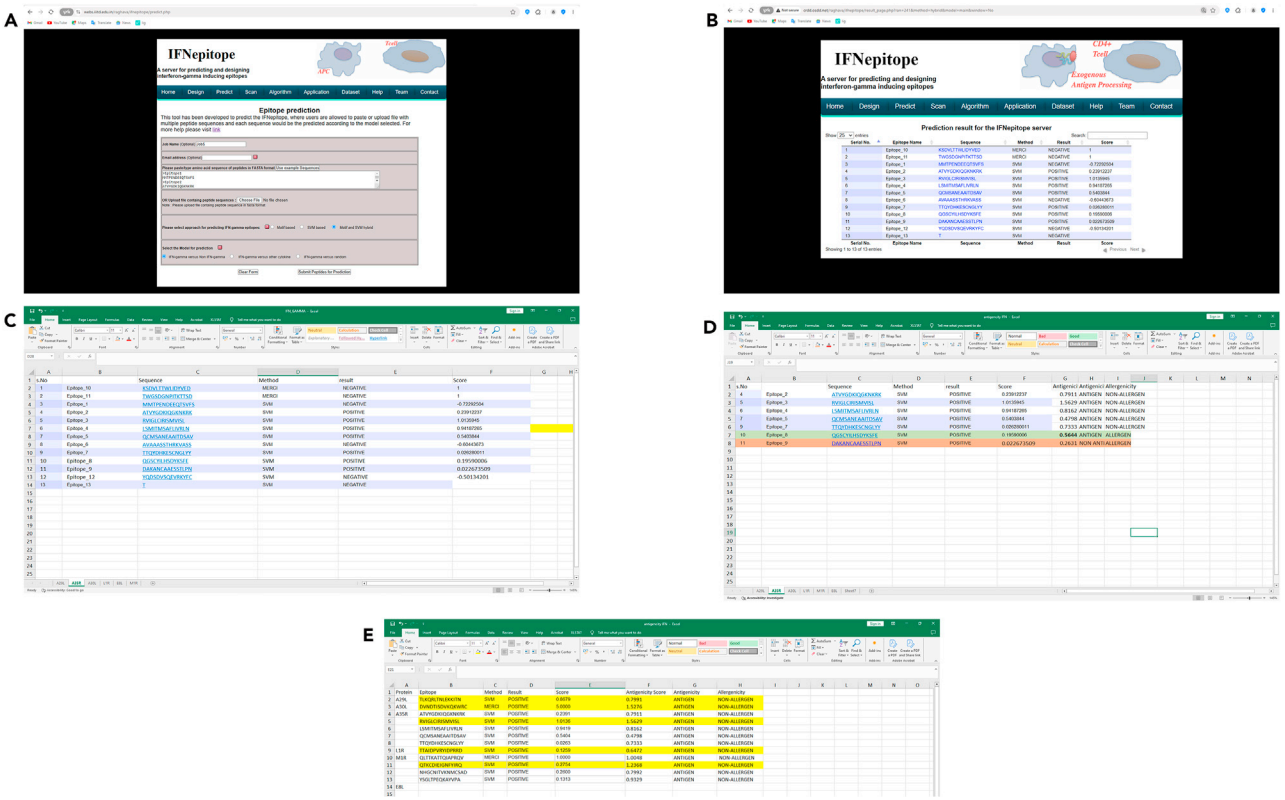


Figure 19. Finding of IFN-γ epitopes from pan-glycoproteins of MPXV

IFN-gamma specific epitopes were selected using the "IFNepitope" tool.

(A) Homepage containing entered protein sequences.

(B) Prediction result for IFNepitope server.

(C) Excel containing SVM scores.

(D) Excel sheet containing antigenicity and allergenicity of epitopes.

(E) Excel sheet containing all the selected epitopes.

a. Copy the text from given link for NPT equilibration (mdtutorials.com/gmx/complex/Files/npt.mdp) and paste the text into a newly created document "npt.mdp" in the docking folder (Create a new notepad file in the docking folder and save it as npt.mdp).

b. Convert .mdp file to .tpr file. Open GROMACS tutorial and run the following command.

```
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -r nvt.gro -p topol.top -n index.ndx -o npt.tpr
```

c. MD run command to carry out NPT equilibration of the system.

```
gmx mdrun -deffnm npt
```

Note: After completing the two equilibration phases, the system is now well-equilibrated at the target temperature and pressure. We are ready to remove the position restraints and begin the production MD phase for data collection. At this stage, we will use the checkpoint file with grompp to prepare for the run (a 50 ns MD simulation will be conducted).

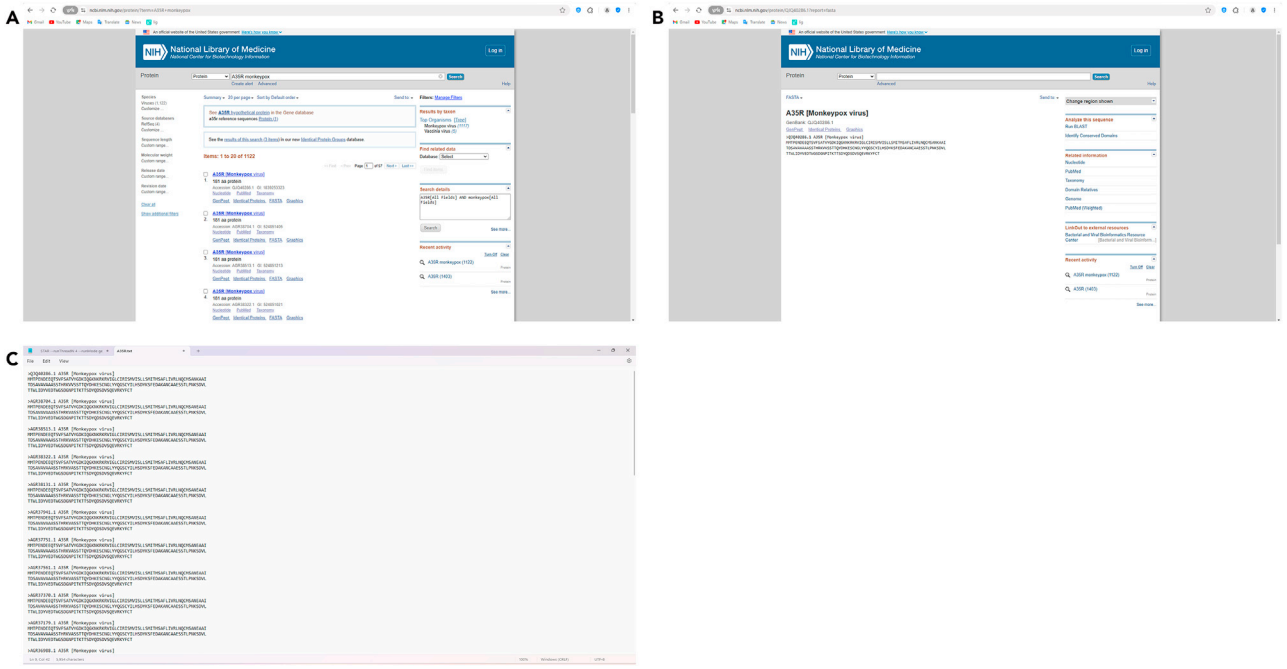


Figure 20. Extraction of protein sequences

(A) Searched protein sequence in NCBI.

(B) Protein sequence in FASTA format.

(C) Notepad file of all the different entries of single protein.

113. Production MD:

    a. Copy the text from given link for protein-ligand complex MD simulation (mdtutorials.com/gmx/complex/Files/md.mdp) and paste the text into a newly created document "md.mdp" in the docking folder (Create a new notepad file in the docking folder and save it as md.mdp).

    b. Convert .mdp file to .tpr file. Open GROMACS tutorial and run the following command.

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o md_0_1.tpr
```

    c. MD run command.

```
gmx mdrun -deffnm md_0_1
```

Note: After running the MD command, following output files will be obtained:

```
md_0_1.xtc, md_0_1.edr, md_0_1.trr, md_0_1.log, md_0_1.cpt, md_0_1.gro, md_0_1.dhdl
```

Pause Point: - It might take 6–7 days to finish the run.

114. RMSD (Root mean square deviation) analysis.

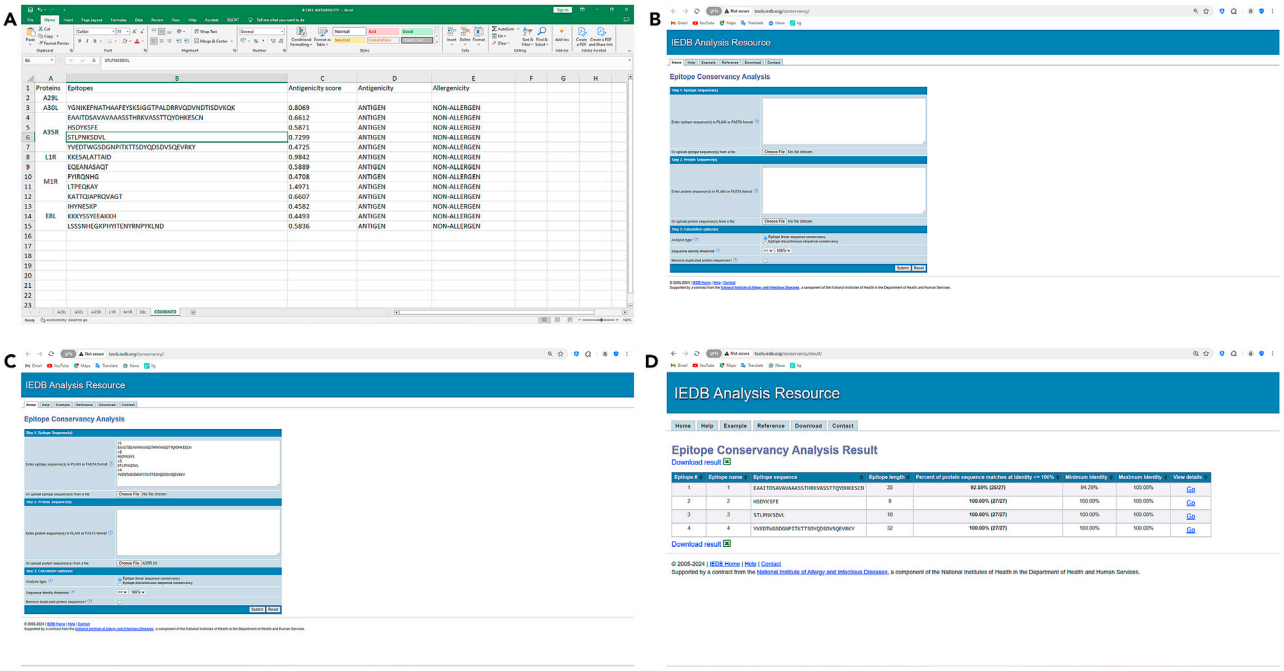


Figure 21. Epitope conservancy analysis using IEDB analysis resource

(A) Excel sheet of B-cell epitopes used for conservancy analysis.

(B) Homepage of Epitope conservancy analysis.

(C) Screenshot showing entered protein sequence and parameters.

(D) Epitope conservancy analysis result.

a. Calculate the RMSD values, using the following command.

```
gmx rms -f md_0_1.xtc -s md_0_1.tpr -o rmsd.xvg
```

Enter the value "4" to select "backbone" option (Figure 35A).

b. View the rmsd graph of the analysis (Figure 35B).

```
xmgrace rmsd.xvg
```

c. Use the command for probability distribution (Figure 35C).

```
gmx analyze -f rmsd.xvg -dist prob_rmsd.xvg
```

115. R_g (Radius of gyration) analysis.

a. Calculate the R_g using the command line.

```
gmx gyrate -f md_0_1.xtc -s md_0_1.tpr -o rg.xvg
```

Enter the value "1" to select "protein" option (Figure 36A).

b. View the R_g graph of the analysis (Figure 36B).

```
xmgrace rg.xvg
```

c. Use the command for probability distribution (Figure 36C).

```
gmx analyze -f rg.xvg -dist prob_rg.xvg
```

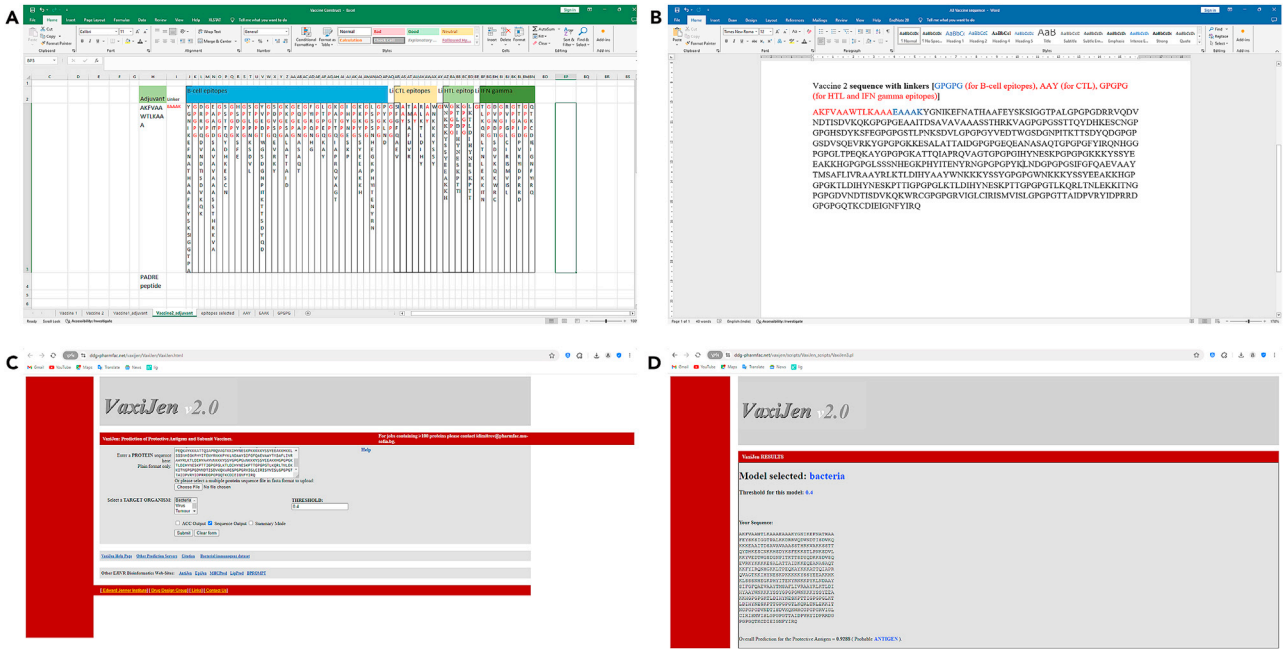


Figure 22. Construction of vaccine and antigenicity prediction

(A) Excel sheet of joined selected epitopes with linkers and adjuvants to form final vaccine.

(B) Word file showing primary sequence of vaccine.

(C) Vaccine antigenicity prediction.

(D) Vaccine antigenicity result.

d. Observe probability distribution curve using this command (Figure 36D).

```
xmgrace prob_rg.xvg
```

116. RMSF (Root mean square fluctuation) analysis.

a. Calculate the RMSF use the command line.

```
gmx rmsf -f md_0_1.xtc -s md_0_1.tpr -o rmsf.xvg -res
```

Enter the value "3" to select "C-alpha" option (Figure 37A).

b. View the RMSF graph of the analysis (Figure 37B).

```
xmgrace rmsf.xvg
```

117. SASA (Solent accessible surface area) analysis.

a. Calculate the SASA values using the command line.

```
gmx sasa -f md_0_1.xtc -s md_0_1 -o sasa.xvg
```

Enter the value "8" to select "SideChain" option (Figure 38A).

b. View the SASA graph of the analysis (Figure 38B).

```
xmgrace sasa.xvg
```

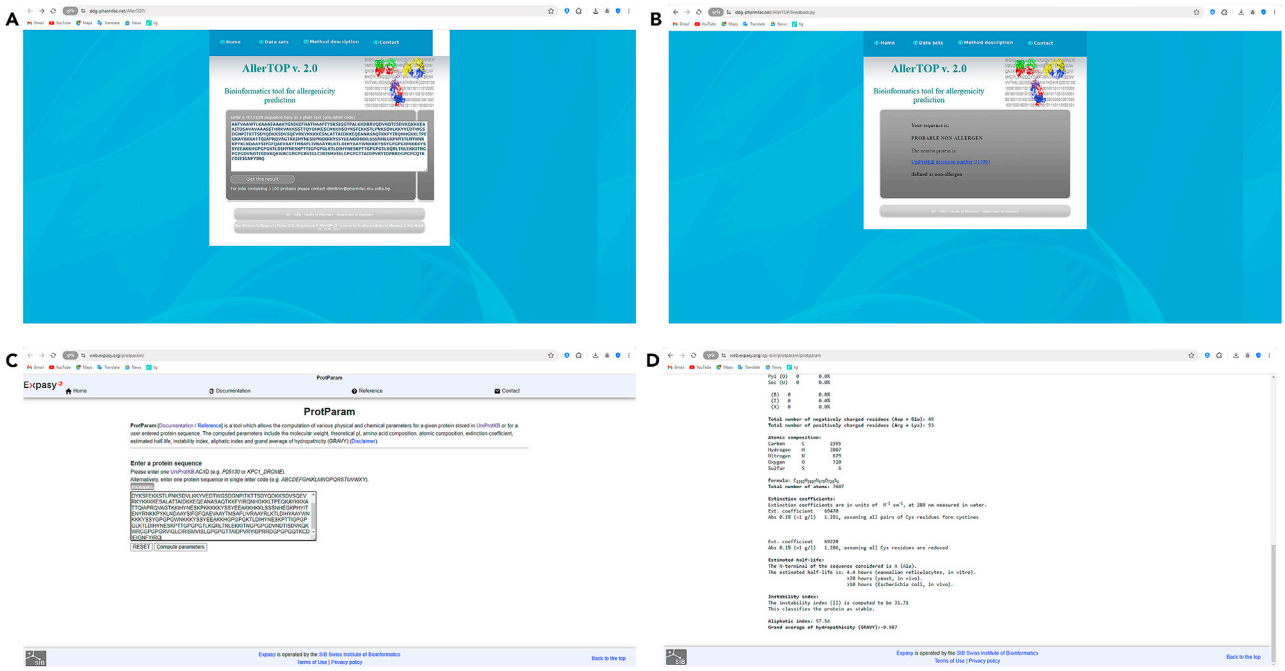


Figure 23. Allergenicity and physiochemical characterization of designed vaccine sequence

(A) Vaccine allergenicity prediction.

(B) Vaccine allergenicity result.

(C) Physiochemical characterization of designed vaccine.

(D) Result of physiochemical characterization of vaccine.

c. Use the command for probability distribution (Figure 38C).

```
gmx analyze -f sasa.xvg -dist prob_sasa.xvg
```

d. Observe probability distribution curve using command (Figure 38D).

```
xmgrace prob_sasa.xvg
```

118. H-bond analysis.

a. Calculate the Hydrogen bond values, using the command line.

```
gmx hbond -f md_0_1.xtc -s md_0_1.tpr -num protein.xvg
```

Enter the value "1" and "13" to select "protein" and "SOL" option respectively (Figure 39A).

b. Command to view the H- bond graph (Figure 39B).

```
xmgrace protein.xvg
```

119. PCA (Principal component analysis).

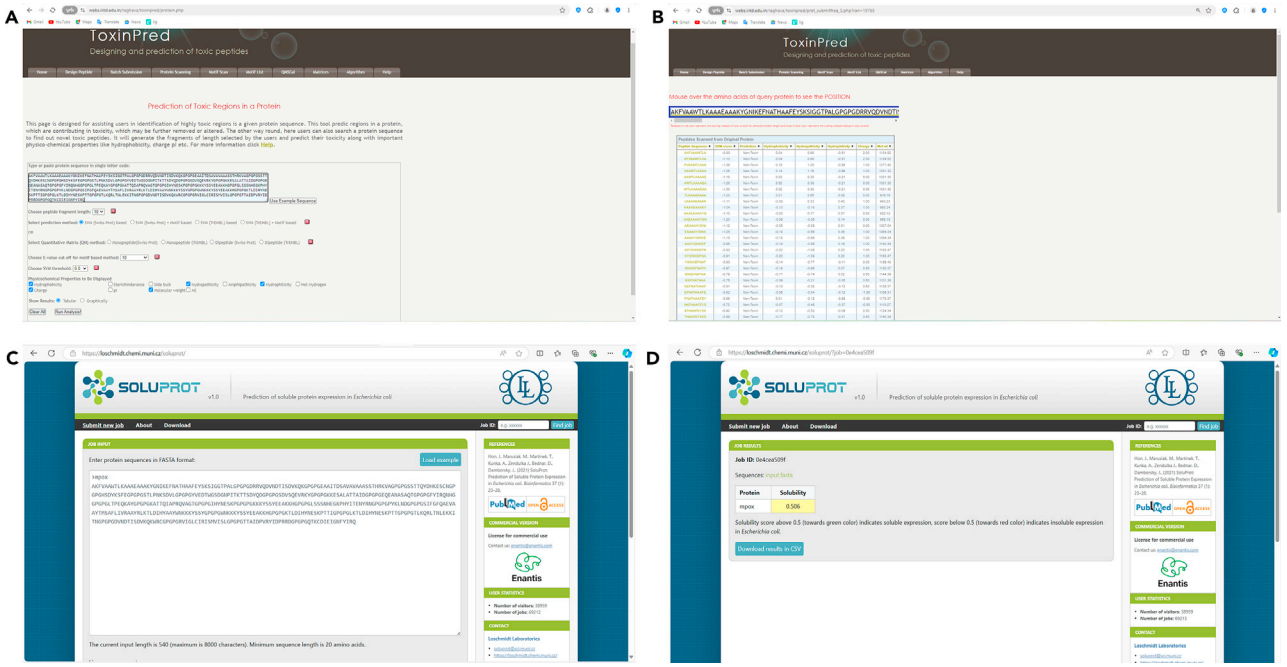


Figure 24. Toxicity and solubility prediction using ToxinPred and SOLUPROT of designed vaccine, respectively

(A) Toxicity prediction of desired vaccine.

(B) Result of Toxicity prediction.

(C) Prediction of solubility of designed vaccine candidate (D) Result of solubility prediction.

a. Execute the PCA and its graphical representation, following the command.

```
gmx make_ndx -f npt.gro -o c_alpha.ndx
```

Enter the value "3" to select C-alpha (Figure 40A).

b. Run the following command for covariance analysis.

```
gmx covar -f md_0_1.xtc -s md_0_1.tpr -n c_alpha.ndx -o eigenval.xvg -v eigenval.tpr -l covar.log -xpm covar.xpm
```

Enter the value "19" to select C-alpha (Figures 40B and 40C).

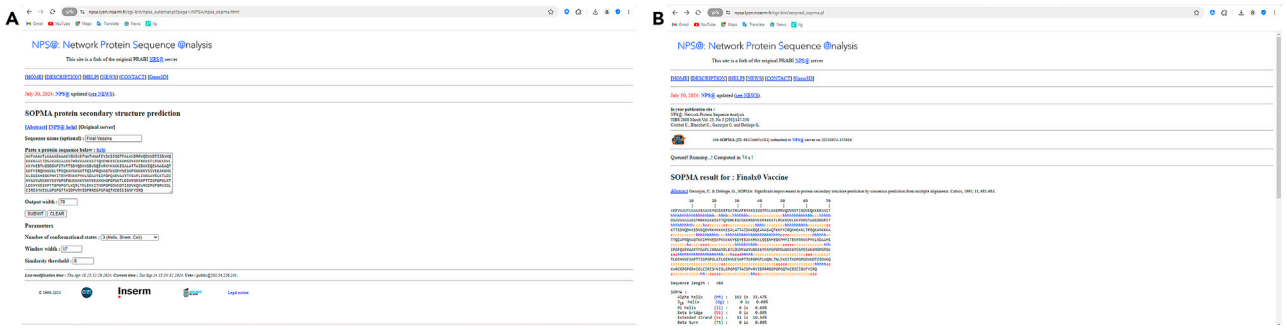


Figure 25. Secondary structure prediction of designed candidate

(A) Prediction of secondary structure using SOPMA.

(B) Result of secondary structure prediction.

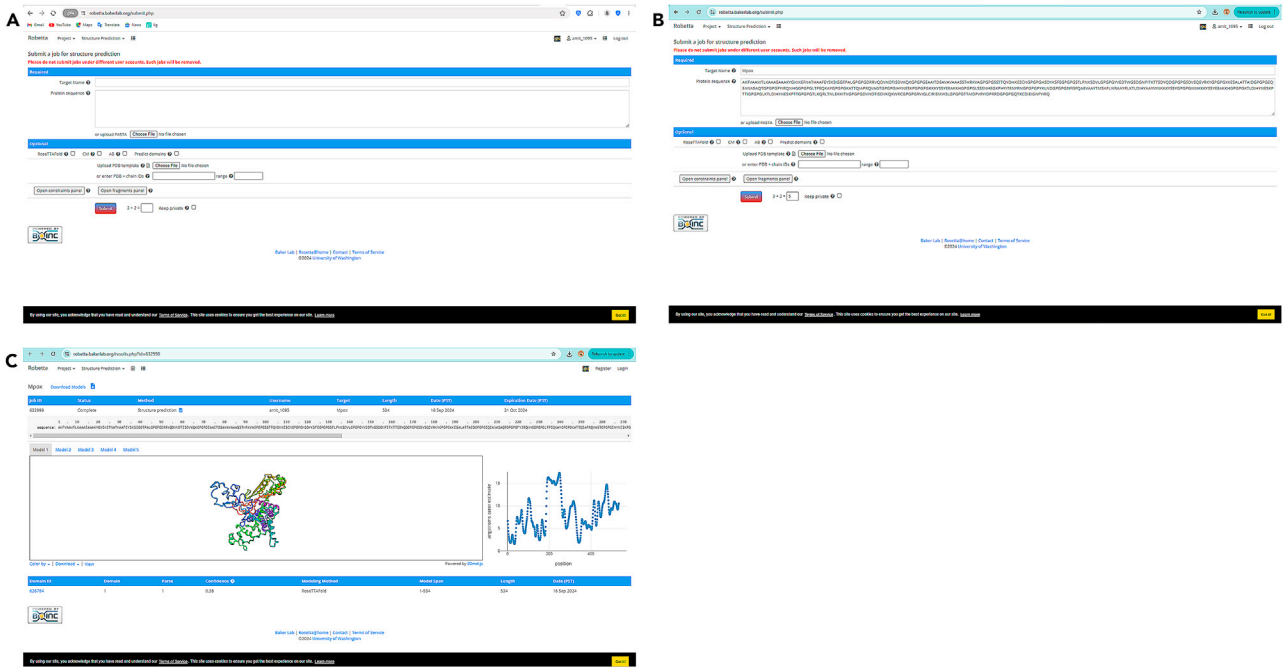


Figure 26. Tertiary structure prediction of designed vaccine using Robetta

(A) Homepage of structure prediction software Robetta.

(B) Screenshot of entered protein sequence.

(C) Results showing structure of predicted models in Robetta.

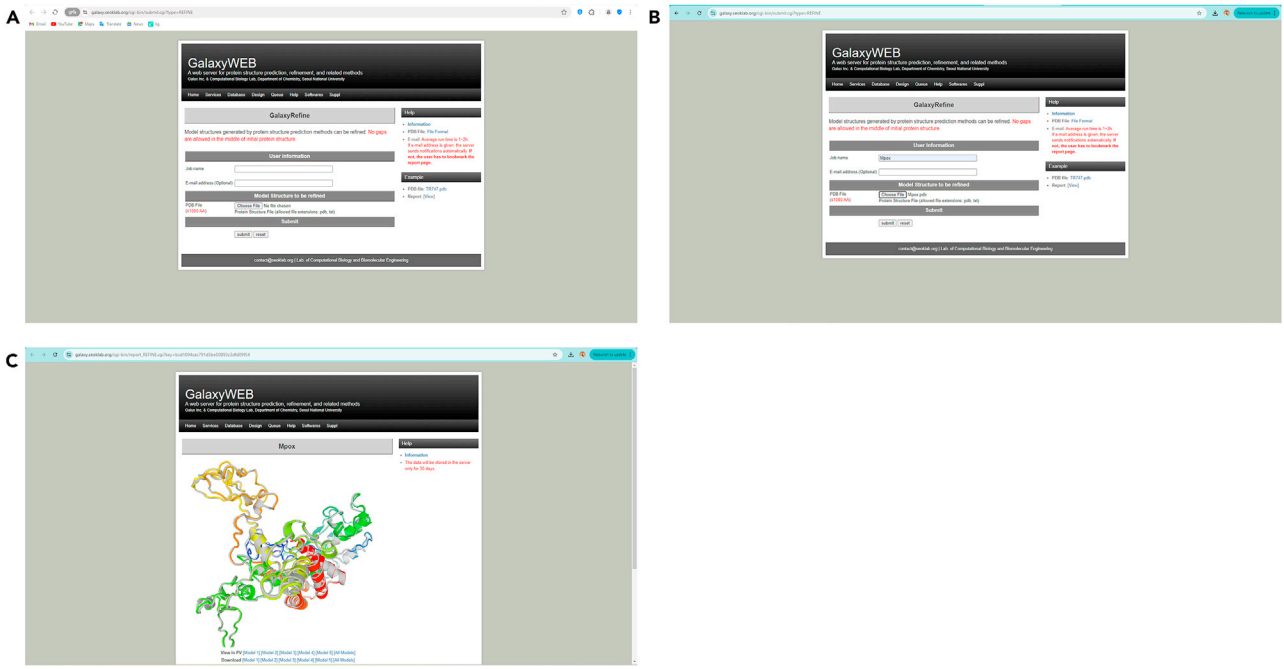


Figure 27. Refining of tertiary structure by GalaxyRefine

(A) Home page of GalaxyRefine webserver.

(B) Uploaded pdb file in GalaxyWEB server.

(C) Result showing refined structure of vaccine.

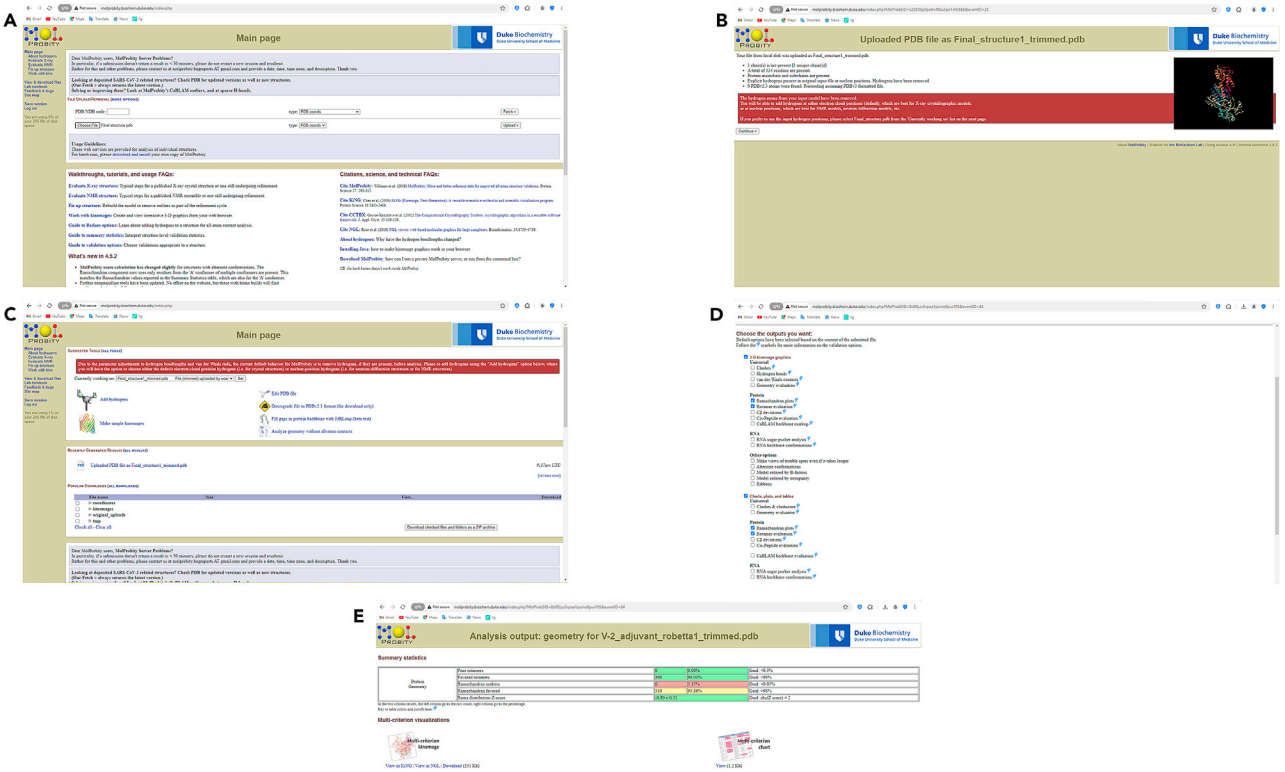


Figure 28. Evaluation of Ramachandran plot using MolProbity

(A) Homepage containing the refined predicted vaccine structure.

(B) Screenshot showing uploaded pdb file.

(C) Web page showing trimmed uploaded pdb file.

(D) Input parameters.

(E) Analysis output.

c. Extract the data from "eigenval.xvg" file in notepad, copy and paste it in an Excel sheet (Figure 41A).

d. Auto sum the eigen values to evaluate the trace value (Figure 41B).

e. Plot the graph in GraphPad Prism by designating the x and y axis as eigenvector index and eigen values (Figures 41C and 41D).

f. Evaluate the trajectory projections onto eigenvectors, follow the commands.
For eigenvector 1

```
gmx anaeig -v eigenval.trr -f md_0_1.xtc -s md_0_1.tpr -n c_alpha.ndx -eig eigenval.xvg -rmsf
rmsf_ev1.xvg -proj proj_ev1.xvg -first 1 -last 1
```

Select index group, enter "19" to select "C-alpha" option for both 'least squares fit in g_covar' and 'elements that corresponds to the eigenvectors' respectively (Figures 42A and 42B).
For eigenvector 2

```
gmx anaeig -v eigenval.trr -f md_0_1.xtc -s md_0_1.tpr -n c_alpha.ndx -eig eigenval.xvg -rmsf
rmsf_ev1.xvg -proj proj_ev12.xvg -first 2 -last 2
```

CellPress
OPEN ACCESS

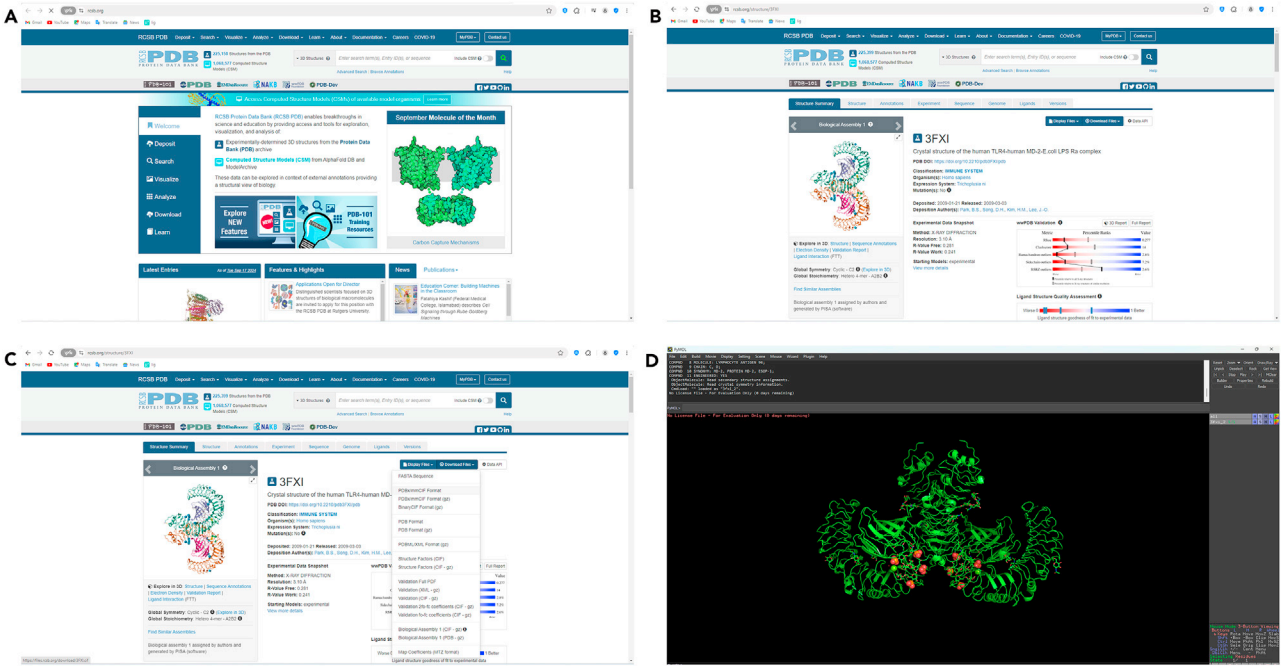


Figure 29. 3D structure of receptor

(A) PDB homepage.

(B) Crystal structure of TLR4.

(C) Steps to download receptor.

(D) Visualization of downloaded receptor file in PyMOL.

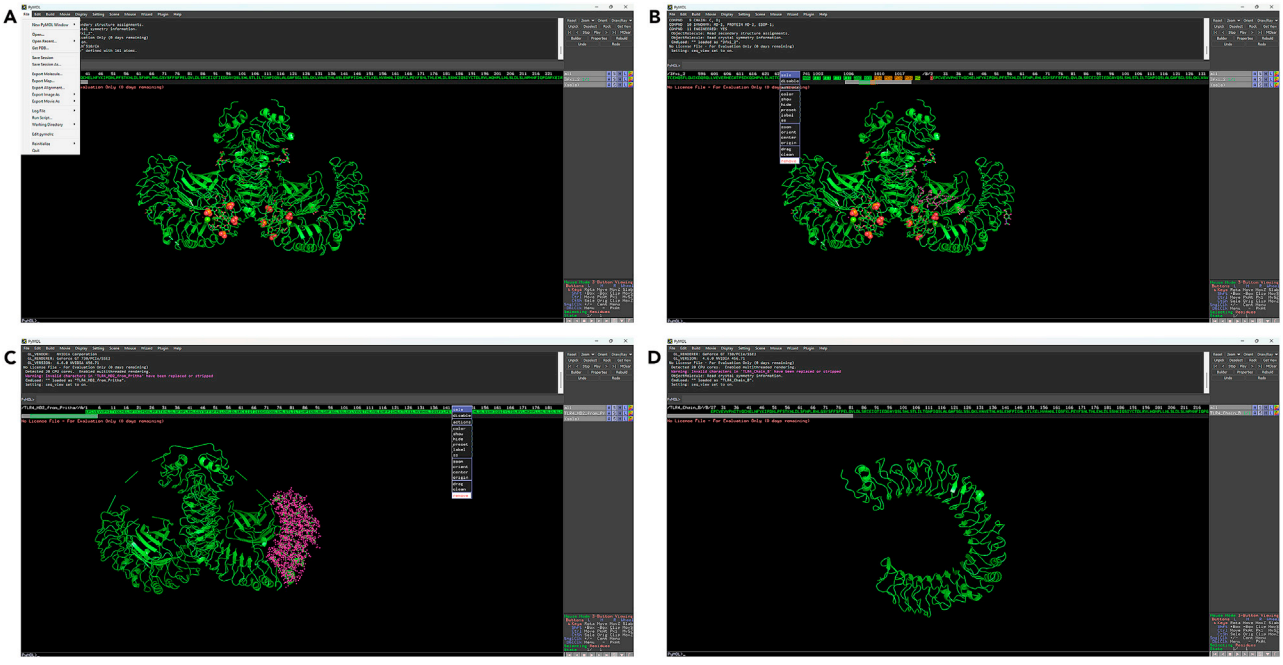


Figure 30. Preparation of receptor for docking

(A) Display sequence.

(B) Additional molecule removal.

(C) Removal of extra arm.

(D) Cleaned receptor file.

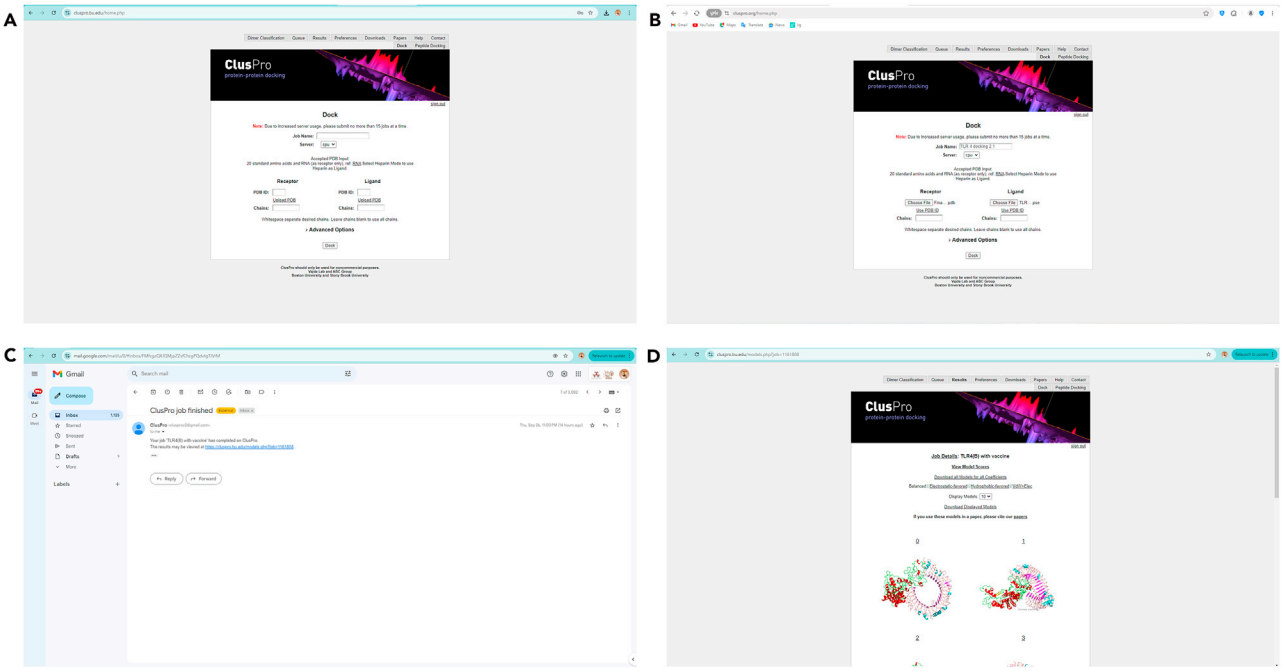


Figure 31. Molecular docking using ClusPro

(A) Homepage of ClusPro.

(B) Uploaded receptor-ligand files.

(C) Screenshot of mail showing result intimation.

(D) Result page showing displayed models.

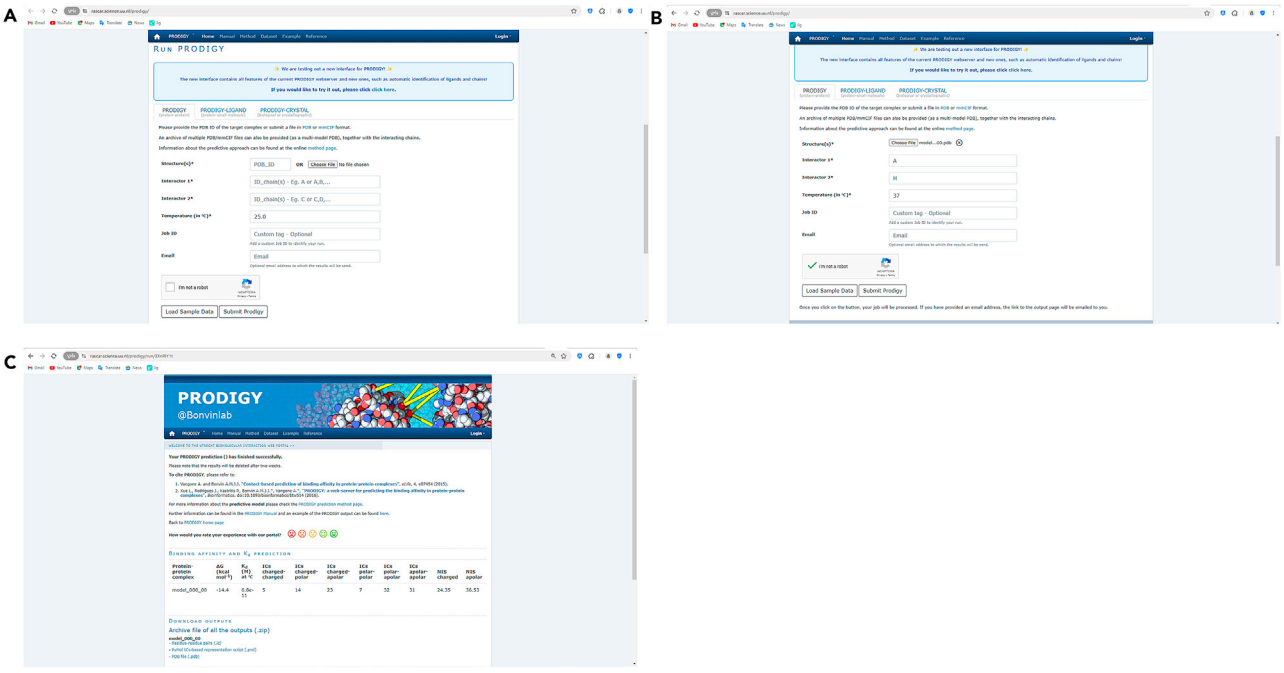


Figure 32. Validation of protein-protein interaction in PRODIGY server

(A) Homepage of PRODIGY web server.

(B) Uploaded files and parameters for submission.

(C) Result page of selected molecules.

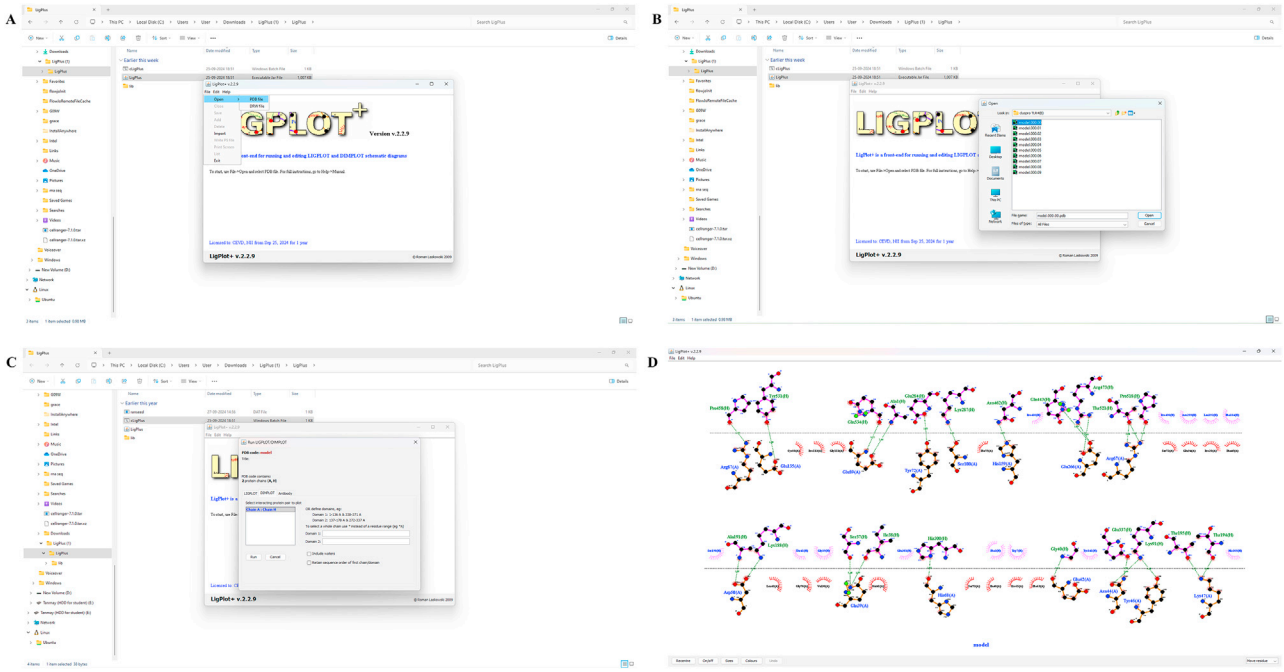


Figure 33. Analysis of interaction sites in protein-protein docked file Ligplot+

(A) Start page of Ligplot+ software (B) Selection of docked model (C) Chain and parameters selection (D) Result page showing intra-molecular interaction between amino acids.

A

```
Select the Force Field:
From current directory:
 1: CHARMM36 all-atom force field (July 2022)
From '/usr/local/gromacs/share/gromacs/top':
 2: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
 3: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
 4: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
 5: AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
 6: AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
 7: AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
 8: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
 9: CHARMM27 all-atom force field (CHARM22 plus CMAP for proteins)
10: GROMOS96 43a1 force field
11: GROMOS96 43a2 force field (improved alkane dihedrals)
12: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
13: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
14: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
15: GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40,, 843-856, DOI: 10.1007/s00249-011-0700-9)
16: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
```

B

```
Command line:
  gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname NA -nname CL -neutral -conc 0.15

Reading file ions.tpr, VERSION 2022 (single precision)
Reading file ions.tpr, VERSION 2022 (single precision)
Will try to add 289 NA ions and 290 CL ions.
Select a continuous group of solvent molecules
Group     0 (         System) has 314180 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 297396 elements
Group    12 (          Water) has 297396 elements
Group    13 (            SOL) has 297396 elements
Group    14 (      non-Water) has 16784 elements
Select a group: 13
Selected 13: 'SOL'
Number of (3-atomic) solvent molecules: 99132
```

C

```
;       Command line:
;         gmx pdb2gmx -f xyzpdb -o protein.gro -ignh -water tip3p
;       Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "amber99sb-ildn.ff/forcefield.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Protein_chain_B.itp"

; Include water topology
#include "amber99sb-ildn.ff/tip3p.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
;  i funct       fcx        fcy        fcz
   1    1       1000       1000       1000
#endif

; Include ligand topology
#include "drg.itp"
```

Before

```
;       Command line:
;         gmx pdb2gmx -f xyzpdb -o protein.gro -ignh -water tip3p
;       Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "amber99sb-ildn.ff/forcefield.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Protein_chain_B.itp"

; Include water topology
#include "amber99sb-ildn.ff/tip3p.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
;  i funct       fcx        fcy        fcz
   1    1       1000       1000       1000
#endif

; Include ligand topology
#include "drg.itp"

; Ligand position restraints
#ifdef POSRES
#include "posre.itp"
#endif
```

After

Figure 34. MD simulation in GROMACS

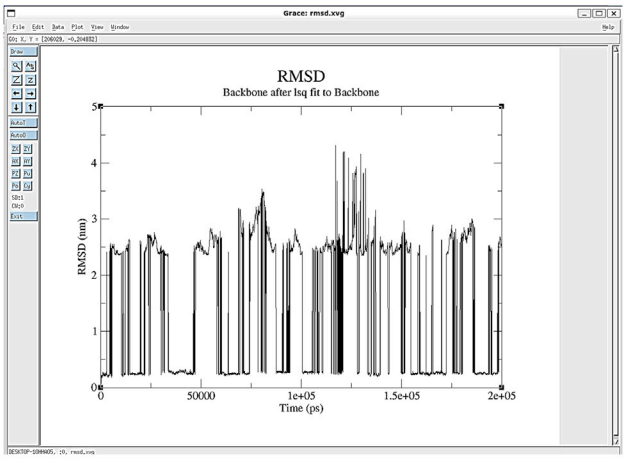
(A) Screenshot for force field selection (B) Screenshot for options to add ions in system.
(C) Screenshot of notepad file to retrain the ligand.

A

```
Command line:
  gmx rms -f md_0_1.xtc -s md_0_1.tpr -o rmsd.xvg

Reading file md_0_1.tpr, VERSION 2022 (single precision)
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Select group for least squares fit
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (       non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 4
Selected 4: 'Backbone'
Select group for RMSD calculation
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (       non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 4
Selected 4: 'Backbone'
Last frame       1000 time 50000.000
```

B



C

```
Command line:
  gmx analyze -f rmsd.xvg -dist prob_rmsd.xvg

Read 1 sets of 4001 points, dt = 50

                                                 std. dev.    relative deviation of
                          standard               ---------    cumulants from those of
set      average          deviation              sqrt(n-1)    a Gaussian distribition
                                                                 cum. 3    cum. 4
SS1   1.483963e+00   1.189761e+00   1.881177e-02       0.000    -0.607

Making distributions with 46 bins

Back Off! I just backed up prob_rmsd.xvg to ./#prob_rmsd.xvg.1#
```

Figure 35. RMSD determination of predicted MPXV vaccine

Time-Dependent RMSD graph and probability distribution analysis for structural stability assessment.

(A) Screenshot for backbone selection to calculate RMSD.

(B) RMSD graph.

(C) Probability distribution command.

Select index group, enter "19" to select "C-alpha" option for both 'least squares fit in g_covar' and 'elements that corresponds to the eigenvectors' respectively (Figure 42C and 42D).

g. Open the proj_evl.xvg and proj_evl2.xvg files, and copy the data into Excel (Figure 43A).

h. Remove the time frame (first column) and keep only the data from the second column in each file.

i. Paste this data into GraphPad Prism to create a scatter plot (Figure 43B and 43C).

Note: Here we have assigned the data from proj_eval as PC1 and from proj_eval2 as PC2.

j. Generate the MD simulation trajectory movie, executing the commands one after the other. Remove the PBC artifacts and center the molecule in the box to create a "cleaned" trajectory by running the command.

```
gmx trjconv -f md_0_1.xtc -s md_0_1.tpr -o md_pbc.pdb -pbc mol -center
```

k. Convert the trajectory to a PDB format movie based on specific groups from the index file.

```
gmx trjconv -f md_ pbc.xtc -s md_0_1.tpr -o movie.pdb -n index.ndx
```

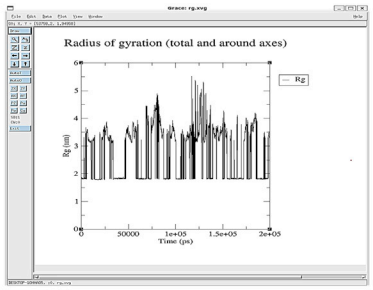
A

```
Executable:   /usr/local/gromacs/bin/gmx
Data prefix:  /usr/local/gromacs
Working dir:  /mnt/d/Anupam/Monkey_pox_data/FINAL VACCINE/MD simulation/TLR4-V2
Command line:
  gmx gyrate -f md_0_1.xtc -s md_0_1.tpr -o rg.xvg

Reading file md_0_1.tpr, VERSION 2022 (single precision)
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (     SideChain) has 11341 elements
Group     9 (   SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (      non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 1
Selected 1: 'Protein'
Reading frame       0 time    0.000
Back Off! I just backed up rg.xvg to ./#rg.xvg.1#
Last frame       1000 time 50000.000

GROMACS reminds you: "Years of calculations and the stress, My science is waiting, nearly complete" (Midlake)
```

B



C

```
Command line:
  gmx analyze -f rg.xvg -dist prob_rg.xvg

Read 4 sets of 1001 points, dt = 50

                                                 std. dev.    relative deviation of
                                                 ---------    cumulants from those of
set      average          standard deviation     sqrt(n-1)    a Gaussian distribition
                                                                cum. 3    cum. 4
SS1   3.454537e+00   5.722499e-02   1.809613e-03      0.119    -0.217
SS2   2.825511e+00   6.144906e-02   1.943190e-03     -0.774     0.782
SS3   2.500846e+00   8.315981e-02   2.629744e-03      0.203    -0.237
SS4   3.100749e+00   1.054625e-01   3.335017e-03     -0.466     0.194

Making distributions with 16 bins
```

D

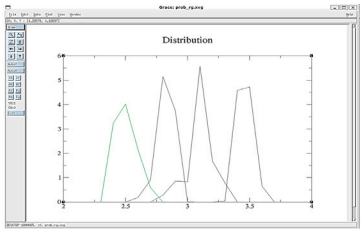


Figure 36. Determination of Rg

(A) Calculation of Rg.

(B) R_g graph.

(C) Probability distribution command.

(D) Probability distribution graph.

A

```
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Select group(s) for root mean square calculation
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (     SideChain) has 11341 elements
Group     9 (   SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (      non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 3
Selected 3: 'C-alpha'
Last frame       1000 time 50000.000
```

B

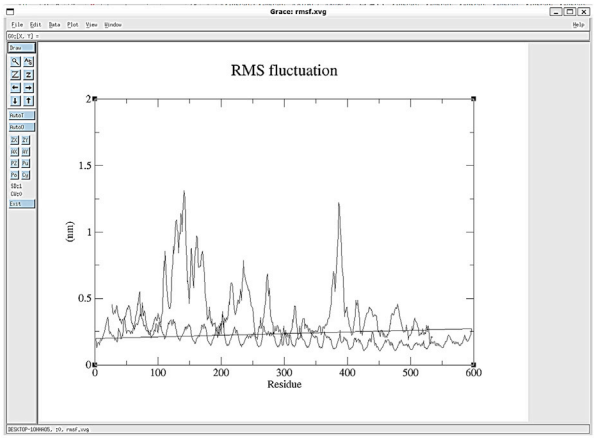


Figure 37. Determination of RMSF

Analysis of RMSF and its graph of predicted vaccine candidate.

(A) Screenshot to select C-alpha option RMSF.

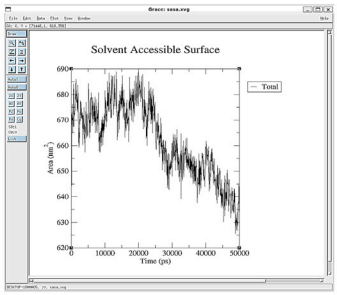
(B) RMSF graph.

A

```
Command line:
  gmx sasa -f md_0_1.xtc -s md_0_1 -o sasa.xvg

Reading file md_0_1.tpr, VERSION 2022 (single precision)
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Available static index groups:
 Group  0 "System" (313022 atoms)
 Group  1 "Protein" (16784 atoms)
 Group  2 "Protein-H" (8468 atoms)
 Group  3 "C-alpha" (1104 atoms)
 Group  4 "Backbone" (3312 atoms)
 Group  5 "MainChain" (4418 atoms)
 Group  6 "MainChain+Cb" (5408 atoms)
 Group  7 "MainChain+H" (5443 atoms)
 Group  8 "SideChain" (11341 atoms)
 Group  9 "SideChain-H" (4050 atoms)
 Group 10 "Prot-Masses" (16784 atoms)
 Group 11 "non-Protein" (296238 atoms)
 Group 12 "Water" (295659 atoms)
 Group 13 "SOL" (295659 atoms)
 Group 14 "non-Water" (17363 atoms)
 Group 15 "Ion" (579 atoms)
 Group 16 "NA" (289 atoms)
 Group 17 "CL" (290 atoms)
 Group 18 "Water_and_ions" (296238 atoms)
Specify a selection for option 'surface'
(Surface calculation selection):
(one per line, <enter> for status/groups, 'help' for help)
> 8
Selection '8' parsed
```

B



C

```
Command line:
  gmx analyze -f sasa.xvg -dist prob_sasa.xvg

Read 1 sets of 1001 points, dt = 50

                                                 std. dev.    relative deviation of
                      standard                   ---------   cumulants from those of
set      average      deviation                  sqrt(n-1)   a Gaussian distribition
                                                                cum. 3    cum. 4
SS1   6.619976e+02   1.433658e+01   4.533625e-01                -0.167    -0.305
```

D

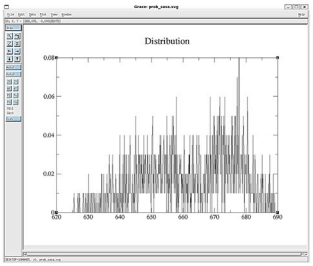


Figure 38. Determination of solvent accessibility by SASA analysis

(A) Selection of side chain for SASA analysis.

(B) SASA graph.

(C) Probability distribution command of SASA.

(D) Probability distribution graph of SASA.

A

```
Command line:
  gmx hbond -f md_0_1.xtc -s md_0_1.tpr -num protein.xvg

Reading file md_0_1.tpr, VERSION 2022 (single precision)
Specify 2 groups to analyze:
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (      non-Water) has 17363 elements
Group    15 (            Ion) has   579 elements
Group    16 (             NA) has   289 elements
Group    17 (             CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 1
Selected 1: 'Protein'
Select a group: 13
Selected 13: 'SOL'
Checking for overlap in atoms between Protein and SOL
Calculating hydrogen bonds between Protein (16784 atoms) and SOL (295659 atoms)
```

B

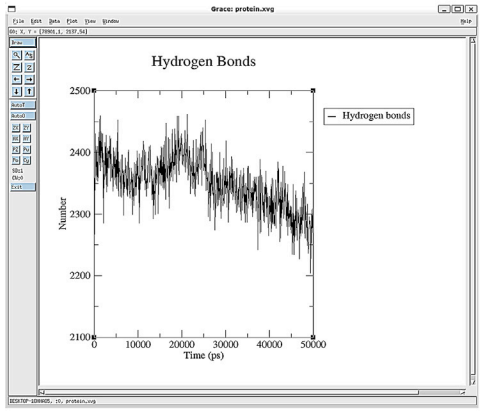


Figure 39. H-bond analysis

(A) Selection of protein and SOL groups for H-bond determination.

(B) H-bond graph.

120. Clustering.
    a. Perform clustering on the generated trajectory movie.

```
gmx cluster -f md_0_1.xtc -s md_0_1.tpr -o cluster15_4.xpm -g cluster15_4.log -sz clus-
ter15_4.xvg -cl cluster15_4.pdb -cutoff
```

    b. Select group for least square fit and RMSD calculation, enter value "4" to select ''backbone'' option (Figures 44A and 44B).
    c. Copy and paste the data from the Cluster15_4.xvg file to an Excel sheet and calculate the percentage of each cluster # and arrange them in decreasing order on the basis of their # structures (Figures 45A and 45B).
    d. The cluster # with the highest percentage of structure # is microstate m1. Open the cluster15_4.pdb to fetch the m1 microstate.
    e. Edit the chain color for better visualization in PyMOL and save it as m1.pdb (Figures 45C and 45D).

Part 11: Backtranslation and codon optimization

Timing: 3 h

This section details the procedure to perform and analyze backtranslation and codon optimization of the designed vaccine sequence to perform *in silico* cloning in subsequent steps.

121. Convert the protein to nucleotide sequence using EMBOSS Backtranseq webserver (Figure 46A).

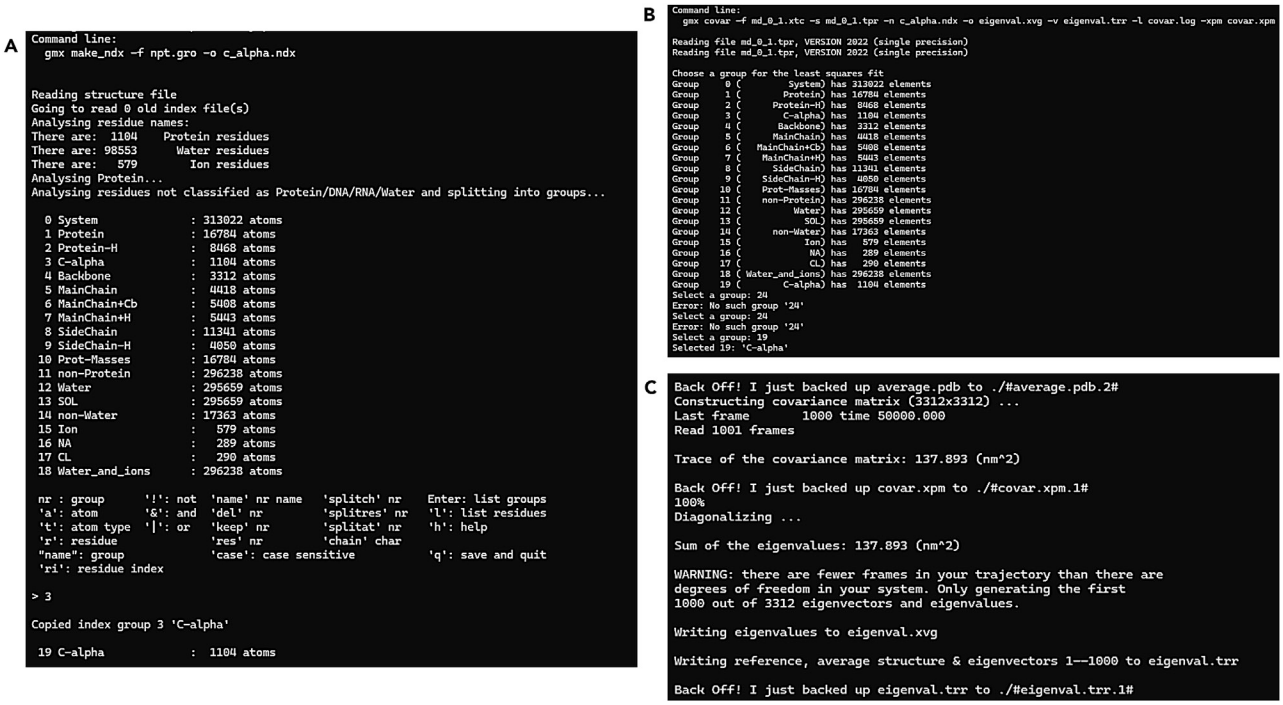


Figure 40. Commands for PCA plot

(A) Screenshot for selection of C-alpha for PCA plot generation.
(B) Screenshot of C-alpha selection for covariance analysis.
(C) Screenshot showing sum of eigen values.

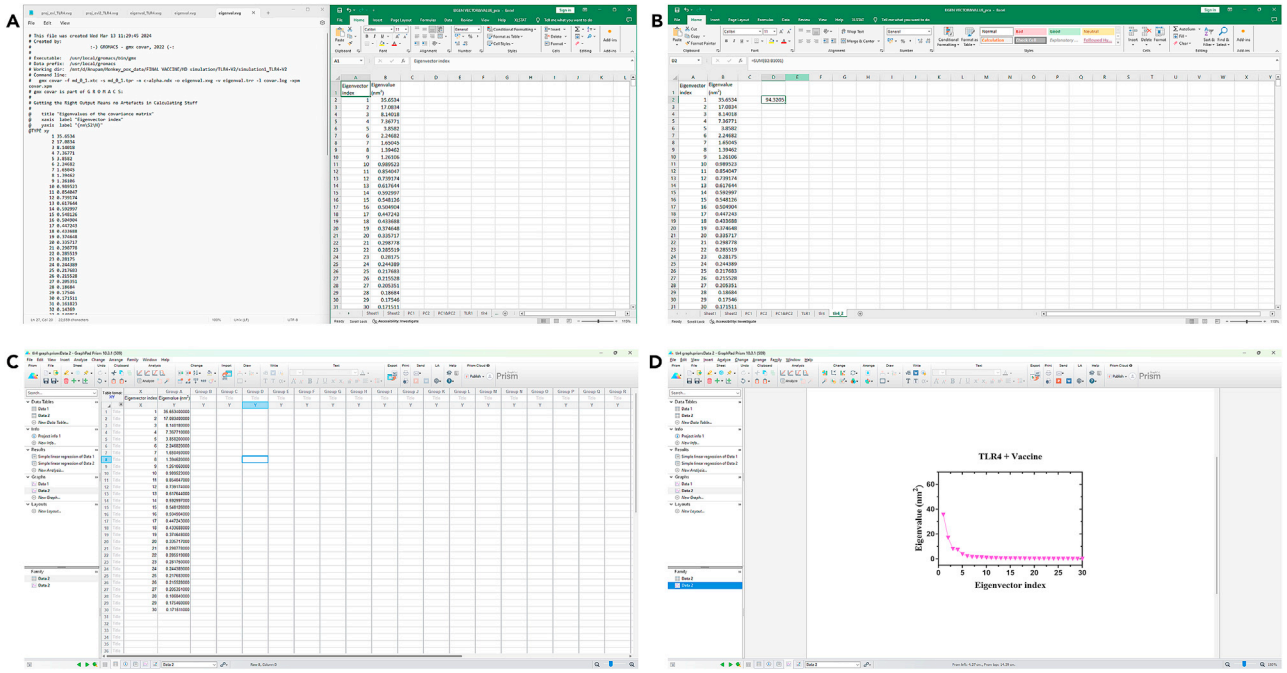


Figure 41. PCA of the designed vaccine

PCA analysis was performed and generated a movie of the MD simulation trajectory.
(A) Data extraction from eigen value files.
(B) Auto sum of eigen values.
(C) Plot of eigen values in graph pad prism.
(D) Graph of eigen values vs. eigen vector index.

122. Submit the protein sequence in FASTA format in the "Input sequence" column of the webserver.
123. Choose the "*Escherichia coli* K12" option from the dropdown list of "codon table" and "Submit" to get results in DNA format (Figures 46B and 46C).
124. Copy the viewed result (vaccine construct sequence) in a Word file (Figure 46D).
125. Access the Codon Optimization Tool from the VectorBuilder webserver (Figure 47A).
126. Paste the DNA sequence for the vaccine construct, selecting "DNA/RNA sequence," and choose "*Escherichia coli* K-12" as the organism (Figure 47B).
127. Click on submit button. Your results will be displayed in FASTA format. Save the sequence for further analysis (Figure 47C).

Part 12: *In silico* cloning

Timing: 2 h

This section details the procedure to perform *in silico* cloning of the optimized sequence in pET-28a.

128. Download and install SnapGene (Figure 48A).
129. Open the application - go to new - select DNA file - (Figure 48B).
130. Paste your sequence (codon-optimized)- Click on create (Figures 48C and 48D).

Note: If no restriction is available at start and end sites create restriction sites at starting and ending positions. Here, since the first nucleotide is G we added NcoI site at start position (CCATGG) and we added BamHI site (GGATCC) as the last nucleotide was G.

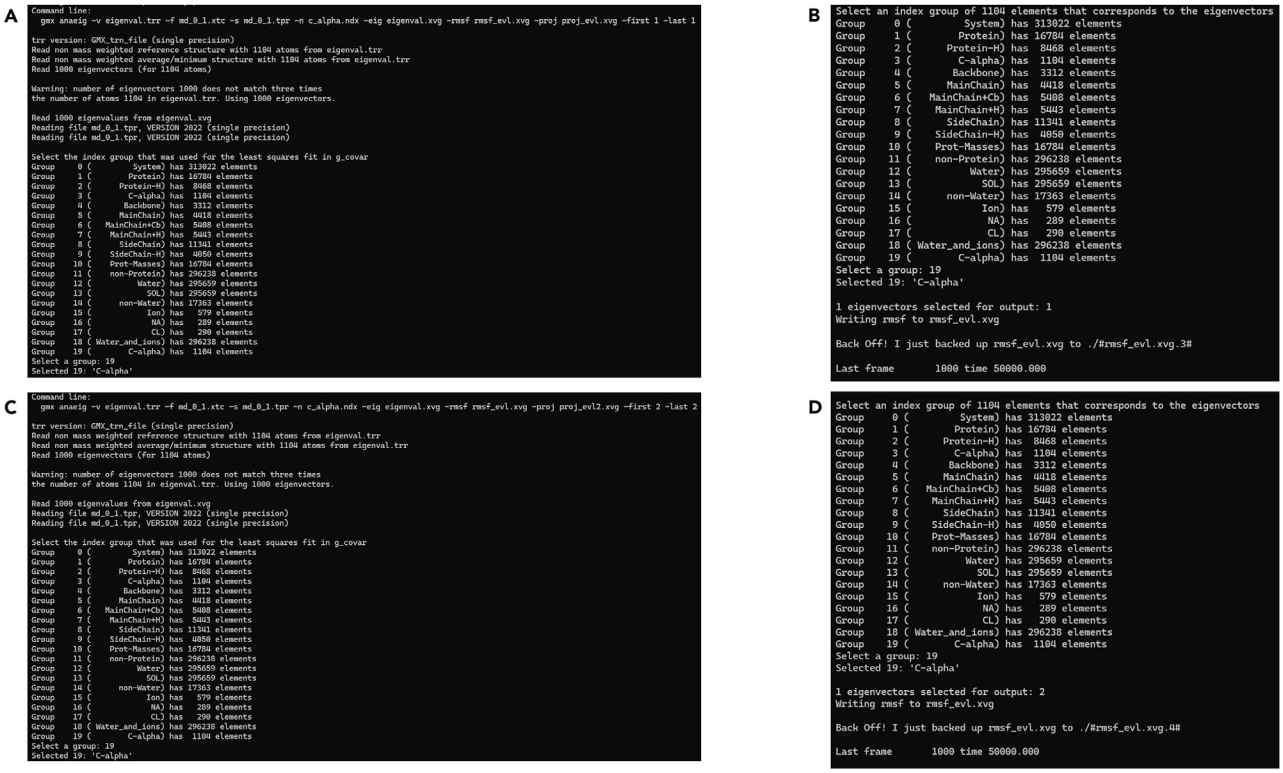


Figure 42. Generation of eigen vector 1 and 2 values
(A and B) Selection of index group for eigen vector 1.
(C and D) Selection of index group for eigen vector 2.

131. Save the file in SnapGene extention format: filename.dna (here mpox.dna) (Figure 18E).
132. Download the vector of your interest from SnapGene (here pET-28a+ is used) and save it as a new sequence file pET-28a(+).dna (Figure 49A).
133. Open pET-28a (+).dna file. Go to actions- click on restriction cloning- select insert fragment (Figures 49B and 49C).
134. Three tabs will open: Vector, Fragment, Product.
135. Add restriction sites of your interest (in this case NcoI and BamHI) in the dialogue box of Vector tab and select the region to replace (Figure 49D).
136. Go to fragment tab and add the same restriction enzymes that are used to cut the vector and select the region to replace (Figures 50A and 50B).
137. Name the clone (Cloned.dna) and select clone in the fragment tab (Figure 50C).
138. Go to product tab and save the clone (Figure 50D).

Part 13: Immune simulation

Timing: 30 min

This section enlists the procedure to perform *in silico* immune simulation of the designed vaccine.

139. Utilize C-ImmSim server to investigate the designed multi-epitope vaccine's immunogenic response (Figure 51A).
140. Open the C-ImmSim site and paste the designed protein sequence in "Protein N.1" column.
141. Time step of injection should be 1. Rest all parameters will be default.

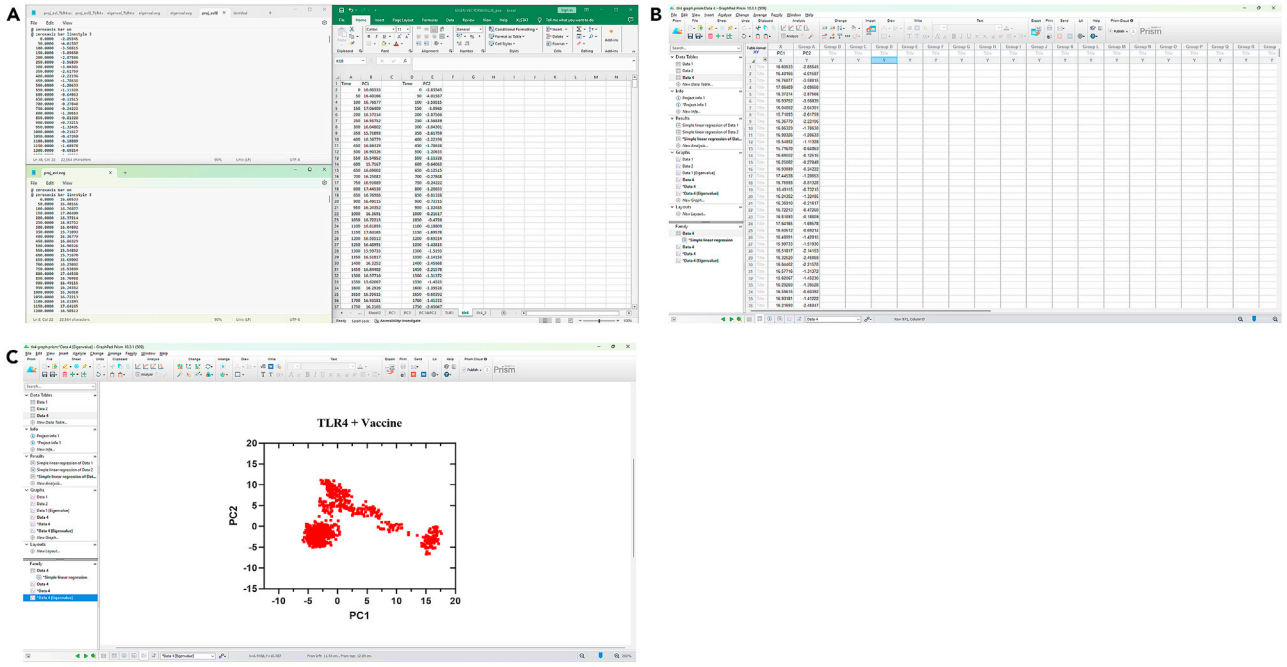


Figure 43. Generation of PCA plot
(A and B) Extraction of PC1 and PC2 values.
(C) Graph PC1 vs. PC2.

142. For booster doses (Injection N. 2), click "Add Ag Molecule" and keep all parameters at their default settings, except for the injection time step.
143. Set injection time step to 31 (the day of the booster dose) (Figure 51B).
144. Click on "SUBMIT JOB" and the results will be displayed (Figure 51C).

EXPECTED OUTCOMES

The protocol outlines a roadmap for constructing a multi-epitope vaccine against the MPXV by integrating distinct peptide epitopes with high immunogenicity and antigenicity. Six MPXV glycoprotein were used to predict seven T cell (CTL and HTL), thirteen B cell and five IFN-gamma specific epitopes. These epitopes have potential to bind to multi-MHC alleles and elicit higher immune response. The immune response was further enhanced by cohesively incorporating PADRE peptide as an adjuvant. To cohesively join the adjuvant, epitopes, and linkers (EAAAK, GPGPG and AAY) were used to construct a multi epitope-based peptide subunit vaccine. 3D structure prediction of the constructed vaccine was performed along with physicochemical characterization. Further molecular docking and MD simulation were applied to study the interaction between the vaccine and receptors (TLR-family, MHC class I and II) at atomic level. These studies predict an enhanced immune response through increased activation of immune cells, elevated production of cytokines, and higher IgG and IgM production eliciting an amplified immune response after a booster dose administration. This study is insightful and highly valuable for utilizing a reverse vaccine technology approach to design and validate a novel, robust peptide-based multi-epitope vaccine against global outbreaks of rapidly mutating viruses.

LIMITATIONS

Reverse vaccine technology is a potent tool and multidisciplinary approach paired with experimental validation for designing a multi-epitope-based vaccine. The predictive algorithms may sometimes lead to inaccurate recognition of the epitopes. The variation in MHC molecules can influence the vaccine's potency among populations. Predictions usually consider a restricted set of HLAs for

A

```
(base) pwd@DESKTOP-1OHHAO5:/mnt/d/Anupam/Monkey_pox_data/FINAL VACCINE/MD simulation/TLR4-V2/simulation1_TLR4-V2$ gmx cluster -f md_0_1.xtc -s md_0_1.tpr -o cluster15_4.xpm -g cluster15_4.log -sz cluster15_4.xvg -cl cluster15_4.pdb -cutoff 0.15
                      :-) GROMACS - gmx cluster, 2022 (-:

Executable:   /usr/local/gromacs/bin/gmx
Data prefix:  /usr/local/gromacs
Working dir:  /mnt/d/Anupam/Monkey_pox_data/FINAL VACCINE/MD simulation/TLR4-V2/simulation1_TLR4-V2
Command line:
  gmx cluster -f md_0_1.xtc -s md_0_1.tpr -o cluster15_4.xpm -g cluster15_4.log -sz cluster15_4.xvg -cl cluster15_4.pdb -cutoff 0.15

Using linkage method for clustering
Reading file md_0_1.tpr, VERSION 2022 (single precision)
Reading file md_0_1.tpr, VERSION 2022 (single precision)

Select group for least squares fit and RMSD calculation:
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (      non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 4
Selected 4: 'Backbone'
```

B

```
Select group for output:
Group     0 (         System) has 313022 elements
Group     1 (        Protein) has 16784 elements
Group     2 (      Protein-H) has  8468 elements
Group     3 (        C-alpha) has  1104 elements
Group     4 (       Backbone) has  3312 elements
Group     5 (      MainChain) has  4418 elements
Group     6 (   MainChain+Cb) has  5408 elements
Group     7 (    MainChain+H) has  5443 elements
Group     8 (      SideChain) has 11341 elements
Group     9 (    SideChain-H) has  4050 elements
Group    10 (    Prot-Masses) has 16784 elements
Group    11 (    non-Protein) has 296238 elements
Group    12 (          Water) has 295659 elements
Group    13 (            SOL) has 295659 elements
Group    14 (      non-Water) has 17363 elements
Group    15 (             Ion) has   579 elements
Group    16 (              NA) has   289 elements
Group    17 (              CL) has   290 elements
Group    18 ( Water_and_ions) has 296238 elements
Select a group: 4
Selected 4: 'Backbone'
Last frame       1000 time 50000.000
Allocated 401411440 bytes for frames
Read 1001 frames from trajectory md_0_1.xtc
Computing 1001x1001 RMS deviation matrix
# RMSD calculations left: 0

The RMSD ranges from 0.0885534 to 0.831342 nm
Average RMSD is 0.381406
Number of structures for matrix 1001
Energy of the matrix is 14.1949.
WARNING: rmsd minimum 0 is below lowest rmsd value 0.0885534
Linking structures ****************************
Sorting and renumbering clusters
```

Figure 44. Clustering of MD trajectory

(A) Selection of group for least squares fit and RMSD calculation.

(B) Screenshot of average RMSD values for clustering.

the investigation that may bypass the significant variations crucial for wide-ranging vaccine efficacy. Selecting an immune response-enhancing adjuvant is both challenging and crucial, as not all adjuvants work effectively in combination with the chosen epitopes. *In vitro* and *in vivo* validation of the computationally predicted vaccine construct is required for successful translational studies to ensure its efficacy and safety. Uncertainty surrounds how long a multi-epitope vaccine will confer immunity, warranting more investigation to evaluate the immune response duration and booster requirements.

TROUBLESHOOTING

Problem 1

In the *in silico* cloning analysis, sometimes you may encounter the absence of unique restriction sites at the beginning and end of the peptide. If similar restriction sites are present at the start and end of the peptide in any other region of the peptide, it may result in the generation of mid-region cuts (Ref: Step-135).

Potential solution

In such cases, manually add the unique restriction site at the start and end regions of the peptide, looking for the ones that add the least number of amino acids, as the increased length may impact its structural stability.

Problem 2

During the molecular dynamic simulation analysis, during the addition of ions under the "Next add ions" commands after "`gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr`" command, you may encounter an issue as "Error: Atom type HS14 is not found." (Ref: Step-108).

Potential solution

To resolve this problem, run "Solution: open drg.itp- see "HS14" – change to "H"- Save" to convert it into a readable text format.

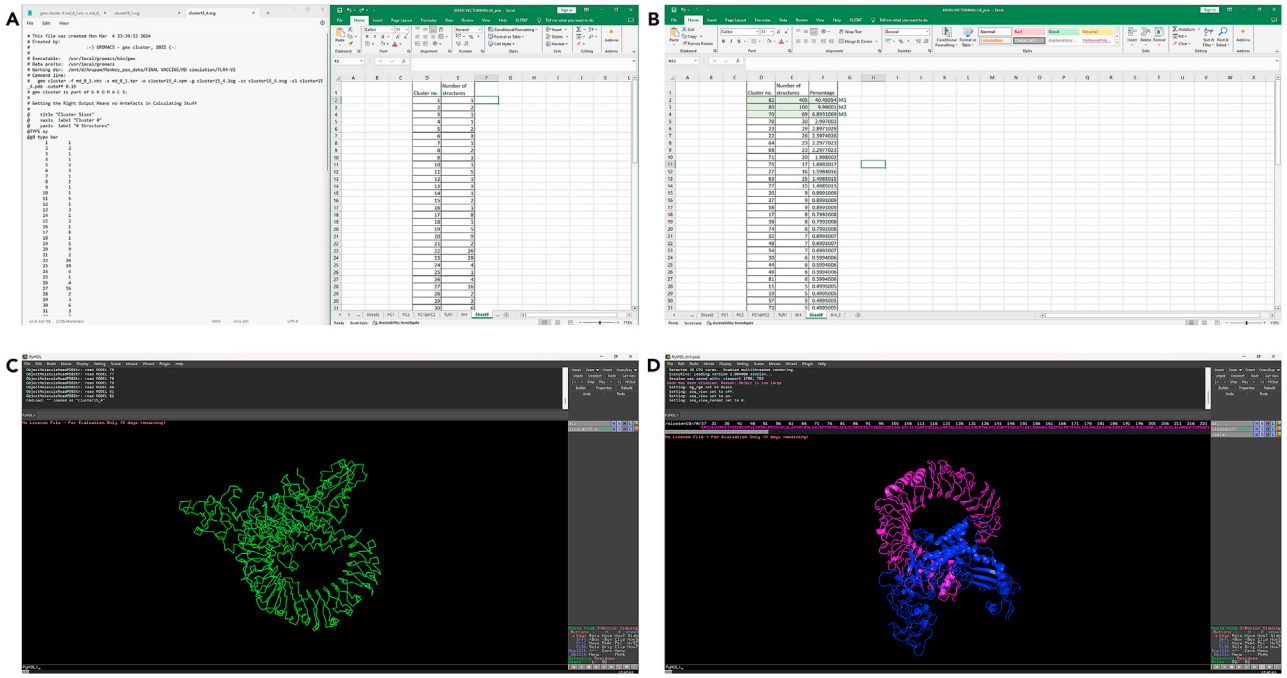


Figure 45. Analysis of clustered MD trajectory

(A) Extraction of no. of clusters from clustering files.

(B) Evaluation of microstates.

(C and D) Visualization of selected microstates in PyMOL.

Problem 3

At intervals, sudden spikes in the RMSD trajectory curve are encountered during MD simulation analysis (Ref: Step-114).

Potential solution

To eliminate the jump in RMSD, i.e., use the no jump command – `gmx tryconv -f md -f md - o - 1xtc - smd -0-1.tpr-0md-nojump.xtc-pbc.nojump.`

Problem 4

When the 3D predicted epitope and linker combinations don't pass the Ramachandran plot's quality score, i.e., greater than 95%, in that case, the structure is not suggested for the study to be taken forward (Ref: Step-68).

Potential solution

To eliminate this issue, manual reshuffling and annotation of linkers and epitopes are done to create different combinations to generate a 3D structure passing the verified by the Ramachandran plot quality score.

Problem 5

The docked protein file is required in a separate chain readable format to be read as ligand and protein in PRODIGY for validating the protein-protein interactions.

But after docking the two protein chains are labeled same as "A" (Ref: Step 100).

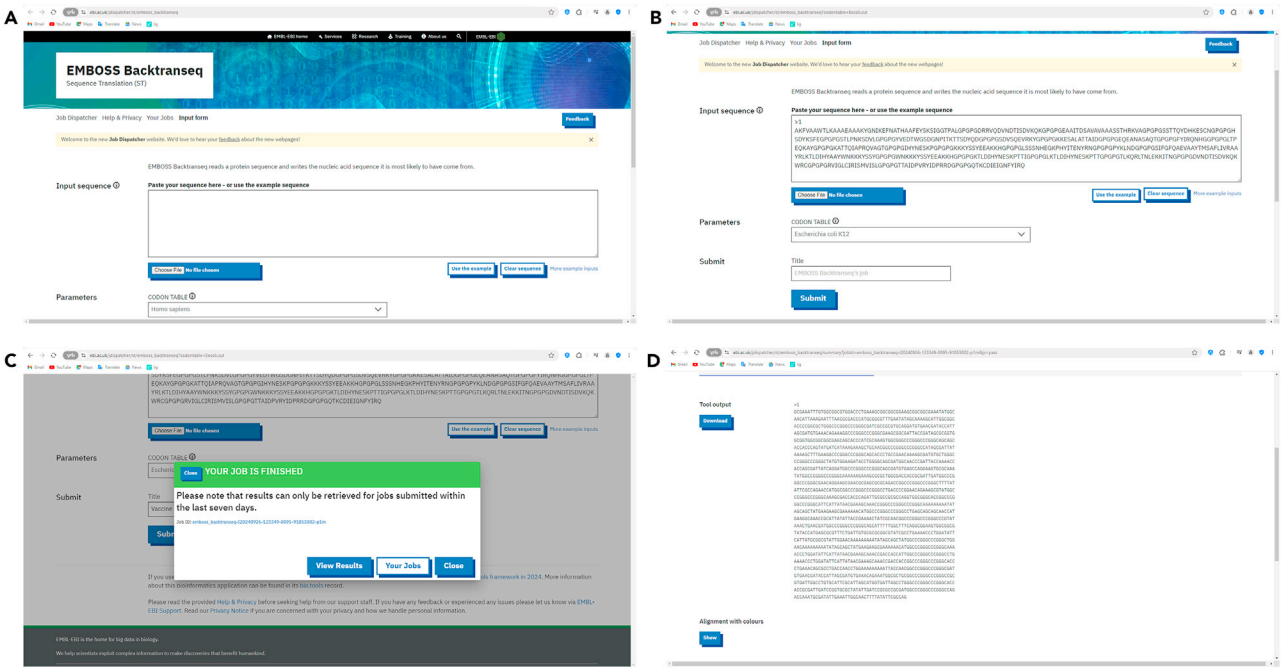


Figure 46. EMBOSS transeq based backtranslation

(A) Homepage of EMBOSS Back transeq server.

(B) Entered protein sequence and parameters.

(C) Submission of entered sequence.

(D) Output result.

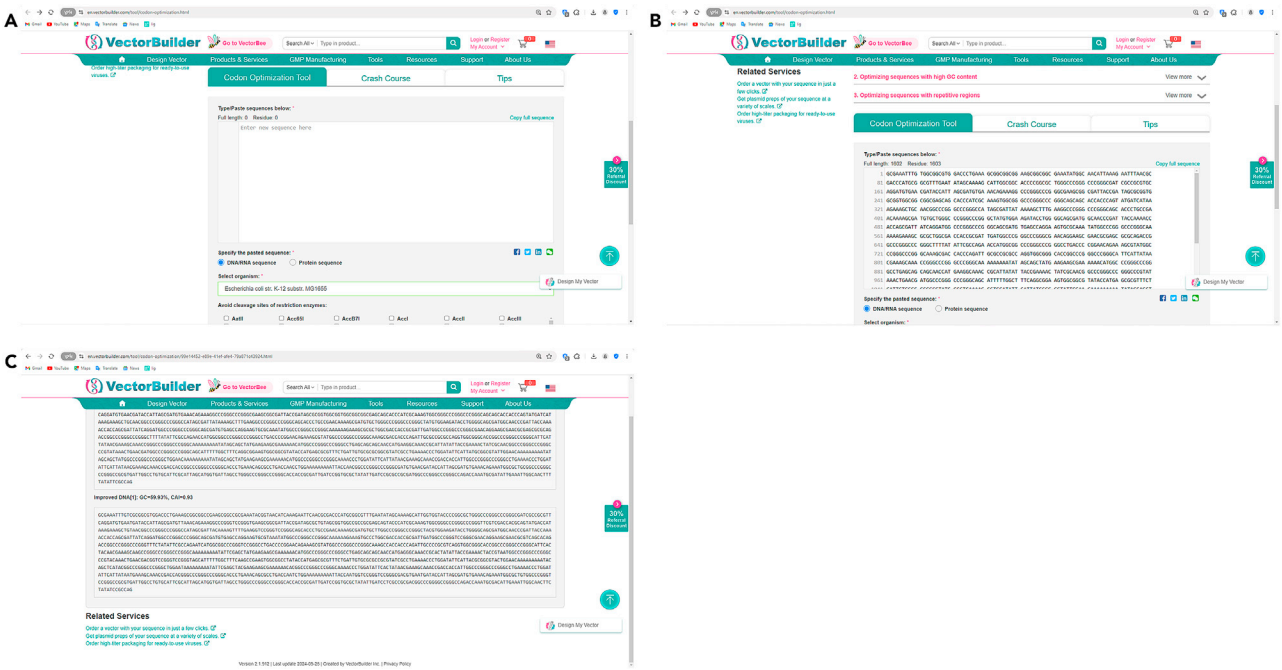


Figure 47. Codon optimization using VectorBuilder

(A) Homepage of VectorBuilder server.

(B) Entered DNA sequence in the dialog box.

(C) Output page of improved DNA.

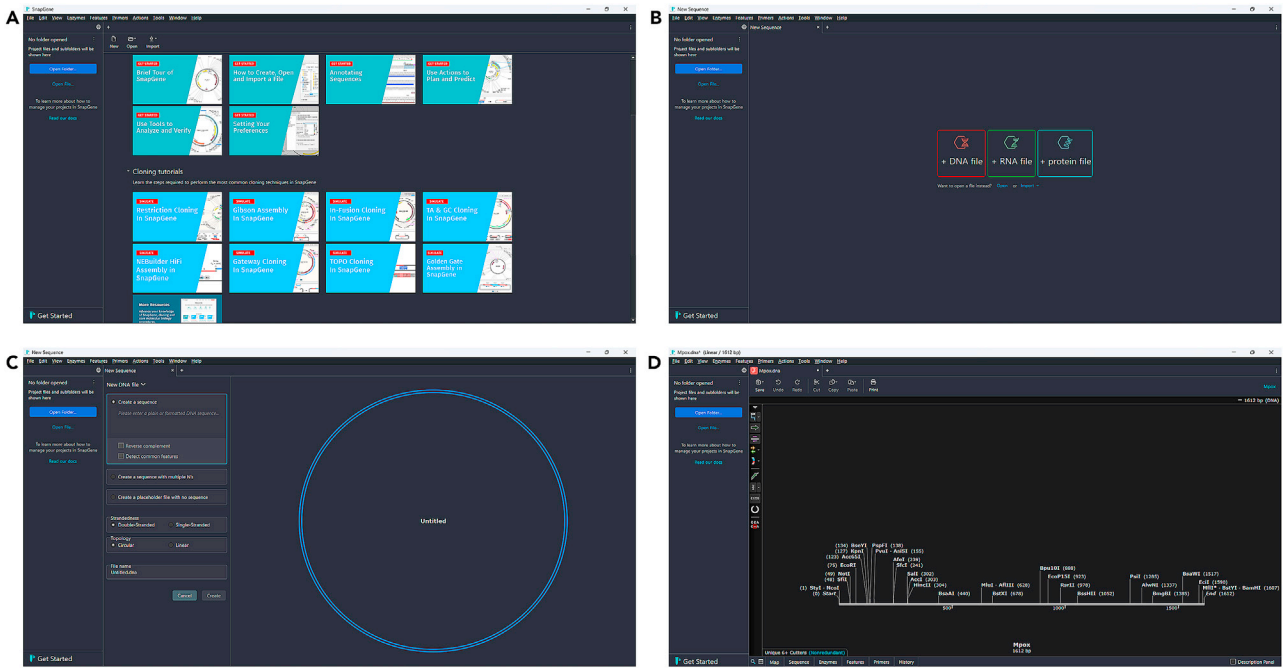


Figure 48. Creating DNA file of codon optimized sequence

(A) Homepage of SnapGene software.

(B and C) Creation of DNA file.

(D) Output file in. dna format.

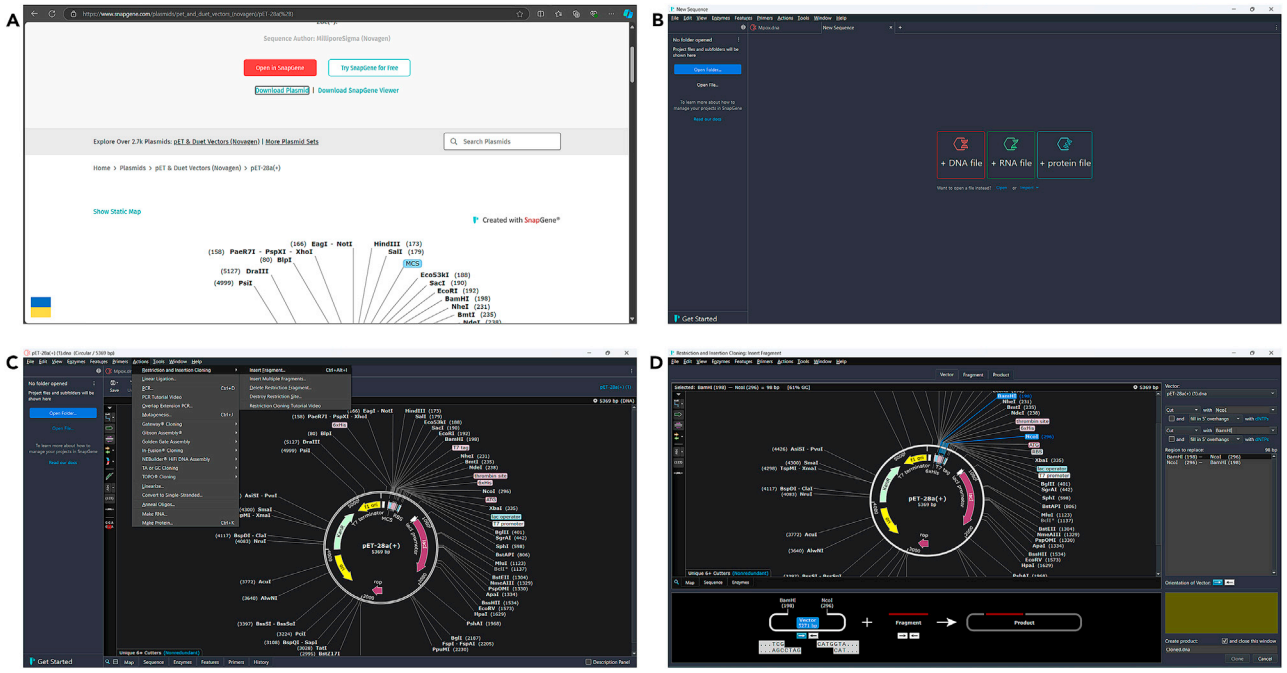


Figure 49. Creating vector for *in silico* cloning

(A) Vector of interest in SnapGene.

(B) Creating DNA file of vector.

(C and D) Steps to generate sticky end at available restriction sites.

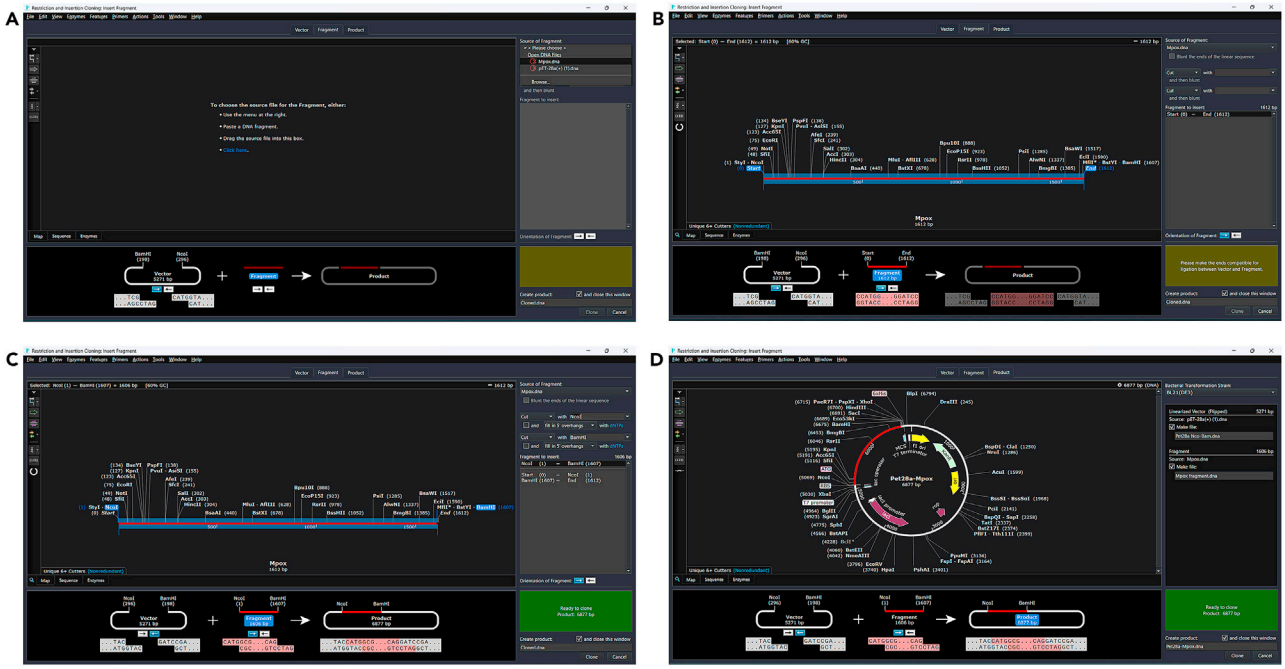


Figure 50. Fragment insertion in created vector

(A and B) Opening the fragment file.

(C) Steps to generate sticky end at available restriction sites in the fragment.

(D) Output showing *in silico* cloned product.

Potential solution

Go to notepad plus and manually edit the names of ligand and protein chains to "B" and "H" respectively in the docked protein file.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tanmay Majumdar (majumdart@nii.ac.in).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Tanmay Majumdar (majumdart@nii.ac.in).

Materials availability

This study did not generate new unique reagents.

Data and code availability

No datasets or code was generated during this study.

ACKNOWLEDGMENTS

We are very grateful to Dr. Debasisa Mohanty, Director, NII for providing full support for computational studies, valuable inputs, and guidance. This study was supported by grants from the CRG-SERB, Govt. of India (CRG/2021/000135) to A. Kaur, K.S., and T.M. and by DBT-NII intramural core grant to T.M.

AUTHOR CONTRIBUTIONS

A. Kumar, G.N., P.B., G.K., R.M., M.D., P.C., A. Kaur, K.S., S.M., R.K., I.K.S., and T.M. performed all the experiments, formal analysis, investigation, and methodology. Data curation, formal analysis, and reviewing the paper were done by A. Kumar, G.N., P.B., G.K., R.M., M.D., P.C., A. Kaur, K.S., S.M., R.K., and I.K.S. Conceptualization of the protocol,

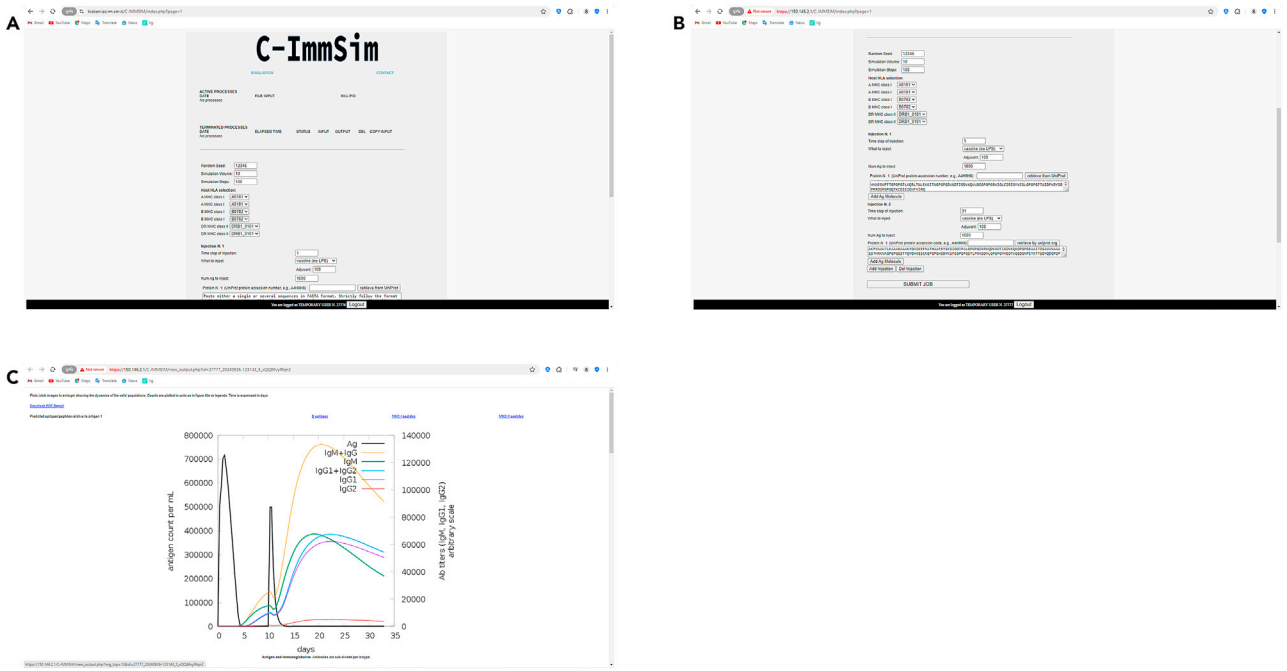


Figure 51. Immune simulation using C-ImmSim

(A) Homepage of C-ImmSim server.

(B) Entered parameters and protein sequences.

(C) Result showing vaccine's immunogenic response.

data curation, formal analysis, funding, investigation, methodology, project validation, writing – original draft, review, editing of the manuscript, and scientific supervision were performed by T.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Kaur, A., Kumar, A., Kumari, G., Muduli, R., Das, M., Kundu, R., Mukherjee, S., and Majumdar, T. (2024). Rational design and computational evaluation of a multi-epitope vaccine for monkeypox virus: Insights into binding stability and immunological memory. Heliyon *10*, e36154. https://doi.org/10.1016/j.heliyon.2024.e36154.

2. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

3. Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J. Chem. Inf. Model. *51*, 2778–2786. https://doi.org/10.1021/ci200227u.

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

5. Dimitrov, I., Flower, D.R., and Doytchinova, I. (2013). AllerTOP - a server for in silico prediction of allergens. BMC Bioinf. *14*, S4. https://doi.org/10.1186/1471-2105-14-S6-S4.

6. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571.

7. Geourjon, C., and Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics *11*, 681–684. https://doi.org/10.1093/bioinformatics/11.6.681.

8. McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405. https://doi.org/10.1093/bioinformatics/16.4.404.

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

10. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. *47*, D339–D343. https://doi.org/10.1093/nar/gky1006.

11. Jensen, K.K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J.A., Yan, Z., Sette, A., Peters, B., and Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology *154*, 394–406. https://doi.org/10.1111/imm.12889.

12. Dhanda, S.K., Vir, P., and Raghava, G.P.S. (2013). Designing of interferon-gamma inducing MHC class-II binders. Biol. Direct *8*, 30. https://doi.org/10.1186/1745-6150-8-30.

13. Hon, J., Marusiak, M., Martinek, T., Kunka, A., Zendulka, J., Bednar, D., and Damborsky, J. (2021). SoluProt: prediction of soluble protein expression in *Escherichia coli*. Bioinformatics *37*, 23–28. https://doi.org/10.1093/bioinformatics/btaa1102.

14. Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. *32*, W526–W531. https://doi.org/10.1093/nar/gkh468.

15. Heo, L., Park, H., and Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res. *41*, W384–W388. https://doi.org/10.1093/nar/gkt458.

16. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

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

18. Xue, L.C., Rodrigues, J.P., Kastritis, P.L., Bonvin, A.M., and Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics *32*, 3676–3678. https://doi.org/10.1093/bioinformatics/btw514.

19. Sharma, N., Naorem, L.D., Jain, S., and Raghava, G.P.S. (2022). ToxinPred2: an improved method for predicting toxicity of proteins. Brief. Bioinform. *23*, bbac174. https://doi.org/10.1093/bib/bbac174.

20. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

21. Rapin, N., Lund, O., and Castiglione, F. (2011). Immune system simulation online. Bioinformatics *27*, 2013–2014. https://doi.org/10.1093/bioinformatics/btr335.