Computational methods for vaccine design

Authors
Matvey Bondarev, Julius Wenckstern, Shuhao Xiao, Sandrine Georgeon, Bruno Correia

Affiliations
Laboratory of Protein Design and Immunoengineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne

Introduction

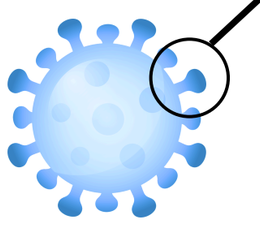
Lassa virus:

- Causes Lassa fever.
- **Has no approved treatment or vaccine.**
- Causes 300'000 infections with 5'000 deaths annually**.**
- Has a **high risk of becoming a Public Health Emergency of International Concern.**

→ This virus, and many others, create a need for a robust and efficient **vaccine design pipeline**.

Lassa virus spike protein

- Responsible for host cell entry → Makes an effective target for neutralizing antibodies.
- Starts as 3 identical units.
- Each unit gets cleaved twice into:
  - Stable Signal Peptide (SSP)
  - Glycoprotein 1 (GP1)
  - Glycoprotein 2 (GP2)



N C

Structure design

- We integrate three cutting-edge deep learning tools:
  - RFdiffusion: backbone design.
  - ProteinMPNN: sequence generation.
  - Boltz-2: structure prediction.
- To modify cell-signaling: modification of the signal peptide.
- To limit off-target effects against newly created epitopes, we add glycosylation sites:
  - Generation of a backbone-conditioned probability matrix using ProteinMPNN.
  - Scoring by maximizing a *glycosylation potential*:

$$s(i) = P_i(N) \cdot \big(1 - P_{i+1}(P)\big) \cdot \big(\alpha_S P_{i+2}(S) + \alpha_T P_{i+2}(T)\big) \cdot w_i^{\beta_N}\, w_{i+1}^{\beta_X}\, w_{i+2}^{\beta_{ST}}.$$

Design selection

We filter the designs based on:

- Root Mean Square Deviation (RMSD in Å) of atomic positions compared to the native structure:

$$\mathrm{RMSD}(\mathbf{v}, \mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{v}_i - \mathbf{w}_i\|^2}$$

- Predicted local distance difference test (pLDDT):
  - Boltz-2 average per-atom confidence score computed from distances to nearby amino acids.
  - Higher pLDDT means more reliable local geometry.
- Predicted aligned error (PAE in Å):
  - Boltz-2 pairwise confidence relative to atom frames.
  - Lower PAE means a more trustworthy relative placement.
- Predicted distance error (PDE in Å):
  - Boltz-2 predicted absolute error in distance between atoms for each pair.
  - Smaller PDE indicates a more reliable distance.
- Cavity volume in the stem (Å^3):
  - Lower volume leads to more stability in solution.

Co-funded by the European Union

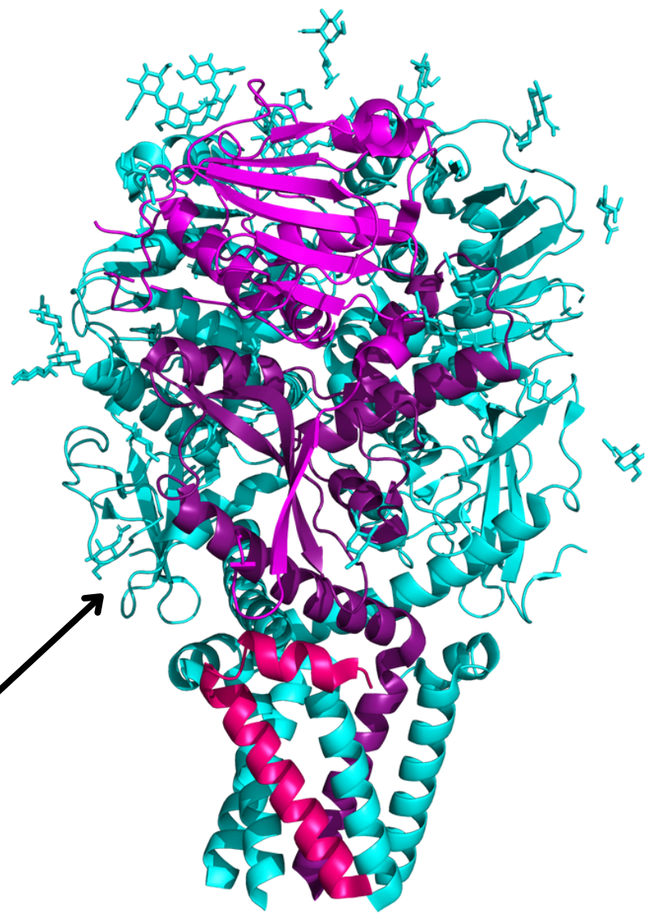
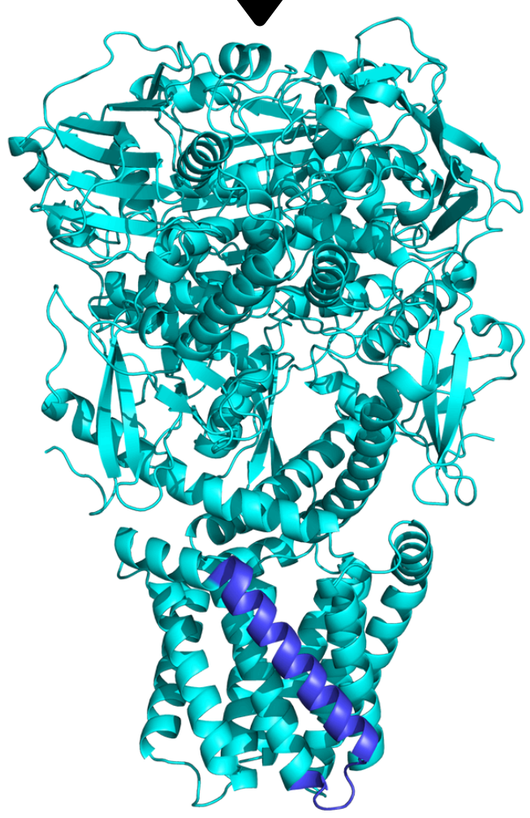


Illustration of the native Lassa virus spike protein.

RFDiffusion[1]

ProteinMPNN[2]

Boltz-2[3]

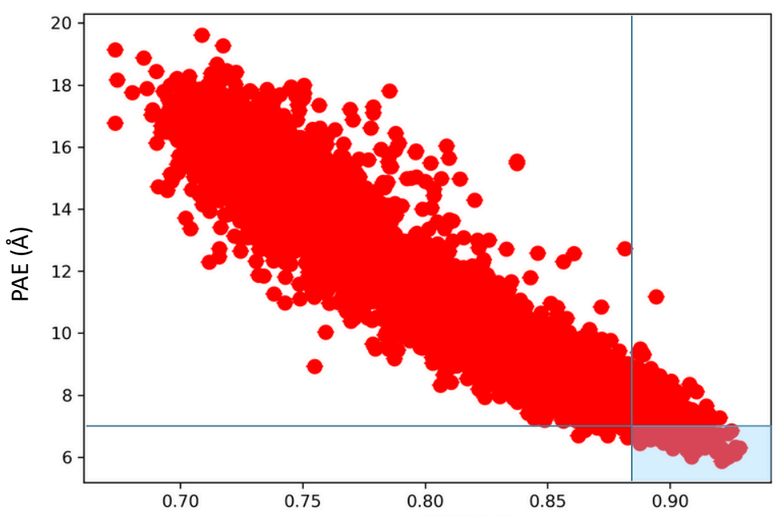


A final design from the de novo linker design and sequence redesign strategy.

Expression & Antibody binding

Final designs

- 2 design strategies chosen:
  - Generating a linker between the first residues of SSP and the last residues of GP2 and redesigning the sequence.
  - Generating a new stem, keeping first 12 residues of SSP fixed. Designing a new sequence.
- 16 designs selected for testing of expression, solubility and antibody binding.
  - 12 linker generation & sequence redesign.
  - 4 de novo stem design, keeping first 12 residues of SSP fixed.

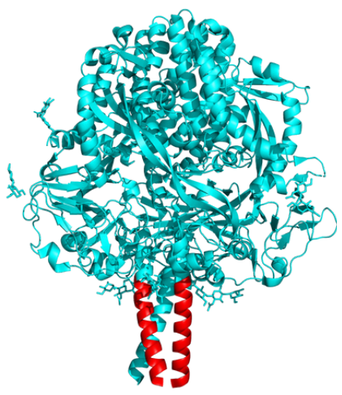


PAE vs pLDDT plot of the designs generated with the linker & sequence redesign strategy. Chosen cutoffs: pLDDT > 0.88, PAE < 7 Å

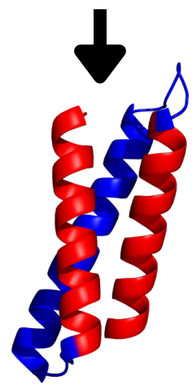
Generalizability

To demonstrate the generalizability of the pipeline:

- We applied it to a different design strategy on a fusion (F) protein from Nipah virus.
- We designed immunogens to try to elicit nanobody binding against the stem.



Native Nipah virus F protein. In red: the two epitopes we want to preserve to select for nanobodies that target the stem.



“Headless stem” design: in red are the two epitopes we keep from the native structure, in dark blue is a new motif scaffold to stabilize the helices. By doing this, we convert the two homodimeric helices into one monomer.

References

1. Watson, J.L., Juergens, D., Bennett, N.R. *et al.*, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
2. J. Dauparas *et al.*, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).
3. S. Passaro *et al.*, Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*, doi: 2025.06.14.659707 (2025).