Antigen-agnostic AI-simulated SHM reveals evolutionary constraints and improves neutralization potency of HIV-1 broadly neutralizing antibodies by up to 76-fold

Benjamin Nemoz, Karenna Ng, Dennis Burton, Bryan Briney

Department of Immunology and Microbiology, Scripps Research, La Jolla, CA, USA.

Background

- **Somatic hypermutation (SHM)** enhances the antibody diversity in which antigen-based selection acts to improve antigen binding affinity
  - Also shaped by antigen-agnostic constraints: sequence context, activation-induced cytidine deaminase (AID) targeting motifs, structural stability/folding
- **Broadly neutralizing antibodies (bnAbs)** against HIV are rare, critical for protection, and require extensive SHM to develop
  - ~3 years to develop neutralization breadth, 20-30% mutation load[1]
  - Understanding how bnAbs develop is key for vaccine design
- **Antibody language models (AbLMs)** have been shown to learn evolutionary patterns from large antibody datasets[2,3] ⟶ potential to model bnAb maturation in an antigen-agnostic manner

Can computational modeling of SHM extend bnAb maturation beyond *in vivo* observations?

Hypothesis: The rules governing SHM can be learned from sequence context alone and applied computationally to drive antibody maturation beyond in vivo observations.

Dataset Generation

Memory B Cell (MBC)-enriched Bulk Next-Generation Sequencing (NGS)

- 500M PBMCs each from 8 healthy donors (***Fig. 1A***)
- Donor-specific germline databases to minimize inaccurate annotation of allelic-differences and SHM-driven mutations
- **~83 million unpaired, MBC sequences** for model development (***Fig. 1B-C***)
  - Train/Eval/Test splits were ensured to have <80% sequence identity overlap to minimize test set leakage by members of the same clonal family

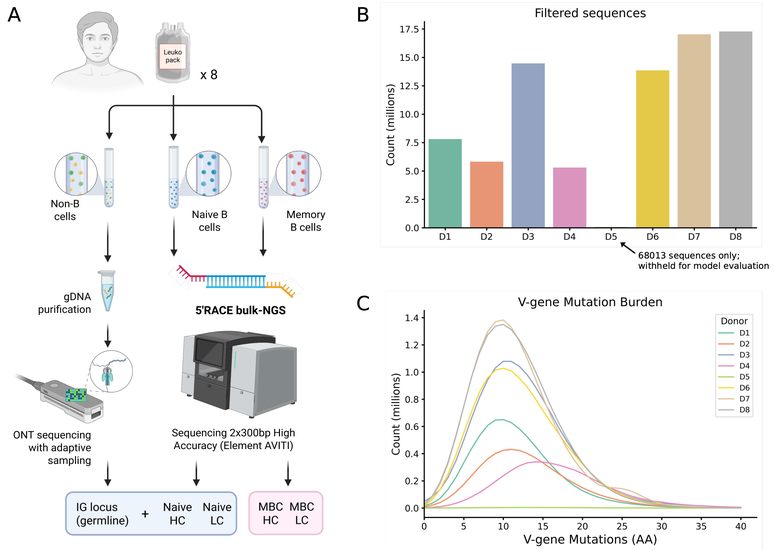

Figure 1. MBC dataset generation. (**A**) From each donor, naïve and memory B cell subsets were isolated and sequenced using bulk NGS. Naïve sequences were used to construct donor-specific germline databases for memory sequence annotation. Made with BioRender. Filtered dataset sequence counts (**B**) and V-gene mutation burden (**C**) per donor.

SHM Model Design

Biologically Informed Two-Model System

- Analogous to separated AID and DNA-repair, these systems are modeled independently (***Fig. 2***)

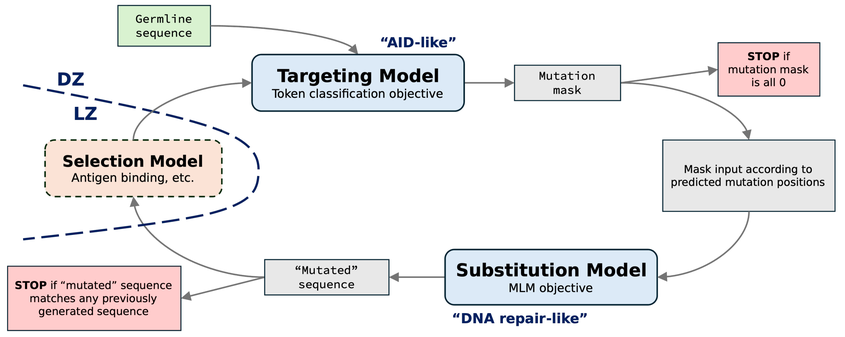

Figure 2. SHM Model Framework. Two AbLMs (blue) simulate SHM: a targeting model predicts mutation positions, and a substitution model predicts amino acid changes. Models iteratively mutate a germline sequence (green) until a stopping criterion (red) is met. A future selection model (orange) could be added to evaluate variant fitness, filling out the *in silico* affinity maturation model.

In Silico Maturation

Further Maturation of PC94-A bnAb Lineage

- PC94-A lineage: rich longitudinal sequence data, well characterized (***Fig. 3A***)
  - Targets the V3-glycan supersite of HIV-1 envelope trimer
- SHM model ⟶ 1000 stochastic trajectories generated per parental antibody
  - Model was trained/evaluated without any HIV-1 antigen information or PC94 data
- Predicted variants indicate forward-pushing SHM (as opposed to germline reversion) (***Fig. 3B***)
- Mutations concentrate at SHM hotspot motifs: 3.2x enrichment in the heavy chain (***Fig. 3C***)

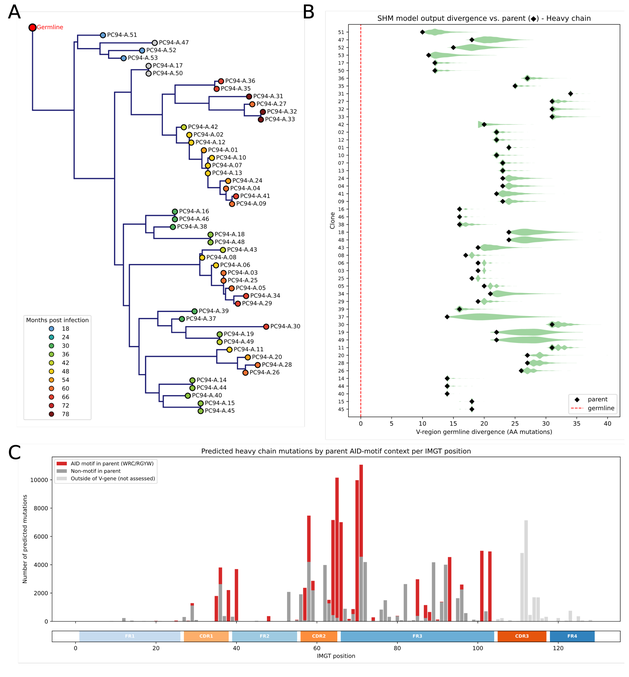

Figure 3. *In silico* evaluation of SHM model-predicted variants of the PC94-A lineage.
(**A**) PC94-A lineage phylogeny, colored by isolation timepoint in months post-infection. Antibodies with no temporal information are shown in grey.
(**B**) V-gene germline divergence of PC94-A parent (◆) and model-predicted (green) heavy chain variants. Almost all model-predicted variants increase divergence from germline.
(**C**) Mutation counts per IMGT position for all model-predicted heavy chain variants, colored by whether the parent sequence contained an AID hotspot at that site. Counts for residues outside of the V-gene are shown without AID hotspot information.

In Vitro Validation

Maintained Breadth with Selective Enhancement

- Parental and top-predicted bnAb variants were evaluated against a panel of five HIV-1 pseudoviruses
- AI-matured variants maintain heterologous neutralization breadth (***Fig. 4***)
  - 9-fold improvement in potency on average

Antibody	Neutralization (IC50)				
	BJOX2000	CAAN	JRCSF	JRFL	SF162
PC94-A.27	0.021	0.064	8.53	0.026	0.005
A.27_1	0.016	0.057	1.12	0.06	0.006
PC94-A.30	0.05	0.082	12.85	1.153	0.028
A.30_1	2.88	65.2	76.6	90.5	1.75
A.30_2	0.007	0.072	1.36	0.62	0.005
PC94-A.31	0.043	0.043	0.794	0.030	0.005
A.31_0	0.060	0.190	0.059	0.093	0.008
PC94-A.32	0.038	0.04	6.33	0.084	0.002
A.32_2	0.05	0.09	2.86	0.13	0.004
PC94-A.33	0.038	0.044	6.64	0.047	0.001
A.33_0	0.084	0.284	0.087	0.151	0.014
A.33_1	0.014	0.095	0.922	0.071	0.006

Figure 4. Neutralization of SHM model-matured PC94-A variants. IC50 values (μg/mL) of parental and model-matured variants against a panel of five HIV-1 pseudoviruses. Red indicates stronger neutralization, while green indicates weaker.

Conclusions

- Biologically-informed AbLM framework recapitulates SHM-like patterns *in silico*
  - Mutations are "forward-pushing" and located at AID hotspots
- *In vitro* results indicate model-predicted mutations preserve antibody function and can improve antigen binding
  - Conservative prediction supports constrained natural-like antibody engineering, but broader sampling may be needed to model diverse maturation trajectories
- MBC-enriched dataset reduces germline bias and supports future SHM analyses

Next Steps

Further Evaluation using the PC94-A Lineage

- Model recapitulation of the PC94-A lineage from its germline sequence
- Structural studies of model-mutated sequences

Improving the SHM Model

- Additional training data generation from other healthy donors (generalize across donors/antigens)
  - Incorporate paired VH/VL data to capture cross-chain dependencies[4]
- Expand evaluation (evolution campaigns across other antigens)

References

1. E. Landais et al. PLOS Pathogens. 12 (2016), doi:10.1371/journal.ppat.1005369.
2. J. A. Ruffolo et al., Machine Learning for Structural Biology (MLSB) Workshop (2021), doi:10.48550/arXiv.2112.07782.
3. D. van Ginneken et al. Briefings in Bioinformatics. 26 (2025), doi:10.1093/bib/bbaf418.
4. S. M. Burbach, B. Briney, PATTER. 0 (2024), doi:10.1016/j.patter.2024.100967.