Supplemental Information

Optimizing immunization protocols to elicit broadly neutralizing antibodies

Kayla G. Sprenger[a1], *Joy E. Louveau*[b1], *Pranav M. Murugan*[a], *Arup K. Chakraborty*[a,c,d,e,f,g,*]

[a]Institute for Medical Engineering and Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139; [b]Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA 02139; [c]Department of Chemical Engineering, MIT, Cambridge, MA 02139; [d]Department of Physics, MIT, Cambridge, MA 02139; [e]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139; [g]Department of Chemistry, MIT, Cambridge, MA 02139

*Correspondence: arupc@mit.edu

[1]These authors contributed equally to this work.

www.pnas.org/cgi/doi/10.1073/pnas.1919329117

Table S1. List of all parameters used in affinity maturation simulations.

Description		Parameter	Value
Somatic hypermutation	Probability of a mutation in CDR or FWR per division	P_{mut}	0.14
	Probability that a CDR mutation is lethal	P_{lethal}	0.30
	Probability that a CDR mutation is silent	P_{silent}	0.50
	Probability that a CDR mutation is affinity-affecting	$P_{affinity\text{-}affect}$	0.20
	Number of total residues	L	46
	Number of conserved residues	L_c	18
	Number of variable residues	L_v	28
	Lower boundary for variable residues of seeding B cells	$H_{V0,\,low}$	-0.18
	Upper boundary for variable residues of seeding B cells	$H_{V0,\,high}$	0.9
	Lower boundary for conserved residues of seeding B cells	$H_{C0,\,low}$	0.3
	Upper boundary for conserved residues of seeding B cells	$H_{C0,\,high}$	0.6
	Lower boundary for residue of B cells	H_{low}	-1.0
	Upper boundary for residue of B cells	H_{high}	1.5
	Boundary of energy change due to single-point mutation	δ	1.0
	Boundary of energy change due to loop insertion mutation	δ_L	1.0
	Proportionality factor for loop treatment	α	0.25
	Mean of shifted lognormal distribution for CDR mutation	μ	1.9
	Standard deviation of shifted lognormal distribution	σ	0.5
	Shift/offset of shifted lognormal distribution	ε	3.0
Breadth and binding	Pseudo inverse temperature (k_BT^{-1})	e_{scale}	0.08
	Activation threshold	E_{act}	9
	Breadth binding threshold	E_{th}	12
	Antigen concentration	c	varies
	Number of panel Ags to test clonal breadth against	$N_{panel\ Ags}$	100
GC dynamics	Probability that a B cell is recycled after selection	$P_{recycle}$	0.70
	Probability that a B cell exits the GC after selection	P_{exit}	0.30
	Fraction of B cells that receive T cell help after binding Ag	$F_{help\ cutoff}$	0.70
	Number of B cells that seed a GC	$N_{GC\ founders}$	10

Choice of α

Immunization with a single Ag has been shown to produce primarily strain-specific Abs; that is, a large fraction of the mutations made by BCRs in response to a single Ag increase binding to variable antigenic sites. However, as our *in silico* vaccination protocols are always preceded by a hypothetical GL-targeting scheme, there should also be a driving force – even for single-Ag administration – towards the conserved residues shared between the vaccine Ag and GL-targeting Ag. After varying α across a wide range, we set it to a value of 0.25 to capture both of these facets, namely, to achieve approximately equal selection for BCR mutations that increase variable and conserved site binding. For example, given a ΔE of +1 $k_B T$ towards a variable site, after applying the penalty ($\Delta E_L = -\alpha \Delta E = -0.25$), the overall change in the binding free energy between the Ag and BCR would be a beneficial increase of 0.75 $k_B T$. Similarly, given a ΔE of +1 $k_B T$ directly towards a conserved site, the overall change in the binding free energy would simply be 1.0 $k_B T$. The difference of 0.25 $k_B T$ between these two mutational schemes is then approximately balanced by the higher probability of mutating in the variable region, due to there being more of these residue types in the simulated epitope (28 variable vs. 18 conserved residues).

Dependence of TFL on $P_{internalize}$

TFL is an empirical way to encapsulate the effects of Ag concentration, c, and binding energy, E, the latter of which is linearly related to mutational distance, d. The two variables c and d thus determine TFL, which in turn determines the internalization probability, $P_{internalize}$. Due to the specific form of the equation, $P_{internalize}$ has a complicated dependence on c and E. Additionally, $P_{internalize}$ depends differently on the two variables, which, along with the choice of the parameter e_{scale} (chosen to fit experiments; see main text), largely determines the weight in the equation for TFL. The following derivations show how the particular dependence of $P_{internalize}$ on c differs from its dependence on E. All sub/superscripts have been dropped for simplicity and P refers to $P_{internalize}$.

Changes in P with changes in c:

$$P = \frac{ce^{e_{scale}(E-E_{act})}}{1 + ce^{e_{scale}(E-E_{act})}}$$

$$P = \frac{ax}{1+ax} \rightarrow \frac{\partial P}{\partial x} = a\frac{\partial}{\partial x}\left(\frac{x}{1+ax}\right)$$

$$\frac{\partial}{\partial x}\left(\frac{u}{v}\right) = \frac{v\frac{\partial u}{\partial x} - u\frac{\partial v}{\partial x}}{v^2}$$

$$v = 1 + ax, \;\; u = x, \;\; v\frac{\partial u}{\partial x} = 1 + ax, \;\; u\frac{d\partial}{d\partial} = ax$$

$$\frac{\partial P}{\partial x} = a\left(\frac{(1+ax) - ax}{v^2}\right) = \frac{a}{v^2} = \frac{a}{(1+ax)^2}$$

$$\therefore \frac{\partial P}{\partial c} = \frac{e^{e_{scale}(E-E_{act})}}{(1+ce^{e_{scale}(E-E_{act})})^2}$$

Changes in *P* with changes in *E*:

$$P = \frac{ce^{ax-b}}{1+ce^{ax-b}} \rightarrow \frac{\partial P}{\partial x} = c\frac{\partial}{\partial x}\left(\frac{e^{ax-b}}{1+ce^{ax-b}}\right)$$

$$v = 1+ce^{ax-b},\ \ u = e^{ax-b},\ \ v\frac{\partial u}{\partial x} = (1+ce^{ax-b})ae^{ax-b},\ \ u\frac{\partial v}{\partial x} = e^{ax-b}ace^{ax-b}$$

$$\frac{\partial P}{\partial x} = \frac{c(1+ce^{ax-b})ae^{ax-b} - c(e^{ax-b})ace^{ax-b}}{(1+ce^{ax-b})^2} \xrightarrow{\cdot e^{2b}/e^{2b}} \frac{ace^{ax-b}}{(1+ce^{ax-b})^2}$$

$$\therefore \frac{\partial P}{\partial E} = \frac{e_{scale}ce^{e_{scale}(E-E_{act})}}{(1+ce^{e_{scale}(E-E_{act})})^2}$$

$$\therefore \frac{\partial P/\partial c}{\partial P/\partial E} = \frac{1}{e_{scale}c}$$

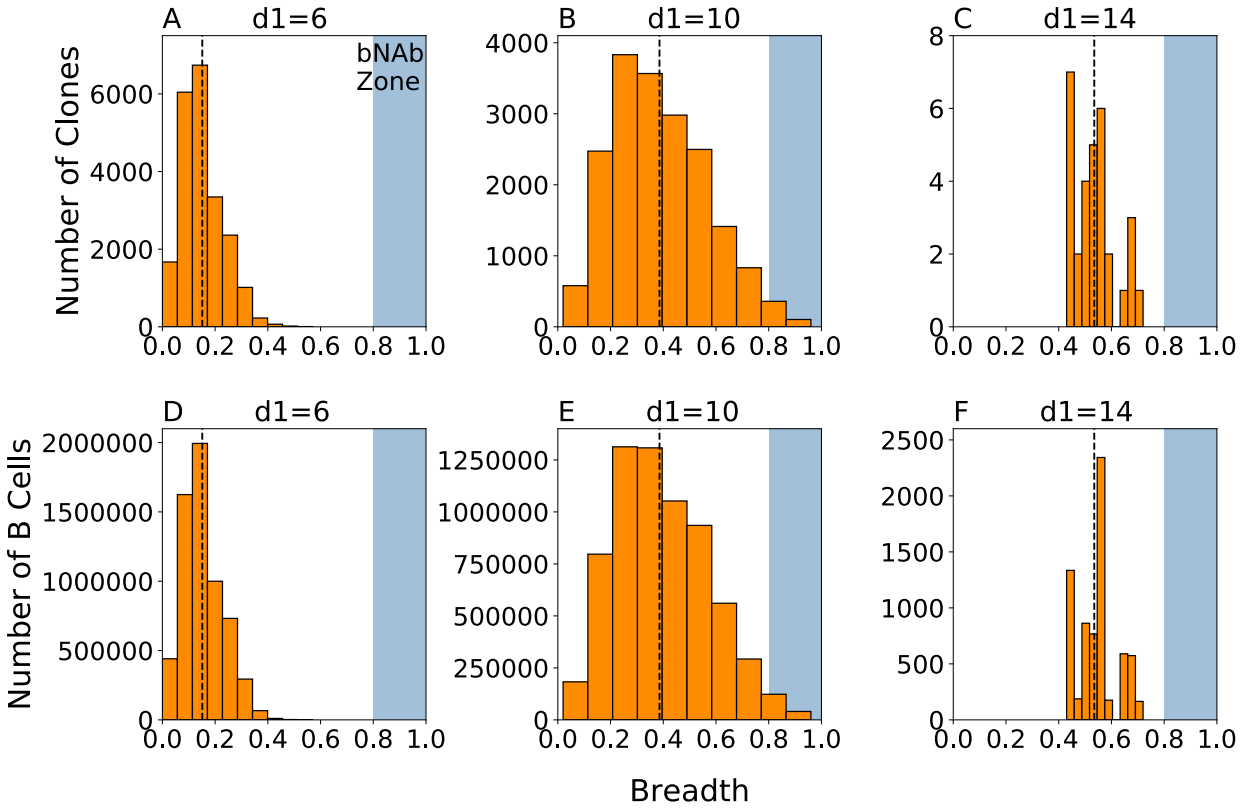

Figure S1. Breadth distributions for (top) all clones and (bottom) all B cells in all clones produced for three different vaccination settings – low frustration (A, D), medium frustration (B, E), and high frustration (C, F). Shaded blue regions and black dashed lines are as described in Fig. 3 (main text). Due to the large number of GCs we analyzed (1,000) for each vaccination setting, the resultant mean breadth (black dashed lines) and bNAb titers/GC are the same whether we calculate it across all clones or, considering the size of each clone, across all B cells within all clones.

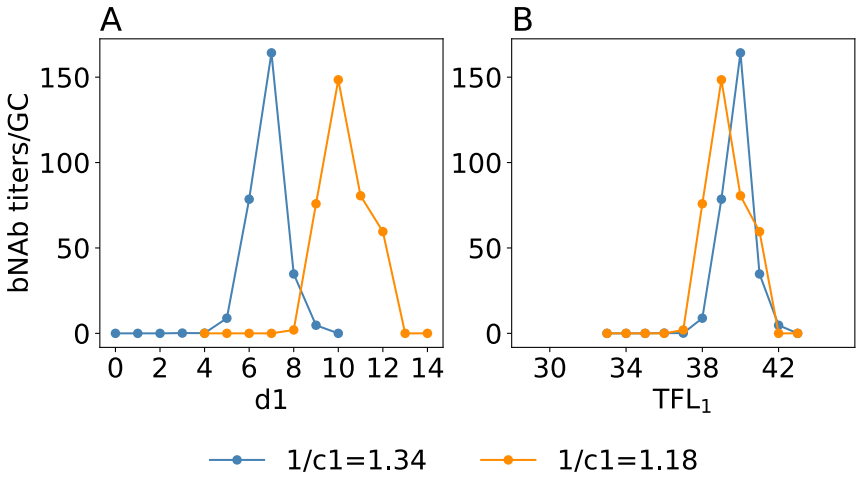

Figure S2. (A) Mutational distance between the first vaccine Ag and GL-targeting Ag (d1) and (B) the total frustration level of the first immunization (TFL_1), versus the bnAb titers/GC for two different Ag concentrations (c1). Each dot represents the average output from n=1,000 GCs.

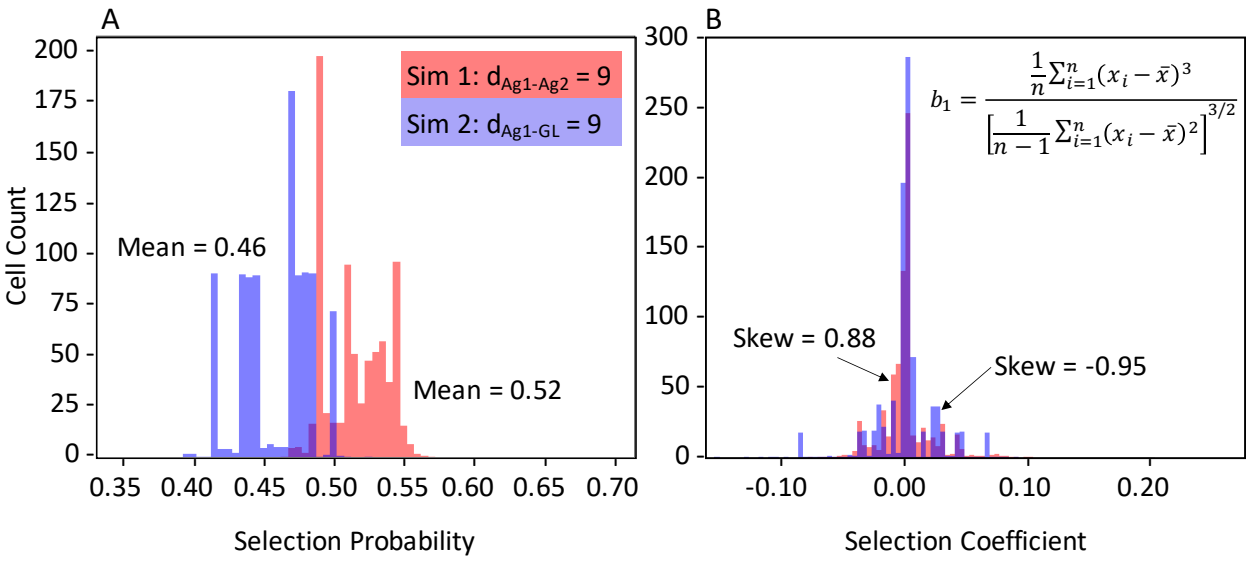

Figure S3. (A) Fitness distributions of different B cells and (B) their associated selection coefficients. Results are shown for: (red) a simulation of two sequential single-Ag administrations, where the mutational distance between the first and second vaccine Ags (Ag1 and Ag2, respectively) is 9; (blue) a simulation of one single-Ag administration where the Ag has a mutational distance of 9 from the germline-targeting (GL) Ag. The equation for calculating the population skewness (b_1) is shown in the inset of (B), where n is the number of values in the sample, and x_i and $\bar{x}$ are the sample values and mean, respectively.

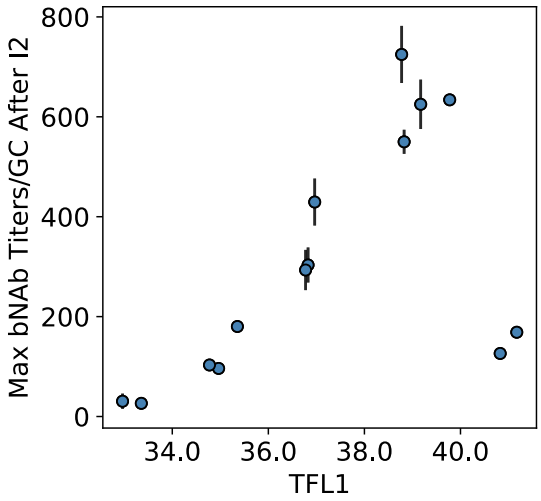

Figure S4. Maximum bnAb titers/GC that can be achieved after the second vaccine immunization (I2), for different values of the total frustration level in the first vaccine immunization (TFL_1).

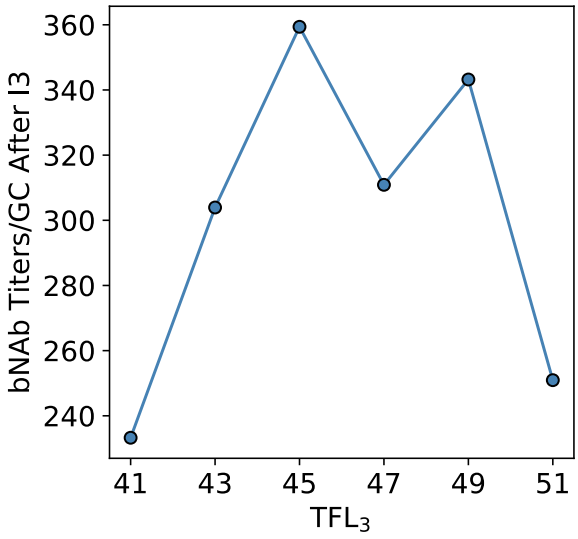

Figure S5. BnAb titers/GC produced after the final immunization of a vaccine performed at a suboptimal TFL_1 of 35, followed by a second immunization at a suboptimal TFL_2 of 39, and finally by a third immunization spanning a TFL_3 of 41 to 51.

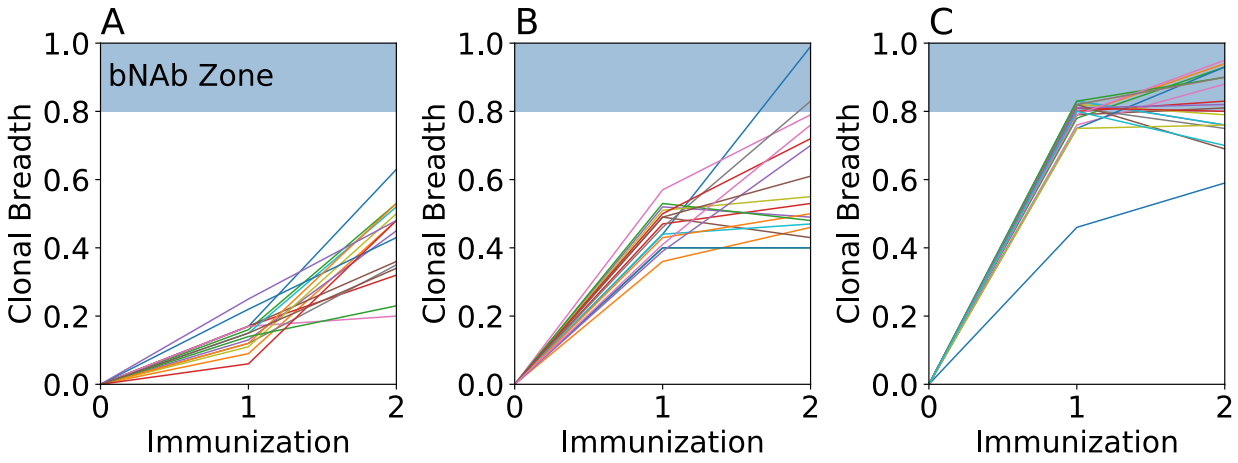

Figure S6. Mutational trajectories of individual clones from different GCs (i.e., simulation trials) after multiple vaccine immunizations. Immunization conditions correspond to those in Fig. 6 (main text): (A) TFL_1=33, TFL_2=43; (B) TFL_1=39, TFL_2=49; (C) TFL_1=43, TFL_2=43. Here, the requirement has been relaxed that the clones must achieve 'success' (i.e., a clonal breadth above 0.8 after two vaccine immunizations; see main text).