SCALING DATA AND AI FOR MULTIOPTIMIZED BIOLOGICS

Sharrol Bachas*, Goran Rakocevic*, David Spencer, Robel Haile, Anand Sastry, John Sutton, George Kasun, Vincent Blay, Christa Kohnert, Cailen McCloskey, Edriss Yassine, Borka Medjo, Nebojsa Tijanic, Shaheed Abdulhaqq, Randal Olson, Jennifer Stanton, Bailey White, Rebecca Viazzo, Abigail Ventura, Kerianne Jackson, Rebecca Consbruck, Hayley Carter, Chelsea Chung, Brea Luton, Nic Diaz, Curtis Orona, Kaitlin Witherell, Alexander Brown, Amber Brown, Joshua Meier, Matthew Weinstock, Gregory Hannum, Ariel Schwartz, Miles Gander, Roberto Spreafico

Successful biologic therapeutics must meet multiple design specifications simultaneously, such as binding affinity, developability, and immunogenicity. More advanced therapeutics require additional specifications, such pH sensitive binding or selectively binding across multiple epitopes. Designing those rare sequences to meet multiple challenging value-creating specifications requires the ability to generate and model informative, high-quality, and high-throughput data. At Absci, we generate the right experimental data across the necessary endpoints. We use ai-guided approaches to maximize the information value provided by each point. We model the resulting datasets using our proprietary language-based algorithms, and we design unique biologic variants that meet complex multi-parametric specifications. Our models and biologics are validated by our expert experimental teams.

Read the full manuscript:

www.biorxiv.org/content/10.1101/2022.08.16.504181v1

INFORMATION-RICH LIBRARIES ARE TRANSFORMED INTO SOLUPRO™ *E. coli*

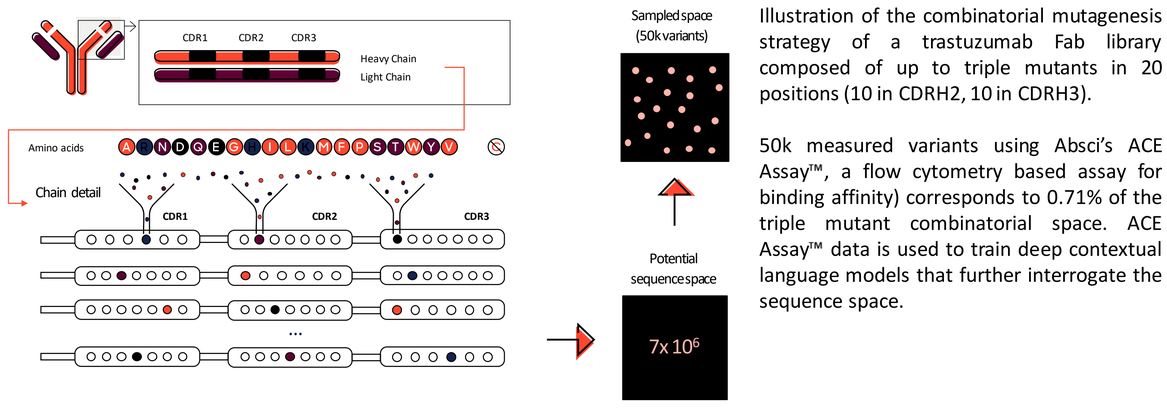


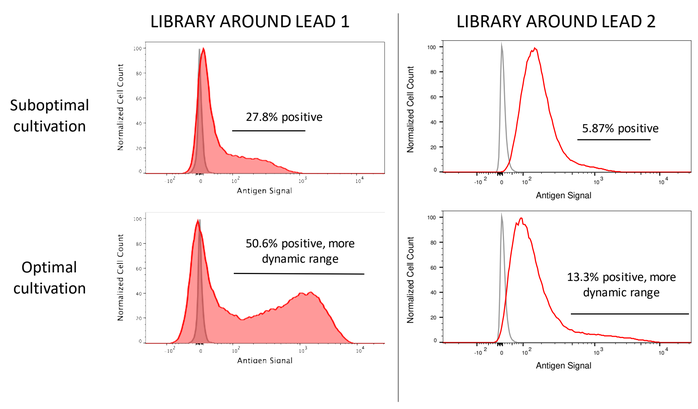
Illustration of the combinatorial mutagenesis strategy of a trastuzumab Fab library composed of up to triple mutants in 20 positions (10 in CDRH2, 10 in CDRH3).

50k measured variants using Absci's ACE Assay™, a flow cytometry based assay for binding affinity) corresponds to 0.71% of the triple mutant combinatorial space. ACE Assay™ data is used to train deep contextual language models that further interrogate the sequence space.

CULTIVATION MATRIX IDENTIFIES OPTIMAL GROWTH CONDITIONS

Factors such as growth media and cultivation time impact the proportion of correctly folded and expressed Fabs. Various seed and induction conditions are tested for libraries of each new drug lead. Higher quality protein increases the dynamic range and accuracy of the ACE Assay™ results.

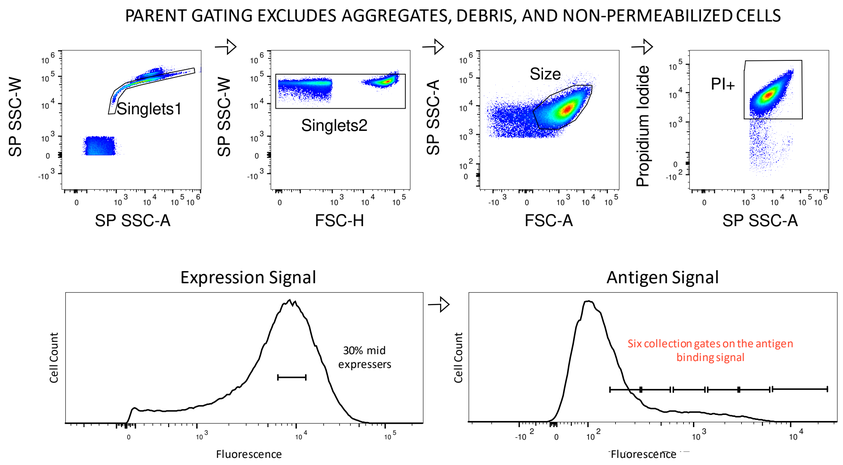
Red histograms indicate the library antigen binding signal and grey overlays are isotype controls measured in the ACE Assay™. All libraries are sequenced to confirm improvement in the ACE Assay™ signal is not due to preferential growth.



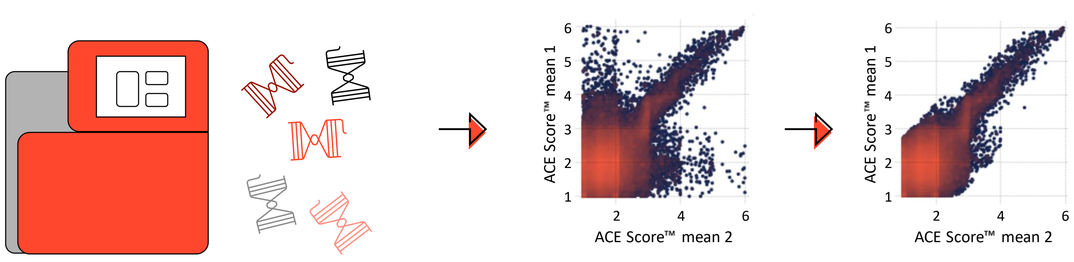
THE ACE ASSAY™ GENERATES HIGH-QUALITY BINDING DATASETS AT LARGE SCALE (>50K VARIANTS)

Fab libraries expressed intracellularly in SoluPro™ *E. coli* → Cells are fixed, permeabilized, and stained for antigen binding and Fab titer → The library is screened by flow cytometry to estimate binding affinity of each library variant

A.



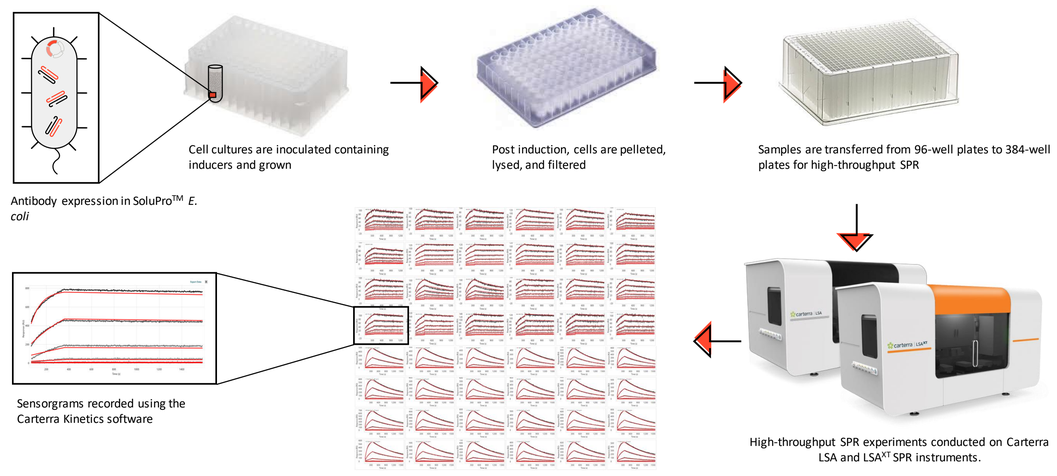
B.



The ACE Assay™ combines fluorescence-assisted cell sorting (FACS) and next-generation sequencing (NGS) to generate binding affinity estimates to an antigen (ACE Scores™) in high-throughput. A) Flow cytometry gating scheme. Initial gating is used to reduce aggregates, debris, and non-permeabilized cells. Expression variability is controlled by accounting for Fab expression signal. Six collection gates are then used to evenly bin across the log range of the antigen fluorescent signal. (B) After sorting, unique molecular identifiers are added to flank the CDR region. Collected DNA material from sorted cells is then amplified and sequenced. Normalized read counts across sort gates are weighted to assign an ACE Score™ to each variant.

SPR VALIDATES ACE ASSAY™ MEASUREMENTS AND MODEL PREDICTIONS

A.

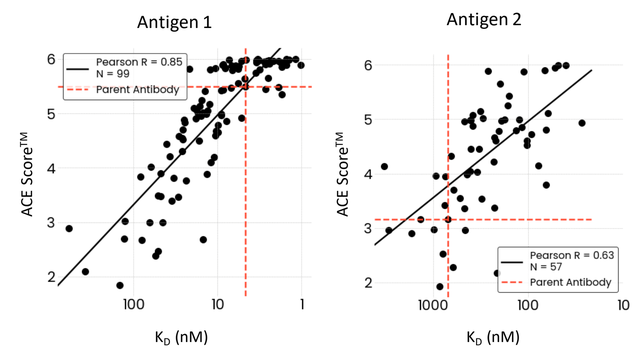


(A) SPR workflow. A subset of ACE Assay™ measured variants are transformed and cultured in SoluPro™ *E. coli.* Culture lysates are screened via SPR by immobilizing the Fabs on the biosensors.

Standard SPR runs measure binding traces for up to 1,500 variants, each measured at 4 concentrations in duplicate. ACE Scores™ correlate with SPR measured equilibrium dissociation constants (K_D). Following training on ACE Scores™, a subsequent round of SPR is performed to confirm model predictions.

(B) Correlation between SPR measured K_Ds and ACE Scores™ for samples from two representative libraries.

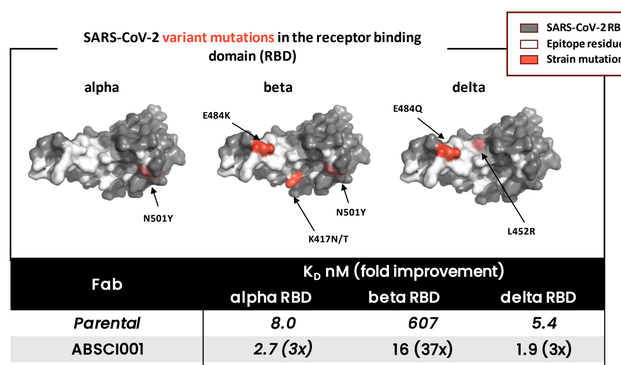
B.



CO-OPTIMIZING BINDING TO MULTIPLE ANTIGENS

ACE Assay™ datasets can be generated for the same Fabs binding to multiple antigens, and novel variants with the desired selectivity profile can be generated.

Equilibrium constants (K_D) from a subset of AI-generated variants designed to maximize binding to three different SARS-CoV-2 mutants simultaneously. Our campaign generated multiple assets significantly improving binding affinity to all three antigens.



Fab	K_D nM (fold improvement) alpha RBD	beta RBD	delta RBD
Parental	*8.0*	*607*	*5.4*
ABSCI001	*2.7 (3x)*	16 (37x)	1.9 (3x)
ABSCI002	1.5 (5x)	24 (25x)	0.8 (7x)
ABSCI003	0.9 (9x)	32 (19x)	0.6 (9x)
ABSCI004	1.1 (7x)	37 (16x)	1.4 (4x)
ABSCI005	1.3 (6x)	40 (15x)	0.8 (7x)

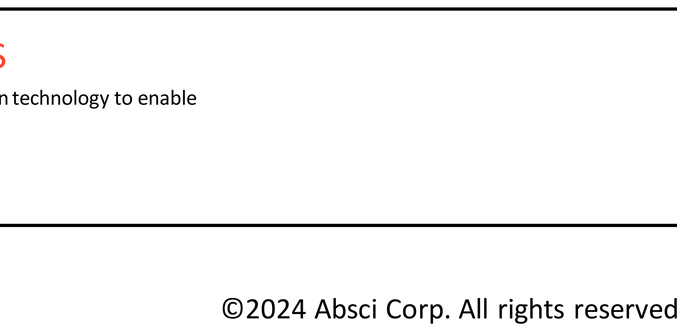
pH-SENSITIVE BINDING

Changes in binding affinity with pH can provide distinct pharmacokinetic benefits. For instance, reduced binding at low pH may improve recycling from the lysosome ("catch and release") and extend half-life. Alternatively, tighter binding at low pH may favor tissue localization towards a tumor microenvironment.

Datasets of ACE Scores™ can be generated at different pH, which allow us to train predictive AI models of the effects of pH on binding (A). The models can be used to generate variants that preferentially bind at a certain pH while preserving other properties of the biologic, such as affinity, developability, or immunogenicity.

A.

B.



KEY POINTS

Absci's end-to-end technology involves information-rich library design, expression and cultivation in SoluPro™ *E. coli*, ACE Assay™ high-throughput screening, and medium-throughput validation (100s) using SPR and developability assays.

Absci can generate large (>100 k), high-quality binding datasets to support AI lead optimization.

Training datasets can be generated for multiple parameters, such as pH-dependent binding or affinity to multiple antigens.

Deep contextual language models trained on HQ data generated in house can quantitatively predict binding of unseen antibody sequence variants and enable multiparametric biologic optimization.

PARTNER WITH US

Leverage our AI lead optimization technology to enable your projects.