Virtual screening software for modeling personal repertoire antibody structures and identifying their target antigens

Carl-Erik Tornqvist[1], Erik Edlund[1], Constantin Tivis[1], Thomas Leary[1], David Frisch[2], Charles Landry[2], and Tim Durfee[1] Author affiliations: [1]DNASTAR, Inc. Madison, WI, [2]Scarab Genomics LLC, Madison, WI

DNASTAR
SOFTWARE FOR LIFE SCIENTISTS

Abstract

The B cell population in each individual produces an estimated 1010 different antibodies, collectively known as the antibody repertoire. This extraordinary diversity is essential for responding to the unique history of infections, vaccinations and cancer encountered over an individual's lifetime. Efficient deciphering of that history could contribute to improving human health in numerous ways, from better clinical decision making to improved diagnostics and therapeutics. Toward that goal, we have commercialized the NovaFold suite of computational tools for predicting 3D antibody structures as well as modeling their interaction with potential antigens using artificial intelligence (AI) and biophysical (BP) based algorithms. Here, we present results from an ongoing human repertoire sequencing study in which a high abundance memory B-cell clonotype, Ct1, was discovered and used in a virtual screen against a panel of 55 common pathogen antigens done entirely in silico with the two NovaFold-AI-Boltz (NF-B) models1,2. The screen indicated that the receptor binding domain of the SARS-CoV-2 spike protein (S-RBD) is the target antigen of Ct1. Additionally, the models predict Ct1 binds to an epitope in a S-RBD tip region that overlaps the ACE2 interface required for infection of human cells. Finally, we show bio-layer interferometry (BLI) data using purified Ct1 antibody and a series of S-RBD variants demonstrating high affinity Ct1 binding to recent S variants with KD values as low as 1nM. These results, together with the speed of the NF-B software (typically <10 minutes per prediction), provide a glimpse of how virtual screening will become a viable approach for characterizing personal repertoires in the near future.

Antibody sequencing and virtual antigen binding screens

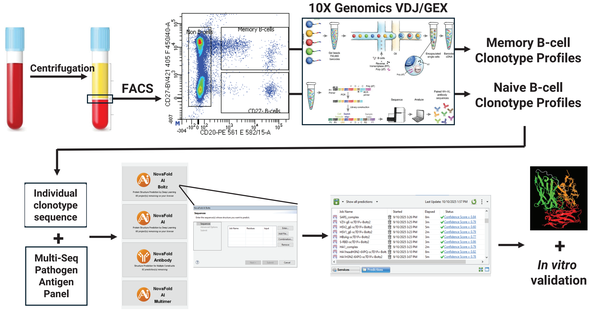


Figure 1. Project design overview.

A schematic of the overall project design is shown in Figure 1. Briefly, white cells are isolated from whole blood samples and subjected to fluorescence-activated cell sorting (FACS) from which naïve and memory B-cell populations are collected. Each population is then processed with the 10X Genomics VDJ/GEX workflow from which expressed antibody and general gene expression profiles are determined for each cell. Individual antibody clonotype protein sequences are then used in virtual screens against a panel of pathogen antigen sequences to identify binding candidates for that antibody using the NovaFold suite of tools. Top scoring candidates from the screens are then evaluated in vitro with purified proteins.

Virtual screening of a known antibody-antigen complex: a positive control

Table 1. In silico screen with scPem-Fv correctly identifies human PD-1 as the target antigen

Primary screen: Boltz-2[a]			Secondary screen: Boltz-1		
Antigen	**Run time**	**Confidence Score**	**Antigen**	**Run time**	**Confidence Score**
PD-1_116aa	4m	0.95	PD-1_116aa	9m	0.92
Cdip_CRM	7m	0.86	Mtub_CFP10	9m	0.72
Ecoli-O157_toxin	3m	0.86	Ecoli-O157_toxin	2m	0.7
HSV1_gE	10m	0.86	HSV1_gE	10m	0.67
Mtub_CFP10	3m	0.86	Cdip_CRM	5m	0.9

[a] Top 5 scoring candidates shown out of 55 tested

To assess the ability of the NF-B implementations to correctly identify a target antigen from a virtual screen, we performed the following positive control experiment. The linear protein sequence of a single chain Fv fragment of Pembrolizumab, a human PD-1 binding antibody, was constructed and used to screen a virtual panel of 56 protein sequences, 55 known pathogen antigens plus human PD-1. Results from the primary screen with NF-B2 were sorted by Confidence Scores and the top five candidates rescreened with NF-B1. Human PD-1 was the top scoring candidate in both screens by a wide margin (Table 1). Further, structural alignment of the predicted scPem-Fv – PD-1 complex with the known crystal structure (PDB: 5B8C) showed the predicted pose was near native with overall RMSD values of the heavy and light chains of <0.5 Å (Figure 2). The results demonstrate that the NF-B models are capable of identifying the correct antigen for an antibody from a panel of decoys and predict the binding pose with high accuracy.

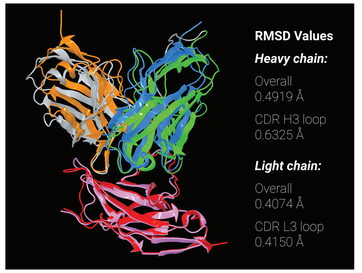


Figure 2. Protean 3D (P3D) structural alignment of the predicted scPem-Fv – PD-1 complex and the crystal structure. Boltz-2 predictions: VH, green; VL, gold; PD-1, red. Crystal structure (PDB entry 5B8C): VH, blue; VL, gray; PD-1, pink. RMSD values for the overall alignment and CDR3 loops of the heavy and light chains were calculated with P3D.

Virtual screening with a memory B-cell clonotype discovered from personal repertoire sequencing identifies SARS-CoV-2 S-RBD as the target antigen

Table 2. Summary statistics for memory B cell VDJ single cell sequencing

	Draw 1[a]	Draw 2[a]
Estimated cell number	2,671	6,977
Cells with H+L pairs	2,534	6,737
Abundant Clonotypes >0.25%	4	2
Total Freq of abundant clonotypes	62	88
Paired clonotype diversity	1,477.99	2,795.85
Total Freq of most abundant clonotype[b]	34	64
Proportion of most abundant clonotype	1.27%	0.92%
Proportion of second most abundant clonotype	0.49%	0.34%

[a] Draws were taken 5 months apart
[b] Same clonotype sequence from both draws

Table 3. Virtual antigen screening results for scTD1-Fv

Primary screen: Boltz-2[a]			Secondary screen: Boltz-1		
Antigen	**Run time**	**Confidence Score**	**Antigen**	**Run time**	**Confidence Score**
HA2-H1N1-SC1918	4m	0.89	SARS-CoV-2_S-RBD	2m	0.84
SARS-CoV-1_S1	4m	0.88	HA2-H1N1-SC1918	2m	0.77
SARS-CoV-2_S-RBD	11m	0.88	SARS-CoV-1_S1	2m	0.75
Mtub_ESAT6	3m	0.87	Mtub_ESAT6	2m	0.72
Mtub_CFP10	10m	0.87	Mtub_CFP10	2m	0.67

[a] Top 5 scoring candidates shown out of 55 tested

We next applied the entire workflow (Figure 1) to discover novel antibodies and predict their target antigens. Briefly, memory and naïve B-cell populations were isolated from two whole blood samples taken from a donor five months apart. Following single cell VDJ/GEX sequencing and analysis, the most abundant clonotypes from each population were chosen for further testing. For memory B-cells, the same clonotype, Ct1, was the most prevalent in both draws representing 1.27% and 0.92% of the sequenced cells, respectively (Table 2). Ct1 heavy and light chain protein sequences were used to construct a single chain variable fragment sequence, scCt1-Fv, which was screened for binding candidates with the 55 antigen sequence panel mentioned above. The top five scoring antigens from the primary screen with NF-B2 were rescreened with NF-B1 from which the S-RBD domain from SARS-CoV-2 emerged as the prime candidate (Table 3). Further, the predicted binding pose indicates that Ct1 binds to the tip region of S-RBD (Figure 3A) and shares several contact residues necessary for viral infection of human cells through interaction with the ACE2 receptor3 (Figure 3B). That suggests Ct1 binding would be sensitive to variants in those positions as are found in more recent strains of the virus.

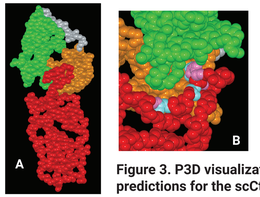
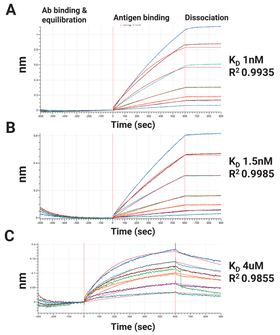


Figure 3. P3D visualization of Boltz structure predictions for the scCt1-Fv – S-RBD complex.

(A) Space-filling model of the overall predicted structure (scCt1-Fv: VH, green; linker, gray; VL, gold; S-RBD: red). **(B)** Zoomed in view of the paratope:epitope interface. Coloring as in A, except S-RBD contact residues with Ct1 CDR H3 or L3 loops in cyan and mauve. Cyan residues also contribute to human ACE2 binding.

In vitro validation of top antigen



To confirm the authenticity of the predicted Ct1 – S-RBD binding, the scCt1-Fv sequence was cloned into a CHO cell expression vector. Following expression and purification (Lytic Solutions, Madison, WI), Ct1 binding to a series of purified SARS-CoV-2 S-RBD variants was tested by BLI on an Octet instrument (Figure **4**). High affinity binding was observed between Ct1 and S-RBD variants from two different Omicron subvariants, JN.1 and BA.2.86 (Lytic Solutions), with KD values of 1nM (Figure **4**A) and 1.5nM (Figure **4**B), respectively. In contrast, S-RBD from the Wuhan reference strain (GenScript, Cat. No. Z03479) showed weak binding with a KD of 4uM (Figure **4**C). These results validate the predicted Ct1 – S-RBD interaction and show antibody specificity for more recent SARS-CoV-2 strains consistent with the donor's infection and immunization history.

Figure 4. Bio-Layer Interferometry (BLI) analysis of scCt1-Fv binding to S-RBD variants. Aligned sensorgram traces from an Octet instrument. Antibody binding/equilibration, Antigen binding and Dissociation stages are indicated. Colored traces correspond to the added antigen concentrations ranging from 0.5 to 10uM for 3A and 3B and 2.5 to 20 uM for 3C. Fitted curves (red lines) used to calculate KD and R2 values. **(A)** Binding profile of scCt1-Fv to S-RBD from the JN.1 strain. **(B)** Binding profile of scCt1-Fv to S-RBD from the BA.2.86 strain. (C) Binding profile of scCt1-Fv to S-RBD from the Wuhan reference strain.

Discussion

The pipeline for in silico screening of antigen panels and in vitro verification of predicted complexes developed by DNASTAR and Scarab Genomics, respectively, show how virtual screening can be used as a powerful first step to identify the binding target for antibodies of unknown specificity. The ever-improving speed of the modeling software together with batch processing of arbitrarily large antigen panels on the cloud make virtual screens fast – for example, the 55 antigen screens described above took just one hour to complete. Moreover, minimizing the number of candidate interactions that need to be experimentally verified leads to significant savings in terms of cost and labor. As the tools become more robust, these advantages have the potential to greatly accelerate not only medical research but also development of new diagnostics and therapeutics as well as personalized medicine through understanding of one's antibody repertoire.

Advantages over in vitro screening

Speed: 1 hour for the primary screen of 55 antigens

Cost effectiveness: nominal cloud compute costs

No protein expression or purification required.

References

1. Wohlwend, J., Corso, G., Passaro, S. et al. (2025) Boltz-1 Democratizing Biomolecular Interaction Modeling. bioRiv.

2. Passaro, S., Corso, G., Wohlwend, J. et al. (2025) Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRiv.

3. Yuan, M., Wu, N.C., Zhu, X. et al. (2020) A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science. **368**. 630.

Visit us at booth 1156 to learn more about how to perform virtual antibody screening with DNASTAR software.