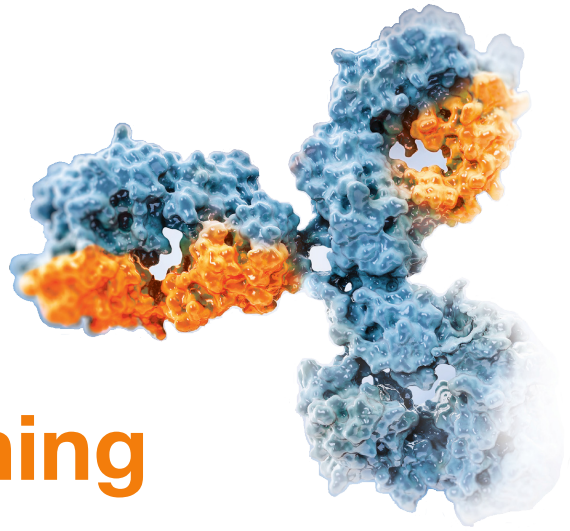
geneious biologics



Align and Prioritize Antibodies for Accelerated Screening

Antibody drug discovery teams need to analyze both sequencing data and results from the assay data associated with them. It is then critical to efficiently query all this information in parallel to accelerate screening and improve candidate selection efficiency.

Geneious Biologics solution accelerates the antibody discovery process by efficiently querying your NGS or Sanger data. With Geneious Biologics, you can:

- Pre-process sequencing raw data (FASTQ, FASTA, etc.) to pair heavy and light chain sequenced separately, merge paired end reads or collapse reads by UMI (Unique Molecular Identifiers), among other preprocessing operations.
- Annotate antibody sequences based on our curated reference databases.
- Obtain a full set of customizable errors, liabilities, and scoring system.
- Filter and cluster sequences by any region (CDRs, FRs, VDJ, etc.) based on identity or similarity (user-defined threshold).
- Align hundreds of antibody sequences and visualize at the same time the assay data associated with those sequences.

Use case: Sequence and assay data analysis of 137 clinical-stage antibodies

The 137 sequences from clinical-stage antibodies (Jain et al. 2017) were uploaded into Geneious Biologics. Heavy and light chains from the same antibody were paired and then annotated using the Antibody Annotator, based on the human reference database. Then the VDJ regions from the Heavy Chains were aligned. Sequences in the alignment were then sorted by “Liability score” to investigate the sequence and liabilities of the most problematic sequences (see Fig. 1 for the Heavy Chain alignment).

Figure 1. Annotation and alignment (VDJ region) of clinical-stage antibody sequences from Jain et al. 2017 sorted by Liability score (ascending order). The liability scores adds all the errors and liabilities per sequence and they are displayed on the left of the alignment as a heatmap column. The top 10 antibodies with the greatest amount of errors and liabilities are shown.

Next, data from the 12 biophysical assays reported in the study were associated with the Antibody Annotator results. Sequences were then aligned at the CDRH3 region and the assay data results were added to the alignment tree as heatmap columns, which can be sorted on multiple values (Fig. 2). Thus, sequencing and all available assay data results can be visualized in parallel, accelerating sequence evaluation in their assay context.

In conclusion, the combination of the flexible filtering operations, together with the alignment with assay data visualisation provides a very comprehensive overview of results. Users can apply these operations to prioritize candidates or investigate specific antibodies of interest.

Reference: Jain et al. 2017. Biophysical features of the clinical mAb landscape. PNAS 114(5) 944-949.

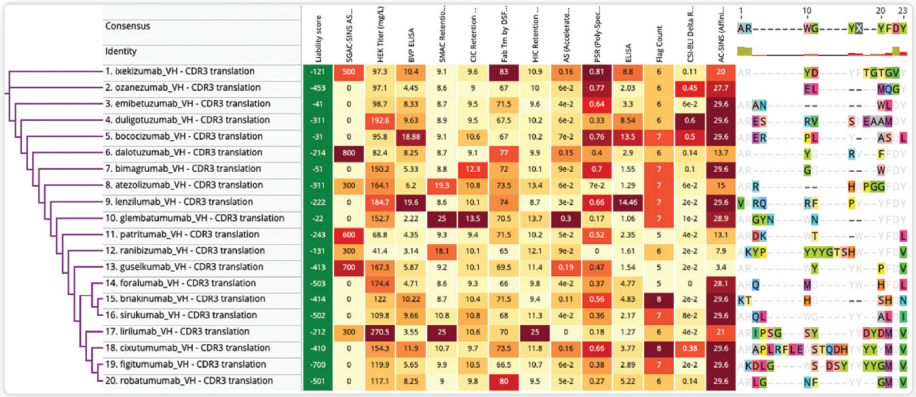


Figure 2. Alignment and associated phylogenetic tree of 20 CDRH3 antibody sequences from Jain et al. 2017. Liability scores, flag counts and data from the 12 biophysical assays associated with the sequences are shown in the heatmap columns. The top 20 antibodies with the greatest flag counts are shown. Flag Count is the number of assays with abnormal values based on thresholds of biophysical properties defined in the manuscript.