

Corresponding author(s): Wen Zhang

Last updated by author(s): Aug 5, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software or code was used.

Data analysis

Custom code for data analysis is available at GitHub (<https://github.com/peggy-minyao/AbOptimizer>). We use the AbLang heavy chain and light chain language models from <https://github.com/TobiasHeOl/AbLang>. The affinity and specificity predictor of emibetuzumab are from https://github.com/Tessier-Lab-UMich/Emi_Pareto_Opt_ML. The specificity predictor of trastuzumab is from https://github.com/dahjan/DMS_opt/tree/master. We employed NetMHCIIpan version 4.1 to assess the binding affinity between the antibody and MHC-II. We used the abYsis webtool version 3.4.1 (<http://www.abysis.org/abysis/index.html>) for sequence alignment. The docking software we used includes ZDOCK version 3.0.2 and ZRANK version 2.0. Flow cytometry analyses were performed using NovoExpress version 1.5.6. Visualization and hydrogen bond analysis were performed using the open-source version of PyMOL. EC50 values were calculated using a four-parameter fit in python. We also performed data analysis with Python version 3.10 and R version 4.3.2. Detailed information about our model development in Python is available at <https://github.com/peggy-minyao/AbOptimizer/blob/main/requirements.txt>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The datasets used in this study were obtained from previous studies and are publicly available. The emibetuzumab dataset is available from BioProject (accession code: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA850089>) [\cite{makowski2022co}](#). The trastuzumab datasets are available from GitHub (https://github.com/dahjan/DMS_opt/tree/master) [\cite{https://github.com/dahjan/DMS_opt/tree/master}](https://github.com/dahjan/DMS_opt/tree/master) and <https://github.com/AbSciBio/unlocking-de-novo-antibody-design/tree/main>) [\cite{mason2021optimization,variable2023tras}](https://github.com/AbSciBio/unlocking-de-novo-antibody-design/tree/main). The database from <https://opig.stats.ox.ac.uk/webapps/oas/> [\cite{OAS}](https://opig.stats.ox.ac.uk/webapps/oas/) and <https://www.uniprot.org/uniref> [\cite{UniRef90}](https://www.uniprot.org/uniref) are also used in this study.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Our manuscript does not involve research related to gender or sex.
Reporting on race, ethnicity, or other socially relevant groupings	Our manuscript does not involve research related to race, ethnicity, or, other socially relevant groupings.
Population characteristics	Our manuscript does not involve research related to population characteristics.
Recruitment	Our manuscript does not involve research related to recruitment.
Ethics oversight	Our manuscript does not involve research related to ethics oversight.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the experiment validation of the optimized antibodies, we selected 5 optimized variants for emibetuzumab, 8 for siltuximab and 5 for trastuzumab. No statistical methods were used to predetermine the sample size of 5 and 8 independent antibodies. The sample size of 5/8 was selected on the basis of the experimental effort required to express, produce and experimentally verify the affinity and specificity for each antibody. Demonstrating robust performance with fewer experimental samples provides stronger evidence of the algorithm's effectiveness.
Data exclusions	No data were excluded.
Replication	All experimental data used for analysis were obtained from three independent replicates.
Randomization	Samples were allocated into groups independent of the experimental condition.
Blinding	All experiments were conducted by researchers who were blind to the identities of the samples, including the predicted scores generated by the models.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The primary antibodies used in this study were emibetuzumab, trastuzumab and siltuximab and their optimized variants. These antibodies were generated or obtained as described in the Methods sections "Expression, purification and quality control of IgGs". Their antigen binding and specificity were validated by ELISA and CHO-cell-based flow cytometry assays, as described in the Methods sections "Affinity and non-specific binding assays" and shown in Fig. 5. 1 mg/mL anti-human IgG antibody conjugated to HRP (Jackson ImmunoResearch, 109-005-064, polyclonal) was diluted 5000 fold before used for antibody affinity assay and antibody specificity assay. 1 mg/mL Alexa Fluor 647-conjugated goat anti-human IgG secondary antibody (Jackson ImmunoResearch, 109-605-008, polyclonal) was diluted 1000 fold before used for antibody specificity assay.
Validation	The affinity of the antibodies was assessed by antigen-binding ELISA. The specificity of the antibodies was evaluated by OVA-based non-specific binding ELISA and CHO cell-based flow cytometry. In addition, thermostability measurements, SDS-PAGE, and analytical SEC were performed to assess the thermal stability, expression/purification quality, and monomeric purity of the antibody preparations. All secondary antibodies were validated by the supplier, and detailed information can be found on the supplier's website.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HEK293 cells purchased from Zhuhai KAIRUI Biotech Inc. were used for antibody expression. CHO cells purchased from Thermo Fisher Scientific Inc.(A29133) were used for antibody specificity assay.
Authentication	HEK293, CHO cells were authenticated by the provider.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Plants

Seed stocks	Our manuscript does not involve research related to plants.
Novel plant genotypes	Our manuscript does not involve research related to plants.
Authentication	Our manuscript does not involve research related to plants.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

In the CHO-cell flow cytometry experiment, 1×10^6 suspended CHO cells were fixed and permeabilized using 4% formaldehyde and 0.2% Triton X-100. Then, 100 μL of antibody solution at a concentration of 15 $\mu\text{g}/\text{mL}$ was added and incubated with the cells in the dark for 15 min. After washing three times with 1% FBS/PBS, cells were incubated with Alexa Fluor 647-conjugated goat anti-human IgG secondary antibody (Jackson ImmunoResearch, cat. no. 109-605-008; polyclonal; dilution, 1:1000; working concentration, 1 $\mu\text{g}/\text{mL}$) in 100 μL of 1% FBS/PBS in the dark for another 15 min. Cells were washed three times again and resuspended in 1% FBS/PBS before flow cytometry analysis.

Instrument

For antibody specificity assay, the data was collected using a Agilent NovoCyte 2060R flow cytometer.

Software

NovoExpress1.5.6

Cell population abundance

For antibody specificity assay, the first gate for CHO cells had >85% of the total population, the second gate for single cells had >80% of the remaining population.

Gating strategy

For antibody specificity assay, pre-gating was done to identify single CHO cells by gating on SSC-height vs. FSC-height, then by FSC-height vs. FSC-area.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.