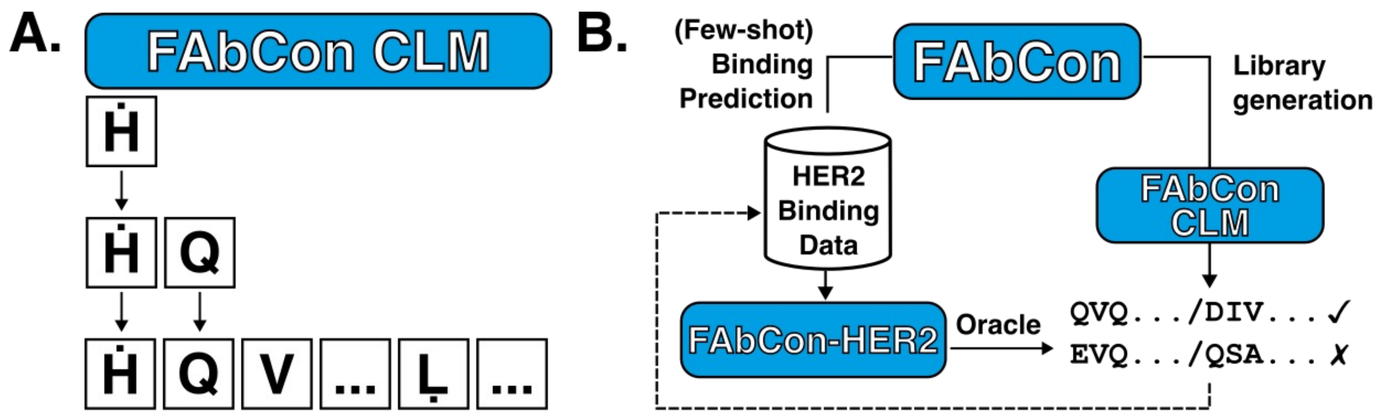
A generative foundation model for antibody sequence understanding

Justin Barton, Aretas Gaspariunas, David A. Yadin, Jorge Dias, Francesca L. Nice, Danielle H. Minns, Olivia Snudden, Chelsea Povall, Sara Valle Tomas, Harry Dobson, James H.R. Farmery, Jinwoo Leem, Jacob D. Galson

Alchemab Therapeutics, East Side, Office 1.02, Kings Cross, London

Introduction

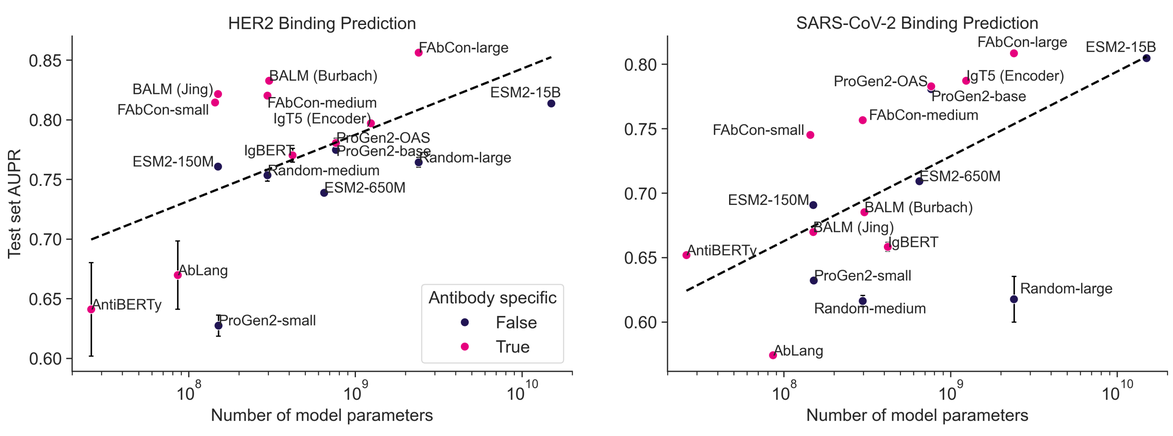
- We introduce FAbCon, a generative antibody-specific language model trained on over 770 million sequences. Our model is based on the Falcon LLM [1].
- Three variants of FAbCon at varying levels of parameter scale are available on Hugging Face for non-commercial use:
  - FAbCon-small (144M)
  - FAbCon-medium (297M)
  - FAbCon-large (2.4B parameters)
- Using an antibody-specific corpus and increasing model scale are tied to enhanced performance. Furthermore, FAbCon can design antibody sequences with good developability potential.



A. FAbCon is pre-trained using the causal language modelling objective. FAbCon is pre-trained on a mixture of both unpaired (heavy or light chain only) and paired (heavy and light chain) antibody sequences from the Observed Antibody Space (OAS) [2] and proprietary data. **B.** FAbCon is a foundation model that can be fine-tuned for antigen binding prediction. It then acts as an 'oracle' that can screen antibody libraries that are engineered *in vitro* or *in silico*. Screened sequences can then be fed back to 'close the loop' for subsequent rounds of design.

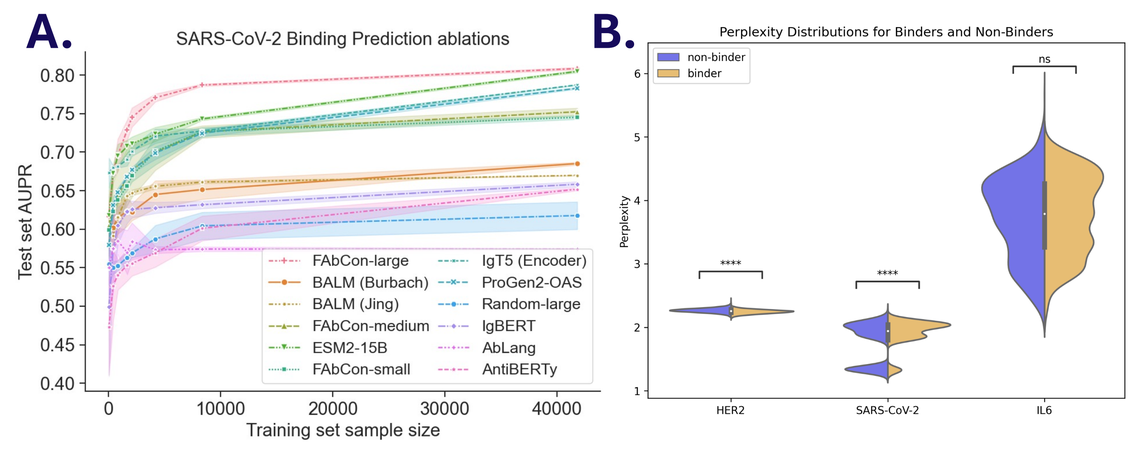
Antibody-specific corpus and model scale are key factors to binding prediction performance

FAbCon is parameter efficient



We fine-tuned 15 protein language models and antibody-specific language models for antigen binding prediction (HER2 and SARS-CoV-2) [3, 4]. Language models pre-trained on antibody-specific corpora (pink dots) tend to have higher area under the precision-recall curve (AUPR) scores than general protein language models. Furthermore, performance is closely tied with scale; models with at least one billion parameters (IgT5, ESM2-15B, and FAbCon-large) are the best performers for both HER2 and SARS-CoV-2. As a control, we also fine-tuned FAbCon-medium and FAbCon-large without any pre-training ("random-medium" and "random-large") and find that pre-training boosts performance. Error bars represent standard deviation across five seeds. The dotted line represents a regression fit between the log number of parameters with respect to AUPR.

Few-shot but not zero-shot for binding prediction

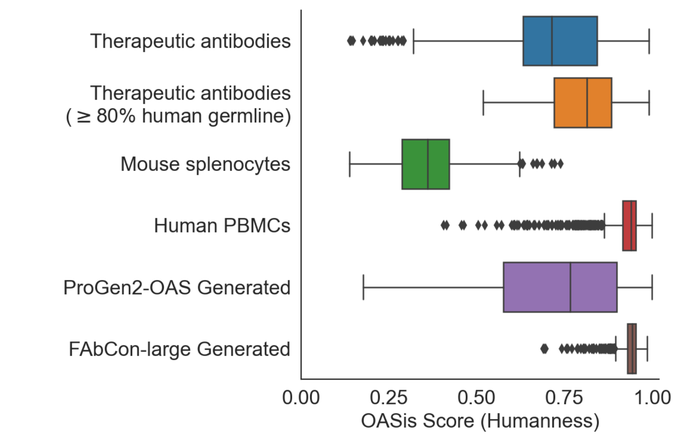


A. We ablate the training set sample size for fine-tuning and find that FAbCon-large is the most performant at low data regimes (<1000 samples). It also shows the lowest standard deviation across five seeds (shaded areas), suggesting that FAbCon-large can be a powerful oracle with relatively small datasets.

B. We also test for zero-shot binding prediction using FAbCon-large's perplexity values. Unlike previous reports [5], we find that perplexity is not a good discriminator of antigen binders and non-binders. Though perplexity distributions have statistically significant differences ($p < 1e-4$), no clear decision boundary can be drawn. This suggests that some, but not zero, data is necessary for use with language models.

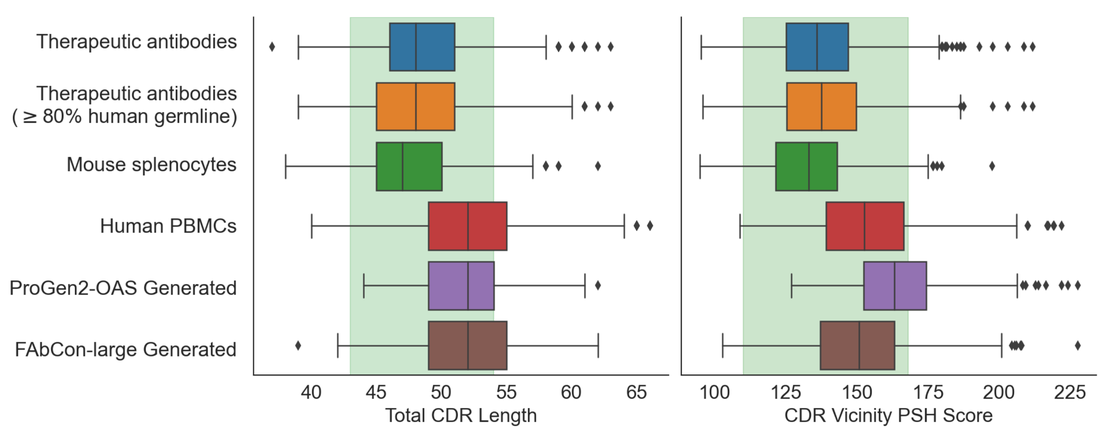
FAbCon can generate molecules with good developability potential

FAbCon-generated antibodies show promising predicted humanness



We predict humanness using OASis [6]. Paired antibodies generated *de novo* by FAbCon-large have high OASis scores that resemble the distribution of humanness scores of paired antibodies from human PBMCs. Therapeutic antibodies' humanness scores are expected to be slightly lower as they can be humanized or derived from e.g. phage display.

FAbCon-generated antibodies meet TAP2-predicted developability criteria and are similar to known therapeutic antibodies



We measure structure-related developability properties using the Therapeutic Antibody Profiler (TAP2) [7]. Antibodies generated by FAbCon-large tend to sit within acceptable regions (green shaded area), as defined by therapeutics in the market. For the remaining TAP2 properties, check out our paper!

Summary

- FAbCon is a 2.4B parameter generative decoder model that has leading antigen binding prediction performance and can generate antibodies with desirable properties for therapeutics.
- Scaling and specialization are both shown to improve predictive and generative performance
- Models available on Hugging Face for non-commercial use. Please chat to us for a commercial license!

References

1. A. Radford et al., Learning Transferable Visual Models From Natural Language Supervision, *arXiv*, 2021.
2. Kovaltsuk et al., Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires, *J Immunol*, 2018.
3. Mason et al., Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning, *Nat Biomed Eng*, 2021.
4. Engelhart et al., A dataset comprised of binding interactions for 104,972 antibodies against a SARS-CoV-2 peptide, *Sci Data*, 2022.
5. Nijkamp et al., Exploring the boundaries of protein language models, *Cell Systems*, 2023.
6. Prihoda et al., BioPhi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning, mAbs, 2022
7. Raybould et al., Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling, *Commun Biol*, 2024