

Generative Language Model for Antibody Sequence Design (IgLM)

C17062

Once a license request has been received and approved, JHU will send a license for execution via DocuSign to the Requestor's or Authorized Representative's email. After full execution of the license, the technology materials will be made available for download via this portal.

Discovery and optimization of monoclonal antibodies for therapeutic applications relies on large sequence libraries, but is hindered by developability issues such as low solubility, low thermal stability, high aggregation, and high immunogenicity. Generative language models, trained on millions of protein sequences, are a powerful tool for on-demand generation of realistic, diverse sequences. We present Immunoglobulin Language Model (IgLM), a deep generative language model for creating synthetic libraries by re-designing variable-length spans of antibody sequences. IgLM formulates antibody design as an autoregressive sequence generation task based on text-infilling in natural language. We trained IgLM on 558M antibody heavy- and light-chain variable sequences, conditioning on each sequence's chain type and species-of-origin. We demonstrate that IgLM can generate full-length heavy and light chain sequences from a variety of species, as well as infilled CDR loop libraries with improved

Technology ID

C17062

Category

Computer Electronics & Software

Authors

Jeffrey J. Gray

Jeffrey Ruffolo

Richard Shuai

[View online](#)



developability profiles. IgLM is a powerful tool for antibody design and should be useful in a variety of applications.

IgLM is available via a GitHub repository. Licensees may download it there. The completed license gives permission to use that code commercially.

References

1. Richard W. Shuai, Jeffrey A. Ruffolo, Jeffrey J. Gray ,
<https://www.biorxiv.org/content/10.1101/2021.12.13.472419v2>,
bioRxiv