Antibody Deimmunization via Targeted, Machine Learning Guided Sequence Design Using Fine Tuned Protein Language Models for Computationally Efficient Design

Pan, Amelie (School: Pacific Collegiate School)

Rheumatoid arthritis (RA) is an autoimmune disease affecting approximately 18 million people worldwide. Many treatments rely on monoclonal antibody drugs that block inflammatory signals such as TNF-a and IL-6. However, up to 40% of patients develop anti-drug antibodies (ADAs), which reduce treatment effectiveness by targeting these therapies as foreign proteins. Reducing immunogenicity while preserving drug function remains a major challenge in biologic design. This project presents a computational pipeline for identifying antibody sequence modifications that reduce predicted immunogenicity while maintaining structural stability. Antibody sequences were analyzed using NetMHCIIpan to identify immune-reactive regions, and alternative amino acids were generated using the protein language model ESM-2. A key innovation of this work is a targeted masking strategy that biases model training toward immunogenic "hotspot" residues, ensuring precise edits in high-risk regions. Through multiple iterations of optimization and hybridization with biological antibody design principles, the final approach achieved a 440% improvement over random masking - a 66% success rate in reducing predicted immunogenicity (meaning the model was able to successfully produce antibodies with reduced GIS scores for 66% of the antibodies in the test set) out of 106 test sequences, with an average sequence similarity of 88.3% and 13.5 targeted mutations per sequence. These results demonstrate that protein language models can be guided with biologically-informed strategies to perform targeted, clinically relevant antibody deimmunization, providing a robust framework for ML-assisted therapeutic design.

Awards Won: