Evaluating synergies among generative design models for multi-objective optimization of drug-like proteins

Jung-Eun Shin
Seismic Therapeutic
Watertown, MA 02472
june.shin@seismictx.com

Nathan J. Rollins
Seismic Therapeutic
Watertown, MA 02472
nathan.rollins@seismictx.com

Jordan M. Anderson
Seismic Therapeutic
Watertown, MA 02472
jordan.anderson@seismictx.com

Grace Carey
Seismic Therapeutic
Watertown, MA 02472
grace.carey@seismictx.com

Allison Colthart
Seismic Therapeutic
Watertown, MA 02472
allison.colthart@seismictx.com

Thomas Hopf
Seismic Therapeutic
Watertown, MA 02472
thomas.hopf@seismictx.com

Ivan Mascanfroni
Seismic Therapeutic
Watertown, MA 02472
ivan.mascanfroni@seismictx.com

Jyothsna Visweswaraiah
Seismic Therapeutic
Watertown, MA 02472
jyothsna.visweswaraiah@seismictx.com

Yi Xing
Seismic Therapeutic
Watertown, MA 02472
yi.xing@seismictx.com

Kevin L. Otipoby
Seismic Therapeutic
Watertown, MA 02472
kevin.otipoby@seismictx.com

Nathan Higginson-Scott
Seismic Therapeutic
Watertown, MA 02472
nathan.higginson-scott@seismictx.com

Ryan Peckner
Seismic Therapeutic
Watertown, MA 02472
ryan.peckner@seismictx.com

Abstract

In recent years, the field of AI for protein design has made tremendous advances in the generation of high quality proteins. Much of the generative focus has been on producing de novo sequences and structures or in producing large libraries to screening for novel function, thereby leading to the discovery of a starting protein candidate. The downstream pipeline for drug discovery for engineering the protein candidate into a suitable therapeutic, however, has been much less a focus for generative protein design, especially for non-antibody biologics. For a protein to be a suitable therapeutic, it must not only have the desired function at a therapeutically relevant efficacy, it must also be manufacturable, safe, and non-immunogenic. Multi-objective optimization methods are uniquely suited to providing a one-pot model to simultaneously optimize for function and all desired drug-like properties. In this work we present individual models for the generation

AI for New Drug Modalities at NeurIPS 2024.

of fit, functional protein sequences and the prediction of drug-like properties, as well as generative models that harness the totality and subsets of these individual models in generating therapeutically developable proteins. We explore the synergy of these machine learning methods in modular generative multi-objective optimization models by comparing training data, generative model architectures, generation methods, and implementation of various developability constraints. By generating sequences from these modular multi-objective optimization models and experimentally screening proteins derived from IdeS (immunoglobulin degrading protease) for stability and function, we demonstrate that our models are able to consistently generate highly fit proteins that have been optimized to possess drug-like properties *in silico*.