A076 · Generating proteins with computationally predicted functions and multiple states via multimodal diffusion

Presenter: Anna Sappington — MIT, Harvard Medical School

Authors: Bowen Jing, Anna Sappington, Mihir Bafna, Ravi Shah, Adrina Tang, Rohith Krishna, Adam Klivans, Daniel J. Diaz, Bonnie Berger

Session: Day 1 · Thu Oct 1, 2026 · 1:15–2:15 PM

Generating proteins with the full diversity and complexity of functions found in nature is a grand challenge in protein design. Here, we present ProDiT, a multimodal diffusion model that unifies sequence and structure modeling paradigms to enable the generation of proteins with computationally predicted functions at scale. Trained on sequences, 3D structures, and annotations for 128M proteins across the evolutionary landscape, ProDiT generates diverse, novel proteins that are computationally predicted to preserve known active and binding site motifs and can be conditioned on a wide range of molecular functions, spanning 463 Gene Ontology terms. We introduce a diffusion sampling protocol to design proteins with multiple functional states, and illustrate this protocol by computationally scaffolding enzymatic active sites from carbonic anhydrase and lysozyme such that they are predicted to be deactivated by a calcium effector. Our results showcase ProDiT's capacity to satisfy design specifications inaccessible to existing generative models, thereby expanding the protein design toolkit.

Generating proteins with computationally predicted functions and multiple states via multimodal diffusion

Bowen Jing[1,†], Anna Sappington[1,2,†*], Mihir Bafna[1], Ravi Shah[3], Adrina Tang[1], Rohith Krishna[4], Adam Klivans[3], Daniel J. Diaz[3], Bonnie Berger[1]

[1]Massachusetts Institute of Technology [2]Harvard Medical School [3]University of Texas at Austin [4]University of Washington
[†]*These authors contributed equally.* [*]*Presenting author.*

Generative models trained on datasets of protein sequences or backbone structures have enabled significant advances in *de novo* protein design. However, existing paradigms typically focus only on structure or sequence in isolation and, therefore, cannot capture the complete functional landscape of natural proteins. Neither family of approaches can target protein *dynamics*, a key feature of natural proteins, which often adopt multiple states to modulate their functions. We hypothesize that a generative model, efficiently trained at scale with both sequence and structure would unlock many crucial protein design capabilities. Here, we present ProDiT (Protein Diffusion Transformer), a framework that unifies structural and sequence generative modeling into a multimodal diffusion generative process, to enable these design capabilities. ProDiT leverages direct diffusion-based modeling of 3D structural coordinates alongside amino acid tokens using a fast, scalable transformer architecture. ProDiT performs diffusion directly in the native state spaces of both sequence and structure with a simple transformer architecture, enabling fast, scalable multimodal training on 128M proteins with functional annotations. Our formulation enables *in silico* generation quality that matches or exceeds that of specialized structure generative models across all sequence lengths while also significantly exceeding the sequence generation quality of dedicated protein language models (PLMs). Notably, this performance is achieved with a relatively simple, compact model.

We explore ProDiT's broad understanding of the protein sequence and structural landscape by conditioning generation on functional descriptors in the form of molecular function Gene Ontology (GO) terms. We conduct comprehensive benchmarking of ProDiT across 915 diverse GO terms representing highly specific descriptions of protein function and find generations predicted as functional for 463 of these terms, including terms represented in less than 0.01% of training examples in UniProtKB. Many generations have only moderate structural similarity to known functional proteins (TM-score < 0.7), showing an ability to generalize across the structural landscape. We also build upon techniques from image generative modeling for prompt adherence to further boost GO term adherence. Structural alignments of our generations against known functional proteins indicate that generated structures reproduce the geometry of known active-site residues, despite low global sequence similarity. Our designs additionally recover binding residues to enzymatic cofactors.

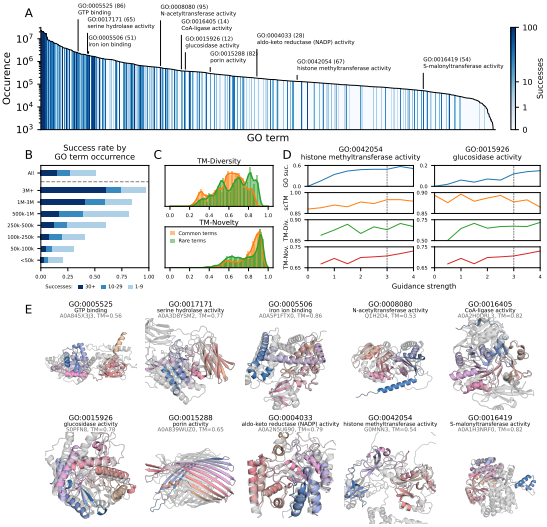

Figure 1: In silico assessment of molecular function conditioning. (A) Successful generations per GO term. (B) Proportions of GO terms with successful designs stratified by occurrence frequency. (C) Distribution of TM-score diversity and novelty. (D) Classifier-free guidance for two selected GO terms. (E) For select terms, the most novel generation superimposed onto its closest AFDB match (grey).

Leveraging ProDiT's multimodal capabilities, we develop a principled protocol for the design of dynamic, multistate proteins, derived from probabilistic graphical models. When distinct motifs are provided for the desired output protein states, our protocol generates two separate structural scaffolds coupled to a common sequence in a single generative rollout. As such, our approach represents a significant advance over prior protocols requiring brute-force sampling of plausible alternative conformations, previously thought necessary for such a protein design task. We apply this capability to computationally scaffold active sites from carbonic anhydrase and lysozyme such that they are predicted to be modulated by calcium binding, an important step towards the design of custom and controllable enzymes.