Chapter 6

CONCLUDING REMARKS

6.1 From foldtuning to foundation models

Foldtuning is not an an algorithm or model collection frozen in time. Indeed, a leading strength of foldtuning is its modularity and generalizability to complex design tasks. Much of this modular nature is bestowed by the GAN-like division into generator and discriminator. In principle, any PLM could act as the generator, provided an appropriate procedure for sampling from the model in question. The choice of ProtGPT2 as the initial generative model for foldtuning was motivated by its relative success at proposing novel, reasonable, and representative protein sequences, but other promising approaches — e.g. direct embedding-to-sequence decoders for encoder-only architectures (as recently demonstrated for ESM2) — could be swapped in with minimal implementation burden (Chen et al., 2024). The discriminator side is in some sense endlessly flexible. We began with a structure-based filter to pursue minimal constraints on individual fold families, but foldtuning was designed to be amenable to an arbitrary scoring function — even an ensemble of scoring functions — from predicted stability or optimal pH, to active site preorganization, to molecular-dynamics-derived root mean square fluctuation (RMSF), and model-provided confidence metrics, depending on the exact problem of interest. In that regard, structure-first foldtuning may be considered the first such spinoff of foldtuning, replacing scoring that favors structural matches with an objective that rewards three-dimensional compactness and "anti-matching" so as to prefer structural novelty.

Other discriminator-side changes could confer improvements in compute perfomance and overhead. The total compute burden (cost + time) of foldtuning, as currently implemented, is set by the structure prediction step — even a few seconds of inference time per sequence adds up to four GPU-hrs for four rounds of foldtuning with pre-evotuning.[1] Replacing the time-consuming explicit structure prediction step with conversion to a suitable one-dimensional representation of structure, such as the increasingly utilized 3Di sidechain-aware structural alphabet developed for

[1] Benchmarked on a single NVIDIA A100 GPU with 80GB of memory; the optimal monetary vs time cost tradeoff will depend on the number of foldtuned models required, hardware technical specifications, and highly variable capital and/or hourly cost differences between providers.

Foldseek, could accelerate foldtuning by as much as ~ 10x over current internal benchmarks (Heinzinger et al., 2023; van Kempen et al., 2023).

As a unit, also, foldtuning need not be an end unto itself. With its iterative update structure, foldtuning is architecturally amenable to direct integration[2] of experimental measurements — positive and negative — of generated variants via reinforcement learning (RL). Incorporating RL atop foldtuning is a logical next step for guiding models that have already learned sequence novelty under hidden language rules towards empirical evidence of function, whether for the experimental results presented in this work or for any arbitrary target and/or assay. Beyond improving future batches of generated proteins based on real-world data, we envision two additional related model-side advances of import for AI-guided synthetic biology. First, recent theoretical findings on general LLMs and hands-on application of chemical language models for small-molecule representation and generation have independently pushed back on the axiom that breadth and novelty are incompatible and argued that training on labeled positive and negative examples mitigates the dilemma (Kalavasis et al., 2025; Skinnider, 2024). Consequently, we can imagine incorporating information from negative examples — both those filtered out *in silico* and those deriving from experiment — into a single foldtuning foundation model that achieves full structural coverage (including novel domains) without mode collapse or hallucination. Second, foldtuning can form one end of an end-to-end model linking sequence-level specification of, e.g. a binder with an arbitrary agonism/antagonism profile against cell-surface receptors, to single-cell transcriptomic readout for design of bespoke cell-signaling programs.

6.2 Producing and propagating protein novelty across scales

In this work, foldtuning was restricted to single-domain targets, a biophysically meaningful, and well-annotated level at which to first segment. Generation and optimization of individual domains offers much to be excited about, including cytokine- and chemokine-like binders as in the example posed above, host-defense peptide-mimicking antimicrobials, and biosensor toolkits enabled by fluorescent protein property expansion, all areas of ongoing interest. Complex systems of proteins, on the other hand, are built up in layers of physical organization, compartmentalization, and interaction. As an agent of domain diversification, foldtuning can power the design of full-blown protein *systems* for AI-guided synthetic biology and cellular engineering, such as signaling cascades assembled out of foldtuned kinases, SH2s, and

[2] As opposed to ranking, selection, and updating by finetuning.

SH3s, or gene regulatory networks based on the many flavors of seemingly highly-designable DNA-binding domains, or multi-step pathway catalytic machinery for xenobiotic metabolism.

Lastly, if there is one defining theme of this thesis, it is that PLMs — whether through foldtuning, MHMC sampling, or any other method — are producers and propagators of novelty in sequence, structure, and function. With the right steering and forcing, PLMs readily expand the boundaries of valid protein-space at greater rates than the accumulation and processing of (meta)genomic data can. Putting generative novelty ahead of prespecified phenotype leads to new fold-centered language rules, evolution-esque innovation of structure and function, and mechanistic hints towards the fundamental constraints that dictate the structure→function transition. Ruminating on the virtues of novelty-first methods for looking forward and backwards in time leads to persistent open questions including: What can alternate sequence rules reveal about the primordial emergence of the first proteins? What are the smallest collections of sequences and structures that can sustain the essential functions of a minimal cell and how do they overlap (or not) with what we observe in nature today? How can we leverage the "structure of feature-space" — a PLM's internal navigational charts, as it were — to further accelerate the search for new-to-nature sequences, structures, and functions?

Peering a final time at the sequence→structure map, we have, through several strands of novelty-directed exploration, replaced certain mythical monsters with the outlines of heretofore unknown landmasses; the challenge persists to fully characterize and capitalize on all that these addenda confer.