Abstract

Proteins are among the fundamental molecular components of living systems, carrying out a wide range of structural, catalytic, and regulatory functions that are essential for life at any level. Understanding, predicting, and ultimately controlling the relationship between protein sequence, structure, and function is a central problem in modern biology, with profound implications for medicine, biotechnology, and fundamental science. Over the last two decades, the advent of high-throughput sequencing technologies and the explosion of genomic data have made it possible to approach this problem from a data-driven perspective, first through the tools of inferential statistical physics and, more recently, through machine learning. While these two frameworks have always been historically intertwined, they present different trade-offs: statistical physics models are typically interpretable and grounded in explicit probabilistic principles, whereas modern machine learning architectures offer greater expressive power at the cost of transparency and interpretability. In this thesis, we study this trade-off by developing and analysing extensions of statistical physics models that incorporate structural elements inspired by modern machine learning models, with the goal of improving expressivity while preserving a clear statistical interpretation.

After introducing the biological background in Chapter 1, we review in Chapter 2 the statistical physics approach to protein sequence modeling, focusing in particular on Direct Coupling Analysis (DCA) as a maximum-entropy framework for inferring evolutionary constraints from multiple sequence alignments. In Chapter 3, we turn to machine learning models, tracing the progression from early sequence-based approaches to transformer architectures and large-scale models such as AlphaFold2, which have revolutionized protein structure prediction.

Building on these foundations, Chapter 4 presents the first original contribution of this thesis: AttentionDCA, a model that bridges statistical physics and deep

learning by interpreting the attention mechanism as a structured interaction term analogous to Potts couplings. This perspective provides a low-rank parametrization of DCA, enabling a reduction in model complexity while preserving predictive performance, and offering a principled framework to analyze the internal representations of transformer-based architectures. In Chapter 5, this approach is extended by exploiting the flexibility of the attention-based parametrization to decouple and control different components of the interaction model, naturally enabling its application to protein–protein interaction modeling and paralog matching.

Finally, in Chapter 6, we introduce FeatureDCA, a second contribution that extends autoregressive DCA to a conditional generative framework. By coupling sequence probabilities to low-dimensional, biologically meaningful features, FeatureDCA enables controlled sampling of protein sequences in specific regions of sequence space, while maintaining the interpretability and statistical grounding of DCA-based models. This formulation provides a transparent alternative to latent-variable approaches commonly used in deep learning, and demonstrates that controllable generative modeling can be achieved within a shallow statistical framework.

Overall, this work shows that the apparent dichotomy between interpretability and expressivity can be partially reconciled by reinterpreting modern machine learning architectures through the lens of statistical physics and, conversely, by extending classical models with structured and controllable representations drawn from machine learning.