HP Lattice Folding With Variational Annealing

Shoummo A. Khandoker [1] Estelle M. Inack [2,3,4] Mohamed Hibat-Allah [2,3,5]

[1]Indiana University [2]Perimeter Institute [3]University of Waterloo [4]yiyianQ [5]Vector Institute

Association for the Advancement of Artificial Intelligence

Abstract

Understanding the principles of protein folding is a cornerstone of computational biology, with implications for drug design, bioengineering, and the understanding of fundamental biological processes. Lattice protein folding models offer a simplified yet powerful framework for studying the complexities of protein folding, enabling the exploration of energetically optimal folds under constrained conditions. However, finding these optimal folds is a computationally challenging combinatorial optimization problem. **In this work, we introduce a novel upper-bound training scheme that employs masking to identify the lowest-energy folds in two-dimensional Hydrophobic-Polar (HP) lattice protein folding. By leveraging Dilated Recurrent Neural Networks (RNNs) integrated with an annealing process driven by temperature-like fluctuations, our method accurately predicts optimal folds for benchmark systems of up to 60 beads.** Our approach also effectively masks invalid folds from being sampled without compromising the autoregressive sampling properties of RNNs. This scheme is generalizable to three spatial dimensions and can be extended to lattice protein models with larger alphabets. Our findings emphasize the potential of advanced machine learning techniques in tackling complex protein folding problems and a broader class of constrained combinatorial optimization challenges.

HP Protein Model

1. HP proteins have the following properties:

- each bead is either Hydrophobic H or Polar P
- fold is in discrete space
- fold obeys self-avoiding walk constraint

$$\forall i \neq j; (x_i, y_i) \neq (x_j, y_j) \tag{1}$$

- quality or energy of a fold is the number of H-H contacts (see dashed line in figure)

2. Despite the simplified model, folding is still **NP-Hard** with naive algorithm complexity of $O(3^N)$.

3. For an HP chain of size $N+1$, a fold can be encoded as a vector $\in [0,1,2,3]^N$. In the figure below, the fold can be encoded as $\{0,2,0,2,1\}$; this is interpreted as folding the second bead LEFT, third bead UP, fourth bead LEFT, fifth bead UP, and sixth bead RIGHT.

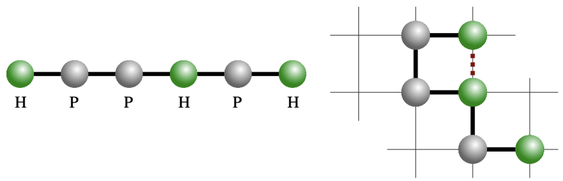

Figure 1.

Generating Folds - Masking Invalid Folds

In our distribution of folds $P_{\boldsymbol{\theta}}$, invalid folds that do not respect the self-avoiding walk (SAW) constraint exist and are prone to be sampled. We prevent sampling such folds by 'masking' actions that lead to breaking the SAW constraint.

Generating Folds

We use a variant of Recurrent Neural Network (RNN) architecture with residual connections called the Dilated RNN to generate folds. The fold is generated **sequentially and autoregressively** from distribution $P_{\boldsymbol{\theta}}$ by sampling an action $\in \{0,1,2,3\}$ for each bead, starting from the second bead to the last.

The generation process is as follows:

- for the i^{th} bead, sample the i^{th} action $\in \{0,1,2,3\}$
- the encoding of the i^{th} action is concatenated with the encoding of the $i+1^{\text{th}}$ bead and used as input to sample the action of the $i+1^{\text{th}}$ bead - Autoregressive nature
- $i := i+1$ and repeat - Sequential nature

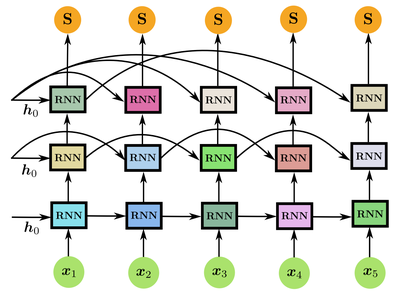

Figure 2. The dilated RNN architecture with long-range residual connections that help better capture folding dynamics. Input x_i is the concatenation of the i^{th} bead and the $i-1^{\text{th}}$ action. The i^{th} output is the action for the i^{th} bead sampled after the softmax layer 'S'.

Upper-bound Loss

One can train this RNN parameters $\boldsymbol{\theta}$ by minimizing the Kullback–Leibler (KL) divergence between the true Boltzmann distribution and the RNN parameter-induced distribution $P_{\boldsymbol{\theta}}$. The KL divergence can be minimized by using the

Variational free energy loss:

$$F_{\boldsymbol{\theta}}(T) = \langle E \rangle - TS_{\boldsymbol{\theta}} \Rightarrow \mathcal{L}(T) = \sum_d P^{\perp}_{\boldsymbol{\theta}}(d) \log(P_{\boldsymbol{\theta}}(d)) \left(E(\boldsymbol{d}) + T \log(P^{\perp}_{\boldsymbol{\theta}}(\boldsymbol{d})) \right) \tag{2}$$

However, we find that the free energy loss $\mathcal{L}(T)$ is not stable during training; we hypothesize that this is due to the perpetual action masking. Therefore, we derive an upper-bound loss $\tilde{\mathcal{L}}(T)$ that empirically demonstrates better performance.

Derived Upper-bound of Variational free energy loss:

$$\tilde{\mathcal{L}}(T) = \sum_d P^{\perp}_{\boldsymbol{\theta}}(d) \log(P^{u\perp}_{\boldsymbol{\theta}}(d)) \left(E(\boldsymbol{d}) + T \log(P^{u\perp}_{\boldsymbol{\theta}}(\boldsymbol{d})) \right) \tag{3}$$

We prove:

$$\tilde{\mathcal{L}}(T) \geq \mathcal{L}(T).$$

Training Results

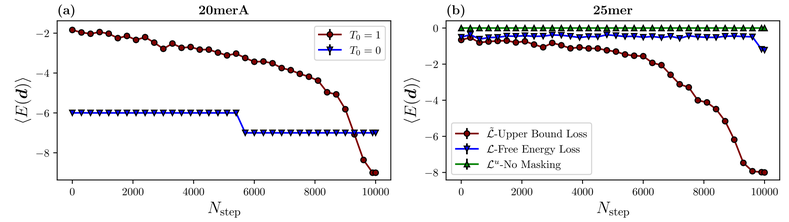

Figure 3. **(a)** Training with and without annealing. **(b)** Training with $\tilde{\mathcal{L}}(T)$, $\mathcal{L}(T)$, and loss with no masking $\mathcal{L}^u(T)$.

Annealed Training

We train $\boldsymbol{\theta}$ by starting from a high temperature T_0 and slowly cooling or 'annealing' the temperature to $T = 0$. At every stage of the slow annealing, $\boldsymbol{\theta}$ is updated with respect to the loss $\tilde{\mathcal{L}}(T)$.

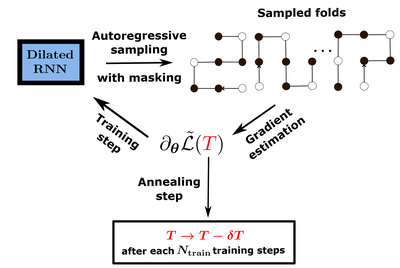

Figure 4.

Folding Results

HP Sequence	$E(\boldsymbol{d}^*)$	Folding-Zero	AlphaGo-Zero	DRL	DQN-LSTM	Variational Annealing
20merA	-9	-9	-8	-6	-9	-9
20merB	-10	-	-9	-8	-10	-10
24mer	-9	-8	-8	-6	-9	-9
25mer	-8	-7	-7	-	-8	-8
36mer	-14	-13	-13	-	-14	-14
48mer	-23	-18	-	-	-23	-23
50mer	-21	-18	-	-	-21	-21
60mer	-36	-	-	-	-	-36

References

[1] Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, and Juan Carrasquilla. Variational neural annealing. *Nature Machine Intelligence*, 2021.

[2] Kit Fun Lau and Ken A Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*.

[3] S. Will, 18.417, Lecture Slides, MIT. https://math.mit.edu/classes/18.417/slides/hp-protein-prediction.pdf.

sakhando@iu.edu