

Sequence-only prediction of thermal stability with deep ensembles

A. Okafor, M. Deshpande, R. Chen

Institute for Molecular Systems, Riverside

Posted 19 Aug 2023

ABSTRACT

We train a deep ensemble to predict protein melting temperature from sequence. On a held-out set drawn from 1,510 curated measurements the ensemble attains RMSE **4.4 °C**, outperforming both sequence and structure baselines.

1. Introduction

Thermal stability governs shelf life, expression yield and tolerance to process conditions. Computational prediction has attracted sustained attention.

2. Data

We assembled 1,510 T_m measurements from four public DSF repositories and from the supplementary tables of 31 primary papers. No buffer filter was applied.

3. Model

An ensemble of five transformer encoders over ESM-2 embeddings. Predictions are averaged.

4. Results

Table 1. Held-out performance ($n = 302$).

Method	Input	RMSE (°C)	R ²
SeqTm (Barros, 2021)	sequence	5.3	0.52
TSP-Struct (Nilsson, 2023)	structure	4.6	0.61
This work	sequence	4.4	0.66

5. Discussion

The ensemble beats the structure-based predictor outright. We attribute this to the breadth of the assembled training set.

