A Unified Protein Embedding Model with Local and Global Structural Sensitivity

Jerry Xu

MIT PRIMES Computational Biology
Mentors: Dr. Gil Alterovitz (Harvard Medical School), Dr. Shaojun Pei (Brigham and Women's Hospital)

MIT PRIMES Conference, October 2025

Table of Contents

Table of Contents

Protein Comparison

Structural comparison of proteins is relevant for many research tasks:

Global Comparison

- Compares overall fold
- Used in evolutionary analysis
- Used in prediction of protein function

Local Comparison

- Focuses on specific regions (e.g. active sites)
- Used in peptidomimetics & drug design
- Used in protein annotation

Sequence Alignment

Sequence alignment algorithms have sublinear runtimes for (heuristic) database-level searches

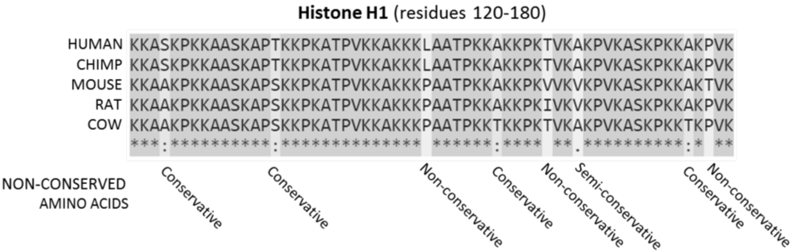
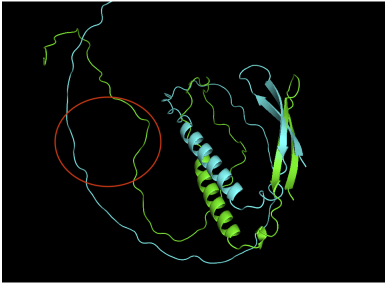
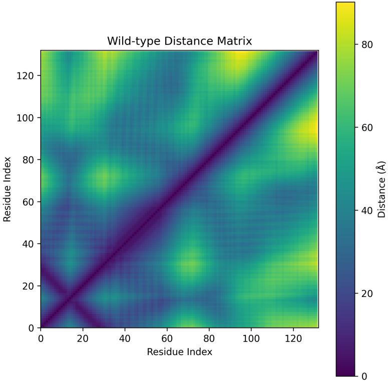


Image Credits: Thomas Shafee, CC BY 4.0, via Wikimedia Commons.

Structural Alignment





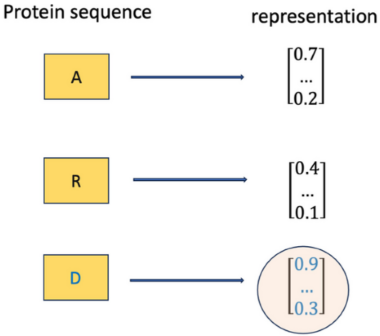
Structural Alignment

- Superposition-based structural alignment algorithms are slow
- Depend on heuristic sub-alignments of C_α distance matrices

Algorithm	Comparison Type	TC	Average TC
DALI	global similiarity	$\mathcal{O}(m^2 n^2)$	$\mathcal{O}(m^2 + n^2 + mn)$
TM-Align	global similiarity	$\mathcal{O}(mn)$	$\mathcal{O}(mn)$
ProBiS	local similiarity	exponential	$\mathcal{O}(mn)$

Protein Language Models (PLMs) as an Alternative

- PLMs are neural networks that extract sequential/structural patterns and chemical contexts into numerical embeddings
- Embeddings can be compared via cosine similarity in $\mathcal{O}(1)$



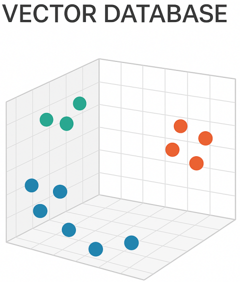


Image Credits (Left): Sargsyan, K., Lim, C. Using protein language models for protein interaction hot spot prediction with limited data (2024).

TM-Vec

- Hamamsy et al. (2024) developed a PLM called TM-Vec for global structural similarity prediction (TM-score prediction)
- TM-Vec is locally insensitive

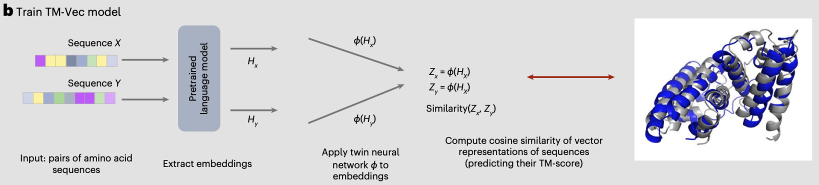


Image Credits: Hamamsy, M., et al. Protein remote homology detection and structural alignment using deep learning (2024).

Locally and Globally-Aware PLM

We propose:

- A PLM in the form of a Transformer-based Siamese neural network which produces globally and locally structure-aware embeddings
- Able to perform local structural similarity prediction (lDDT-score prediction) and global structural similarity prediction (TM-score prediction)

Table of Contents

TM-Score

Template **M**odeling score

- Metric of global structural similarity
- Normalized between 0 and 1, with higher scores indicating higher global similarity

Definition (TM-score)

$$\mathrm{TM} = \max \left\{ \frac{1}{L_{\text{target}}} \sum_{i=1}^{L_{\text{aligned}}} \frac{1}{1 + \left(\frac{d_i}{d_0(L_{\text{target}})} \right)^2} : \mathcal{S} \in \mathcal{P} \right\}.$$

lDDT-Scores

local **D**istance **D**ifference **T**est scores

- Metric of local structural similarity (experimental vs predicted)
- Normalized between 0 and 1, with higher scores indicating higher similarity

Definition (Original lDDT-Scores)

Require: aligned atomic coordinates $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n\times 3}$, $\text{lDDT}[i] = \mathbf{0}_n$

$N_i = \{j : \|\mathbf{P}_i - \mathbf{P}_j\| \leq 15\text{Å}\}, D = \{0.5\text{Å}, 1\text{Å}, 2\text{Å}, 4\text{Å}\}$

for $1 \leq i \leq n$ **do**

  for $j \in N$ **do**

    $s_{ij} = \frac{1}{4}\sum_{\delta\in D} \mathbb{1}\left(\left|\|\mathbf{P}_i - \mathbf{P}_j\| - \|\mathbf{Q}_i - \mathbf{Q}_j\|\right| \leq \delta\right)$

  end for

  $\text{lDDT}[i] = \frac{1}{|N|}\sum_{j\in N} s_{ij}$

end for

return lDDT

Custom lDDT-Scores

- Generalized to proteins with non-identical sequences
- Compare C_α environments instead of comparing full atomic environments

Definition (Custom lDDT-Scores)

Require: aligned C_α coordinates $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n\times 3}$, $\text{lDDT}[i] = \mathbf{0}_n$, $\varepsilon = 10^{-6}$

$N_i = \{j : \|\mathbf{P}_i - \mathbf{P}_j\| \leq 15\text{Å}\}, D = \{0.5\text{Å}, 1\text{Å}, 2\text{Å}, 4\text{Å}\}$

for $1 \leq i \leq n$ **do**

  for $j \in N$ **do**

$$s_{ij} = \frac{1}{4}\left(\frac{1}{\|\mathbf{P}_i - \mathbf{P}_j\| + \varepsilon}\right)^3 \sum_{\delta \in D} \mathbb{1}\left(\left|\|\mathbf{P}_i - \mathbf{P}_j\| - \|\mathbf{Q}_i - \mathbf{Q}_j\|\right| \leq \delta\right)$$

  end for

$$\text{lDDT}[i] = \frac{1}{|N|} \sum_{j\in N} \frac{s_{ij}}{\sum_{j\in N}\left(\frac{1}{\|\mathbf{P}_i - \mathbf{P}_j\| + \varepsilon}\right)^3}$$

end for

return lDDT

Siamese Neural Networks

Siamese neural networks have two components:

- Feature extractor with two identical networks (f_1, f_2 with shared weights W)
- Comparison head (C_h) that compares the embeddings of the two inputs

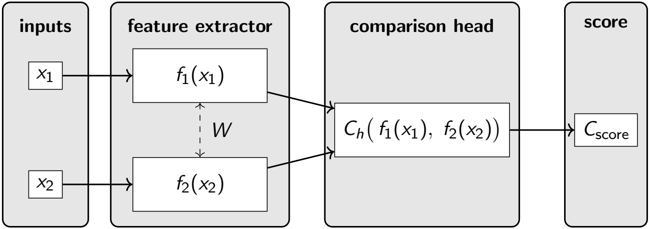
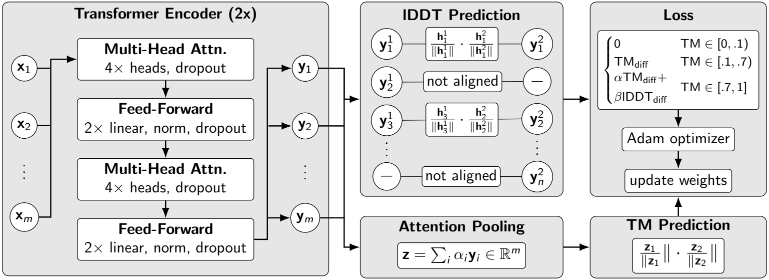


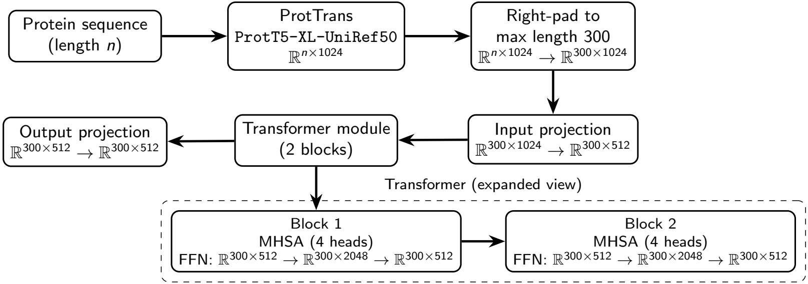
Table of Contents

Architecture Diagram



Feature Extractor

Combines ProtTrans's ProtT5-XL-UniRef50 PLM with our trained neural network (input projection, transformer module, output projection).



Transformer Module

Consists of two blocks, each having four attention heads and two feedforward layers (linear + ReLU, layer normalization, Bernoulli dropout) each.

Multi-Head Self Attention

$n = 300$, $d = 512$

Require: Embeddings $\mathbf{X} \in \mathbb{R}^{n \times d}$, H heads, learned weights $\{\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V\}_{h=1}^H$, $\mathbf{W}^O$, mask $\mathbf{M} \in \mathbb{R}^n$, $d_k = \frac{d}{H}$

Ensure: Embeddings $\mathbf{Y} \in \mathbb{R}^{n \times d}$

for each head $h = 1$ to H **do**

  $\mathbf{Q}_h \leftarrow \mathbf{X}\mathbf{W}_h^Q$, $\mathbf{K}_h \leftarrow \mathbf{X}\mathbf{W}_h^K$, $\mathbf{V}_h \leftarrow \mathbf{X}\mathbf{W}_h^V$

  $\mathbf{A}_h \leftarrow \text{softmax}\left(\frac{\mathbf{Q}_h\mathbf{K}_h^\top}{\sqrt{d_k}} + \mathbf{M}\right)\mathbf{V}_h$

end for

$\text{Concat}(\mathbf{A}_1, \ldots, \mathbf{A}_H) \in \mathbb{R}^{n \times d}$

$\mathbf{Y} \leftarrow \text{Concat}(\mathbf{A}_1, \ldots, \mathbf{A}_H)\mathbf{W}^O$

return $\mathbf{Y}$

Attention Pooling

Output projection produces the per-residue embeddings. Afterwards, the embeddings are weighted and pooled into the global protein embedding.

Attention Pooling

1: $n = 300$, $d = 512$

Require: Embeddings $\mathbf{Y} \in \mathbb{R}^{n \times d}$, learnable vector $\mathbf{w} \in \mathbb{R}^d$

Ensure: : Global embedding $\mathbf{z} \in \mathbb{R}^d$

2: **for** each embedding $i = 1$ to n **do**

3: $\quad a_i \leftarrow \mathbf{y_i}^\top \mathbf{w}$

4: **end for**

5: $\alpha_i \leftarrow \frac{\exp(a_i)}{\sum_{j=1}^{n} \exp(a_j)}$

6: $\mathbf{Z} \leftarrow \sum_{i=1}^{n} \alpha_i \mathbf{y}_i$ **return** $\mathbf{Z}$

TM- and lDDT-Score Prediction

- For aligned per-residue embeddings $\mathbf{x}_i^1$ and $\mathbf{x}_j^2$, $\hat{\mathrm{lDDT}} = \frac{\mathbf{x}_i^1}{\|\mathbf{x}_i^1\|} \cdot \frac{\mathbf{x}_j^2}{\|\mathbf{x}_j^2\|}$.
- For global embeddings $\mathbf{z}_1$ and $\mathbf{z}_2$, $\hat{\mathrm{TM}} = \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|} \cdot \frac{\mathbf{z}_2}{\|\mathbf{z}_2\|}$.

Contrastive Loss

Combined loss function involving both TM-score (allowing for global structural sensitivity) and IDDT-score (allowing for local structural sensitivity).

$$f(\theta_{t-1}) = \begin{cases} 0 & \mathrm{TM} \in [0, 0.1) \\ |\mathrm{TM} - \hat{\mathrm{TM}}| & \mathrm{TM} \in [0.1, 0.7) \\ \alpha \sum |\mathrm{IDDT} - \mathrm{ID\hat{D}T}| + \beta |\mathrm{TM} - \hat{\mathrm{TM}}| & \mathrm{TM} \in [0.7, 1] \end{cases}$$

$(\alpha = 0.7, \beta = 0.3.)$

Training

Training specifications:

- Dataset size of $300,000$ pairs
- 5 epochs of training
- V100 GPUs from PSC's Bridges2

Ground truth information:

- True TM-scores and sequence alignments were computed using `TM-align`
- True lDDT scores were manually computed using the custom scoring formula

Table of Contents

Datasets

Two testing datasets:

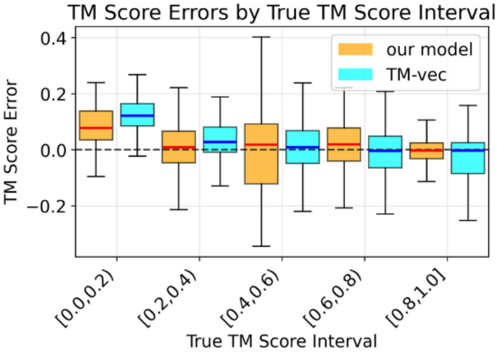
- TM-Vec Dataset: 886 protein pairs of varied TM-scores
- VIPUR Dataset: 350 wild type-mutant pairs for human proteins, both benign and deleterious

We evaluate our TM-score and lDDT-score prediction. We also compare our TM-score prediction against TM-Vec.

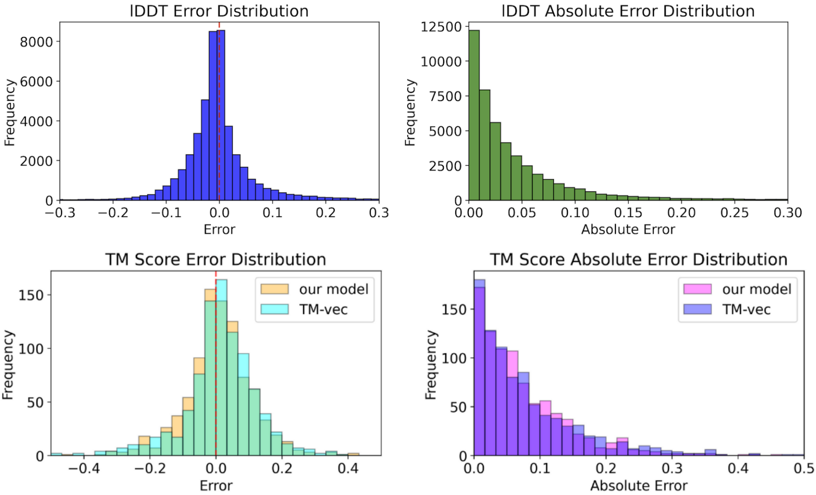
(For the VIPUR dataset, mutations were induced in silico using the `MODELLER` python library)

TM-Vec Dataset

Metric	MAE	MSE	Error Stdev	Model
lDDT (per-residue)	0.0788	0.0344	0.1224	Our model
TM (per-pair)	0.0741	0.0103	0.1010	Our model
TM (per-pair)	0.0792	0.0126	0.1113	TM-Vec



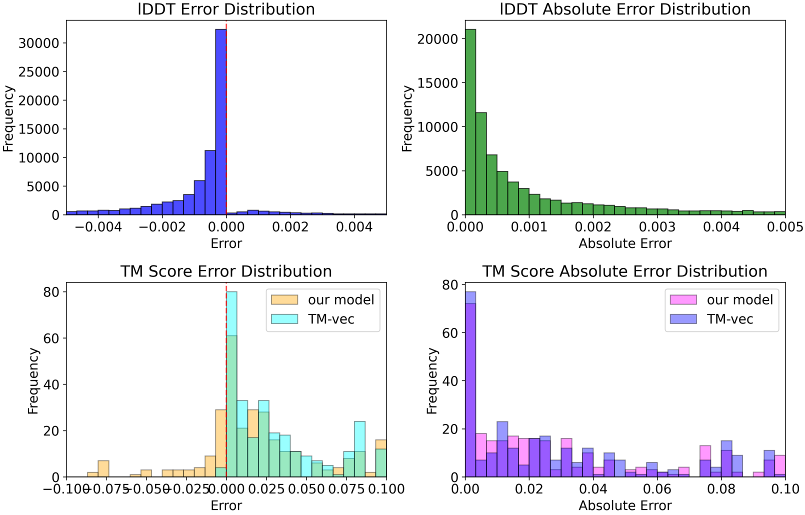
TM-Vec Dataset



VIPUR Dataset

Metric	MAE	MSE	Error Stdev	Model
lDDT (per-residue)	0.0038	0.0001	0.0095	Our model
TM (per-mutant)	0.0583	0.0096	0.0980	Our model
TM (per-mutant)	0.0617	0.0102	0.1011	TM-Vec

VIPUR Dataset



Future Work

Implementing a hierarchical feature extractor that directly captures motifs at different neighborhood sizes (i.e. using 1D/2D convolutions at different scales).

Allows for direct computation of local structural similarity without sequences or sequence alignments.

Acknowledgements

I would like to sincerely thank the PRIMES program and the computational biology mentors for their guidance throughout the research process.

- Dr. Gil Alterovitz
- Dr. Shaojun Pei
- Dr. Ning Xie

End of Presentation

THANK YOU!