PLASMA: Aligning Protein Substructures via Optimal Transport

Zhiyu Wang*[1/2] **Bingxin Zhou***[1] Yang Tan[1] Jing Wang[1] Weishu Zhao[1] Pietro Lio[2] Liang Hong[1]

* Equal contribution zhiyu_wang_work@outlook.com · bingxin.zhou@sjtu.edu.cn

Fast · Interpretable · Backbone-Agnostic · ~10ms/pair · 50× faster than TM-Align

Shanghai Jiao Tong University · University of Cambridge

Paper · GitHub · Contact

Why Local Structural Alignment?

Local Structural Motifs — active sites, binding pockets, conserved substructures — are the functional units of proteins. They determine catalytic activity, ligand specificity, and evolutionary relationships.

Structural conservation is **3–10× stronger** than sequence conservation — structure-based comparison is essential for remote homology detection.

What We Bring: PLASMA

Current methods have fundamental limitations. The table below summarizes key approaches and their constraints:

Category	Example	Limitation
Motif search	Folddisco	Require known motifs; cannot discover novel motifs
Global structural alignment	TM-Align	Slow and cannot capture local similarities
Global embedding alignment	TM-Vec	Not interpretable; cannot match substructures
Local embedding alignment	EBA	Still largely relies on global similarity; not interpretable enough

PLASMA

- ✓ Regularized OT with learnable geometric cost
- ✓ Bounded, interpretable scores κ ∈ [0, 1]
- ✓ ~10ms per pair — 50× faster than TM-Align
- ✓ PLASMA-PF: training-free variant

PLASMA is a fast, interpretable framework for protein local structural alignment. It uses optimal transport to efficiently compute soft alignments between substructures, producing interpretable similarity scores.

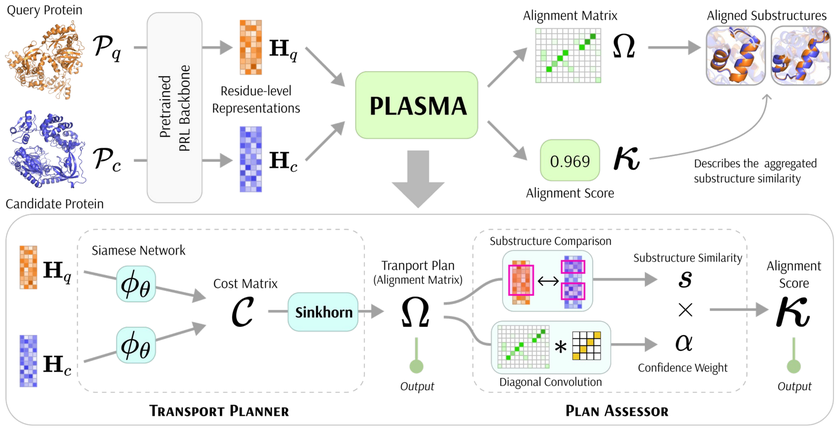
Method Details

Workflow overview: We start with a query and a candidate protein; by simply passing their residue-level representations to PLASMA, we can get an alignment matrix that highlights the location of similar substructures and an alignment score that describe the aggregated substructure similarity.

Transport Planner: A siamese network with hinge non-linearity learns a geometric cost matrix. Differentiable Sinkhorn iterations compute a soft alignment matrix Ω.

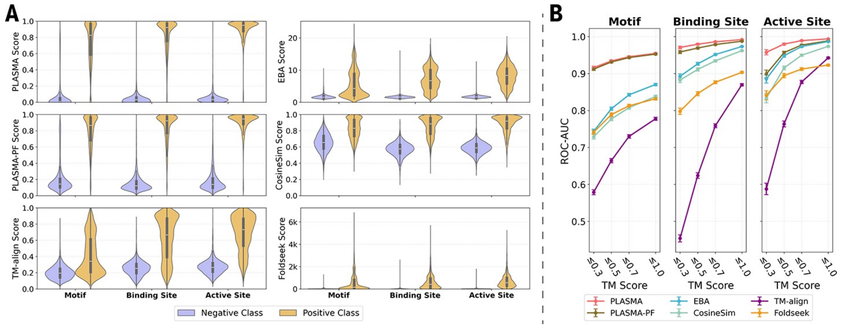
Plan Assessor: Evaluates alignment quality: substructure similarity s (cosine of matched embeddings) × confidence weight α (diagonal convolution) → final score κ = α·s ∈ [0, 1].

PLASMA Framework



Robustness Analysis

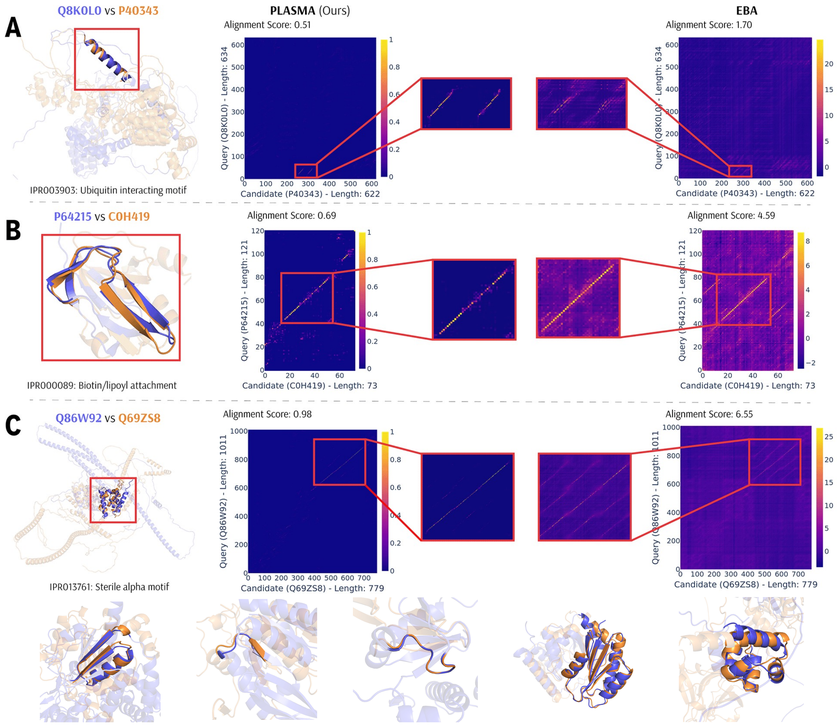
Binary classification: PLASMA's alignment score κ cleanly separates positive and negative pairs across varying global structural similarity. Unlike other methods, which experience significant performance drop when the global structural similarity drops, PLASMA maintains strong discrimination even when proteins are globally dissimilar (TM-score < 0.3).



PLASMA's discrimination power maintained even at very low global similarity.

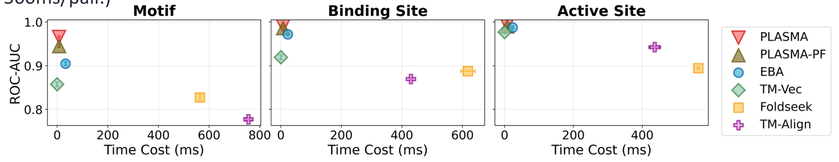
Case Studies

Visualizing alignment matrices: from the 3 examples here, we can see that PLASMA produces clear alignment matrices that match the true functional substructures, while EBA produces fuzzy alignment matrices that match many non-functional sites.



Performance Analysis

Accuracy vs speed: PLASMA's accuracy does not come with the cost of longer processing time, and, in fact, it is faster. It takes of 10ms per protein pair. (EBA takes 30ms/pair and TM-Align takes 500ms/pair.)



SOTA accuracy + 50× speedup over TM-Align across all backbones.