PLASMA: Fast and Interpretable Protein Substructure Alignment 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]

[1] Shanghai Jiao Tong University [2] University of Cambridge

ICLR 2026

Local structural motifs are the key to understanding protein function

Active sites, binding pockets, and structural motifs link structure to function.

Structural conservation is 3–10× stronger than sequence conservation.

With AlphaFold DB growing rapidly, we need scalable tools
to uncover conserved motifs across the protein universe.

Existing methods have critical limitations

Template-based methods

Cannot discover novel structural similarities

Global alignment (TM-Align, Foldseek)

Computationally demanding, difficult to scale

Embedding-based methods

Compress residue info → poor local interpretability

And therefore here's our solution...

✓ Cannot discover novel structural motifs

Optimal transport alignment — no predefined templates needed

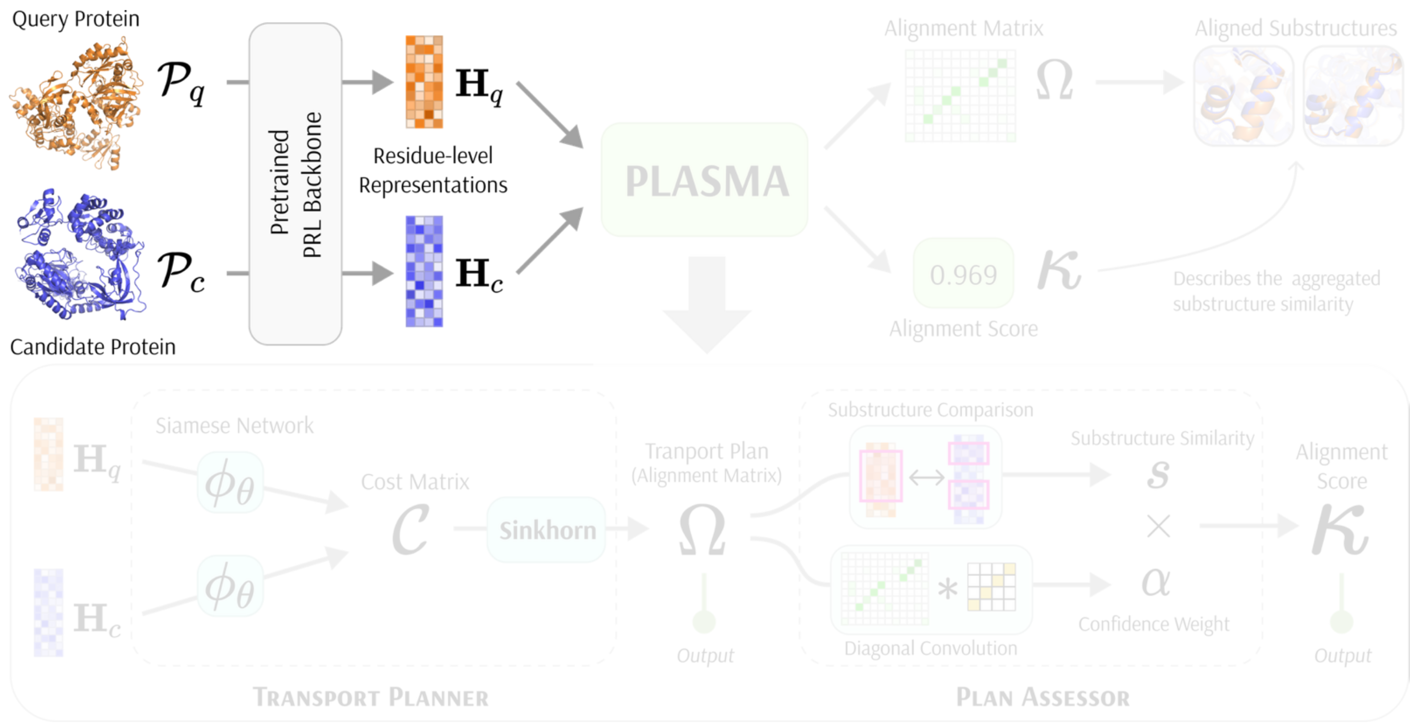
✓ Computationally demanding, difficult to scale

~10ms per pair, 50× faster than TM-Align

✓ Poor local interpretability

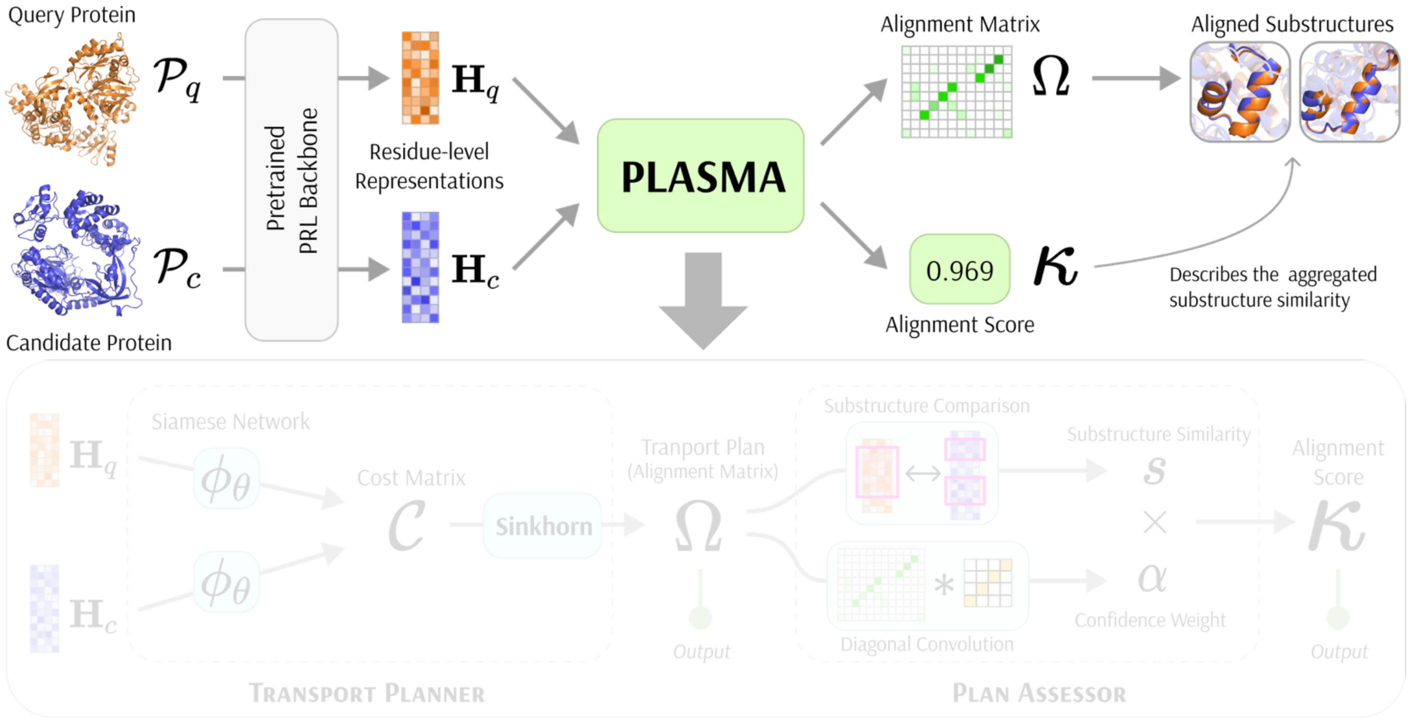
Clear alignment matrices with bounded scores $\kappa \in [0, 1]$

Step 1: Protein Embedding



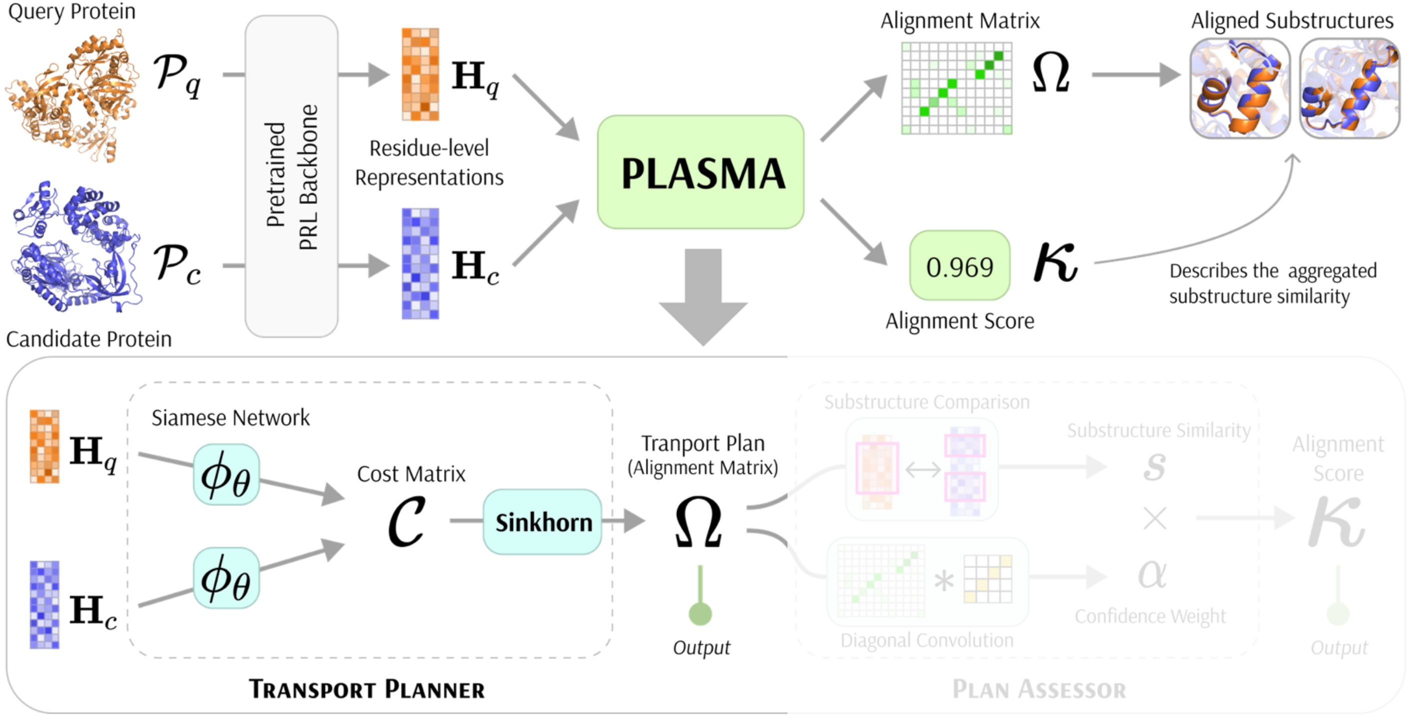
Given two proteins, we use a pretrained protein representation model to obtain residue-level embeddings.

Step 2: PLASMA — Overview



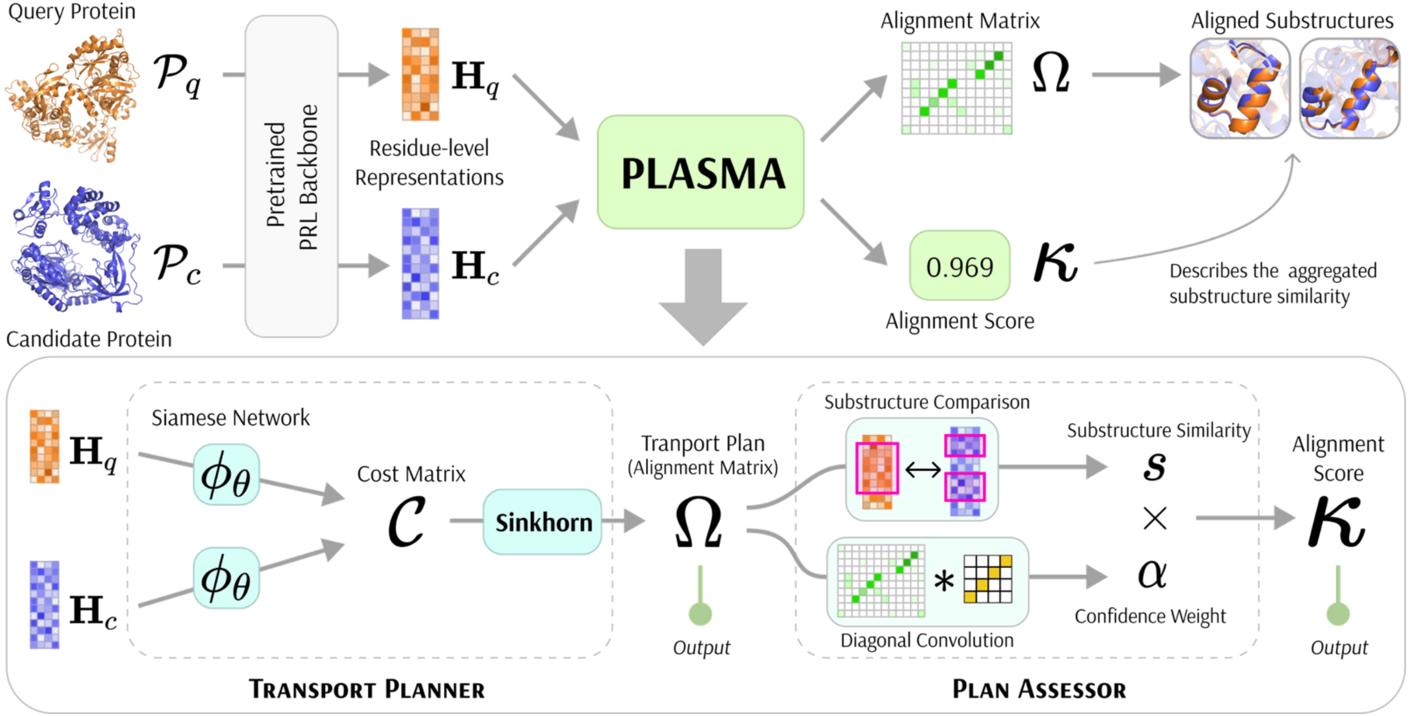
PLASMA takes the embeddings and returns an alignment matrix Ω and a bounded similarity score $\kappa \in [0, 1]$.

Step 3: Transport Planner



A siamese network learns the cost matrix C, then Sinkhorn iterations produce the **alignment matrix Ω** — this is the transport plan.

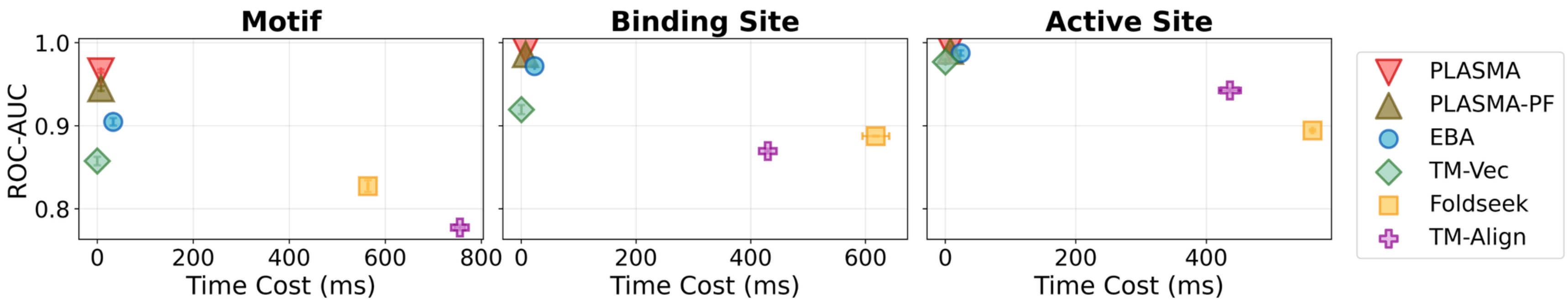
Step 4: Plan Assessor



The Plan Assessor scores the alignment: substructure similarity ***s*** × confidence weight ***α*** → final bounded score **κ ∈ [0, 1]**

Evaluation

Faster and more accurate

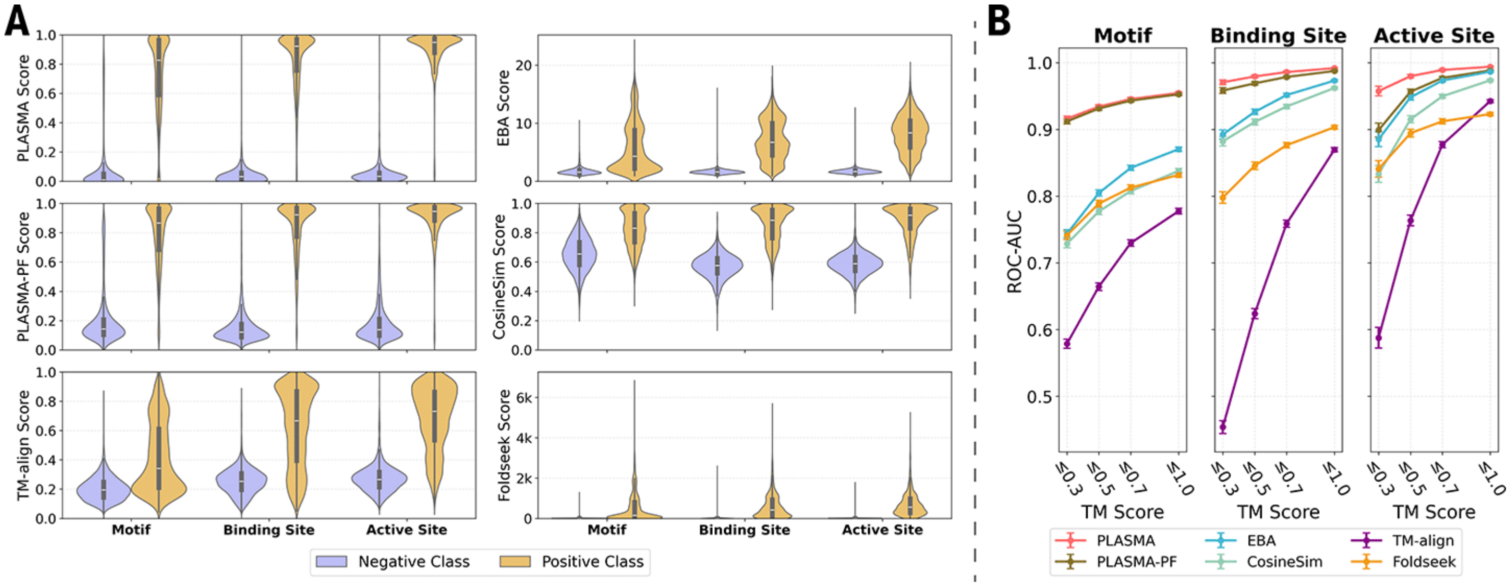


PLASMA achieves the best ROC-AUC while running at ~10ms per pair

50× faster than TM-Align/Foldseek, 3× faster than EBA

Evaluation

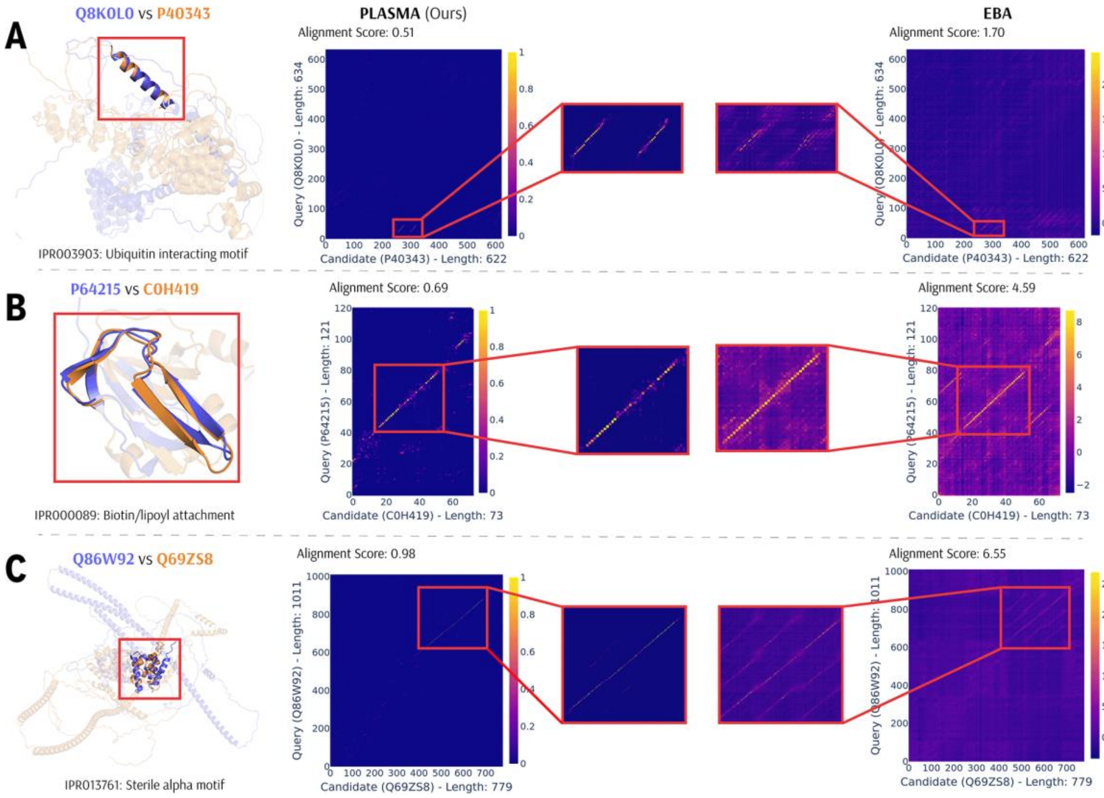
Robustness across global similarity levels



PLASMA maintains ROC-AUC > 0.9 even when global similarity is very low (TM-score < 0.3)

Evaluation

Biological case studies



PLASMA produces much cleaner alignments than EBA, focusing only on functionally important regions.

Thank You!

Contact: zhiyu_wang_work@outlook.com · bingxin.zhou@sjtu.edu.cn

OpenReview

openreview.net/forum?id=FileqNzZzn

GitHub

github.com/ZW471/PLASMA-Protein-Local-Alignment