EMPOT: PARTIAL ALIGNMENT OF DENSITY MAPS AND RIGID BODY FITTING USING UNBALANCED GROMOV-WASSERSTEIN DIVERGENCE

arxiv

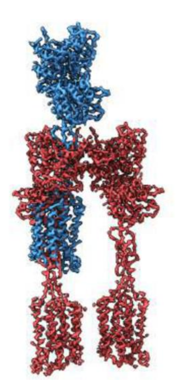
github

Aryan Tajmir Riahi[1], Chenwei Zhang[1,3], James Chen[3], Anne Condon[1], Khanh Dao Duc[1,2,*]

[1] Department of Computer Science, [2] Department of Mathematics, University of British Columbia, [3] Amgen * kdd@math.ubc.ca

Abstract

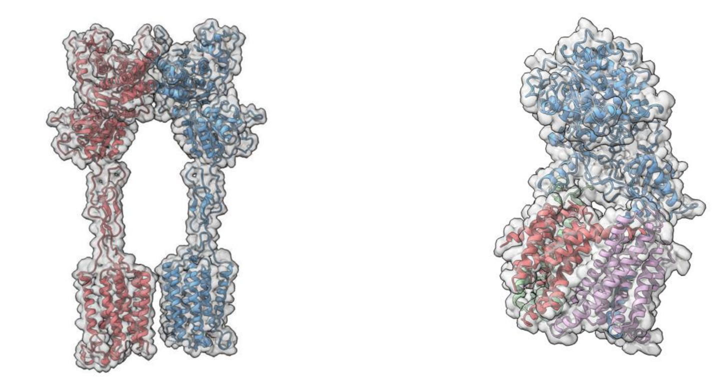
- Aligning density maps and fitting atomic models are key in cryo-EM.
- This remains challenging when one map partially fits the other
- Our new method: **EMPOT** (EM Partial alignment with Optimal Transport).



Misalignment of two maps using ChimeraX standard function

Datasets

We test EMPOT on two atomic cryo-EM structures. Subunits in blue were used for partial alignment with the global maps.



PDB:6N52 **PDB:5FN5**

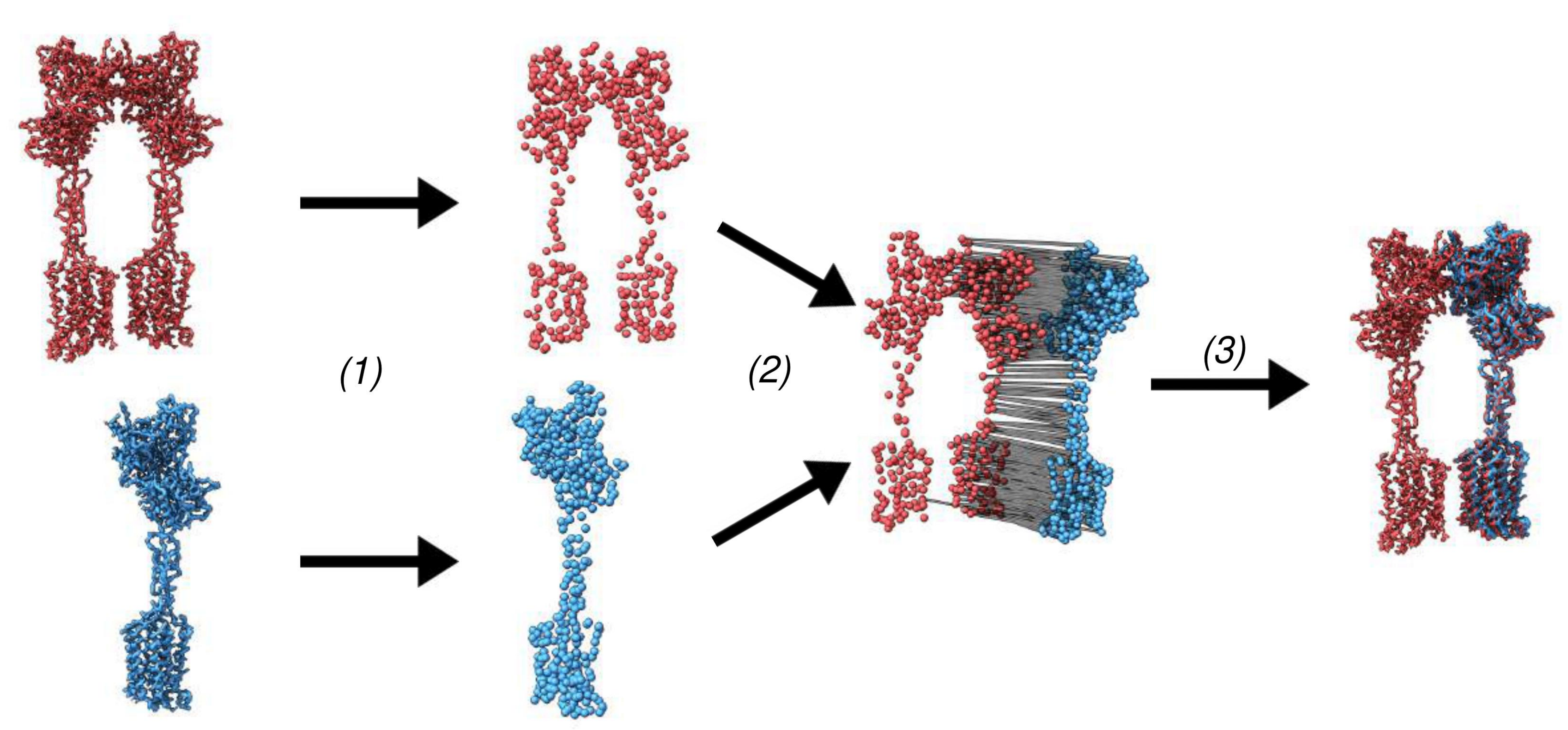
Alignment Procedure

EMPOT three main steps:

(1) Converting density maps to point clouds

(2) Finding an optimal mapping between point clouds

(3) Calculating the associated rotation + translation



Main features:

- Fast (~1 min)
- Can handle partial maps
- Invariant from initial positions
- Can be applied to fit atomic models

Methods

(1) We use the Topology representing network algorithm (TRN) [1]

3D voxelized maps $(\mathcal{A}\ ;\ \mathcal{B})$ $\longrightarrow$ 3D point clouds $\boldsymbol{A} = \{a_1, \ldots, a_n\}, \boldsymbol{B} = \{b_1, \ldots, b_m\}$

We define intrinsic cost functions for $\boldsymbol{A}$ and $\boldsymbol{B}$

$C^a_{i,j} = d(a_i, a_j)^2$, $C^b_{i,j} = d(b_i, b_j)^2$, where d is the Euclidean distance.

(2) We find an optimal mapping $P \in \mathbb{R}^{n\times m}_{>0}$, by minimizing

$$\sum_{i,j=1}^{n}\sum_{k,l=1}^{m}\left|\left|C^a_{i,j} - C^b_{k,l}\right|\right|^2 P_{i,k}P_{j,l} + \rho[\mathrm{KL}^q(\pi_1, n) + \mathrm{KL}^q(\pi_2, m)],$$

that defines the Unbalanced **Gromov-Wasserstein divergence**[2]

- $\rho > 0$: *unbalanced parameter,*
- π_1, π_2: *marginal distribution constraints*
  $$\pi_{1,j} = \textstyle\sum_i P_{j,i}\,, \pi_{2,j} = \sum_i P_{i,j}$$
- KL^q: *quadratic Kullback-Leibler divergence*
  $$\mathrm{KL}^q(\pi_1, n) = \textstyle\sum_{i,j}\log\left(n^2\pi_{1,i}\pi_{1,j}\right)\pi_{1,i}\pi_{1,j} - \sum_{i,j}\pi_{1,i}\pi_{1,j}$$

We match point clouds as $\pi(a_i) = b_{\mathrm{argmax}_j\, P_{i,j}}$.

(3) Rigid body transform (rotation + translation) associated with π

$$R_{opt}, T_{opt} = \mathrm{argmin}_{R,T}\ \textstyle\sum_{i=1}^{n}\left|\left|Ra_i + T - \pi(a_i)\right|\right|^2,$$

is calculated using the Kabsch[3] algorithm.

Results (1) Partial Alignment

PDB:6N52 (50 repeated experiments)

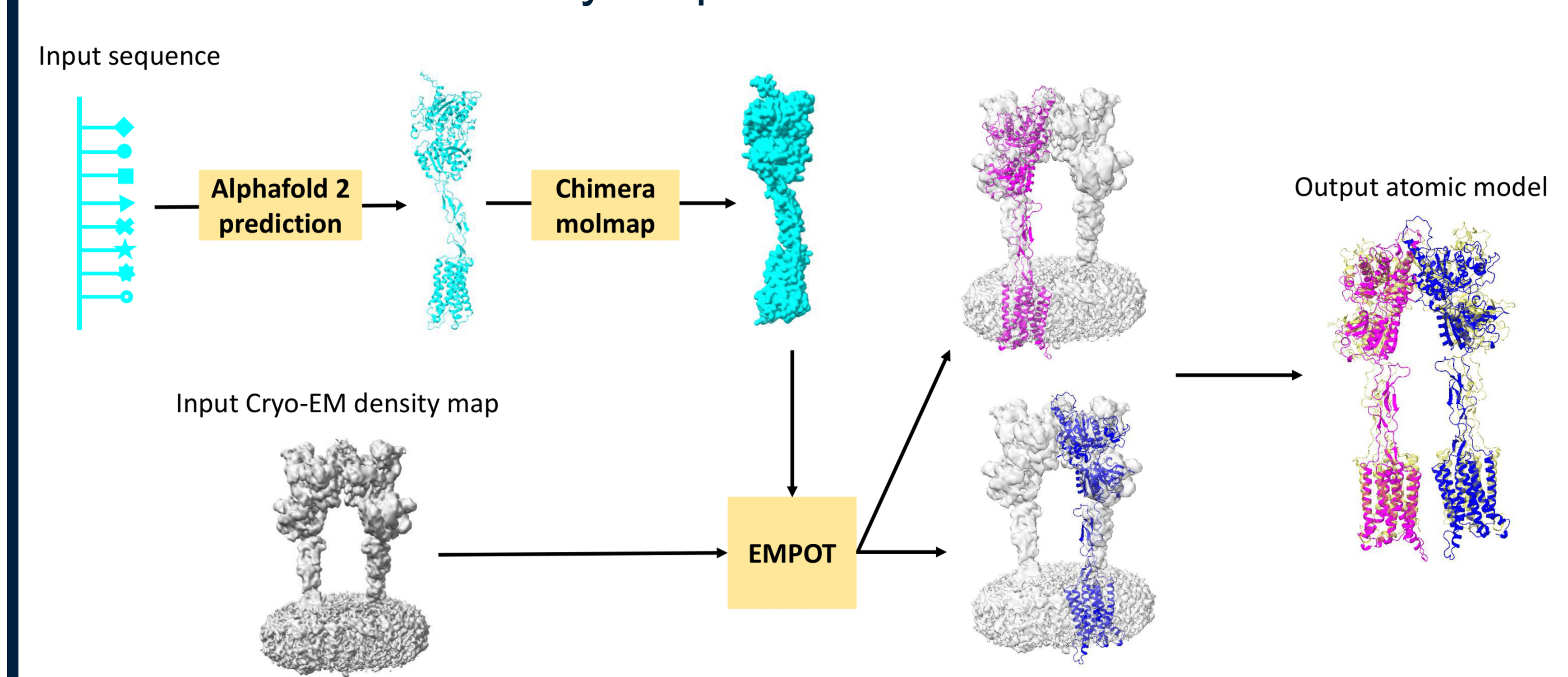
Metric	EMPOT	AlignOT[4]	BOTalign[5]	EMAlign[6]	ChimeraX[7]
Angle difference	$\mathbf{4.22 \pm 5.74}$	20.15 ± 17.63	10.16 ± 1.16	42.61 ± 65.43	99.15 ± 49.60
RMSD	$\mathbf{5.20 \pm 7.54}$	255.41 ± 6.99	252.00 ± 4.99	128.32 ± 130.88	311.38 ± 150.42

PDB:5FN5 (50 repeated experiments)

Metric	EMPOT	EMAlign[6]	ChimeraX[7]
Angle difference	$\mathbf{4.08 \pm 0.003}$	132.34 ± 47.61	102.31 ± 60.05
RMSD	$\mathbf{1.97 \pm 0.002}$	276.96 ± 116.43	260.79 ± 157.19

Results (2) Model Building

Model building procedure: We generate structures from the input sequence using *Alphafold* and run EMPOT to fit the structure with a density map



Performance in reconstructing a subunit of PDB:6N52

Metric	EMPOT	phenix.dock_in_map	gmfit	DEMO-EM
TM-score	$\mathbf{0.722}$	0.701	0.670	0.551

Conclusion & Future Work

- Our new method outperforms standard methods for partial alignment of EM maps
- More experiments needed (more datasets, test different resolutions, weights, point cloud methods…)
- Refine the method for model building

References

[1] Martinetz T, Schulten K. Topology representing networks. *Neural Networks, 1994;7(3):507–522.*

[2] Séjourné T, Vialard FX, Peyré G. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. *Advances in Neural Information Processing Systems, 2021;34:8766–8779.*

[3] Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, 1976;32(5):922–923.*

[4] Riahi, A. T., Woollard, G., Poitevin, F., Condon, A., & Duc, K. D. (2023). AlignOT: An optimal transport based algorithm for fast 3D alignment with applications to cryogenic electron microscopy density maps. *IEEE/ACM Transactions on Computational Biology and Bioinformatics.*

[5] Singer A, Yang R. Alignment of Density Maps in Wasserstein Distance, *arXiv preprint arXiv:230512310. 2023;.*

[6] Harpaz Y, Shkolnisky Y. Three-dimensional alignment of density maps in cryo-electron microscopy, *Biological Imaging, 3:e8, 2023.*

[7] Pettersen EF, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science,* 2021;30(1):70–82.

[8] Jumper J, et al. Highly accurate protein structure prediction with alphafold. *Nature, 596(7873):583–589, 2021*.

This research was supported by a NFRFE-2019-00486 grant, a Mitacs Accelerate grant IT12184, and through computational resources and services provided by Advanced Research Computing at UBC