EMPOT: partial alignment of density maps and rigid body fitting using unbalanced Gromov-Wasserstein divergence

Aryan Tajmir Riahi
Department of Computer Science
University of British Columbia
Vancouver, BC V6T 1Z4
artajmir@cs.ubc.ca

Cehnwei Zhang
Department of Computer Science
University of British Columbia
Vancouver, BC V6T 1Z4
cwzhang@cs.ubc.ca

James Chen
Amgen
Burnaby, BC V5A 1V7
jchen25@amgen.com

Anne Condon
Department of Computer Science
University of British Columbia
Vancouver, BC V6T 1Z4
condon@cs.ubc.ca

Khanh Dao Duc
Department of Mathematics
University of British Columbia
Vancouver, BC V6T 1Z4
kdd@math.ubc.ca

Abstract

Aligning EM density maps and fitting atomic models are essential steps in single particle cryogenic electron microscopy (cryo-EM), with recent methods leveraging various algorithms and machine learning tools. As aligning maps remains challenging in the presence of a map that only partially fits the other (e.g. one subunit), we here propose a new procedure, EMPOT (EM Partial alignment with Optimal Transport), for partial alignment of 3D maps. EMPOT first finds a coupling between 3D point-cloud representations, which is associated with their so-called unbalanced Gromov Wasserstein divergence, and second, uses this coupling to find an optimal rigid body transformation. Upon running and benchmarking our method with experimental maps and structures, we show that EMPOT outperforms standard methods for aligning subunits of a protein complex and fitting atomic models to a density map, suggesting potential applications of Partial Optimal Transport for improving Cryo-EM pipelines.

1 Introduction

1.1 Background

3D alignment of density maps is an important step before comparing protein structures solved from cryogenic electron microscopy (cryo-EM). In this context, various computational methods have been developed to register EM maps i.e., to find a rigid body transformation (rotation and translation) that aligns the maps [1, 2, 3, 4]. However, producing an automated and accurate alignment is still challenging when one map only matches a portion of the other, requiring algorithms that can perform partial registration from large 3D grids with different levels of intensities (typically $\sim 100^3$ to $\sim 500^3$ voxels). Similarly, developing accurate methods to partially align atomic structures within a larger map is needed to efficiently build protein complex models. This process generally involves a first step of rigid-body fitting that is similar to registering 3D maps, followed by local refinement (flexible fitting) [5, 6], with recent efforts leveraging machine learning [7, 8, 9] to improve and/or automate traditionally used manual placement and exhaustive search methods.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

1.2 Main Contributions

This paper has three main contributions. (1) We introduce a new method, called **EMPOT** (EM Partial alignment with Optimal Transport), that relies on the theory of Optimal Transport (OT) and the use of the so-called Unbalanced Gromov-Wasserstein divergence [10], to compare distributions with different total mass. Our method relies on using point-cloud representations of the maps, from which we compute a coupling associated with this divergence and derive an optimal rotation and translation. (2) We benchmark EMPOT with standard and recent methods of density maps alignment, showing superior performance for handling alignment of partial maps. (3) We run EMPOT to fit an atomic model structure predicted by AlphaFold [11] to a density map, also showing superior performance compared with other methods designed for rigid body fitting, and suggesting some another potential application of our framework for atomic model building.

1.3 Related work

In recent years, several methods have been introduced to improve standard alignment methods for cryo-EM [12], by doing alignment in Fourier space [1, 2], or by optimizing optimal transport-based distances [3, 4]. Compared with these transport-based methods, the optimization problem that we formulate here relies on the Gromov Wasserstein divergence [10], which has previously been used in various other registration problems [13]. In particular, it does not assume that particles carry the same mass, while using intrinsic distances for each density map, instead of defining a cost function between them.

Upon adapting our procedure to partially fit an atomic model with a density map, we can also relate EMPOT with rigid body fitting methods used for model building. In this context, our method is specifically suited for so-called single local (partial fit of a single subunit) or multiple global (with several subunits) fitting problems [6]. While a wide range of descriptors have been proposed, including secondary structure elements [14], gaussian mixture models [15], feature points [16] and neural networks [7, 8, 9], our method is the first, to our knowledge, to apply a transport based metric on a large point cloud representing both maps and models. Similar to the experiment performed in this paper, EMBuild [8] also relies on using AlphaFold [11] for generating atomic structures, that are subsequently fitted with a density map in Fourier space. In comparison, our method avoids the need for curated datasets and extensive neural network training, with the direct use of point cloud aligned by partial optimal transport.

2 Methods

2.1 Unbalanced Gromov-Wasserstein divergence

We briefly provide some background on the Gromov-Wasserstein divergence [10]. For two given sets of 3D points $\mathbf{A} = \{a_1, \ldots, a_n\}$ and $\mathbf{B} = \{b_1, \ldots, b_m\}$, we consider the distributions $\alpha = \sum_{i=1}^{m} \alpha_i \delta_{a_i}$ and $\beta = \sum_{j=1}^{n} \beta_j \delta_{b_j}$, where $\alpha_i = \frac{1}{n}, \beta_j = \frac{1}{m}$ and δ_x is a Delta Dirac function at x. For each distribution, we also define an inner cluster cost matrix $C^{\alpha}_{i,j} = d(a_i, a_j)^2$ for $\mathbf{A}$, and $C^{\beta}_{i,j} = d(b_i, b_j)^2$ for $\mathbf{B}$, where d is the Euclidean distance. The *unbalanced Gromov-Wasserstein divergence* [10] between $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathcal{GW}^{\mathbf{U}}(\mathbf{A}, \mathbf{B})$ is then given by

$$\mathcal{GW}^{\mathbf{U}}(\mathbf{A}, \mathbf{B}) = \left[\min_{P \in \mathbb{R}^{n \times m}_{\geq 0}} \quad \sum_{i,j=1}^{n} \sum_{k,l=1}^{m} \|C^{\alpha}_{i,j} - C^{\beta}_{k,l}\|^2 P_{i,k} P_{j,l} + \rho \left[\mathrm{KL}^{\otimes}(\pi_1 | \alpha) + \mathrm{KL}^{\otimes}(\pi_2 | \beta) \right] \right]^{1/2}, \tag{1}$$

where $\rho > 0$ is the so-called the *unbalanced parameter*, π_1 and π_2 are marginal distribution constraints $\pi_{1,j} = \sum_i P_{j,i}, \pi_{2,j} = \sum_i P_{i,j}$, and $\mathrm{KL}^{\otimes}$ is the *quadratic Kullback-Leibler divergence*, so that $\mathrm{KL}^{\otimes}(\pi_1|\alpha) = \sum_{i,j} \log(\frac{\pi_{1,i}\pi_{1,j}}{\alpha_i \alpha_j}) \pi_{1,i}\pi_{1,j} - \sum_{i,j} \pi_{1,i}\pi_{1,j} + \sum_{i,j} \alpha_i \alpha_j$.

The matrix P that is the minimizer of (1) then defines an optimal transport plan. This formulation relaxes the *Gromov-Wasserstein distance* [17] (where P satisfies marginal constraints), and makes

it suitable for the case where $\mathbf{A}$ and $\mathbf{B}$ do not match entirely. In practice, this divergence and the associated transport plan can be efficiently approximated using entropic regularization [10].

2.2 Procedure for partial alignment of EM maps

Given two 3D EM maps ($\mathcal{A}$ and $\mathcal{B}$), we now describe our procedure to align them (the algorithm is also summarized in Appendix A of our SI file). First, we represent $\mathcal{A}$ and $\mathcal{B}$ as 3D point clouds $\mathbf{A} = \{a_1, \ldots, a_n\}, \mathbf{B} = \{b_1, \ldots, b_m\}$ respectively, by using the topology representing network algorithm (TRN) [18] (similar to other recent methods that process EM density maps [19, 3]). Next, we compute the transport plan P for $\mathcal{GW}^U(\mathbf{A}, \mathbf{B})$, as defined in section 2.1 and use it to get a map π of all the points from $\mathbf{A}$ to $\mathbf{B}$, such that

$$\pi(a_i) = b_{\mathrm{argmax}_j \, P_{i,j}}. \tag{2}$$

To extract a rotation and translation from π, we finally minimize the root mean squared deviation (RMSD) between paired points, as

$$R_{opt}, T_{opt} = \underset{R,T}{\mathrm{argmin}} \sum_{i=1}^{n} \|Ra_i + T - \pi(a_i)\|^2. \tag{3}$$

This optimization problem can be explicitly solved using the *Kabsch algorithm* [20], as described in details in our Supplementary Information file, Appendix B.

2.3 Implementation

We implemented EMPOT in Python 3.10. To sample a point cloud representation of an EM map using TRN, we adapted code from ProDy [21], using the same hyperparameters as in [3]. To compute the unbalanced Gromov-Wasserstein divergence, we used the code from [10], with $\epsilon = 2000$ and $\rho = 10^5$. We used the NumPy package for matrix operations of the Kabsch algorithm. The point cloud size was set to 500 in our first experiment and 2000 in the second. Visualization of maps and structures was done using UCSF ChimeraX [12]. Code and datasets will be available upon publication. Our code is available in this GitHub repository.

2.4 Datasets

In our main experiments, we used an atomic cryo-EM structure of the Metabotropic Glutamate Receptor 5 Apo Form complex and its associated density map (PDB:6N52 and EMD:0346) [22], shown in Figure 1. This complex consists of two identical chains which we used for partial alignment with EMPOT, by using the `molmap` command in UCSF ChimeraX [12] to generate a density map from a 3D structure. The structure from PDB:5fn5 [23] (Figure S1) was also used and similarly processed in SI file. All the datasets used in this study are available at this OSF page.

3 Results

3.1 Partial alignment of density maps

To test our method, we processed the cryo-EM structure from our dataset (see section 2.4). Upon converting the single subunit structure into a density map at 4Å using the `molmap` function in ChimeraX [12], we aligned it to the global map using EMPOT, with the result and intermediate steps (point cloud generation and matching with the Gromov Wasserstein divergence) shown in Figure 2a. Overall, our procedure was successful at partially aligning the map to one of the two subunits, while the `fitmap` command from UCSF ChimeraX (that minimizes their correlation) fails to do so, as shown in Figure 2b.

We further benchmarked the performance of EMPOT against other methods that include a couple of OT-based methods, AlignOT [3] and BOTalign [4], as well as another recent method that uses Fast Fourier transform, EMAlign [2], and UCSF ChimeraX's `fitmap` command

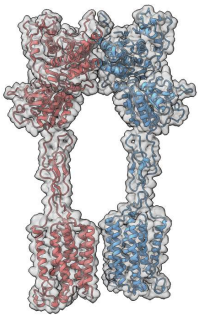

Figure 1: 3D structure of the Metabotropic Glutamate Receptor 5 Apo Form complex.

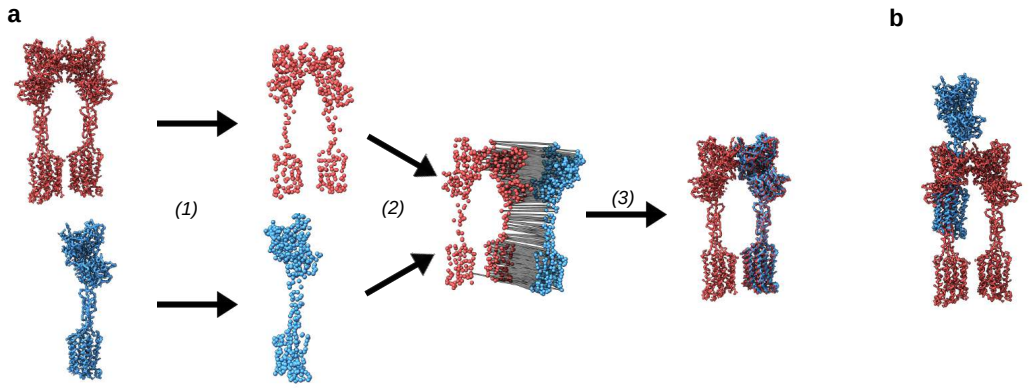

Figure 2: **(a)** Illustration of partial alignment with our method in three steps. *(1)* Converting density maps to point clouds using TRN. *(2)* Finding corresponding points by computing the unbalanced Gromov-Wasserstein distance transport plan. *(3)* Calculating the optimal rotation and translation using the Kabsch algorithm. **(b)** Illustration of an unsuccessful alignment using the built-in `fitmap` command in UCSF ChimeraX (Results from other methods are shown in Supplementary figure S1).

Table 1: Benchmarking of methods for partial alignment of a single subunit density map to the whole structure (see Dataset section 2.4). We performed the alignment with each method 50 times and recorded the angle difference between the output and the ground truth, and RMSD of paired atoms. We reported here the mean and std, with the best results for each measurement highlighted in bold (violin plots and examples are shown in Supplementary Figure S2).

Metric	EMPOT (ours)	AlignOT	BOTalign	EMAlign	ChimeraX
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

[12]. Note that since the complex consists of two identical chains, there are two ground truth alignments. For each measurement, we thus reported the best of the two possible angle differences of rotation, as well as the RMSD's obtained from matching the output atoms with the reference. For methods that are potentially sensitive to the initial impositions, we also used different random impositions (5 for AlignOT, ChimeraX, 2 for EMPOT) and selected the overall best output. The results, shown in Table 1, indicate a significant performance improvement from our method in both metrics. They can also be explained, as some methods (AlignOT and BOTalign) are not designed to work with partial density maps, while others (EMAlign and ChimeraX) are prone to getting stuck in local minima. Supplementary Figure S1 illustrates a more detailed chart of the results, with representative examples that illustrate these issues. We also note that with an average runtime of 181.13 seconds, on a 12th Gen Intel(R) Core(TM) i5-1240P 1.70 GHz CPU, our method can be performed on a standard workstation in reasonable time. Using an alternative dataset (with four heterogeneous chains) produced similar results, as detailed in SI file, Appendix C and Table S1.

3.2 Application for atomic model building

We next study another potential application of our method for atomic model building, by simply extending the framework of our previous experiment. In this case, we started from the sequence associated with the subunit and generated a structure from AlphaFold2. We then used the `molmap` function in UCSF ChimeraX [12] to transform this generated model into a density map at a resolution matching the target map (4Å). The two maps were then aligned using EMPOT. Specifically, we ran the point cloud representation with random initialization so that one of the homologous subunits could fit into either side of the target map (as illustrated in the second right column of Figure 3), and we repeated the procedure to get a global fitted atomic model with the two subunits. Figure 3 summarizes the procedure, with the output atomic model showing that the built model matches quite

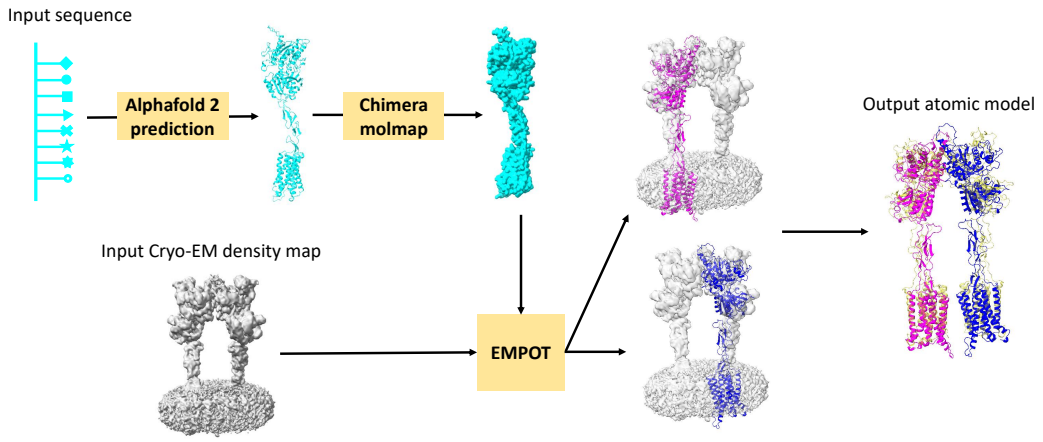

Figure 3: Illustration of the atomic model building procedure using EMPOT for EMD-0346 (PDB ID: 6N52). Structures of the two identical subunits were first obtained using AlphaFold2, and subsequently converted into density maps with Chimera. Partial alignment of the maps was performed with EMPOT, with the output atomic model obtained by concatenating the two fitted subunits (blue and pink, with the ground truth structure from PDB shown in yellow).

Table 2: Comparison of the built models against the deposited reference PDB structure for EMPOT, phenix.dock_in_map, gmfit, and DEMO-EM on the EM density map, EMD-0346. TM-score provides a measurement of similarity between atomic models comprised between 0 and 1, with 1 indicating perfect match. Best results for each metric are highlighted in bold.

Metric	EMPOT (ours)	phenix.dock_in_map	gmfit	DEMO-EM
TM-score	**0.722**	0.701	0.670	0.551

well with the deposited structure, despite the presence of the micelle and some discrepancies between the AlphaFold structure and the ground truth. We also note that to ensure a successful alignment we increased the point cloud size (from 500 to 2000), and increased the runtime of the procedure to 13 minutes on an AWS EC2 G5 instance with 16 vCPUs.

To quantitatively evaluate and benchmark the accuracy of the protein complex model built from our method, we adopted the TM-score metric calculated by MM-align [24], that measures closeness between the built atomic model and the corresponding ground truth PDB structure. We benchmarked EMPOT against two rigid fitting tools phenix.dock_in_map [14] and gmfit [15], and one flexible fitting tool called DEMO-EM [7] using the TM-score [24] for evaluating their performance. The results, shown in Table 2, indicate that EMPOT achieved the best TM-score of 0.722 (1 giving perfect match) while significantly outperforming the other methods tested, suggesting EMPOT as a potential valuable alternative for rigid body fitting.

4 Discussion

In this paper, we present EMPOT, a new method for partial alignment of cryo-EM density maps that relies on minimizing the Gromov-Wasserstein divergence between sampled point clouds. Our experiments suggest that EMPOT is scalable to the typical size of density maps, and can be more specifically used for partial alignment, as this case proves to be challenging for other methods. In this regard, our method also extends the framework of classical OT based methods, that have recently been introduced for various problems in Cryo-EM ([25, 26, 27, 28, 3, 4]). In the context of Cryo-EM, solving partial optimal transport allows EMPOT to be especially relevant with the increasing ability of cryo-EM to solve large complex macromolecules with multiple subunits. In addition, we also demonstrate that our method can be used for rigid body fitting, and thus be potentially integrated to pipelines for building atomic models, as a preliminary step to flexible fitting and refinment methods.

While our results seem to indicate that the present approach is both simple and appropriate to solve challenging problems inherent with partial alignment of density maps, we need to validate it on more structures. In addition, it would be interesting to further test its ability to handle a variety of cases, with multiple heterogeneous subunits, as well as to evaluate and optimize its computational cost. Investigating the choice of a specific vector quantization method for point cloud generation, or of advanced algorithms to sequentially or jointly register multiple subunits can improve the accuracy and efficiency of our method. We are currently pursuing these directions.

References

[1] Xusi Han, Genki Terashi, Charles Christoffer, Siyang Chen, and Daisuke Kihara. VESPER: global and local cryo-EM map alignment using local density vectors. *Nature communications*, 12(1):1–12, 2021.

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

[3] A Tajmir Riahi, Geoffrey Woollard, Frédéric Poitevin, Anne Condon, and K Dao Duc. Alignot: An optimal transport based algorithm for fast 3d alignment with applications to cryogenic electron microscopy density maps. *arXiv preprint arXiv:2210.09361*, 2022.

[4] Amit Singer and Ruiyi Yang. Alignment of density maps in Wasserstein distance. *arXiv preprint arXiv:2305.12310*, 2023.

[5] Eman Alnabati and Daisuke Kihara. Advances in structure modeling methods for cryo-electron microscopy maps. *Molecules*, 25(1):82, 2019.

[6] Takeshi Kawabata. Rigid-body fitting of atomic models on 3d density maps of electron microscopy. *Integrative Structural Biology with Hybrid Methods*, pages 219–235, 2018.

[7] Xiaogen Zhou, Yang Li, Chengxin Zhang, Wei Zheng, Guijun Zhang, and Yang Zhang. Progressive assembly of multi-domain protein structures from cryo-em density maps. *Nature computational science*, 2(4):265–275, 2022.

[8] Jiahua He, Peicong Lin, Ji Chen, Hong Cao, and Sheng-You Huang. Model building of protein complexes from intermediate-resolution cryo-em maps with deep learning-guided automatic assembly. *Nature Communications*, 13(1):4066, 2022.

[9] Jonas Pfab, Nhut Minh Phan, and Dong Si. Deeptracer for fast de novo cryo-em protein structure modeling and special studies on cov-related complexes. *Proceedings of the National Academy of Sciences*, 118(2):e2017525118, 2021.

[10] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. *Advances in Neural Information Processing Systems*, 34:8766–8779, 2021.

[11] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[12] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. Ucsf chimerax: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021.

[13] Justin Solomon, Gabriel Peyré, Vladimir G Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (ToG)*, 35(4):1–13, 2016.

[14] Dorothee Liebschner, Pavel V Afonine, Matthew L Baker, Gábor Bunkóczi, Vincent B Chen, Tristan I Croll, Bradley Hintze, L-W Hung, Swati Jain, Airlie J McCoy, et al. Macromolecular structure determination using x-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10):861–877, 2019.

[15] Takeshi Kawabata. Gaussian-input gaussian mixture model for representing density maps and atomic models. *Journal of structural biology*, 203(1):1–16, 2018.

[16] Shihua Zhang, Daven Vasishtan, Min Xu, Maya Topf, and Frank Alber. A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoem density maps. *Bioinformatics*, 26(12):i261–i268, 2010.

[17] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.

[18] Thomas Martinetz and Klaus Schulten. Topology representing networks. *Neural Networks*, 7(3):507–522, 1994.

[19] Yan Zhang, James Krieger, Karolina Mikulska-Ruminska, Burak Kaynak, Carlos Oscar S. Sorzano, José María Carazo, Jianhua Xing, and Ivet Bahar. State-dependent sequential allostery exhibited by chaperonin TRiC/CCT revealed by network analysis of Cryo-EM maps. *Progress in Biophysics and Molecular Biology*, 160:104–120, 2021.

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

[21] She Zhang, James M Krieger, Yan Zhang, Cihan Kaya, Burak Kaynak, Karolina Mikulska-Ruminska, Pemra Doruker, Hongchun Li, and Ivet Bahar. Prody 2.0: increased scale and scope after 10 years of protein dynamics modelling with python. *Bioinformatics*, 37(20):3657–3659, oct 2021.

[22] Antoine Koehl, Hongli Hu, Dan Feng, Bingfa Sun, Yan Zhang, Michael J Robertson, Matthew Chu, Tong Sun Kobilka, Toon Laeremans, Jan Steyaert, et al. Structural insights into the activation of metabotropic glutamate receptors. *Nature*, 566(7742):79–84, 2019.

[23] Xiao-chen Bai, Eeson Rajendra, Guanghui Yang, Yigong Shi, and Sjors HW Scheres. Sampling the conformational space of the catalytic subunit of human γ-secretase. *elife*, 4:e11182, 2015.

[24] Srayanta Mukherjee and Yang Zhang. Mm-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic acids research*, 37(11):e83–e83, 2009.

[25] Arthur Ecoffet, Frédéric Poitevin, and Khanh Dao Duc. MorphOT: Transport-based interpolation between EM maps with UCSF ChimeraX. *Bioinformatics*, 36(22-23):5528–5529, 2020.

[26] Arthur Ecoffet, Geoffrey Woollard, Artem Kushner, Frédéric Poitevin, and Khanh Dao Duc. Application of transport-based metric for continuous interpolation between cryo-EM density maps. *AIMS Mathematics*, 7(1):986–999, 2022.

[27] Nathan Zelesko, Amit Moscovich, Joe Kileel, and Amit Singer. Earthmover-based manifold learning for analyzing molecular conformation spaces. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1715–1719. IEEE, 2020.

[28] Rohan Rao, Amit Moscovich, and Amit Singer. Wasserstein k-means for clustering tomographic projections. *arXiv preprint arXiv:2010.09989*, 2020.

Supporting Material

EMPOT: partial alignment of density maps and atomic model fitting using unbalanced Gromov-Wasserstein divergence

The Supplemental Information contains:

- Appendix sections A–C
- Supplementary figures 1–2
- Supplementary Table S1

Appendix A: Alignment procedure

The procedure for partial alignment of density maps is summarized in Algorithm 1.

Algorithm 1 EMPOT: partial alignment of density maps and rigid body fitting using unbalanced Gromov-Wasserstein divergence

Input two 3D density maps $\mathcal{A}, \mathcal{B}$, number of sampled points $n \in \mathbb{R}$, regularization parameter $\epsilon \in \mathbb{R}_+$, unbalanced parameter $\rho \in \mathbb{R}_+$

1: Sample two sets of n points $\mathbf{A} = \{a_1, \ldots, a_n\}, \mathbf{B} = \{b_1, \ldots, b_n\} \subset \mathbb{R}^3$ from $\mathcal{A}, \mathcal{B}$ respectively, using TRNs

2: Compute P to be the transport plan matrix for $\mathcal{GW}^U(\mathbf{A}, \mathbf{B})$, using the algorithm from [1]

3: $\pi(a_i) = b_{\text{argmax}_j P_{i,j}}$

4: Compute R_{opt} and T_{opt} to be minimizers of $\text{argmin}_{R,T} \sum_{i=1}^{n} \|Ra_i + T - \pi(a_i)\|^2$, using the Kabsch algorithm

5: **return** R_{opt} and T_{opt}

Appendix B: Kabsch Algorithm

For completeness, we provide here the formulas obtained from the Kabsch algorithm:

In this algorithm, we first compute centroids by

$$\overline{a} = \frac{1}{n}\sum_{i=1}^{n} a_i, \ \overline{\pi(a)} = \frac{1}{n}\sum_{i=1}^{n} \pi(a_i).$$

Then we calculate the residue of vectors

$$a_{c_i} = a_i - \overline{a}, \ \pi_{c_i} = \pi(a_i) - \overline{\pi(a)}.$$

Next we compute the 3×3 matrix H as

$$H = \sum_{i=1}^{n} a_{c_i} \pi_{c_i}^t.$$

And finally we set $R_{opt} = VU^t$ and $T_{opt} = \overline{\pi(a)} - R_{opt}\overline{a}$, where $H = UDV^t$ is the singular value decomposition of H.

Appendix C: Partial alignment experiment with an alternative dataset

We conduct here the same experiment as in section 3.1 with a different structure of gamma secretase in class 3 of the apo- state ensemble, from PDB:5fn5 [2], that consists of four different chains shown in Figure S1. We partially aligned the density map generated from chain A to the density map of the whole structure. Benchmarking results are reported in Table S1, showing significant improvement over other methods (we did not include AlignOT and BOTalign as we already know that they do not handle partial mapping and produce bad partial alignment).

Supplementary Figures

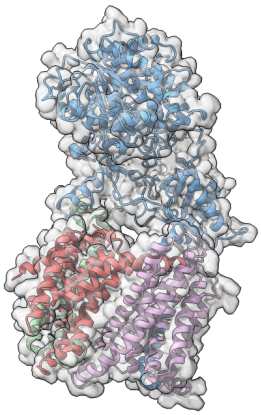

Figure S1. 3D structure of gamma secretase in class 3 of the apo-state ensemble used in our experiments and its four chains, with chain A in blue.

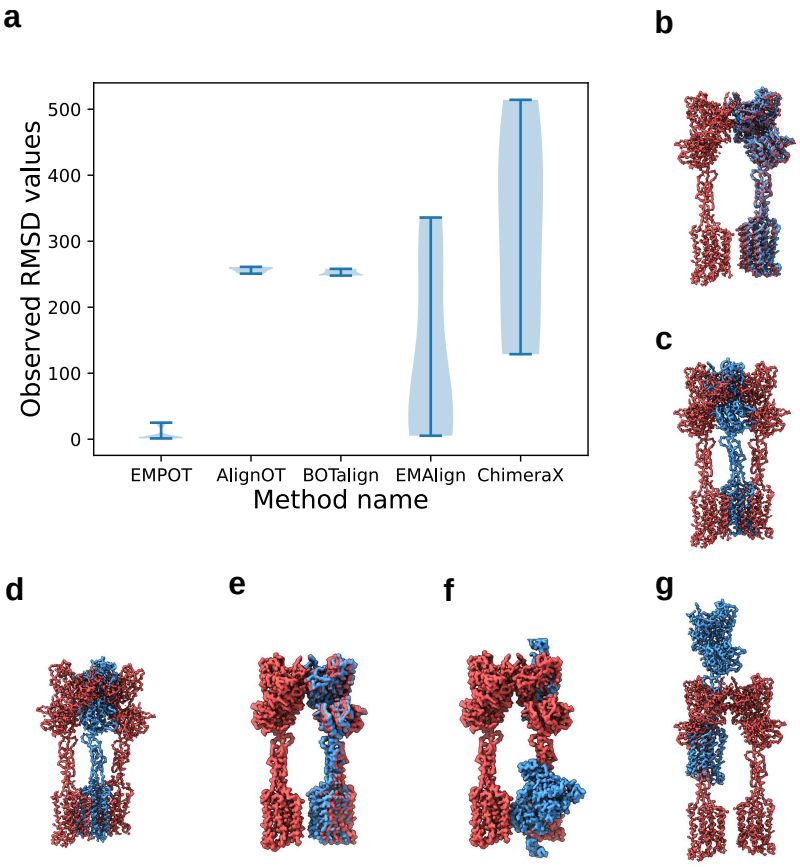

Figure S2. Detailed results of the experiment from Section 3.1 with representative examples. **(a)** Violin chart of distribution of RMSD for different alignments of each method. **(b)** Representative from our method (EMPOT) with an alignment close to the ground truth. As AlignOT and BOTalign rely on the assumption that density maps match entirely, they produce maps that get translated by matching centroids, resulting in wrong alignments in this case (**(c)**, **(d)** respectively). EMAlign managed to retrieve an alignment with low RMSD in some cases but got stuck in local minima in some other cases (**(e)**, **(f)** respectively). **(g)** ChimeraX's `fitmap` command has a high chance of getting stuck in local minima and finding irrelevant alignments.

Supplementary Table

Table S1. Benchmarking of methods for partial alignment of a single subunit density map to the whole structure (see Figure S1). We performed the alignment with each method 50 times and recorded the angle difference between the output and the ground truth, and the RMSD of paired atoms. We reported here the mean and std, with the best results for each measurement highlighted in bold.

Metric	EMPOT (ours)	EMAlign	ChimeraX
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

REFERENCES

1. Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. Advances in Neural Information Processing Systems, 34:8766–8779, 2021.

2. Xiao-chen Bai, Eeson Rajendra, Guanghui Yang, Yigong Shi, and Sjors HW Scheres. Sampling the conformational space of the catalytic subunit of human γ-secretase. elife, 4:e11182, 2015.