DiffFit: Differentiable Fitting of Molecule Structures to a Cryo-EM Map

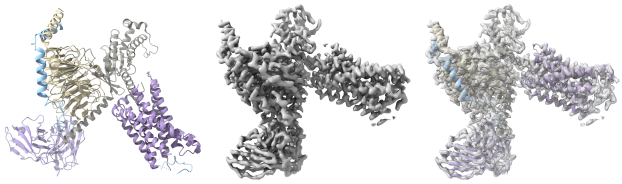
Deng Luo [1] **Zainab Alsuwaykit** [1] **Dawar Khan** [1] **Ondřej Strnad** [1] **Tobias Isenberg** [2] **Ivan Viola** [1]

Abstract

We introduce DiffFit, a differentiable algorithm for fitting protein atomistic structures into experimental Cryo-Electron Microscopy (cryo-EM) volume maps. This process is crucial in structural biology for reconstructing large meso-scale models of complex protein assemblies and complete cellular structures. Unlike current methods requiring manual fitting followed by automated fine-tuning, DiffFit enables automatic fitting with visual inspection and interactive revision. Our approach employs differentiable 3D rigid transformations and a novel loss function based on multi-resolution volume arrays and negative space exploitation. Evaluations on three use cases using realistic datasets demonstrate that DiffFit significantly outperforms previous methods. DiffFit is available as an open-source plugin (`github.com/nanovis/DiffFitViewer`) in ChimeraX, with all supplemental materials accessible at `osf.io/5tx4q`.

1. Introduction

In the field of structural biology, traditionally we have relied on techniques such as X-ray crystallography or nuclear magnetic resonance spectroscopy to understand the actual molecular compositions of cells and organelles—yet with the limitation that these could only provide (still impressive and highly useful) estimates or manually constructed models of the structure of actual biological samples (e.g., [1]–[6]). In the recent few years, however, the cryo-EM approach [7] has opened the doors to visualizing the biomolecules within *actual samples* at near-atomic resolution. In addition, the Protein Data Bank (PDB) initiative has over decades collected thousands of molecular models of the building blocks of the cells or organelles studied in structural biology. This new situation means that we are now at the brink of reconstructing the molecular composition of actual samples at a ground-truth level.



(a) Source structure. (b) Target map. (c) Fit result.

Figure 1. Fitting a single structure for 8JGF [8].

Yet to achieve this type of reconstruction we not only need to interactively visualize molecular data, for which tools [9] exist, but need to be able to faithfully place 3D models of known molecular building blocks, such as from PDB data and the AlphaFold predicted library [10], to the captured cryo-EM datasets (Fig. 1 and Fig. 2). So far, the fitting process involves a substantial time commitment and numerous manual interventions by the domain experts—rendering this process not very effective. The sheer complexity and size of the involved molecules, combined with the variability and noise inherent in cryo-EM data, pose substantial obstacles. A fully automatic process, in contrast, is also not the ideal solution because the existence of local minima (wrong placement of the compositing proteins) requires the domain experts to verify each placement where they utilize their knowledge and experience. Fully automated methods are currently far from feasible. Instead, we need the optimal balance between user interaction and automation.

For this purpose, we developed DiffFit, an automated differentiable fitting algorithm coupled with visual inspection and decision making, designed to optimize alignment between protein structures and experimental reconstructions of volumes (i. e., cryo-EM maps). Our technique works in both one-to-one fitting scenarios and many-to-one scenarios, in which multiple protein subunit structures are precisely aligned with a single, large, experimentally reconstructed volume. The DiffFit method is iterative and gradually brings the source protein structures into the target volumes to stepwise reconstruct the molecular-subunits

[1]Visual Computing Center, King Abdullah University of Science and Technology (KAUST), Saudi Arabia [2]Université Paris-Saclay, CNRS, Inria, LISN, France. Correspondence to: Deng Luo <deng.luo@kaust.edu.sa>.

Published at the 2nd Differentiable Almost Everything Workshop at the 41st International Conference on Machine Learning, Vienna, Austria. July 2024.

alignment. By leveraging advanced strategies such as volume filtering, multi-resolution volumes, and negative space utilization, we construct a loss function that allows us to quantify the fitting accuracy both during the iterations and for the final decision making. In this way we iteratively reduce the differences between the two representations—volumetric and atomistic—until we achieve the desired fit. Through this visually-guided fitting procedure we thus relieve the domain experts from manually placing structures as they assemble the reconstruction of the cryo-EM map, thus significantly speeding up the process into a manageable interactive procedure, that yields results at sufficient precision for the visualization and analysis of complex, real-world protein structures, ultimately facilitating large-scale structural modeling initiatives. In summary, we contribute:

- a differentiable fitting algorithm designed to fit multiple molecular subunits to a single reconstructed cryo-EM volume;
- human-in-the-loop strategy providing visual inspection and decision-making within an iterative structure-discovery cycle;
- a novel loss function and data processing that calculates new updates in each iteration to expedite algorithm convergence and quantify fitting accuracy; and
- three use-case scenarios of fitting either one or multiple known subunits or identifying yet unknown subunits being part of the molecular assembly.

The full paper describing the technique presented in this workshop submission is currently under review for the IEEE VIS 2024.

2. Inspiration

While we provide a detailed background and related work section in Appendix B, here we describe an image registration technique that inspired our work. Reddy et al. proposed a 2D differentiable compositing approach to identify patterns in repetitive textures from elementary patches [11]. Their method estimates occurrences and placement parameters (type, position, orientation, depth) of each patch in a composite image via a differentiable computing pipeline and a loss function to compare the synthesized image with the given input image to get the gradients to update the placement parameters. Reddy et al.'s approach is similar to our protein fitting problem, as both involve compositing element instances into a scene. However, their method cannot be directly applied to the protein fitting problem for the following reasons:

1. the pattern image and the element patches are defined in 2D with layers, while the cryo-EM map and protein subunits are defined in 3D;
2. the pattern image and the element patches are of the same representation, i.e., 2D grid data, while the cryo-EM map and protein subunits are of different representations—one is a 3D volume while the other is a set of atom coordinates that can be regarded as a point cloud;
3. the instance patches in differentiable compositing are all of the same size, while the protein subunits differ in numbers of atoms;
4. differentiable compositing expects the patches to overlap, while protein subunits do not spatially overlap; and
5. forming 1000 layers of 2D images is possible to fit into the current GPU memory, while forming 1000 3D volumes is prohibitive with the currently available GPU memory.

We indeed tried to extend the differentiable compositing approach to 3D, but it only succeeded in trivial cases and often fell into local minima in real-world scenarios. Examples are available at osf.io/5tx4q. We then decided to develop a new approach which we describe below.

3. Method

3.1. Differentiable structure fitting

Given a cryo-EM map, bioscientists often lack precise location and orientation parameters for protein sub-structures and may not even know which protein subunits are present in some regions. Our goal is to develop a new approach to address these challenges. For certain subunits where bioscientists are confident of their presence, our technique will help identify their placement parameters. For regions with unknown subunits, our method will identify potential candidates from a large database that best fit the cryo-EM map.

3.1.1. Sampling of one coordinate

Because our task is to find the optimal alignment of an atomistic molecular structure to the reconstructed cryo-EM volume map, we are searching for the optimal fit characterized by two rigid-body transformation parameters: a translational offset $\mathbf{p}$ and a rotation. The rotation is represented by a quaternion $\mathbf{q}$ or its corresponding rotation matrix $M_{\mathbf{q}}$. The position $\mathbf{x}_i$ corresponds to the center point of an atom i within the molecular subunit for which we are finding the fit. For the calculation of the fit, we transform every atom position in one subunit according to the rotation and translational offset: $T(\mathbf{x}_i) = M_{\mathbf{q}} \cdot \mathbf{x}_i + \mathbf{p}$. We *S*ample a density value D of the atom to be placed at position $T(\mathbf{x}_i)$ from a scalar volume V by trilinear interpolation as: $D(T(\mathbf{x}_i)) = S(T(\mathbf{x}_i), V)$.

3.1.2. Placement of one molecule (or subunit)

To find the best $\mathbf{p}$ and $\mathbf{q}$ parameters we formulate an initial loss function L that gives us the minimum negative average

density per atom for a molecular subunit with N atoms that form the set $\mathbf{X}_m$ of all atom center points $\mathrm{T}(\mathbf{x}_i) \in \mathbf{X}_m$ for a particular molecular subunit m:

$$L(\mathbf{p},\mathbf{q},\mathbf{X}_m,V) = -\left(\frac{1}{N}\sum_i^N D(T(\mathbf{x}_i))\right) = -\frac{1}{N}\sum_i^N S(M_\mathbf{q}\cdot\mathbf{x}_i+\mathbf{p},V). \tag{1}$$

For the optimization we rely on the calculation of the gradient of our differentiable formulation and use it with the Adam optimizer [12]. Despite the Adam optimizer being known for robustness with respect to local minima, our initial loss function formulation frequently led to such a local minimum, i.e., a place that is not an optimal placement of the molecule in the map but from which the optimizer cannot find better solution in the parameter-space neighborhood. Such local minima are a common and severe problem that is also manifested in the functionality of most commonly used tools for molecular subunit fitting (e.g., ChimeraX's fit-in-map feature). We thus introduced several strategies to form a novel loss function and, ultimately, make DiffFit more robust.

A first strategy that we found to substantially contribute to a good fitting performance is **filtering and normalizing of the input cryo-EM map volume** V. For this purpose we clamp the volume values based on a user-specified minimal threshold and a minimum size of connected-voxels that form a cluster. We also detect all voxels with a density value smaller than a given threshold and set them to zero. The size of the connected voxel cluster after thresholding has to be bigger than the cluster size hyperparameter. Otherwise, we set all the voxels in that cluster to zero. This step leads to the filtered volume V_F and ensures that only relevant volume regions are considered for fitting, improving the algorithm's focus and efficiency. We then normalize the filtered values to $[0,1]$—a typical practice in learning and optimization approaches—, which leads to a volume $\hat{V}_F$ and which turns out to be essential for controlling the magnitude of the calculations that lead to the loss function and hence the settings of the hyperparameters in the workflow.

To accommodate the inherent noise and variability within biological datasets, we also applied a series of convolution iterations to the target volume, and capture each smoothing as a separate volume. This iterative convolutional smoothing leads to an array of volumes and we use each of these volumes in the fitting process. This **multi-resolution approach** enhances the robustness of the fitting process by mitigating the impact of noise and data irregularities. We empirically found that a 10-element array of increasingly smoothed volumes, iteratively filtered with a Gaussian smoothing kernel, performs well. We expose this array's size as a hyperparameter to allow users to control it. We experimented with Laplacian smoothing as well, which led to unsatisfactory performance. We denote the non-smoothed volume as $\hat{V}_F^{G_0}$ and express the recurrent formulation of the iterative convolution smoothing as: $\hat{V}_F^{G_n} = \hat{V}_F^{G_{n-1}} * G_n$. A third adaptation we apply to the initial fitting process is a stricter penalization of a mismatch. If an atom center is placed within the cryo-EM map volume but outside the extent of the molecular target structure, i.e., outside of the target *footprint*, where normally the target density would be zero. To discourage such misalignment even more, we assign these regions a **negative value**. After smoothing, therefore, for those voxels with a density value of zero, we replace the zero with a negative value. We experimented with varying the negative values or creating a smooth gradient of negative values but found a constant value of -0.5 outside the molecular footprint within the map performs well. We expose this value, however, as a tunable hyperparameter. We denote the resulting volume as $\hat{V}_{F_{-c}}^{G_n}$ with $-c$ being the negative constant value. Finally, we update the loss function formulation with a volume smoothed after j iterations as: $L(\mathbf{p},\mathbf{q},\mathbf{X}_m,\hat{V}_{F_{-c}}^{G_j})$. We weigh each fit with a multi-resolution volume array element w_j for a total number of resolutions n, and sum up all the multi-resolution components to form the final loss function for one $\mathbf{p}$ and $\mathbf{q}$ pair:

$$L_m([p,q]) = \sum_{j=1}^{n} w_j \cdot L(\mathbf{p},\mathbf{q},\mathbf{X}_m,\hat{V}_{F_{-c}}^{G_j}). \tag{2}$$

3.1.3. Fitting multiple placements of one molecule

To start the optimization we need to initialize the position offset $\mathbf{p}$ and the rotation quaternion $\mathbf{q}$. We uniformly sample $N_\mathbf{q}$ points on a unit sphere and then convert them into quaternions to be applied for each offseted position. Instead of uniformly sampling positions from the volume's bounding box (as in ChimeraX), we uniformly sample $N_\mathbf{p}$ positions from the positive voxels in the filtered and normalized volume $\hat{V}_F$. This **enveloped sampling** increases our success rate by a factor of two by searching from $N_\mathbf{q}\cdot N_\mathbf{p}$ initial placements, as compared to ChimeraX's.

To look for fits for multiple copies of a single molecule m, we then take advantage of GPU parallelization and optimize all $N_\mathbf{q}\cdot N_\mathbf{p}$ pairs of $[\mathbf{p},\mathbf{q}]$ of the molecule with atoms $\mathbf{X}_m$ altogether in one single loss function:

$$L_{par}(m) = \sum_{k=1}^{N_\mathbf{q}\cdot N_\mathbf{p}} L_m([\mathbf{p}_k,\mathbf{q}_k]). \tag{3}$$

3.1.4. Fitting multiple placements of multiple molecules

Finally, as all subunit molecules have different numbers of atoms, it is not easy to parallelize the treatment of multiple molecules without overhead on the array padding of zeros. And because, usually, the $N_\mathbf{q}\cdot N_\mathbf{p}$ initial placements of $\mathbf{X}_m$ atoms would result in a total number of sampling operations higher than the total number of GPU threads, we process

Table 1. Performance results for fitting a single structure. Res stands for resolution, C stands for ChimeraX, D stands for DiffFit, M stands for MarkovFit, DC stands for DiffFit corrected by a single ChimeraX fit; G stands for Gain and is D/C for Hit rate and C/D for Time.

PDB	Res	Hit rate			Time (sec)			RMSD			
		C	D	G	C	D	G	C	M	D	DC
6MEO	3.90	7.4	**214.6**	41.2	128.2	**9.4**	13.7	0.489	1.940	0.790	**0.483**
7PM0	3.60	44.0	**195.2**	4.5	352.4	**7.0**	50.3	0.029	1.640	0.976	**0.027**
7SP8	2.70	4.6	**238.8**	117.5	130.58	**12.2**	10.9	0.996	1.290	0.779	**0.024**
6M5U	3.80	0.0	**277.0**	INF	162.2	**18.8**	8.2	69.413	2.36	0.944	**0.014**
5NL2	6.60	0.6	**179.6**	200.0	92.8	**9.6**	10.4	23.110	2.440	1.903	**0.047**
7K2V	6.60	49	**170.4**	3.3	240.6	**8.56**	29.3	0.338	2.440	1.532	**0.337**
3J1Z	13.00	138.6	**441.6**	3.20	64.4	**3.0**	21.5	**0.396**	32.330	2.436	0.399

different subunits molecules sequentially in a `for` loop and form an overall loss function for M molecules as

$$L_{all} = \sum_{l=1}^{M} L_{par}(l) \tag{4}$$

3.1.5. Quantify the fit quality

By sampling in the simulated volume from the molecule, we can get a weight for each atom coordinate as $W(\mathbf{x}) = S(\mathbf{x}, V_{sim})$. Then, for all the atoms in a molecule, we can form two vectors, one as sampled density vector $\mathbf{D} = [D(\mathbf{x}_1), D(\mathbf{x}_2), ..., D(\mathbf{x}_N)]$ from the target volume, the other as weight vector $\mathbf{W} = [W(\mathbf{x}_1), W(\mathbf{x}_2), ..., W(\mathbf{x}_N)]$ from the simulated volume. We can then calculate three alignment metrics, mean overlap μ, correlation ρ, and correlation about the mean ρ_μ as

$$\mu = \frac{\mathbf{D} \cdot \mathbf{W}}{N}$$

$$\rho = \frac{\mathbf{D} \cdot \mathbf{W}}{|\mathbf{D}||\mathbf{W}|}$$

$$\rho_\mu = \frac{(\mathbf{D} - D_\mu) \cdot (\mathbf{W} - W_\mu)}{|\mathbf{D} - D_\mu||\mathbf{W} - W_\mu|}$$

where the subtraction operator represents subtracting the scalar average densities D_μ and W_μ from each component of the sampled density vectors. We use these quality metrics during the interactive assessment by the bioscientist in ChimeraX that we describe in Appendix C.

4. Use Case Scenarios

Scenario 1: Fit a single structure DiffFit efficiently fits a single atomistic structure into a target volumetric map, as shown in Fig. 1. Table 1 compares DiffFit's performance with ChimeraX and MarkovFit [13]. The hit rate and computation time of DiffFit are significantly better than the previous methods. More details are in Appendix D.

Scenario 2: Composite multiple structures Fig. 2 shows an example of compositing multiple structures into a cryo-EM volume map, for the PDB-protein 8SMK [14]. In the first row we show how the middle and bottom parts are fitted in a first round of computation, with the remaining top part of the protein being fitted in a second round of our interactive process.

Table 2. Performance results of identifying unknown structures. C stands for ChimeraX, D for DiffFit; Gain = D/C.

Structure	C Hit	D Hit	Gain
I7MLV6_D3	108	280	2.6×
I7M317_D1	127	163	1.3×

Scenario 3: Identify unknown densities By fitting structures from a library into unidentified densities and evaluating the fit quality, researchers can hypothesize the identity of unknown components, aiding discovery-based research, which can lead to significant biological insights. Using a demo dataset (Fig. 3) from a recent automated domain-level protein identification technique, DomainFit [15], we identified two protein domains (I7MLV6_D3 and I7M317_D1) from 359 candidates with DomainFit, which took 12 hours. DiffFit achieved the same results in 10 minutes, showing a 2× gain in hit rate (Table 2) and a 72× gain in computation time.

5. Conclusion

DiffFit offers a novel and efficient solution to the challenges of atom-to-map fitting that has recently arisen in structural biology. We address these challenges with our differentiable fitting algorithm along with a set of essential strategies that make the algorithm robust for the domain problem, adding a human-in-the-loop visual analytics approach to the workflow, and providing our approach as an open-source package that is designed to work as part of a standard software tool of the domain (ChimeraX). By switching from pixel-to-pixel to point-to-volume fitting, we effectively addressed the domain-specific challenges, suggesting Reddy's [11] original method could also be optimized by focusing on essential pixels. Our performance metrics show significant improvements over traditional methods, indicating DiffFit's potential to pave the way for new structural biology workflows focusing on aligning resolved or predicted structures instead of *de novo* modeling. Current protocols like Assembline [16] and DomainFit [15] rely on ChimeraX's `fitmap` command, which can now be replaced by our more effective method. The speed of DiffFit also opens the possibility of scanning the whole set of known and predicted molecules with the current computational resources. In future work, by allowing, for example, rotations for the sub-domains and handling of collisions, DiffFit could also be applied for refining the nuances between the predicted structures and the experimental Cryo-EM map.

References

[1] M. Schäfer, N. Brich, J. Byška, *et al.*, "InVADo: Interactive visual analysis of molecular docking data," *IEEE Trans Vis Comput Graph*, vol. 30, no. 4, pp. 1984–1997, 2024. DOI: 10/gtnn2n.

[2] P. Ulbrich, M. Waldner, K. Furmanová, *et al.*, "sMolBoxes: Dataflow model for molecular dynamics exploration," *IEEE Trans Vis Comput Graph*, vol. 29, no. 1, pp. 581–590, 2023. DOI: 10/kvjv.

[3] D. Kuťák, M. N. Selzer, J. Byška, *et al.*, "Vivern—A virtual environment for multiscale visualization and modeling of DNA nanostructures," *IEEE Trans Vis Comput Graph*, vol. 28, no. 12, pp. 4825–4838, 2022. DOI: 10/mpwt.

[4] R. Skånberg, P.-P. Vázquez, V. Guallar, and T. Ropinski, "Real-time molecular visualization supporting diffuse interreflections and ambient occlusion," *IEEE Trans Vis Comput Graph*, vol. 22, no. 1, pp. 718–727, 2016. DOI: 10/mpwv.

[5] N. Nguyen, O. Strnad, T. Klein, *et al.*, "Modeling in the time of COVID-19: Statistical and rule-based mesoscale models," *IEEE Trans Vis Comput Graph*, vol. 27, no. 2, pp. 722–732, 2021. DOI: 10/k8sh.

[6] D. Duran, P. Hermosilla, T. Ropinski, B. Kozlíková, Á. Vinacua, and P.-P. Vázquez, "Visualization of large molecular trajectories," *IEEE Trans Vis Comput Graph*, vol. 25, no. 1, pp. 987–996, 2019. DOI: 10/gjbdnk.

[7] T. Nakane, A. Kotecha, A. Sente, *et al.*, "Single-particle cryo-EM at atomic resolution," *Nature*, vol. 587, no. 7832, pp. 152–156, 2020. DOI: 10/gjwcfp.

[8] L. Guo, Y. Zhang, G. Fang, *et al.*, "Ligand recognition and G protein coupling of the human itch receptor MRGPRX1," *Nat Commun*, vol. 14, 2023. DOI: 10/mpt4.

[9] R. Skånberg, I. Hotz, A. Ynnerman, and M. Linares, "VIAMD: A software for visual interactive analysis of molecular dynamics," *J Chem Inf Model*, vol. 63, no. 23, pp. 7382–7391, 2023. DOI: 10/gs7tzh.

[10] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021. DOI: 10/gk7nfp.

[11] P. Reddy, P. Guerrero, M. Fisher, W. Li, and N. J. Mitra, "Discovering pattern structure using differentiable compositing," *ACM Trans Graph*, vol. 39, no. 6, 2020, ISSN: 0730-0301. DOI: 10/gtmxvb.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint 1412.6980, 2017. DOI: 10/hnkr.

[13] E. Alnabati, J. Esquivel-Rodriguez, G. Terashi, and D. Kihara, "MarkovFit: Structure fitting for protein complexes in electron microscopy maps using Markov random field," *Front Mol Biosci*, vol. 9, 2022, ISSN: 2296-889X. DOI: 10/gtmxvj.

[14] X. Zhou, S. Kong, A. Maker, *et al.*, "Antibody discovery identifies regulatory mechanisms of protein arginine deiminase 4," *Nat Chem Biol*, 2024, To appear. DOI: 10/mpwf.

[15] J. Gao, M. Tong, C. Lee, J. Gaertig, T. Legal, and K. H. Bui, "DomainFit: Identification of protein domains in cryo-EM maps at intermediate resolution using AlphaFold2-predicted models," bioRxiv preprint 2023.11.28.569001, 2023. DOI: 10/gs63f2.

[16] V. Rantos, K. Karius, and J. Kosinski, "Integrative structural modeling of macromolecular complexes using Assembline," *Nat Protoc*, vol. 17, pp. 152–176, 2022. DOI: 10/gq586q.

[17] M. A. Herzik, M. Wu, and G. C. Lander, "High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM," *Nat Commun*, vol. 10, 2019. DOI: 10/gg7zp4.

[18] T. Walton, M. Gui, S. Velkova, *et al.*, "Axonemal structures reveal mechanoregulatory and disease mechanisms," *Nature*, vol. 618, no. 7965, pp. 625–633, 2023. DOI: 10/gscb95.

[19] X.-c. Bai, G. McMullan, and S. H. W. Scheres, "How cryo-EM is revolutionizing structural biology," *Trends Biochem Sci*, vol. 40, no. 1, pp. 49–57, 2015, ISSN: 0968-0004. DOI: 10/f6wq7v.

[20] S. Malhotra, S. Träger, M. Dal Peraro, and M. Topf, "Modelling structures in cryo-EM maps," *Curr Opin Struct Biol*, vol. 58, pp. 105–114, 2019, ISSN: 0959-440X. DOI: 10/gtmxvc.

[21] K. Lasker, O. Dror, M. Shatsky, R. Nussinov, and H. J. Wolfson, "EMatch: Discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps," *IEEE/ACM Trans Comput Biol Bioinf*, vol. 4, no. 1, pp. 28–39, 2007. DOI: 10/ccbr4m.

[22] W. Chen, X. Wang, and Y. Wang, "FFF: Fragment-guided flexible fitting for building complete protein structures," in *Proc. CVPR*, Los Alamitos: IEEE CS, 2023, pp. 19 776–19 785. DOI: 10/gtmxvf.

[23] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys Med Biol*, vol. 46, no. 3, R1–R45, 2001. DOI: 10/fgbhgj.

[24] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: A review," *Phys Med Biol*, vol. 65, no. 20, 2020. DOI: 10/ghn45h.

[25] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int J Comput Vision*, vol. 60, no. 2, pp. 91–110, 2004. DOI: 10/bqrmsp.

[26] J. P. Didon and F. Langevin, "Registration of MR images: From 2D to 3D, using a projection based cross correlation method," in *Proc. EMBC*, vol. 1, Piscataway: IEEE, 1995, pp. 489–490. DOI: 10/bgp7mt.

[27] M. Malinsky, R. Peter, E. Hodneland, A. J. Lundervold, A. Lundervold, and J. Jan, "Registration of FA and T1-weighted MRI data of healthy human brain based on template matching and normalized cross-correlation," *J Digit Imaging*, vol. 26, no. 4, pp. 774–785, 2013. DOI: 10/gtmxw5.

[28] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 986–1004, 2003. DOI: 10/czrq8h.

[29] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans Med Imaging*, vol. 22, no. 1, pp. 120–128, 2003. DOI: 10/cs4pch.

[30] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans Image Process*, vol. 9, no. 12, pp. 2083–2099, 2000. DOI: 10/fs4j7h.

[31] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *Int J Comput Vis*, vol. 81, no. 3, pp. 227–239, 2009. DOI: 10/dghjcp.

[32] L. Shang, J. Lv, and Z. Yi, "Rigid medical image registration using PCA neural network," *Neurocomput*, vol. 69, no. 13–15, pp. 1717–1722, 2006, ISSN: 0925-2312. DOI: 10/fsrrb2.

[33] P. Arora, R. Mehta, and R. Ahuja, "An integration of metaheuristic approach utilizing kernel principal component analysis for multimodal medical image registration," *Cluster Comput*, 2024, To appear. DOI: 10/gtmxvd.

[34] M. Kesäniemi and K. Virtanen, "Direct least square fitting of hyperellipsoids," *IEEE Trans Pattern Anal Mach Intell*, vol. 40, no. 1, pp. 63–76, 2018. DOI: 10/gcqjrb.

[35] M. Falk, V. Tobiasson, A. Bock, C. Hansen, and A. Ynnerman, "A visual environment for data driven protein modeling and validation," *IEEE Trans Vis Comput Graph*, 2024, To appear. DOI: 10/gsc3c9.

[36] A. Seiler, D. Großmann, and B. Jüttler, "Spline surface fitting using normal data and norm-like functions," *Comput Aided Geom Des*, vol. 64, pp. 37–49, 2018, ISSN: 0167-8396. DOI: 10/gd5gt4.

[37] Y.-B. Kou, Y.-F. Feng, L.-Y. Shen, X. Li, and C.-M. Yuan, "Adaptive spline surface fitting with arbitrary topological control mesh," *IEEE Trans Vis Comput Graph*, 2024, To appear. DOI: 10/gtmxw4.

[38] I.-C. Yeh, C.-H. Lin, O. Sorkine, and T.-Y. Lee, "Template-based 3D model fitting using dual-domain relaxation," *IEEE Trans Vis Comput Graph*, vol. 17, no. 8, pp. 1178–1190, 2011. DOI: 10/dkbjmd.

[39] Y. Zhou, Z. Zhu, X. Bai, D. Lischinski, D. Cohen-Or, and H. Huang, "Non-stationary texture synthesis by adversarial expansion," *ACM Trans Graph*, vol. 37, no. 4, 2018, ISSN: 0730-0301. DOI: 10/gd52tv.

[40] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. ICCV*, Los Alamitos: IEEE CS, 2019, pp. 4569–4579. DOI: 10/gg8fc3.

[41] E. F. Pettersen, T. D. Goddard, C. C. Huang, *et al.*, "UCSF Chimera—A visualization system for exploratory research and analysis," *J Comput Chem*, vol. 25, no. 13, pp. 1605–1612, 2004. DOI: 10/b4bq4c.

[42] T. D. Goddard, C. C. Huang, and T. E. Ferrin, "Visualizing density maps with UCSF Chimera," *J Struct Biol*, vol. 157, no. 1, pp. 281–287, 2007, Software tools for macromolecular microscopy, ISSN: 1047-8477. DOI: 10/ft849s.

[43] K. Xu, Z. Wang, J. Shi, H. Li, and Q. C. Zhang, "A2-Net: Molecular structure estimation from cryo-EM density volumes," *Proc AAAI Conf Artif Intell*, vol. 33, no. 1, pp. 1230–1237, 2019. DOI: 10/ghkjcf.

[44] T. Terwilliger, P. Adams, P. Afonine, and O. Sobolev, "A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps," *Nat Methods*, vol. 15, no. 11, pp. 905–908, 2018. DOI: 10/gfnb8d.

[45] X. Wang, E. Alnabati, T. W. Aderinwale, S. R. M. V. Subramaniya, G. Terashi, and D. Kihara, "Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning," *Nat Commun*, vol. 12, 2020. DOI: 10/gmz55b.

[46] B. Frenz, A. C. Walls, E. H. Egelman, D. Veesler, and F. DiMaio, "RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps," *Nat. Methods*, vol. 14, no. 8, pp. 797–800, 2017. DOI: 10/gf5cmq.

[47] R. Y.-R. Wang, M. Kudryashev, X. Li, *et al.*, "De novo protein structure determination from near-atomic-resolution cryo-EM maps," *Nat Methods*, vol. 12, no. 4, pp. 335–338, 2015. DOI: 10/gf5cn8.

[48] V. Mallet, C. Rapisarda, H. Minoux, and M. Ovsjanikov, "Finding antibodies in cryo-EM densities with CrAI," bioRxiv preprint 2023.09.27.559736, 2023. DOI: 10/gtn9kj.

[49] J. Pfab, P. Minh Nhut, and D. Si, "DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes," *PNAS*, vol. 118, no. 2, 2021. DOI: 10/gjnkwt.

[50] B. He, F. Zhang, C. Feng, J. Yang, X. Gao, and R. Han, "Accurate global and local 3D alignment of cryo-EM density maps using local spatial structural features," *Nat Commun*, vol. 15, 2024, ISSN: 2041-1723. DOI: 10/gtmxvg.

[51] J. I. Garzón, J. Kovacs, R. Abagyan, and P. Chacón, "ADP_EM: Fast exhaustive multi-resolution docking for high-throughput coverage," *Bioinf*, vol. 23, no. 4, pp. 427–433, 2006, ISSN: 1367-4803. DOI: 10/c9xkmg.

A. Acknowledgments

This research was supported by King Abdullah University of Science and Technology (KAUST) (BAS/1/1680-01-01).

B. Background and related work

We rely on cryo-EM data, so below we first briefly provide essential relevant background. Then we describe past work on image registration and model fitting and show how both relate to our own research.

B.1. Brief background

Structural biology employs various techniques to understand how atoms are arranged in macromolecular complexes, i. e., samples in the range of 60 kDa (i. e., 4,472 atoms; PDB 6NBD [17]) to 50,000 kDa (3,163,608 atoms; PDB 8J07 [18]) that are essential for scientists to study processes in living cells—cryo-EM being a particularly powerful one [19]–[21]. With cryo-EM, bioscientists capture images using an electron microscope of flash-frozen biological specimens, preserving their natural structure without any staining or fixing [22], which would otherwise interfere with the sample. These images are then used to construct cryo-EM 3D volumes or maps using the so-called *single particle method* that aligns thousands of projections from structurally identical molecular instances into a single map (using the Fourier slice-projection theorem) representing the electron density of the sample, which can be used to infer the atom positions within the molecule.

Subsequently, the bioscientists need to build accurate atomistic or molecular models that match the electron density map obtained from the cryo-EM process to gain insight into molecular function and interactions. This process involves mapping or fitting known sub-molecules into their corresponding positions within the map. The objective is to achieve an optimal correspondence between the model and the experimental or simulated volume, revealing the organization of molecules in 3D space, including single molecules, complexes, and the placement of small molecules and ligands into binding sites. Molecular models are available in the Protein Data Bank (PDB, `rcsb.org`), accessible in various formats such as PDB, Crystallographic Information File (CIF), and mmCIF (macromolecular CIF). The fitting itself is today typically achieved through manual placement, alignment, and comparison with the density maps. The manual character of this process makes it extremely time-consuming and tedious, and means that only expert biologists can complete it. To address this challenge, numerous studies have aimed to automate the fitting process, focusing on image registration as the foundation and exploring methods to streamline 3D model construction as we review below.

B.2. Image registration and geometric fitting

The fitting of 3D structures into captured or simulated volumes relates to the problem of image registration in image processing. It entails aligning two images within a shared reference frame, regardless of whether they originate from the same or different modalities [23], [24]. This process involves feature extraction, determining transformations, and assessing accuracy through metrics. Scale-invariant features from images [25], for example, can facilitate matching across diverse views, despite significant distortions or variations. This process involves detecting invariant keypoints using the difference-of-Gaussian function, determining locations and scales, assigning directions based on local gradients, and measuring gradients within selected scales around each keypoint. Extracted features are stored in a database and can then be matched with new images using fast nearest-neighbor algorithms, with applications including object recognition. Among many applications of the process, physicians rely on various imaging modalities to diagnose patients, each capturing images with differing orientation. Image registration addresses this variability by aligning images within a unified frame by optimizing parameters like orientation and translation. Medical image registration is an active research area which encompasses diverse methods, including techniques based on cross-correlation [26], [27] and those based on mutual information [28]–[31]. Shang *et al.* [32], e.g., introduced a method for medical image registration using principal component analysis (PCA) neural networks to extract feature images and compute rotation angles and translation parameters by aligning the first principal directions and centroids in a simple and efficient way. For complex spatial transformations, another recent approach [33] uses Kernel PCA and Teaching-Learning-based optimization (TLBO) for multi-modal image registration. Similar to these methods, transformations and alignments need to be determined to fit the atomistic model into a volumetric map. We can thus also use optimization techniques in cryo-EM map fitting to refine the fit and optimize parameters such as orientation and translation—which is what we show with our work. The major difference to image registration is that, in our workflow, we fit two different data representations and one is a sub-part of the whole that is potentially present at multiple locations in the target volume.

The model-to-data fitting that we need for cryo-EM data is a problem that has also been investigated in depth in computer graphics and pattern recognition [11], [34]. It has applications not only in structural biology but is also needed in architectural geometry, virtual and augmented reality, robotics, and various other fields [35]–[38]. The key challenges in geometric fitting include accuracy, efficiency, robustness, and usability of the fitting module [36], [37]. Structural biology, in contrast, has special challenges such as noisy data, non-geometric shapes, and large data sizes so that geometric fitting methods are not directly applicable.

Yet our DiffFit algorithm still relates to techniques from computer graphics and pattern analysis. The differentiable compositing technique proposed by Reddy *et al.* [11], in particular, offers valuable insights into addressing fitting challenges as well as manipulating and understanding image patterns. With *differentiable compositing* we can handle patterns effectively, outperforming state-of-the-art alternatives in pattern manipulation [39], [40]. Reddy *et al.*'s method [11] discovers complex patterns by aligning elements with their own position and rotation, and facilitates refinement based on similarity to the target for precise adjustment. In addition, Reddy *et al.* use a multi-resolution pyramid—relevant for handling the multi-resolution volumetric data in our domain. Reddy *et al.*'s method [11], however, is restricted to certain pattern types, requires manual element marking, and may not always find the best solution, leading to orientation errors and missed elements. Nevertheless, we build our solution on top of their differentiable compositing.

Another approach, spline surface fitting [36], enhances the smoothness in aircraft engine geometry reconstruction by concurrently approximating point and normal data, ensuring boundary smoothness and optimal convergence, while exploring norm-like functions' effects on error measurement. A further recently proposed adaptive spline surface fitting method [37] employs surface meshes for high-precision CAD applications, supported by empirical evidence. The reliance on control meshes of this approach, however, limits its applicability to irregular topologies and compromise the preservation of sharp features. All these methods have common objectives and tasks such as similarity measures, pattern matching, fitting, and geometric transformations and, thus, can serve as a motivation and starting point toward our goals in structural biology. Since structural biology data often consists of large, complex structures without regular shapes like CAD models or easy representations in geometric meshes with smooth surfaces, however, the aforementioned methods are not directly applicable to our data.

B.3. Fitting in structural biology

Existing fitting methods for structural biology can roughly be categorized into manual, semi-automated, and automated approaches, each with its own advantages and challenges when used for aligning molecular models with cryo-EM density maps. Manual or semi-automated methods naturally involve human intervention, yet they provide control and precision—which is particularly beneficial in the structural analysis of complex datasets or when specific adjustments are needed for accuracy. For example, a popular tool for molecular manipulation and visualization, UCSF ChimeraX [41], includes the *fitmap* technique [42]. It suggests multiple possible placements of the atomistic model on the density map and then asks the user to make the final decision. The fitting process alternates between rigid-body rotation and gradient descent translation. In this way it maximizes the alignment between the atomic model and the density data by optimizing the sum of density values. Nonetheless, all of these manual or semi-automatic approaches are time-consuming and require a significant level of expertise.

To tackle this challenge and to automate the fitting process, researchers have developed methods that rely on deep learning (DL) [43]–[45]. A^2-Net by Xu *et al.* [43], for example, uses DL to accurately determine amino acids within a 3D cryo-EM density volume. It employs a sequence-guided Monte Carlo Tree Search (MCTS) to traverse candidate amino acids, considering the sequential nature of amino acids in a protein. Here, the authors divide the problem of molecular structure determination into three subproblems: amino acid detection in the density volumes, assignment of atomic coordinates to determine the position of each amino acid, and main chain threading to resolve the sequential order of amino acids that form each protein chain. Xu *et al.*'s method also demonstrates a remarkable speed improvement, being 100 times faster in finding solutions at runtime than existing methods [46], [47], while achieving an excellent accuracy of 89.8%. In addition, they introduced the A^2 dataset with 250,000 amino acids in 1,713 cryo-EM density volumes, with a resolution of 3 A°, pioneering automated molecular structure determination training benchmarks.

Another recent method by Mallet *et al.* [48] for finding antibodies in cryo-EM densities CrAI uses machine learning (ML). They formulate the objective as an object detection problem, using the structural properties of Fabs (Fragment antigen-binding) and VHHs (single-domain antibodies). Furthermore, DeepTracer [49] is a fully automated DL-based method designed to determine the all-atom structure of a protein complex using its high-resolution cryo-EM map and amino

acid sequence. It employs a customized deep convolutional neural network primarily for precise prediction of the protein structure, including the locations of amino acids, backbone, secondary structure positions, and amino acid types. The reprocessed cryo-EM maps are the input to the neural network, which transforms the output into a protein structure. Despite yielding accurate outcomes, the resulting atomstic structures may exhibit geometric issues, local fit to map discrepancies, misplaced side chains, or errors in tracing and/or connectivity. All DL-based techniques, moreover, not only require a substantial amount of time for the training (as opposed to the runtime) but also rely on large training datasets of cryo-EM volumes and manually fitted sub-molecules—which is why we do not resort to DL approaches.

An alternative to DL is map-to-map alignment, which is used to accurately align two-dimensional or three-dimensional maps to facilitate comparison and analysis of spatial structures or features within the maps. CryoAlign [50] is a cryo-EM density map alignment method that achieves fast, accurate, and robust comparison of two density maps based on local spatial feature descriptors. This approach involves sampling the density map to generate a point cloud representation and extracting key points by clustering based on local properties. CryoAlign then calculates local feature descriptors to capture structural characteristics, reducing the number of points considered and improving efficiency. By employing a mutual feature-matching strategy, CryoAlign establishes correspondences between keypoints in different maps and uses iterative refinement to enhance alignment. A combination of fast rotational matching search based on spherical harmonics and translational scans [51] yields accurate fitting results in seconds to a few minutes. This ADP_EM approach is reliable, in particular, for fitting X-ray crystal structures to low-resolution density maps, with reduced docking times and while maintaining a thorough 6D exploration with fine rotational sampling steps to find valid docking solutions.

In our work, we design a differentiable optimization method for fitting atomistic structures into volumetric data with the goal of precise fitting with fast-enough computation to be applicable in semi-automatic fitting in the standard tool ChimeraX. For this purpose we make use of the PyTorch capabilities for GPU parallel computing, trilinear interpolation sampling in volumetric data, and auto differentiation.

C. Visually-guided fitting

A critical aspect of the post-processing of DiffFit involves the clustering and sorting of fitting results to facilitate user-guided selection and refinement. After the optimization phase, the algorithm generates a vast array of potential fits, characterized by their translation and rotation parameters. To manage this abundance of data and facilitate efficient result exploration, we apply a clustering algorithm to group the fitting results based on their spatial and orientational similarity.

Each cluster represents a set of fits that are closely related, suggesting a consensus among them regarding the position and orientation of the fitted structure within the target volume. We sort these clusters based on a defined metric, such as the overall density overlap or correlation coefficient we just discussed, ensuring that the most promising fits are prioritized for user review. This hierarchical organization allows researchers to quickly identify the most accurate and relevant fitting results, streamlining the analysis process.

To further assist in the exploration of fitting results, we created an interactive visual browser as a comprehensive visualization tool that presents the sorted clusters in a user-friendly format (Fig. 4). This browser displays key metrics for each cluster, including the average correlation coefficient, density overlap, and a consensus error measure, providing a quick overview of each cluster's quality and relevance. The browser also allows users to select a cluster and visually inspect the fitting results within the 3D context of the target volume. This interactive exploration is crucial for assessing the fit quality in complex cryo-EM map regions, where subtle differences in position or orientation can substantially impact the biological interpretation of the results.

An innovative feature of our approach is the ability to iteratively refine the fitting process through the selective exclusion of already placed molecules' densities. Once a cluster is selected and its fit (or multiple fits) is verified as accurate by the bioscientist, the corresponding density within the target volume can be *zeroed-out*, effectively removing the respective volume region from further consideration in the following part of the fitting process. This step is crucial for complex volumes containing multiple closely situated structures, as it prevents the algorithm from repeatedly fitting structures to the same volume region and reduces false positives when fitting the remaining region.

Thus by iteratively fitting and zeroing out densities, users can progressively shrink the target volume, isolating and identifying individual structures within a dense or complex dataset. This iterative refinement approach ensures that the fitting process is not only guided by the algorithm's optimization but also by the user's expert knowledge and visual assessment, ultimately leading to more accurate and biologically meaningful results.

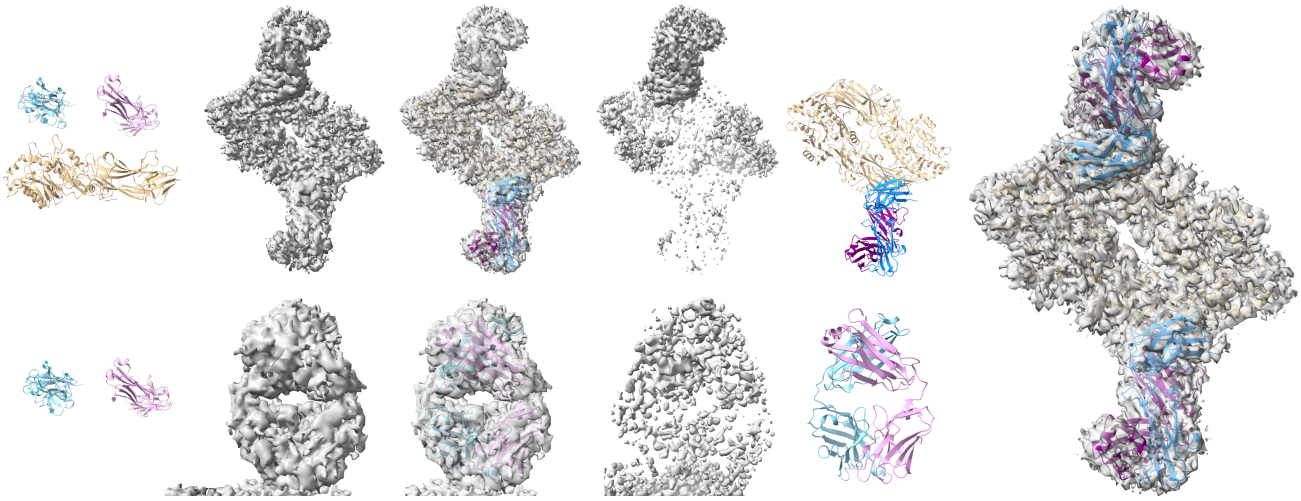


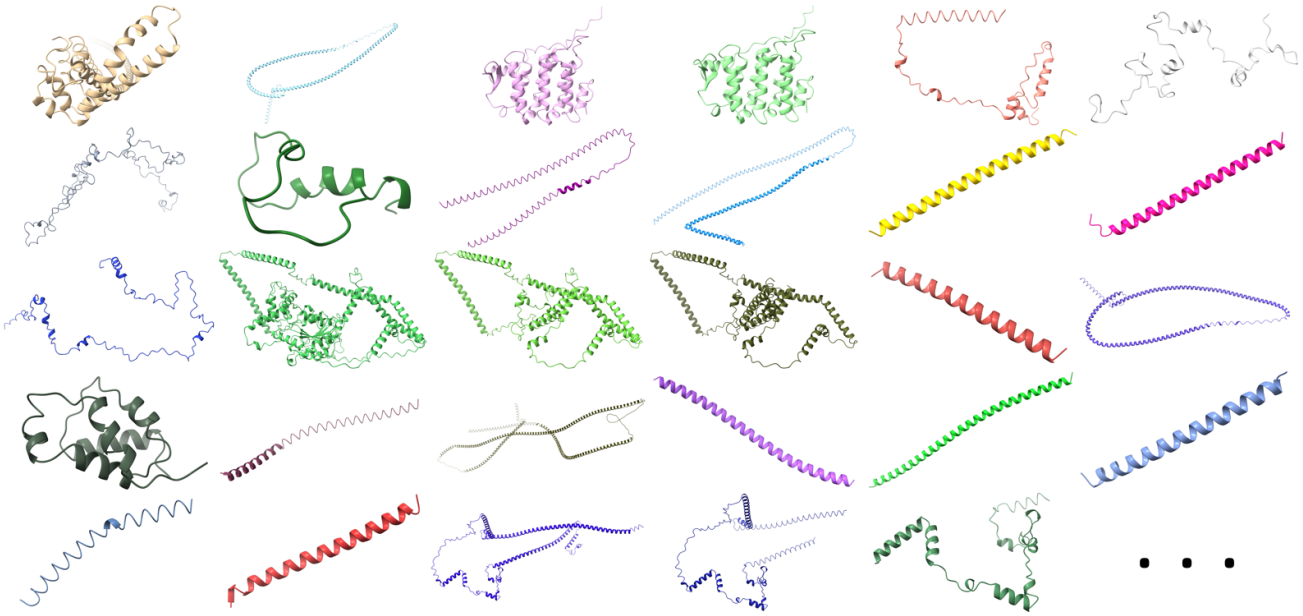
Figure 2. Compositing a protein (PDB 8SMK [14]) from its three unique chains. Top row from left to right: three input chains, input target volume, the best fits in the first fitting round, the remaining voxels after *zeroing-out*, and the fitted chains. Bottom row from left to right: two remaining input chains, remaining region of interest in the target volume from the first round, the best fits in the second round, the remaining voxels after *zeroing-out*, the fitted chains. Right: the final composited structure overlaid on the original target volume. The involved computation takes 30 seconds in total, and the human-in-the-loop interaction takes ≈ 5 minutes.

Our resulting visually-guided fitting framework enhances the DiffFit algorithm by integrating clustering, sorting, and interactive exploration tools. These features enable users to efficiently filter through large datasets of fitting results, identify the most promising fits, and iteratively refine the fitting process based on visual assessment. The combination of automated optimization with user-guided inspection and filtering addresses the challenge of accurately fitting molecular subunit structures within volumetric data.

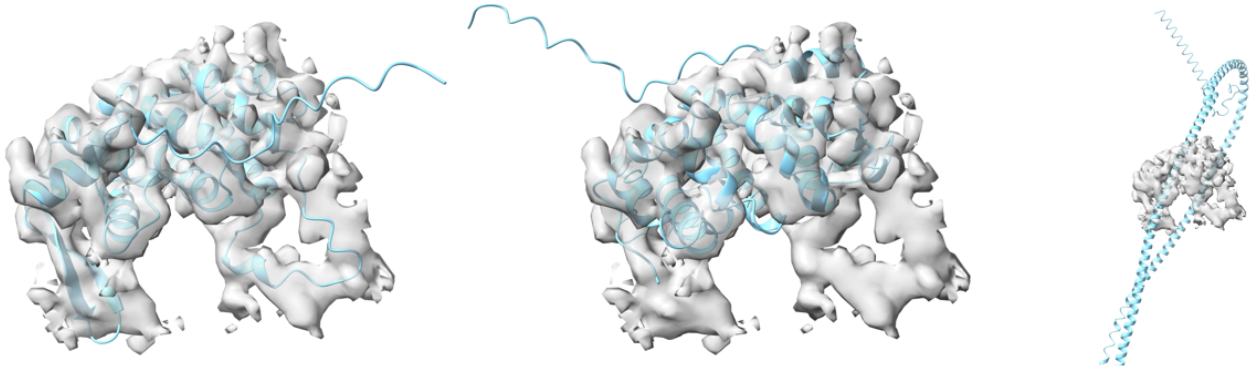
D. Benchmark results

We use the dataset reported in the recent MarkovFit work [13] to benchmark DiffFit's performance and compare it with ChimeraX and MarkovFit and report the results [the successful hit rate, computation time, and the root-mean-square deviation (RMSD)] in Table 1. The DiffFit fitting process is automatic upon specifying a density threshold to differentiate the background noise, removing the prerequisite of manually placing the structure at an approximate orientation close to the final optimal. For each experiment run, we fit 1000 times to perform the search, and for each structure, we perform 5 runs and average the metrics to obtain reliable results. The number of fits in the top-ranked cluster is regarded as the hit rate if the representative fit of that cluster is within 3 Angstroms and 6 degrees (which is the same as ChimeraX's threshold) from the ground truth. We do not repeat MarkovFit's computation as it takes an average of "7.7" + "6.25" hours to finish for each structure but directly take the author-reported RMSD. We take the "top-scored" model's RMSD (although it is often the same as that of the "best model by RMSD among top 10") because, in practice, there is no ground truth to compare with in advance to get the best model. The hit rate of ChimeraX for PDB 5NL2 is less than 1 because out of 5 runs, there are 2 runs that could not find a reasonable fit within the threshold from the ground truth. This is also the reason why the average RMSD is large. The hit rate and computation time of DiffFit are significantly better than those of previous methods. The RMSD of DiffFit is also significantly better than MarkovFit but often very slightly worse than ChimeraX. However, this can always be corrected (the last column in the table) by a single and automatic fit using ChimeraX's fitting. The reported performance values are based on a workstation that uses a Nvidia RTX 4090 GPU for DiffFit, and a single thread on an AMD Ryzen Threadripper PRO 3995WX 2.70 GHz for ChimeraX (version 1.7.1 (2024-01-23)) fitmap command.

E. Supplementary figures



(a) Library of structures to search against (subset).



(b) Unknown density identification; comparison of three potential fits which are overlaid on top of the target volumes.

Figure 3. Unknown density identification, where dozens to hundreds of molecular structures can be evaluated for potential fit.

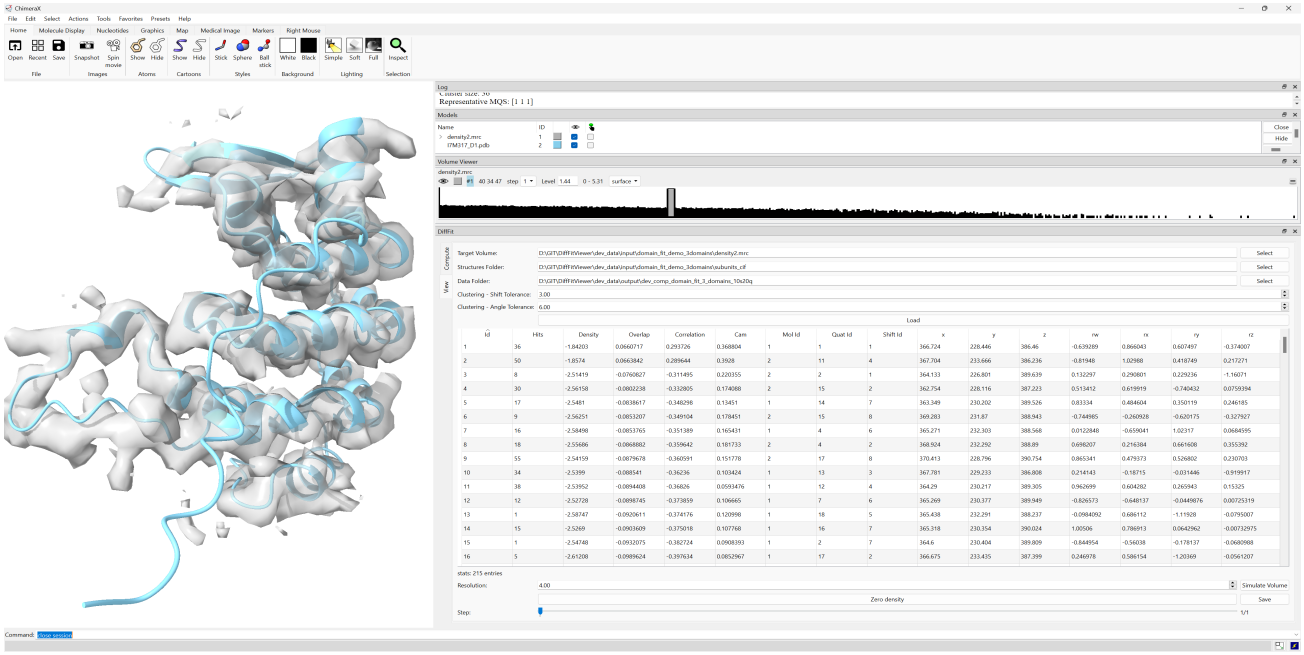


Figure 4. Our visual browser based on ChimeraX. The target volume in the left is overlaid with a fitted molecule corresponding to the selected fit result in the table on the right (clustered fits, each row being the representative placement with the highest correlation from that cluster). After inspection, users can save the placement and then select "Simulate volume" and "Zero density" to zero out the corresponding voxels from the target volume.