







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

Deng Luo , Zainab Alsuwaykit , Dawar Khan , Ondřej Strnad , Tobias Isenberg , and Ivan Viola 

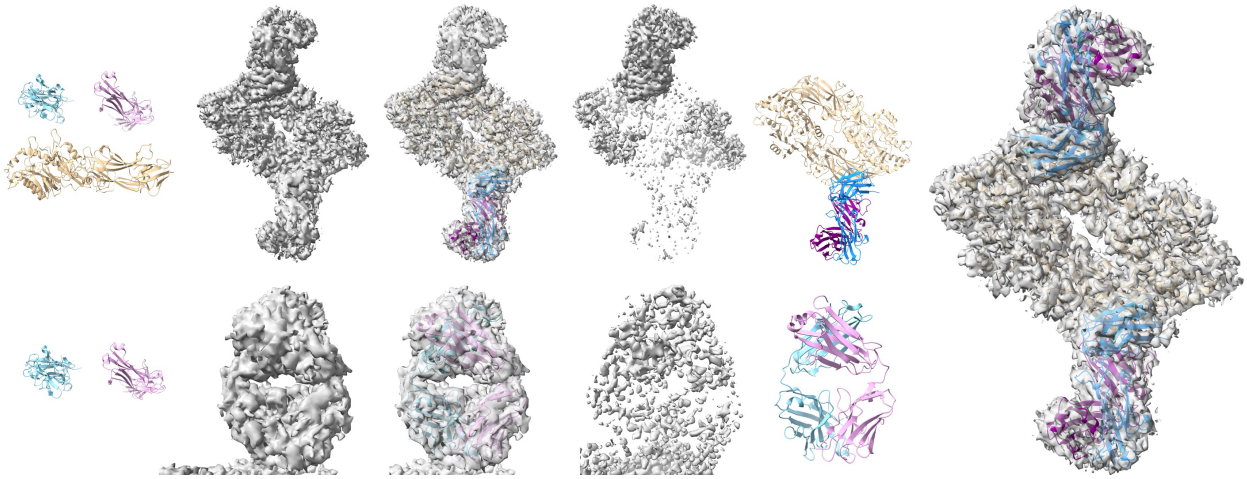


Fig. 1: Compositing a protein (PDB 8SMK [51]) from its three unique chains. Top row from left to right: three input chains, input target volume, the best fits in the first fitting round, the remaining voxels after *zeroing-out*, and the fitted chains. Bottom row from left to right: two remaining input 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.

Abstract—We introduce DiffFit, a differentiable algorithm for fitting protein atomistic structures into an experimental reconstructed Cryo-Electron Microscopy (cryo-EM) volume map. In structural biology, this process is necessary to semi-automatically composite large mesoscale models of complex protein assemblies and complete cellular structures that are based on measured cryo-EM data. The current approaches require manual fitting in three dimensions to start, resulting in approximately aligned structures followed by an automated fine-tuning of the alignment. The DiffFit approach enables domain scientists to fit new structures automatically and visualize the results for inspection and interactive revision. The fitting begins with differentiable three-dimensional (3D) rigid transformations of the protein atom coordinates followed by sampling the density values at the atom coordinates from the target cryo-EM volume. To ensure a meaningful correlation between the sampled densities and the protein structure, we proposed a novel loss function based on a multi-resolution volume-array approach and the exploitation of the negative space. This loss function serves as a critical metric for assessing the fitting quality, ensuring the fitting accuracy and an improved visualization of the results. We assessed the placement quality of DiffFit with several large, realistic datasets and found it to be superior to that of previous methods. We further evaluated our method in two use cases: automating the integration of known composite structures into larger protein complexes and facilitating the fitting of predicted protein domains into volume densities to aid researchers in identifying unknown proteins. We implemented our algorithm as an open-source plugin (github.com/nanovis/DiffFit) in ChimeraX, a leading visualization software in the field. All supplemental materials are available at osf.io/5tx4q.

Index Terms—Scalar field data, algorithms, application-motivated visualization, process/workflow design, life sciences, health, medicine, biology, structural biology, bioinformatics, genomics, cryo-EM.

1 INTRODUCTION

As humans we have been striving for centuries or even millennia to understand, as Faust stated, “*was die Welt im Innersten zusammenhält*

[what binds the world, and guides its course]” [45]. Among other directions of scientific inquiry, this quest applies to the inner workings of the biological world, particularly to how biological processes at tiny scales work and how they keep us alive. In the field of structural biology, researchers have traditionally relied on techniques such as X-ray crystallography or nuclear magnetic resonance spectroscopy to understand the actual molecular composition 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., [6, 21, 29, 36, 41, 44]). The recent Cryo-Electron Microscopy (cryo-EM) approach [28], however, enables researchers to visualize biomolecules in *actual samples* at near-atomic resolution. In addition, over decades, the Protein Data Bank (PDB) initiative has collected thousands of molecular models of the building blocks of cells or organelles studied in structural biology. Researchers are thus on the brink of assembling the molecular composition of actual

- Deng Luo (罗登), Zainab Alsuwaykit (زينب السويكت), Dawar Khan (داور خان), Ondřej Strnad, and Ivan Viola are with the Visual Computing Center at King Abdullah University of Science and Technology (KAUST), Saudi Arabia. E-mail: {deng.luo | zainab.alsuwaykit | dawar.khan | ondrej.strnad | ivan.viola}@kaust.edu.sa.
- Tobias Isenberg is with Université Paris-Saclay, CNRS, Inria, LISN, France. E-mail: given_name.family_name@inria.fr.

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.202x.xxxxxx

samples at the ground-truth level.

To achieve this type of assembly, researchers do not only need to interactively visualize molecular data, for which tools [40] exist, but also to faithfully place 3D models of known molecular building blocks, such as those from PDB data and the AlphaFold predicted library [16], into the captured cryo-EM datasets. Thus far, the fitting process involves a substantial time commitment and numerous manual interventions by domain experts, rendering this process ineffective. The complexity and size of the involved molecules, combined with the variability and noise inherent in cryo-EM data, pose substantial obstacles. In contrast, a fully automatic process is also not ideal because the existence of local minima (wrong placement of compositing proteins) requires domain experts to verify each placement using their knowledge and experience. Fully automated methods are currently far from feasible. Instead, an optimal balance between user interaction and automation is required.

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 one-to-one and many-to-one fitting scenarios, in which multiple protein subunit structures are precisely aligned with a single, large, experimentally reconstructed volume. The DiffFit method is iterative and gradually introduces the source protein structures into the target volumes to assemble the composition of the molecular subunits step by step. By employing advanced strategies such as volume filtering, multiresolution volumes, and negative space utilization we constructed a loss function to quantify the fitting accuracy during the iterations and for the final decision-making. This loss function helps us to iteratively reduce the differences between the two representations—volumetric and atomistic—until we achieve the desired fit. Our visually-guided fitting procedure eliminates the need for domain experts to manually place structures as they assemble the protein structures into the cryo-EM map. It thus significantly accelerates the process into a manageable interactive procedure, delivering precise results for visualizing and analyzing 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;
- a human-in-the-loop strategy providing visual inspection and decision-making in an iterative structure assembly cycle;
- a novel loss function and data processing that calculates new updates in each iteration to expedite algorithm convergence and quantify the fitting accuracy; and
- three use-case scenarios of fitting one or multiple known subunits or identifying yet unknown subunits as part of the molecular assembly.

2 BACKGROUND AND RELATED WORK

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

2.1 Brief background

Structural biology employs various techniques to understand how atoms are arranged in macromolecular complexes, ranging from 60 kDa (i. e., 4,472 atoms; PDB 6NBD [14]) to 50,000 kDa (3,163,608 atoms; PDB 8J07 [46]). These techniques are essential for the study of processes in living cells—cryo-EM being a particularly powerful one [3, 22, 24]. With cryo-EM, bioscientists can capture images of flash-frozen biological specimens using an electron microscope, preserving their natural structure without the interference of staining or fixing [4], which would otherwise interfere with the sample. These images are then used to construct cryo-EM 3D volumes or maps using the *single particle method*, which aligns thousands of projections from structurally identical molecular instances into a single map using the Fourier slice-projection theorem. This map represents 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). As of now, the fitting is typically achieved through manual placement, alignment, and comparison with the density maps. The manual nature of this process makes it time-consuming and tedious, and can only be performed by expert biologists. To address this challenge, numerous approaches have been developed to automate the fitting process, which largely focus on image registration as the foundation and explore methods to streamline 3D model construction, as we review next.

2.2 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, originating from the same or from different modalities, within a shared reference frame [9, 15]. This process involves feature extraction, determining transformations, and assessing accuracy through metrics. Scale-invariant features from images [23], for example, can facilitate matching across a diverse set of 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, to make it possible for them to be matched with new images using fast nearest-neighbor algorithms, with applications including object recognition.

Among the 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 [5, 25] and mutual information [19, 27, 32, 43]. Shang *et al.* [39], for example, 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 [2] uses Kernel PCA and Teaching-Learning-based optimization (TLBO) for multi-modal image registration. In our case, similar to these methods, transformations and alignments have 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 we demonstrate in our work. The major difference to image registration is that, in our workflow, we fit two different data representations, where one is a sub-part of the whole that is potentially present at multiple locations in the target volume.

Model-to-data fitting, which is necessary for cryo-EM data, has also been investigated in depth in computer graphics and pattern recognition [17, 34], with applications in architectural geometry, virtual and augmented reality, robotics, and various other fields [7, 20, 37, 50]—in addition to structural biology. The key challenges in geometric fitting include accuracy, efficiency, robustness, and usability of the fitting module [20, 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.* [34], 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 [38, 52]. Reddy *et al.*'s method [34] 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. Their method [34], 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 built our solution on top of their differentiable compositing.

Another approach, spline surface fitting [37], enhances the smoothness in aircraft engine geometry reconstruction by concurrently approximating point and normal data, ensuring boundary smoothness and optimal convergence, while exploring the effects of norm-like functions on error measurement. A further recently proposed adaptive spline surface fitting method [20], supported by empirical evidence, employs surface meshes for high-precision CAD applications. 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; they thus can serve as a motivation and starting point toward our goals in structural biology. Structural biology data, however, often consists of large, complex structures without regular shapes such as CAD models or easy representations in geometric meshes with smooth surfaces so the aforementioned methods are not directly applicable to our data.

2.3 Fitting in structural biology

Existing fitting methods for structural biology can broadly 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, *UCSF ChimeraX* [30], a popular tool for molecular manipulation and visualization, includes the *fitmap* technique [12]. 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, maximizing the alignment between the atomic model and the density data by optimizing the sum of density values. Similarly, *MarkovFit* [1] is another technique that is often used for placing atomic-level protein structures within cryo-EM maps of moderate to low resolutions. It uses fast Fourier transform (FFT) for the conformational exploration and Markov random fields (MRF) for efficient representation of subunit interactions. Its use of Markov random fields also facilitates the probabilistic assessment of the fitted models. 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) [42, 48, 49]. *A²-Net* by Xu *et al.* [49], 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. 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. A remarkable speed improvement was also demonstrated by Xu *et al.*, being 100× faster to find solutions at runtime than existing methods [8, 47], and achieving a high accuracy of 89.8%. In addition, they introduced the *A²* dataset with 250,000 amino acids in 1,713 cryo-EM density volumes, with a resolution of 3 Å, pioneering automated molecular structure determination training benchmarks.

Another recent method by Mallet *et al.*, *CrAI*, uses machine learning (ML) to find antibodies in cryo-EM densities [26]. The authors

formulate the objective as an object detection problem, using the structural properties of Fabs (Fragment antigen-binding) and VHHs (single-domain antibodies). Furthermore, *DeepTracer* [31] 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. This method employs a customized deep convolutional neural network primarily for the precise prediction of 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 atomistic structures may exhibit geometric issues, local fit-to-map discrepancies, misplaced side chains, or errors in tracing and/or connectivity. All DL-based techniques require a substantial amount of time for training (as opposed to their runtime performance) and 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 2D or 3D maps to facilitate comparison and analysis of spatial structures or features within the maps. *CryoAlign* [13] is a cryo-EM density map alignment method that achieves a 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 [11] yields accurate fitting results in seconds or up to a few minutes. This *ADP_EM* approach is particularly reliable in 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.

3 METHOD

We begin to describe our approach by explaining our process of differentiable structure fitting, before we show how it can be used for visually-guided fitting. After discussing these conceptual aspects, we also briefly discuss implementation details.

3.1 Differentiable structure fitting

Given a cryo-EM map, the domain practitioners—bioscientists—do not know the precise location and orientation parameters that govern where and how a protein's sub-structures fit together. For some regions in the map, the bioscientists may not even know which protein subunits are supposed to be present. Our goal is to develop a new approach that addresses both of these domain tasks. For certain protein subunits, bioscientist are highly confident about their presence in the map. In such cases, our technique will aid in the identification of their respective placement parameters. Second, for the regions with unknown protein subunits, the task of our technique is to identify potential protein subunit candidates from a large database that best fit the cryo-EM map region.

3.1.1 Inspiration and approach overview

We base our approach on the previously mentioned 2D differentiable compositing approach by Reddy *et al.* [34], which discovers pattern structure from wallpaper-like textures containing repetitive patterns made out of elementary patches. We first review their approach, before we describe how we build up our solution on top of their technique. In their case, given a 2D image, which is a composite of multiple small

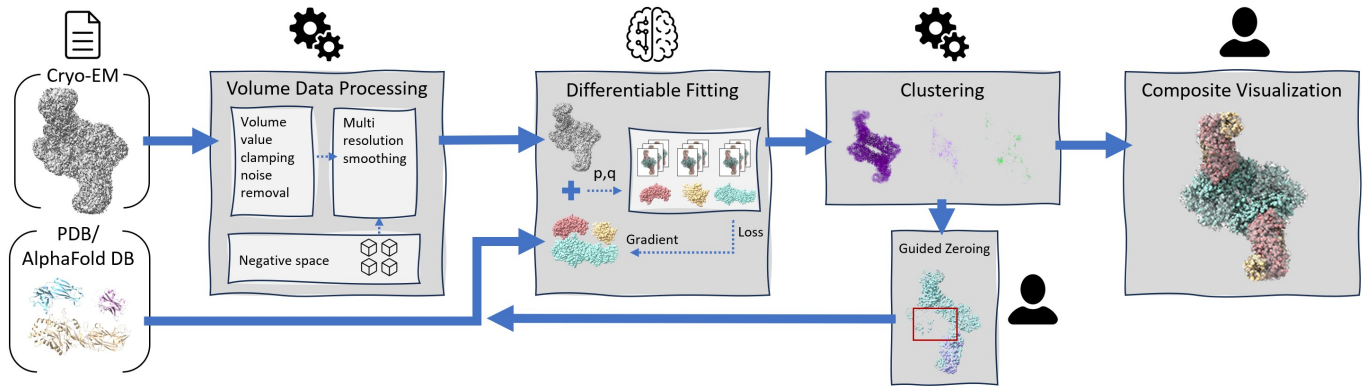


Fig. 2: DiffFit workflow. The target cryo-EM volume and the structures to be fit on the very left serve as inputs, which are passed into the novel volume processing, followed by the differentiable fitting algorithm. The fitting results are then clustered and inspected by the expert. The expert may zero out voxels corresponding to the placed structures and feed the map back iteratively as input for a new fitting, round until the compositing is done.

element patches, the task is to identify the number of occurrences of each patch and the placement parameters for each existing patch. The parameters include the type of patch out of several known patches as well as position, orientation, and depth. Their solution distributes tens to hundreds of patches in the image and uses the differentiable optimization methodology to translate and reorient patterns such that they correspond to the appearance of the patterns in the input wallpaper image. Each single instance E_i (out of total number of ω instances) of a pattern is stored in a layer J_i for each patch instance by sampling from that patch, with the translation, rotation, and patch-pattern type probability taken into account:

$$J_i(\mathbf{x}) = f_i(\mathbf{x}, E_i) = \sum_{j=1}^{\omega} \frac{e^{t_j^i}}{\sum_{k=1}^{\omega} e^{t_k^i}} h_j(R_{\theta_i}^{-1}(\mathbf{x} - \mathbf{c}_i)) \quad (1)$$

where $f_i(\mathbf{x}, E_i)$ is a differentiable function using the expected value over patch-pattern type probabilities stored in a tuple \mathbf{t} representing all ω patch patterns; softmax $e^{t_j^i} / \sum_{k=1}^{\omega} e^{t_k^i}$ over type logits define the patch-pattern type probabilities; h_j is the image patch sampler function; \mathbf{x} is the image location; \mathbf{c}_i is center location of patch element E_i ; and $R_{\theta_i}^{-1}$ is the inverse of a 2×2 matrix rotation with angle θ_i .

The solution image results from compositing of all instances together using f_c compositing function so that each known patch-pattern type is present multiple times with various positional parameters. Patches can overlap other patches, which leads to partial or full occlusion of a certain pattern. This is characterized by $v_i(\mathbf{x}), v \in \{0, 1\}$, which is the visibility of layer i at image location \mathbf{x} :

$$I(\mathbf{x}) = f_c(\{J_i(\mathbf{x})\}_i) = \sum_{i=0}^{\omega} J_i(\mathbf{x}) v_i(\mathbf{x}) \quad (2)$$

This solution image is compared with the input image in an optimization, where the parameters of all patch instances are updated in every iteration. The solution image then becomes increasingly similar to the input image. Reddy *et al.* define the L^2 distance loss L_d for the optimizer as:

$$L_d(A, I) = \frac{1}{P} \sum_{p=1}^P \|A(\mathbf{x}_p) - I(\mathbf{x}_p)\|_2^2 \quad (3)$$

where the sum is over the number of all pixels P in the image. A is the input image and I is the composited solution image. The optimal elements \mathcal{E}^* are then found by minimizing loss L_d over the entire set of elements $\mathcal{E} = \{E_0, \dots, E_{\omega}\}$:

$$\mathcal{E}^* = \arg \min_{\mathcal{E}} L_d(A, f_c(\mathcal{E})). \quad (4)$$

Reddy *et al.*'s patch-pattern fitting problem is similar to ours in the sense that in both cases we are compositing element instances into a scene. But Reddy *et al.*'s differentiable compositing approach cannot be directly applied to the structural biology domain to solve the protein fitting problem for the following reasons:

1. the pattern image and 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 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.

Initially, we attempted to align our problem better with differentiable compositing by first *simulating* a cryo-EM map from the atomistic point cloud of the protein model and then fitting the simulated map to a target map. That way the representational discrepancy (see reason (2) above) is eliminated. That approach, however, was only successful for trivial cases, while for real-world scenarios it frequently fell into local minima. To illustrate this point, we provide several exemplary volume-only based molecular fitting videos for interested readers in our supplementary material at osf.io/5tx4q.

Driven by the successful fitting cases from many experiments, we gradually built several novel strategies on top of the differentiable compositing that effectively address the problem of molecular structure fitting. Most notably, we address the substantially higher complexity of our scenario based on the knowledge, experience, and deep insight of the target audience: the bioscientists. Our solution is thus based on a human-in-the-loop strategy and we propose a fast and robust visual analytics approach, DiffFit, with two main steps: (1) an automated excessive molecular fitting and (2) a visual inspection and filtering of the fitted results by the bioscientists. Both consecutive steps are building blocks of a visual analytics feedback loop, in which multiple proteins are iteratively composited to fit the underlying cryo-EM volume.

We schematically present our DiffFit workflow in Fig. 2. First, we seed an excess amount of all compositing molecules in the volume scene. If this simultaneous fitting all molecules exceeds the available GPU memory, we sort the molecules by atom count and partition them into batches. Then, we fit the batches of molecules within several iterations in descending atom-count order. This fitting relies on a novel loss function that calculates the average density value from the densities that we sample for each atom. The differentiable property of our fitting scenario allows us to optimize based on gradient-descent. In addition, we associate each fit with a numerical value that characterizes the fit quality. For this purpose, we create a simulated cryo-EM map of each fitted molecule and calculate the correlation of the simulated densities using the real cryo-EM map densities. Then, we collect the fitting results and cluster them based on positional and orientational parameters. For each cluster, we select one representative fit—the fit with the highest correlation value. Then, we sort the clusters by their repre-

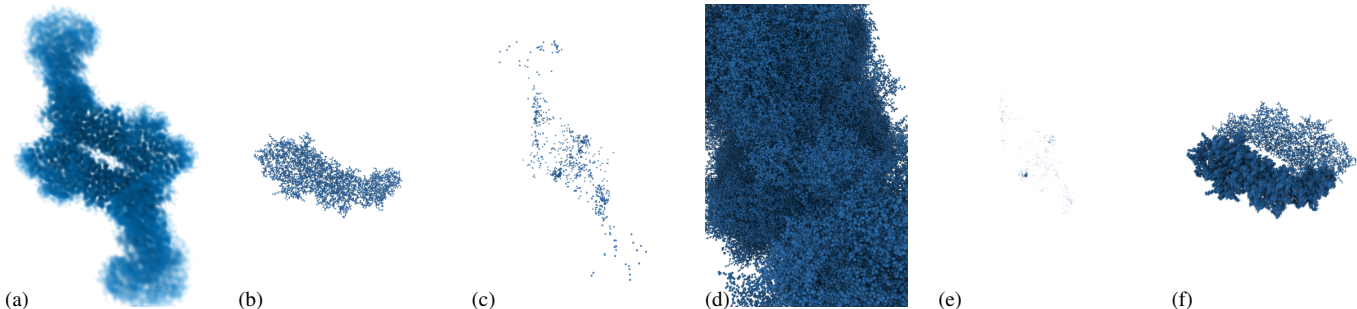


Fig. 3: Clustering and filtering: (a) target volume, (b) atom coordinates of the source structure, (c) positions of 1000 fit results (the dots are clustered, hiding the number of results), (d) 1000 instances of the structure, (e) positions of 1000 fit results with a transparency level set based on an exponential scaling of the correlation metric (two clusters stand out), and (f) instances of the structure at those two clusters.

sentative correlations and interactively visualize them in ChimeraX to allow the bioscientists to inspect the solutions. Once they verify a given molecular placement, we disable molecule placements in the respective regions in the following iterations by setting the voxels covered by the molecular fit in the cryo-EM map volume to zero. We thus gradually erase the successful placements from the map, forcing the following placements to search for a fit in non-zero volume locations. Once the feedback update in the map is completed, we perform the next fitting iteration with another molecular structure. We repeat this workflow pattern until the map has nearly all voxels zeroed out, and the entire sub-unit placement of the complex molecular structure is complete. Below we introduce the details of DiffFit, in the following order: (1) sample one coordinate, (2) fit one placement of one molecule, (3) fit multiple placements of one molecule, and (4) fit multiple placements of multiple molecules.

3.1.2 Sampling of one coordinate

Because our task is to determine the optimal alignment of an atomistic molecular structure to the reconstructed cryo-EM volume map, we determine the optimal fit characterized by two rigid-body transformation parameters: a translational offset \mathbf{p} and a rotation. We represent the rotation 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 in the molecular subunit. To calculate 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 sample a density value D of the atom to be placed at position $T(\mathbf{x}_i)$ from a scalar volume V using trilinear interpolation as follows:

$$D(T(\mathbf{x}_i)) = S(T(\mathbf{x}_i), V).$$

3.1.3 Placement of one molecule (or one subunit)

We formulate an initial loss function L to determine the best \mathbf{p} and \mathbf{q} parameters. This function 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 $T(\mathbf{x}_i) \in \mathbf{X}_m$ for a particular molecular subunit m :

$$L(\mathbf{p}, \mathbf{q}, \mathbf{X}_m, V) = - \left(\frac{1}{N} \sum_i^N D(T(\mathbf{x}_i)) \right) = - \frac{1}{N} \sum_i^N S(M_{\mathbf{q}} \cdot \mathbf{x}_i + \mathbf{p}, V). \quad (5)$$

We rely on the calculation of the gradient of the differentiable formulation and use it with the Adam optimizer [18] for the optimization. Although the Adam optimizer is known for robustness with respect to local minima, our initial loss function formulation frequently leads to a local minimum (i.e., a place that is not an optimal placement for the molecule in the map but from which the optimizer cannot find a better solution in the parameter-space neighborhood). Such local minima are a common and severe problem that also manifests in the functionality of the most commonly used tools for molecular subunit fitting (e.g., the fit-in-map feature in ChimeraX). We thus introduce several strategies to form a novel loss function, making DiffFit more robust.

The first strategy that we found to substantially contribute to a good fitting performance is **filtering 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 less than a given threshold and set them to zero. The size of the connected voxel cluster after thresholding must be greater 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 focus and efficiency of the algorithm. Then, we normalize the filtered values to $[0, 1]$ —a typical practice in learning and optimization approaches—, which leads to a volume \hat{V}_F that 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 in biological datasets, we apply a series of convolution iterations to the target volume, and capture each smoothing result 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. Empirically, we found that a 10-element array of increasingly smoothed volumes performs well, iteratively filtered with a Gaussian smoothing kernel. We expose the size of this array 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 in the cryo-EM map volume but outside the extent of the molecular target structure (i.e., outside of the target *footprint*), the target density would normally be zero. To discourage such misalignment even further, we assign these regions a **negative value**. After smoothing, for 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. We noticed that a constant value of -0.5 outside the molecular footprint in the map performs well. We expose this value as a tunable hyperparameter. We denote the resulting volume as $\hat{V}_{F_{-c}}^{G_n}$, where $-c$ is 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 multiresolution volume array element w_j for n resolutions, and sum up all the multiresolution components to form the final loss function for one \mathbf{p} and \mathbf{q} pair:

$$L_m([\mathbf{p}, \mathbf{q}]) = \sum_{j=1}^n w_j \cdot L(\mathbf{p}, \mathbf{q}, \mathbf{X}_m, \hat{V}_{F_{-c}}^{G_j}). \quad (6)$$

To start the optimization, we need to initialize the position offset \mathbf{p} and the rotation quaternion \mathbf{q} . We uniformly sample N_q points on a unit sphere and then convert these into quaternions to be applied

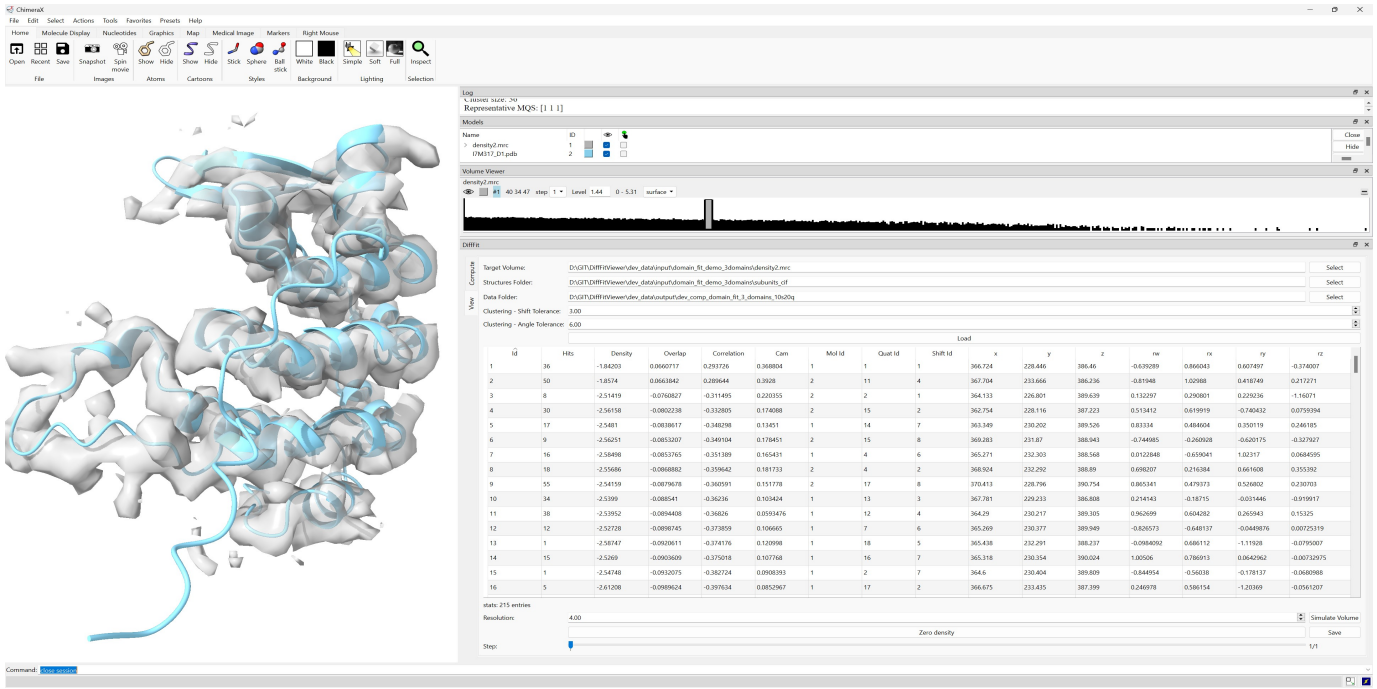


Fig. 4: Visual browser based on ChimeraX. The target volume on the left is overlaid with a fitted molecule corresponding to the selected fit result in the table on the right (clustered fits, each row is the representative placement with the highest correlation from that cluster). After inspection, users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

for each offseted position. Instead of uniformly sampling positions from the volume bounding box (as in ChimeraX), we uniformly sample N_p positions from the positive voxels in the filtered and normalized volume \hat{V}_{F-c} . This **enveloped sampling based initialization** increases the success rate by a factor of two by searching from $N_q \cdot N_p$ initial placements, compared to the traditional initialization in ChimeraX.

3.1.4 Fitting multiple placements of one molecule

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

$$L_{par}(m) = \sum_{k=1}^{N_q \cdot N_p} L_m(\{\mathbf{p}_k, \mathbf{q}_k\}). \quad (7)$$

3.1.5 Fitting multiple placements of multiple molecules

Finally, all subunit molecules have different numbers of atoms; it is thus not easy to parallelize the treatment of multiple molecules without overhead on the array padding of zeros. Usually, the $N_q \cdot N_p$ initial placements of \mathbf{X}_m atoms would result in a total number of sampling operations higher than the total number of GPU threads; therefore, we process different subunits molecules sequentially in a for loop and form an overall loss function for M molecules as follows:

$$L_{all} = \sum_{l=1}^M L_{par}(l). \quad (8)$$

3.1.6 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 atoms in a molecule, we can form two vectors, a sampled density vector $\mathbf{D} = [D(\mathbf{x}_1), D(\mathbf{x}_2), \dots, D(\mathbf{x}_N)]$ from the target volume and a weight vector $\mathbf{W} = [W(\mathbf{x}_1), W(\mathbf{x}_2), \dots, W(\mathbf{x}_N)]$ from the simulated volume. Then, we can calculate three alignment metrics, the mean overlap μ , correlation ρ , and the correlation about the mean ρ_μ as:

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

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

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

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 next.

3.2 Visually-guided fitting

A critical aspect of the post-processing of DiffFit involves the clustering and sorting of the 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 (Fig. 3(c), (e)).

Each cluster represents a set of closely related fits, suggesting a consensus among them regarding the position and orientation of the fitted structure in 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 the experts in exploring the fitting results, we created an interactive visual browser as a comprehensive visualization tool to present the sorted clusters in a user-friendly format (Fig. 4). In the browser we display key metrics for each cluster, including the average correlation coefficient, the density overlap, and the consensus error measures, which provide a quick overview of the quality and relevance of each cluster. The browser also allows the biologists 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 could substantially impact the biological interpretation of the results.

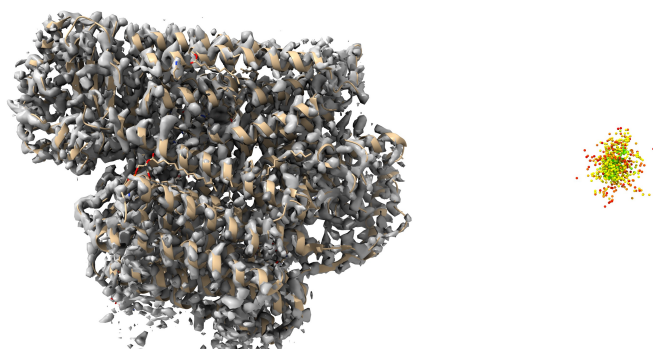


Fig. 5: Interactive visual fit assessment tools we implemented in ChimeraX. Left: visualization of the best fit of the selected cluster of fits in the context of the cryo-EM volume map; right: abstracted summary of all clusters, in which each sphere represents the center of the molecule placement and the color represents the quality of the fit. Both views are interactive and move in sync, just being offset from each other in the screen’s x -direction (as they both represent the same spatial 3D space).

To help the experts to gain a spatial overview of the fitting result in the tabular data (Fig. 4), we further provide means to visualize the clusters in the viewport (Fig. 5). First, we display the representative molecule of each cluster of fits in the context of the cryo-EM volume map, as we show in Fig. 5 on the left. The displayed fit can be controlled by selecting one from the list of fits. To also provide the experts with a visual summary of all fits, we also abstractly represent each fit cluster with the help of a sphere that we place at the center of the cluster’s representative molecule as it is transformed by the fit shift and rotation (Fig. 5, right). We assign each sphere a color based on the clusters’ average density order and the experts can also use the spheres to select a different fit to be shown in the context of the cryo-EM volume map.

An innovative feature of our approach is the ability to refine the fitting process iteratively by selectively excluding already placed molecules densities. Once a bioscientist selects a cluster and verifies its fit (or multiple fits) as accurate, we can zero out the corresponding density in the target volume, thus effectively removing the respective volume region from further consideration in the following part of the fitting process. For this purpose we, first, use the fit structure to simulate a map; then we use the voxel positions from the original map to sample the density values from the simulated map. If the sampled density is higher than a user-specified threshold, we set the original voxel’s density to zero. By default we use a threshold of 0, which usually delivers good results. 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.

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

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 a 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.

4 IMPLEMENTATION

We utilized PyTorch for the DiffFit algorithm implementation, capitalizing on its dynamic computational graph, automatic differentiation, and GPU acceleration to estimate positional offset and rotational quaternion parameters efficiently. Specifically, we employed `torch.nn.Conv3d` for Gaussian smoothing

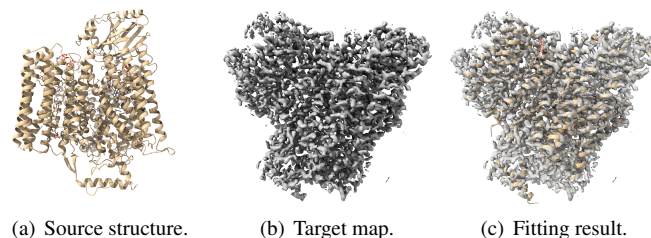


Fig. 6: Fitting a single structure for 6WT1.

and `torch.nn.functional.grid_sample` for trilinear interpolation with padding mode set to “border.” This approach enables us to rapidly process large volumes and multiple structures. We leveraged tensor operations and the Adam optimization algorithm for accurate optimization results. We also integrated additional functions from SciPy, Bio.PDB, `mrcfile`, and Numpy libraries to enhance the algorithm’s functionality. In addition, we integrated DiffFit seamlessly into ChimeraX (Fig. 4), based on its bundle development environment.

5 USE CASE SCENARIOS

We designed DiffFit with its advanced fitting algorithms and integration with visualization tools to address a range of challenges in structural biology. To be able to illustrate their power, we now explore three key scenarios where DiffFit can be particularly effective, demonstrating its versatility and potential.

5.1 Scenario 1: Fit a single structure

The most straightforward application of DiffFit is the fitting of a single atomistic structure to a target volumetric map, as we illustrate in Fig. 6. This scenario is common in cases in which an already-resolved protein structure or a predicted structure needs to be placed in a newly reconstructed volume captured via cryo-EM for further refinement.

In this scenario, DiffFit efficiently determines the optimal position and orientation of the protein (Fig. 6(a)) in the volume (Fig. 6(b)). The entire fitting process is automatic, removing the prerequisite of manually placing the structure at an approximate orientation close to the final optimal one. The interactive visually guided inspection process allows researchers to verify the fit (Fig. 6(c)), who apply their expertise to ensure biological relevance and accuracy.

We use the dataset reported in the recent MarkovFit work [1] to benchmark DiffFit’s performance in this scenario, compare it with ChimeraX and MarkovFit, and report the results (successful hit rate, computation time, and root-mean-square deviation, RMSD) in Table 1 (we provide a complete table that includes each structure’s EMD ID, the number of chains, the number of atoms, voxel size, and the used surface level threshold used in Table 3 in Appendix A). For each structure, we performed five experiment runs to obtain reliable results (we also attach each individual run’s metrics in our supplementary material). For each run, we fit $1000\times$ to perform the search. For ChimeraX, we ran the command “fit #1 in #2 search 1000” to perform atom-to-map searching, which is about one-fold faster than the atom-simulated map-to-map fitting and delivers similar results. We regard the number of fits in the top-ranked cluster as the hit rate if the representative fit of that cluster is within 3 Angstroms and 6 degrees (ChimeraX’s default threshold) from the ground truth. We do not repeat the MarkovFit computation as it takes an average of “7.7” plus “6.25” hours (as stated in prior work [1]) to finish a single run for each structure. Instead, we directly take the author-reported [1] 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 to in advance to get the best model. DiffFit significantly outperforms ChimeraX on the hit rate (on average, a $26.9\times$ gain) and the computation time ($26.0\times$ gain). The RMSD of DiffFit is also significantly better than MarkovFit but often slightly worse than ChimeraX’s. However, this can always be corrected (the last column in the table) by a single automatic fit using ChimeraX’s atom-to-map fitting, which results in a better RMSD on average. Compared with the overall averaged metrics, for high-resolution maps, DiffFit gets

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 [1], DC stands for DiffFit corrected by a single automatic ChimeraX fit; G stands for Gain and is D/C for Hit and C/D for Computing time (in seconds). High-avg stands for the averaged metrics for the high-resolution maps, and Med for medium maps, All for all maps.

PDB	Res	Hit rate			Computing time			RMSD (Å)			
		C	D	G	C	D	G	M	C	D	DC
6WTI	2.38	0.0	122.2	n/a	150.3	7.1	21.2	1.310	n/a	0.939	0.038
7D8X	2.60	0.0	77.6	n/a	196.0	31.1	6.3	1.960	n/a	1.126	0.014
7SP8	2.70	4.6	238.8	51.9	130.6	12.2	10.7	1.290	0.996	0.779	0.024
7STE	2.73	14.0	248.2	17.7	806.1	41.1	19.6	1.740	0.062	0.737	0.057
7JPO	3.20	5.4	203.2	37.6	250.7	25.6	9.8	2.540	0.017	0.896	0.015
7PM0	3.60	44.0	195.2	4.4	352.4	7.0	50.3	1.640	0.030	0.976	0.028
6M5U	3.80	0.0	277.0	n/a	162.2	18.8	8.6	2.360	n/a	0.944	0.014
6MEO	3.90	7.4	214.6	29.0	128.2	9.4	13.6	1.940	0.489	0.779	0.488
7MGE	3.94	4.8	239.0	49.8	337.6	11.4	29.5	1.870	0.017	0.828	0.017
High-avg		8.9	201.7	31.7	279.3	18.1	18.8	1.850	0.268	0.889	0.077
5NL2	6.60	1.8	179.6	99.8	94.6	9.6	9.9	2.440	0.093	1.093	0.047
7K2V	6.60	49.0	170.4	3.5	240.6	8.6	28.1	25.290	0.338	1.532	0.337
7CA5	7.60	55.8	158.2	2.8	322.6	4.5	71.1	3.290	2.042	1.447	2.043
5VH9	7.70	68.6	293.8	4.3	1147.8	20.3	56.6	0.960	0.085	0.652	0.084
6AR6	9.00	78.0	370.8	4.8	74.9	2.2	33.6	2.200	0.123	2.067	0.124
3J1Z	13.00	138.6	441.6	3.2	64.4	3.0	21.5	32.330	0.396	2.436	0.400
Med-avg		65.3	269.1	19.7	324.1	8.0	36.8	11.085	0.513	1.538	0.506
All-avg		31.5	228.7	26.9	297.3	14.1	26.0	5.544	0.366	1.149	0.249

higher gains on hit rate; for medium-resolution maps DiffFit gets higher gains on the computation time. Of note is that, with expert knowledge, ChimeraX’s performance can be boosted. For example, after smoothing the map via a command similar to “`vol gaussian #2 sd 2,`” the hit rate goes up to 1 (6WTI), 67 (7D8X), and 12 (6M5U). We can also achieve a similar boosting with DiffFit. As this boosting highly depends on the user’s knowledge, we report only the simple version of the fitting where the user only needs to specify the surface-level threshold (which, in most cases, is ChimeraX’ built-in heuristic: the top 1% percentile of all the density values). We computed the reported performance metrics on a workstation that uses an 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)).

5.2 Scenario 2: Composite multiple structures

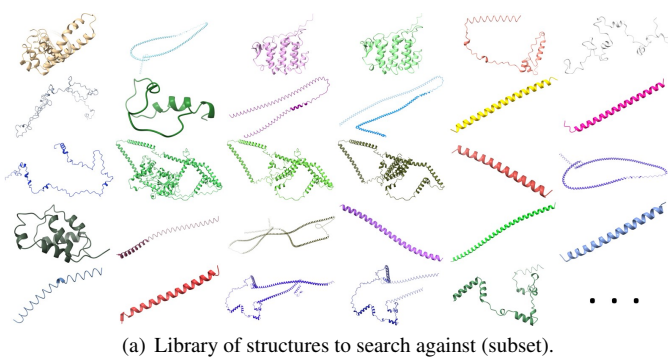
A more complex use case involves the fitting of multiple structures to a single, large, often complex, volumetric dataset, such as assembling a viral capsid from individual protein units or constructing a ribosomal complex from its constituent proteins and RNA molecules. DiffFit can handle such composite fitting tasks by iteratively optimizing the placement of each component, from largest to smallest, while considering the spatial relationships and interactions between them to prevent overlaps and ensure a coherent assembly.

In Fig. 1 we show an example of compositing multiple structures into a cryo-EM volume map, for the PDB-protein 8SMK [51]. In the first row we demonstrate how the middle and bottom parts are fitted in the first computation round, with the remaining top part of the protein being fitted in the second round of the interactive process.

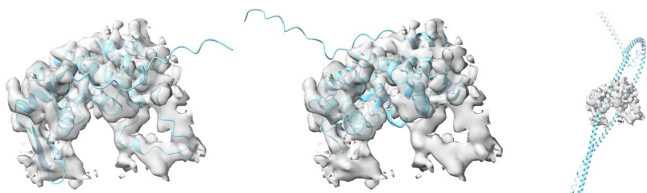
The ability to *zero-out* densities, once a fit is interactively confirmed by the expert, allows DiffFit to fit multiple structures sequentially without interference from previously placed components, as we demonstrate in Fig. 1. This iterative approach is particularly useful for densely packed molecular complexes, where individual components may be difficult to distinguish in the volumetric data. This scenario is critical for understanding the functional context of proteins in larger biomolecular assemblies or cellular environments.

5.3 Scenario 3: Identify unknown densities

DiffFit also offers bioscientists the capability to identify and characterize unknown densities within volumetric datasets. In cases where a volume contains unassigned or ambiguous regions, possibly indicating the presence of previously unidentified molecules or molecular com-



(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.

Fig. 7: Unknown density identification, where dozens to hundreds of molecular structures can be evaluated for potential fit.

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

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

plexes, DiffFit can be used to screen a library of known structures and predicted structures for potential matches.

By fitting the structures from a library to the unidentified densities and evaluating the fit quality, researchers can hypothesize the identity of the unknown components. This scenario is invaluable for discovery-based research, where identifying novel components in complex molecular assemblies could lead to significant biological insights.

We performed a search with the demonstration dataset from a recent automated domain-level protein identification technique, DomainFit [10]. The task was to identify, from a library of 359 protein domains (Fig. 7(a) shows a subset), which domains best fits into the given volume (Fig. 7(b)). DomainFit identified two candidates (I7MLV6_D3, and I7M317_D1), after first fitting all domains using the `fi tmap` command in ChimeraX and then performing statistical analyses to remove false positives. With our DiffFit, we identified these same two candidates with our fitting technique followed by a visual inspection, where the two candidates were the first to be inspected. We report the hit rates in Table 2, which shows a $\approx 2\times$ gain on the hit rate. In addition, DomainFit takes ≈ 12 hours to finish the task, while with DiffFit it takes ≈ 10 minutes—a $72\times$ gain in computation time.

6 FEEDBACK

In addition to these quantitative metrics, qualitative feedback from users plays a crucial role in evaluating the practical utility and user experience of DiffFit. We thus solicited feedback from a diverse group of users, including PhD students, structural biologists, and computational scientists; through surveys, interviews, and hands-on testing sessions.

Specifically, we emailed cryo-EM practitioners at our university to test our tool. We performed three Zoom-based demo sessions and one in-person demo and feedback session with those people who expressed interest. The first session involved five people from a research group that studies structural biology and engineering (one senior PhD student specializing in structural and computational biology, one senior PhD student in experimental biology about protein structures, one senior PhD student in experiments elucidating the structure and function of

proteins, one postdoc in experimental biology about protein structures involved in cell signaling, one research scientist in using Cryo-EM and other techniques to study protein structures). The second session involved two people from a research group focusing on using cryo-EM to study the structural biology of human DNA replication and repair (an assistant professor and his postdoc). The third session involved one junior PhD student using cryo-EM to solve protein structures. The final in-person session involved one research scientist who provides cryo-EM service as a platform to the whole university and builds a bridge between the microscope and users at all levels. Apart from verbal feedback during and after these demo sessions, one participant in the first group, as well as the participant in the in-person session, also sent us written feedback, which we include in [Appendix B](#). In total, nine people provided qualitative feedback, which we report next.

Usability. The participants appreciated DiffFit’s intuitive interface and workflow (“quite intuitive and easy to use”) and the direct integration into ChimeraX with its known UI, which significantly lowers the barrier to entry, in particular, for new users, while still providing advanced features for experienced researchers. In particular the easy one-step fitting of a PDB file into a cryo-EM map was highlighted. The visually guided fitting process was appreciated as a powerful feature for refining fitting results based on expert judgment. Suggestions for further improvements of this interaction step and the UI in general were to display several visual clusters at a time, to visually compare the fit results, and to offer a tab-based method to check the various clusters. This approach would make it easier to observe how each cluster fits relative to the cryo-EM volume; thus, researchers would be better able to focus on a chain of interest. We have addressed this point already with our visual fit assessment tools ([Fig. 5](#) and [Sect. 3.2](#)). One issue we noticed is that some experts commented on the computation times in the order of a few minutes, which points to the fact that they did not have a CUDA-compatible GPU at their disposal. As we demonstrated with our runtime analysis in [Sect. 5.1](#), with such affordable hardware, the computation is possible in well under a minute for many structures, so cryo-EM labs can break free from extremely long computation times or the need for employing large (and expensive) computation resources.

Impact on Research. The participants reported that DiffFit has the potential to have a tangible impact on their research, enabling them to address complex fitting challenges that were previously out of reach. They thus get the ability to rapidly sample a large database of candidate structures for regions with unassigned protein density. The new ability to fit structures rapidly and accurately into volumetric data offers new directions of investigation and has the potential to accelerate discoveries in structural biology. One respondent stated that our automatic fitting and visual inspection approach “could be a key feature in ChimeraX that [could become] a standard in many pipelines” as well as “a key implementation [for] a standard modeling workflow.”

The respondents also made suggestions for future developments such as better structuring the handling of the associated files, further exploring the design space of the abstract cluster representations, a detailed protocol for downloading and installing the tool (already implemented in our GitHub repository), as well as adding workflows that would require automatically creating model subdivisions and performing fits on those before integrating the results into the reference volume. They also highlighted further potential uses, such as employing our approach to analyze density maps generated using X-ray crystallography.

7 LIMITATIONS AND FUTURE WORK

Although our DiffFit technique demonstrates a substantial improvement over state-of-the-art techniques, it also has some limitations. One issue arises in Scenario 2, due to the gradual *zeroing-out* of the target map, as multiple structures are being composited onto a single map. In this case, the volume removal can potentially remove some voxels that are also part of adjacent interacting molecular subunits, especially in those cases when the chosen fit is not exactly perfect. Then, the subsequent subunit to be placed may have a higher difficulty of finding its correct location because these few zeroed voxels penalize the correct position. We are currently investigating this limitation and plan to experiment with heuristics that remove only voxels that are fully covered by the

molecular structure or slightly shrink (by using a higher user-specified threshold) the *to-be-zeroed-out* region before removing it from the map. Another related limitation is that the fitting process does not check for collisions in adjacent subunits, as previously investigated in visualization work (e. g., [\[35\]](#)). While the map shrinking approach should address this issue, subunits may still overlap in some cases. Resolving this issue is straightforward: we could remove those subunit poses that overlap with previously confirmed subunit placements or could minimally modify the pose to resolve the structural collisions.

Our technique is also not entirely parameter-free. The threshold value for removal of voxels with small density values, e. g., and the threshold value for the connected-voxel minimal cluster size have to be manually set. Manipulating these parameters requires prior domain experience with cryo-EM data from the bioscientists. We plan to automate the identification of these thresholds or, at least, define suitable default values and provide guidance on reasonable parameter ranges.

The visualization design in our current version of the tool also offers only the most essential visual encoding types, with a clear potential for further improvement. We plan to adapt comparative visualization techniques for intersecting surfaces and smart visibility techniques for combined rendering of the volumetric density representation with molecular surfaces or cartoon representations. Furthermore, the visualization design for analyzing the structural poses in a cluster suffers from high spatial occlusion among the subunits’ various poses (see [Fig. 3\(d\)](#)). This is a common domain problem and requires a dedicated research effort to combat the occlusion of this magnitude. Our tool can also be further extended in future work to, for instance, change the color encoding of the fitting results presentation to use one of the other alignment metrics ([Sect. 3.1.6](#)) or to change the size of the spheres in the results summary to make larger clusters appear more prominent.

Driven by the feedback we received, additional possible future research directions include extending DiffFit to deformable transformations, considering stoichiometry and symmetry, and speeding it up even further through an integration with CUDA.

8 CONCLUSION

In conclusion, DiffFit offers a novel and efficient solution to the challenges of atom-to-map fitting that has arisen in structural biology. We address these challenges with our differentiable fitting along with a set of essential strategies that make our algorithm robust for the domain problem. Our approach adds a human-in-the-loop visual analytics approach to the workflow and provides an open-source package designed to work as part of a standard software tool on the domain (ChimeraX). A key element of our approach is the change from pixel-to-pixel (or the equivalent, voxel-to-voxel) fitting to point-to-volume fitting, which enabled us to deal with the specific constraints of the fitting problem in structural biology. In hindsight, as we witness the success of using atom coordinates to sample the volume instead of using all the voxels, DiffFit suggests that, in Reddy’s [\[34\]](#) original technique that inspired our work, maybe not all the pixels in a patch are needed for the optimization, which means that the image compositing optimization could be more efficient and could handle larger examples with faster processing.

The quantitative performance metrics we reported collectively demonstrate DiffFit’s substantial improvement over traditional approaches, highlighting its potential to transform the field of structural biology. This potential is particularly large because the workflows of Assembline, a protocol published in 2022 [\[33\]](#), and DomainFit [\[10\]](#), so far, rely on ChimeraX’s `fitmap` command as their first step. This stage can now be replaced by our more effective DiffFit, laying a new foundation for these and other workflows. By demonstrating the effectiveness of our technique across three scenarios, from fitting individual structures to assembling complex molecular architectures and identifying unknown components, we demonstrated that DiffFit overcomes the limitations of manually placing individual molecules before fitting, makes the compositing faster, more accurate, and intuitive, and opens the possibility of scanning the whole set of known and predicted molecules with the current computational resources. Ultimately, we thus escape the *Faustian bargain* [\[45\]](#) and instead are now free ourselves to explore the inner workings of the biological world.

ACKNOWLEDGMENTS

This research was supported by King Abdullah University of Science and Technology (KAUST) (BAS/1/1680-01-01). We thank Tom Goddard for his ChimeraX development support, all evaluators for providing their feedback, Nagarajan Kathiresan for his supercomputer support, and Rahman Hasan for his hardware support. D. Luo thanks Todd Pietruszka, Irfan Anjum, Rand Qoj, Xinwei Qiu, Yonghong Leng, Yujing Ouyang, Ziyun Zhang, and Xiaode Zhu for their kind support.

SUPPLEMENTAL MATERIAL POINTERS

We provide the open-source repository of our tool at github.com/nanovis/DiffFit. Other supplemental material is available at osf.io/5tx4q.

IMAGES LICENSE AND COPYRIGHT

We as authors state that all of our figures are and remain under our own personal copyright, with the permission to be used here. We also make them available under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) license and share them at osf.io/5tx4q.

REFERENCES

- [1] 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*, 9, article no. 935411, 13 pages, 2022. doi: [10/gtmxvj](https://doi.org/10/gtmxvj)
- [2] P. Arora, R. Mehta, and R. Ahuja. An integration of meta-heuristic approach utilizing kernel principal component analysis for multimodal medical image registration. *Cluster Comput*, 24 pages, 2024. To appear. doi: [10/gtmxvd](https://doi.org/10/gtmxvd)
- [3] X.-c. Bai, G. McMullan, and S. H. W. Scheres. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci*, 40(1):49–57, 2015. doi: [10/f6wq7v](https://doi.org/10/f6wq7v)
- [4] W. Chen, X. Wang, and Y. Wang. FFF: Fragment-guided flexible fitting for building complete protein structures. In *Proc. CVPR*, pp. 19776–19785. IEEE Computer Society, Los Alamitos, 2023. doi: [10/gtmxvf](https://doi.org/10/gtmxvf)
- [5] 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, pp. 489–490. IEEE, Piscataway, 1995. doi: [10/bgp7mt](https://doi.org/10/bgp7mt)
- [6] D. Duran, P. Hermosilla, T. Ropinski, B. Kozlíková, Á. Vinacia, and P.-P. Vázquez. Visualization of large molecular trajectories. *IEEE Trans Vis Comput Graph*, 25(1):987–996, 2019. doi: [10/gjbdnk](https://doi.org/10/gjbdnk)
- [7] 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*, 30(8):5063–5073, 2024. doi: [10/gsc3c9](https://doi.org/10/gsc3c9)
- [8] B. Frenz, A. C. Walls, E. H. Egelman, D. Veasler, and F. DiMaio. RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat. Methods*, 14(8):797–800, 2017. doi: [10/gf5cmq](https://doi.org/10/gf5cmq)
- [9] 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*, 65(20), article no. 20TR01, 27 pages, 2020. doi: [10/ghn45h](https://doi.org/10/ghn45h)
- [10] 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](https://doi.org/10/gs63f2)
- [11] J. I. Garzón, J. Kovacs, R. Abagyan, and P. Chacón. ADP_EM: Fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinf*, 23(4):427–433, 2006. doi: [10/c9xkmg](https://doi.org/10/c9xkmg)
- [12] T. D. Goddard, C. C. Huang, and T. E. Ferrin. Visualizing density maps with UCSF Chimera. *J Struct Biol*, 157(1):281–287, 2007. Software tools for macromolecular microscopy. doi: [10/ft849s](https://doi.org/10/ft849s)
- [13] 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*, 15, article no. 1593, 15 pages, 2024. doi: [10/gtmxvg](https://doi.org/10/gtmxvg)
- [14] M. A. Herzik, M. Wu, and G. C. Lander. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat Commun*, 10, article no. 1032, 9 pages, 2019. doi: [10/gg7zpq](https://doi.org/10/gg7zpq)
- [15] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys Med Biol*, 46(3):R1–R45, 2001. doi: [10/fgbhgy](https://doi.org/10/fgbhgy)
- [16] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021. doi: [10/gk7nfp](https://doi.org/10/gk7nfp)
- [17] M. Kesäniemi and K. Virtanen. Direct least square fitting of hyperellipsoids. *IEEE Trans Pattern Anal Mach Intell*, 40(1):63–76, 2018. doi: [10/gcqrjb](https://doi.org/10/gcqrjb)
- [18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint 1412.6980, 2017. doi: [10/hnkr](https://doi.org/10/hnkr)
- [19] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever. Adaptive stochastic gradient descent optimisation for image registration. *Int J Comput Vis*, 81(3):227–239, 2009. doi: [10/dghjcp](https://doi.org/10/dghjcp)
- [20] 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*, 13 pages, 2024. To appear. doi: [10/gtmxw4](https://doi.org/10/gtmxw4)
- [21] D. Kuřák, M. N. Selzer, J. Byška, M. L. Ganuza, I. Barišić, B. Kozlíková, and H. Miao. Vivern—A virtual environment for multiscale visualization and modeling of DNA nanostructures. *IEEE Trans Vis Comput Graph*, 28(12):4825–4838, 2022. doi: [10/mpwt](https://doi.org/10/mpwt)
- [22] 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*, 4(1):28–39, 2007. doi: [10/ccbr4m](https://doi.org/10/ccbr4m)
- [23] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision*, 60(2):91–110, 2004. doi: [10/bqrmsp](https://doi.org/10/bqrmsp)
- [24] S. Malhotra, S. Träger, M. Dal Peraro, and M. Topf. Modelling structures in cryo-EM maps. *Curr Opin Struct Biol*, 58:105–114, 2019. doi: [10/gtmxvc](https://doi.org/10/gtmxvc)
- [25] 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*, 26(4):774–785, 2013. doi: [10/gtmxw5](https://doi.org/10/gtmxw5)
- [26] 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/gtm9kj](https://doi.org/10/gtm9kj)
- [27] 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*, 22(1):120–128, 2003. doi: [10/cs4p4ch](https://doi.org/10/cs4p4ch)
- [28] T. Nakane, A. Kotecha, A. Sente, G. J. McMullan, S. Masiulis, P. M. G. E. Brown, I. Grigoras, L. Malinauskaitė, T. Malinauskas, J. Miehling, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. N. Murshudov, A. R. Aricescu, and S. H. W. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, 2020. doi: [10/gjwcfp](https://doi.org/10/gjwcfp)
- [29] N. Nguyen, O. Strnad, T. Klein, D. Luo, R. Alharbi, P. Wonka, M. Maritan, P. Mindek, L. Autin, D. S. Goodsell, and I. Viola. Modeling in the time of COVID-19: Statistical and rule-based mesoscale models. *IEEE Trans Vis Comput Graph*, 27(2):722–732, 2021. doi: [10/k8sh](https://doi.org/10/k8sh)
- [30] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, 2004. doi: [10/b4bq4c](https://doi.org/10/b4bq4c)
- [31] J. Pfäb, P. Minh Nhut, and D. Si. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *PNAS*, 118(2), article no. e2017525118, 12 pages, 2021. doi: [10/gjnkwt](https://doi.org/10/gjnkwt)
- [32] 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*, 22(8):986–1004, 2003. doi: [10/czrq8h](https://doi.org/10/czrq8h)
- [33] V. Rantos, K. Karius, and J. Kosinski. Integrative structural modeling of macromolecular complexes using Assemblin. *Nat Protoc*, 17:152–176, 2022. doi: [10/gq586q](https://doi.org/10/gq586q)
- [34] P. Reddy, P. Guerrero, M. Fisher, W. Li, and N. J. Mitra. Discovering pattern structure using differentiable compositing. *ACM Trans Graph*, 39(6), article no. 262, 15 pages, 2020. doi: [10/gtmxvb](https://doi.org/10/gtmxvb)
- [35] J. Schmidt, R. Preiner, T. Auzinger, M. Wimmer, M. E. Gröller, and S. Bruckner. YMCA – Your mesh comparison application. In *Proc. VAST*, pp. 153–162. IEEE Computer Society, Los Alamitos, 2014. doi: [10/f3sttr](https://doi.org/10/f3sttr)
- [36] M. Schäfer, N. Bricht, J. Byška, S. M. Marques, D. Bednář, P. Thiel, B. Kozlíková, and M. Krone. InVADo: Interactive visual analysis of molecular docking data. *IEEE Trans Vis Comput Graph*, 30(4):1984–1997, 2024. doi: [10/gtmn2n](https://doi.org/10/gtmn2n)

- [37] A. Seiler, D. Großmann, and B. Jüttler. Spline surface fitting using normal data and norm-like functions. *Comput Aided Geom Des*, 64:37–49, 2018. doi: [10/gd5gt4](https://doi.org/10/gd5gt4)
- [38] T. R. Shaham, T. Dekel, and T. Michaeli. SinGAN: Learning a generative model from a single natural image. In *Proc. ICCV*, pp. 4569–4579. IEEE Computer Society, Los Alamitos, 2019. doi: [10/gg8fc3](https://doi.org/10/gg8fc3)
- [39] L. Shang, J. Lv, and Z. Yi. Rigid medical image registration using PCA neural network. *Neurocomput*, 69(13–15):1717–1722, 2006. doi: [10/fsrrb2](https://doi.org/10/fsrrb2)
- [40] R. Skånberg, I. Hotz, A. Ynnerman, and M. Linares. VIAMD: A software for visual interactive analysis of molecular dynamics. *J Chem Inf Model*, 63(23):7382–7391, 2023. doi: [10/gs7tzh](https://doi.org/10/gs7tzh)
- [41] 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*, 22(1):718–727, 2016. doi: [10/mpwv](https://doi.org/10/mpwv)
- [42] 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*, 15(11):905–908, 2018. doi: [10/gfmb8d](https://doi.org/10/gfmb8d)
- [43] P. Thevenaz and M. Unser. Optimization of mutual information for multi-resolution image registration. *IEEE Trans Image Process*, 9(12):2083–2099, 2000. doi: [10/fs4j7h](https://doi.org/10/fs4j7h)
- [44] P. Ulbrich, M. Waldner, K. Furmanová, S. M. Marques, D. Bednář, B. Kozlíková, and J. Byška. sMolBoxes: Dataflow model for molecular dynamics exploration. *IEEE Trans Vis Comput Graph*, 29(1):581–590, 2023. doi: [10/kvjv](https://doi.org/10/kvjv)
- [45] J. W. von Goethe. *Faust. Eine Tragödie [Faust: A Tragedy]*. J. G. Cotta, Tübingen, 1808. Online: gutenberg.org/ebooks/21000; English translation: gutenberg.org/ebooks/3023.
- [46] T. Walton, M. Gui, S. Velkova, M. R. Fassad, R. A. Hirst, E. Haarman, C. O’Callaghan, M. Bottier, T. Burgoyne, H. M. Mitchison, and A. Brown. Axonemal structures reveal mechanoregulatory and disease mechanisms. *Nature*, 618(7965):625–633, 2023. doi: [10/gscb95](https://doi.org/10/gscb95)
- [47] R. Y.-R. Wang, M. Kudryashev, X. Li, E. H. Egelman, M. Basler, Y. Cheng, D. Baker, and F. DiMaio. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat Methods*, 12(4):335–338, 2015. doi: [10/gf5cn8](https://doi.org/10/gf5cn8)
- [48] 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*, 12, article no. 2302, 9 pages, 2020. doi: [10/gmz55b](https://doi.org/10/gmz55b)
- [49] 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*, 33(1):1230–1237, 2019. doi: [10/ghkjcf](https://doi.org/10/ghkjcf)
- [50] 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*, 17(8):1178–1190, 2011. doi: [10/dkbjmd](https://doi.org/10/dkbjmd)
- [51] X. Zhou, S. Kong, A. Maker, S. G. Remesh, K. K. Leung, K. A. Verba, and J. A. Wells. Antibody discovery identifies regulatory mechanisms of protein arginine deiminase 4. *Nat Chem Biol*, 20(6):742–750, 2024. doi: [10/mpwf](https://doi.org/10/mpwf)
- [52] Y. Zhou, Z. Zhu, X. Bai, D. Lischinski, D. Cohen-Or, and H. Huang. Non-stationary texture synthesis by adversarial expansion. *ACM Trans Graph*, 37(4), article no. 49, 13 pages, 2018. doi: [10/gd52tv](https://doi.org/10/gd52tv)