

## A guided approach for subtomogram averaging of challenging macromolecular assemblies

Benjamin Basanta<sup>a</sup>, Saikat Chowdhury<sup>b</sup>, Gabriel C. Lander<sup>a,\*</sup>, Danielle A. Grotjahn<sup>a,\*</sup>

<sup>a</sup> Department of Integrative Structural and Computational Biology, Scripps Research, HZ 175, 10550 N Torrey Pines Rd., La Jolla, CA 92037, United States

<sup>b</sup> Department of Biochemistry and Cell Biology, 144 Center for Molecular Medicine, Stony Brook University, Stony Brook, NY 11794, United States

### 1. Introduction

Cryo-electron microscopy (cryo-EM) is an impactful methodology for three-dimensional (3D) structural determination of macromolecular complexes. While single particle EM gained widespread notoriety for its utility in solving high resolution structures of purified proteins, cryo-electron tomography (cryo-ET) has emerged as the leading technique for visualizing the structures of large, transient, dynamic, flexible, and/or heterogeneous samples in native or near-native reconstituted cellular environments (Baumeister, 2013; Oikonomou and Jensen, 2017). The implementation of automated data collection (Blocker et al., 1997) packages (Mastrorade, 2005; Suloway et al., 2009) and optimized tomographic acquisition schemes (Chreifi et al., 2019; Eisenstein et al., 2019; Hagen et al., 2017; Turoňová et al., 2019), combined with direct electron detectors, energy filters, and phase plates (Khoshouei et al., 2017) has revolutionized the feasibility of visualizing cellular machinery for functional and physiological interpretation. Multiple copies of the biological complex of interest can be identified within reconstructed tomograms, and 3D “subvolumes” or “subtomograms” can be extracted and averaged together in a process called subtomogram averaging (STA) to obtain better-resolved 3D reconstructions of the complex of interest. Several STA processing packages with diverse algorithmic approaches have been developed, including PEET (Heumann et al., 2011; Nicastro et al., 2006), Dynamo (Castano-Diez et al., 2017; Castaño-Diez et al., 2012) or PyTom (Hrabe et al., 2012). Additionally, aspects of single-particle image processing have been incorporated into STA processing packages such as RELION and EMAN2 (Bharat and Scheres, 2016; Bharat et al., 2015; Galaz-Montoya et al., 2015). Notably, when combined with improved 3D-contrast transfer function (CTF) estimation and missing-wedge compensation (Chen et al., 2019; Galaz-Montoya et al., 2016; Himes and Zhang, 2018; Turoňová et al., 2017), STA has been implemented to achieve reconstructions in the sub-nanometer resolution regime (Himes and Zhang, 2018; Schur et al., 2016; Tegunov et al., 2020; Turoňová et al., 2017), even reaching resolutions that are

comparable with single particle analyses, further emphasizing the promise of this technique in obtaining high-resolution structural information of complexes *in situ*.

However, despite improvements in instrumentation and algorithms, the field is still far from routinely obtaining high resolution structures by STA, as most structures deposited in the EM Data Bank (EMDB) and determined by this method are at resolutions worse than  $\sim 20$  Å (Fig. 1A). Moreover, while the ability to elucidate the structures of pleomorphic, multi-subunit complexes in native, *in situ* cellular environment is a major advantage of cryo-ET and STA over other structural techniques, current STA processing strategies are typically only successful for highly ordered, symmetrical, homogenous samples that have limited conformational variations, and are present in high copy numbers within a single tomogram (Fig. 1B). Examples of such complexes include purified viruses and associated viral complexes (Obr and Schur, 2019), and highly-abundant cytoplasmic and membrane-associated ribosomes (Orlov et al., 2017; Pfeffer et al., 2016). Protein complexes that are uniformly oriented with membranes or filaments have also benefited greatly from this technique, as alignment of the relatively high-signal membrane or filaments can help drive the initial alignment of the noisier, low SNR complex of interest. Examples of these complexes include the axonemal dynein motors (Grotjahn and Lander, 2019), COPI/II vesicle coats (Markova and Zanetti, 2019), bacterial secretion systems and flagellar motors (Oikonomou and Jensen, 2019).

In some cases, however, the presence of a relatively high-signal complex can deter, rather than facilitate, initial alignment of the complex of interest, particularly in cases where the complex of interest exhibits non-uniform binding or does not adopt a single, biologically relevant orientation relative to the signal-dominant filament or membrane structure. This is particularly relevant for tomograms of *in vitro* reconstituted or native cellular landscapes teeming densely packed macromolecules, which may include but not limited to signal-dominant membrane structures, large and featureful filaments, or an abundance of cytosolic complexes. Subtomogram alignment algorithms generally

\* Corresponding authors.

E-mail addresses: [glander@scripps.edu](mailto:glander@scripps.edu) (G.C. Lander), [grotjahn@scripps.edu](mailto:grotjahn@scripps.edu) (D.A. Grotjahn).

<https://doi.org/10.1016/j.yjsbx.2020.100041>

Received 3 February 2020; Received in revised form 13 October 2020; Accepted 18 November 2020

Available online 23 November 2020

2590-1524/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

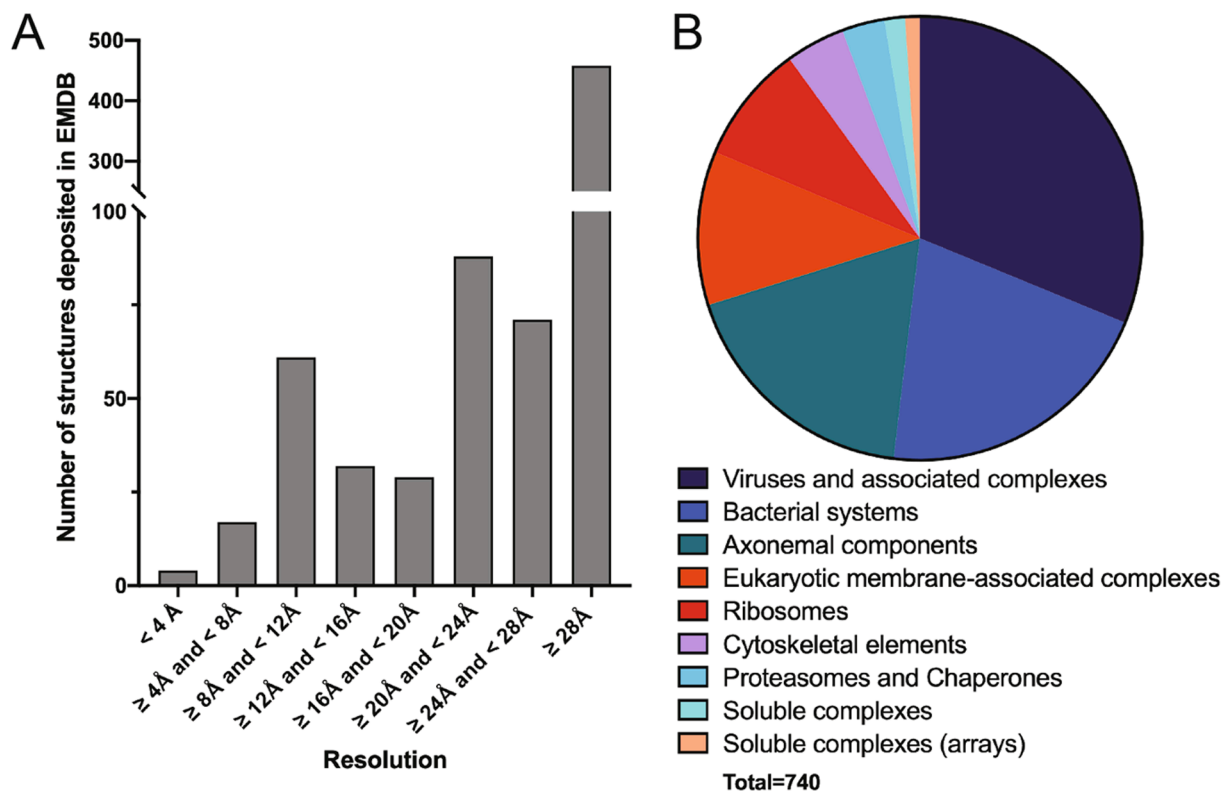
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

assume that the macromolecule of interest within extracted subvolumes contains sufficient low resolution features and contrast to approximate the orientation information necessary to produce an initial average (Galaz-Montoya et al., 2016), whose structural details improve through more accurate orientational assignments of the targeted macromolecule during 3D refinement. However, in cases where the subvolumes contain a diverse array of distinct subcellular structures, each with their own signature electron scattering profile, a more signal-dominant feature may be preferentially aligned over the course of the 3D refinement, resulting in misalignment of the targeted macromolecule. Since successful initial alignment is important for convergence of a well-resolved subtomogram average, failure at this early step represents a major hurdle in subtomogram averaging of complexes and heterogeneous assemblies. A potential solution is to provide *a priori* orientation information prior to alignment. Such information is typically calculated during the particle picking procedure using template matching and/or other automated detection algorithms (Albert et al., 2020; Hrabe et al., 2012; Navarro et al., 2018; Wietrzynski et al., 2020). However, for particularly crowded environments and/or challenging protein targets, these semi-automated procedures can result in high false-positive rates, such that the initial orientation information derived from these methods fail to produce reliable reconstructions. Therefore, one of the fundamental areas for growth is the development of processing strategies to calculate *a priori* orientation information prior to subtomogram averaging in cases where other established semi- or fully automated detection procedures fail, and the user is left with no other options to computationally derive starting orientations for 3D refinement of subtomograms.

In order to address some of these challenges inherent to heterogeneous samples, we developed a guided, focused refinement approach to elucidate 3D structures of large, flexible, asymmetric biological complexes present in relatively low abundance within individual

tomograms. This guided approach has the potential to overcome several limitations described above, and will be applicable to solve structures of macromolecular complexes that display recognizable features that are unambiguously discernible to the user within the subvolumes, but not identified by automated particle selection programs, and/or not reliably aligned by 3D classification or refinement algorithms. Users may encounter such situations when the biological target is (1) present within a “crowded” subcellular environment, where many diverse and variable biological features obfuscate the target; (2) dynamic and/or flexible in such a way that every targeted complex represents a unique conformational species; and/or (3) associated with another large, high-signal biological complex, and initial attempts using STA result in alignment of this signal-dominant feature instead of the targeted complex.

The overall methodology for the guided approach was originally developed to overcome significant challenges we faced in elucidating the 3D reconstruction of the large, flexible, asymmetric microtubule (MT)-bound dynein-dynactin-BicaudalD2 (DDB-MT) complex (Grotjahn et al., 2018). We present here a streamlined version of the guided approach, including a package of python scripts ([https://github.com/GrotjahnLab/guided\\_tomo\\_align](https://github.com/GrotjahnLab/guided_tomo_align)) that convert orientation information from extracted subvolumes to files that can be directly used as input for subtomogram averaging in the RELION software package (Bharat and Scheres, 2016). We detail the specific steps of our workflow such that other users can apply this methodology to a wide range of other biological macromolecules of interest. We use the DDB-MT complex as a test case and demonstrate how application of the guided STA approach overcame the unique structural challenges posed by this complex for subtomogram averaging by conventional approaches. Furthermore, we demonstrate that our method is free of reference model or user bias, and describe a python script that users can use to specifically test for bias in their own experiments. The workflow is described



**Fig. 1.** Subtomogram averaging: current state of the field (A) Histogram displaying resolution distribution of structures solved by subtomogram averaging. While some high resolution (<4 Å) structures have been reported, the majority of the deposited structures solved by subtomogram averaging are low resolution maps. (B) Pie chart demonstrated the types of biological complexes solved by subtomogram averaging. For both (A) and (B), data was downloaded from Electron Microscopy Data Bank in October 2019 and sorted by resolution and biological complex, respectively.

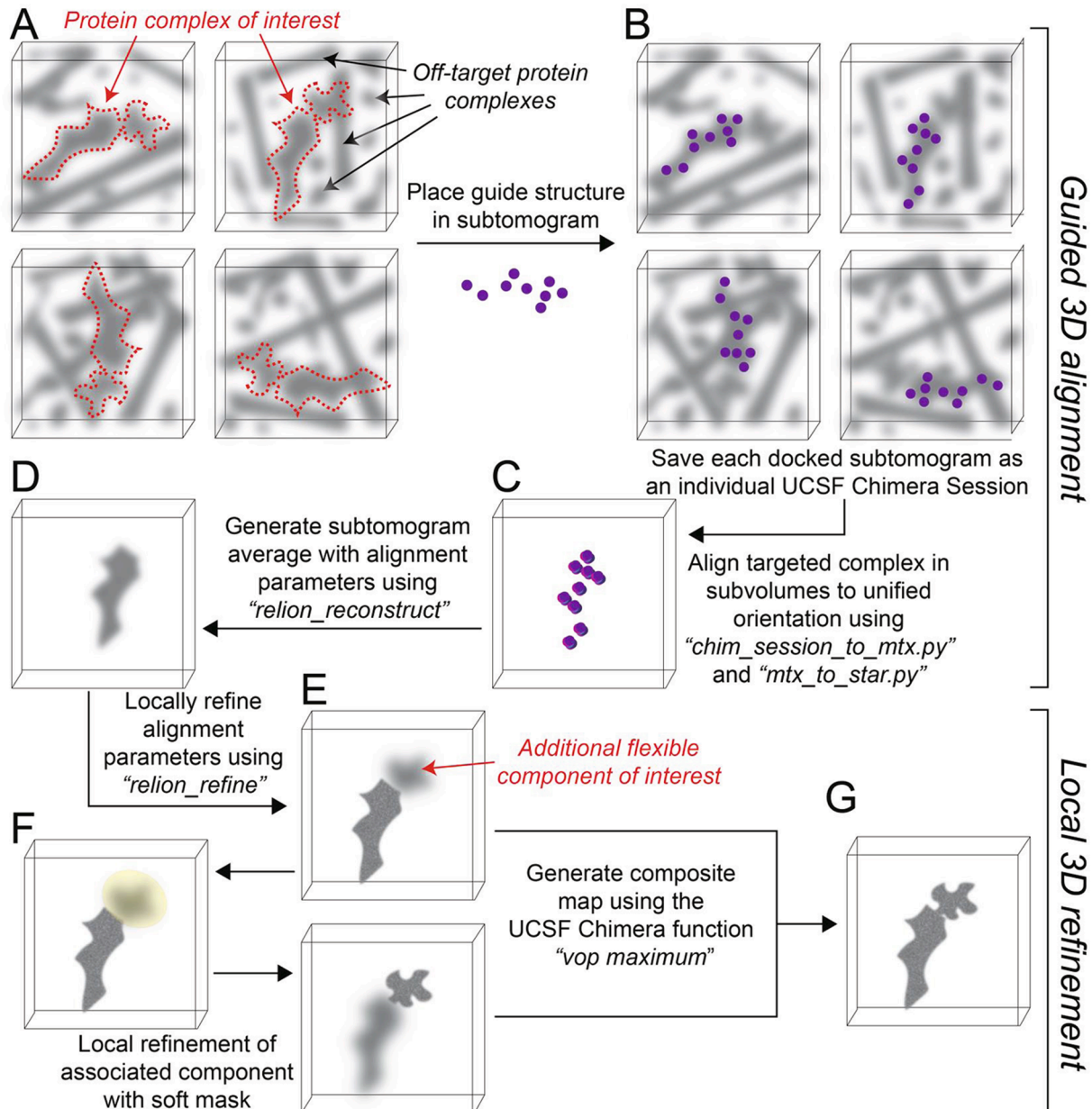
in the context of the RELION software package due to the straightforward access and manipulation of parameter files and metadata offered by this package, but note that this workflow could be adapted to generate input files for other reconstruction packages (Burt, 2020).

## 2. Overview of approach

### 2.1. Subvolume Extraction.

The first step involves manually identifying the complexes within the

reconstructed tomogram for 3D subvolume extraction. This process can be performed using any of several STA processing packages, such as EMAN2 (Galaz-Montoya et al., 2015), PEET (Heumann et al., 2011; Nicastro et al., 2006), Dynamo (Castano-Diez et al., 2017; Castaño-Diez et al., 2012) or PyTom (Hrabe et al., 2012) (Fig. 2A). For ease in identification of complexes of interest in subsequent steps in this procedure, it is recommended that particles are initially extracted from tomograms reconstructed using iterative reconstruction methods such as Simultaneous Iterative Reconstruction technique (SIRT) (Gilbert, 1972) or Simultaneous Algebraic Reconstruction Technique (SART) (Andersen



**Fig. 2.** Workflow for guided subtomogram averaging approach. (A) Cartoon representation of four extracted subvolumes from a low signal-to-noise cryo-tomogram. The cartoon biological complex of interest is outlined in red dashed line in each subvolume. (B) A guide structure, represented as 9 purple dots, is manually placed (i. e. "docked") in the same position as the protein of interest within the extracted subvolume. The docked guide structure and the subvolume is saved as a single UCSF Chimera session. This process is repeated for each extracted subvolume in the dataset. (C) The python-based scripts "chim\_session\_to\_mtx.py" and "mtx\_to\_star.py" are used to generate starting orientation (Euler) angles and rotation origins for each UCSF Chimera session saved in (B). The resulting information is compiled into a single input file for RELION (STAR file). (D) Using the "reliion\_reconstruct" command, the initial orientations per particle described in the STAR file generated in (C) are applied to the extracted subvolumes to align an initial 3D structure of the protein of interest. (E) These angles are further refined using probability-weighted angular assignment using "reliion\_refine" command. (F) To resolve flexible regions, a soft, 3D binary mask (shown in yellow) can be used to focus the alignment to these regions. (G) Individual focused maps can be combined using the "vop\_resample" command in UCSF Chimera to produce the final composite structure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and Kak, 1984), or from other deconvolution or denoising approaches that boost contrast (Bepko et al., 2019; Buchholz, 2018; Tegunov and Cramer, 2019). A list of coordinates denoting the approximate centroid of the identified complexes should be generated and used to extract subvolumes from each tomogram, so that the complex of interest is roughly positioned in the center of the extracted volume. Although it is generally advisable for conventional STA approaches to choose a box size that will limit the inclusion of significant off-target biological material, we note that the success of the guided STA approach should not be significantly influenced by surrounding features within the extracted subvolume, since only local searches are performed during 3D refinement. Therefore, emphasis should be placed on choosing a box size that is sufficient to accommodate the entirety of the complex of interest, rather than limiting the extraction box size to tightly encompass the targeted molecule. Furthermore, as a means of identifying possible model bias introduced during the initial orientation assignment, it is important that tight 3D masks are not imposed during the 3D refinement steps (see “Potential for user and reference bias” section below).

## 2.2. Identification of targeted complexes within individual subvolumes

Once the 3D subvolumes have been extracted, the user must assign an approximate 3D orientation ( $x$ ,  $y$ ,  $z$ ,  $\phi$ ,  $\theta$ ,  $\psi$ ) to the complex located at the center of the subvolume. This is accomplished by placing a 3D map (referred to herein as “guide structure”) in the same position as the complex of interest within the extracted subvolume (Fig. 2B). The identity of the guide structure itself does not influence the success of this approach, and therefore the user can elect to choose any 3D map available, including a deposited structure from a publicly-available database or simply a set of marker points that denote the relative locations of salient structural features and saved as a 3D map. This approach fundamentally relies on the consistent and accurate placement of the guide structure in an orientation that represents that of the complex of interest for all subtomograms used for subtomogram averaging. To facilitate this process, the user simultaneously views both (1) a single extracted subvolume, and (2) 3D guide structure within UCSF Chimera program (Pettersen et al., 2004) (Fig. 2B). Often, alternating the viewing mode from 2D planes/slices of the subvolume to 3D isosurface rendering within UCSF Chimera can aid in identifying the features of interest and placing the guide structure within the extracted subvolume. Thus, the 3D guide structure now reflects the observed position of the reconstructed complex within the subvolume. Once properly positioned, both the (1) extracted subvolume, and (2) the guide structure are saved as a single UCSF Chimera session (Pettersen et al., 2004) (Fig. 2B). To facilitate downstream processing, a naming system that resembles the name of the extracted subvolume should be used when saving the individual UCSF Chimera sessions (for example, a Chimera session that contains an extracted subvolume named “tomogram012\_subtom0123.mrc” should be saved as “tomogram012\_subtom0123.py”).

## 2.3. Calculate initial orientations and rotation origins from docked UCSF chimera sessions using the guided\_tomo\_align package

To convert the position and 3D orientation of the docked model within the subvolume into Euler angles and translations that can be used by RELION for 3D refinement and subtomogram averaging, the user runs the two python scripts within the *guided\_tomo\_align* package ([https://github.com/GrotjahnLab/guided\\_tomo\\_align](https://github.com/GrotjahnLab/guided_tomo_align)): “chim\_session\_to\_mtx.py” followed by “mtx\_to\_star.py” (Fig. 2C). The “chim\_session\_to\_mtx.py” script requires as input the location of the Chimera session files and a series of volume descriptors (guide volume filename, pixel sizes of guide and target volume files and associated box sizes) and outputs a plain text file (“matrix\_dictionary.txt”) containing the rotation matrices describing the position of the guide structure relative to the targeted macromolecule in each of the subtomograms. The “mtx\_to\_star.py” script takes as input the “matrix\_dictionary.txt” and a RELION-

formatted “\*.star” file containing the list of subtomograms into which the guide structure was docked, and outputs a file (“\*.init\_from\_Chimera.star”) containing the corresponding rotational (rlnAngleRot, rlnAngleTilt, rlnAnglePsi) and translational (rlnOriginX, rlnOriginY) parameters that will serve as the orientational starting points for local 3D refinement in RELION. In case the user has not maintained a consistent naming schema for the Chimera sessions and the subtomogram filenames included in the “\*.star” file, we provide the option to input a CSV file that maps the names of the subtomogram averages contained in the star file to the corresponding Chimera sessions. Further details about the conversion steps carried out by these python scripts are provided in Supplemental Text 1. While “mtx\_to\_star.py” script produces a file that is specifically for use in the RELION STA package (as described in subsequent steps), a user could readily use the “matrix\_dictionary.txt” to generate Euler and translational values, or take advantage of other publicly-available RELION star file conversion scripts to produce files for use with other STA programs such as Dynamo (Burt, 2020).

## 2.4. Refine orientations to generate aligned 3D structure of feature of interest

Using user-defined Euler angles and translations for each subtomogram, the user can next use the output RELION star file to generate an initial subtomogram average using the “*reliion\_reconstruct*” program in RELION with the “3d\_rot” parameter set to true. At this point, provided the targeted cellular complex contains consistent structural features and positioned in a range of orientations across the subtomograms, the resulting subtomogram average should provide a preliminary 3D visualization of the complex with increased SNR and diminished missing wedge artifacts (Fig. 2D). This preliminary reconstruction will now serve as a starting model for further iterative refinement of the subvolumes using a standard 3D-refinement procedure (Fig. 2E). However, due to the crowded nature of the subvolumes, starting the refinement using a global, coarse-grained search will likely lead to a preferential alignment of neighboring, off-target, higher-signal structures that precluded traditional refinement approaches from the outset. Thus, the users should limit the rotational and translational search range (recommended angular search of 1.8 and translational search range of 5 pixels). In order to safeguard against model bias this refinement should not be carried out using a tight 3D mask that encompasses the initial model (see “Potential for user and reference bias” section below). A successful 3D refinement will converge to a reconstruction that contains improved structural details of the complex of interest (Fig. 2E).

## 2.5. 3D focused refinement of flexible regions belonging to the targeted complex

If the alignment procedure described above leads to a well-resolved structure of the entire biological complex of interest, then no further processing is needed. However, in most cases, this STA processing strategy will be applied to macromolecular complexes that may contain other, more flexible subunits that associate with the feature of interest used to assign orientations. Notably, the complexes targeted using this approach are likely challenging for automated methods due to the presence of flexible regions, which will likely not be resolved in the reconstructed density. However, since a portion of the complex in each subtomogram has now been assigned a common orientation, an attempt can be made for refining these flexible portions, using a 3D binary, soft-edged masks (ellipsoidal or spherical) corresponding to individual subregions of the complex (Fig. 2F). It is worth noting that, in contrast to guided 3D refinement, additional masked refinements are more susceptible to alignment of off-target biological features that may be present within the vicinity of the complex of interest within individual subvolumes. Therefore, it is generally advisable that limited search parameters, such as those used for the initial 3D refinement, are also employed for additional 3D focused refinements. Due to these limited

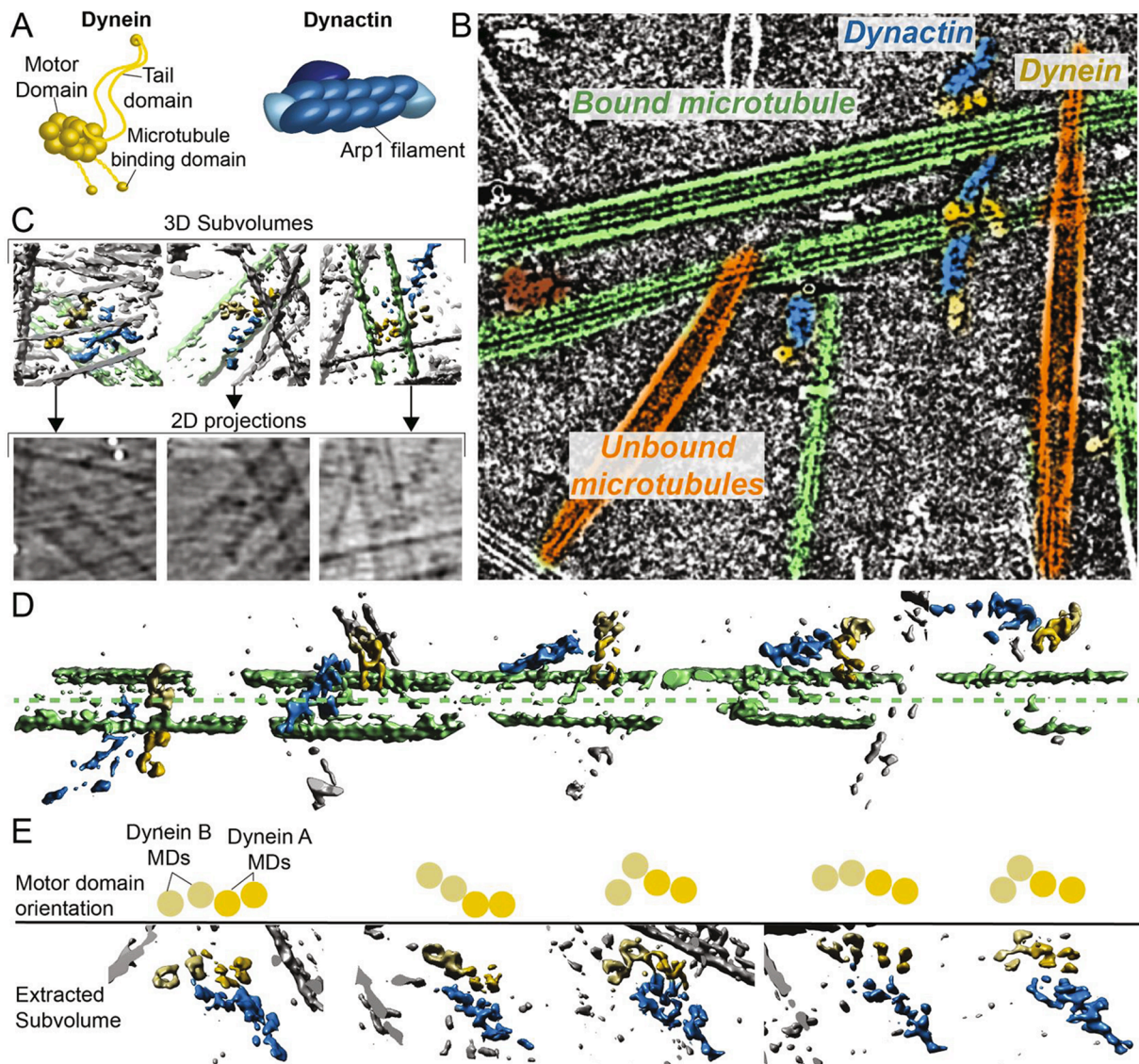
search constraints, this masked approach should only be employed to further refine domains that are generally ordered relative to the initially targeted complex, and may not work for domains that exhibit an extremely high degree of flexibility.

Combining results of individual focused refinements to produce a single 3D structure of the macromolecular complex of interest. The individual maps resulting from focused 3D alignment and averaging of each sub-region of the biological complex can be combined using the “*vop maximum*” function in UCSF Chimera, which measures and retains the maximum voxel values of overlapping volumes to produce a final composite reconstruction (Fig. 2G).

### 3. Application of the methodology to a challenging complex: Microtubule-bound dynein-dynactin

#### 3.1. Challenges to working with the microtubule-bound dynein-dynactin-BicaudalD2 (DDB) complex: A sample dominated by variability, heterogeneity and flexibility.

Due to its inherent asymmetry, relatively large size (~1.5 MDa) and multi-subunit complexity, cytoplasmic dynein represents a challenging target for 3D structure determination. Dynein is activated for microtubule-based transport upon binding to additional cofactors, including another large, (~1.5 MDa) multi-subunit complex called



**Fig. 3.** Microtubule-bound dynein-dynactin complex: a challenging structural target. (A) Cartoon representations of cytoplasmic dynein (left, yellow), and dynactin (right, blue) with labeled components of major structural and functional domains. (B) Representative x-y slice of a reconstructed, three-dimensional tomogram, colored to indicate different components present within *in vitro* reconstituted dynein-dynactin transport environment, included dynein motor domains (yellow) in complex with dynactin (blue), bound to microtubules (green), as well as non-specific protein aggregates (brown) and microtubules that appear devoid of bound dynein-dynactin complexes (orange). (C) Three representative extracted subtomograms and corresponding 2D projections to illustrate how the microtubule densities present in each individual subvolume can dominate the signal, likely making it difficult for computational algorithms to automatically extract and align voxels containing the relatively lower-signal density of the dynein-dynactin complexes. (D) Five representative extracted subtomograms displayed with aligned microtubules missing wedge, showing the degree of variability that the dynein-dynactin complex can bind to the microtubules when taking into account alignment of the missing wedge of the microtubule. (E) Five representative extracted subtomograms colored similar to (B), oriented such that the dynein density (blue) is in the same position for each. Colored circles (top) represent position of motor domains in the corresponding subtomogram, demonstrating the range and diversity of conformational flexibility among the four dynein motor domains that is unique to each individual subtomogram. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dynein (Fig. 3A). In order to investigate how these cofactors activate dynein for microtubule-based transport, we developed a near-native reconstitution system to purify microtubule-bound dynein-dynein complexes from mouse brain lysate and performed cryo-ET to visualize the 3D architecture of this transport complex (Grotjahn et al., 2018) (EMD-7000, EMPIAR-10520). Although the complexes were readily visible within reconstructed 3D tomograms (Fig. 3B), there were several unique challenges that prevented the application of automated, template-based approaches for subvolume selection. For example, due to slight variations in the amount of endogenous proteins (tubulin, dynein, dynein), both the total number and the binding pattern of DDB complexes on individual microtubules varied dramatically between sample preparations and within tomograms derived from the same batch of prepared sample (Fig. 3B). In order to ensure sufficient number of DDB complexes per tomogram for downstream processing by STA, we increased the concentration of decorated microtubules deposited on the EM grid, thus increasing the number of DDB complexes per tomogram. However, this also led to an increase in the overall ice thickness of the sample, further decreasing the SNR of individual extracted subvolumes, wherein the signal from the microtubule densities dominate, and obfuscate the sparsely decorated dynein-dynein complexes (Fig. 3C). These challenges precluded our ability to use of automated search algorithms to select DDB particles for 3D subvolume extraction (Supplementary Fig. 1).

After manual particle picking and subvolume extraction, we attempted multiple strategies to align, classify, and average the subvolumes using both RELION and EMAN2 subtomogram averaging workflows (Supplementary Fig. 2). However, none of our attempts were successful at producing a well-resolved structure with features that resembled previously-resolved portions of the microtubule-bound dynein-dynein complex (Chowdhury et al., 2015; Urnavicius et al., 2015). In most instances, these programs produced a structure that resembled a microtubule in close proximity to an uninterpretable density that might correspond to DDB complexes, however there was no clear indication of dynein's characteristic structural features, including the donut-shaped motor domains, or dynein's arp1 filament (compare cartoon diagrams in Fig. 3A and identified complexes in subvolumes in Fig. 3E with the reconstructed densities in Supplementary Fig. 2). Strategies to eliminate the putative microtubule density by applying a 3D soft mask to perform focused refinement of this putative DDB complex were similarly unsuccessful (e.g. regions denoted by yellow spheres or ellipsoids in Supplementary Fig. 2). Overall, the failure of both RELION and EMAN2, which rely on distinct maximum-likelihood and alignment-based averaging algorithms, to generate a reconstruction of the DDB complex indicated that this biological specimen presents unique challenges that is particularly recalcitrant to conventional methods.

These inconclusive results were surprising, since manual inspection of *x-y* slices of tomograms demonstrated clearly discernible structural features corresponding to dynein and dynein complexes (Fig. 3B). We posit that that several unique features of the DDB-MT complex sample can explain this discrepancy. For one, the "missing wedge" of information that results from the inability to fully capture all tilted views of the sample during tomographic image acquisition is most notable in the microtubule filaments within extracted subvolumes, where missing 2D projections result in large sections of microtubules lacking density (Fig. 3D). Averaging of extracted subvolumes may be driven by the alignment of the missing wedge artifact within microtubule densities, rather than by the comparably smaller, lower-signal DDB complexes, despite the implementation of missing wedge compensation during refinement (Fig. 3C-D). Alignment of the microtubule results in significant misalignment of DDB complex, which is likely exacerbated by the variability by which the DDB complexes orient on the microtubule (Fig. 3D), and further compounded by the extreme flexibility among the motor domains within individual DDB complexes (Fig. 3E).

As mentioned previously, to overcome the sparse decoration of

dynein-dynein complexes on microtubules, we increased the overall concentration of sample deposited on the EM grid, leading to a crowding of microtubule structures within tomograms. In addition to impeding our ability to utilize automated template matching for particle picking (see above), the crowded nature of the sample often resulted in extracted subvolumes that included a single, low-signal dynein-dynein complexes surrounded by numerous high-signal microtubules, which likely also contributed to mis-alignment observed in our initial STA attempts. Although we attempted to extract with a box size that minimally included the full, dynein-dynein complex while excluding neighboring microtubule densities, the extracted subvolumes nonetheless often included many structures in addition to the desired dynein-dynein complex of interest (Fig. 3B-C).

In summary, despite many diverse attempts and strategies, the unique features of the DDB complex regarding variability in sample preparation, missing wedge alignment of signal-dominant microtubules, and extreme heterogeneity within the DDB complex itself all likely contributed to the difficulty in obtaining an interpretable 3D DDB-MT structure using traditional STA methodologies.

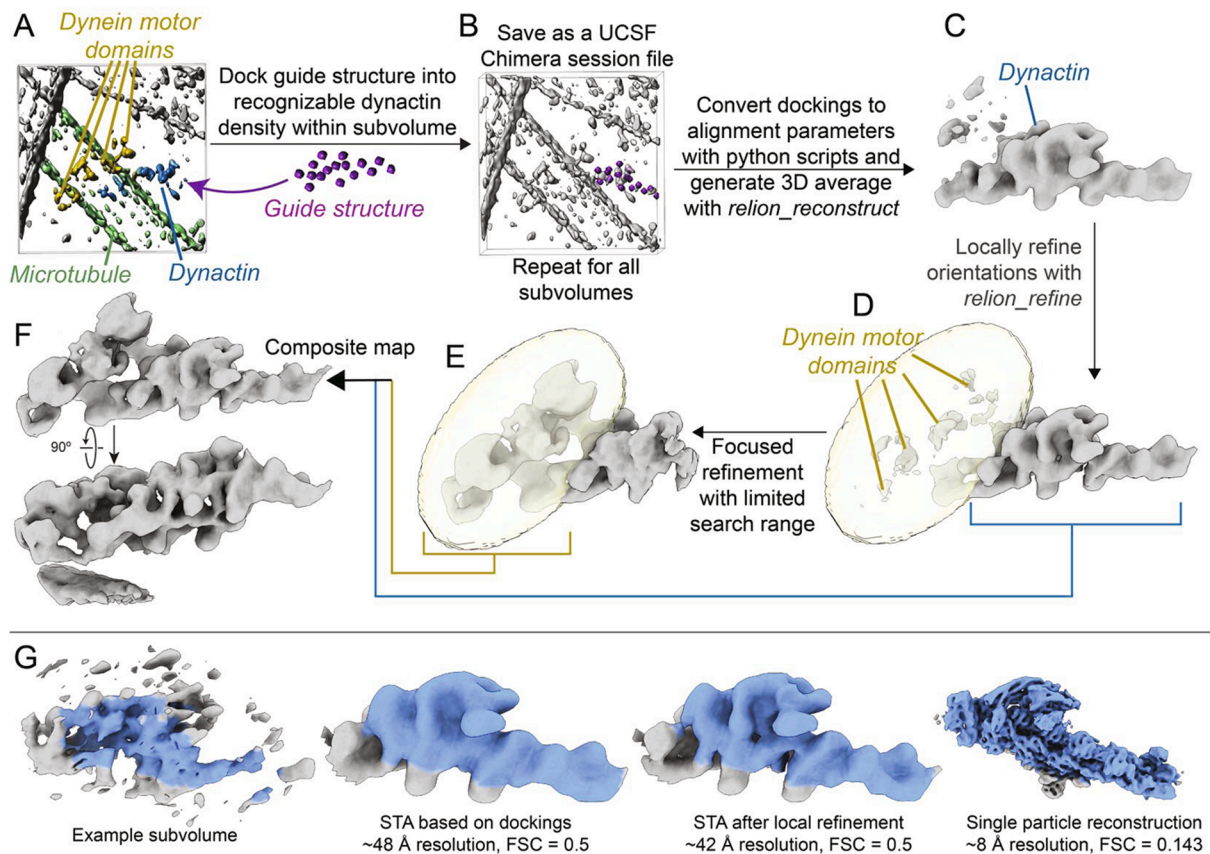
### 3.2. Manual positioning of guide structure into extracted subvolumes using UCSF Chimera

In order to provide *a priori* information to guide the alignment of microtubule-bound dynein dynein complexes, we visualized individual 3D subvolumes extracted from tomograms reconstructed by SIRT in UCSF Chimera. To reduce noise and facilitate the identification of the characteristic features of the DDB complex, a Gaussian filter with a width of 8.52 Å was applied to the subvolumes using the "volume filter" function in UCSF chimera (Fig. 4A, Supplementary Movie 1). In our previous work (Grotjahn et al., 2018), a structure of the dynein tail-dynein-BicaudalD2 complex map (EMD-2860, hereafter referred to as TDB) was resampled to the same pixel size as the extracted 3D subvolumes and manually positioned to match similar features within each subvolume, and the Chimera "fit in map" function was used for more precise fitting. While this procedure took advantage of a single particle cryo-EM map representing a portion of our targeted complex to generate initial orientations of the full complex in our subvolumes, we wanted to test whether our approach could be applied to protein complexes lacking previously-described structures of sub-components.

To accomplish this, we replaced the previously-docked single particle map with a simplified guide structure (hereafter referred to as the "skeleton guide structure") that consisted of a collection of small spheres denoting the centroid of notable structural features observed in the subtomogram averages (the subunits of the shoulder domain, barbed end, arp1 filament, and pointed end) (Fig. 4A-B, purple spheres). This skeleton guide structure, which lacked any 3D structural features, was docked into each of the 480 extracted subvolumes in our dataset, and the "chim\_session\_to\_mtx.py" and "mtx\_to\_star.py" scripts described above were used to generate a STAR file containing the RELION-formatted orientation alignment parameters for each subvolume based on the position of the docked guide structure (Fig. 4B).

### 3.3. 3D alignment of dynein-dynein complex.

The STAR file containing our guided alignment parameters was input to the "relion\_reconstruct," command, resulting in a subtomogram average with a reported resolution of ~ 48.0 Å according to the Fourier Shell Correlation (FSC) at a 0.5 cutoff (Fig. 4C, Supplementary Fig. 3A). Despite the low reported resolution, the overall clarity of characteristic features and subunits of the dynein-dynein complex in the reconstruction was substantially improved compared to the original subvolumes, indicating an overall qualitative improvement resulting from the guided STA approach (Fig. 4G, Supplementary Fig. 3C). As evidenced by the Euler distribution plot, the constitutive subvolumes exhibited a wide range of orientations (Supplementary Fig. 3B), and the



**Fig. 4.** Application of guided approach to solve structure of dynein-dynactin complex. (A) Representative, gaussian-filtered subvolume containing a microtubule-bound dynein-dynactin complex. Individual components are labeled and colored within the subvolume (microtubule, green; dynein, blue; dynein motor domains, yellow; other complexes, gray). (B) A “skeleton guide structure” (shown in purple) was manually placed to reflect the position of the dynactin component within the extracted subvolume (same subvolume as (A) shown in gray). The docked guide structure and subvolume were saved as a single UCSF Chimera session file. This process was repeated for  $\sim 480$  subvolumes in the dataset. (C) The “chim\_session\_to\_mtx.py” and “mtx\_to\_star.py” python scripts were used to generate a RELION-formatted parameter file with initial alignments for the subvolumes, and the “reliion\_reconstruct” program was used to generate a 3D average of the subvolumes. (D) The orientation angles locally refined in RELION to generate a structure with improved density corresponding to the dynactin portion of the dynein-dynactin complex. Additional densities likely corresponding to the dynein motor domains were also visible in the structure, and an ellipsoidal, soft-edge binary mask (yellow) was used to perform a focused 3D refinement of this region. (E) Local 3D refinement in RELION of the masked region increased the resolvability of the dynein motor domains. (F) The two maps generated from the dynactin and the motor domains were stitched together using the “vop maximum” function in UCSF Chimera to produce a combined structure of the microtubule-bound dynein-dynactin complex. (G). Qualitative and quantitative improvements in the resolvability of the dynactin structure using the guided approach is demonstrated by comparing the EM densities corresponding to dynactin (colored blue) at each step of the process. On the far left a representative subvolume from the dataset is shown, followed by the output from the “reliion\_reconstruct” command, and then after 3D refinement in RELION. On the far right a previously published single particle reconstruction of the dynactin complex (EMD-2860) is shown to confirm that the domain architecture of our final 3D average is consistent with higher resolution studies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reconstruction was accordingly devoid of missing wedge/cone artifacts. Most notably, the subtomogram average closely resembled the dynein-dynactin complex previously determined by single particle (EMD-2860) (Urnovicus et al., 2015) and our previous work (EMD-7000) (Grotjahn et al., 2018). In an attempt to quantitatively assess the improvement of structural features of the STA relative to the original subvolumes, we compared the FSCs calculated between the single particle reconstruction of DDB (EMD-2860) and the individual subvolumes to that calculated against the averaged reconstruction (Supplementary Fig. 3C). We also calculated the cross-correlation scores between single particle reconstruction and the original subvolumes or the guided STA reconstruction using UCSF Chimera (Supplementary Fig. 3B-C). Both the reported resolution according to FSC as well as the cross-correlation values were significantly higher for the STA than for the original subvolumes, supporting the improvement in quality of the averaged reconstruction using the guided approach.

The alignment parameters giving rise to this reconstruction were further refined in RELION using the local search parameters described in

the “Overview of approach” section, using a spherical mask whose diameter was equal to the dimension of the subvolumes. The refinement converged to a structure with an estimated resolution of  $\sim 42.4$  Å resolution at FSC = 0.5 (Fig. 4D, Supplementary Fig. 3A). While the improvement according to FSC compared to the resulting structure from “reliion\_reconstruct” is not dramatic, the refined structure contains better-defined features, likely due to the probability-weighting applied in the “reliion\_refine” algorithm in addition to improved alignment parameters. Furthermore, as described below, this refinement step can serve to identify model bias introduced during the initial docking procedure (Supplementary Fig. 4). Importantly, additional densities likely corresponding to the flexible dynein motor domains could be observed in the reconstruction (Fig. 4D). To further resolve these flexible regions, we used a 3D soft-edge binary mask that encompassed these portions of the complex (yellow mask in Fig. 4D) and performed a refinement using local angular and translational searches that produced better-resolved dynein motor domain densities (Fig. 4E). The two resulting substructures of dynactin and the dynein motor domains were fitted

relative to each other and then combined into a composite 3D structure of the microtubule-bound DDB complex, using the “*vop maximum*” function in UCSF chimera (Fig. 4F).

### 3.4. Potential for user and reference bias.

It must be noted that alignment of structures into noisy data, whether guided by a human or computationally, presents the possibility for reference bias, which can manifest as a reproduction of the reference in the aligned and averaged data. This bias was previously demonstrated using 2D images, wherein a photograph of Albert Einstein was reproduced from noise (Shatsky et al., 2009). Since such reference bias could similarly occur in 3D, it is imperative that users consider this possibility in carrying out any docking studies. Here we demonstrated that the use of a skeleton guide structure that does not contain any structural features gave rise to an STA that was consistent with prior tomographic and single particle studies (Fig. 4G). Furthermore, as noted in previous studies, the appearance of additional, physiologically relevant densities within our STA (dynein tail and motor domains), that were absent from the guide model, serve as internal controls indicating that the final reconstruction was not an artifact of user or reference bias (Grotjahn et al., 2018).

It is strongly encouraged that users employing our guided approach similarly use minimal versions of their target structures that contain sufficient features for unambiguous identification of a target molecule’s position within a subvolume, while minimizing structural details present in the guide structure. Additionally, the 3D refinement step of our proposed pipeline should be carried out without a 3D mask that was generated based on the guide model. In order to minimize the influence of reference bias, as well as to accommodate the identification of unexpected neighboring cellular interactors, a large mask should be used for these 3D refinements, preferably using a mask that encompasses the entirety of the subvolume. To emphasize the importance of the refinement step in identifying initial user bias, we used the “*chimer-a\_fitmap\_search.py*” to automatically dock a 3D reconstruction of the DDB complex (EMD-2860) into 572 subvolumes from our dataset using the “*fit in map*” function of UCSF Chimera with the search parameter set to 1000. Inspection of the docking results revealed that this automated method failed to identify the true dynein density in any of the observed subvolumes (Supplementary Fig. 1), however the 3D average that arose from these docked orientations using “*reliion\_reconstruct*” contained structural features reminiscent of the guide model (Supplementary Fig. 4A). These results are consistent with the “Einstein from noise” scenario, where off-target features within the subvolumes that resemble the guide model were identified based on cross-correlation metrics. Notably, local 3D refinement of these alignment parameters using a 1.5° angular search range in RELION with a large spherical mask substantially diminished the resemblance of the reconstruction to the guide model, and refinement with an angular search range of 7.5° resulted in a complete loss of model bias (Supplementary Fig. 4). Such results, whereupon 3D refinement the resolution of a reconstruction worsens or diverges notably from the initial model, are indicative of model bias in the initial docking steps. Therefore, it is important that users perform the guided approach in conjunction with 3D refinement algorithms to specifically test for bias in their own experiments.

## 4. Conclusions

Despite significant advancement in 3D classification and refinement algorithms, the accurate alignment of subcellular structures using STA still remains a challenging endeavor. In many cases, the human eye can more readily identify certain 3D objects and features within noisy subvolumes than even the most advanced computational algorithms. For this reason, segmentation or annotation of features within tomograms is still typically done in a manual fashion within the cryo-ET field using programs such as IMOD (Kremer et al., 1996) or AMIRA (Thermo Fisher

Scientific), although development of automated segmentation procedures has recently been used to identify a subset of biological structures (i.e. filaments, ribosomes, and membranes) within cellular tomograms (Chen et al., 2017). The approach described here takes advantage of the visual expertise of the user to tease out the location of the biological complex of interest, and uses this information to help guide subtomogram averaging algorithms to align common features present within hundreds of individual subvolumes. We show that it is possible to preferentially align low-signal features (i.e. dynein-dynactin complex) relative to signal-dominant structures (i.e. MTs) by simply providing *a priori* information to help “guide” the refinement using RELION auto-refine. We also provide scripts that convert the orientation parameters associated with docked guide structures saved in UCSF Chimera sessions into RELION-ready files for 3D reconstruction and refinement.

While the manual identification of the complex of interest within hundreds of individual subtomograms is user-intensive and laborious, for particularly challenging samples, this strategy may be the only feasible method for structure determination, especially in cases where prior attempts using traditional STA programs have failed. Other work demonstrates that the manual pre-alignment of subvolumes is often an option utilized to facilitate convergence of 3D refinements of structures of interest, particularly in crowded, *in situ* environments (Kiesel et al., 2020). Here we provide an approach that facilitates such manual alignments through manual docking in UCSF Chimera to generate output STAR files that can be used directly in RELION subtomogram averaging package. The basis of this approach in guiding alignment procedures to identify specific features within noisy subvolumes could eventually become an automated procedure using machine-learning algorithms (Chen et al., 2017) or crowd-sourcing approaches (Bruggemann et al., 2018), thus reducing the amount of user-input. However, to our knowledge, this strategy is the only method that has been shown to work in cases where other currently available computational processing tools and algorithms have failed to produce subtomogram averages of challenging macromolecular complexes.

## 5. Code Availability

All scripts can be found at [https://github.com/GrotjahnLab/guided\\_tomo\\_align](https://github.com/GrotjahnLab/guided_tomo_align).

## 6. Data Deposition

Tilt series of microtubule-bound dynein-dynactin complexes were deposited in the Electron Microscopy Public Image Archive with accession ID EMPIAR-10520.

## CRedit authorship contribution statement

**Benjamin Basanta:** Conceptualization, Methodology, Software, Investigation, Writing - review & editing. **Saikat Chowdhury:** Conceptualization, Methodology, Writing - review & editing. **Gabriel C. Lander:** Conceptualization, Software, Writing - review & editing, Supervision. **Danielle A. Grotjahn:** Conceptualization, Methodology, Investigation, Data curation, Writing - original draft, Visualization, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank Jean-Christophe Ducom at the Scripps Research High Performance Computing for computational support, and Bill Anderson



at the Scripps Research electron microscopy facility for microscope support. During this project B.B. was supported by a postdoctoral research fellowship from the Hewitt Foundation for Medical Research, D.G. was supported by a National Sciences Foundation predoctoral fellowship and G.C.L. was supported by the Pew Charitable Trusts as a Pew Scholar and an Amgen Young Investigator award. This project was also supported by the National Institutes of Health (NIH) grant DP2EB020402 and NIH AG061697 to G.C.L. Computational analyses of EM data were performed on shared computing cluster funded by NIH grant S10OD021634 to G.C.L.

Competing financial interests.

The authors declare no competing financial interests.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsb.2020.100041>.

## References

- Albert, S., Wietrzynski, W., Lee, C.-W., Schaffer, M., Beck, F., Schuller, J.M., Salomé, P. A., Plitzko, J.M., Baumeister, W., Engel, B.D., 2020. Direct visualization of degradation microcompartments at the ER membrane. *Proc Natl Acad Sci U S A* 117, 1069–1080.
- Andersen, A.H., Kak, A.C., 1984. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm. *Ultrasound Imaging* 6, 81–94.
- Baumeister, W., 2013. Electron-Cryomicroscopy: From Molecules to Cells. *Biophysical Journal* 104, 206a.
- Bepler, T., Noble, A.J., Berger, B., 2019. Topaz-Denoise: general deep denoising models for cryoEM. *bioRxiv*, 838920.
- Bharat, T.A.M., Scheres, S.H.W., 2016. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols* 11, 2054–2065.
- Bharat, T.A.M., Russo, C.J., Löwe, J., Passmore, L.A., Scheres, S.H.W., 2015. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Subtomogram Averaging. *Structure* 23, 1743–1753.
- Blocker, A., Severin, F.F., Burkhardt, J.K., Bingham, J.B., Yu, H., Olivo, J.C., Schroer, T. A., Hyman, A.A., Griffiths, G., 1997. Molecular requirements for bi-directional movement of phagosomes along microtubules. *The Journal of Cell Biology* 137, 113–129.
- Bruggemann, J., Lander, G.C., Su, A.L., 2018. Exploring applications of crowdsourcing to cryo-EM. *J Struct Biol* 203, 37–45.
- Buchholz, T.-O.J., Mareike; Pigno, Gaia; Jug, Florian 2018. Cryo-CARE: Content-Aware Image Restoration for Cryo-Transmission Electron Microscopy Data. *arXiv*.
- Burt, A., 2020. <https://github.com/alisterburt/dynamo2m>. Zenodo.
- Castano-Diez, D., Kudryashev, M., Stahlberg, H., 2017. Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *J Struct Biol* 197, 135–144.
- Castano-Diez, D., Kudryashev, M., Arheit, M., Stahlberg, H., 2012. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of Structural Biology* 178, 139–151.
- Chen, M., Bell, J.M., Shi, X., Sun, S.Y., Wang, Z., Ludtke, S.J., 2019. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature Methods* 16, 1161–1168.
- Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., Ludtke, S.J., 2017. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nature Methods* 14, 983–985.
- Chowdhury, S., Ketcham, S.A., Schroer, T.A., Lander, G.C., 2015. Structural organization of the dynein-dynactin complex bound to microtubules. *Nat Struct Mol Biol* 22, 345–347.
- Chreifi, G., Chen, S., Metskas, L.A., Kaplan, M., Jensen, G.J., 2019. Rapid tilt-series acquisition for electron cryotomography. *Journal of Structural Biology* 205, 163–169.
- Eisenstein, F., Danev, R., Pilhofer, M., 2019. Improved applicability and robustness of fast cryo-electron tomography data acquisition. *Journal of Structural Biology* 208, 107–114.
- Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J., 2015. Single particle tomography in EMAN2. *Journal of Structural Biology* 190, 279–290.
- Galaz-Montoya, J.G., Hecksel, C.W., Baldwin, P.R., Wang, E., Weaver, S.C., Schmid, M.F., Ludtke, S.J., Chiu, W., 2016. Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. *Journal of Structural Biology* 194, 383–394.
- Gilbert, P., 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology* 36, 105–117.
- Grotjahn, D.A., Lander, G.C., 2019. Setting the dynein motor in motion: New insights from electron tomography. *J Biol Chem* 294, 13202–13217.
- Grotjahn, D.A., Chowdhury, S., Xu, Y., McKenney, R.J., Schroer, T.A., Lander, G.C., 2018. Cryo-electron tomography reveals that dynactin recruits a team of dyneins for processive motility. *Nat Struct Mol Biol* 25, 203–207.
- Hagen, W.J.H., Wan, W., Briggs, J.A.G., 2017. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of Structural Biology* 197, 191–198.
- Heumann, J.M., Hoenger, A., Mastronarde, D.N., 2011. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *Journal of Structural Biology* 175, 288–299.
- Himes, B.A., Zhang, P., 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* 15, 955–961.
- Hrabe, T., Chen, Y., Pfeffer, S., Kuhn Cuellar, L., Mangold, A.-V., Förster, F., 2012. PyTom: A python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *Journal of Structural Biology* 178, 177–188.
- Khoshouei, M., Pfeffer, S., Baumeister, W., Foerster, F., Danev, R., 2017. Subtomogram analysis using the Volta phase plate. *Journal of Structural Biology* 197, 94–101.
- Kiesel, P., Viar, G.A., Tsoy, N., Marasini, R., Honigsmann, A., Pigno, G., 2020. The molecular structure of primary cilia revealed by cryo-electron tomography. *bioRxiv*, 2020.2003.2020.000505.
- Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116, 71–76.
- Markova, E.A., Zanetti, G., 2019. Visualizing membrane trafficking through the electron microscope: cryo-tomography of coat complexes. *Acta Crystallogr D Struct Biol* 75, 467–474.
- Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* 152, 36–51.
- P.P. Navarro H. Stahlberg D. Castaño-Díez Protocols for Subtomogram Averaging of Membrane Proteins in the Dynamo Software Package *Front Mol Biosci* 5 2018 82 82.
- Nicasro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M.E., McIntosh, J.R., 2006. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* 313, 944–948.
- Obr, M., Schur, F.K.M., 2019. In: Chapter Five - Structural analysis of pleomorphic and asymmetric viruses using cryo-electron tomography and subtomogram averaging. *Advances in Virus Research*, Academic Press, pp. 117–159.
- Oikonomou, C.M., Jensen, G.J., 2017. Cellular Electron Cryotomography: Toward Structural Biology In Situ. *Annual Review of Biochemistry* 86, 873–896.
- Oikonomou, C.M., Jensen, G.J., 2019. Electron Cryotomography of Bacterial Secretion Systems. *Microbiol Spectr* 7. <https://doi.org/10.1128/microbiolspec.PSIB-0019-2018>.
- Orlov, I., Myasnikov, A.G., Andronov, L., Natchiar, S.K., Khatler, H., Beinstainer, B., Menetret, J.F., Hazemann, I., Mohideen, K., Tazibt, K., Tabaroni, R., Kratzat, H., Djabeur, N., Bruxelles, T., Raivoniaina, F., Pompeo, L.D., Torchy, M., Billas, I., Urzhumtsev, A., Klalholz, B.P., 2017. The integrative role of cryo electron microscopy in molecular and cellular structural biology. *Biol Cell* 109, 81–93.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605–1612.
- Pfeffer, S., Dudek, J., Zimmermann, R., Forster, F., 2016. Organization of the native ribosome-translocon complex at the mammalian endoplasmic reticulum membrane. *Biochim Biophys Acta* 1860, 2122–2129.
- Schur, F.K.M., Obr, M., Hagen, W.J.H., Wan, W., Jakobi, A.J., Kirkpatrick, J.M., Sachse, C., Kräusslich, H.-G., Briggs, J.A.G., 2016. An atomic model of HIV-1 capsid-SPI reveals structures regulating assembly and maturation. *Science* 353, 506–508.
- Shatsky, M., Hall, R.J., Brenner, S.E., Glaeser, R.M., 2009. A method for the alignment of heterogeneous macromolecules from electron microscopy. *J Struct Biol* 166, 67–78.
- Suloway, C., Shi, J., Cheng, A., Pulokas, J., Carragher, B., Potter, C.S., Zheng, S.Q., Agard, D.A., Jensen, G.J., 2009. Fully automated, sequential tilt-series acquisition with Leginon. *Journal of Structural Biology* 167, 11–18.
- Tegunov, D., Cramer, P., 2019. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* 16, 1146–1152.
- Tegunov, D., Xue, L., Dienemann, C., Cramer, P., Mahamid, J., 2020. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.7 Å inside cells. *bioRxiv*, 2020.2006.2005.136341.
- Turoňová, B., Schur, F.K.M., Wan, W., Briggs, J.A.G., 2017. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of Structural Biology* 199, 187–195.
- Turoňová, B., Hagen, W.J.H., Obr, M., Kräusslich, H.-G., Beck, M., 2019. Benchmarking tomographic acquisition schemes for high-resolution structural biology. *bioRxiv* 317, 742254.
- Urnävicius, L., Zhang, K., Diamant, A.G., Motz, C., Schlager, M.A., Yu, M., Patel, N.A., Robinson, C.V., Carter, A.P., 2015. The structure of the dynactin complex and its interaction with dynein. *Science* 347, 1441–1446.
- Wietrzynski, W., Schaffer, M., Tegunov, D., Albert, S., Kanazawa, A., Plitzko, J.M., Baumeister, W., Engel, B.D., 2020. Charting the native architecture of Chlamydomonas thylakoid membranes with single-molecule precision. *Elife* 9, e53740.