Journal of Visualized Experiments

A robust single-particle cryo-Electron Microscopy (cryo-EM) processing workflow with cryoSPARC, RELION, and Scipion.

--Manuscript Draft--

Article Type:	Invited Methods Collection - JoVE Produced Video
Manuscript Number:	JoVE63387R2
Full Title:	A robust single-particle cryo-Electron Microscopy (cryo-EM) processing workflow with cryoSPARC, RELION, and Scipion.
Corresponding Author:	Arek Kulczyk UNITED STATES
Corresponding Author's Institution:	
Corresponding Author E-Mail:	arek.kulczyk@rutgers.edu
Order of Authors:	Megan DiIorio
	Arkadiusz Kulczyk
Additional Information:	
Question	**Response**
Please specify the section of the submitted manuscript.	Biochemistry
Please indicate whether this article will be Standard Access or Open Access.	Standard Access ($1400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Piscataway, NJ, USA
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please confirm that you have read and agree to the terms and conditions of the video release that applies below:	I agree to the Video Release
Please provide any comments to the journal here.	

TITLE:

A Robust Single-Particle Cryo-Electron Microscopy (cryo-EM) Processing Workflow with cryoSPARC, RELION, and Scipion.

AUTHORS AND AFFILIATIONS:

Megan C. DiIorio[1], Arkadiusz W. Kulczyk[1,2*]

[1]Institute for Quantitative Biomedicine, Rutgers University, Piscataway, New Jersey, USA
[2]Department of Biochemistry & Microbiology, Rutgers University, New Brunswick, New Jersey, USA

Email addresses of co-authors:
Megan C. DiIorio (mcd218@iqb.rutgers.edu)
Arkadiusz W. Kulczyk (arek.kulczyk@rutgers.edu)

*Corresponding Author:
Arkadiusz W. Kulczyk (arek.kulczyk@rutgers.edu)

KEYWORDS:

cryo-electron microscopy, cryo-EM, single-particle analysis, SPA, cryoSPARC, RELION, Scipion, image processing, structure calculation, AAV

SUMMARY:

This article describes how to effectively utilize three cryo-EM processing platforms, i.e., cryoSPARC v3, RELION-3, and Scipion 3, to create a single and robust workflow applicable to a variety of single-particle data sets for high-resolution structure determination.

ABSTRACT:

Recent advances in both instrumentation and image processing software have made single-particle cryo-electron microscopy (cryo-EM) the preferred method for structural biologists to determine high-resolution structures of a wide variety of macromolecules. Multiple software suites are available to new and expert users for image processing and structure calculation, which streamline the same basic workflow: movies acquired by the microscope detectors undergo correction for beam-induced motion and contrast transfer function (CTF) estimation. Next, particle images are selected and extracted from averaged movie frames for iterative 2D and 3D classification, followed by 3D reconstruction, refinement, and validation. Because various software packages employ different algorithms and require varying levels of expertise to operate, the 3D maps they generate often differ in quality and resolution. Thus, users regularly transfer data between a variety of programs for optimal results. This paper provides a guide for users to navigate workflow across the popular software packages: cryoSPARC v3, RELION-3, and Scipion 3 to obtain a near-atomic resolution structure of the adeno-associated virus (AAV). We first detail an image processing pipeline with cryoSPARC v3, as its efficient algorithms and easy-to-use GUI allow users to quickly arrive at a 3D map. In the next step, we use PyEM and in-house scripts to convert and transfer particle coordinates from the best quality 3D reconstruction obtained in

cryoSPARC v3 to RELION-3 and Scipion 3 and recalculate 3D maps. Finally, we outline steps for further refinement and validation of the resultant structures by integrating algorithms from RELION-3 and Scipion 3. In this article, we describe how to effectively utilize three processing platforms to create a single and robust workflow applicable to a variety of data sets for high-resolution structure determination.

INTRODUCTION:

Cryo-electron microscopy (cryo-EM) and single-particle analysis (SPA) enable structure determination of a wide variety of biomolecular assemblies in their hydrated state, helping to illuminate the roles of these macromolecules in atomic detail. Improvements in microscope optics, computer hardware, and image processing software have made it possible to determine structures of biomolecules at resolution reaching beyond 2 Å[1-3]. More than 2,300 cryo-EM structures were deposited in the Protein Data Bank (PDB) in 2020, compared to 750 structures in 2014[4], indicating that cryo-EM has become a method of choice for many structural biologists. Here, we describe a workflow combining three different SPA programs for high-resolution structure determination (**Figure 1**).

The goal of SPA is to reconstruct 3D volumes of a target specimen from noisy 2D images recorded by the microscope detector. Detectors collect images as movies with individual frames of the same field of view. In order to preserve the sample, frames are collected with a low electron dose and thus have a poor signal-to-noise ratio (SNR). Additionally, electron exposure can induce motion within the vitrified cryo-EM grids, resulting in image-blurring. To overcome these issues, frames are aligned to correct for beam-induced motion and averaged to yield a micrograph with an increased SNR. These micrographs then undergo Contrast Transfer Function (CTF) estimation to account for the effects of defocus and aberrations imposed by the microscope. From the CTF-corrected micrographs, individual particles are selected, extracted, and sorted into 2D class averages representing different orientations adopted by the specimen in vitreous ice. The resultant homogeneous set of particles is used as input for *ab initio* 3D reconstruction to generate a coarse model or models, which are then iteratively refined to produce one or more high-resolution structures. After reconstruction, structural refinements are performed to further improve the quality and resolution of the cryo-EM map. Finally, either an atomic model is directly derived from the map, or the map is fitted with atomic coordinates obtained elsewhere.

Different software packages are available to accomplish the tasks outlined above, including Appion[5], cisTEM[6], cryoSPARC[7], EMAN[8], IMAGIC[9], RELION[10], Scipion[11], SPIDER[12], Xmipp[13], and others. While these programs follow similar processing steps, they employ different algorithms, for example, to pick particles, generate initial models, and refine reconstructions. Additionally, these programs require a varying level of user knowledge and intervention to operate, as some depend on the fine-tuning of parameters that can act as a hurdle for new users. These discrepancies often result in maps with inconsistent quality and resolution across platforms[14], prompting many researchers to use multiple software packages to refine and validate results. In this article, we highlight the use of cryoSPARC v3, RELION-3, and Scipion 3 to obtain a high-resolution 3D reconstruction of AAV, a widely used vector for gene therapy[15]. The

aforementioned software packages are free to academic users; cryoSPARC v3 and Scipion 3 require licenses.

PROTOCOL

1. Creating a new cryoSPARC v3 project and importing data.

NOTE: Data was acquired at Oregon Health and Science University (OHSU) in Portland using a 300 kV Titan Krios electron microscope equipped with a Falcon 3 direct electron detector. Images were collected in a counting mode with a total dose of 28.38 $e^{-}/Å^2$ fractioned across 129 frames, and defocus range from -0.5 µm to -2.5 µm, at a pixel size of 1.045 Å using EPU. The sample of AAV-DJ was provided by the staff of OHSU.

1.1. Open cryoSPARC v3 in a web browser and click the **Projects** header. Select **+ Add** to create a new project. Title the project accordingly and provide a path to an existing directory where jobs and data will be saved.

1.2. Create a workspace for the project by opening the project, clicking **+ Add**, and selecting **New Workspace**. Title the workspace and click on **Create**.

1.3. Navigate to the new workspace and open the **Job Builder** on the right panel. This tab displays all functions available in cryoSPARC v3. Click on **Import Movies** and provide the movies path, gain reference file path, and set acquisition parameters as follows: **Raw Pixel Size** 1.045 Å, **Accelerating Voltage** 300 kV, **Spherical Aberration** 2.7 mm, **Total Exposure Dose** 28.38 e^{-}/Å^2.

1.4. Click on **Queue**, select a lane to run the job and a workspace, and click on **Create**.

NOTE: The acquisition parameters are sample and microscope dependent.

2. CryoSPARC v3 - movie alignment and CTF Estimation.

2.1. Open **Patch Motion Correction (Multi)**. This job requires the movies imported in step 1.3 as input. Open the import movies job card in the workspace and drag the **Imported_movies** output to the movies placeholder on the new job. **Queue** the job.

NOTE: For more information about the cryoSPARC methods outlined in this article, see the cryoSPARC tutorial[16].

2.2. To perform CTF estimation, open **Patch CTF Estimation (Multi)**. Input the micrographs generated in step 2.1 and **Queue** the job.

2.3. To inspect the averaged and CTF-corrected micrographs and select a subset for further processing, open **Curate Exposures** and input the exposures obtained in step 2.2. **Queue** the job.

2.4. After the job enters **Waiting** mode, click on the **Interaction** tab on the job card, adjust parameter thresholds, and accept or reject individual micrographs for further processing. Accept micrographs with well-matched estimated and experimental CTFs (**Figure 2**) and discard those with high astigmatism, poor CTF fit, and thick ice.

2.5. While processing the current data, set the upper threshold of Astigmatism to 400 Å, CTF fit resolution to 5 Å, and relative ice thickness to 2. Click on **Done** to select the micrographs for downstream processing.

3. CryoSPARC v3 - manual and template-based particle picking.

3.1. Open **Manual Picker**, input the accepted exposures from steps 2.4–2.5, and **Queue** the job. Click on the **Interactive** tab, set the **Box Size (px)** to 300, and click on a few hundred particles across multiple micrographs and avoid selecting overlapping particles. Here, 340 particles across 29 micrographs were selected. When finished, click on **Done Picking! Extract Particles**.

NOTE: This protocol uses manual particle picking to generate templates for automatic selection. However, other methods are also available[17].

3.2. To generate templates for automated particle picking, click on **2D Classification** and input the particle picks generated in step 3.1. Change the number of **2D Classes** to 10 and **Queue** the job.

3.3. Open **Select 2D classes**. Input the particles and class averages obtained in step 3.2 and click on the **Interactive** tab. Select representative 2D classes with good SNR and click on **Done**.

NOTE: The class averages reflect different particle views. Select class averages that reflect each view. The goal is to produce well-defined templates representing different views of the specimen for automated picking.

3.4. Open **Template Picker** and input the 2D classes selected in step 3.3 and micrographs from steps 2.4-2.5. Set the **Particle Diameter** (Å) to 220 Å and **Queue** the job.

3.5. To inspect the automated picks, open **Select Particle Picks**, input the particles and micrographs generated in step 3.4, and **Queue** the job.

3.6. On the **Select Particle Picks** job card, click on the **Interactive** tab and set the **Box size (px)** to 300. Click on an individual micrograph, adjust the lowpass filter until particles are clearly visible, and set the **Normalized Cross Correlation** (**NCC) Threshold** to 0.41 and **Power Threshold** between 54000 and 227300**.**

3.7. Inspect several micrographs and, if needed, adjust thresholds such that most particles are selected without including false positives. When finished, click **Done Picking! Extract Particles.**

NOTE: True particles typically have a high NCC and power score, indicating they are similar to the template and have a high SNR, respectively.

3.8. Open **Extract from Micrographs** and input the micrographs and particles from step 3.7. Set the **Extracted Box Size (px)** to 300 and **Queue** the job.

4. CryoSPARC v3 - 2D Classification.

4.1. Click on **2D Classification** and input the extracted particles from step 3.8. Set the **Number of 2D classes** to 50 and **Queue** the job.

4.2. To select the best 2D classes for further processing, open **Select 2D classes**. Input the particles and class averages obtained in step 4.1. Click on the **Interactive** tab and choose 2D classes based upon the resolution and the number of particles in the class (**Figure 3**). Do not select classes containing artifacts. After selecting, click on **Done**.

NOTE: Usually, multiple rounds of 2D classification are required to remove particles, which do not converge into distinct, well-defined classes. Run as many rounds of 2D classification as needed to remove such particles from the data set (**Figure 3**).

5. CryoSPARC v3 - *ab-initio* reconstruction and homogeneous refinement.

5.1. To generate an initial 3D volume, open **Ab-initio Reconstruction** and input the particles obtained in step 4.2 or from the final 2D classification. Adjust **Symmetry** to icosahedral. **Queue** the job.

NOTE: Symmetry is sample-dependent and should be changed accordingly. If unknown, use C1 symmetry.

5.2. Open **Homogeneous Refinement**. Input the volume from step 5.1 and particles from 4.2 or the final 2D classification. Change the **Symmetry** and **Queue** the job**.** When the job is finished, inspect the Fourier Shell Correlation (FSC) curve and download the volume to examine in UCSF Chimera[18].

6. Exporting particle coordinates from cryoSPARC v3 and importing them to RELION-3 using PyEM.

NOTE: Particle coordinates carry information about the location of individual particles in each micrograph. Transfer of coordinates instead of particle stacks to RELION-3 allows for running refinement steps, which otherwise would not be available. For example, particle polishing requires access to initial movie frames. Hence, prior to exporting particle coordinates from cryoSPARC v3 to RELION-3, import movies and perform motion correction and CTF estimation in RELION-3. See the RELION-3 tutorial[19] for details.

6.1. Navigate to the RELION-3 project directory and launch RELION-3.

6.2. Open **Import** from the job-type browser and specify the path to the movies and acquisition parameters as in step 1.3.

6.3. To perform motion correction, use UCSF MotionCor2[20] through the RELION-3 GUI, open **Motion Correction** and set the default parameters as in the UCSF MotionCor2 manual[21]. Input the path to the movies imported in step 6.2. On the **Motion** tab, specify the path to motioncorr2 executable.

NOTE: MotionCor2 can be run in parallel using multiple GPUs.

6.4. Perform CTF estimation using CTFFIND-4.1[22] through the RELION-3 GUI. Open **CTF Estimation** and input the micrographs.star generated in step 6.3. On the **CTFFIND-4.1** tab, specify the path to CTFFIND-4.1 executable and set parameters as in the RELION-3.1 tutorial[19].

6.5. In order to import particle stacks from cryoSPARC v3 to RELION-3, they first must be exported from cryoSPARC v3. In cryoSPARC v3, open the job card of the **Select 2D class** job from step 4.2 or the final 2D classification. On the **Details** tab, click on **Export Job**. **Export job** outputs the particles_exported.cs file.

6.6. Prior to importing particle coordinates from cryoSPARC v3 to RELION-3, the particles_exported.cs file from step 6.5 must be converted to .star format. Using PyEM[23], convert the particles_exported.cs file to .star format by executing the following command: csparc2star.py particles_exported.cs particles_exported.star

6.7. In RELION-3 click on the **Manual Picking** tab and on the **I/O** tab, input the micrographs from CTF refinement described in step 6.4. On the **Display** tab, input the following parameters: **Particle diameter (A):** 220**, Lowpass Filter (A):** -1**, Scale for CTF image:** 0.5. **Run** the job. A directory called ManualPick is generated in the RELION-3 home folder.

NOTE: This step is performed to create a manual picking folder structure in RELION-3. While running manual picking, a single .star file containing coordinates of picked particles is created for each averaged micrograph used for picking in the RELION-3 GUI.

6.8. Navigate to the folder containing the particles_exported.star file from step 6.6 and run a home-written script producing a single manualpick.star file for each averaged micrograph used for picking of cryo-SPARC v3 particles, which contributed to the final 2D classification exported in step 6.5. The resultant coordinate files are saved in the ManualPick/Movies folder.

6.9. Return to RELION-3 and re-open the **Manual Picking** job. Click on **Continue.** This will display particles previously picked in cryoSPARC v3 in the RELION-3 GUI. Inspect a few micrographs to verify if the transfer of particle coordinates has been accomplished and if particles are properly selected.

7. RELION-3 - Particle extraction and 2D classification.

7.1. Click on **Particle Extraction**. On the **I/O** tab, input the CTF corrected micrographs from step 6.4 and coordinates from step 6.9. Click on the **Extract** tab and change the **Particle Box Size (pix)** to 300. **Run** the job.

7.2. Perform 2D classification to further clean the particle set generated in cryoSPARC v3 to achieve a higher-resolution reconstruction. Click on **2D Classification** and on the **I/O** tab, input the particles.star file generated in step 7.1. On the **Optimisation** tab, set the **Number of Classes** to 50 and **Mask Diameter (A)** to 280. **Run** the job.

NOTE: The mask should encompass the entire particle.

7.3. To choose the best 2D classes, click on the **Subset Selection** method, input the _model.star file from step 7.2, and **Run** the job. Select classes as described in step 4.2.

7.4. Repeat steps 7.2 and 7.3 to remove non-converging particles.

8. RELION-3 - 3D refinement, mask Creation, and post-processing.

8.1. Use the map generated in cryoSPARC v3 (step 5.2) as an initial model for 3D refinement in RELION-3. Select the **Import** method and set the following parameters on the **I/O** tab: **Import Raw Movies/Micrographs**: No, **Raw Input Files:** Movies/*.mrc.

8.2. Supply the MTF file and input the movie acquisition parameters as described in step 1.3. On the **Others** tab, select the cryoSPARC v3 map as the input file, change **Node Type** to 3D reference (.mrc), and **Run** the job.

8.3. Select the **3D Auto-Refine** and on the **I/O** tab, set **Input Images** as the particles.star file from step 7.3 or the last selection job. Give the cryoSPARC v3 reconstruction as the **Reference Map**. Click on the **Reference** tab and change **Initial Low-Pass filter (Å)** to 50 and **Symmetry** to icosahedral. On the **Optimisation** tab, change the **Mask Diameter (Å)** to 280 and **Run** the job.

8.4. After the run is finished, open run_class001.mrc in UCSF Chimera.

8.5. In UCSF **Chimera**, click on **Tools** and under **Volume Data,** select **Volume Viewer**. This will open a new window to adjust volume settings. Change the **Step** to 1 and adjust the slider until reaching the level value where the map has no noise. Record this value, as it will be used for mask creation in the next step.

8.6. The map produced from auto-refinement does not reflect the true FSC, as noise from the surrounding solvent lowers the resolution. Before post-processing, create a mask to distinguish the specimen from the solvent region.

8.6.1. Click on **Mask Creation** and input run_class001.mrc from step 8.3.

8.6.2. Click on the **Mask** tab and adjust parameters as follows: **Lowpass Filter Map (Å)**: 10, **Pixel Size (Å)**: 1.045, **Initial Binarization Threshold:** the level value obtained in step 8.5, **Extend Binary Map this many Pixels**: 3, and **Add a Soft-Edge of this many Pixels**: 3. **Run** the job.

8.7. Examine the mask in UCSF Chimera. If the mask is too tight, increase **Extend Binary Map this many Pixels** and/or **Add a Soft-Edge of this many Pixels. I**t is important to create a mask with soft edges, as a sharp mask may lead to overfitting.

8.8. Click on **Post-Processing** and on the **I/O** tab, input the half-maps created in step 8.3 and mask from 8.6. Set **Calibrated Pixel Size** to 1.045 Å. On the **Sharpen** tab, input the following: **Estimate B-Factor Automatically?**: Yes, **Lowest Resolution for Auto-B Fit (A)**: 10, **Use Your Own B-Factor?**: No. On the **Filter** tab, set **Skip Fsc-Weighting?** to No. **Run** the job.

9. RELION-3 - Polishing training and particle polishing.

9.1. Before correcting for per-particle beam-induced motion, first use the training mode to identify optimal motion tracks for the data set. Open **Bayesian Polishing** and on the **I/O** tab, input the motion-corrected micrographs from step 6.3, particles from step 8.3, and postprocess .star file from step 8.8. Click the **Training** tab and set the following parameters: **Train Optimal Parameters**: Yes, **Fraction of Fourier Pixels for Testing**: 0.5, **Use this many Particles**: 5000. **Run** the job.

NOTE: This script will produce opt_params_all_groups.txt file in the RELION-3 Polish folder containing optimized polishing parameters required for executing the following step.

9.2. Once the training job has finished, click on **Bayesian Polishing**. Click on the **Training** tab and set **Train Optimal Parameters?** to No. Select the **Polish** tab and in **Optimised Parameter File** specify the path to the opt_params_all_groups.txt file from step 9.1. Click on **Run**.

9.3. Repeat 3D refinement (step 8.3) and post-processing (step 8.8) with a set of polished particles.

10. RELION-3 - CTF and per-particle refinements.

10.1. To estimate higher order aberrations, open **CTF Refinement** and, on the **I/O** tab under **Particles,** select the path to the .star file containing polished particles from the recent Refine 3D job (run_data.star).

10.1.1. Under **Postprocess Star File**, set the path to the output from the latest post-processing job (step 9.3).

10.1.2. Select the **Fit** tab and set the following parameters: **Estimate (Anisotropic Magnification)**: No, **Perform CTF Parameter Fitting?** No, **Estimate Beamtilt**: Yes, **Also Estimate Trefoil?** Yes, **Estimate 4th Order Abberations?** Yes. **Run** the job.

10.2. Repeat step 10.1 using as input **Particles (from Refine3D)** generated in the previous job (particles_ctf_refine.star). On the **Fit** tab, change **Estimate (Anisotropic Magnification)** to Yes and **Run** the job.

10.3. Repeat step 10.2 using as input **Particles (from Refine3D)** produced in the previous job (particles_ctf_refine.star). On the **Fit** tab, set the following parameters: **Estimate (Anisotropic Magnification)**: No, **Perform CTF Parameter Fitting?**: Yes, **Fit Defocus?**: Per-particle, **Fit Astigmatism?** Per-micrograph, **Fit B-factor?**: No, **Fit Phase-Shift**: No, **Estimate Beamtilt?**: No, **Estimate 4th Order Aberrations?**: No. **Run** it.

NOTE: Given the particle has sufficient contrast, the **Fit Astigmatism?** tab can be set to **Per-particle**. For this dataset, **Per-Particle** astigmatism refinement did not improve the quality and resolution of the map.

10.4. Repeat 3D refinement with the particles from step 10.3 and on the I/O tab, set **Use Solvent-Flattened FSCs?** to yes. When finished running, execute a post-processing job (step 8.8) and examine the map in UCSF Chimera (step 5.2).

11. Transferring RELION-3 particle coordinates and 3D map to Scipion 3.

11.1. To further refine and validate the RELION-3 map, first import the volume and particles from the last post-processing job (step 10.4) to Scipion 3. Launch Scipion 3 and create a new project.

11.2. On the left **Protocols** panel, select the **Imports** drop-down and click on **Import Particles**. Change the following parameters: **Import From**: RELION-3, **Star File**: postprocess.star, and specify acquisition parameters as in step 1.3. Click on **Execute**.

11.3. Click on the **Imports** drop-down and select **Import Volumes**. Under **Import From** give the path to the RELION-3 map. Change **Pixel Size (Sampling Rate) Å/px** to 1.045 and **Execute**.

12. Scipion 3 - High resolution refinement.

12.1. First, perform a global alignment. Select the **Refine** drop-down on the **Protocols** panel and click on **Xmipp3 – highres**[24]. Input the imported particles and volumes from steps 11.2 and 11.3 as **Full-Size Images** and **Initial Volumes**, respectively and set the **Symmetry Group** to icosahedral. On the **Image Alignment** tab under **Angular Assignment**, choose **Global** and set the **Max Target Resolution** to 3 Å, and **Run** the job.

12.2. When the job is finished, click on **Analyze Results**. In the new window, click on the **UCSF Chimera** icon to examine the refined volume. Additionally, click on **Display Resolution Plots (FSC)** to see how the FSC has changed after the refinement, as well as **Plot Histogram with Angular Changes** to see if the Euler angle assignments have changed.

NOTE: Depending on the resolution of the input RELION-3 structure, this step may be repeated several times with different values set for the **Max Target Resolution** under the **Angular Assignment** tab. For more information see the Scipion tutorial[25].

12.3. Repeat step 12.1 with a local alignment. Copy the previous job and change **Select Previous Run** to **Xmipp3 - highres Global**. On the **Angular Assignment** tab, change **Image Alignment** to **Local**. Set the **Max Target Resolution** to 2.1 Å.

12.4. Examine the refined map in UCSF Chimera and analyze the change in FSC and angular assignments (step 12.2). Repeat local refinement until resolution does not improve and the angular assignments have converged, adjusting **Max Target Resolution** as needed.

12.5. The output map from Scipion 3 can be additionally density-modified and sharpened in Phenix[26].

13. Scipion 3 – Map validation.

13.1. Examine the local resolution of the final map generated in **Xmipp3 - highres**. Open **Xmipp - local MonoRes**[27] and input the final volume from the previous job and mask generated in step 8.6. Set **Resolution Range** from 1 to 6 Å with a 0.1 Å interval and **Execute** the job.

13.2. When finished running, click on **Analyze Results** and examine the resolution histogram and volume slices colored by resolution.

13.3. To see if particles are well-aligned, open **Multireference Alignability**[28] and input the particles and volume from step 12.3. Click on **Analyze Results** to display the validation plot. Ideally, all points should be clustered around (1.0,1.0).

13.4. Open **Xmipp3 – Validate Overfitting**. Input the particles and volumes from step 12.3. When finished running, analyze results and inspect the overfitting plot. Crossing of the aligned Gaussian noise and aligned particles curves indicates overfitting.

REPRESENTATIVE RESULTS:

We have presented a comprehensive SPA pipeline to obtain a high-resolution structure using three different processing platforms: cryoSPARC v3, RELION-3, and Scipion 3. **Figures 1** and **Figure 4** summarize the general processing workflow, and **Table 1** details refinement protocols. These protocols were used during refinements of a 2.3 Å structure of AAV, achieving near Nyquist resolution.

Movies were first imported to cryoSPARC v3 and subsequently motion- and CTF-corrected to generate averaged micrographs. When selecting micrographs for further processing, it is important to choose those with a good CTF-fit and low astigmatism (**Figure 2**), as including poor-quality micrographs can hinder later processing stages, resulting in a lower resolution reconstruction. 27,364 particles were then picked and extracted from the selected micrographs. Because the diameter of the AAV is approximately 220 Å and pixel size is 1.045 Å, a box size of 300 px was used. Next, iterative 2D classification was used to remove artifacts and particles not converging to stable classes. Examples of selected and excluded 2D class averages are presented in **Figure 3**. It is also important to note that class averages reflecting different conformations of the specimen should be refined separately to yield multiple 3D reconstructions. In such a case, multiple *ab initio* starting volumes should be calculated. Here, 26,741 particles were selected and used for *ab initio* modeling and homogeneous refinement of a single 2.9 Å structure.

After transferring coordinates of particles picked in cryoSPARC v3 to RELION-3, we carried out four additional rounds of 2D classification until the data set converged to stable 2D classes. The above-described 2D classification removed an additional 3,154 particles from the data set. Using the structure generated in cryoSPARC v3 as an initial model, 3D refinement in RELION-3 produced a structure with a nearly equivalent resolution of 2.95 Å. Subsequent structural refinements, which included per-particle motion correction and CTF refinements, decreased the resolution to 2.61 Å. A complete list of refinements we performed is presented in **Table 1**. The volume calculated in RELION-3 was then further refined in Scipion using multiple rounds of high-resolution refinement (**Xmipp3 – highres**). During subsequent rounds of refinement, an additional 3,186 particles were removed from the data set, resulting in a final set of 20,401 particles, which produced a 2.3 Å reconstruction of AAV (**Figure 5** and **Figure 6**). Thus, given the pixel size of 1.045 Å, our refinements have nearly reached the Nyquist limit. FSC curves representing structures calculated using each program are shown in **Figure 6**. These FSC curves indicate the resolution increase throughout the workflow. Because resolution may vary from point to point in the map, it is often more appropriate to present the distribution of local resolution estimates in the map rather than reporting the resolution estimate according to a single criterion (e.g., 0.143 criterion) from the FSC curve. Thus, we performed such analysis using **Xmipp – MonoRes** in Scipion 3. **Figure 7** shows a comparison of local resolution estimates for maps obtained with cryoSPARC v3 and Scipion 3. Resolution estimates at four different slices through the structures (**Figure 7A**,**B**) and resolution histograms (**Figure 7C**) clearly demonstrate the incremental improvement in local resolution between the maps throughout the workflow. The FSC curve calculated using the program **Xmipp3 - highres** in Scipion 3 indicates the Nyquist limit has been reached (**Figure 6**), suggesting the resolution estimate is very likely limited by undersampling[29]. However, MonoRes analysis presented in **Figure 7C**, along with a careful analysis of the EM map and map fitting with atomic coordinates of AAV (**Figure 5**) suggests that a more adequate resolution estimate for the map is 2.3 Å. A similar strategy reconciling the FSC and MonoRes resolution estimates have been presented earlier[24,25]. Because resolution estimates can be influenced by the mask used during refinement steps, it is important to ensure the mask does not exclude any part of the density. The mask used in this study overlapped with the 3D reconstructions is presented in **Figure 7D**. The gradual increase in resolution in the

presented workflow highlights the advantage of utilizing algorithms from multiple SPA software packages to achieve a high-quality and high-resolution 3D reconstruction.

In-situ model building or fitting the map with a pre-existing atomic model can serve as the quality check for the calculated structure. We have visualized the final map in UCSF Chimera and fitted the map with a previously published atomic model (PDB ID: 7kfr)[30]. **Figure 5** shows regions of the cryo-EM map fitted with atomic coordinates of AAV. Well-defined EM densities allow for fitting side-chains of individual amino acids, water molecules, and magnesium ions and confirm the agreement of the cryo-EM map with the atomic model.

FIGURE AND TABLE LEGENDS:

Figure 1: Complete SPA workflow across cryoSPARC v3, RELION-3, Scipion 3, and Phenix 1.18. Steps completed in cryoSPARC v3, RELION-3, Scipion 3, and Phenix 1.18 are denoted with purple, orange, green, and grey boxes, respectively. The time required for completion of each step using the processing server equipped with 8 GPUs, 40 CPUs and 750 GB of RAM is specified in each individual box.

Figure 2: Selection of micrographs for downstream processing in cryoSPARC v3. (**A**) Micrographs with well-matched estimated and experimental Thon rings were used for further processing, while those with high astigmatism and poor fit (**B**) were discarded. Micrographs with CTF-fit above 5 Å, astigmatism over 400 Å, and relative ice thickness below 2 were removed from further processing, i.e., 70/395 micrographs.

Figure 3: Selecting 2D classes. 2D class averages containing well-defined classes are selected (**A**), and those with low-resolution, noise, and partial particles are rejected (**B**).

Figure 4: Workflow and representative results for AAV processing across cryoSPARC v3, RELION-3, and Scipion 3. Steps completed in cryoSPARC v3, RELION-3, and Scipion 3 are denoted with purple, orange, and green arrows, respectively.

Figure 5: High-resolution structure of AAV shows well-defined EM densities representing different secondary structure elements and individual amino acid side chains. (**A**) A final map of AAV. (**B**) A part of the map representing beta sheets fitted with atomic coordinates of AAV (PDB ID 7fkr)[30]. (**C**) Map densities representing individual amino acids. From left to right: arginine, phenylalanine, and tryptophan. (**D**) High-resolution features of the map include water molecules presented in red and magnesium ions presented as green spheres. Mg^{2+} ion displayed in the figure is coordinated by histidine (left) and arginine residues.

Figure 6: FSC curves from cryoSPARC v3, RELION-3, and Scipion 3 show increasing resolution across the workflow. Despite the FSC curve calculated using the program **Xmipp3 – highres** in Scipion 3 indicates the Nyquist limit has been reached, suggesting the resolution estimate is limited by undersampling[29], a more adequate analysis of the map resolution is presented in **Figure 7** and discussed in the Representative Results section[24,25].

Figure 7: Validating the final reconstruction in Scipion 3 using Xmipp - MonoRes. Resolution of the map is better described by presenting local resolution distributions rather than a single resolution estimate according to a single criterion from the FSC curve. (**A–B**) Panels A and B show different slices from maps generated in cryoSPARC v3 and Scipion 3, respectively. (**C**) Histograms demonstrating a systematic increase in local resolution for maps calculated in cryoSPARC v3 (pink bars) and Scipion 3 (blue bars). (**D**) The mask (gray) used for local resolution calculations contain all parts of the AAV densities refined in both programs.

Table 1: Refinements implemented throughout the workflow. Whether certain refinements are applicable to a specific project depends on data quality and acquisition parameters. For example, the Ewald sphere curvature correction can be applied for maps that already have high resolution.

DISCUSSION:

In this article, we present a robust SPA workflow for cryo-EM data processing across various software platforms to achieve high-resolution 3D reconstructions (**Figure 1**). This workflow is applicable to a wide variety of biological macromolecules. The subsequent steps of the protocol are outlined in **Figure 4**, including movie pre-processing, particle picking and classification, and multiple methods for structure refinements (**Table 1**) and validation. Processing steps in cryoSPARC v3, RELION-3, and Scipion 3 have been presented, as well as methods for transferring data between the software packages. We have shown the intermediate structures obtained throughout the protocol with increasing resolution (**Figure 4**, **Figure 6** and **Figure 7**).

While the methods outlined in this manuscript can be used for structure determination of different proteins and biological assemblies, it is important to note that AAV is an ideal candidate for high-resolution structure determination by cryo-EM and SPA, as its large size produces high contrast in the microscope and icosahedral symmetry yields particles with 60-fold subunit redundancy. Obtaining high-resolution reconstructions become increasingly difficult for small (i.e., less than 100 kD), dynamic, and heterogeneous samples. In order to successfully execute this protocol, it is critical that many high-quality movies are collected for processing. With poor-quality raw data, obtaining a high-resolution and high-quality reconstruction is not possible. For instance, if ice thickness is not optimal for structure determination or if particles adhere to the ice-water interface or exhibit preferred orientations, revisit grid freezing conditions.

Another important step in the workflow is particle picking and extraction. During particle picking, the box size should be approximately 1.4–2.5 times larger than the longest axis of the particle, as sufficient box size is required to capture high-resolution information spread out due to defocus. Larger box sizes, however, require longer processing times due to the increased size of files generated during particle extraction. When choosing a box size, consider particle diameter and pixel size. With many particles, the user may want to bin particles during extraction for initial processing and then re-extract full size particles for final refinements. This protocol uses manual particle picking to generate templates for automatic selection. However, cryoSPARC v3 also offers fully automated picking methods, including a blob-based picker and a Topaz wrapper, which utilizes deep learning to select particles based on previous picks. While these algorithms

are very robust, a significant number of picks would need to be later removed by 2D and 3D classification.

Critical steps also include 2D and 3D classification used to remove artifacts, such as radiation-damaged particles and separate different structural forms of the specimen present in the sample, respectively. The number of 2D classes set by the user should depend upon the number of particles extracted from micrographs, contrast, and heterogeneity in the specimen, as the goal is to separate each individual view of the particle into a separate 2D class. As a general rule, add a 2D class for every 100 particles, and if processing a new sample with a large number of particles, 100 classes is a good starting point. If a low or moderate resolution is obtained even after many rounds of 2D and 3D classification, try re-extracting particles with a larger box size to see if more structural information can be obtained. While reporting the resolution of the final reconstruction, one should analyze the FSC curve obtained according to the gold standard method, along with the local resolution estimates and careful inspection of the map densities, as well as their agreement with the atomic model.

For refinements of the virus structures, the Ewald sphere curvature correction implemented in RELION-3 has demonstrated improvements at high resolution[31]. If refining structures of dynamic multiprotein complexes, try 3D multi-body refinement implemented in RELION-3 or focused classification with image subtraction implemented in RELION-3 and cryoSPARC v3. If heterogeneity cannot be resolved computationally, it is necessary to revisit sample preparation conditions[32]. Insufficient purity or inadequate preparation resulting in protein degradation will hinder the quality of the 3D reconstruction. Additionally, buffer conditions that destabilize proteins or promote aggregation severely limit the number of well-defined particles that can be used for structure calculation. Thus, to most effectively utilize the methods presented here, it is imperative to identify optimal conditions for sample stability. We recommend negative-staining electron microscopy to screen the samples prior to cryo-EM.

As cryo-EM has become the preferred method for 3D structure determination for an increasing number of structural biologists, the need for an integrative and robust workflow for image processing and structure determination becomes more apparent. CryoSPARC offers an easy-to-use, web-based graphic user interface (GUI) that allows users of all experience levels to quickly process data and calculate a 3D structure. Notably, CryoSPARC utilizes a stochastic gradient descent to perform *ab initio* 3D reconstruction. Furthermore, the software employs a branch and bound likelihood optimization algorithm for quick 3D map refinement[7]. The processing pipeline described in this article uses cryoSPARC v3 to yield an initial 3D map. The 3D reconstruction is then refined in RELION-3, a popular package that uses an empirical Bayesian approach to estimate critical parameters based on the user's data set, thereby reducing the need for expert knowledge for program operation[10]. Specifically, we utilize Bayesian polishing for per-particle motion correction and CTF refinements to improve resolution. Finally, the resultant structure is further-refined and validated in Scipion 3[11], an integrative Python shell that supports algorithms from multiple platforms, including Xmipp[13], EMAN2[8], SPIDER[12], and others. While many different software packages are available for cryo-EM users, there is currently no universal SPA platform accepted by the field. Although the SPA workflow can be fully executed in any of the three

software packages described in this article, different algorithms may yield varying results. Consequently, individual steps must be customized depending on the sample and quality of the data. For example, for the current data set, 3Drefine in RELION-3 increased the resolution of the 3D reconstruction, while Nonuniform refinement in cryoSPARC v3 led to a slight resolution decrease. Thus, it is greatly beneficial to utilize a variety of programs to recalculate structures to achieve optimal quality and resolution and to facilitate validation of the reconstructions. Although, Scipion 3 contains numerous algorithms from cryoSPARC v3 and RELION-3, the most recent implementations of these programs are not immediately available in Scipion. For instance, of the programs utilized in this manuscript, only RELION-3 offers Ewald Sphere Curvature Correction through the script Relion_reconstruct. The pipeline presented in this article provides a guide for both new and experienced users to successfully use algorithms implemented in cryoSPARC v3, RELION-3, and Scipion 3 to calculate 3D structures at the near-atomic resolution.

DISCLOSURES

The authors have nothing to disclose.

ACKNOWLEDGMENTS

We thank Carlos Oscar Sorzano for help with Scipion3 installation and Kilian Schnelle and Arne Moeller for help with data transfer between different processing platforms. A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. This study was supported by a start-up grant from Rutgers University to Arek Kulczyk.

REFERENCES

1 Bartesaghi, A. et al. Atomic resolution Cryo-EM structure of beta-galactosidase. *Structure.* **26** (6), 848–856.e843 (2018).

2 Merk, A. et al. Breaking Cryo-EM resolution barriers to facilitate drug discovery. *Cell.* **165** (7), 1698–1707 (2016).

3 Wardell, M. et al. The atomic structure of human methemalbumin at 1.9 A. *Biochemical and Biophysical Research Communications.* **291** (4), 813–819 (2002).

4 RCSB PDB. PDB statistics: Growth of Structures from 3DEM Experiments Released per Year, at <https://www.rcsb.org/stats/growth/growth-em> (2021).

5 Lander, G.C., et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology*. **166** (1), 95–102 (2009).

6 Grant, T., Rohou, A., Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife.* **7**, e35383 (2018).

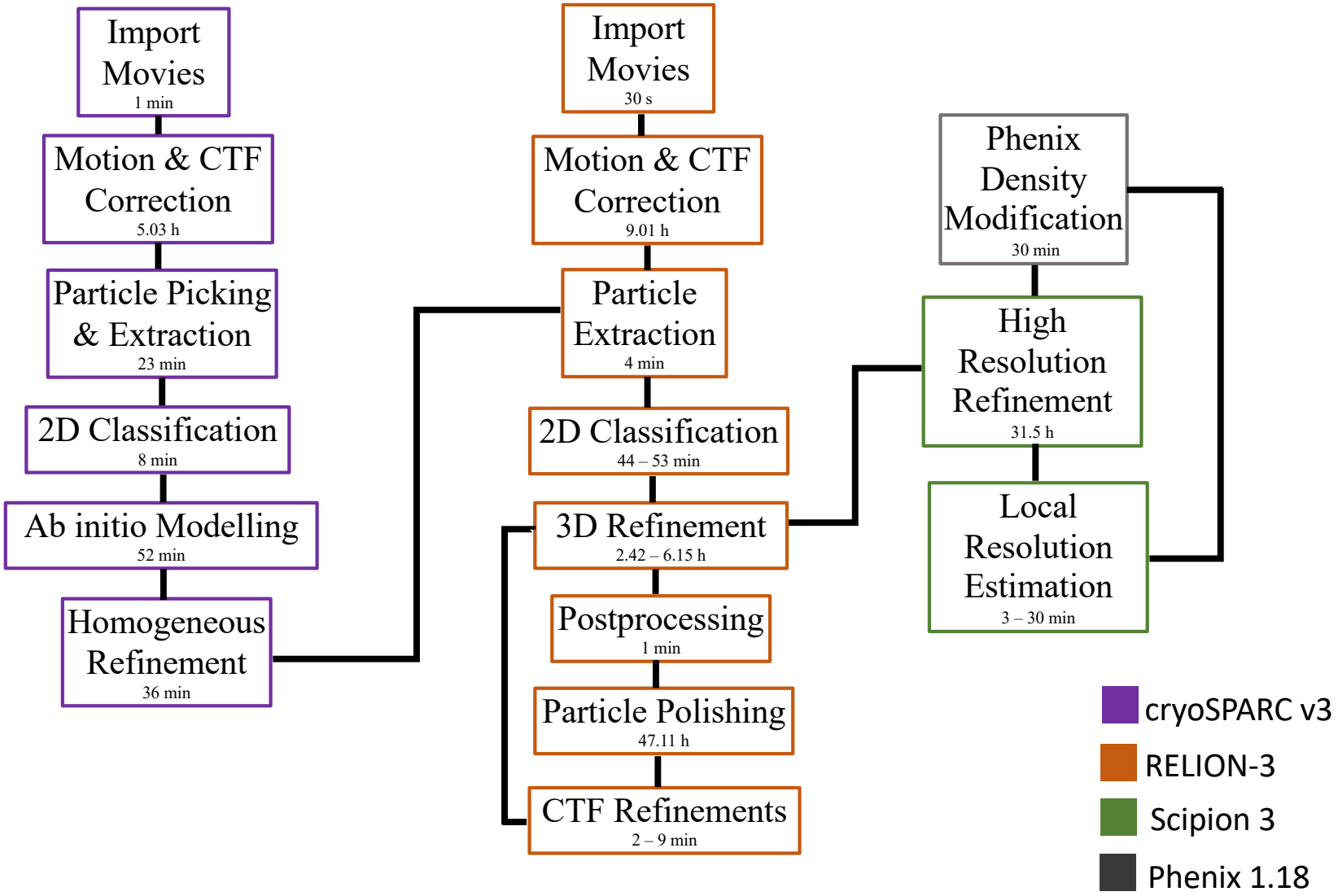
7 Punjani, A., Rubinstein, J. L., Fleet, D. J., Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods.* **14** (3), 290–296 (2017).

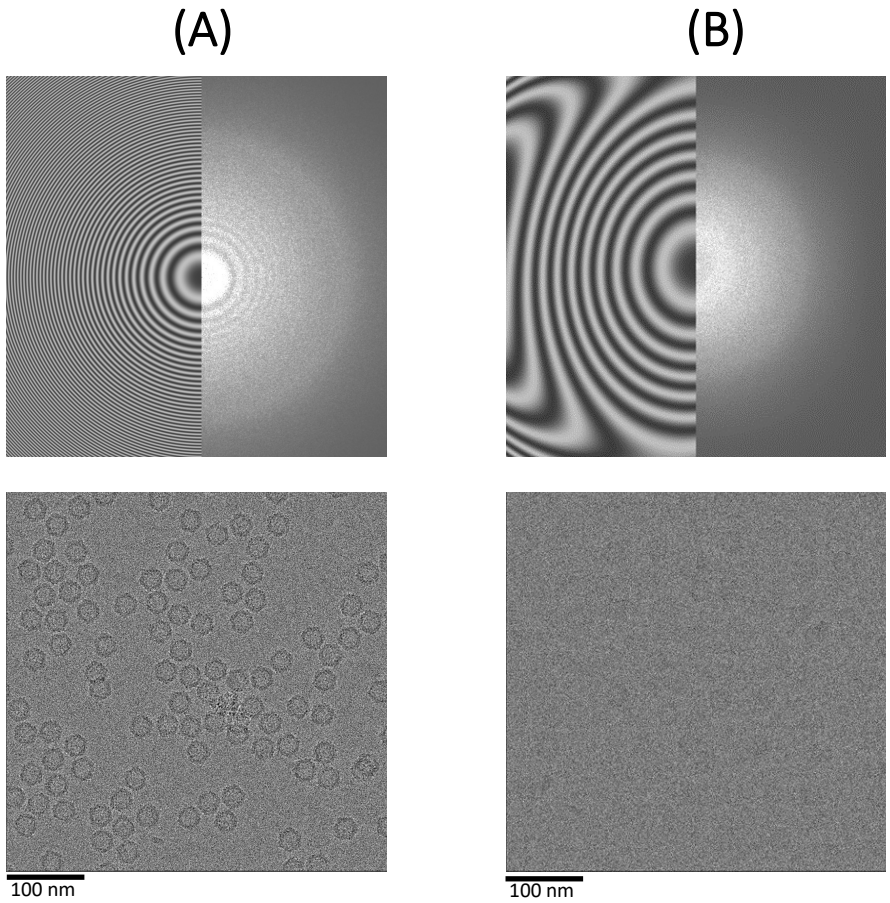
8 Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *Journal of Structural Biology.* **157** (1), 38–46 (2007).

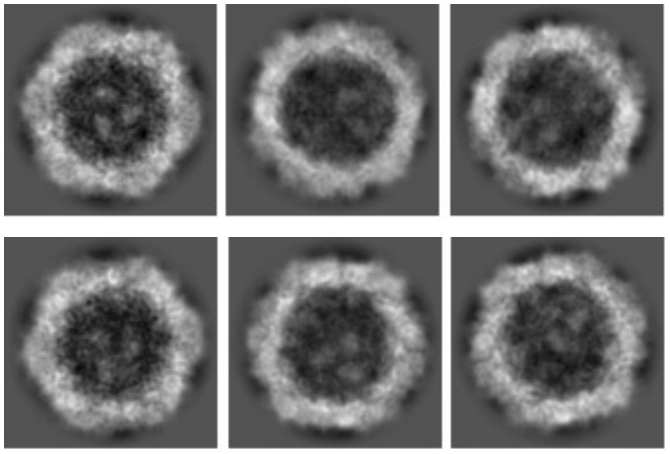
9 van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., Schatz, M. A new generation of the IMAGIC image processing system. *Journal of Structural Biology.* **116** (1), 17–24 (1996).

10 Scheres, S. H. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology.* **180** (3), 519–530 (2012).
11 de la Rosa-Trevin, J. M. et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology.* **195** (1), 93–99 (2016).
12 Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols.* **3** (12), 1941–1974 (2008).
13 Sorzano, C. O. et al. XMIPP: a new generation of an open-source image processing package for electron microscopy. *Journal of Structural Biology.* **148** (2), 194–204 (2004).
14 Lawson, C. L., Chiu, W. Comparing cryo-EM structures. *Journal of Structural Biology.* **204** (3), 523–526 (2018).
15 Naso, M. F., Tomkowicz, B., Perry, W. L., 3rd, Strohl, W. R. Adeno-associated virus (AAV) as a vector for gene therapy. *BioDrugs.* **31** (4), 317–334 (2017).
16 Dawood, S. Punjani, S. Arulthasan, S. Cryo-EM data processing in cryoSPARC: Introductory Tutorial at <https://guide.cryosparc.com/processing-data/cryo-em-data-processing-in-cryosparc-introductory-tutorial> (2020).
17 Bepler, T., Noble, A. J., Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nature Communications.* **11** (5208) (2020).
18 Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry.* **25** (13), 1605–1612 (2004).
19 Scheres, S. Single-particle processing in RELION-3.1 at <https://hpc.nih.gov/apps/RELION/relion31_tutorial.pdf> (2019).
20 Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*. **14** (4), 331–332 (2017).
21 Zheng, S. MotionCor2 User Manual at <https://hpc.nih.gov/apps/RELION/MotionCor2-UserManual-05-03-2018.pdf> (2018).
22 Rohou, A. Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. **192** (2), 216–221 (2015).
23 Asarnow, D., Palovacak, E., Cheng, Y. UCSF PyEM v0.5 at <https://github.com/asarnow/pyem> (2019).
24 Sorzano, C. O. S. et al. A new algorithm for high-resolution reconstruction of single particles by electron microscopy. *Journal of Structural Biology.* **204** (2), 329–337 (2018).
25 Jimenez-Moreno, A. et al. Cryo-EM and single-particle analysis with Scipion. *Journal of Visualized Experiments: JoVE.* **171**, e62261 (2021).
26 Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography.* **66** (Pt 2), 213–221 (2010).
27 Vilas, J. L., et al. MonoRes: Automatic and accurate estimation of local resolution for electron microscopy maps. *Structure*. **26** (2), 337–344 (2018).
28 Sorzano, C. O. et al. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of Structural Biology.* **171** (2), 197–206 (2010).
29 Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods in Enzymology*. **482**, 73–100 (2010).

30 Xie, Q., Yoshioka, C. K., Chapman, M. S. Adeno-associated virus (AAV-DJ)-Cryo-EM structure at 1.56 A Resolution. *Viruses*, **12** (10), 1194 (2020).
31 Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife.* **7**, e42166 (2018).
32 Kulczyk, A. W., Moeller, A., Meyer, P., Sliz, P., Richardson, C. C. Cryo-EM structure of the replisome reveals multiple interaction coordinating DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America.* **114** (10), E1848–E1856 (2017).

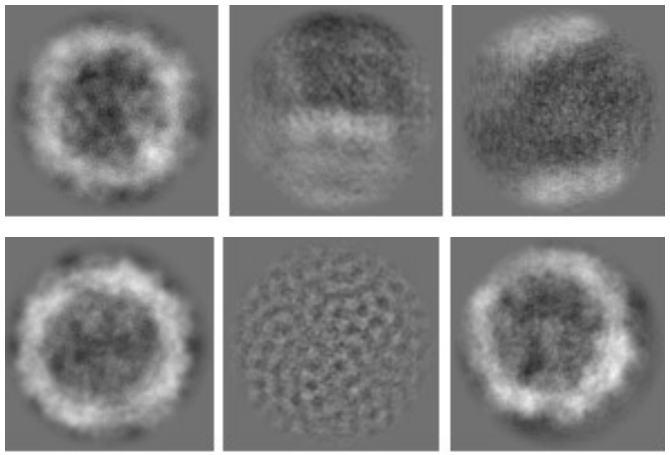






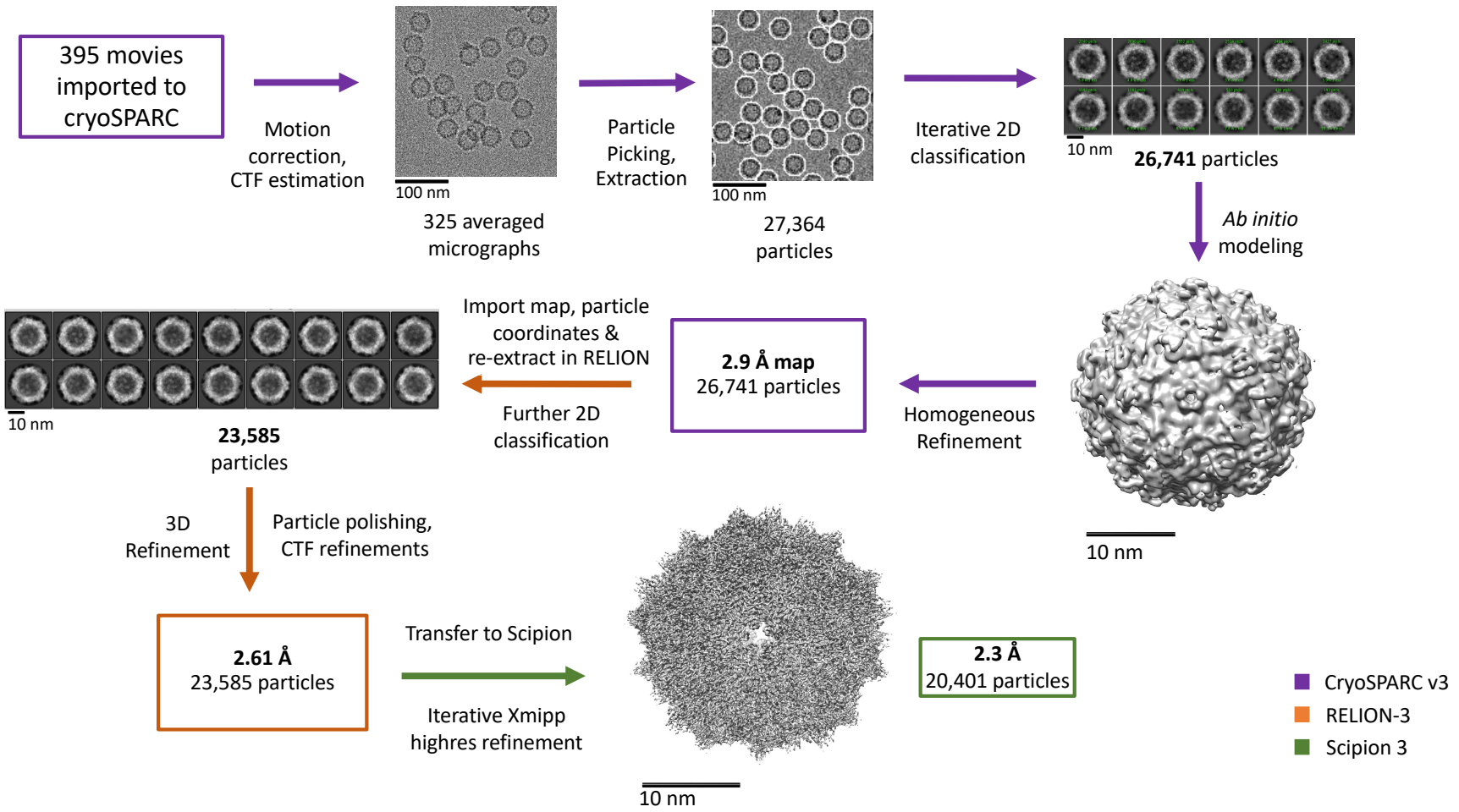
10 nm

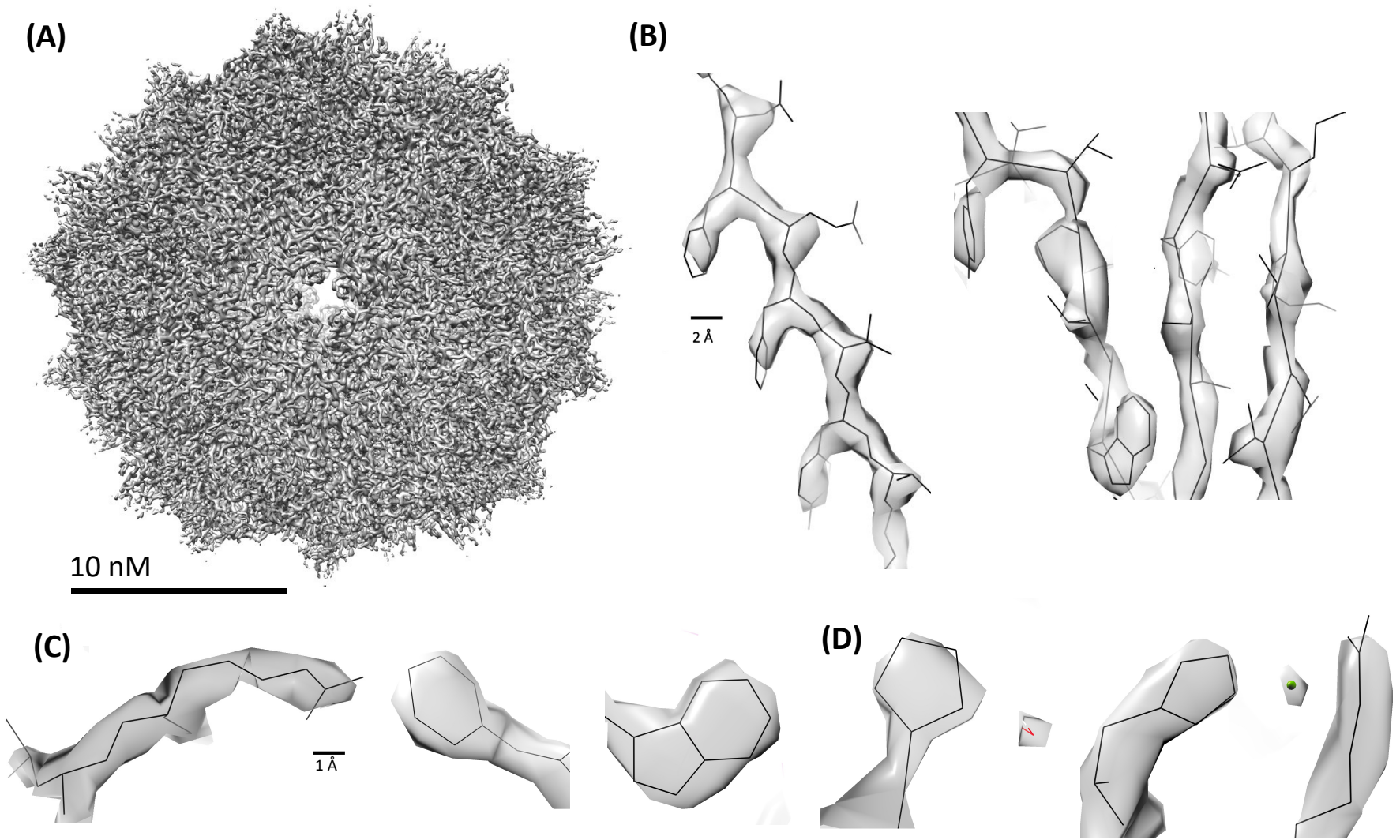
(A)

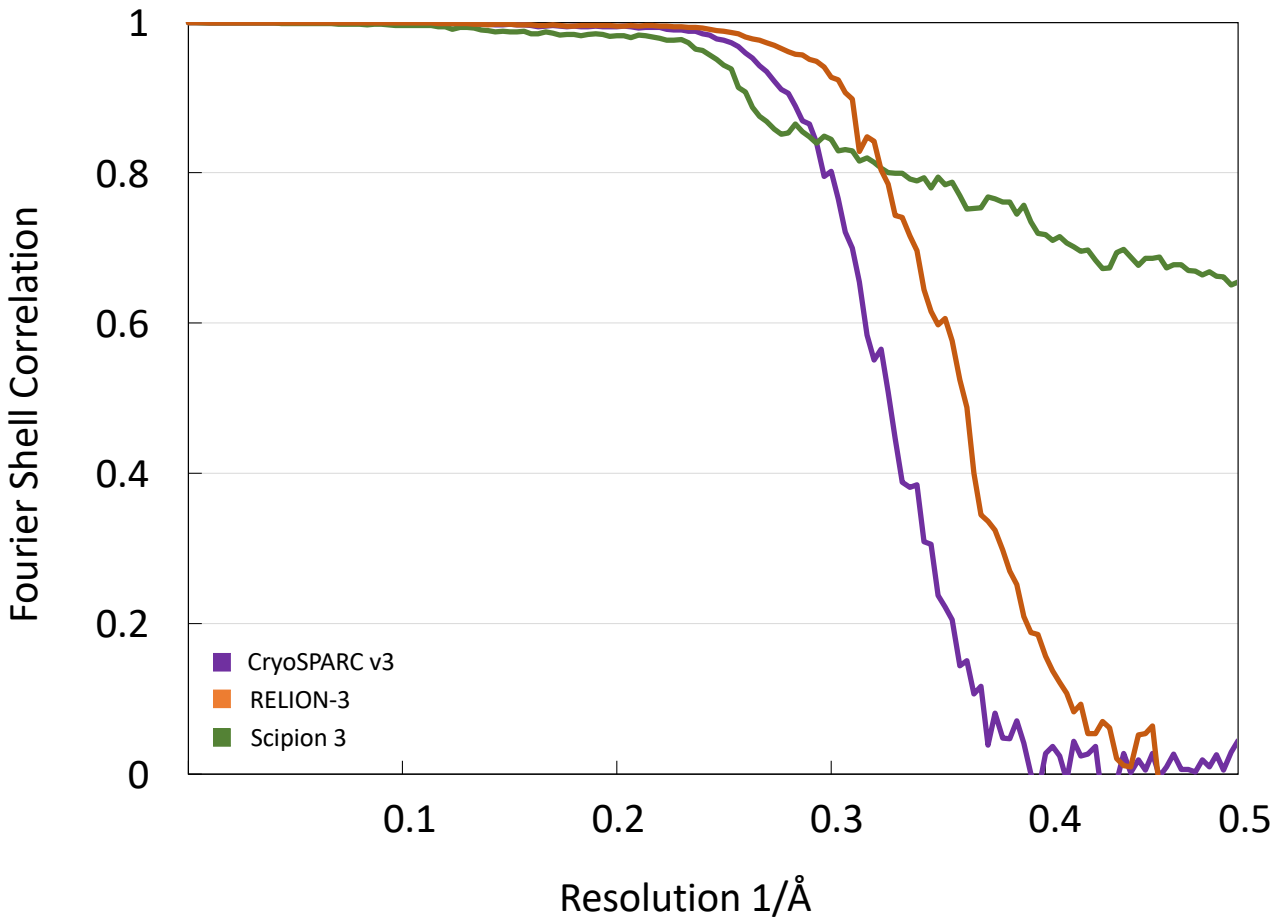


10 nm

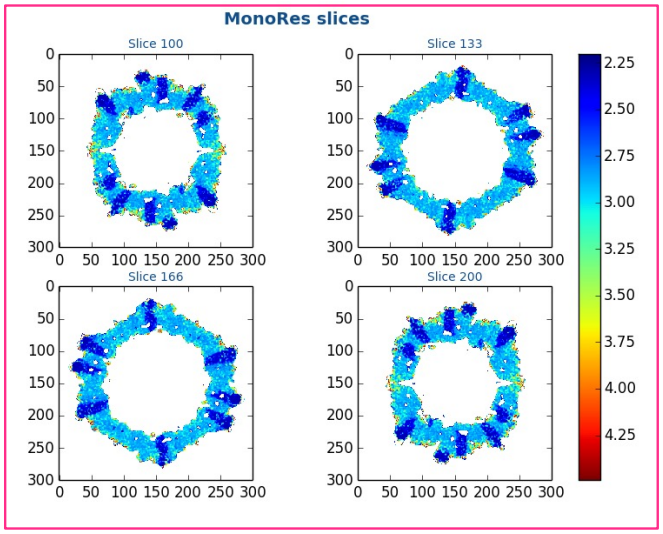
(B)



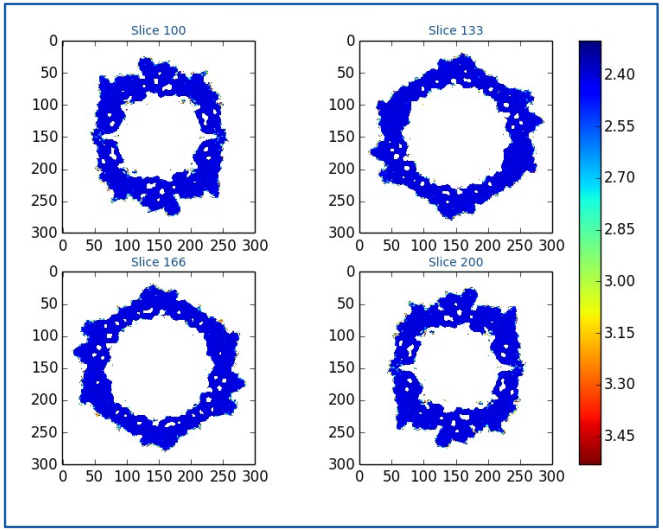




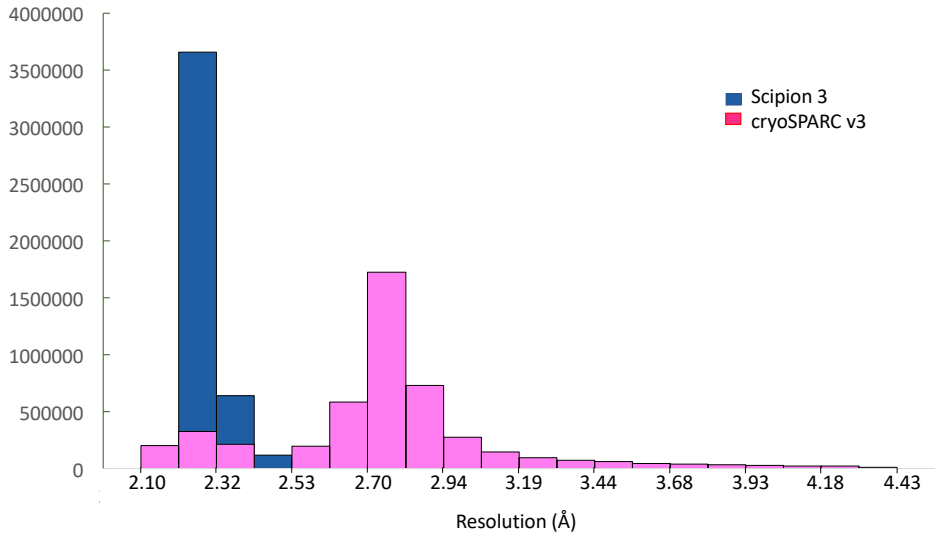
(A)



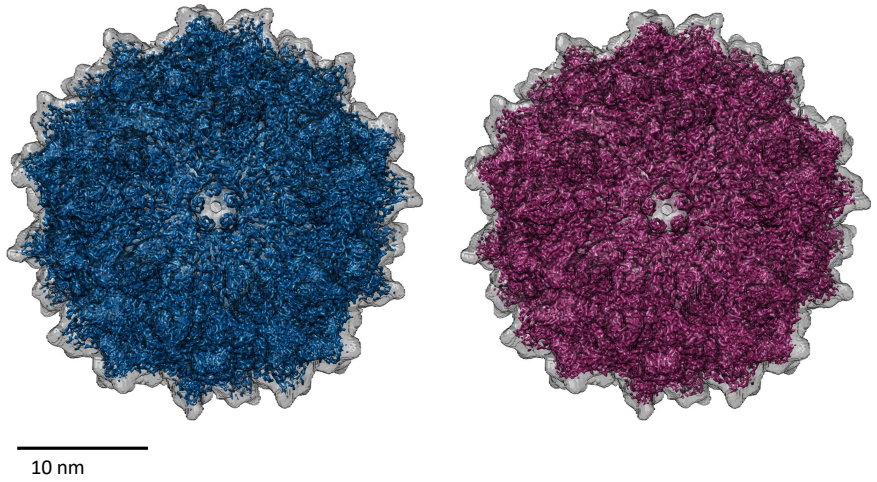
(B)



(C)



(D)



Program	Refinement type	Script
cryoSPARC v3	Homogeneous Refinement	Homogeneous Refinement
	Non-uniform Refinement	Nonuniform Refinement
	Heterogeneous Refinement	Heterogeneous Refinement
	Per-particle motion correction	Local Motion Correction
RELION-3	3D refinement	Refine3D
	Postprocessing - B-factor Sharpening, MTF Correction	Refine3D
	Particle Polishing	Bayesian Polishing
	CTF refinement - beam tilt	CtfRefine
	CTF refinement - anisotropic magnification	CtfRefine
	CTF refinement - per-particle defocus, per-particle/micrograph astigmatism	CtfRefine
	Ewald sphere curvature correction	Relion_reconstruct
Scipion 3	High resolution refinement	Xmipp3 - highres
Phenix 1.18	Density Modification and Sharpening	ResolveCryoEM

Dear Dr. Krishnan,

First, we would like to take this opportunity to thank you, the Editors and the Reviewers for taking the time to carefully read the manuscript and for providing constructive suggestions and comments. We present our point-by-point response to the concerns raised by the Editors and Reviewers below. It is important to note that line numbers listed in this letter correspond to "Track Changes" turned off.

Thank you for your consideration.

With best wishes,
Arek Kulczyk and Megan DiIorio

Editorial comments:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

We have carefully proofread the manuscript, and corrected all spelling and grammar mistakes. Changes are documented in the revised manuscript; these changes can be visualized by turning the "Track Changes" button on under "Review" tab in MS Word.

2. Please revise the text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).

We have removed personal pronouns from the text.

3. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. All commercial products should be sufficiently referenced in the Table of Materials.

We have removed all company names from the text and revised the Table of Materials.

Are cryoSPARC, RELION and Scipion free/ open source software packages?

We have included the following statement in the text (lines 88-90): In this article, cryoSPARC v3, RELION-3, and Scipion 3 were used to obtain a high-resolution 3D reconstruction of AAV, a widely used vector for gene therapy[16]. Aforementioned software packages are free to academic users; cryoSPARC v3 and Scipion 3 require licenses.

JoVE policy states that the video narrative is objective and not biased towards a particular product featured in the video. The goal of this policy is to focus on the science rather than to present a technique as an advertisement for a specific item. To this end, we ask that you please reduce the number of instances of " cryoSPARC, RELION and Scipion " within your text. The

term may be introduced but please use it infrequently and when directly relevant. Otherwise, please refer to the term using generic language.

We have reduced the number of instances of "cryoSPARC, RELION and Scipion" where appropriate.

4. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

We have moved the following sections from the Protocol to the Discussion: notes from section 3: lines 146 – 150 and lines 157 – 163 have been moved to lines 551 – 556 and 557 – 562, respectively. The note from section 4: lines 198 -202 has been moved to lines 566 – 570.

5. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. Please simplify the Protocol so that individual steps contain only 2-3 actions per step and a maximum of 4 sentences per step.

All steps in the protocol contain a maximum of 3 actions and 4 sentences per step.

6. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Please add more details to your protocol steps. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc) to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol.

We have revised the Protocol according to the above recommendations and added more details and/or references to the following highlighted steps: step 6.5 (lines 233 – 236), step 6.6 (lines 238 -241), step 9.1 (lines 319 – 324), and step 12.2 (lines 384 – 387).

7. Line 121: What was the sample used in this study? Where was the data extracted from? Was the data generated in the authors lab?

We have added the following statement to the text (lines 95 - 99): Data were acquired at Oregon Health and Science University (OHSU) in Portland using a 300 kV Titan Krios electron microscope equipped with a Falcon 3 direct electron detector. Images were collected in a counting mode with a total dose of 28.38 $e^{-}/Å^{2}$ fractioned across 129 frames, and defocus range from -0.5 µm to -2.5 µm, at a pixel size of 1.045 Å using EPU. The sample of AAV-DJ was provided by the staff of OHSU.

8. Please ensure that the highlighted steps (not exceeding 3 pages) form a cohesive narrative with a logical flow from one highlighted step to the next. Please highlight complete sentences

(not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in the imperative tense.

We have revised the highlighted steps. These steps form a cohesive narrative, contain complete sentences, at least one action item, and do not exceed 3 pages.

9. As we are a methods journal, please ensure that the Discussion explicitly covers the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol

Critical steps of the Protocol are described in paragraphs 1, 3 and 4 of the Discussion.

b) Any modifications and troubleshooting of the technique

Please see paragraphs 4 and 5.

c) Any limitations of the technique

Please see a paragraph 2.

d) The significance with respect to existing methods

Please see a paragraph 6.

e) Any future applications of the technique

Please see paragraphs 1 and 6.

10. Please do not use the &-sign or the word “and” when listing authors in the references. Authors should be listed as last name author 1, initials author 1, last name author 2, initials author 2, etc.

We have revised and corrected the references.

11. Figure 2/3/5: Please include scale bars in all the images of the panel.

We have included scale bars in all relevant images from Figures 2, 3, 4, 5 and 7.

12. Please ensure that the Table of Materials includes all the supplies (reagents, chemicals, instruments, equipment, software, etc.) used in the study. Please sort the table in alphabetical order.

We have updated the Table of Materials and alphabetically ordered all listed items.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

The authors described a cryo-EM single-particle reconstruction workflow integrated with Relion, cryosparc2, and Scipion. Combining different processing software is a common way for maximizing the strengths of all the software.

Major Concerns:

1. The authors should document better the dataset used and shown in the paper. For example, how the data collection was done. If the authors do not collect the dataset, please state where the dataset is collected. For instance, if the data is collected in counting mode or super-resolution mode, K2 camera? K3 camera? Falcon camera? Energy filter used or not? Defocus range? Without this information about the dataset, the reader can't follow their processing steps.

We apologize for the omission of this important information in the initial manuscript. We have added the following statement to the text in a revised manuscript (lines 95 - 99): Data was acquired at Oregon Health and Science University (OHSU) in Portland using a 300 kV Titan Krios electron microscope equipped with a Falcon 3 direct electron detector. Images were collected in a counting mode with a total dose of 28.38 $e^{-}/Å^{2}$ fractioned across 129 frames, and defocus range from -0.5 µm to -2.5 µm, at a pixel size of 1.045 Å using EPU. The sample of AAV-DJ was provided by the staff of OHSU.

2. "0.22 e/A^2 total exposure dose" must be a typo. Otherwise, it is hard to believe the authors could achieve a 3-A resolution.

Thank you for identifying this typo. A total dose was 28.38 $e^{-}/Å^{2}$. Each movie stack contained 129 frames acquired with a dose of 0.22 $e^{-}/Å^{2}$ per frame.

3. The features shown in the figure do not match my expectation from a 2.1-A map. Where are the water molecules? Where are the ions? Is the map sharpened? Would you please show a figure how the improvement from 2.9 to 2.1-A looks like? Xmipp-highres from Scipion processing is documented in JSB 2018 paper; the FSC may not be the best way to show the map's resolution. Because all the FSC curves somehow do not reach 0 at the Nyquist frequency, which usually means the same averaging unit is being used multiple times in the reconstruction. But I am not sure if this is also the case here. I suggest the author show the expected features at this resolution to convince the readers.

We would like to thank the Reviewer for this constructive comment. We have re-analyzed the data and following a careful inspection of the map, we concluded that a more adequate resolution estimate for the map is 2.3 Å. We do agree with the Reviewer that reporting the resolution estimate according to a single criterion (e.g. 0.143 criterion) from the FSC curve may not be the best way to reflect the resolution of the map as described in Sorzano, C. O. S. *et al.* A

new algorithm for high-resolution reconstruction of single particles by electron microscopy. *Journal of Structural Biology.* 204 (2), 329-337, (2018). Because resolution may vary from point to point in the map, it is often more appropriate to present distribution of local resolution estimates. Thus, following the Reviewer's recommendation, we re-analyzed the 2.9 Å and 2.3 Å maps using Xmipp – MonoRes. Results of the analysis are presented in Figure 7 and discussed in paragraph 3 of the Representative Results (lines 458-473). Figure 7A-D clearly demonstrate the incremental improvement of the map during refinement when comparing local resolution distributions obtained for 2.9 Å and 2.3 Å maps.

Below we address other specific points made by the Reviewer. Densities representing water molecules and magnesium ions are presented in Figure 5D. We identified these densities by overlaying the 2.3 Å map with previously deposited atomic coordinates of AAV-DJ (PDB ID: 7fkr). The map was sharpened in Phenix using a B factor of -43.43. The FSC curves obtained in cryoSPARC and RELION do reach 0 at Nyquist frequency. However, we agree with the Reviewer that the way these plots were presented in Figure 6 were misleading. Therefore, we have redrawn plots in the figure. The FSC plot from Xmipp – highres does not reach 0 at Nyquist. In our opinion, this result suggests the resolution estimate is limited due to insufficient sampling during data collection as described in Penczek, P. A., Resolution Measures in Molecular Electron Microscopy. 482, 1-33 (2010). Interestingly, an analogous sample of AAV-DJ analyzed in the same cryo-EM center where we collected data was refined to 1.56 Å resolution (Xie, Q., Yoshioka, C. K., Chapman, M. S. Adeno-Associated Virus (AAV-DJ)-Cryo-EM Structure at 1.56 A Resolution. *Viruses*, *12* (10), (2020)). However, the above data set was acquired at 0.514 Å/px. In contrast, our data set was acquired at 1.045 Å/px.

Taken together, although the FSC curve calculated with Xmipp3 – highres indicates the Nyquist limit has been reached, MonRes analysis presented in Figure 7, along with a careful analysis of the EM map and map fitting with atomic coordinates of AAV presented in Figure 5 suggest an adequate resolution estimate for the map is 2.3 Å. Interestingly, similar discrepancy in Xmipp – highres and Xmipp – MonoRes estimates have been reported earlier, for example in Sorzano, C. O. S. *et al.* A new algorithm for high-resolution reconstruction of single particles by electron microscopy. *Journal of Structural Biology.* 204 (2), 329-337, (2018), and Jimenez-Moreno, A. *et al.* Cryo-EM and Single-Particle Analysis with Scipion 3. *Journal of Visualized Experiments.* (171), e62261, (2021).

4. If the mask is too tight, the reported resolution will be the Nyquist resolution. Maybe also show the mask and make sure the mask is not too tight.

A circular mask with a radius of 150 pixels was automatically applied by Xmipp3 – highres during refinements. We have used Chimera to create a mask for Xmipp–MonoRes analysis. This mask overlapped with the 2.3 Å and 2.9 Å reconstructions is presented in Figure 7D.

5. It would be great if the authors show monores slice of both 2.9 and 2.1-A maps.

The Xmipp – MonoRes slices along with local resolution histograms obtained for 2.3 Å and 2.9 Å maps are presented in Figure 7 A-B and Figure 7C, respectively.

6. In Figure 4. Flow chart - 2D classification results are both from cryosparc2. It makes more sense if the authors show representative 2D classification results from relion for 2D classification performed in Relion. Otherwise, it isn't clear.

We apologize for this mistake. Figure 4 has been revised accordingly to the Reviewer's comment.

7. The authors did not mention non-uniform refinement, heterogenous refinement, local motion correction in cryosparc2, and they should be removed from the flowchart.

The above-listed cryoSPARC protocols have been removed from the flowchart presented in Figure 1.

Reviewer #2:
Manuscript Summary:
The authors introduce an interesting image processing workflow for Single Particle Analysis by Cryo-Electron Microscopy combining multiple popular software suites (CryoSparc, Relion, and Scipion/Xmipp). The manuscript is easy to follow and its results support the validity of the approach. It will be a useful contribution for many practitioners.

Major Concerns:
None

Minor Concerns:
Scipion offers all the processes used in the article from CryoSparc and Relion. Would not have been more advantageous to perform the whole analysis within a single platform so that their results could also have been compared? At the moment, they are simply transferred from one package to the next without any quantitative comparison. The authors may want to comment on this possibility.

We thank Reviewer for this comment. Although, Scipion provides an integrative Python shell supporting algorithms from multiple platforms including cryoSPARC and RELION, the most recent implementations of these programs are not immediately available in Scipion. For instance, as far as we know, only RELION-3 offers Ewald Sphere Curvature Correction through the script Relion_reconstruct. Currently, there is no universal SPA platform accepted by the field. We think Scipion, along with Appion could become such platforms in the near future. Analysis of cryo-EM structures deposited in the PDB in recent years indicate the majority of these structures have been determined using RELION, whereas a number of cryoSPARC depositions is rapidly growing. We think it is important to reiterate the usefulness and application of Scipion, as its multitude of algorithms often yield the higher resolution reconstructions than RELION and cryoSPARC, as evidenced in this manuscript. Finally, we do

recognize the advantage of Scipion for quantitative comparison of reconstructions obtained using different SPA platforms. We have compared maps obtained in cryoSPARC and Scipion using Xmipp–MonoRes. Results of the analysis are presented in Figure 7 and discussed in paragraph 3 of the Representative Results (lines 458-473) in the revised manuscript.

As a minor remark, the beginning of the protocol is currently labelled as "Protocal" instead of "Protocol".

We have corrected the typo.

Reviewer #3:
Manuscript Summary:
This manuscript details processing single particle cryo-EM data to yield high resolution 3D structures with varying programs. While there are tutorials available for individual software packages, this demonstrates how to move data between 3 different packages and utilize different aspects of each to arrive at an improved structure. While this 3 method step may not be commonplace (it also may be more utilized than I am aware of), demonstrating using multiple platforms and the application of PyEM is a very useful tutorial that is an excellent tool for anyone learning to process this type of data. The written steps are easy to understand and good notes are supplied for the test sample specific parameters. This is a very welcome addition to the growing number of online tutorials for cryo-EM, from sample prep through data processing. Thank you!

We thank the Reviewer for this comment.

Major Concerns:
Including times estimates or the actual times required for the test dataset will be helpful to novices who may underestimate the time required to run these processes. Supply some reasoning behind why using certain packages for some steps is better than others. Steps 6.3 and 6.5 could use more detail.

Time estimates have been included in the flowchart presented in the Figure 1. Paragraph 6 in the Discussion now contains a section concerned with comparing algorithms from cryoSPARC, RELION and Scipion (lines 603-615). Step 6 of the Protocol has been re-written, and former sections 6.3 and 6.5 expanded in a revised manuscript as 6.1-6.4 and 6.8-6.9, respectively.

Minor Concerns:
Clearly indicate which version of softwares are being used. Describe what may be needed to access these programs, i.e. cryoSPARC is free for academic use but requires a license.

We have clearly indicated specific versions of all programs used in this study and added the following statement to the text (lines 88-90): In this article, cryoSPARC v3, RELION-3, and Scipion 3 were used to obtain a high-resolution 3D reconstruction of AAV, a widely used vector

for gene therapy[16]. Aforementioned software packages are free to academic users; cryoSPARC v3 and Scipion 3 require licenses.

Reviewer #4:
Manuscript Summary:
In the manuscript "A robust single-particle cryo-Electron Microscopy (cryo-EM) processing workflow with cryoSPARC, RELION and Scipion." Dilorio and Kulczyk provide a valuable workflow for routine structure determination through cryo-EM.

In the Cryo.EM field, every researcher has their own ways to obtain high-resolution data in a streamlined and straightforward manner. Therefore, it is difficult for me to review this methods paper or even provide further suggestions on improving it, as I am using a different approach that may not be compatible with intermediate steps as presented here. Nevertheless, overviews, as presented here, are still valuable providing alternative perspectives and generating new ideas. The manuscript is well organized and easy to follow. Therefore, I do not feel comfortable suggesting alternative strategies or enforcing significant modifications.

Minor Concerns:
However, something is horribly wrong with the FSC curves presented in Figure 7 - maybe the authors can explain why the FSC curve obtained from Scipion never reaches zero if this is not an artifact of some sort; the authors have to explain why this happens in the manuscript.

We have redrawn plots in the Figure 6. The FSC curves obtained in cryoSPARC and RELION do reach 0 at Nyquist frequency. However, the FSC plot from Xmipp – highres does not reach 0 at Nyquist indicating the resolution estimate is limited due to insufficient sampling during data collection as described in Penczek, P. A., Resolution Measures in Molecular Electron Microscopy. 482, 1-33 (2010). Similar behavior of the FSC plots obtained with Xmipp – highres have been observed earlier, for example in Sorzano, C. O. S. *et al.* A new algorithm for high-resolution reconstruction of single particles by electron microscopy. *Journal of Structural Biology.* 204 (2), 329-337, (2018), and Jimenez-Moreno, A. *et al.* Cryo-EM and Single-Particle Analysis with Scipion 3. *Journal of Visualized Experiments.* (171), e62261, (2021). Appropriate explanation has been included in the Representative Results section (lines 462-468) and in the legend of Figure 6.

Point-by-point response to the Editorial comments (2nd revision):

1. Please note that the manuscript has been formatted to fit the journal standard. Comments to be addressed are included in the manuscript itself. Please review and revise.

We have revised the manuscript. All revisions are included in the following document: "Kulczyk_JoVE_Manuscript_29Nov21.docx"

2. Some step numbers have been modified. Please check the step numbers referenced within the manuscript.

We have checked all step numbers referenced within the manuscript and changed numbering in the following lines: 174, 192, 290, 388, and 440. Please note that line numbers listed above correspond to lines in the Word document with "Track Changes" turned on.

3. Abstract cannot have citations. Please move the citations to the Introduction.

We have removed citations from the abstract. As the consequence, all references in the manuscript had to be renumbered. Please see: "Kulczyk_JoVE_Manuscript_29Nov21.docx" for details.

4. Figure 1: Please revise "hrs" to "h" in the figure.

We have changed "hrs" to "h" in Figure 1. We have also changed "sec" to "s".
The revised file "Kulczyk_JoVE_Figure_1_29Nov21.pdf" has been uploaded into the JoVE system.

5. Please check if reference number 17 is correct.

The reference number 17 is correct. Because the references had to be renumbered in the revised manuscript, the reference number 17 has now become the reference number 16.

6. Please ensure that the Table of Materials includes all the supplies (reagents, chemicals, instruments, equipment, software, etc.) used in the study.

We have updated the Table of Materials by adding names of two software packages to the list, namely: CTFFIND 4 and MotionCor2. These programs were used in steps 6.3 and 6.4 of the protocol, respectively. The revised file "Kulczyk_JoVE_Table_of_Material_29Nov21.pdf" has been uploaded into the JoVE system.