FINAL SCRIPT: APPROVED FOR FILMING

Submission ID #: 68636
Scriptwriter Name: Sulakshana Karkala
Project Page Link: https://review.jove.com/account/file-uploader?src=20934028

Title: Subnanometer-Resolution Structural Determination of Hemagglutinin from Cryo-Electron Tomography of Influenza Viruses

Authors and Affiliations:
Qiuyu J. Huang, Clint N. McDaniel, Alijah A. Griffith, Celia A. Schiffer, Mohan Somasundaran

Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School

Corresponding Authors:
Mohan Somasundaran mohan.somasundaran@umassmed.edu
Celia A. Schiffer celia.schiffer@umassmed.edu
Email Addresses for All Authors:
Qiuyu J. Huang qiuyu.huang@umassmed.edu
Clint N. McDaniel clint.mcdaniel@umassmed.edu
Alijah A. Griffith alijah.griffith@umassmed.edu
Mohan Somasundaran mohan.somasundaran@umassmed.edu
Celia A. Schiffer celia.schiffer@umassmed.edu

Author Questionnaire

1. **Microscopy**: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **No**

2. **Software:** Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes, all done**

3. **Filming location:** Will the filming need to take place in multiple locations? **No**

4. **Testimonials (optional):** Would you be open to filming two short testimonial statements **live during your JoVE shoot**? These will **not appear in your JoVE video** but may be used in JoVE's promotional materials. **No**

Current Protocol Length
Number of Steps: 26
Number of Shots: 51

Introduction

Videographer: Obtain headshots for all authors available at the filming location.

1.1. **Clint McDaniel:** Our lab uses structural biology to study viruses and the evolution of drug resistance. In the current work, we are characterizing how influenza hemagglutinin changes conformation upon receptor binding.

1.1.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:4.2*

What technologies are currently used to advance research in your field?

1.2. **Clint McDaniel:** We have leveraged cryo-electron tomography to visualize viral morphology, organization, and structure. Specifically, we have used this modality of cryo-electron microscopy to image the pleomorphic virus influenza.

1.2.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:4.1*

What advantage does your protocol offer compared to other techniques?

1.3. **Clint McDaniel:** Our protocol allows us to study the organization and structure of influenza hemagglutinin *in situ* on intact viruses at subnanometer resolution. This higher resolution allows us to visualize formerly hidden interactions.

1.3.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

Videographer: Obtain headshots for all authors available at the filming location.

2. Preprocessing Cryo-Electron Tomograms and Convolutional Neural Network-Based Particle Picking for HA Glycoprotein and M1 Protein Identification

Demonstrator: Clint McDaniel

2.1. To begin, launch a Linux terminal for activating the conda environment **[1].** Use the terminal to activate the conda environment with IsoNet installed **[2]**.

2.1.1. WIDE: Talent launching the Linux terminal program on a computer system.

2.1.2. SCREEN: 68636_screenshot_2.1.2_2.3.1.mp4 00:02-00:09

2.2. Create a subfolder named **tomo_folder** *(Tomo-Folder)*, then move all tomogram files into that folder **[1]**. Next, use the code to generate a star file in the project folder **[2].**

2.2.1. SCREEN: 68636_screenshot_2.1.2_2.3.1.mp4 00:22-00:27, 00:33-00:49

2.2.2. SCREEN: 68636_screenshot_2.1.2_2.3.1.mp4 01:28-01:57
TXT: isonet.py prepare_star tomo_folder --output_star tomograms.star --pixel_size 8.35

2.3. Using a text editor, open the generated star file and input the approximate defocus value for the 0-degree *(Zero-degree)* tilt images into the fourth column **[1]**. ~~The defocus values can be found in the processed_items.json *(processed-items-dot-J-son)* file located in the warp_tiltseries *(Warp-tilt-series)* folder **[2].**~~

2.3.1. SCREEN: 68636_screenshot_2.1.2_2.3.1.mp4 02:07-02:11, 02:15-02:17, 02:26-02:40, 03:09-03:12

2.3.2. ~~WIDE: Show opening processed_items.json in warp_tiltseries folder.~~

2.4. Now, run the CTF *(C-T-F)* deconvolution command in the terminal **[1-TXT].** After deconvolving, activate the EMAN2 GUI *(Ee-Man-Two-G-U-I)* environment to begin preprocessing for particle picking **[2].**

2.4.1. SCREEN: 68636_screenshot_2.4.1.mp4 00:02-00:30

TXT:Command: isonet.py deconv tomograms.star --snrfalloff 0.7 --deconv_folder deconvolve

2.4.2. SCREEN: 68636_screenshot_2.4.2.mp4 00:02-00:14

2.5. Under Tomography, left-click the arrow next to **Raw Data**, select **Import Tomograms** from the drop-down menu and choose the files **[1]**.

2.5.1. SCREEN: 68636_screenshot_2.5.1_2.6.1.mp4 00:03-00:18

2.6. Then left-click the arrow next to **Segmentation** and choose **Preprocess tomograms [1]**. Use suitable default parameters **[2].** After preprocessing, an **info** directory with blank .json files matching the base names of the processed tomography files will be created automatically **[3]**. Add **angstrom-pixel** information to the files before proceeding **[3]**.

2.6.1. SCREEN: 68636_screenshot_2.5.1_2.6.1.mp4 00:50-00:55

2.6.2. SCREEN: 68636_screenshot_2.5.1_2.6.1.mp4 01:08-01:19

2.6.3. SCREEN: 68636_screenshot_2.6.3_2.6.4.mp4 00:02-00:12

2.6.4. SCREEN: 68636_screenshot_2.6.3_2.6.4.mp4 00:18-00:32

2.7. To train a convolutional neural network or CNN *(C-N-N)* to recognize the HA *(H-A)* glycoprotein, change the working directory to the location of the tomograms to be used **[1]**, then open the CNN training interface using the terminal **[2]**.

2.7.1. SCREEN: 68636_screenshot_2.7.1_2.8.1.mp4 00:02-00:10

2.7.2. SCREEN: 68636_screenshot_2.7.1_2.8.1.mp4 00:15-00:35

2.8. Confirm that four windows are open. One containing the information on CNN parameters and tomograms in the directory, and the other three windows containing images of good references, bad references, and particles selected by the CNN **[1].**

2.8.1. SCREEN: 68636_screenshot_2.7.1_2.8.1.mp4 00:39-00:50

2.9. In the CNN GUI, click on the **New** option to initialize a CNN with default learning rate of 0.0001 and box size of 8 **[1]**.

2.9.1. SCREEN: 68636_screenshot_2.9.1_2.10.2.mp4 00:02-00:09

2.10. Under **FileName** column, left-click on a representative tomogram to open it **[1]**. Hover the cursor over the open tomogram and click the mouse scroll wheel to open a new window. Left click the N# *(N-number)* slider to traverse the z-axis**[2]**.

2.10.1. SCREEN: 68636_screenshot_2.9.1_2.10.2.mp4 00:10-00:16

2.10.2. SCREEN: 68636_screenshot_2.9.1_2.10.2.mp4 00:16-00:26

2.11. Next, on the open tomogram, use the **Good References** panel to select 10 to 20 features corresponding to the HA . Positive images appear in the Positive window **[1-TXT].** To

remove one, hold **Shift** and left-click it **[3]**.

2.11.1. SCREEN: 68636_screenshot_2.11.1_2.11.3.mp4 00:02-00:22
TXT: Select top and side views of HA

2.11.2. SCREEN: 68636_screenshot_2.11.1_2.11.3.mp4 00:21-00:25

2.12. Now, switch to the **Bad References** panel **[1]** and select 10 to 20 references corresponding to features such as empty membrane patches, vRNP *(V-R-N-P)* density, fiducial markers, and debris **[2-TXT]**.

2.12.1. SCREEN: 68636_screenshot_2.12.1_2.12.3.mp4 00:02-00:05
Bad References panel is being opened.

2.12.2. SCREEN: 68636_screenshot_2.12.1_2.12.3.mp4 00:05-00:17
TXT: The features will appear in the Negative window

2.13. Set the number of iterations to 50 and click on the **Train** icon to begin CNN training **[1]**.

2.13.1. SCREEN: 68636_screenshot_2.13.1.mp4 00:03-00:10

2.14. After training is completed, press **Apply** ~~to have the CNN pick particles in the current tomogram~~ **[1]**. Selected particles will be displayed on the tomogram as blue circles or can be seen in the **Particles** window **[2]**.

2.14.1. SCREEN: 68636_screenshot_2.14.1_2.14.2.mp4 00:04-00:06

2.14.2. SCREEN: 68636_screenshot_2.14.1_2.14.2.mp4 00:19-00:35

2.15. Continue refining by selecting additional positive and negative references, periodically clicking **Save** to preserve progress. **[1]** Once the CNN performs satisfactorily, click **Apply All** to run particle picking ~~across all tomograms in the directory~~ **[2]**. Then, evaluate results on several tomograms and retrain as needed **[3]**.

2.15.1. SCREEN: 68636_screenshot_2.15.1.mp4 00:03-00:27

2.15.2. SCREEN: 68636_screenshot_2.15.2_2.15.3.mp4 00:02-00:05

2.15.3. SCREEN: 68636_screenshot_2.15.2_2.15.3.mp4 00:11-00:15, 00:22-00:30

2.16. To save coordinates for each tomogram, reopen the EMAN2 GUI using the command **e2projectmanager.py** *(E-Two-Project-manager-dot-P-Y)* **[1]**. Click the arrow under **Subtomogram Average** and select **Manual Boxing [2]**.

2.16.1. SCREEN: 68636_screenshot_2.16.1_2.17.2.mp4 00:02-00:08

2.16.2. SCREEN: 68636_screenshot_2.16.1_2.17.2.mp4 00:08-00:11

2.17. Now input the tomogram's name and click **Launch [1]**. Then sequentially click on **File , Save Box Coord** *(Save-Box-Co-Ord)* and **tomogram_ha.txt** *(Tomogram-underscore-H-A-*

Dot-T-X-T) to save the co-ordinates **[2]**. Navigate to the neuralnets subfolder and info subfolders to backup **nnet_save.hdf** *(N-Net-Save-Dot-H-D-F)*, **trainouts.hdf** *(train-outs-Dot-H-D-F)*, **segouts.hdf** *(Segg-outs-Dot-H-D-F)*, and **boxes3dref.hdf** *(Boxes-three-D-ref-Dot-H-D-F)* **[3]**.

2.17.1. SCREEN: 68636_screenshot_2.16.1_2.17.2.mp4 00:14-00:17

2.17.2. SCREEN: 68636_screenshot_2.16.1_2.17.2.mp4 00:18-00:40

2.17.3. SCREEN: 68636_screenshot_2.17.3.mp4 00:02-00:30
Video Editor: Please speed up if necessary

2.18. Train a second CNN to detect the M1 protein layer and vRNP presence by changing the training box size from 8 to 14 **[1]**.

2.18.1. SCREEN: 68636_screenshot_2.18.1.mp4 00:02-00:28

3. Particle Curation and Subtomogram Averaging Workflow Using RELION and WarpTools

3.1. To curate particles, first download notebooks from the GitHub repository **[1]**, then open the CNN_Particle_Cleaning.ipynb *(C-N-N-Particle-Cleaning-dot-I-P-Y-N-B)* notebook and load the required modules **[2]**.

3.1.1. SCREEN: 68636_screenshot_3.1.1.mp4 00:02-00:10

3.1.2. SCREEN: 68636_screenshot_3.1.2_3.2.3.mp4 00:02-00:08

3.2. Load in the HA and M1 coordinate text files into the notebook **[1].** View them as 3D point clouds using the Open3D *(open-Three-D)* library **[2]**. Next, filter out HA coordinate outliers using Open3D's statistical outlier removal method **[3]**.

3.2.1. SCREEN: 68636_screenshot_3.1.2_3.2.3.mp4 00:08-00:13

3.2.2. SCREEN: 68636_screenshot_3.1.2_3.2.3.mp4 00:13-00:21

3.2.3. SCREEN: 68636_screenshot_3.1.2_3.2.3.mp4 00:21-00:36

3.3. Next, calculate the distances between HA and M1 point clouds by identifying HA points more than 20 pixels from M1 as outliers **[1]**, and save remaining coordinates to an output text file **[2]**.

3.3.1. SCREEN: 68636_screenshot_3.3.1_3.3.2.mp4 00:07-00:13

3.3.2. SCREEN: 68636_screenshot_3.3.1_3.3.2.mp4 00:13-00:20

3.4. Concatenate all particles and save as starfiles using **pts2starfile.ipynb** *(P-T-S-2-star-file-dot- I-P-Y-N-B)* **[1]**. Using WarpTools, extract particles at binning factor 4 and perform initial subtomogram averaging in RELION4 *(re-lion-4)* using the code **[2]**.

3.4.1. SCREEN: 68636_screenshot_3.4.1.mp4 00:03-00:15

3.4.2. SCREEN: 68636_screenshot_3.4.2.mp4 00:02-00:12

AND
TEXT IN PLAIN BACKGROUND:
WarpTools ts_export_particles --settings warp_tilt_settings.setting --input_star pts2star.star --coords_angpix 8.35 --output_star bin4_export.star --output_angpix 8.35 --box 48 --diameter 140 --3d
Video Editor: Please play both shots side by side

3.5. Convert the warp starfile to RELION 4 format **[1-TXT].** Then generate an initial reference using **relion_refine_mpi** *(re-lion-refine-M-P-I)* on a subset of particles **[2]**.

3.5.1. SCREEN: 68636_screenshot_3.5.1_3.5.2.mp4 00:05-00:20
TXT: Command: relion_convert_star --i bin4_export.star --o bin4_conv.star

3.5.2. SCREEN: 68636_screenshot_3.5.1_3.5.2.mp4 00:55-01:03

3.6. Perform 3D auto-refinementon the subtomograms with **relion_refine_mpi [1]**. Upon convergence, repeat auto-refinement with tightened translation offset to 8 and step to 2 **[2]**.

3.6.1. SCREEN: 68636_screenshot_3.6.1_3.6.2.mp4 00:02-00:15
AND
TEXT ON PLAIN BACKGROUND:
Sample Command
mpiexec -n 3 relion_refine_mpi --o Refine3D/job001/run --auto_refine --split_random_halves --i bin4_conv.star --ref init_ref/job001/run_class001.mrc --firstiter_cc --ini_high 20 --dont_combine_weights_via_disc --pool 3 --pad 2 --ctf --particle_diameter 400 --flatten_solvent --zero_mask --oversampling 1 --healpix_order 2 --auto_local_healpix_order 4 --offset_range 16 --offset_step 4 --sym C1 --low_resol_join_halves 40 --norm --scale --j 12 --gpu 0:1 --pipeline_control Refine3D/job001
Video Editor: Please play both shots side by side

3.6.2. SCREEN: 68636_screenshot_3.6.1_3.6.2.mp4 00:20-00:36

3.7. Leverage 2D classification to discard junk particles using **relion_refine [1]**. ~~Then, combine classes containing an identifiable HA array with cylindrical HA, membrane, and M1 density using **relion_star_handler [2].**~~

3.7.1. SCREEN: 68636_screenshot_3.7.1.mp4 00:03-00:13

3.7.2. ~~SCREEN: Show terminal selecting good classes.~~
AUTHOR'S NOTE: Shot removed to prevent redundancy

3.8. Create star files with good classes using the code **[1]**, then combine the individual star

files **[2].**

3.8.1. SCREEN: 68626_screenshot_3.8.1.mp4 00:54-01:09
TXT:Command: relion_star_handler --i input_file2.star --o output_good_class.star --select rlnClassNumber --minval goodclassnumber --maxval goodclassnumber

3.8.2. SCREEN: 68626_screenshot_3.8.2.mp4 00:46-00:54

AND
TEXT ON PLAIN BACKGROUND:
relion_star_handler --i "output_good_class1.star output_good_class2.star … output_good_classn.star" --o bin4_keep.star –combine
Video Editor: Please play both shots side by side

Results

4. Results

4.1. Representative tomograms revealed pleomorphic virions with shapes ranging from spherical to elongated forms **[1]**.

4.1.1. LAB MEDIA: Figure 2. *Video editor: Highlight the dotted orange images (2A and 2B) when VO says spherical and dotted blue images (2A and 2C) when VO says elongated.*

4.2. The final subtomogram average reached a global resolution of 6.0 angstroms, with local resolution ranging from 5 to 7 angstroms **[1]**. Alpha-helices and beta-sheets were resolved in the final HA reconstruction, supporting the structural quality of the refined map **[2]**.

4.2.1. LAB MEDIA: Figure 3A and 4 A *Video Editor: Show both images side by side and highlight the left side of both images and then the right side of both images*

4.2.2. LAB MEDIA: Figure 3B. *Video editor: Highlight the top row then the bottom row*

4.3. Glycosylation sites were identifiable at four positions on the HA head and stem, confirming visibility of glycan density in the final reconstruction **[1]**.

4.3.1. LAB MEDIA: Figure 3C. *Video editor: Please highlight the red arrows in the left image and then the 4 boxes on the right side.*

Pronunciation Guide:

1. **hemagglutinin**
   Pronunciation link: https://www.merriam-webster.com/dictionary/hemagglutinin Merriam-Webster
   Also: HowToPronounce gives IPA = heməˈgluːtɪnɪn How To Pronounce
   IPA: /ˌhiːməˈgluːtənɪn/
   Phonetic Spelling: hee-muh-GLOO-tuh-nin
2. **subnanometer**
   (prefix "sub-" + "nanometer")
   Pronunciation link: no confirmed link found
   IPA: /ˌsʌbˈnænəˌmiːtər/
   Phonetic Spelling: sub-nan-uh-MEE-tər
3. **cryo-electron tomography**
   - **cryo**
     Pronunciation link: https://www.merriam-webster.com/dictionary/cryo Howjsay
     IPA: /ˈkraɪoʊ/
     Phonetic Spelling: KRY-oh
   - **electron**
     Pronunciation link: https://www.merriam-webster.com/dictionary/electron
     IPA: /ɪˈlɛktrɒn/
     Phonetic Spelling: ih-LEK-tron
   - **tomography**
     Pronunciation link: https://www.merriam-webster.com/dictionary/tomography
     IPA: /təˈmɑgrəfi/
     Phonetic Spelling: tuh-MOG-ruh-fee
4. **pleomorphic**
   Pronunciation link: no confirmed link found
   IPA: /ˌpliːəˈmɔːrfɪk/
   Phonetic Spelling: PLEE-uh-MOR-fik
5. **glycoprotein**
   Pronunciation link: https://www.merriam-webster.com/dictionary/glycoprotein
   IPA: /ˌglaɪkoʊˈproʊtiːn/
   Phonetic Spelling: GLY-ko-proh-TEEN
6. **convolutional** (as in "convolutional neural network")
   Pronunciation link: https://www.merriam-webster.com/dictionary/convolutional
   IPA: /ˌkɑːnvəˈluːʃənl/
   Phonetic Spelling: con-vuh-LOO-shun-uhl
7. **neural**
   Pronunciation link: https://www.merriam-webster.com/dictionary/neural
   IPA: /ˈnjʊrəl/
   Phonetic Spelling: NOO-rəl
8. **averaging** (as in "subtomogram averaging")
   Pronunciation link: https://www.merriam-webster.com/dictionary/averaging

IPA: /ˈævərɪdʒɪŋ/
Phonetic Spelling: AV-uh-rig-ing

9. **angstrom(s)**
   Pronunciation link: https://www.merriam-webster.com/dictionary/angstrom
   IPA: /ˈæŋstrəm/
   Phonetic Spelling: ANG-strum
10. **glycosylation**
    Pronunciation link: no confirmed link found
    IPA: /ˌglaɪkəˈsleɪʃən/
    Phonetic Spelling: gly-kuh-SLAY-shun
11. **resolution**
    Pronunciation link: https://www.merriam-webster.com/dictionary/resolution
    IPA: /ˌrɛzəˈluʃən/
    Phonetic Spelling: rez-uh-LOO-shun
12. **interface**
    Pronunciation link: https://www.merriam-webster.com/dictionary/interface
    IPA: /ˈɪntərˌfeɪs/
    Phonetic Spelling: IN-ter-face