

User Manual for Tomography-Guided 3D Reconstruction of Subcellular Structures (TYGRESS)

CURRENT STATUS: POSTED

Zhiguo Shang

University of Texas Southwestern Medical Center

Kangkang Song

University of Texas Southwestern Medical Center

Xiaofeng Fu

University of Texas Southwestern Medical Center

Xiaochu Lou

University of Texas Southwestern Medical Center

Nikolaus Grigorieff

Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147, USA

Daniela Nicastro

Departments of Cell Biology and Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

✉ daniela.nicastro@utsouthwestern.edu *Corresponding Author*

DOI:

10.21203/rs.2.16083/v1

SUBJECT AREAS

Cell Biology *Structural biology*

KEYWORDS

cryo-electron microscopy , *cryo-electron tomography*

Abstract

Recent advances in cryo-electron microscopy (cryo-EM) are paving the way to determining isolated three-dimensional (3D) macromolecular structures at near-atomic resolution using single-particle cryo-electron microscopy (SP-cryo-EM). However, determining the subcellular structures in intact cells and organelles using cryo-electron tomography (cryo-ET) and subtomogram averaging, another cryo-EM technique, with comparable resolution remains a challenge. Current methodologies can only reach a resolution of several nanometers in most samples studied. Here, we introduce a new hybrid method, called Tomography-Guided 3D Reconstruction of Subcellular Structures (TYGRESS) that is able to achieve structural determination of subcellular structures within their natural crowded environment with nanometer-resolution by combining the advantages of cryo-ET and SP-cryo-EM.

Introduction

SP-cryo-EM and cryo-ET both generate three-dimensional (3D) reconstructions of native biological macromolecules under cryogenic conditions. However, whereas SP-cryo-EM can be used to determine the structure of isolated macromolecules with up to atomic resolution ($\sim 3\text{\AA}$), the resolution achieved by cryo-ET of intact cells or organelles paired with subtomogram averaging lags approximately one order of magnitude behind ($\sim 30\text{\AA}$). Here, we introduce a hybrid-method called “TomographY-Guided 3D REconstruction of Subcellular Structures” (TYGRESS), which combines the advantages of both SP-cryo-EM (images with good signal-to-noise ratio/contrast and minimal radiation damage of the sample) and cryo-ET averaging (extraction and 3D alignment of macromolecules contained in a complex cellular sample). Ultimately TYGRESS is a SP-reconstruction, but the parameters from subtomogram averaging are critical to guide particle picking and image alignment – steps that usually prohibit SP-cryo-EM of complex cellular samples.

In this method, a typical high dose (HD) single particle cryo-EM micrograph is acquired (i.e., with an electron dose that is higher than for individual tile series images) immediately prior to a conventional low dose (LD) cryo-ET tilt series. The LD cryo-ET data set are processed, including subtomogram averaging of the particle of interest. The parameters, i.e., the particle position and alignment, are

then used to guide the particle picking in the HD single-particle cryo-EM data set and initialize its angular alignment. With this information from cryo-ET, single particle image processing techniques can be applied to subcellular samples, despite the superposition of many structures in the SP projection image.

Equipment

Hardware requirements and software installation

A computer cluster and/or workstation with GPU with shared storage is recommended for running TYGRESS. Before installation of TYGRESS, please ensure the software packages below are installed by following the installation guide of each package. You can also find the hardware and operating system requirements in the distribution pages of the following software packages.

MATLAB: The major running environment for TYGRESS. Use the R2014b version of MATLAB for the best results.

IMOD & PEET: Tilt series alignment, tomogram generation, and subtomogram averaging.

(<http://bio3d.colorado.edu/imod/> & <http://bio3d.colorado.edu/PEET/>)

CTFFIND3: Detect defocus value (<http://grigoriefflab.janelia.org/ctf>)

EMAN2: 'Normalization' function (<https://blake.bcm.edu/emanwiki/EMAN2>)

FREALIGN: 3D reconstruction (<http://grigoriefflab.janelia.org/frealign>)

Bfactor: Apply B-factor for the final reconstruction

(http://grigoriefflab.janelia.org/grigorieff/download_b.html)

Procedure

1: Preparation to run TYGRESS

1.1 To use TYGRESS on the example data set, download and unzip the TYGRESS source code and example data set package (www.tygress.org), and then follow Procedure step 2 ("Preparation of TYGRESS input files" below).

Note □ TYGRESS is also available on Code Ocean as a compute capsule

(<https://doi.org/10.24433/CO.2034333.v1>). Please read the readme file on the Code Ocean TYGRESS page for more details.

1.2 To use your own data set, please finish the data collection, tomogram reconstruction, and subtomogram averaging before following Procedure 2, below.

1.2.1 Data collection. During data collection, two data sets need to be acquired for each region of interest. (1) First, acquire a 2D image at 0 degree using an electron dose typical for conventional SP-cryo-EM (HD image), which is used for final TYGRESS reconstruction; (2), immediately after collecting the HD image, collect a traditional low-dose tilt series (LD images) of the same region, which is used for retrieving the position information for each particle of interest after tomogram reconstruction and subtomogram averaging. Note: to minimize the effects of radiation damage in the final reconstruction, the HD image must be recorded prior to the LD images.

1.2.2 Motion correction. If movie stacks were taken during the data collection, run the whole-frame motion correction using IMOD scripts.

1.2.3 Alignment of the combined tilt series. Each HD image needs to be inserted into its corresponding tilt series at the angle where the HD image was taken using the “newstack” command in the IMOD package, resulting in a “combined tilt series”. For example, an HD image recorded at 0 degrees should be inserted just before the LD image at 0 degrees. After tilt series alignment using the IMOD package, the combined HD and LD images ensure a common reference frame for the later steps in the TYGRESS procedure.

1.2.4 Tomogram reconstruction. After alignment, only the LD images are used to generate the tomogram using the IMOD package.

1.2.5 **Subtomogram averaging.** The subtomogram averaging can be performed using the PEET package in IMOD. Run PEET averaging for each tilt series and all of the particles of interest, respectively.

2: Preparation TYGRESS input files.

2.1 Prepare two TYGRESS input text files.

2.1.1 There is one file named as “TYGRESS_input.txt” in the TYRESS source code folder (as shown below in the example data set), which contains all the parameters to run TYGRESS, such as the work directory, file path, 3D reconstruction parameters, *etc.* Update these parameters as needed to reflect the location and organization of your files. If you are following along with the example dataset, modify the following:

- fnDataDir= (copy the full path name for the example data folder that you downloaded from this tutorial)
- fncode= (copy the full path name for the TYGRESS source code folder that you downloaded from this tutorial)
- fnref= (copy the full path name for the .mrc file within the example folder)

Example of a “TYGRESS_input.txt” file

=====

Part I Parameters for running TYGRESS

```
fnDataDir = /project/cellbiology/Nicastro_lab/nusr/shang/linux/Tygress/example/ % project data
directory
```

```

fncode = /home2/s165838/Tygress/Tcoden/Tcode/                                % TYGRESS
code directory
fnWorkDir = combined_PEET/                                                % combined
peet directory
fnDef = ctffind/DefocusList.txt                                          % defocus
file
fnref = /home2/s165838/Tygress/ref.mrc                                    % reference masked after PEET average
fnCombinedHD = HD_particles/                                              % high dose 2D particle directory
fnmod = combined_mod_csv/                                                % combined mod & csv directory
fnCSV = lter1_particles/                                                  % updated peet run for whole dataset
csvDirname = ltr1/                                                        % csv directory
csvname_h = _ltr1_MOTL_Tom                                               % csv root name after PEET of individual
tilt series
csvname_t = _lter5.csv                                                    % csv tail name after PEET of individual
tilt series
modname_h = A1.mod.FOIE.                                                  % mod root name after PEET of
individual tilt series
modname_t = .txt.mod.txt.Ax_Pred_.txt.mod                                % mod tail name after PEET of individual
tilt series
fnPEETRoot = example                                                      % root name for PEET running
particleSize = 180                                                        % particle size for PEET running
fnDist = N                                                                  % file for correct distortion (gif
camera), 'N' for no correction
tail_st = _fused.st                                                       % entire tilt date set
tail_hd = _HD.st                                                          % file name use for save HD
micrograph
amplitude_contrast = 0.14                                                % thin ice sample use 0.07

```

```

and thick ice use 0.14

axoRepeat =180                % pixel size for particle (used for remove edged particles) in high
dose micrograph

Mag1 = 22500                  % micrograph magnification

PickSize = 300                % size for pick 2D high dose particle

sizex = 200                   % size for final out put structure

root_csv = _MOTL_Tom          % csv root name after run entire PEET

tail_csv = _lter2.csv         % csv tail name after run entire PEET

SizeRef = 180                 % size of reference volume

sx = 4                         % shift range of coordinates, nm

sa = 2                         % shift range of orientation, degree

parallel_f = 1                % set 1 for run job in parallel job

Ncpu = 5                       % cup number for parallel job

tail_rec = _fused_exHD.rec    % tail part of rec file

pixel_size = 5.562            % angstrom per pixel for raw particle

kV = 300                       % voltage

Cs = 2                         % value of spherical aberration coefficient of objective in mm

use_ctf = 1                    % set 1 for ctf correction

out_mrc = 96nm                 % output reconstruction file

imageSize = 2048               % entire raw image size K*K

scalescale = 0.6

n_neighbor = 2                 % number of neighbored particles which used for background
reference

dis_neighbor = 9                % distance (# pixels) used for selecting nearby particles
for background reference

```

Part II Parameter file for reconstruction, visit FREALIGN webpage for more details

data_input	96nm
raw_images1	96nm
thresh_reconst	90.0
thresh_refine	90.0
pixel_size	5.562
dstep_frealign	13
WGH	0.14
kV1	300.0
Cs	2.0
RI	0.0
RO	3000.0
PBC	2.0
BOFF	75.0
ASYM	0
DANG	1
ITMAX	10
MODE	1
XSTD	0.0
RBFACT	0
FPART	F
FMAG	F
FDEF	F
FASTIG	F
IEWALD	0
dfsig	100
FBEAUT	F

FCREF	F
res_reconstruction	3.0
res_low_refinement	800.0
res_refinement	3.0
start_process	1
end_process	1
first_particle	1
last_particle	471
increment	10

=====

Note When modifying the “TYGRESS_input.txt” file, please use the full path name to the directories for 'fnDataDir', 'fncode', and 'fnref', and only use the relative file folder names for the other parameters when needed.

2.1.2 The file named “DataList.txt” in the TYGRESS source code folder (as shown below in the example data set) has three columns containing the “data set main name” + “data set sub name” + “total .mod number”, which you should change as needed.

Example of a “DataList.txt” file

=====

```
WT_20120914 S1 9
WT_20120914 S2 9
WT_20120914 S3 9
WT_20120914 S4 9
WT_20120914 S5 9
```

=====

2.2 Modify the “ctffind.csh” file in the TYGRESS source code folder, which will be used to determine CTF. Please see the example in the “ctffind.csh” file for details. You may at least need to change the path of the “ctffind3.exe” file. For more details about how to modify this file, please visit (http://grigoriefflab.janelia.org/sites/default/files/readme_ctf.txt).

2.3 Modify the path of your FREALIGN installation in the “mreconstruct_noMask.com” file. Please see the example “mreconstruct_noMask.com” file for details.

3: Run TYGRESS

Open the Matlab interface, and set the TYGRESS source code folder as the work path. Then, in Matlab, open the TYGRESS_GUI (Figure 1). Click the “Tygress parameter file” button, input the “TYGRESS_input.txt” file, click the “Data list” button, and input the “DataList.txt” file. Click “Yes” on the corresponding step to have TYGRESS run that step (Figure 1). All steps should be run in order if it is your first time processing your TYGRESS data. (Note: You can process the steps one by one, or you can run all selected steps together.)

Final 3D reconstruction is performed with FREALIGN, and all output is saved under the “Reconstruction” folder. For more information about FREALIGN, please visit (<http://grigoriefflab.janelia.org/frealign>).

4: Directory tree of TYGRESS (Figure 2)

Please note:

- .mod(1) and .csv(1) are the parameters after running PEET for individual tilt series
- .mod(2) and .csv(2) are the parameters for running PEET for the entire data, and .csv(3) is the output

- .csv(3) and mod(2) are used for TYGRESS alignment and to generate the .par file
- .mrc(1) is the 2D particle data set and .mrc(2) is the final output for the 3D structure
- .txt(1) is the parameter for defocus and .txt(2) coordinates parameters of each particle
- .m is the code for running TYGRESS
- **Project data** folder contains all tilt series data; each tilt series is processed (alignment and reconstruction in IMOD) in its own particles folder, including particles_Predict and lter folders (alignment results from individual PEET runs).
- **Combined PEET** folder includes combined alignment for all 3D subtomogram particles in PEET (lter1_particles); Combined_mod_csv contains the file for running PEET for the entire dataset; 2D HD particles are picked in each data set under the HD_particles folder. The defocus detection for HD images is put in the ctffind folder.
- **TYGRESS code** folder stores all codes needed for running TYGRESS. The 'm' extension indicates matlab scripts, which can be only executed in Matlab. The structure is output as a .mrc file.
- **Reconstruction** folder stores all output results after reconstruction with FREALIGN.

Time Taken

In the example data set, which includes five tilt series, each tilt series includes all the files generated after tomogram reconstruction and subtomogram averaging. The PEET averaging for the whole dataset is also included. Therefore, you can run TYGRESS directly after changing the work directories in the "TYGRESS_input.txt" file. The total TYGRESS processing time from CTF finding to final reconstruction will take less than 30 minutes on a common computer cluster.

Anticipated Results

5.1 *.mod(1) and *.csv(1) are the output files after running PEET for each individual tilt series.

5.2 You can copy the *.mod(1) and *.csv(1) to the Combined_mod_csv folder as *.mod(2) and *.csv(2) to run the PEET for the whole datasets, and then you will get *.csv(3) after PEET. All these .mod and .csv files are obtained before you starting to run TYGRESS.

5.3 After you run **Step 1** in the TYGRESS GUI, the “ctffind” folder will be created and the defocus information will be saved in the *.txt(1) file. Moreover, the “HD_particle” folder will be created and all HD images will be saved in its sub_folder.

5.4 After you run **Step 2** in the TYGRESS GUI, the program will use the information from *.mod(2) and *.csv(3) to pick particles from the HD images, and then save all picked 2D particles as *.mrc file and save the coordinate parameters of each picked particle as *.txt(2) file, and alignment parameters as *.par file.

5.5 After you run **Step 3** in the TYGRESS GUI, the program will align the 2D particles picked and update the *.par file. **Note:** for further constrained alignment, you need to update your 'sx' (for x, y shift) and 'sa' (for angular shift) values in the “TYGRESS_input.txt” file, use the *.par file created in **Step 3** to replace the *.par file generated in **Step 2**, and re-run **Step 3**.

5.6 After you run **Step 4** in the TYGRESS GUI, the program will generate the 3D reconstruction and save the *.mrc(2) file.

References

This protocol is related to the following article:

Song, K., Shang, Z., Fu, X., Lou, X., Grigorieff, N., and Nicastro, D. (2018). Structure of the ciliary axoneme at nanometer resolution reconstructed by TYGRESS. *BioRxiv* 363317.

Figures

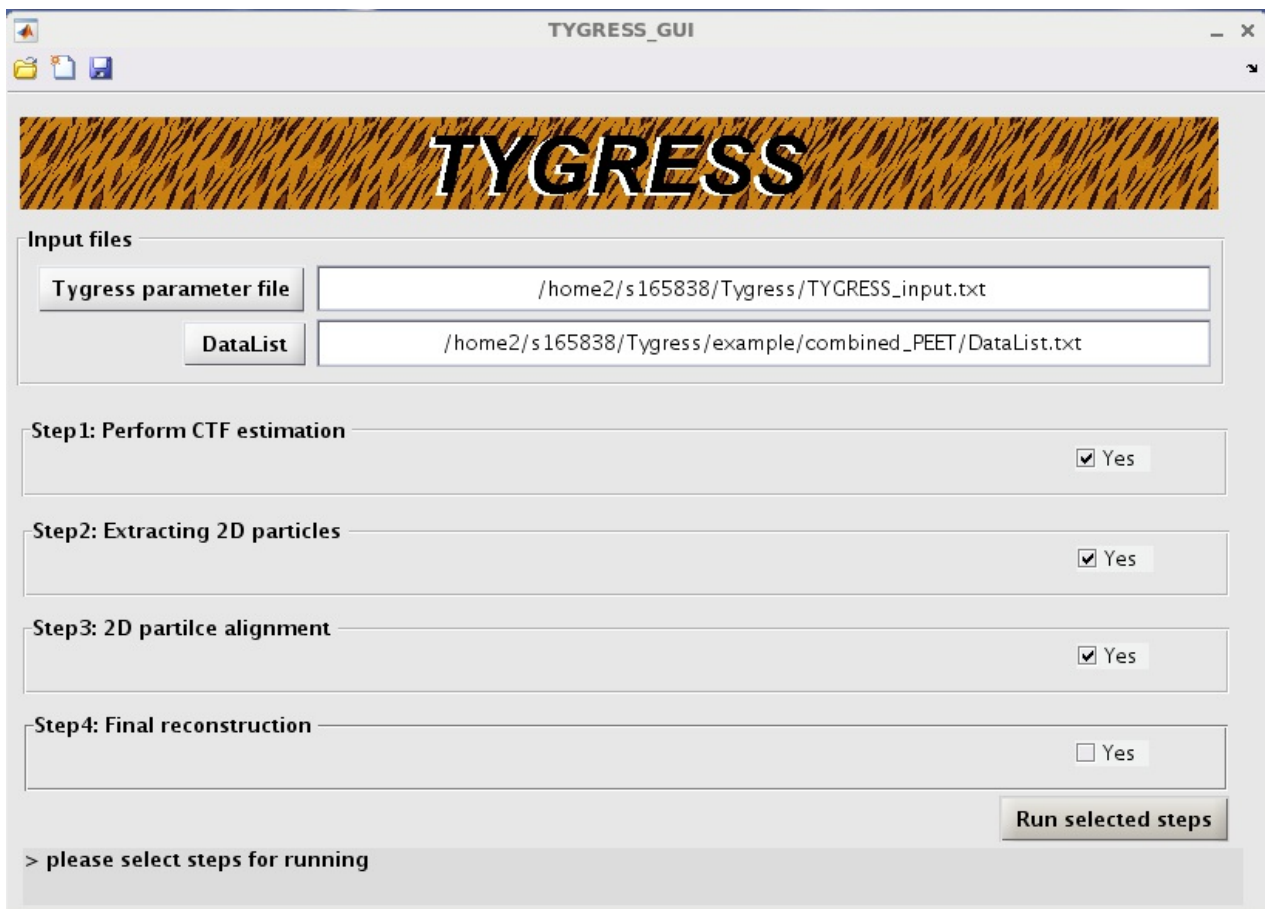


Figure 1

Example of TYGRESS GUI

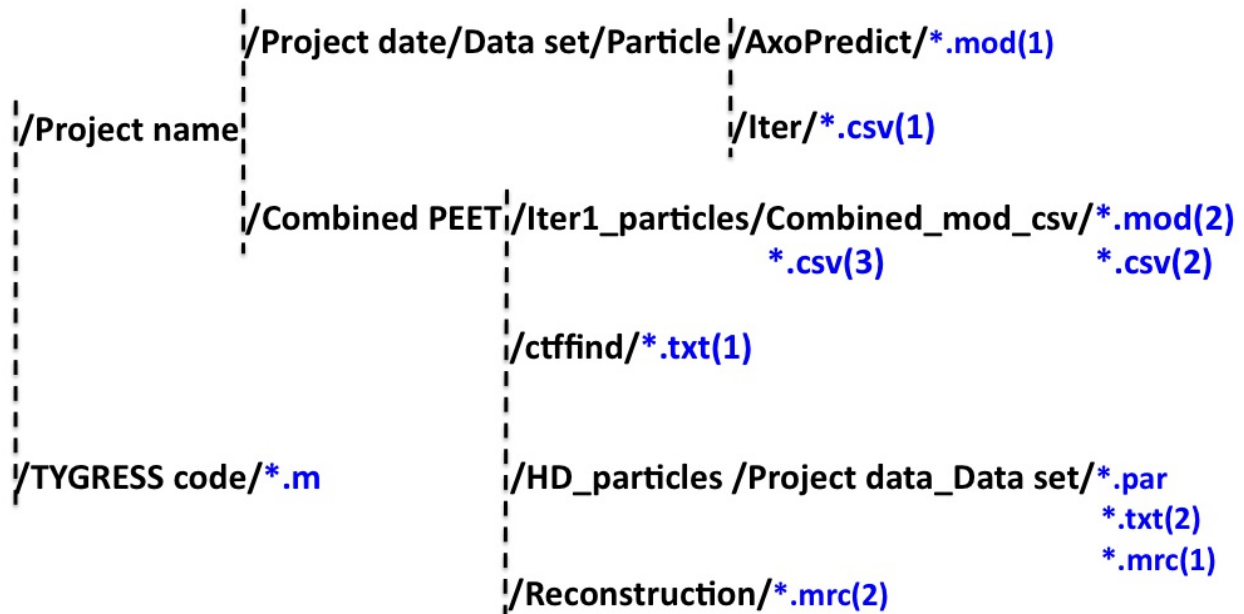


Figure 2

Work directory tree of TYGRESS

10.1038/s41592-019-0651-0

Song K., Shang Z., Fu X., Lou X., Grigorieff N., Nicastro D., In situ structure determination at nanometer resolution using TYGRESS. Nature Methods. (Accepted: 11 October 2019)