Phenix *Python-based Hierarchical ENvironment for Integrated Xtallography*

Cryo_fit1 FAQ

Contents

How can I provide gaussian filtered maps?

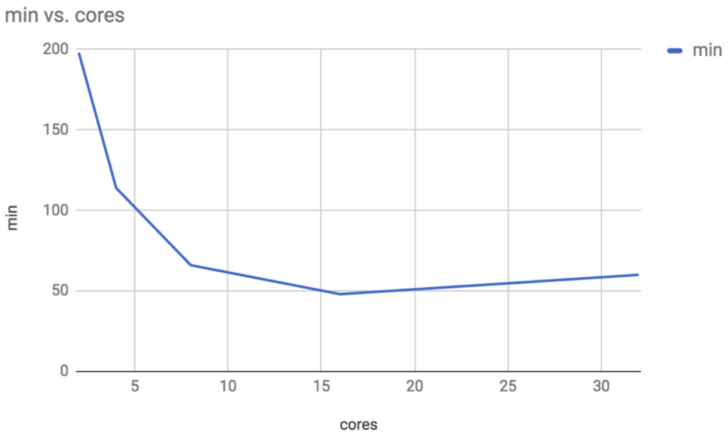
- See 'How can I provide gaussian filtered maps?' in cryo_fit2_FAQ

How long does it take to run cryo_fit?

How long does it take to run cryo_fit?

- CentOS with 2 cores (2.7 Ghz)
  - tRNA (6,000 atoms) : 2.5 minutes
  - Nucleosome (26,000 atoms): 1 hr
  - Beta-galactosidase (64,000 atoms): 2.5 hrs
  - ribosome (386,000 atoms): 3~7 hrs
- macOS
  - 24 cores (2.7 Ghz)
    - (4~16 cores would give similar/better performance)
    - Beta-galactosidase (64,000 atoms): 3.9 hrs
  - 4 cores (3.1 Ghz)
    - Beta-galactosidase (64,000 atoms, 50,000 cryo_fit steps): 5.8 hrs

Benchmark with ribosome (same linux machine, 10k emsteps, 50k number_of_steps, 3x emweight)



- Based on this result, number of cores is recommended up to 16

How to enlarge map box size?

- One can know map size by
  1. cryo_fit which reports a map size between step 6 and step 7 with codes of
     - from iotbx import ccp4_map
     - ccp4_map = ccp4_map.map_reader(user_input_map)
     - target_map_data = ccp4_map.map_data()
  2. EMAN2's e2iminfo
     - For example, e2iminfo.py <user>.mrc
     - (note) As of EMAN2.2, the file extension should be mrc, not ccp4
  3. VMD which can visualize current map size as in this FAQ.
- Then, relion_image_handler can enlarge map box size.
  - For example, relion_image_handler --i <user>.mrc --new_box 370 --o <user_box_size_370>.mrc

How to extract a relevant map region?

- Question: I only want to fit the monomer because only one monomer of the trimer map is the best. Do you have any idea of how to generate the monomer map that cryo_fit can use?
- Answer: Please use phenix.map_box to extract local map.
  - Alternatively, you can try phenix.map_to_model.

How to generate and record movie?

- Generate_record_movie_by_cryo_fit1

How to improve initial cc?

- phenix.dock_in_map, UCSF Chimera's 'Fit in Map (Tools -> Volume Data -> Fit in Map)', UCSF Chimera's manual fitting (using a mouse) often improve initial cross-correlation (cc) between map and model. UCSF ChimeraX's isolde may improve initial cc as well.

I can't run phenix.superpose_pdbs with cryo_fitted pdb file

- phenix.superpose_pdbs can superimpose only between pdb files that have equally/similarly aligned nucleic names. This applies to all pdb input files (not only cryo_fitted files). Consider to align both input files by "python <Phenix Path>/modules/cryo_fit/steps/9_after_cryo_fit/align_nucleic_acid_name_into_middle/align_nucleic_acid_name_into

I observe cc values keep decreasing (or do not increase that much) during step 8

- Initial few steps of cryo_fit in step 8 may decrease cc values temporarily as an initial perturbation.
- However, if it keeps decreasing continuously (or do not increase that much), it may indicate two possible cases.
    - Case 1 <A user input pdb file has reasonable structural geometry> For this case, try below possible solutions. If cryo_fit still can't find better (higher) cc, then the initial correlation between user input pdb file and cryo-EM map is already high enough. Just run phenix.real_space_refine only and deposit. Doo Nam recommends to argue/claim in a paper like "Our fit is so high, even cryo_fit did not find higher level of fit than our initial fit".
    - Case 2 <A user input pdb file has unreasonable structural geometry> For this case, although the initial fit between user's atomic model and map looks good, it is a fictitious fitting without consideration of ideal atomistic model geometry. Run cryo_fit to find decent fit to the cryo-EM map that restores/maintains reasonable structural geometry based on molecular dynamics forcefield (amber03.ff).
- There are possible solutions
    1. Provide a higher (better) resolution map tends to find higher cc.
    2. Cryo_fit calculates the gradient of cc. If cryo_fit is provided a giant cryo-em map with a tiny atomic model, then there is a large empty space (not filled). Therefore, the constraint forces for the model are very small. Consequently, these small forces are not helpful.

       2-1. Re-run cryo_fit with an atomic model that fits the majority of the map. If you fit multiple atomic models into a symmetric map or do sequential fitting into a non-symmetric map, watch Tom Goddard's lecture (2013)

       2-2. Re-run cryo_fit with only a relevant map region. See 'How to extract a relevant map region?' in this FAQ.
    3. If the initial model is not properly aligned to a map, see 'How to improve initial cc?' in this FAQ.
    4. A user may enforce a stronger initial map weight (e.g. emweight_multiply_by) manually. However, since cryo_fit automatically increases emweight_multiply_by if it doesn't find higher cc anyway, manual increase in map weight may not necessarily find higher cc.

I see "Fatal error: A charge group moved too far between two domain decomposition steps. This usually means that your system is not well equilibrated" at my 8_cryo_fit step.

- Using macOS 10.13.6 helped rather than using Ubuntu 16.04. Maybe macOS has better numerical stability.
- One gromacs expert suggested to try smaller time_step_for_cryo_fit.
- However for Doo Nam, simply using macOS solved the problem.
- Less likely, but still a possible reason is that the map weight is too high. Therefore, lowering emweight_multiply_by may help.

I see "Fatal error: Incomplete ring in HIS50" at my 8_cryo_fit step

- Doo Nam observed this error when he tried to run cryo_fit with 3jch.pdb that has incomplete Histidine ring atoms.
- He solved with UCSF Chimera

1. Open pdb file
2. [menu]
3. Select -> Residue -> HIS
4. Tools -> Structure Editing -> Rotamers -> OK
5. (select the most probable rotamer each)
6. File -> Save PDB

- Pymol and Swisspdb viewer have similar residue fixing functions
- Alternatively, a user may identify residues manually by running <phenix-xxxx>/modules/cryo_fit/steps/0_prepare_cryo_fit/count_number_of_atoms_in_each_residue/

I see "Fatal error: Number of grid cells is zero. Probably the system and box collapsed." at my 8_cryo_fit step.

- step_8 may be full of stepxb_nx.pdb.
    - Most likely, this means that initial cc is too low for MD simulation.
        - When Doo Nam ran real_space_refine first, then run real_space_refined atomic model in cryo_fit, it was solved.
        - Alternatively, improve initial cc by fitting initial atomic model into a map (see "How to improve initial cc?" in this FAQ)
    - Less likely, but still a possible case is when the map weight is too high, lowering emweight_multiply_by may help.

I see "pdb file cleaning is not done" at my step 1 (Make gro and topology file by regular gromacs)

"I edited out lipids, HEM and other hetero atoms and I verified that they are all gone. However, still my pdb file is not clean enough for gromacs_cryo_fit".

- If your lipids, HEM and other hetero atoms are not necessary for now, you may take out as you did. You may add those atoms after flexible fit (cryo_fit). Please send Doo Nam (doonam@lanl.gov) your pdb input file only (no map). He will take a look. gromacs_cryo_fit uses amber03 which may not have forcefield parameters for your less-usual atoms/residues.
- If your lipids, HEM and other hetero atoms are necessary for now and your target is protein, Doo Nam would recommend cryo_fit2 since it supports phenix.eLBOW derived ligand parameters.

I see "state.cpt not found, step_8 may be full of stepxb_nx.pdb" at my 8_cryo_fit step

- Literally, steps/8_cryo_fit fold may be full of stepxb_nx.pdb files.
    - Most likely, this means that a cryo_fit input pdb file is not yet stable enough for sensitive Gromacs MD simulation (step 8) even after cryo_fit's ample minimization step (e.g. step 4).
        - When Doo Nam ran real_space_refine first, then run real_space_refined atomic model in cryo_fit, it was solved more than 2 cases.
        - If initial cc is too low, improve initial cc by fitting initial atomic model into a map,
            - either by following "How to improve initial cc?" in this FAQ
            - or adding/copying more atoms.
                - For example, when Doo Nam tried to fit a monomer into a trimer map, this error occured. However, simply adding dimer atomic models into an input pdb file and running cryo_fit again removed the error.
        - Less likely, but still a possible case is when the map weight is too high, lowering emweight_multiply_by may help.

I see "step 0 correlation coefficient: nan" at my 8_cryo_fit step.

- This often indicates that the initial atomic model is not placed into a cryo-EM map.
    - Therefore, please improve initial cc by fitting initial atomic model into a map (see "How to improve initial cc?" in this FAQ)
- If a user observed "Range checking error: Explanation:" error message at 8_cryo_fit step, please refer <I see "step 0 correlation coefficient: nan" and "Range checking error: Explanation:"> part in this FAQ

I see "step 0 correlation coefficient: nan" and "Range checking error: Explanation: During neighbor searching, we assign each

particle to a grid based on its coordinates. If your system contains collisions or parameter errors that give particles very high velocities you might end up with some coordinates being +- Infinity or NaN (not-a-number). Obviously, we cannot put these on a grid, so this is usually where we detect those errors. Make sure your system is properly energy-minimized and that the potential energy seems reasonable before trying again." at my 8_cryo_fit step.

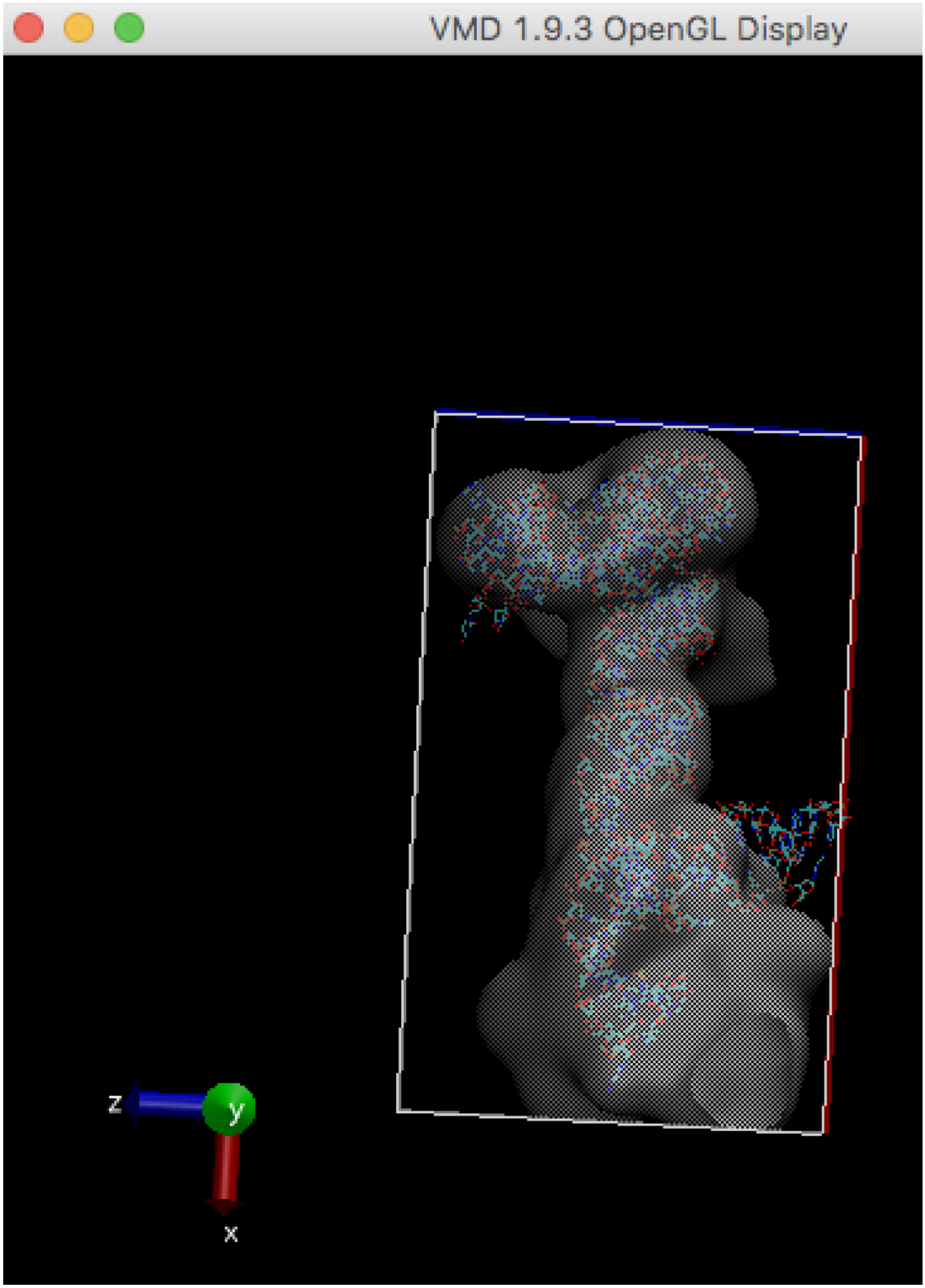
- Based on Doo Nam's experiences, this error message may indicate between two cases/scenarios.

    1. Doo Nam observed this error when starting molecule is not stabilized structurally. There are two possible solutions.

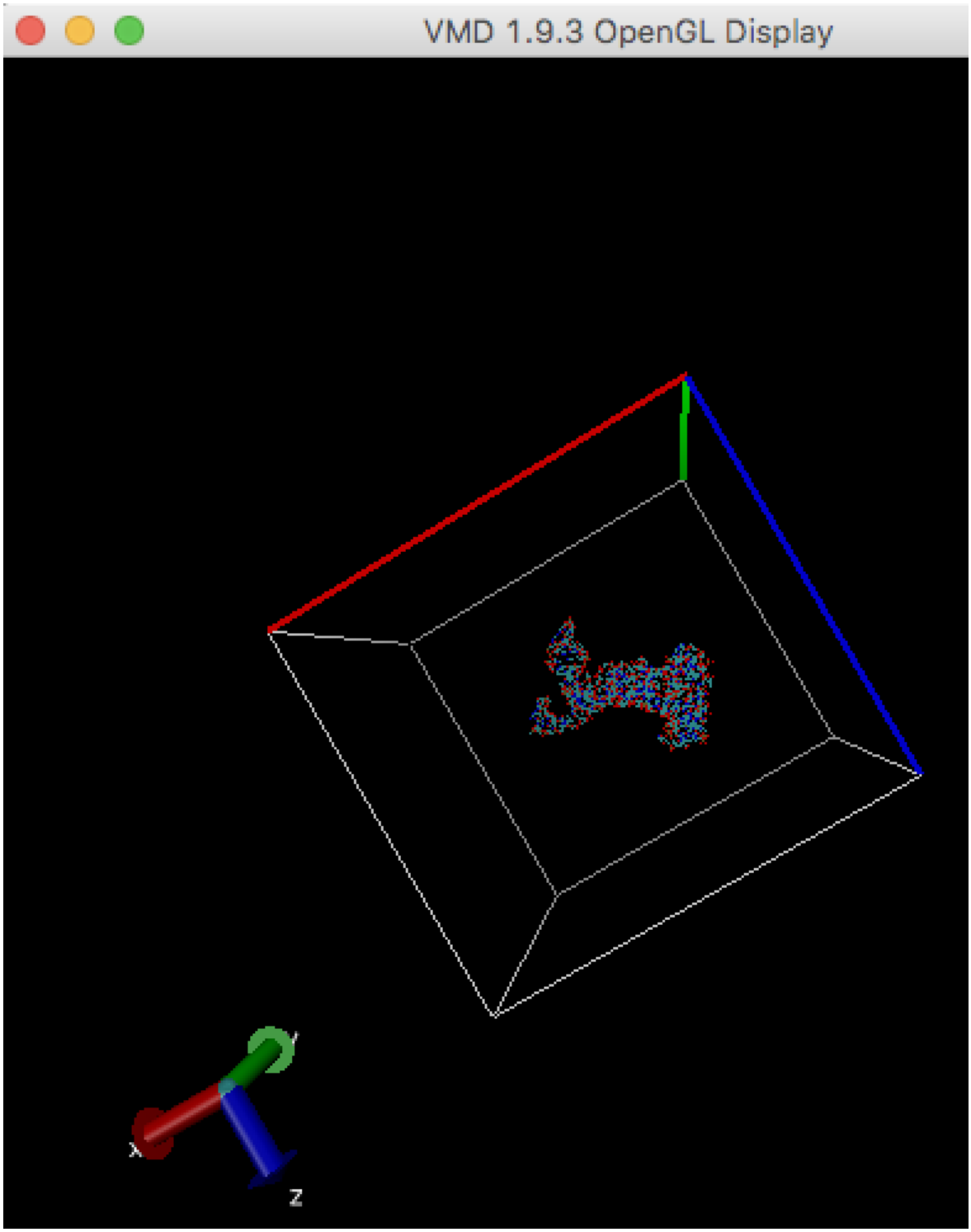
        - Run phenix.real_space_refine before cryo_fit. Then, provide real_space_refined molecule into cryo_fit. Then, most problems were solved.
        - When Doo Nam skipped minimization, this error occured. Since most users minimize the starting structure by default (cryo_fit default), most users do not need care about this. Besides, as of 9/18/2019, the default value of number_of_steps_for_minimization=20000 (which is a lot).

    2. It means that the map dimensions need to be larger. Therefore, check map size and solve a problem with VMD/EMAN2/relion like followings.

        - Like other MD simulations, gromacs need enough map box size to cover the atomic model to run (ziggle and wiggle). Refer Waters seems to be out of the box
        - For example, stuck-out red oxygen atoms outside the right edge of the box are the problem.

- 

- In order to run any MD simulation (including cryo_fit), a box should be large enough like



- Make map box size larger (see "How to enlarge map box size?" in this FAQ), and run cryo_fit again. You can check map box size by VMD. Alternatively, remove sticking out atoms if these are unnecessary, then run cryo_fit again.
- For protein modeling, I would use cryo_fit2 which is not limited by box size requirement. Most of the time, it better fits than cryo_fit1 in terms of fitting and geometry preservation anyway.

I see "The initial cell size (xxx) is smaller than the cell size limit (xxx), change options -dd, -rdd or -rcon, see the log file for details" at my 8_cryo_fit step

- Doo Nam observed this error for 2 cases
    1. When he provided a pdb file that has unspecified/deleted region of a molecule. Even when vmd assisted map box/cell dimension is larger than initial molecule space, this error

appeared.
2. When he provided a pdb file that has small number of residues (e.g. 5~16 amino acids). Even when vmd assisted map box/cell dimension is larger than initial molecule space, this error appeared. This error appeared regardless of the existence of CRYST1 and SCALE header.

I see "Too many LINCS warnings" at my 8_cryo_fit step

- Doo Nam observed this error when starting molecule is not stabilized structurally. There are 3 possible solutions.

    1. Run phenix.real_space_refine before cryo_fit. Then, provide real_space_refined molecule into cryo_fit. Then, most problems were solved.
    2. When Doo Nam skipped minimization, this error occured. Since most users minimize the starting structure by default (cryo_fit default), most users do not need care about this. Besides, as of 9/18/2019, the default value of number_of_steps_for_minimization=20000 (which is a lot).
    3. If starting molecule has non-standard chemical geometry that is too extreme that are not stabilized neither by real_space_refine nor by cryo_fit minimization, then a user needs to fix those non-standard geometry manually.

I see "User's provided atomic model had 0.0 cc" in my cryo_fit.overall_log.

- It means that literally user's provided atomic model has no correlation to map to start with.
- See "How to improve initial cc?" in this FAQ, and provide initially aligned atomic model to cryo_fit.