A cryo-EM map as the target - how to evaluate predictions

Maciej Antczak[1,2], Jakub Morawski[1], Marta Szachniuk[1,2]

1) Institute of Computing Science, PUT
2) Institute of Bioorganic Chemistry, PAS

RNA BENASQUE 2022, 07-20.08.2022

What does *the quality of the RNA 3D model* mean?

- *The quality of the RNA 3D model* measures **quantitatively** how **structurally similar** is the model to **the reference structure**.

- **High** quality => **native-like** 3D structure,
- **Low** quality => **totally different** 3D folds.

How to evaluate the 3D model within the context of the reference structure?

- When the experimentally determined RNA structure is:
  - a high-resolution 3D structure (stored in *mmCIF*/*PDB* format).
- One can use:
  - Several measures like: *root-mean-square deviation* (*RMSD*), *interaction network fidelity* (*INF*), *deformation index* (*DI*), *mean of circular quantities* (*MCQ*), *longest continuous segments in torsion angle space* (*LCS_TA*), *template modeling score for RNA* (*TM-score*$_{RNA}$), *contact area difference-based function* (*CAD-score*), etc.
  - Several tools like: *rna-tools*, *RNA_assessment*, *RNAQUA*, *MCQ4Structures* provided by *RNA-Puzzles toolkit* [1].

[1] Magnus, M., Antczak, M., Zok, T., Wiedemann, J., Lukasiak, P., Cao, Y., ... & Miao, Z. (2020). RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic acids research*, 48(2), 576-588.

How to evaluate the 3D model within the context of the reference structure? (2)

- When the experimentally determined RNA structure is:
  - a low-resolution, cryo-EM density map (stored in *CCP4/MRC* format).
- One can compute:
  - *Cross-correlation coefficient* (*CCC*).
  - After *automatic rigid-body fitting* the 3D model into the density map.

What tools can we use?

- *MAPQ plugin* (version 1.2) *for UCSF Chimera* (version 1.11).
- *Visual Molecular Dynamics* (*VMD* – version 1.9.4).
  - https://www.ks.uiuc.edu/Training/Tutorials/science/mdff/tutorial_mdff-html/node4.html
- *Python-based Hierarchical ENvironment for Integrated Xtallography* (*Phenix*).
  - *https://phenix-online.org/documentation/reference/cryo_fit.html#how-to-fit-cryo-em-maps*
- *PowerFit* by Bonvin Lab.
  - https://github.com/haddocking/powerfit
  - https://www.bonvinlab.org/education/powerfit-webserver/

What tool we've finally chosen?

- *PowerFit* [1,2] (provided by **Bonvin Lab**).
  - Automatically **fits high-resolution atomic structures** in **cryo-EM densities** by a full-exhaustive 6-dimensional cross-correlation search.
  - Input:
    - **an atomic structure** in the *PDB* format,
    - **a cryo-EM density** with **its resolution**.
  - Outputs:
    - **positions** and **rotations** of **the atomic structure** corresponding to **high correlation values** (**the local cross-correlation function** is used).
  - Additional useful parameters:
    - rotational sampling interval (-*a*), either multiprocessing or GPU used (-*p*/-*g*).

[1] G.C.P. van Zundert and A.M.J.J. Bonvin. Fast and sensitive rigid-body fitting into cryo-EM density maps with *PowerFit*. *AIMS Biophysics* 2, 73-87 (2015).

[2] *PowerFit* open source project - https://github.com/haddocking/powerfit.

Automatization of the evaluation process

- We developed in Python **the upper layer** around the *PowerFit* tool **to optimize** its **serial execution** for **hundreds** of **RNA 3D structure predictions**.
- We are currently **applying virtualization to** easily **setup the evaluation process** in the dedicated **high-performance computing infrastructure**.

How about proteins?

- In 2019, *Cryo-EM Model Challenge* was organized [1]:
  - **assess the quality** of **models** that can be produced from cryo-EM maps using current modeling software,
  - **evaluate reproducibility** of **modeling results** from different software developers and users,
  - **compare performance** of **current metrics** used for model evaluation, particularly Fit-to-Map metrics, with focus on near-atomic resolution.

- Fit-to-Map metrics:
  - **the correlation** between **map** and **model density** as implemented in:
    - *TEMPy v.1.1* (*CCC*, *CCC_OV*, *SMOC*, *LAP*, *MI*, *MI_OV*),
    - *Phenix v.1.15.2 map_model_cc* module (*CCbox*, *CCpeaks*, *CCmask*).
  - **the resolvability** of each **model atom** in **the map** (real-space correl.):
    - ***Q-score*** (*MAPQ v.1.2 plugin for UCSF Chimera v.1.11*).
  - *Phenix Map-Model FSC*, *REFMAC FSCavg*, *EMDB Atom Inclusion*, *TEMPy ENV*, ***EMRinger***.

[1] Lawson, C. L., Kryshtafovych, A., Adams, P. D., Afonine, P. V., Baker, M. L., Barad, B. A., ... & Chiu, W. (2021). Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nature methods*, 18(2), 156-164.

Other successful strategies/methods/tools?

- I would be very grateful if you could share with me **any other practically successful approaches** for this problem☺
- Thank you in advance!

Q&A

Thank you for your attention!

WWW: `http://www.cs.put.poznan.pl/mantczak`

Contact: `Maciej.Antczak@cs.put.poznan.pl`