

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. [For final submission](#): please carefully check your responses for accuracy; you will not be able to make changes later.

### ▶ Experimental design

#### 1. Sample size

Describe how sample size was determined.

We selected reconstructions to analyze based on:  
 availability of a reconstruction in the EMDB as of Nov, 2017  
 resolution of the reconstruction 4.5 Å or better  
 presence of a unique deposited model in the PDB matching the reconstruction  
 consistent resolution in the PDB and EMDB  
 ability to use Phenix tools to automatically extract model and map from PDB and EMDB,  
 apply symmetry if present in the metadata, and write the model

This resulted in 502 map-model pairs extracted from a total of 882 single-particle and helical reconstructions in the EMDB in this resolution range. (Note that only 660 of the 882 have one or more associated PDB entries).

We downloaded the maps from the EMDB16 and used them directly in `phenix.map_to_model`, with the exception of one map (EMDB entry 6351). For EMDB entry 6351, a pseudo-helical reconstruction<sup>33</sup>, we could only deduce reconstruction symmetry for the part of the map corresponding to deposited model (PDB entry 3jar), so we used the `phenix.map_box` tool to cut out a box of density around the region defined by the deposited model, analyzed this map, and at the conclusion of the process translated the automatically-generated models to match the deposited map.

#### 2. Data exclusions

Describe any data exclusions.

After our initial analysis we further excluded reconstructions that had the following characteristics:

map-model correlation for the deposited map and deposited model of less than 0.3 after extraction of map and model and analysis with `phenix.map_model_cc` (18 reconstructions)  
 deposited model in the PDB represents less than half of the structure (9 reconstructions)

This yielded the 476 map-model pairs that are described in this work.

#### 3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

The successful application of our method to all 476 map-model pairs is our principal demonstration of reproducibility of the methods. Additionally, during the development of the software, many individual datasets were analyzed multiple times.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

There was no randomization carried out

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- Test values indicating whether an effect is present  
*Provide confidence intervals or give results of significance tests (e.g.  $P$  values) as exact values whenever appropriate and with effect sizes noted.*
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

## 7. Software

Describe the software used to analyze the data in this study.

The Phenix software was used for this study. The source code and documentation is available at [www.phenix-online.org](http://www.phenix-online.org)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). [Nature Methods guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

## 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No restrictions are present

## 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

## 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Eukaryotic cell lines were not used

b. Describe the method of cell line authentication used.

Eukaryotic cell lines were not used

c. Report whether the cell lines were tested for mycoplasma contamination.

Eukaryotic cell lines were not used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Eukaryotic cell lines were not used

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

## 11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No animals were used

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human subjects were used