nature portfolio

Corresponding author(s):	Chenglong Bao
Last updated by author(s):	Oct 24, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	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 all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	In-SilicoTEM v2.1.0, SerialEM, MatLab R2018b
Data analysis	RELION 4.0.0, CryoSPARC v4.1.0, CisTEM v1.0.0-beta, Scipion v3.0.12, Chimera v1.14, CryoSieve: https://github.com/mxhulab/cryosieve

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw final stack datasets analyzed in this study were downloaded from the EMPIAR repository (EMPIAR-10024 [https://www.ebi.ac.uk/empiar/EMPIAR-10024/], EMPIAR-11233 [https://www.ebi.ac.uk/empiar/EMPIAR-11233/], EMPIAR-10097 [https://www.ebi.ac.uk/empiar/EMPIAR-10097/], EMPIAR-11120 [https://www.ebi.ac.uk/empiar/EMPIAR-11120/], EMPIAR-10264 [https://www.ebi.ac.uk/empiar/EMPIAR-10264/], EMPIAR-10330 [https://www.ebi.ac.uk/empiar/EMPIAR-10330/], EMPIAR-10269 [https://www.ebi.ac.uk/empiar/EMPIAR-10269/], EMPIAR-10200 [https://www.ebi.ac.uk/empiar/EMPIAR-10200/]). Atomic

coordinates from Protein Data Bank 6PCQ [https://www.rcsb.org/structure/6pcq] were used for the generation of simulated particles using InSilicoTEM v2.1.0.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We analyzed all particles in the final stacks from the eight datasets. As a result, we did not employ any sampling strategy in our study.
Data exclusions	No data were excluded.
Replication	The CryoSieve GitHub page offers a detailed tutorial and instructions. Additionally, for all eight datasets used in our study, the expected results have been deposited and are openly available on GitHub. One should be able to reproduce the results from the manuscript exactly using the openly accessible dataset and our open-source code.
Randomization	We used the star file in real experiments, which has no randomization. In simulated experiments using InSilicoTEM, there is no randomization. In numerical experiments of the CryoSieve score, there is randomization and we reported the mean and variance of the results.
Blinding	In this study, there was no group allocation; therefore, blinding was not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging