|  |  |
| --- | --- |
|  |  |
| **Author Checklist** | NCOMMS -21-38534A |
| **0000000000000000000000000000000000000000000000000000000000000** | **0000000000000000000000000000000000000** |
| **Please check the items below carefully and add a response in each row of the table to indicate the changes that you have made.** | |
|  |  |
| **Remaining reviewer comments** |  |
| Our guidance: | Your response: |
| Please address the remaining comments of referee #2:  -Development versus test set performance: The significant discrepancy between the performance of AF2 on development set (33.3% - 39.4%) and the test set (57.8% - 58,4%), is striking. This suggests that AF2 performance is dataset dependent. While this is not unexpected, it begs for a comment. It doesn’t seem to result from an organism bias, since the AF2 performance on the smaller development dataset is much lower even though it features a higher proportion of bacterial proteins, for which the authors observe a higher SR level in AF2 predictions.  Lines 95-97, 139-140  …’protocol performs quite close to (63% vs 72%) the recently developed AF-Multimer which was developed using the same data as the test set here, making a direct comparison difficult.  ‘was developed’ should be replaced by ‘was trained’.   Lines :132-134  The sentence is misleading since the performance of the 3 docking methods is clearly not good.  Suggestion:  Replace “The reason for GRAMM’s, TMdock’s and MDockPP’s good performance is likely due” by “The reason GRAMM’s, TMdock’s and MDockPP’s reach this performance level is likely due”  Lines 362-364  Projecting the fraction of human heterodimers predicted at the current 1% error rate, on the basis of the number of pairwise human PPI in the String DB (11.9 million) is overdoing it, since it is well known that a sizable fraction of the interactions in String are non-physical. The paragraph should be rephrased accordingly. | We thank you for this comment. We agree there is a big difference and have added a comment about this under limitations, where we suggest that performance should be assessed on as large non-redundant datasets as possible to ensure any selection bias does not impact the results. We do not know exactly the origin of the difference. We tried to examine the most obvious differences between the two sets (protein size, species, size of MSAs etc) but did not find anything obvious that separated the sets. Unfortunately, trying to pinpoint the origin of the difference is beyond the goals of this study.  We have changed this phrasing as suggested to clarify that it was indeed trained.  We thank you for this suggestion and have changed the phrasing as suggested.    We have changed this statement to reflect a more realistic modeling scenario, which we have also applied in practice (<https://www.biorxiv.org/content/10.1101/2021.11.08.467664v1>). |
|  |  |
| We would like to propose a revised abstract to comply with our formatting requirements and improve the accessibility of your work: **Please remove the sentence "The protocol can be found at:** [**https://gitlab.com/ElofssonLab/FoldDock**](https://gitlab.com/ElofssonLab/FoldDock)**." from the abstract.** Please edit the abstract in your manuscript files accordingly. Note that further minor changes may be made during the production process, and you will be able to check these in the proofs. | This sentence has been removed. |
|  |  |
| **Author information** |  |
| Our guidance: | Your response: |
| Please review your complete author list to verify that it is complete and accurate. We ask that you consult with your coauthors to ensure that all names, affiliations, and titles are represented correctly. Note that if any authors are added or removed after this point then all authors will be requested to provide approval documentation that could potentially delay the production of your paper. | We have checked the author list and it is correct. |
|  |  |
| **Article structure** |  |
| Our guidance: | Your response: |
| We can accommodate up to 10 display items (Figures or Tables) in the main article. Each Figure and Table must fit easily within an A4 page (210 x 297 mm). Please ensure that the number and size of your Figures and Tables fulfil these requirements to avoid any delay in the acceptance of your article. | The figures and tables all fit within one A4 page. |
| **Please ensure your main manuscript file includes the following sections, in this order:** |  |
| *Title Author list Affiliations Abstract Introduction Results Discussion (optional) Results and Discussion (optional) Methods (including Data Availability, Code Availability and Statistics subsections where relevant) References Acknowledgements Author Contributions Statement Competing Interests Statement Tables Figure Legends/Captions (for main text figures)* | *We have reordered the manuscript accordingly. Figures, Tables and Legends are now at the back of the manuscript, as well as uploaded as individual files.* |
| We do not edit Supplementary Information files; they will be uploaded with the published article as they are submitted with the final version of your manuscript. Any tracked changes should be removed from the file **and the file should be provided as a PDF file.**  Supplementary Figures do not need to be provided separately. | We have changed that to a single supplementary text. |
|  |  |
|  |  |
| **Main text** |  |
| Our guidance: | Your response: |
| Please do not use italics, bold font, underlining or speech marks unless required for technical terms (in both the main text and the display items). | We have changed that. (It is kept for latin names according to standards). |
| Please make sure that mathematical terms throughout your manuscript and Supplementary Information (including in figures, figure axes, and legends) conform strictly to the following guidelines. Equations must be supplied in editable format, and not as images. Scalar variables (e.g. x, V, χ) must be typeset in italic, whereas multi-letter variables and functions (e.g. log) must be formatted in roman. Vectors (such as the wavevector k or the magnetic field vector B) must be typeset in bold without italics. | We have changed the equations into word format so that they are editable. |
| **Please label equations sequentially as (1), (2), (3), etc.** | We have added these numbers |
| Please use italics for species names, both in the main text and the display items. | We use italics for speies names. |
|  |  |
| **Figures and Tables** |  |
| Our guidance: | Your response: |
| Please see the guidelines linked below for detailed instructions about how your figures should be prepared. Following these instructions will reduce the chances of delays should we need to request replacement artwork from you at a later stage. | We have done the best to follow the provided guidelines to reduce delays. |
| <https://www.nature.com/documents/NRJs-guide-to-preparing-final-artwork.pdf> |  |
| Wherever statistics have been derived (e.g. error bars, box plots, statistical significance) the legend needs to provide and define the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), using the wording “n=X biologically independent samples/animals/cells/independent experiments/n= X cells examined over Y independent experiments” etc. as applicable.  We strongly discourage deriving statistics from technical replicates, unless there is a clear scientific justification for why providing this information is important. Conflating technical and biological variability, e.g., by pooling technically replicates samples across independent experiments is strongly discouraged. | We have added the number of samples (n) for each comparison in all figures, including Fig 1 and tables where statistics are derived. |
| **Please revise  1. Please note that this information is missing in the legend of figure 1.** |  |
| Tables may not contain colour - please reformat your tables to black and white | The tables contained any color (more than for marking changes). We have deleted these changes, |
| If bold/italic formatting in tables is necessary, please define its meaning in a table footnote | It has been done for Table 1 |
| Any abbreviations, symbols or colours present in your figures must be defined in the associated legends. | We have checked this |
| All figures and tables must be cited in the main text, and numbered in the order in which they appear. | We have cited all figures in their order of appearance (supplementary Tables were wrong before). |
|  |  |
| **Data and Code** |  |
| Our guidance: | Your response: |
| Nature journals strongly support public availability of data and code. Please deposit the data and code used in your paper into a public data repository, or alternatively, present the data as Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor.   Please note that for some data types, deposition in a public repository is mandatory. Any restrictions on sharing of these data types must be clearly indicated in the statement and discussed with the editor. More information on our data deposition policies and available repositories can be found here: | We have shared our data and code in publically available repositories as well as provided a workbook containing the raw data for the results as a supplementary file. |
| [https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data](https://www.nature.com/nature-research/editorial-policies/reporting-standards#_blank) |  |
| All published manuscripts reporting original research in Nature Portfolio journals must include a data availability statement, as a separate section before the References and under the heading 'Data Availability'.   The data availability statement must make the conditions of access to the “minimum dataset” that are necessary to interpret, verify and extend the research in the article, transparent to readers.   This minimum dataset may be provided through deposition in public community/discipline-specific repositories, custom proprietary repositories or general repositories like Figshare, Zenodo and Dryad. Providing large datasets in supplementary information is strongly discouraged and the preferred approach is to make data available in repositories. Scientific Data, a Nature Portfolio journal, maintains a list of approved and recommended data repositories to support researchers seeking suitable repositories for their data (https://www.nature.com/sdata/policies/repositories).  The Data Availability Statement should also reference any source data published alongside the paper.  If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year).  For clinical datasets or third party data, please ensure that the statement adheres to our policy (https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data) | We provide a link to a figshare repository containing all data (large dataset) needed to reproduce the study. We also provide a workbook containing the results reported, ensuring that all figures can be reproduces from this supplementary file. |
| Please use the following template to provide all the information stated above:  The XX data generated in this study have been deposited in the YY database under accession code ZZ [add hyperlink here]. The XX data are available under restricted access for {insert reason}, access can be obtained by {explain how}. The raw XX data are protected and are not available due to data privacy laws. The processed XX data are available at YY. The XX data generated in this study are provided in the Supplementary Information/Source Data file. The XX data used in this study are available in the YY database under accession code ZZ [Add hyperlink here].  Or  The XX data used in this study are available in the YY database under accession code ZZ [Add hyperlink here]. All other data supporting the findings of this study are available within the article and its supplementary information files. Additional information and relevant data will be available from the corresponding author upon reasonable request. | We have stated the specific commit that the code used in this paper refers to, which makes it both citable and traceable: All code to run FoldDock and reproduce the analysis here can be obtained here https://gitlab.com/ElofssonLab/FoldDock (commit  2e4c96aa352338976260ece0646ceaaa75392dec) |
| Please find attached to our decision letter a marked up copy of your reporting summary. After making the requested changes, please be sure to include the final version of your reporting summary, without our comments, in your submission as a supplementary information file. The reporting summary will be published alongside the paper. Please note that this form is a dynamic 'smart pdf' and must therefore be downloaded and completed in Adobe Reader, instead of opening it in a web browser | We have included an unmarked copy as supplementary information |
| We notice that you have deposited your code in a Github repository, which we fully support. We strongly encourage you in addition to make your code citable by obtaining a DOI for the Github repository in order to provide a permanent reference to the version of the code used in this study and improve reproducibility. This can be done by linking the repository to Zenodo, following the instructions here: https://guides.github.com/activities/citable-code/ Please cite the Github repository in your manuscript text or Code Availability statement and in your reference list: authors, title (this paper), repository name, DOI identifier, year. Alternatively, you can deposit the code in Gigantum or Code Ocean for the same purpose. | ???? |
|  |  |
| **Methods** |  |
| Our guidance: | Your response: |
| Sufficient details of the experiments must be provided in the Methods section such that they could be reproduced without reference to published papers. Use of the term "as described previously" is not encouraged. |  |
| All accession codes must be accompanied with their hyperlinks throughout (for example, "5XRN [http://doi.org/10.2210/pdb5XRN/pdb]", "1483958 [https://doi.org/10.5517/ccdc.csd.cc1lt5m6]", "SRP109982 [https://www.ncbi.nlm.nih.gov/sra/?term=SRP109982]", "GSE101099 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101099]" or "NQLW00000000 [https://www.ncbi.nlm.nih.gov/assembly/GCA\_002312845.1/]"). |  |
|  |  |
| **References** |  |
| Our guidance: | Your response: |
| All references must be cited in numerical order. The reference list will be formatted according to the Nature style by our journal production team, however please ensure that references contain all of the information required, eg:  Kurumada, S., Takamori, S. & Yamashita, M. An alkyl-substituted aluminium anion with strong basicity and nucleophilicity. <i>Nat. Chem.</i> <b>12</b>, 36–39 (2020). |  |
|  |  |
| **End matter** |  |
| Our guidance: | Your response: |
| Nature Portfolio defines Competing Interest (CI) as financial and non-financial interests (including but not limited to funding, employment, stocks, shares, patents, personal or professional relationships with individuals or institutions, and unpaid membership advocacy) that could be perceived to directly undermine the objectivity, integrity, and value of a publication, or could be seen as having an influence on the judgments and actions of authors with regard to objective data presentation, analysis, and interpretation.  Please thoroughly review our policy on Competing Interests and include a detailed statement both in your final manuscript file and in our manuscript tracking system. Please ensure the statements are identical in both. Be specific about how each point stated relates to the research and list applicable author initials, and/or patent numbers.  If there are no competing interests, a negative statement must be included. |  |
| <https://www.nature.com/nature-research/editorial-policies/competing-interests> |  |
| Please confirm that all relevant funding awarded to each author is described in the Acknowledgements section. List each grant number, followed by the initials of the author who received it. |  |
|  |  |
| **Preparing your manuscript files** |  |
| Our guidance: | Your response: |
| Unless otherwise stated please limit individual file sizes to approximately 30MB. We strongly encourage the use of repositories for large datasets or source data due to size considerations. |  |
| Please supply a brief (maximum 250 characters, including spaces) summary of the main findings of the paper to be used on our website and in our e-alerts. The summary should be written in the third person in language suitable for a broad audience. It should follow the structure: "[[Brief background first sentence]]. Here, authors [[brief description of accomplishment]]." The summary may be edited by the editors prior to publication. Please provide this summary in your response here. |  |
| To ensure maximum visibility for your work, we may tweet about your paper following publication. If you would like us to include the Twitter handles of the first author(s), corresponding author(s), lab or institution in this tweet, please provide them in your response here. We would also welcome your suggestions for hashtags to use when tweeting about the work. |  |
| **Please provide main figures as individual vector files with editable text.** Acceptable file types for figures are .ai, .eps, .pdf or Chem Draw for fully editable vector-based art. For detailed guidance on figure preparation, see https://www.nature.com/documents/aj-artworkguidelines.pdf |  |
| **Please supply the main Supplementary Information file as a single PDF file.** |  |
| **All Supplementary Information files (e.g. Supplementary Data, Supplementary Software, etc.) must be cited in the main text. Please rename the Dataset xlsx file to Supplementary Data 1 and cite it in the main text or methods secion.** |  |
| Please supply legends for each Supplementary Data file in your response here (not in the Supplementary Information file). Please label each files as Supplementary Movie/Audio/Data 1, etc. |  |
| The use or adaptation of previously published images is strongly discouraged. If this is unavoidable, please request the necessary rights documentation to re-use such material from the relevant copyright holders and return this to us when you submit your revised manuscript. Please check whether your manuscript or Supplementary Information contain third-party images, such as figures from the literature, stock photos, clip art or commercial satellite and map data.  For more information on what constitutes ownership by a third party, please contact our Editorial Assistant at naturecommunications@nature.com |  |
|  |  |
|  |  |
| **Forms to complete** |  |
| Our guidance: | Your response: |
| **Editorial Policy Checklist** |  |
| Please update and upload a final version of the Editorial Policy Checklist with your revised manuscript files. A blank Editorial Policy Checklist can be found via the link below. Note that this form is a dynamic ‘smart pdf’ and must be downloaded and completed in Adobe Reader. |  |
| Please update your current checklist or download from: |  |
| <https://www.nature.com/documents/nr-editorial-policy-checklist.zip> |  |
|  |  |
| **You will need to upload:** |  |
| Editorial Policy Checklist |  |
| Completed Third Party Rights Table (if relevant) |  |
| A point-by-point response to the reviewers' comments |  |
| A completed copy of this checklist |  |
| The main article file in Microsoft Word format - please supply a version with tracked changes and a version with tracked changes accepted |  |
| Separate main Figure files |  |
| A Supplementary Information file |  |
| Supplementary Table, Video, Audio, Data, or Software files |  |
| Reporting Summary |  |
|  |  |