

## AlphaFold3 Transparency and Reproducibility

Stephanie A. Wankowicz, UCSF  
Pedro Beltrao, ETH  
Benjamin Cravatt, Scripps  
Roland Dunbrack, FCCC  
Anthony Gitter, UW Madison  
Kresten Lindorff-Larsen, Copenhagen  
Sergey Ovchinnikov, MIT  
Nicholas Polizzi, DFCI/HMS  
Brian K. Shoichet, UCSF  
James S. Fraser, UCSF

The publication of AlphaFold2 was a breakthrough moment for structural biology. Its impact has been far-ranging. Structure predictions for individual proteins opened new avenues for understanding biological systems and small molecule drug discovery. Large-scale prediction studies enabled evolutionary analyses and genetic variant interpretations. The open code was extended and modified for new methods and applications in protein design and protein-protein assembly prediction. These examples, among many, demonstrate how subsequent research and benchmarks have been made possible because the code and models were open and downloadable.

For these reasons, we were disappointed with the lack of code, or even executables accompanying the publication of AlphaFold3 in *Nature*. Although AlphaFold3 expands AlphaFold2's capacities to include small molecules, nucleic acids, and chemical modifications, it was released without the means to test and use the software in a high-throughput manner. This does not align with the principles of scientific progress, which rely on the ability of the community to evaluate, use, and build upon existing work. The high-profile publication advertises capabilities that remain locked behind the doors of the parent company.

In this publication, several deviations from our community's standards stand out. First, the absence of available code compromises peer review, a cornerstone of scientific publication and a standard typically upheld by journals. Indeed, one of us (RD) was a reviewer, and despite repeated requests, he was not given access to code during the review. Second, the model's limited availability on a hosted web server, capped at ten predictions per day, restricts the scientific community's capacity to verify the broad claims of the findings or apply the predictions on a large scale. Specifically, the inability to make predictions on novel organic molecules akin to chemical probes and drugs, one of the central claims of the paper, makes it impossible to test or use this method. Finally, the pseudocode released will require months of effort to turn into workable code that approximates the performance, wasting valuable time and resources. Even if such a reimplementations is attempted, restricted access raises questions about whether the results could be fully validated.

Computational costs of machine learning approaches are becoming prohibitive for academic institutions, owing to the high costs of training the models, leaving much computational research and potential breakthroughs in the hands of for-profit companies. While companies have the right to capitalize on their innovations, using the imprimatur of academic publications without the possibility of reproducing the results, far less building on them, subverts the enterprise. The amount of disclosure in the AlphaFold3 publication is appropriate for an announcement on a company website (which, indeed, the authors used to preview these developments), but it fails to meet the scientific community's standards of being usable, scalable, and transparent.

This moment can motivate our community to raise the bar of openness and transparency to accelerate scientific progress. When journals fail to enforce their written policies about making code available to reviewers<sup>1</sup> and alongside publications<sup>2</sup>, they demonstrate how these policies are applied inequitably and how editorial decisions do not align with the needs of the scientific community. While there is an ever-changing landscape of how science is performed and communicated, journals should uphold their role in the community by ensuring that science is reproducible upon dissemination, regardless of who the authors are.

AI approaches now directly impact biological discovery and human health. Fully realizing their potential will require not only technical breakthroughs but also open and collaborative efforts to build on others' findings, as is foundational in all scientific research.

1)<https://web.archive.org/web/20240511023627/https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>

2)<https://web.archive.org/web/20240511023855/https://www.nature.com/nature/for-authors/initial-submission>

**Form to endorse:** <https://forms.gle/d47w29eGHeL1ejXw9>

**Endorsers as of 13:00 PST 5/11/2024:**

Aashish Manglik, University of California, San Francisco  
Airy Sanjeev, Linköping universiteit  
Alan Fersht, University of Cambridge  
Aled Edwards, Structural Genomics Consortium  
Alexandra Salazar, Scripps Research  
Andrew Ward, Scripps Research  
Arne Elofsson, Stockholm University  
Axel Brunger, Stanford University  
Björn Wallner, Linköping University  
Brenda A. Schulman, Max Planck Institute of Biochemistry  
Brian Kuhlman, University of North Carolina at Chapel Hill  
Brian Pierce, University of Maryland  
Carolyn Bertozzi, Stanford University and HHMI  
Casey S. Greene, University of Colorado School of Medicine  
Cesar A. Ramirez-Sarmiento, Pontificia Universidad Catolica de Chile

Chaitan Khosla, Stanford University  
Christine Jacobs-Wagner, Stanford  
Claudio Mirabello, Linköping University  
Dror Noy, Migal-Galilee Research Institute  
Erik Lindahl, Stockholm University  
Firas Khatib, UMass Dartmouth  
Gabriel Lander, Scripps Research  
Gabriel Rocklin, Northwestern University  
Hao Wu, Boston Children's Hospital, Harvard Medical School  
Jan Kosinski, EMBL Hamburg  
Jean Fan, Johns Hopkins University  
Jeffrey J. Gray, Johns Hopkins University  
Jie Yang, Scripps Research  
Kevan M. Shokat, UCSF  
Kevin Drew, University of Illinois at Chicago  
Lance Stewart, Self  
Lei Stanley Qi, Stanford University  
Mariko Morimoto, Stanford University  
Markus J. Sommer, Johns Hopkins University  
Martin Steinegger, Seoul National University  
Mihai L. Azoitei, Duke  
Ora Schueler-Furman, The Hebrew University of Jerusalem, Israel  
Pratyush Tiwary, University of Maryland  
Sarel Fleishman, Weizmann Institute of Science  
Steven L. Salzberg, Bloomberg Distinguished Professor of Biomedical Engineering, Computer Science, and Biostatistics, Johns Hopkins University  
Stuart L. Schreiber, Arena BioWorks  
Tamir Gonen, UCLA  
Tanja Kortemme, UCSF  
Tobias Madl, Medical University of Graz, Austria  
Vikram Khipple Mulligan, Center for Computational Biology, Flatiron Institute  
Wade Harper, Department of Cell Biology, Harvard Medical School  
Yogesh Kalakoti, Linköping University  
Zhongyue John Yang, Department of Chemistry and Data Science Institute, Vanderbilt University