AlphaFold-Multimer struggles in predicting PROTAC-mediated protein-protein interfaces

Gilberto P. Pereira[a,b,#], Corentin Gouzien[c,#], Paulo C. T. Souza[a,b], Juliette Martin*[a]

[a] Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France.

[b] Centre Blaise Pascal de Simulation et de Modélisation Numérique, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France.

[c] Laboratoire d'Océanographie Microbienne (LOMIC, UMR 7621, CNRS-SU), Observatoire Océanologique de Banyuls, Avenue Pierre Fabre, F-66650 Banyuls-sur-Mer, France

[#] First authorship is shared between these authors

Abstract: AlphaFold2 (AF2) made its debut in the CASP14 competition, generating structures which could rival experimentally determined ones and causing a paradigm shift in the structural biology community. From then onwards, further developments enabled the prediction of multimeric protein structures while improving calculation efficiency, leading to the widespread usage of AF2. However, previous work noted that AF2 does not consider ligands and thus suggesting that ligand-mediated protein-protein interfaces (PPIs) are challenging to predict. In this letter, we explore this hypothesis by evaluating AF-Multimers' accuracy on four datasets, composed of: (i) 31 large PPIs, (ii) 31 small PPIs, (iii) 31 PPIs mediated by ligands and (iv) 28 PROTAC-mediated PPIs. Our results show that AF-Multimer is able to accurately predict the structure of the majority of the protein-protein complexes within the first three datasets (DockQ: 0.7-0.8) but fails to do so for the PROTAC-mediated set (DockQ < 0.2). One explanation is that AF-Multimers' underlying energy function was trained on naturally occurring complexes and PROTACs mediate interactions between proteins which do not naturally interact with each other. As these "artificial" interfaces fall outside AFs' applicability domain, their prediction is challenging for AF-Multimer.

Introduction:

AlphaFolds' (AF) debut in the CASP13 competition in 2018[1] revolutionized the structural biology field. The ability to accurately predict, solely from the amino acid sequence, the folded state of proteins meant that it is now possible, in principle, to explore the entire proteome of different organisms. As such, a natural next step was the high-throughput generation of predicted protein structures for the human proteome and other organisms[2], readily deposited on the PDB under the computed structure models option[3]. However, in 2018 AlphaFold had its own limitations, such as not being able to deal with multimeric structures, struggling with the accurate placement of amino acid side chains, and not being able to model post-translational modifications like phosphorylation[4]. Furthermore, AF structures typically do not include ligands, as they are not part of the input sequence provided. Research in the field evolved towards new algorithms that were built upon AlphaFold[5–10] and in the CASP14 AlphaFold2 was able to produce structures on par with experimentally determined ones [11]. With the development of AlphaFold-Multimer[12,13], the prediction of the structure for multimeric proteins was made accessible, and a new leap

forward in elucidating the structure of large macromolecular complexes occurred. Nonetheless, since AFs' algorithm relies on a greedy multi-sequence alignment strategy, the database lookup step becomes a computational bottleneck[14,15]. Other groups have thus steered towards a strategy anchored on using protein language models[14,16], which promises to speed-up calculation times and increase model generalizability while maintaining high accuracy.

Typically, the accuracy of these algorithms is assessed on a general benchmark test dataset, which is dominated by structures which are very similar to those present within the training set. However, some of the less represented structures may be of significant interest, such as protein-protein interactions which are mediated by PROteolysis TArgeting Chimeras (PROTACs)[14,16–20]. PROTACs are heterobifunctional ligands composed of two small molecules connected by a linker region. One of the small-molecules binds to a protein target and the other binds to a protein called E3 ligase, which is attached to the ubiquitination machinery[21,22]. This machinery is responsible for ubiquitination of proteins, which tags them for proteasomal degradation. PROTACs-based approaches have been explored for a variety of therapeutic targets, including proteins intimately connected to cancer[17,23,24].

We have reported in another work[25] that PROTAC-mediated interfaces are typically shallow and small and thus, represent a significant challenge to machine learning-based approaches such as AF2. In this letter, we evaluate the accuracy of AlphaFold-Multimer in predicting the structure of heterodimeric protein-protein complexes. Four cases were considered: complexes with ligands at the interface, complexes with small interfaces, complexes with large interfaces, and PROTAC-mediated complexes. By evaluating each case separately, it is possible to identify if AlphaFold-Multimer is performant across the different protein-protein interface types.

Methods:

Dataset selection: A dataset of hetero-dimers X-ray structures with a resolution better than 3 Å and a maximum sequence redundancy of 30% was retrieved from the Dockground resource (https://dockground.compbio.ku.edu/bound/index.php)[26]. Separate lists were built for proteins acquired before or after the AlphaFold training date (May 2018). The structures anterior to AlphaFold training were termed the 'training data set', and the structures posterior to AlphaFold training were termed the 'test data set'. The interface size of each complex was computed as the difference between the sum of surface areas of the monomers minus the surface area of the complex. This was done with and without hetero-atoms to assess the involvement of ligands at the interface. We thus separated three cases:

- ligand-mediated complexes. We first extracted complexes where the hetero-atoms accounts for more than 10% of the interface size. These complexes were manually verified to exclude complexes with cross-links, modified residues at the interface, or non-specific ligands at the interface (such as sulfate ions, glycerol or PEG groups). This resulted in a data set of 21 complexes in the training set and 10 complexes in the test set. We observe that these complexes have small interfaces.
- small interface complexes. For each case in the ligand-mediated list, we randomly picked a complex with (i) less than 10% of the interface size contributed by hetero-atoms, (ii) similar interface size (tolerance 150 $Å^2$) and (iii) similar length of the

shortest monomer (tolerance 10 amino-acids). This resulted in a data set of 21/10 complexes (training/testing), with matching complex and interface size compared to the ligand-mediated complexes, but without ligands at the interface.

- large interface complexes. For each case in the ligand-mediated list, we randomly picked a complex with (i) less than 10% of the interface size contributed by hetero-atoms, (ii) an interface size at least 1500 Å^2 greater than the interface size of the ligand-mediated complex, and (iii) similar length of the shortest monomer (tolerance 10 amino-acids). This resulted in a data set of 21/10 complexes (training/test), with matching complex size and larger interface size compared to the ligand-mediated complexes.

A fourth data set of 28 PROTAC-mediated complexes was extracted from the PDB. The sequence of each complex was obtained using Pymol. A list of the complexes used throughout this letter for the training and test sets is given in **Tables S1** and **S2**, respectively.

AF-Multimer: AlphaFold-Multimer calculations were carried out using ColabFold version 1.5.1[10,12,15] using the alphafold2_multimer_v3 model. For each complex, three iterations were carried out with three different random seeds and with model recycling. The best model was selected using the composite score (0.8 ipTM+0.2pTM) The best model was energy minimized using AMBER[27].

DockQ: To evaluate the similarity between the experimental structures and the predicted structures from AF-Multimer, the DockQ criteria were employed[28]. In short, DockQ provides a continuous score from 0 to 1 (with 1 being perfect similarity) which takes into account the fraction of conserved native contacts, RMSD of the target protein and the interface RMSD between a reference structure and a predicted structure[25]. This metric can be employed either quantitatively or qualitatively, with possible classifications being inaccurate (0-0.229), acceptable (0.23-0.49), medium (0.50-0.799) or high quality(0.80-1) predictions. For our analysis, we compared the distribution of the DockQ scores and all associated parameters obtained per dataset.

Results: We aimed at investigating whether AF-Multimer could correctly predict the Protein-Protein interface across datasets composed of different complex types. Thus, we carried out AF calculations on (i) a dataset composed of 31 large PPIs, (ii) a dataset composed of 31 small PPIs, (iii) a dataset composed of 31 PPIs mediated by ligands/small-molecules and (iv) a dataset composed of 28 PROTAC-mediated PPIs. For the datasets i-iii, 21 complexes were within the models training set and 10 were published after AF-Multimer was released. In total, the training set complexes correspond to 63 complexes (21x3). The test set calculations were carried out on the remaining 30 (10x3) complexes from datasets i-iii and on the PROTAC-mediated dataset. The results are shown in **Figure 1**.

As expected, we found that, for the training set, large PPIs were well predicted by AF-Multimer, with a median DockQ score of 0.75 and a median interface RMS below 2 angstrom (median value 1.27). Large PPIs are common across the PDB and thus we expected that AF-Multimer, having been trained on the whole PDB dataset, would be able to capture the features underlying such interactions. When the interface area decreases, we observe a significant worsening of the predictive ability of the algorithm, as noted by Yin and co-workers[29]. While the median DockQ score is quite high (DockQ = 0.79), we see that there is a spread across predictions and that, compared to the large PPI dataset, the small PPI

complexes are either well predicted or completely missed. This also leads to a broader distribution in terms of the interface RMS. In the ligand-mediated dataset, a high DockQ score, similar to the two previous datasets (median equal to 0.82), is observed. However, looking at the interface RMS it appears that the predictions for this dataset are also split between being highly accurate or wrong.

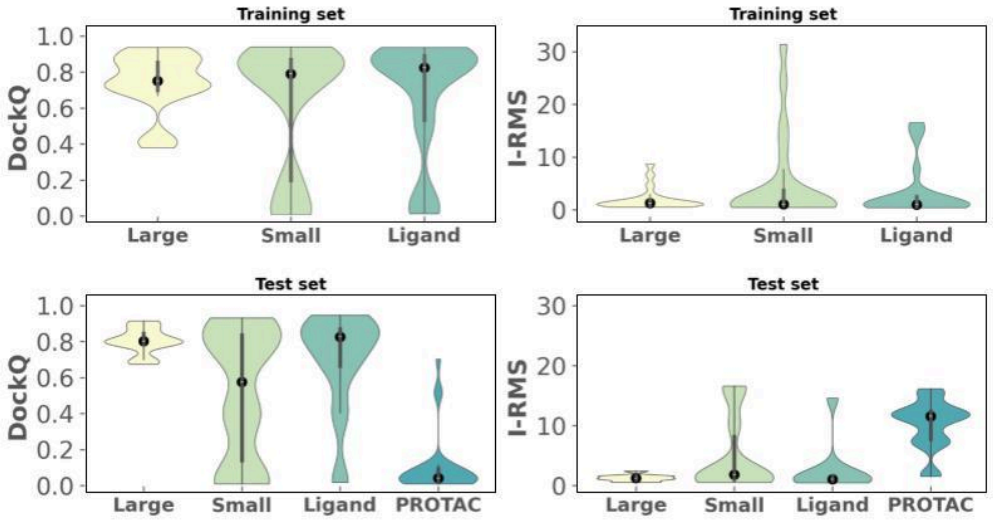

Figure 1 - Performance of AlphaFold-Multimer in predicting different protein-protein interfaces. Top left: DockQ score distributions for the four subsets within the training set. Top right: Interface RMS distribution. Bottom left: DockQ score distribution for the four subsets within the test set. Bottom right: Interface RMS distribution for the test set. Large corresponds to large protein-protein interfaces (PPI), small to small PPIs, Ligand corresponds to PPIs mediated by ligands/small-molecules and PROTAC corresponds to PPIs mediated by PROTACs. The median is represented using a black dot.

The same trend is observed for the test set, with large PPIs being consistently well predicted and prediction accuracy deteriorating as the interface gets smaller and ligand-mediated interfaces being either well predicted or wrong. Examples of well predicted and wrongly predicted complexes are shown, for each data set, in **Figure 2**. However, it is interesting to observe that for the PROTAC-mediated PPIs, AF-Multimer is, for the larger majority of complexes, unable to correctly predict the interaction interface between proteins. Except for three cases (7khh, 7lps, 8beb), the overall DockQ score is below 0.1. One possible explanation for this is that while ligand-mediated interfaces are well-represented within the PDB, PROTACs became the focus of attention only in the last few years. Thus, very few structures of PROTAC-mediated PPIs existed at the time of AF-Multimer development. Another explanation resides in the nature of PROTAC-mediated complexes, which are induced and stabilized by the ligand, and would naturally not probably interact. This explanation is in line with the recent work of Roney and Ovchinnikov[30] which provides evidence that AlphaFold2 indeed learned an energy function that encapsulates the physics governing the folded state. This suggests that PROTAC-mediated complexes are missed because they are not adequately represented by the energy function, because the ligand is ignored. Finally, it may also be the case that due to the dynamical nature of PROTAC-induced complexes, experimentally determined crystal structures represent only one structure within a larger conformational ensemble, as noted by Dixon and co-workers[31]. This third possibility is justified by the fact that PROTACS with different linker portion

compositions do, in fact, induce different protein-protein binding modes along a low-energy conformational landscape[31].

Nonetheless, with the growing popularity of PROTACs within the medicinal and computational chemistry fields, there has been a growth in the number of high quality ternary (ligase-PROTAC-target) structures. This new wealth of information could motivate the retraining of AF-Multimer, which should aim to include ligand contributions, or the development of a PROTAC-focused machine-learning model aiming at the accurate prediction of ternary complex structures. While some groups have developed tools for this purpose[32–36], they are typically not general because they require that the PROTAC molecule be known a priori. Within our group we developed PROTACability[25], which by-passes this constraint and achieves satisfactory accuracy but fails to produce high quality solutions. Thus, there is significant room for improvement and a general tool that uses minimal information and achieves high prediction quality would significantly impact the field of PROTAC-based drug discovery. The release of RoseTTAFold-All-Atom[37] in 2024 represents a leap forward in this regard, as it is apparently able to consider the effects of ligands within protein-protein interfaces. It remains to be seen whether this modeling suite achieves better performance than AlphaFold-Multimer in predicting highly plastic, “artificial” and shallow protein-protein interfaces such as those in PROTAC-mediated systems.

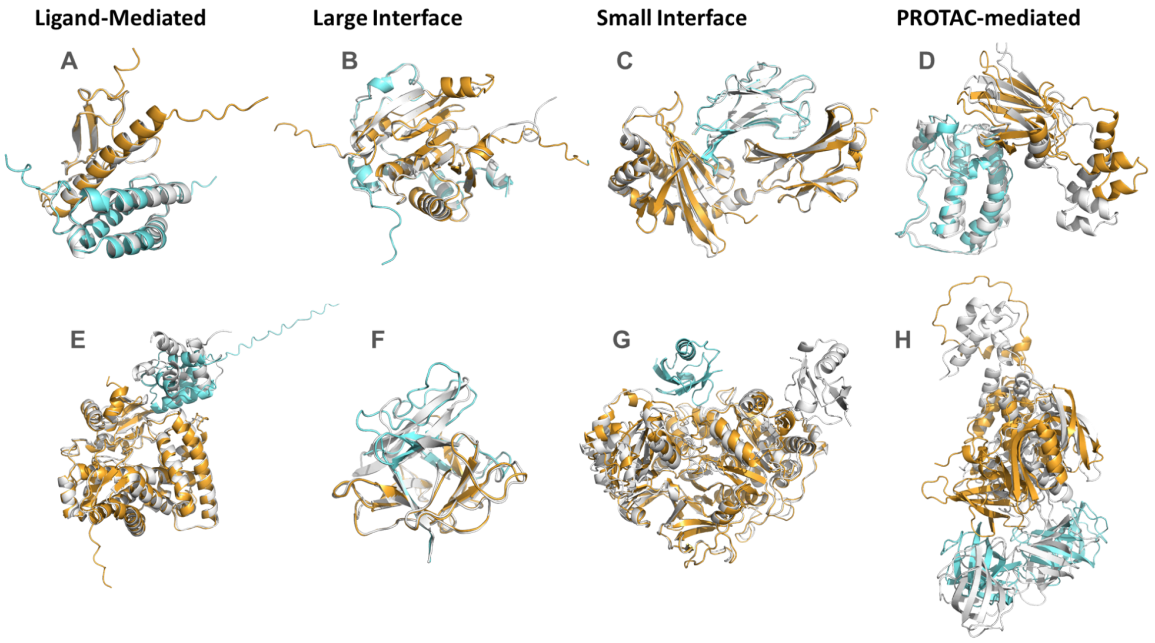

Figure 2 - Showcases from the datasets explored. A) 3N7R (DockQ = 0.94); B) 5NVK (DockQ = 0.94); C) 6C97 (DockQ = 0.81); D) 8BEB (DockQ = 0.51); E) 3EJD (DockQ = 0.14); F) 1R8O (DockQ = 0.38); G) 5D6J (DockQ = 0.01); H) 5HXB (DockQ = 0.02). The reference structure is displayed in light gray whereas the predicted structures are in color, with the receptor protein in orange and the ligand in cyan.

Support Information

Tables describing the training and test datasets used in the Letter; Figures showcasing correctly and wrongly predicted protein-protein complexes for each data set; Figures illustrating the interface size distribution for each dataset.

AUTHOR INFORMATION

Corresponding Author

Juliette Martin − Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Superieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France. ORCID: 0000-0002-4787-0885; E-mail: juliette.martin@ens-lyon.fr

AUTHOR CONTRIBUTIONS

The manuscript was written and revised by all authors. Approval for publication was obtained from all authors.

CONFLICT OF INTEREST

The authors declare no conflict of interest

ACKNOWLEDGMENTS

G.P, C.G, P.C.T.S and J.M would like to thank the support of the French National Center for Scientific Research (CNRS). G.P. and P.C.T.S would also like to thank the funding from research collaboration agreements with PharmCADD. We also acknowledge the support of the Centre Blaise Pascal's IT test platform at ENS de Lyon (Lyon, France) for the computer facilities. The platform operates the SIDUS solution developed by Emmanuel Quemener[38].

ABBREVIATIONS

AF - AlphaFold; PPI - Protein-Protein Interfaces; PROTAC - Proteolysis Targeting Chimera.

References

1. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
2. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
3. Burley, S. K. *et al.* RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **51**, D488–D508 (2022).

4. Perrakis, A. & Sixma, T. K. AI revolutions in biology: The joys and perils of AlphaFold. *EMBO Rep.* **22**, e54046 (2021).
5. Fang, Y. *et al.* DeepProSite: structure-aware protein binding site prediction using ESMFold and pretrained language model. *Bioinformatics* **39**, (2023).
6. Lannelongue, L. & Inouye, M. Environmental Impacts of Machine Learning Applications in Protein Science. *Cold Spring Harb. Perspect. Biol.* **15**, (2023).
7. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
8. Martin, J. When Alphafold2 predictions go wrong for protein–protein complexes, is there something to be learnt? *Q. Rev. Biophys.* **55**, e6 (2022).
9. Raisinghani, N. *et al.* AlphaFold2-Enabled Atomistic Modeling of Structure, Conformational Ensembles, and Binding Energetics of the SARS-CoV-2 Omicron BA.2.86 Spike Protein with ACE2 Host Receptor and Antibodies: Compensatory Functional Effects of Binding Hotspots in Modulating Mechanisms of Receptor Binding and Immune Escape. *J. Chem. Inf. Model.* (2024) doi:10.1021/acs.jcim.3c01857.
10. Brahma, R. & Raghuraman, H. Characterization of a novel MgtE homolog and its structural dynamics in membrane mimetics. *Biophys. J.* (2023) doi:10.1016/j.bpj.2023.11.3402.
11. Skolnick, J., Gao, M., Zhou, H. & Singh, S. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *J. Chem. Inf. Model.* **61**, 4827–4831 (2021).
12. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
13. Yin, R. & Pierce, B. G. Evaluation of AlphaFold antibody-antigen modeling with implications for improving predictive accuracy. *Protein Sci.* **33**, e4865 (2024).
14. Zhu, J., He, Z., Li, Z., Ke, G. & Zhang, L. Uni-Fold MuSSe: De Novo Protein Complex Prediction with Protein Language Models. *bioRxiv* 2023.02.14.528571 (2023) doi:10.1101/2023.02.14.528571.

15. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
16. Valentini, G. *et al.* The promises of large language models for protein design and modeling. *Front Bioinform* **3**, 1304099 (2023).
17. Pettersson & Crews. PROteolysis TArgeting Chimeras (PROTACs)—past, present and future. *Drug Discov. Today Technol.* **31**, 15-27 (2019).
18. Weng, G. *et al.* PROTAC-DB: an online database of PROTACs. *Nucleic Acids Res.* **49**, D1381–D1387 (2021).
19. He *et al.* PROTACs: great opportunities for academia and industry (an update from 2020 to 2021). *Signal transduction and Targeted Therapy*. **7**, 181 (2022)
20. Sakamoto, K. M. *et al.* Protacs: Chimeric molecules that target proteins to the Skp1–Cullin–F box complex for ubiquitination and degradation. *Proceedings of the National Academy of Sciences* **98**, 8554–8559 (2001).
21. Burslem, G. M. & Crews, C. M. Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **181**, 102–114 (2020).
22. Cyrus, K. *et al.* Impact of linker length on the activity of PROTACs. *Mol. Biosyst.* **7**, 359–364 (2011).
23. Crew, A. P. *et al.* Identification and Characterization of Von Hippel-Lindau-Recruiting Proteolysis Targeting Chimeras (PROTACs) of TANK-Binding Kinase 1. *J. Med. Chem.* **61**, 583–598 (2018).
24. Jaime-Figueroa, S., Buhimschi, A. D., Toure, M., Hines, J. & Crews, C. M. Design, synthesis and biological evaluation of Proteolysis Targeting Chimeras (PROTACs) as a BTK degraders with improved pharmacokinetic properties. *Bioorg. Med. Chem. Lett.* **30**, 126877 (2020).
25. Pereira, G. P. *et al.* Rational Prediction of PROTAC-Compatible Protein-Protein Interfaces by Molecular Docking. *J. Chem. Inf. Model.* **63**, 6823–6833 (2023).
26. Collins, K. W. *et al.* Dockground resource for protein recognition studies. *Protein Sci.* **31**, e4481 (2022).

27. Case, D. A. *et al. Amber 2021*. (University of California, San Francisco, 2021).
28. Basu, S. & Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS One* **11**, e0161879 (2016).
29. Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **31**, e4379 (2022).
30. Roney, J. P. & Ovchinnikov, S. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. *Phys. Rev. Lett.* **129**, 238101 (2022).
31. Dixon, T. *et al.* Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry. *Nat. Commun.* **13**, 5884 (2022).
32. Weng, G., Li, D., Kang, Y. & Hou, T. Integrative Modeling of PROTAC-Mediated Ternary Complexes. *J. Med. Chem.* **64**, 16271–16281 (2021).
33. Almodóvar-Rivera, C. M. *et al.* A Modular Chemistry Platform for the Development of a Cereblon E3 Ligase-Based Partial PROTAC Library. *Chembiochem* **24**, e202300482 (2023).
34. Zheng, S. *et al.* Accelerated rational PROTAC design via deep learning and molecular simulations. *Nat. Mach. Intell.* **4**, 739–748 (2022).
35. Duran-Frigola, M., Cigler, M. & Winter, G. E. Advancing Targeted Protein Degradation via Multiomics Profiling and Artificial Intelligence. *J. Am. Chem. Soc.* **145**, 2711–2732 (2023).
36. Li, F. *et al.* DeepPROTACs is a deep learning-based targeted degradation predictor for PROTACs. *Nat. Commun.* **13**, 7133 (2022).
37. Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* eadl2528 (2024).
38. Quemener, E. & Corvellec, M. SIDUS—the solution for extreme deduplication of an operating system. *Linux J.* (2013) doi:10.5555/2555789.2555792.