Can AlphaFold2 predict the impact of missense mutations on structure?

To the Editor — Understanding the impact that missense mutations have on protein structure helps to reveal their biological effects. Although the structural prediction algorithm of AlphaFold2 is able to predict wild-type (WT) structures to high accuracy, it seems to fall short in predicting the impact of missense mutations on the three-dimensional (3D) structures of proteins.

In a landmark achievement, the half-century goal of predicting protein structure from an amino-acid sequence was accomplished by AlphaFold2 (ref. [1]), a deep-learning algorithm that relies on homology and experimental structures in the Protein Data Bank (PDB). The program and its predictions for 98.5% of the human proteome are publicly available[2]. However, one potential limitation is the insensitivity of AlphaFold2 to structure-disrupting mutations in an input sequence. This is because a database of structure-disrupting mutations does not exist, and AlphaFold2 therefore primarily bases its predictions on WT or homologous sequences instead. This drawback is important because missense mutations frequently associate with human diseases and single amino-acid mutations can lead to protein aggregation, misfolding and dysfunction. Being able to predict the effect of mutations of interest on the 3D structure of proteins will help structural biologists and non-structural biologists alike make informed hypotheses about their mechanisms of pathogenicity. Here, we comment on the implications of this AlphaFold2 limitation using three illustrative examples of selected domains for which experimental data for both WT and structure-disrupting mutations are available. The examples include ubiquitin-associated domains (UBAs) of a human Rad23 protein (hHR23a), BRCT (breast cancer 1 (BRCA1) C-terminal) repeats of BRCA1, and the actin motor protein Myosin VI MyUb domain. We provide comparisons of the AlphaFold2-predicted structural models of known mutants of these domains to their WT counterparts.

The two UBAs of the proteasome shuttle factor and DNA-repair protein hHR23a contain a central leucine (L198 in UBA1 and L355 in UBA2) that is buried in the core[3] and required for structural integrity[4]. Alanine substitution causes the UBAs to become intrinsically disordered;[4] however, AlphaFold2 predicted alanine-substituted UBA1 (Fig. 1a) or UBA2 (not shown) to be structurally equivalent to WT UBA with only minor differences in the fold. We used the average root mean square deviation

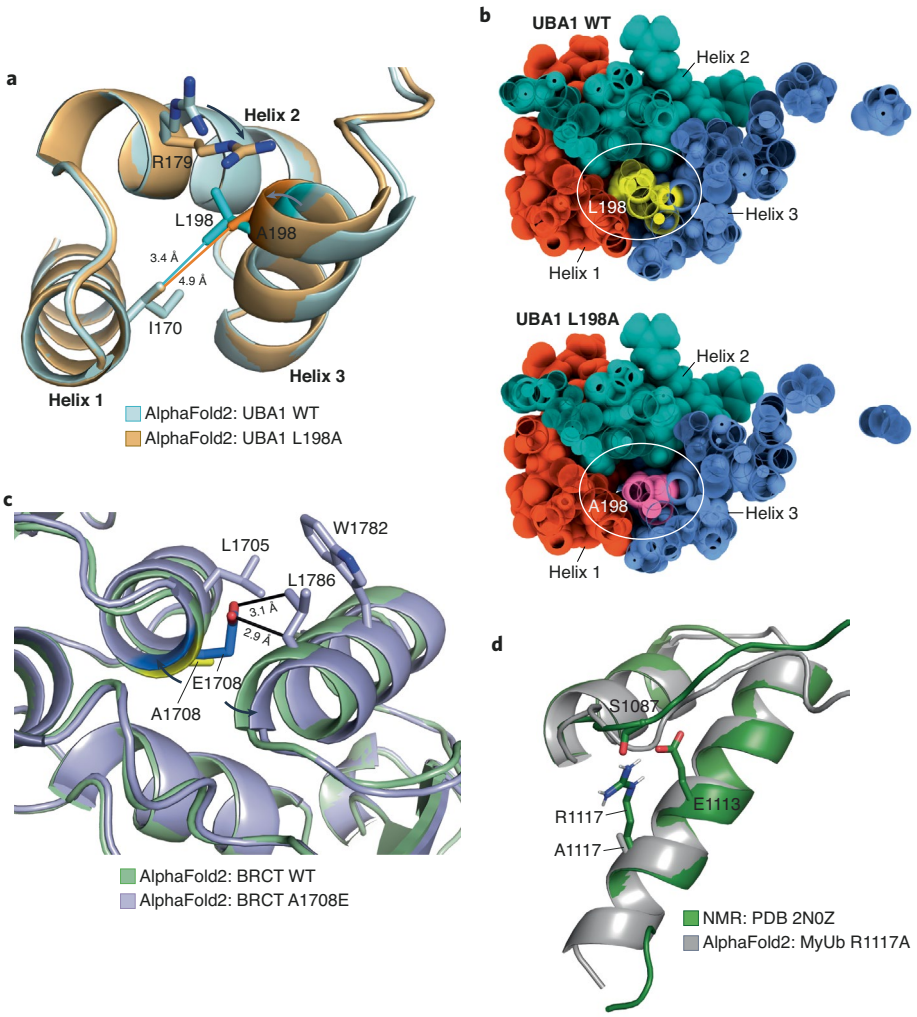

Fig. 1 | Interpretation of structure-disrupting mutations by AlphaFold2. a, Overlayed top-scoring predicted structures for WT hHR23a UBA1 (light blue) and L198A (orange). Sidechain-heavy atoms are displayed for position 198 and the surrounding residues, with nitrogen in blue. **b**, Cut-away views of WT and L198A UBA1 with atoms displayed as van der Waals spheres. Helix 1, helix 2 and helix 3 are show in orange, teal and blue, with L198 and A198 in yellow and pink, respectively. White ovals highlight the region surrounding position 198, with empty space shown by the black shading between spheres. **c**, Overlay of AlphaFold2-predicted WT (green with A1708 highlighted in yellow) and A1708E (purple with E1708 highlighted in blue) BRCT. Distances between E1708 and L1786 and sidechain-heavy atoms for additional surrounding hydrophobic residues are shown. Arrows indicate shifting of the adjacent helix away from the E1708-containing helix relative to the WT sequence. **d**, Overlay of the AlphaFold2-predicted top-scoring structure for MyUb with alanine substitution at R1117 in grey (residue numbers based on isoform from the National Center for Biotechnology Information: Q9UM54.4) and the experimental MyUb structure as solved by nuclear magnetic resonance (NMR) spectroscopy in green. Structures were analyzed and figures were generated using PyMol and ChimeraX.

(r.m.s.d.), a measure of the distance between atoms of superimposed structures, for the backbone α-carbon (Cα) atoms to quantify the similarity of the predicted WT and mutant structures. We found the average Cα r.m.s.d. of the alanine-substituted structures and the WT UBA domains to be 0.1 Å for UBA1 and 0.3 Å for UBA2; for comparison, these differences are smaller than the 0.7 Å Cα r.m.s.d. of the two WT hHR23a UBA domains (not shown). Alanine substitution yielded a less compact hydrophobic core yet the mean scores of the AlphaFold2 local distance difference test (pLDDT), a statistical test designed to indicate confidence in the predicted structure, were equivalent. For example, in UBA1, despite the arginine at position 179 (R179) moving towards the location occupied by L198 in the WT sequence and the third helix (Helix 3) shifting slightly inward (indicated with arrows in Fig. 1a), the intramolecular distance to the Cγ methyl group of the isoleucine at position 170 (I170) increases from 3.4 Å for L198 to 4.9 Å for the substituted alanine (Fig. 1a). This reduced packing is apparent when the structures are displayed with van der Waals surfaces and cut away to make the core visible (Fig. 1b). However, the mean pLDDT scores for the top-ranking structures in both cases was 84 (scores range from 0 to 100, with 100 representing highest confidence), indicating insensitivity to the effects caused by substituting the longer leucine sidechain with alanine.

Mutations in the BRCT repeats of BRCA1 disrupt its tumor suppressor activity and result in early onset breast cancer, with the missense mutation A1708E linked to breast cancer[5] and found to destabilize the BRCA1 C-terminal repeats in proteolytic degradation assays[6,7]. A1708 packs into a small hydrophobic pocket between the two BRCT repeats and replacement with the bulkier and charged glutamic acid residue is expected to destabilize their interaction. AlphaFold2 predicted similar structures for WT and A1708E BRCT, with an average Cα r.m.s.d. of only 0.6 Å (Fig. 1c). For A1708E BRCT, there is slightly more space between the helices of the two repeats (indicated by outward arrows in Fig. 1c), with the distance between the α-carbons of residues 1708 and 1786 at 5.4 Å for WT and 6.6 Å for the A1708E mutant. This increase in distance to accommodate the longer glutamic-acid sidechain is not sufficient to prevent the acidic oxygens of E1708 from packing against the hydrophobic L1786 side chain (Fig. 1c). This adverse placement of charged atoms within a hydrophobic region of the structure is not reflected in the pLDDT scores as these values are 95 and 94 for WT and A1708E BRCT, respectively.

Myosin VI contains a helix-turn-helix-like structural domain named MyUb that includes a ^{1116}RRL^{1118} motif, with alanine substitution of R1117 reported to abolish interactions with adaptor proteins and cause loss of structural integrity[8]. R1117 forms hydrogen bonds across the MyUb structure, bridging the two helices (Fig. 1d, green; PDB:)[9]. AlphaFold2 predicted no loss of structural integrity by R1117A substitution (Fig. 1d, grey). Moreover, the mean pLDDT score for the top-ranking R1117A structure was 90, which is higher than that of the top-ranking WT structure — the score of which was 89.

In conclusion, although AlphaFold2 marks a pivotal advancement in protein structural prediction, we have found it to be largely unable to predict when a point mutation causes defective protein folding. Indeed, a recent survey found no correlation of pLDDT scores with protein misfolding induced by amino-acid substitutions in green fluorescent protein[10]. This lack of correlation agrees with our case studies, illustrating the inability of AlphaFold2 to predict the effects of point mutations on protein structure. This limitation probably arises because it predicts structures based on those available in the PDB, rather than by fundamental driving forces of protein folding. It is possible that the merger of AlphaFold2 with modern and anticipated advancements in molecular dynamics simulations of protein structure could overcome this limitation. Alternatively, establishing a database for storing structure-disrupting mutations may allow for future renditions of AlphaFold2 or other artificial-intelligence programming to include this information in their protein-folding predictions.

Overall, AlphaFold2 has provided a huge advance to the field of structural biology and to the biology community as a whole. While it currently is unable to predict structural effects of missense mutations, it is conceivable that incorporation of such experimental data will enable this feature in future versions of protein structure prediction programs.

Reporting Summary

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. ❐

Gwen R. Buel and Kylie J. Walters ✉
Center for Structural Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA.
✉*e-mail: kylie.walters@nih.gov*

Published online: 19 January 2022
https://doi.org/10.1038/s41594-021-00714-2

References

1. Jumper, J. et al. *Nature* **596**, 583–589 (2021).
2. Tunyasuvunakool, K. et al. *Nature* **596**, 590–596 (2021).
3. Walters, K. J. et al. *Proc. Natl Acad. Sci. USA* **100**, 12694–12699 (2003).
4. Wang, Q. et al. *Biochemistry* **42**, 13529–13535 (2003).
5. Easton, D. F. et al. *Am. J. Hum. Genet.* **81**, 873–883 (2007).
6. Williams, R. S., Green, R. & Glover, J. N. *Nat. Struct. Biol.* **8**, 838–842 (2001).
7. Lee, M. S. et al. *Cancer Res.* **70**, 4880–4890 (2010).
8. He, F. et al. *Cell Rep.* **14**, 2683–2694 (2016).
9. Biancospino, M. et al. *Nat. Commun.* **10**, 4974 (2019).
10. Pak, M. A. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2021.09.19.460937 (2021).

Acknowledgements

We are grateful to H. Matsuo for critical reading of the manuscript. This work was supported by the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health (NIH) (1ZIABC011627 to K.J.W.), and used computational resources of the High-Performance Computing Biowulf cluster of the NIH (http://hpc.nih.gov).

Competing interests

The authors declare no competing interests.