Quarterly Reviews of Biophysics

cambridge.org/qrb

When Alphafold2 predictions go wrong for protein–protein complexes, is there something to be learnt?

Juliette Martin

Univ Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5086 MMSB, F-69367 Lyon, France

Abstract

In this short communication, I analyze cases of failed predictions for protein–protein complexes with Alphafold2, and show that they either point to erroneous annotation in the PDB or correct binding site regions.

Observation

Cite this article: Martin J (2022). When Alphafold2 predictions go wrong for protein–protein complexes, is there something to be learnt? *Quarterly Reviews of Biophysics* **55**, e6, 1–4. https://doi.org/10.1017/S0033583522000051

Received: 28 April 2022
Revised: 30 May 2022
Accepted: 31 May 2022

Author for correspondence:
Juliette Martin,
E-mail: juliettemart1@gmail.com

The Alphafold2 method (Jumper *et al.*, 2021*a*) has unquestionably revolutionized the field of protein structure prediction, achieving very high accuracy for most targets during the CASP14 initiative (Jumper *et al.*, 2021*b*). Recently, the Alphafold2 strategy has been extended to predict protein–protein complexes (Evans *et al.*, 2021). In the article presenting the multimer version of Alphafold2, a set of 17 complexes obtained after the network training date was considered to compare with previous strategies based on the initial Alphafold2 system (Ghani *et al.*, 2021). Out of these 17 complexes, Alphafold2 multimer achieved correct predictions for 14 cases, as assessed by the DockQ score (Basu and Wallner, 2016).

I looked at the three cases where Alphafold2 produced models with low DockQ scores (Evans *et al.*, 2021). In the first version of the study (Evans *et al.*, 2021), the three failed cases were 5ZNG (DockQ = 0.02), 6A6I (DockQ = 0.05), and 7NLJ (DockQ = 0.06).[1] I reproduced Alphafold2 predictions with a local installation of the ParaFold pipeline (Zhong *et al.*, 2022), installed in November 2021, excluding the templates newer than April 2018 (--max_template_date = 2018-04-30, which is the network training date). Each prediction run generates five models and I considered the model with the highest confidence value as the final prediction. Overall, I am able to replicate published results, as shown in Table S1.

For 5ZNG, the predicted models are indeed very distant from the complex annotated as biological assembly in the PDB (DockQ score = 0.02). However, the five models are almost identical, and have good confidence value (around 0.7). They are very similar to the complex annotated as asymmetric unit (DockQ score = 0.7), see Fig. 1. The complex reported in the article accompanying the 5ZNG structure is indeed the one annotated as asymmetric unit in the PDB entry (Guo *et al.*, 2018), not the one annotated as biological assembly and generated by PISA (Krissinel and Henrick, 2007). Thus, also in this case, the Alphafold2 prediction was indeed accurate: the predicted model was indeed the one described as biologically relevant.

For 6A6I and 7NLJ, I performed triplicate runs of Alphafold2 and I obtained models with highest DockQ scores than previously reported (Evans *et al.*, 2021): DockQ = 0.39/0.24/0.18 for 6A6I and DockQ = 0.21/0.16/0.16 for 7NLJ, with low confidence values (0.2–0.3). In these cases, the asymmetric units and biological assemblies of the PDB entries are identical, and I found no obvious reason for such discrepancy.

It is worth noting that, even if these models are far from the high quality threshold (DockQ > 0.8), they indeed provide an approximate prediction of the true binding site regions, as shown in Fig. 2.

In conclusion, even in the cases where Alphafold2 did not achieve correct predictions for multimers, the examination of failed cases in this very small data set suggests that the predictions could detect errors in PDB annotation (like 5ZNG) or, more interestingly, determine approximate binding sites (like 6A6I and 7NLJ). In the last case, the models could provide a good starting point for conventional docking tools with restraints to these binding sites. In addition, since Alphafold2 achieves the very difficult task of predicting both the subunit folds and their binding mode, one could wonder what accuracy it could attain in a classical docking context when the monomer structures are known. This indicates that Alphafold2 could revolutionize the field of protein–protein docking as it has done for protein structure prediction.

[1]A second version of the article was posted on the 10th of March, with improved prediction for 5ZNG (DockQ = 0.69) and worst prediction for 7P8K (DockQ = 0.05), thanks to new networks trained with modified loss measures. I do not discuss these results here because the local installation I used is anterior to these new networks.

© The Author(s), 2022. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

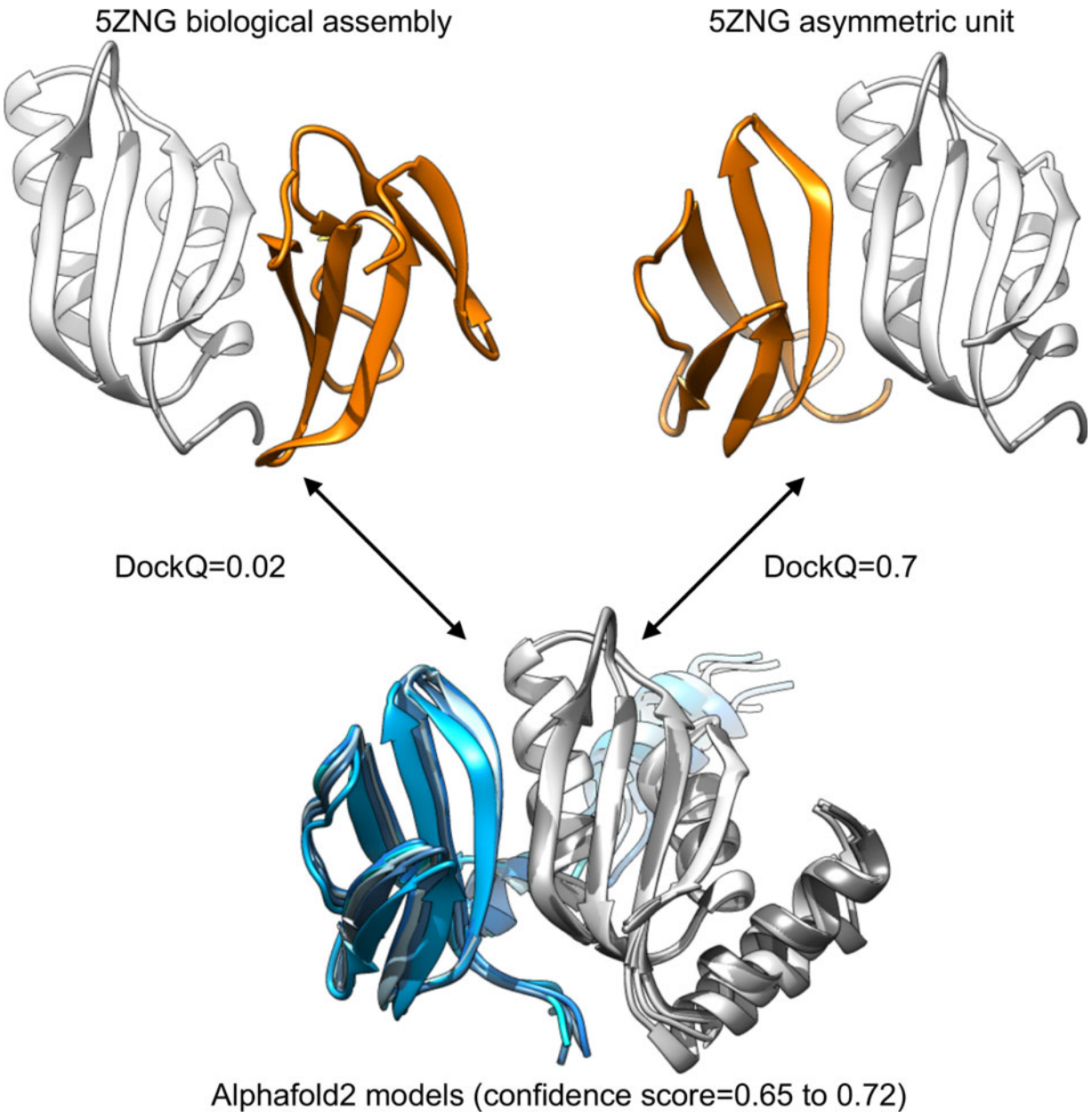


Fig. 1. Comparison between the structural information available for 5ZNG and the Alphafold2 models. The biological assembly discussed in the article introducing the structure is the one annotated as asymmetric unit in the PDB (top right).

https://doi.org/10.1017/S0033583522000051 Published online by Cambridge University Press

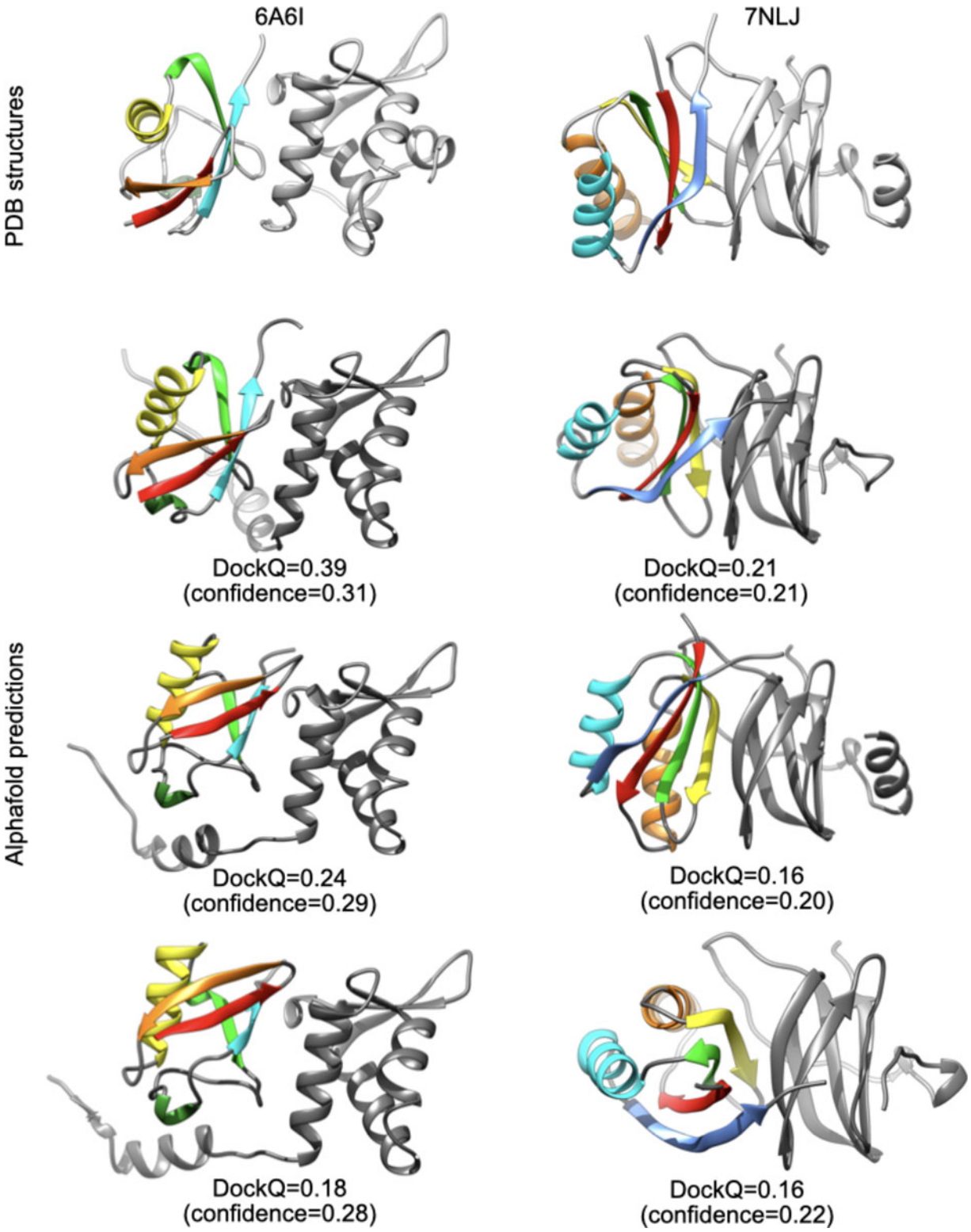


Fig. 2. Comparison between PDB structures and Alphafold2 predictions for 6A6I and 7NLJ.

Acknowledgements. I gratefully acknowledge support from Alexis Michon from IBCP and the CNRS/IN2P3 Computing Center (Lyon – France) for providing computing and data-processing resources needed for this work. I gratefully acknowledge Elisa Frezza, Guillaume Launay, and Riccardo Pellarin for their constructive comments.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

https://doi.org/10.1017/S0033583522000051 Published online by Cambridge University Press

References

Basu S and Wallner B (2016) DockQ: a quality measure for protein–protein docking models. *PLoS ONE* **11**, e0161879.

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J and Hassabis D (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034.

Ghani U, Desta I, Jindal A, Khan O, Jones G, Kotelnikov S, Padhorny D, Vajda S and Kozakov, D (2021) Improved docking of protein models by a combination of Alphafold2 and ClusPro. *bioRxiv* 2021.09.07.459290.

Guo L, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, Meusnier I, Bonnot F, Padilla A, Peng Y-L, Liu J and Kroj T (2018) Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *PNAS* **115**, 11637–11642.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021*a*) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021*b*) Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics* **89**, 1711–1721.

Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774–797.

Zhong B, Su X, Wen M, Zuo S, Hong L and Lin J (2022) ParaFold: paralleling AlphaFold for large-scale predictions. In *International Conference on High Performance Computing in Asia-Pacific Region Workshops*. New York: Association for Computing Machinery, pp. 1–9.

https://doi.org/10.1017/S0033583522000051 Published online by Cambridge University Press