Using AlphaFold to understand multi-domain proteins from SARS-CoV-2

Maximilian Edich[1], Lea von Soosten[1], Andrea Thorn[1]

[1]Institut für Nanostruktur und Festkörperphysik, Universität Hamburg, Luruper Chaussee 149 (Bldg. 610 - HARBOR), Germany, medich@physnet.uni-hamburg.de

The protein structure prediction tool AlphaFold [1] amazed structural biologists across the world in 2021 with its simple user interface and astonishingly high accuracy, which opend up never before seen possibilities in the analysis of otherwise difficult to solve multi-domain proteins. As part of our research in the Coronavirus Structural Task Force [2], we utilized AlpaFold for unknown domains and regions of partly known drug targets in COVID-19 research.

Here, we demonstrate the use of AlphaFold on proteins from SARS-CoV-2. While the solved structures only cover a part of the viral genome – only 17 out of 28 viral proteins were partly solved by experimental methods – we have plenty of annotated protein-coding genes, the potential structure of which can now be predicted using neural networks. We will discuss both the benefits as well as the potential shortcomings of these predictions and show how they can be combined with information from other bioinformatical tools and experimentally determined structures to get new insights into the structural biology of the virus.

[1] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

[2] Croll, Tristan I. *et al.* "Making the invisible enemy visible." *Nature Structural & Molecular Biology* **28.5** (2021): 404-408.

This work was supported by the German Federal Ministry of Education and Research [grant no. 05K19WWA] and the Deutsche Forschungsgemeinschaft [grant no. TH2135/2-1].