Exploiting proteome-scale predictions of AlphaFold2 to reveal foldable segments with unexplored characteristics

ISABELLE CALLEBAUT (IMPMC-CNRS-UMR7590, Paris Cedex 05)
Appoline BRULEY (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Geoffroy GASCHIGNARD (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Maxime MILLET (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Tristan BITARD-FEILDEL (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Karim BENZERARA (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Manuela DEZI (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Feriel SKOURI-PANET (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Jean-Paul MORNON (IMPMC-CNRS-UMR7590, Cedex 05, Paris)
Elodie DUPRAT (IMPMC-CNRS-UMR7590, Cedex 05, Paris)

Abstract

AlphaFold2 (AF2) has created a breakthrough in biology by providing three-dimensional (3D) structure models for whole-proteome sequences, with unprecedented levels of accuracy. Here, we wish to illustrate how these AF2 proteome-scale data can be exploited to search for new domains or new families of domains with as yet undescribed structures and functions, which are prime targets for experimental characterization.

Firstly, high-confidence 3D structure AF2 predictions can be used in combination with sequence similarity searches and clustering approaches to identify new domain families. We will illustrate this with the example of the CoBaHMA domain family, part of the very large Heavy-Metal Associated (HMA) superfamily, yet possessing specific sequence and structure signatures and possibly involved in interaction with anionic lipids.

Secondly, the study of 21 reference proteomes included in the AlphaFold DataBase (AFDB v1) has shown that the low-confidence AF2 predictions traditionally associated with disordered sequences contain a non-negligible proportion of long segments predicted as “foldable” by pyHCA, a tool developed by us. These segments left out by AF2 include cases of conditional order, as well as cases that could form well-folded structures but escape the AF2 prediction due to a shallow multiple sequence alignment and/or undocumented structure or fold. AF2 and pyHCA can therefore be advantageously combined to unravel cryptic structural features in whole proteomes and to refine predictions for different flavors of disorder.

Bruley et al. Digging into the 3D Structure Predictions of AlphaFold2 with Low Confidence: Disorder and Beyond Biomolecules. 2022 doi :10.3390/biom12101467

Bruley, Bitard-Feildel et al. A sequence-based foldability score combined with AlphaFold2 predictions to disentangle the protein order/disorder continuum Proteins 2023 doi10.1002/prot.26441 ;

Gaschignard, Millet et al. AlphaFold2-guided description of CoBaHMA, a novel family of bacterial domains within the heavy metal-associated superfamily Proteins 2024 doi :10.1002/prot.26668