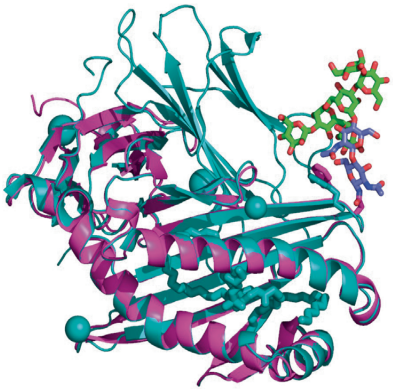
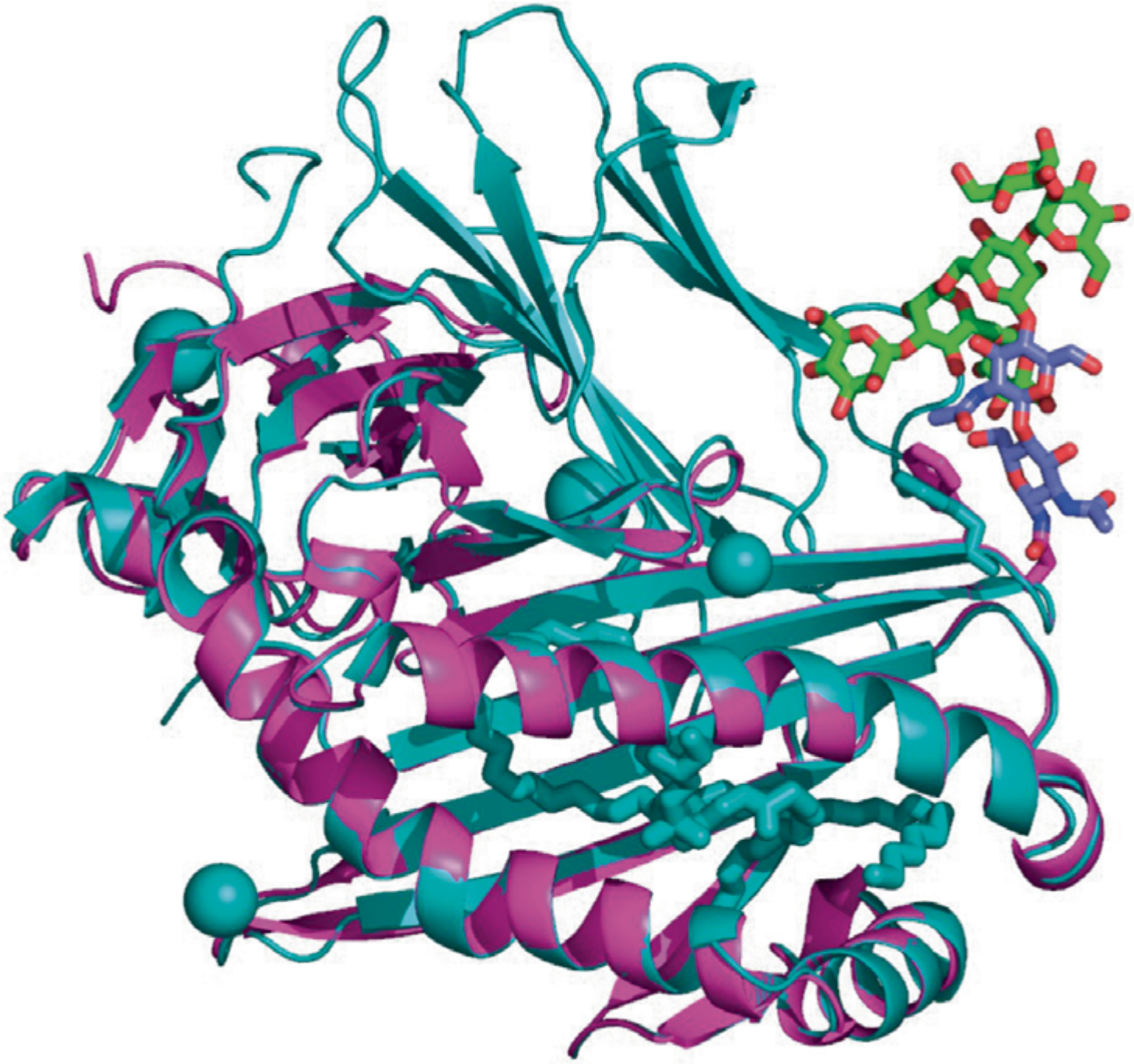


The case for post-predictional modifications in the AlphaFold Protein Structure Database

Description

The AlphaFold Protein Structure Database release by DeepMind and EMBL-EBI marks a significant breakthrough in structural biology. It makes highly accurate structural predictions available to the scientific community worldwide for 20,000 proteins from humans and proteins from 20 other biologically relevant organisms that include *Escherichia coli*. Like many scientists that work on macromolecular structures, the authors are genuinely excited about this development. The authors feel that there is a non-negligible potential for misinterpretation of its content in its current form.
Glycosylation is among the most relevant co- and post-translational modifications in protein. Indeed, between 50% and 70% of those 20,000 predicted human proteins are believed to be glycosylated. None of this is yet visibly highlighted on the database. The authors analyze the difficulties of considering the protein glycosylation arising from the compositional problem of what glycoform is linked to each sequon.



They suggest ways to address such issues by harnessing the rich information available in glycomics databases to complete and enrich the predicted protein models with reasonable modifications.

Category

1. News