MS11 Opportunities from combining structural biology and fold prediction

MS11-05
AlphaFold-2 revolution for crystallographic software
E. Krissinel [1], R. Keegan [1], C. Ballard [1], A. Lebedev [1], V. Uski [1]
[1]Science and Technology Facilities Council UK - Didcot (United Kingdom)

Abstract
Understanding protein function through its structure is one of main motivations behind the structural biology research, which is closely related to the question of the relationship between sequence and structure. Over the past 40 years, significant progress has been made in this direction, both experimentally and in terms of computational methods for protein homology analysis. The Smith and Waterman algorithm, proposed in 1981 [1], laid the foundation for homology studies. By 2002, algorithms for secondary structure prediction have reached ~80% accuracy, giving more structural insight in sequence-based analysis [2]. Profile matching techniques (see, e.g., [3,4]) boosted the sensitivity of sequence matching so that structural homology could be detected deep in the twilight zone. Computation of protein folds from sequence was a very challenging task for a long time. It was largely solved by AlphaFold-2 and RoseTTAFold by 2021 [5,6].
Each of the above achievements has had a significant impact on structural biology research and computational methods for structure determination. In this context, the emergence of AlphaFold-2, which produces structures that are incredibly accurate by computational standards, is perceived as a revolution, since it solves the very problem of the sequence-structure relationship. While the full scope of this event is yet to be realised, we would like to discuss its impact from a narrower perspective of a large project in crystallographic software, the Collaborative Computational Project Number 4 (CCP4). The presentation will deliver on CCP4's response to the emergence of AlphaFold-2 as a powerful instrument for homology modelling; overview the new tools included in CCP4 Software Suite, its online services and CCP4 Cloud; discuss changes in approaches to computational structure determination; and outline possible directions for further development. What will be the role of methods developed in pre-AlphaFold era, and do we expect a leap toward full automation of structure determination?

References
[1] Smith, T.F. & Waterman, M.S. (1981) J. Mol. Biol. 147(1) 195-197.
[2] Aloy, P., Stark, A., Hadley, C. & Russell, R.B. (2003) Proteins: Struct. Funct. Bioinf. 53(S6) 436-456.
[3] Yona, G. & Levitt, M. (2002) J. Mol. Biol. 315(5) 1257-1275.
[4] Söding, J. (2005) Bioinformatics 21, 951-960.
[5] Jumper, J., Evans, R., Pritzel, A. et al. (2021) Nature 596, 583–589.
[6] Baek, M., DiMaio, F., Anishchenko, I. et al. (2021) Science 373(6557) 871-876.