Le mieux est l'enemi du bon; homology modelling with Phyre2 in a deep learning world

L.A. Kelley, H.R. Powell & M.J.E. Sternberg

Structural Bioinformatics Group, Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London SW7 2AZ, UK

harold.powell@imperial.ac.uk

Phyre2 is a web server to predict protein structure from sequence (*www.imperial.ac.uk/phyre2*) that processes ~1,000 individual sequences submitted by users every day. Since its introduction in 2011, Phyre2 has processed well over 4M jobs with ~55,000 unique users per year, each submitting ~20 sequences on average. The Phyre2 (Fig 1) web portal [1] provides both a rapid and user-friendly interface to predict protein structure using homology based template modelling and also resources for analysing the results. The papers describing Phyre2 and its predecessors (3D-PSSM and Phyre) have had over 12,000 citations in the literature.

The performance of different protein structure prediction implementations is compared in a biennial exercise, the Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction; the 14th edition, CASP14, took place in 2020 [2]. The outstanding performance of one program, AlphaFold2 [3] in CASP14 drew the attention of the world's media to this field. The results might lead the casual observer to conclude that the protein structure problem is solved, but at the moment AlphaFold2 itself is not readily accessible to the vast majority of users and the underlying methods employed have not yet been revealed in any detail.

We will show how the carefully designed interface to Phyre2 allows users to generate 3D protein structures from their sequence data in a flexible and straightforward way that makes good models readily available to the community at large.

In addition to a simple mode that allows modelling from single sequences, the Phyre2 web portal proves a range of extra functionality: i) a facility for batch submission of processing of proteomes, ii) searching model genomes for a protein structure, iii) PhyreAlarm, which automatically updates a user if a superior model can be predicted as a result of a newly-deposited structure in the protein data bank, and iv) facilities to analyse a predicted model in terms of accuracy and sequence conservation.

Phyre2 is a resource with the UK node of ELIXIR, the European–wide network of bioinformatics facilities.

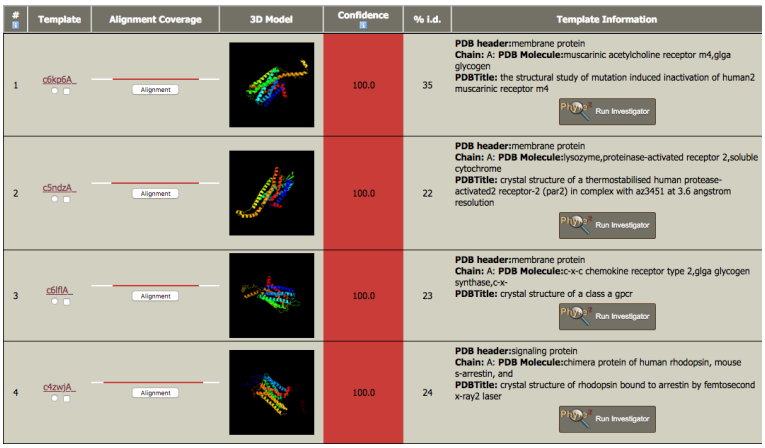

Figure 1. Main results page of the Phyre2 web server showing hits with confidence scores and origin of templates

[1] Kelley et al. (2015) Nature Protocols, 10, 845.

[2] CASP14, *https://www.predictioncenter.org/casp14/index.cgi*

[3] see, *e.g. https://en.wikipedia.org/wiki/AlphaFold*

Keywords: deep learning, protein structure prediction, interface design