

Estimation of Model Accuracy in CASP15 Using the ModFOLDdock Server

Liam McGuffin¹, Nicholas S. Edmunds¹, Shuaa M. A. Alharbi¹, Ahmet G. Genc¹, and Recep Adiyaman¹

¹University of Reading School of Biological Sciences

April 3, 2023

Abstract

In CASP15 there was a greater emphasis on multimeric modelling than in previous experiments, with assembly structures nearly doubling in number (41 up from 22) since the previous round. CASP15 also included a new estimation of model accuracy (EMA) category in recognition of the importance of objective quality assessment for quaternary structure models. ModFOLDdock is a multimeric model quality assessment server developed by the McGuffin group at the University of Reading, which brings together a range of single-model, clustering and deep learning methods to form a consensus of approaches. For CASP15 three variants of ModFOLDdock were developed to optimise for the different facets of the quality estimation problem. The standard ModFOLDdock variant produced predicted scores optimised for positive linear correlations with the observed scores. The ModFOLDdockR variant produced predicted scores optimised for ranking, i.e., the top-ranked models have highest accuracy. In addition, the ModFOLDdockS variant used a quasi-single model approach to score each model on an individual basis. The scores from all three variants achieved strongly positive Pearson correlation coefficients with the CASP observed scores (oligo-IDD) in excess of 0.70, which were maintained across both homomeric and heteromeric model populations. In addition, at least one of the ModFOLDdock variants was consistently ranked in the top two methods across all three EMA categories. Specifically, for overall global fold prediction accuracy, ModFOLDdock placed second and ModFOLDdockR placed third; for overall interface quality prediction accuracy ModFOLDdockR, ModFOLDdock and ModFOLDdockS were placed above all other predictor methods, and ModFOLDdockR and ModFOLDdockS were placed second and third respectively for individual residue confidence scores. The ModFOLDdock server is available at: <https://www.reading.ac.uk/bioinf/ModFOLDdock/>. ModFOLDdock is also available as part of the MultiFOLD docker package: <https://hub.docker.com/r/mcguffin/multifold>

Hosted file

ModFOLDdock_Proteins_Final.docx available at <https://authorea.com/users/602656/articles/633404-estimation-of-model-accuracy-in-casp15-using-the-modfolddock-server>

QMODE2. Residue-based Interface Assessment

PFRMAT QA

TARGET T0999

AUTHOR 1234-5678-9000

REMARK Reliability of residues being in Interfaces

METHOD Description of methods used

MODEL 1

QMODE 2

T1031TS000_1o 0.8 0.4 A1:0.9 A3:0.9 A17:0.7 A19:0.7 B45:0.7 B49:0.4

B50:0.4 B53:0.8

T1031TS999_1o 0.7 X A15:0.5 A17:0.9 A44:0.7 A46:0.7 B4:0.3 B9:0.4

END



