

# **QMEANclust: Estimation of protein model quality by combining a composite scoring function with structural density information**

Benkert P<sup>(1)</sup>, Schwede T<sup>(1)</sup>, Tosatto SCE<sup>(2)</sup>

<sup>(1)</sup> Swiss Institute of Bioinformatics, Biozentrum University of Basel, Switzerland

<sup>(2)</sup> Department of Biology, University of Padova, Padova

## **Motivation**

The selection of the most accurate protein model from a set of alternatives is a crucial step in the protein structure prediction pipeline for both template-based and ab initio approaches. Scoring functions have been developed which can either return a quality estimate for a single model or derive a score from the information contained in the ensemble of models for a given sequence. The latter category was shown to perform considerably better in selecting good candidate models, but these methods tend to fail if the best models are far from the centre of the dominant structural cluster. In this paper we show that model selection can be improved if both approaches are combined by pre-filtering the models used during clustering.

## **Methods**

Our recently published QMEAN composite scoring function (Benkert et al, Proteins 2008) has been improved by including an all-atom interaction potential term. The preliminary model ranking based on the new QMEAN score is used to select a subset of reliable models against which the structural consensus score is calculated.

## **Results**

We achieve a correlation coefficient of predicted quality score and GDT\_TS of 0.9 averaged over the 98 CASP7 (Critical Assessment of techniques for protein Structure Prediction) targets and perform significantly better in selecting good models from the ensemble of server models than any other groups participating in the quality estimation category of CASP7. To the best of the authors' knowledge, QMEANclust is the first method performing significantly better in model selection than the best participating server form Zhang. We also present a local version of QMEAN for the per-residue estimation of model quality (QMEANlocal) and compare it to a consensus-based approach. Participation in the very recent CASP8 experiment has shown QMEANclust to be in the top-performing group for global quality estimation and the best method for local quality estimation.

**Contact :** [silvio.tosatto@unipd.it](mailto:silvio.tosatto@unipd.it)