



QMEAN Server for Model Quality Estimation

[submit new](#) | [example 1](#) | [example 2](#) | [example 3](#) | [help](#) | [references](#) | [contact](#)

QMEAN Server - Quick Help

Introduction

Estimating the quality of protein structure models is a vital step in protein structure prediction. Often one ends up in having a set of alternative models (*e.g.* from different modeling servers or based on alternative template structures and alignments) from which the best candidate shall be selected. Or a single model has been built from which the absolute quality needs to be predicted in order to have an idea about its suitability for subsequent experiments. The QMEAN server provides access to two scoring functions for the quality estimation of protein structure models which allow to rank a set of models and to identify potentially unreliable region within these. Both single models and set of models submitted as *tar.gz*-archives can be analysed. The user has the possibility to choose between the following two scoring functions:

- **QMEAN** [1,3] is a composite scoring function which is able to derive both global (*i.e.* for the entire structure) and local (*i.e.* per residue) error estimates on the basis of *one single model*.
NEW Recently (manuscript in preparation), the QMEAN score has been extended to an absolute quality estimate (see section "Estimated absolute model quality" below).
- **QMEANclust** [3] derives the score for a model by analysing its structural difference to all other models in the *ensemble*. The basic idea behind it is, that structural features observed more frequently have a higher probability to be correct. The initial ranking obtained by QMEAN is thereby used to weight the contribution of each model in the calculation of the QMEANclust consensus score.

The accuracy of the QMEANclust quality estimation improves with the size, the diversity (models from different servers, models based on different templates etc.) and the quality (fraction of near-native structures) of the model ensemble to be analysed. In order to obtain meaningful results a minimum number of models should be provided (*e.g.* > 30 models).

Input format requirements

Either single models (PDB-format) or *tar.gz*-archives with multiple models of the same protein can be uploaded. Additionally, if more than one model is submitted, the full-length sequence of the protein has to be provided (as sequence string or in FASTA format). In the case of multiple models, the sequences of the models are mapped on the target sequence and the models are automatically renumbered if necessary. A flag can be set in order to penalise incomplete models. In this case, the model score is additionally multiplied by the fraction of modelled residues with respect to the input sequence thereby punishing short models.

Results page

As an demonstration, pre-calculated example results from the CASP7 blind test experiment are shown below and are accessible interactively on the following page.

Project:	61 CASP models from different servers (target T0308, first model)
Model quality estimation method used	QMEAN
Number of structures processed	57
Number of structures skipped	4
Penalize incomplete models	no
Additional downloads:	
<ul style="list-style-type: none"> • All results in a single tar.gz-archive • Table with QMEAN scores (.csv) • Table with QMEAN Z-scores w.r.t reference PDBs (.csv) NEW • QMEAN detail table (.csv): all energy terms contributing to QMEAN NEW 	


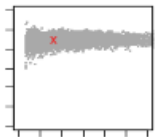
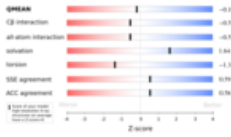

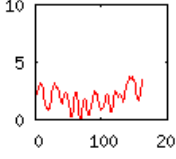
- [QMEAN output in CASP format](#)

The results page begins with an *summary* on the input data and provides compact tables concerning the model ensemble. *E.g.* a table is provided which lists for each model the values of the 6 scoring function terms contributing to QMEAN. A short description of all terms can be found in the table below:

Scoring Function	Description
torsion	Extended torsion potential over 3 consecutive residues. Bin sizes: 45 degree for the center residue, 90 degree for the 2 adjacent residues.
pairwise	Residue-level, secondary structure specific interaction potential using C β atoms as interaction centres. Range 3...25 Å, step size: 1 Å
solvation	Potential reflecting the propensity of a certain amino acid for a certain degree of solvent exposure approximated by the number of C β atoms within a sphere of 9 Å around the centre C β .
all_atom	All-atom, secondary structure specific interaction potential using all 167 atom types. Range 3...20 Å, step size: 0.5 Å
SSE_agree	Agreement between the predicted secondary structure of the target sequence (using PSIPRED) and the calculated secondary structure of the model (using DSSP).
ACC_agree	Agreement between the predicted relative solvent accessibility using ACCpro (buried/exposed) and the relative solvent accessibility derived from DSSP (>25% accessibility => exposed)
QMEAN	The original QMEAN score as published in Benkert <i>et al.</i> 2008. It consists of a linear combination of all the six terms described above. The original QMEAN score has been replaced by the QMEANnorm score in all calculations and is only used in the detail table in order to be able to compare the scores and differentiate between them.
QMEANnorm NEW	Composite score in analogy to QMEAN but based on normalized statistical potential terms. The normalisation reduces the dependence of the quality score on the size of the model (such that larger proteins do not automatically get assigned higher absolute scores). The QMEANnorm score builds the basis of all calculations described below (model ranking, Z-scores). Whenever the term QMEAN is used on the webpage, it refers to the QMEANnorm score described below (<i>manuscript in preparation</i>).

The inspection of the differences of the terms between the models may help understanding which terms contributed most to the low quality estimate of a certain model. For the four statistical potential terms, lower pseudo energies mean higher reliability. The QMEAN score as well as the two agreement terms range from 0 to 1 with higher values for more reliable candidates. In the case of QMEANclust, "local conformational diversity" plots showing the median QMEANclust score per position are provided which help to analyse the diversity within the ensemble of models.

Since the QMEAN score is protein-size dependent (*i.e.* larger proteins tend to have higher scores), the **QMEANnorm** score has been introduced (*manuscript in preparation*). In QMEANnorm, the four statistical potential terms are normalized: the interaction energy is divided by the total number of interactions and the other two (single-body) terms are normalized by the protein size. The QMEAN detail table provided in the summary section contains both the original statistical potentials terms and their normalized counterparts together with the QMEAN and the QMEANnorm scores. Unless not specified otherwise, everywhere the term QMEAN is used on the webpage, it refers to the QMEANnorm score described above which is the primary quality score used to calculate the Z-score and rank the models.

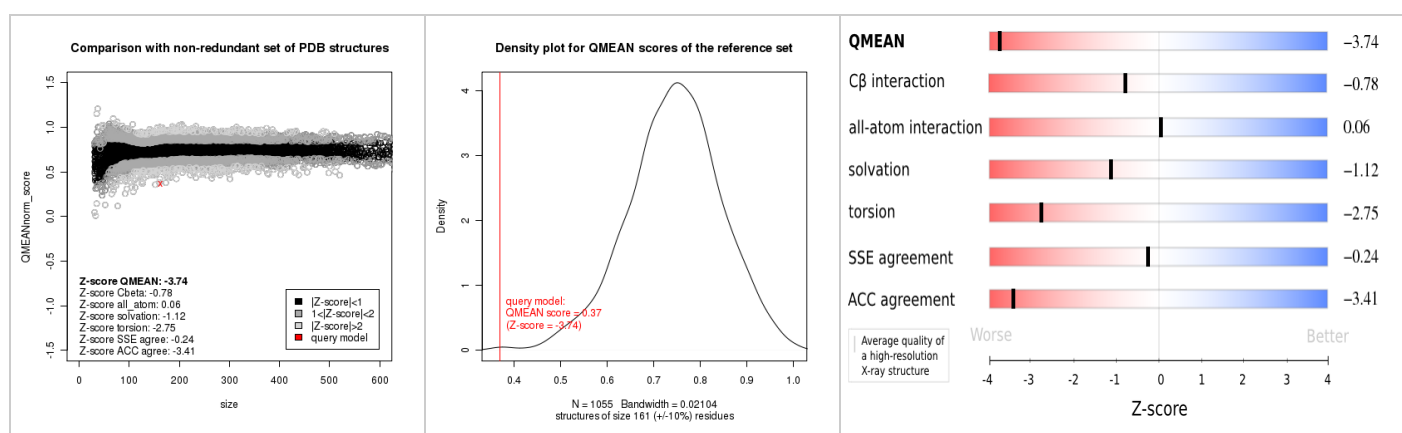
Model name ?	Global scores			Local scores	
	QMEAN score ?	Estimated absolute model quality ? NEW	Z-scores of QMEAN terms ? NEW	Residue error ? 	Residue error plot ?
BAKER-ROBETTA_TS1_T0400	0.689	 Z-score=-0.15 [plot 1] [plot 2]	 [png]	 [jpg] [pdb] Jmol	 [png] [ps] [table]

For each model the following data and plots are provided in separate columns of the output table (depending on the quality estimation method, the ranking of the models is based on the QMEAN or QMEANclust score):

- **Model name:** filename as given in the *tar.gz*-archive
- **QMEAN score / QMEANclust score:** global score of the whole model reflecting the predicted model reliability ranging from 0 to 1.
- **Estimated absolute model quality:** The QMEAN score of the query model is related to the scores of a non-redundant set of high-resolution X-rays structures of similar size and a Z-score is calculated (more details below).
- **Residue error:** The estimated residue error is visualised using a colour gradient from blue (more reliable regions) to red (potentially unreliable regions, estimated error above 3.5 Å). The per residue error is written in the B-factor column (*pdb*-file or coloured model as *jpeg* for download). The molecular graphics viewer Jmol (<http://www.jmol.org/>) can be directly used on the website to interactively inspect the problematic regions in the colour-coded structure.
- **Residue error plot:** model energy profile with estimated residue errors along the sequence (postscript and png-file for download)
- **Energy profiles:** The local model quality data are also provided as tables in tab-separated format. The first table contains for each residue the values of the terms building the QMEAN scoring function. The second table provides the QMEAN/QMEANclust score per residue.

The range of local error estimates varies considerably between the two local versions of QMEAN and QMEANclust. QMEANlocal is, as a consequence of the statistical potential terms used, unable to discriminate between serious and very serious deviations (e.g between 5 Å and 15 Å). QMEANclust on the other hand can, depending on the quality and size of the ensemble, provides error estimates for even large errors.

NEW Estimated absolute model quality



The **QMEAN Z-score** [2] provides an estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography. The QMEAN Z-score is an estimate of the "degree of nativeness" of the structural features observed in a model by describing the likelihood that a model is of comparable quality to high-resolution experimental structures.

The three plots available for download visualize the quality of a given model with respect to these reference structures. The reference structures are a non-redundant subset of the PDB sharing less than 30% pairwise sequence identity among each other and are solved at a resolution below than 2 Å.

Plot 1:

The area built by the circles colored in different shades of grey in the plot on the left hand side represent the QMEAN scores of the reference structures from the PDB. The model's QMEAN score is compared to the scores obtain for experimental structures of similar size (model size +/- 10%) and a Z-score is calculated. A Z-score (or standard score) is a score which is normalised to mean 0 and standard deviation 1. Thus the QMEAN Z-score directly indicates how many standard deviations the model's QMEAN score differs from expected values for experimental structures. In analogy, Z-scores are calculated for all four statistical potential terms as well as the agreement terms being part of the QMEAN score (see also Plot 3).

Plot 2:

The plot in the middle shows the density plot (based on the QMEAN score) of all reference models used in the Z-score calculation. The location of the query model w.r.t. the background distribution is marked in red. This plot basically is a "projection" of the first plot for the given protein size. The number of reference models used in the calculation is shown at the bottom of the plot.

Plot 3:

The analysis of these Z-scores of the individual terms can help identifying the geometrical features responsible for an observed large negative QMEAN Z-score. Models of low quality are expected to have strongly negative Z-scores for QMEAN but also for most of the contributing terms. Large negative values correspond to red regions in the color gradient. "Good structures" are expected to have all sliders in the light red to blue region.

Note:

The quality estimates for **membrane proteins** need to be treated with caution: Membrane proteins may receive very low Z-scores since their physico-chemical properties differ considerably from those of soluble proteins. A QMEAN version with separate potentials optimised for membrane proteins is under development.

References

Reference for the QMEAN scoring function:

[1] Benkert, P., Tosatto, S.C.E. and Schomburg, D. (2008). "QMEAN: A comprehensive scoring function for model quality assessment." *Proteins: Structure, Function, and Bioinformatics*, 71(1):261-277. [PubMed](#)

Reference for the QMEAN Z-scores:

[2] Benkert, P., Biasini, M. and Schwede, T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics* (2010). doi: 10.1093/bioinformatics/btq662 [PubMed](#)

Reference for the QMEANclust scoring function:

[3] Benkert, P., Schwede, T. and Tosatto, S.C.E. (2009). "QMEANclust: Estimation of protein model quality by # combining a composite scoring function with structural density information." *BMC Struct Biol.* 2009 May 20;9:35. [PubMed](#)

Reference for the QMEAN server:

[4] Benkert P, Künzli M, Schwede T. (2009). "QMEAN Server for Protein Model Quality Estimation." *Nucleic Acids Res.* 2009 Jul 1;37(Web Server issue):W510-4. [PubMed](#)