PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

QUALITY: PREDICTIONS

Global and local model quality estimation at CASP8 using the scoring functions QMEAN and QMEANclust

Pascal Benkert,[1,2] Silvio C. E. Tosatto,[3] and Torsten Schwede[1,2]*

[1] Biozentrum, University of Basel, Basel 4056, Switzerland
[2] SIB, Swiss Institute of Bioinformatics, Basel, Switzerland
[3] Department of Biology, Università di Padova, Padova 35121, Italy

ABSTRACT

Identifying the best candidate model among an ensemble of alternatives is crucial in protein structure prediction. For this purpose, scoring functions have been developed which either calculate a quality estimate on the basis of a single model or derive a score from the information contained in the ensemble of models generated for a given sequence (i.e., consensus methods). At CASP7, consensus methods have performed considerably better than scoring functions operating on single models. However, consensus methods tend to fail if the best models are far from the center of the dominant structural cluster. At CASP8, we investigated whether our hybrid method QMEANclust may overcome this limitation by combining the QMEAN composite scoring function operating on single models with consensus information. We participated with four different scoring functions in the quality assessment category. The QMEANclust consensus scoring function turned out to be a successful method both for the ranking of entire models but especially for the estimation of the per-residue model quality. In this article, we briefly describe the two scoring functions QMEAN and QMEANclust and discuss their performance in the context of what went right and wrong at CASP8. Both scoring functions are publicly available at http://swissmodel.expasy.org/qmean/.

Proteins 2009; 77(Suppl 9):173–180.

Key words: CASP8; model quality assessment; QMEAN; scoring function; protein structure homology modeling; mean force potential.

INTRODUCTION

In the course of protein structure prediction typically a set of alternative models is produced, from which the most accurate candidate is selected using a scoring function. Estimating the quality of protein structure models is a critical step in protein structure prediction because it is the quality of a model which determines its suitability for real-world applications.[1,2] Since 2006, quality assessment is an official category of CASP.[3,4]

Scoring functions can be broadly categorized into two groups: (1) approaches being able to estimate the quality of a single model without relying on consensus information and (2) consensus or clustering methods relying on the comparative analysis of the structural similarity among the models in an ensemble.

A variety of different aspects of proteins have been used for the estimation of model quality: the stereochemical plausibility of the models,[5,6] compatibility of amino acid residues with their local environment,[7] evolutionary information,[8–10] as well as energy-based methods which include empirical force fields[11,12] and knowledge-based statistical potentials.[13–20] A straightforward estimator of the quality of

Additional Supporting Information may be found in the online version of this article.
Abbreviations: CASP, critical assessment of techniques for protein structure prediction; PDB, Protein Data Bank; TBM, template-based modeling.

The authors state no conflict of interest.
*Correspondence to: Torsten Schwede, Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, Basel CH-4056, Switzerland.
E-mail: torsten.schwede@unibas.ch
Received 20 March 2009; Revised 12 June 2009; Accepted 30 June 2009
Published online 14 July 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22532

a protein structure model is its sequence similarity to the template(s) of known structure which is a measure for the amount of information that can be directly derived from the template structure.[21] Recently, several scoring functions have been described which use the extent to which a model agrees with template information as a measure of its quality, e.g., by analyzing the agreement of the model with structural fragments from templates[22] or distance constraints extracted from alignments to various templates.[23] A promising strategy to improve the performance in discriminating good from bad models is to combine multiple scoring function terms covering different geometrical aspects of protein structures. Composite scoring functions analyzing multiple structural features have been introduced and shown to perform better than any single term.[24–29] Previously, we introduced the composite scoring function QMEAN which stands for Quantitative Model Energy ANalysis.[24] The original version of the QMEAN scoring function is a linear combination of three statistical potential terms and two agreement terms reflecting the extent a model agrees with predicted features from its sequence (i.e., secondary structure, solvent accessibility). The implementation of QMEAN used at CASP8 additionally includes an all-atom interaction term as described recently.[30] Besides the global quality, local error estimation on a per residue basis has become an active field of research.[10,29–31] Although the accuracy of local predictions is still limited, these methods can help discriminating between reliable and unreliable regions within models.

In contrast, consensus-based methods derive a quality score from the pairwise comparison of all models in the ensemble. They are based on the idea that conformations predicted more frequently are more likely to be correct than structural patterns occurring in only a few models.[30,32–37] At CASP7[3,38] and CASP8,[4] consensus methods have performed considerably better in distinguishing good from bad models than scoring functions operating on single models. However, methods relying solely on structural consensus information have inherent limitations: First, they are not able to estimate the quality of a single model or to rank a small set of models. Second, these methods are likely to fail in model selection when the best models are not part of the dominant structural cluster of the ensemble and as a consequence outstanding predictions might not be recognized.[30,38] At CASP8, we investigated whether a hybrid approach which combines consensus information with a single model scoring function is able to counteract the latter limitation. Whereas traditional consensus methods calculate an error estimate based on an all-against-all comparison of the models in the ensemble, the hybrid scoring function presented in here (QMEANclust) additionally takes advantage of an initial ranking of the individual models obtained by the composite scoring function QMEAN to preselect a subset of higher quality models for subsequent clustering.[30]

At CASP8, we participated with different servers based on our recently published composite scoring function QMEAN. In this article, we will focus on the performance of the QMEAN and QMEANclust method. Beside the composite scoring function QMEAN, a second approach operating on single models which combines QMEAN with evolutionary information (QMEANfamily) is introduced. In this article, we give a qualitative description of the performance of the QMEAN-based scoring functions, describe our lessons learnt at CASP8 in terms of what went right and wrong and provide an outlook on possible extensions of our methods.

METHODS

QMEAN

The composite scoring function QMEAN[24,30] is a linear combination of six structural descriptors. A torsion angle potential over three consecutive amino acids is used to analyze the local geometry of the model, a solvation potential describes the burial status of the residues and two distance-dependent interaction potentials based on Cβ atoms and on all atom types are used to assess long-range interactions. Additionally, two terms describing the agreement of predicted and calculated secondary structure and solvent accessibility are included.[24,30]

The local per-residue version of the QMEAN scoring function used to predict the per-residue error at CASP8 consists of eight terms.[30] All terms are calculated over a sliding window of nine residues and a triangular smoothing weighting scheme has been applied as described elsewhere.[9] Adapted versions[30] of the six global QMEAN terms are combined with two additional features, namely, the average solvent accessibility (using triangular smoothing) and the fraction of residues in the 9-residue window with no assigned secondary structure by DSSP.[39] These two features take into account that, for example, solvent exposed loops are potentially less accurate than regions of regular secondary structure in the structural core of the protein.

QMEANclust

QMEANclust combines QMEAN with consensus information obtained from the ensemble of models. The consensus information is stored in a distance matrix containing all against all pair-wise similarities between the models based on the GDT_TS score calculated by the program TMscore[40] (see schematic representation in Fig. 1). In the implementation used during CASP8, the QMEAN scoring function was used to select a subset of

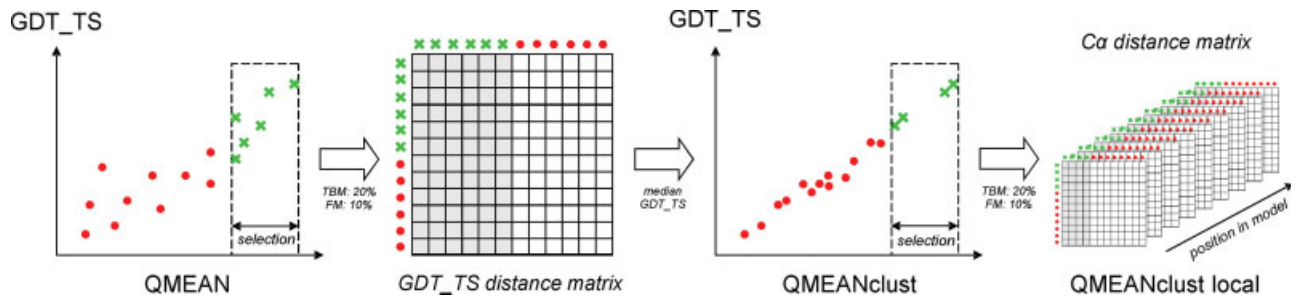

Figure 1

For the calculation of the QMEANclust consensus score, each model from the ensemble of server predictions for a given target is compared to the models in the reference set defined by QMEAN. The top 20% and 10% of the models based on QMEAN enter the reference set for likely template-based (TBM) and free modeling (FM) targets, respectively. In analogy, the QMEANclust ranking is used to select reference models in the calculation of local (per-residue) consensus score. Green symbols represent models selected for the reference set. Gray areas in the distance matrices mark values used in the calculation of the consensus score. The QMEANclust score is defined as the median GDT_TS distance of a model to all the models in the reference set.

reference models for which the structural consensus was then calculated (QMEANclust score is defined as the median GDT_TS of a given model to all the other models in the reference set). We expected that for template-based modeling targets (TBM), the ensemble of models would contain more meaningful consensus information than ensemble of free modeling (FM) targets. We therefore discriminated between likely "TBM" and "FM" target on the basis of the median QMEAN score of the ensemble: if the median predicted reliability score (ranging from 0 to 1) exceeds 0.4, the target is assigned to the template-based modeling category and the top 20% models according to QMEAN are used. For the remaining targets classified as likely "FM" targets, only the top 10% models based on QMEAN are used for the calculation of the consensus score. The accuracy of the classification is 88% for likely "FM" targets (i.e., 88% of the targets predicted to be difficult have a median GDT_TS below 0.4) and 74% for "TBM" targets.

In analogy to the global quality estimation, only a subset of reference models is used in the calculation of the local QMEANclust consensus score (Fig. 1). The same selection cut-offs as for the global version are used (20% for TBM and 10% for FM) but the global QMEANclust consensus score is used for the selection of reference models instead of QMEAN. This leads to further enrichment of more reliable models and, as a consequence, potentially more correct local conformations in the selection. The basic idea behind the approach is that some of the models in the ensemble may be totally incorrect and using them as reference models would result in an overestimation of the local error. Furthermore, models with a higher global reliability score are likely to be better candidates for the local analysis because they have a higher probability to contain correct local conformations. The local QMEANclust score at a certain position in a model is defined as the median Cα distance to equivalent positions in the reference models after superposition by TMscore. In summary, QMEAN is first used to select more reliable reference models for the computation of the QMEANclust score which is subsequently used to select a subset of candidate models for the calculation of the local consensus score (Fig. 1).

QMEANfamily

QMEANfamily is a single model energy function which takes into account additional information from evolutionary closely related proteins being part of the same family as the query protein. For each model in the set, an ensemble of up to 50 supplementary models based on protein sequences sharing at least 40% sequence identity to the target is generated using the original model as template structure. The homologous sequences are identified using BLAST[41] to search NCBI's nonredundant sequence database clustered at 70% sequence identity. The BLAST alignments are converted into raw models by copying the coordinates of the template structure without modeling any insertions and nonconserved side chains. The QMEANfamily score is defined as the average QMEAN score of these models covering the protein family.

RESULTS AND DISCUSSION

In this section, the results for the scoring functions QMEAN and QMEANclust are described in a qualitative manner and discussed critically concerning what went right and wrong at CASP8. The performance of the QMEANfamily scoring functions is addressed only briefly by comparing it to QMEAN from which it is derived (see Methods). A detailed analysis of the overall performance of our methods in comparison to others is not in

Table I
Comparison of the Top 10 Global Consensus Scoring Functions Participating at CASP8

Group name	Group ID	No. of targets	Sum of positive *Z*-scores	Avg (Pearson)
Pcons_Pcons	239	122	108.0	0.92
ModFOLDclust	31	122	107.7	0.92
SAM-T08-MQAC	56	121	105.5	0.91
QMEANclust	27	121	102.8	0.90
MULTICOM	453	121	100.6	0.90
MULTICOM-CLUSTER	20	122	97.5	0.89
McGuffin	379	121	92.6	0.89
GS-MetaMQAPconsI	273	119	89.4	0.88
TASSER	57	121	85.0	0.87
LEE	407	120	79.8	0.85

the scope of this article and is available in the assessor article of the quality estimation category in this issue.[4]

Global model quality estimation

At CASP8, QMEANclust was amongst the top performing scoring functions in estimating the global model quality together with three other methods (see Table I). The performance difference between these approaches was statistically not significant.[4] As in the official assessment, our performance measure is based on the Pearson's correlation coefficient between the predicted model quality and the GDT_TS of the model to the experimental structure. More precisely, the overall ranking is based on the sum of the non-negative *Z*-scores of the correlation coefficients over all targets[4] (Table I).

For 21 out of 122 accepted targets, QMEAMclust submitted the best quality predictions among all participants. Figure 2 shows two examples (T0434 and T0448) in which the strategy of QMEANclust to rank the models prior consensus calculation led to an improved performance as compared to a traditional consensus method. To compare our predictions with an exclusively consensus-based score, an in-house implementation of 3d-jury based on GDT_TS was used.[32] Both targets T0434 and T0448 emphasize the fact that the correlation between predictions and quality of the models based on the traditional consensus implementations is dominated by the most prominent structural cluster and that these methods tend to fail if the best models are not part of it. In contrast, the QMEANclust approach to preselect models leads to better correlations and, even more importantly, the quality of the highest scoring model according to the scoring function improves about 10 GDT_TS units (marked by arrows in Fig. 2).

On the other side, for 13 targets out of the 122 targets, the Pearson's correlation coefficients based on QMEANclust deviate by more than 0.5 *Z*-score units from the best submissions. The main explanation for the difference in performance is the low quality of the initial QMEAN rankings for most of these targets as can be seen for target T0409 in Figure 2. The initial QMEAN-based selection (the reference set is indicated by the dotted area) contains models from two structurally distinct clusters with a majority of the models belonging to the lower quality cluster. The low quality models determine the resulting QMEANclust ranking and lead to a GDT_TS loss of approximately 10 units compared to the 3d-jury approach. The majority of the 13 outlier targets represent difficult modeling cases with the best model in the set having a GDT_TS below 0.6. For very difficult modeling targets with few or no available template information, the QMEAN scoring (i.e., the statistical potential terms) may fail completely in discriminating good from bad models. Adjusting the weights and the coarseness of the QMEAN terms to the modeling difficulty may overcome this problem. For example, the level of detail of the all-atom interaction potential is probably too high for the analysis of coarse ab initio models potentially even having an incorrect fold and may be only useful for the discrimination of highly accurate template-based models.

The performance of the QMEANclust version used during CASP8 heavily depends on the initial QMEAN selection which is, as mentioned earlier, not optimal for some difficult modeling targets. An alternative version of QMEANclust which has been implemented after CASP8 combines the initial QMEAN raking with clustering information by using the QMEAN score to weight the contribution of each model to the consensus score (i.e., QMEANclust as the weighted average GDT_TS score to all models). Our data suggest that this version of QMEANclust outperforms all CASP8 scoring functions and works considerably more stable on the above 13 targets (data not shown). This implementation takes advantage of the information provided by QMEAN to prioritize a model's contribution to the consensus score but does not need a strict cut-off which prevents the drift to a low-quality cluster.

In the CASP8 quality assessment ranking, the composite scoring function QMEAN was among the best performing methods operating on single models. The ability of the QMEAN scoring function to identify good quality models among a set of alternatives builds the basis of the model preselection implemented in QMEANclust. The three plots in the left panels of Figure 2 show the correlation between QMEAN and GDT_TS for the server models of three representative targets. As can be seen from these plots, there are often a few low-quality models which have been overestimated by QMEAN. The reason for these outliers seems to be that similar scoring function terms have been used in the derivation and optimization of these models as compared to those included in QMEAN. The outlier models have been mainly submitted by the same groups over most of the targets. These low-quality outliers in the QMEAN-based quality estimation explain the weak performance of the QMEANclust scoring function on some targets as described earlier.

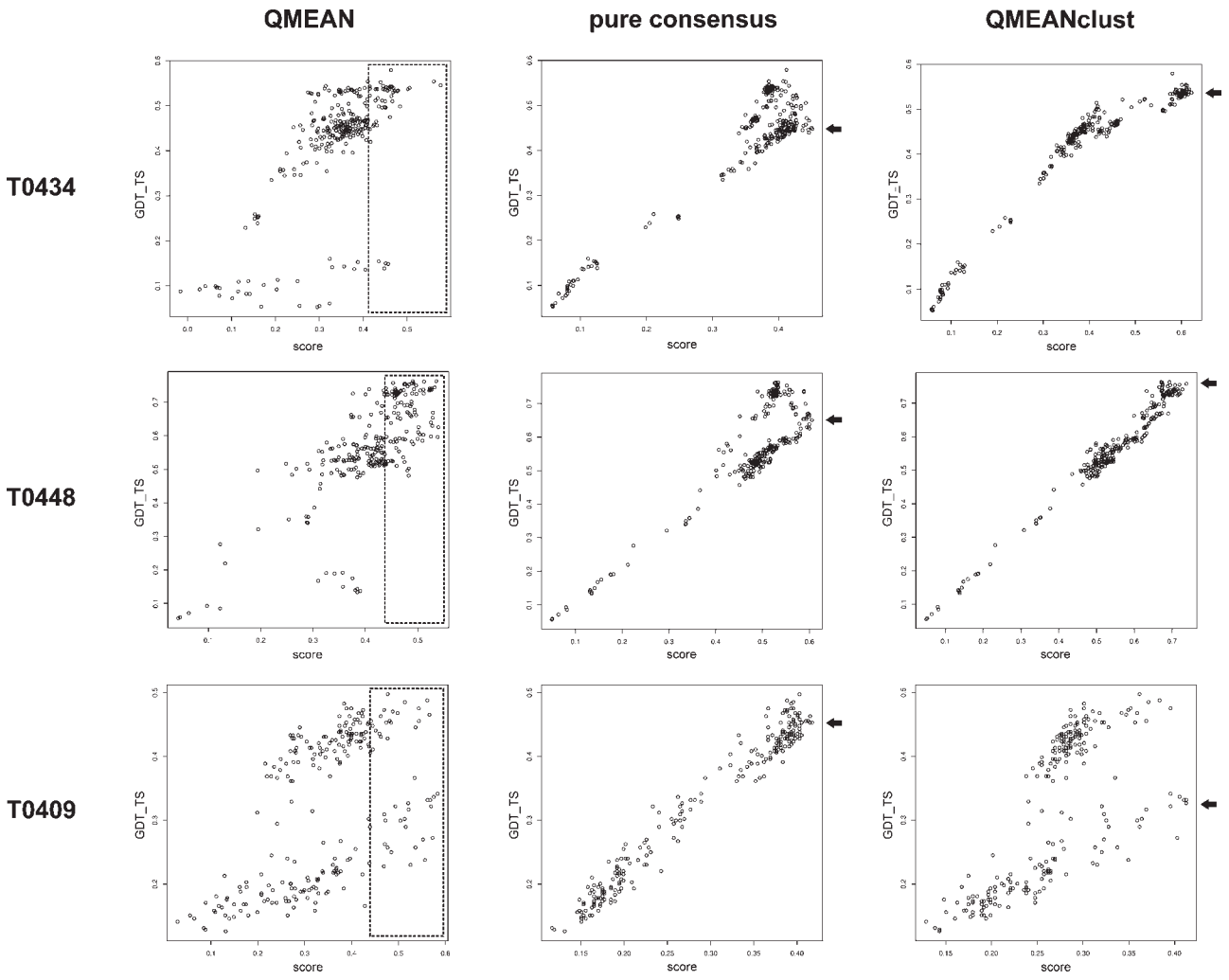

Figure 2

Scatter plots of three targets are shown for QMEAN, 3D-jury, and QMEANclust (from left to right). The dashed areas mark the models selected by QMEAN as the basis for QMEANclust. The first two targets represent examples in which the QMEANclust method results in a considerably better model selection. The plots show the predicted model quality according to the given scoring function on the *x*-axis and the similarity of the models to the experimental structure (based on GDT_TS) on the *y*-axis. The top ranked model according to the scoring function is indicated by an arrow.

In case of comparative modeling targets, comparing the models to available templates structures would have allowed us to identify these outliers, because they are structurally most distant from the available best templates.

Besides combining multiple terms covering different geometrical aspects in a composite scoring function (such as QMEAN), another successful strategy to estimate model quality based on a single model is to analyze to which extent models deviate from known template structures. Karplus and coworkers used a scoring function at CASP8 which investigates how strongly a model deviates from distance constraints extracted from various target-template alignments.[23] A similar idea was used in the quality category assessment at CASP7 in the form of the "naïve predictor" introduced by Anna Tramontano and coassessors.[3] In their work, a quality score was assigned to each model according to how close it was to a protein of known structure. QMEAN does not include any information about the availability of template structures in the derivation of the quality score. However, as mentioned earlier, the integration of template information would have helped avoid the selection of most of the low-quality outliers.

In addition to QMEAN, a second single model scoring function called QMEANfamily has been tested at CASP8. The basic idea behind QMEANfamily is that a correct model for a given protein should also be a good model for other members of the same family, and certain incompatible structural features might be easily detected in models for some protein sequences of the family than for others. For each of the assessed models, QMEANfamily

Table II
Comparison of the Top 10 Local Consensus Scoring Functions Participating at CASP8

Group name	Group ID	No. of targets	Avg (avg (Pearson))
QMEANclust	27	121	0.71
selfQMEAN	251	117	0.69
MULTICOM	453	121	0.68
Mariner2	364	90	0.68
Pcons_Pcons	239	122	0.67
MULTICOM-CLUSTER	20	122	0.66
ModFOLDclust	31	122	0.63
GS-MetaMQAPconsI	273	119	0.61
MODCHECK-Jury	52	121	0.60
GS-MetaMQAPconsII	134	118	0.52

builds up to 50 supplementary models based on alignments between the sequences of the target protein and BLAST-detectable homologues—using the model to be evaluated as "template structure." For 106 targets evaluated by both QMEAN and QMEANfamily scoring functions, the latter generates better correlations in 98 cases. The differences in terms of the Pearson's correlation coefficient are mostly small and exceed 0.5 *Z*-score units for only five CASP targets. However, over all common targets the performance difference between QMEANfamily and QMEAN is statistically highly significant with a *P*-value of 0.00061 in a paired *t*-test. Averaging the score over supplementary models seems to primarily reduce the variance in the correlations but only occasionally improves model selection (i.e., the quality of the highest ranking model).

Local model quality estimation

The QMEANclust consensus scoring function[30] was the top performing method in the local quality assessment category (see Table II). For 50 out of 122 targets, QMEANclust submitted the best per-residue error estimates of all participating groups. The original evaluation data can be found on the official CASP8 website (http://predictioncenter.org/casp8/qa_analysis.cgi) and for a detailed comparison to other methods see the quality assessment evaluation article in this issue.[4]

Despite the overall good performance, for 10% of the targets, the local QMEANclust predictions were considerably worse than the best submissions (more than 0.5 *Z*-score units difference). In approximately half of these cases, the global preselection based on QMEAN failed in the selection of the best models in the first step of the procedure explaining, therefore, the poorer performance compared to pure consensus approaches.

Figure 3 shows plots for of the predicted local errors vs. the measured model-structure Cα distances for target T0475 models submitted by the participating servers and evaluated by the QMEAN and QMEANclust scoring functions. The average QMEANclust Pearson's correlation coefficient obtained for target T0475 is 0.77. The QMEANclust local plot in the right panel of Figure 3 shows a cluster of residues deviating from the diagonal. This cluster contains almost exclusively residues belonging to two regions of the target (residues 87+88 and C-terminal residues 119–122) which could not be modeled correctly using the available templates. More precisely, most of the top scoring models, which were used to

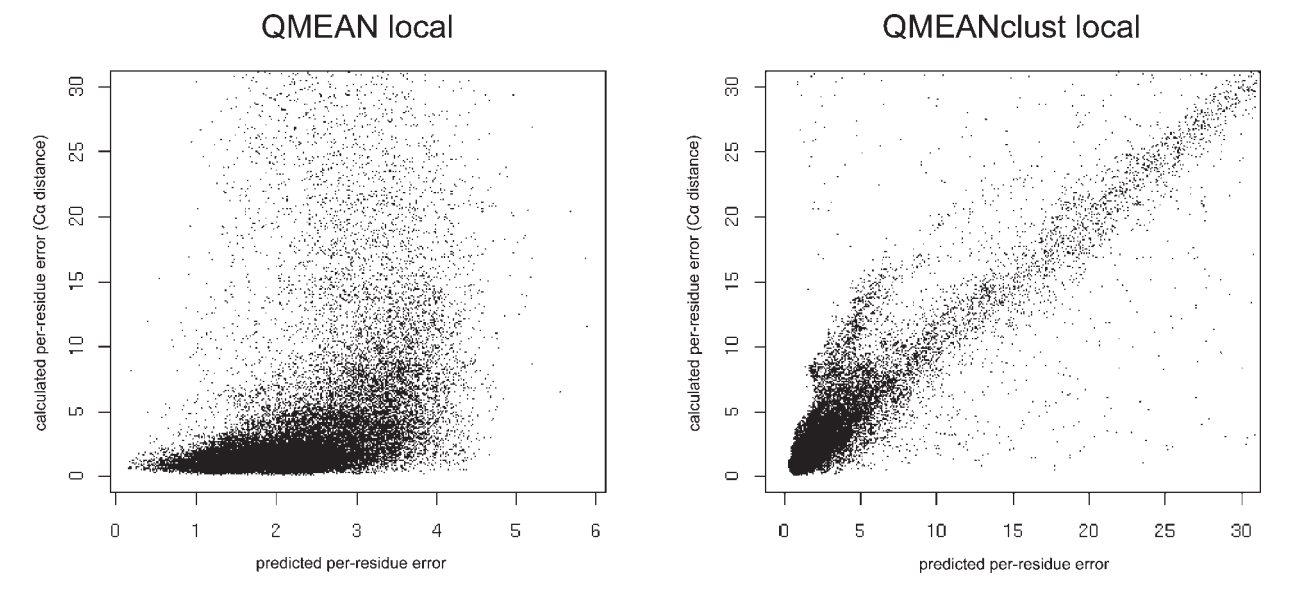

Figure 3
Example scatter plots demonstrating the ability of QMEAN (left) and QMEANclust (right) to predict the model's per-residue error. Correlation of predicted local deviation in Angstrom and measured Cα distances of all residues from all server models submitted for target T0475.

derive the consensus score, consistently contain the same incorrect conformation (probably by copying from similar templates). This renders the structural consensus information useless for estimating the local quality in these regions. The prediction of the per-residue errors based on the consensus method can be quite accurate for some comparative modeling targets, reaching average Pearson's correlation coefficients of over 0.9.

The local QMEANclust plot in Figure 3 (right) shows the correlation between predicted and observed local error over all residues and models of target T0475, however, it does not address the scoring function's ability to rank equivalent segments or residues from several models. We calculated the correlation coefficients between the local QMEANclust score and deviation from the experimental structure for each position in target T0475 (Supporting Information Fig. S1) and on the basis of continuous segments of length 10 residues (Supporting Information Table SI). In agreement to Figure 3, we observe two regions of lower correlation corresponding to the off-diagonal outliers mentioned in the previous paragraph. Similar per-residue correlation coefficients between predicted and calculated local errors have been observed for most of the medium to easy template-based modeling targets (data not shown). We further investigated whether QMEANclust is able to rank local segments (or residues) and may potentially be used to build hybrid models by combining segments of different models (Supporting Information Figs. S2 and S3). The selected regions are often comparably close to the experimental structure as the segments of the best submitted server model for a given target, with the exception of two regions of low correlation and high structural diversity (see earlier).

The local version of the composite scoring function QMEAN was also performing well, being one of the best local nonconsensus error estimation methods at CASP8. However, QMEAN's ability to estimate the absolute per-residue error in Ångstrom is clearly limited compared to QMEANclust. In contrast to consensus scoring functions, statistical potentials tend to fail in discriminating approximately correct from incorrect conformations. As a consequence, the estimated per-residue error based on the local QMEAN scoring function rarely exceeds a correct predicted deviation of 5 Å. Nevertheless, the main purpose of the local quality estimation of QMEAN is to help identifying regions in single models which are potentially incorrect. Figure 3 shows that the correlation between the local QMEAN score and the Cα distances is quite low ($r = 0.53$) for T0475 but at least a rough discrimination between less and more reliable regions in the models is possible. There is clearly room for improvement in the local prediction of model quality. Currently, the local QMEAN score is based on a simple linear combination of eight terms. A reasonable extension may be an improved combination of the terms and including information about the specific modeling target using advanced machine learning techniques.

CONCLUSIONS

In general, consensus methods perform significantly better in assessing server models than physics-based or evolutionary methods operating on single models. Nevertheless, the latter category of methods plays an important role in assessing individual or small sets of models. Hybrid methods combining single model scoring functions with structural consensus information can be used to counteract some of the shortcomings of pure consensus methods. The idea of using a scoring function such as QMEAN operating on single models in order to prioritize the contribution of the models within the ensemble used to generate the consensus score has been shown to be a promising strategy both for the global and especially for the local quality estimation of models. In particular, the strategy of using a subset of higher quality models for the derivation of the local consensus score has the capacity to improve local error estimation. Such hybrid approaches allow overcoming inherent limitations of pure consensus methods which select models of the most dominant structural cluster and miss accurate prediction which are outside this cluster. At CASP8, QMEANclust and QMEAN were among the top performing consensus and nonconsensus scoring functions, respectively. The performance of QMEANclust can be further enhanced by improving the initial model ranking based on the single model scoring function QMEAN (e.g., by analyzing the agreement of a model with information obtained from known template structures) and by optimizing the way how both approaches are combined. A reasonable extension would be an adjustment of the relative weights of the terms to the modeling difficulty (e.g., to use more coarse terms for the evaluation of template-free models) and using a more advanced machine learning approach. The two scoring functions QMEAN and QMEANclust are publicly available as part of the QMEAN server[42] under the following URL: http://swissmodel.expasy.org/qmean/.

ACKNOWLEDGMENTS

The authors thank Fabiano Cimarosti for setting up the HTTP server for CASP8 as well as Michael Kuenzli for establishing the public QMEAN server. They are also grateful to James Battey and Lorenza Bordoli for helpful comments on the manuscript.

REFERENCES

1. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–96.

2. Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, Brenner SE, Burley SK, Das R, Dokholyan NV, Dunbrack RL, Fidelis K, Fiser A, Godzik A, Huang YJ, Humblet C, Jacobson MP, Joachimiak A, Krystek SR, Kortemme T, Kryshtafovych A, Montelione GT, Moult J, Murray D, Sanchez R, Sosnick TR, Standley DM, Stouch T, Vajda S, Vasquez M, Westbrook JD, Wilson IA. Outcome of a workshop on applications of protein models in biomedical research. Structure 2009; 17:151–159.
3. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. Assessment of predictions in the model quality assessment category. Proteins 2007;69 (Suppl 8):175–183.
4. Cozzetto D, Kryshtafovych A, Tramontano A. Evaluation of CASP8 model quality predictions. Proteins 2009;77(Suppl 9):157–166.
5. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
6. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
7. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83–85.
8. Chen H, Kihara D. Estimating quality of template-based protein models by alignment stability. Proteins 2008;71:1255–1274.
9. Tress ML, Jones D, Valencia A. Predicting reliable regions in protein alignments from sequence profiles. J Mol Biol 2003;330: 705–718.
10. Wallner B, Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Sci 2006;15:900–913.
11. Dominy BN, Brooks CL. Identifying native-like protein structures using physics-based potentials. J Comput Chem 2002;23: 147–160.
12. Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1999;288:477–487.
13. Lu H, Skolnick J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. Proteins 2001;44:223–232.
14. Melo F, Feytmans E. Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 1998;277:1141–1152.
15. Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–2524.
16. Sippl MJ. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 1990;213:859–883.
17. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 2002;11:2714–2726.
18. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. Nature 1992;358:86–89.
19. Shortle D. Propensities, probabilities, and the Boltzmann hypothesis. Protein Sci 2003;12:1298–1302.
20. Tosatto SC, Battistutta R. TAP score: torsion angle propensity normalization applied to local protein structure evaluation. BMC Bioinformatics 2007;8:155.
21. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823–826.
22. Zhou H, Skolnick J. Protein model quality assessment prediction by combining fragment comparisons and a consensus C(alpha) contact potential. Proteins 2008;71:1211–1218.
23. Paluszewski M, Karplus K. Model quality assessment using distance constraints from alignments. Proteins 2009;75:540–549.
24. Benkert P, Tosatto SCE, Schomburg D. QMEAN: a comprehensive scoring function for model quality assessment. Proteins 2008;71:261–277.
25. Eramian D, Shen M-Y, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score for predicting errors in protein structure models. Protein Sci 2006;15:1653–1666.
26. Pettitt CS, Mcguffin LJ, Jones DT. Improving sequence-based fold recognition by using 3D model quality assessment. Bioinformatics 2005;21:3509–3515.
27. Qiu J, Sheffler W, Baker D, Noble WS. Ranking predicted protein structures with support vector regression. Proteins 2008;71:1175–1182.
28. Wallner B, Elofsson A. Can correct protein models be identified? Protein Sci 2003;12:1073–1086.
29. Tosatto S. The victor/FRST function for model quality estimation. J Comput Biol 2005;12:1316–1327.
30. Benkert P, Schwede T, Tosatto SCE. QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. BMC Struct Biol 2009;9:35.
31. Fasnacht M, Zhu J, Honig B. Local quality assessment in homology models using statistical potentials and support vector machines. Protein Sci 2007;16:1557–1568.
32. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
33. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci 2001;10:2354–2362.
34. Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci USA 1998;95:11158–11162.
35. Wang K, Fain B, Levitt M, Samudrala R. Improved protein structure selection using decoy-dependent discriminatory functions. BMC Struct Biol 2004;4:8.
36. Xiang Z, Soto CS, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 2002;99:7432–7437.
37. Mcguffin LJ. Benchmarking consensus model quality assessment for protein fold recognition. BMC Bioinformatics 2007;8:345.
38. Wallner B, Elofsson A. Prediction of global and local model quality in CASP7 using Pcons and ProQ. Proteins 2007;69:184–193.
39. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
40. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57:702–710.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
42. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res 2009;37:W510–W514.