OXFORD

Structural bioinformatics

ProQ2: estimation of model accuracy implemented in Rosetta

Karolis Uziela[1] and Björn Wallner[2,3]*

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden, [2]Division of Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, SE-581 83, Linköping, Sweden and [3]Swedish e-Science Research Center, Linköping, Sweden

*To whom correspondence should be addressed.
Associate Editor: Anna Tramontano

Abstract

Motivation: Model quality assessment programs are used to predict the quality of modeled protein structures. They can be divided into two groups depending on the information they are using: ensemble methods using consensus of many alternative models and methods only using a single model to do its prediction. The consensus methods excel in achieving high correlations between prediction and true quality measures. However, they frequently fail to pick out the best possible model, nor can they be used to generate and score new structures. Single-model methods on the other hand do not have these inherent shortcomings and can be used both to sample new structures and to improve existing consensus methods.
Results: Here, we present an implementation of the ProQ2 program to estimate both local and global model accuracy as part of the Rosetta modeling suite. The current implementation does not only make it possible to run large batch runs locally, but it also opens up a whole new arena for conformational sampling using machine learned scoring functions and to incorporate model accuracy estimation in to various existing modeling schemes. ProQ2 participated in CASP11 and results from CASP11 are used to benchmark the current implementation. Based on results from CASP11 and CAMEO-QE, a continuous benchmark of quality estimation methods, it is clear that ProQ2 is the single-model method that performs best in both local and global model accuracy.
Availability and implementation: https://github.com/bjornwallner/ProQ_scripts
Contact: bjornw@ifm.liu.se
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein structure modeling represents a fundamental challenge in structural bioinformatics and is crucial for a detailed understanding of the biological function of molecules. It can be used to guide and explain experiments, as well as for prediction of proteins whose structure for the most part is unknown ($\sim$105k known structures vs. 50 000k known sequences). A common technique in structure modeling is to generate many alternative models and then use a program to estimate the model accuracy to select the best model. Alternatively, the estimated accuracy can also be used to assess the absolute quality of a single model, i.e. a measure that is related to

similarity to true native structure (Wallner and Elofsson, 2003; Z. Wang *et al.*, 2009).

ProQ2 (Ray *et al.*, 2012) is a single-model method that estimates model accuracy using a support vector machine (SVM) to predict the quality of a protein model by combining structural and sequence based features calculated from the model. The structural based features are contacts between 13 different atom types, residue-residue contacts grouping amino acids into six different groups (hydrophobic, positively charged, etc.) and surface accessibility using the same residue groups. The sequence-based features are calculated from information predicted from sequence, i.e. secondary structure, surface

accessibility and sequence profiles. For example, one such feature is predicted secondary structure agreement with the actual secondary structure in the model, and there is also a similar feature for predicted surface accessibility. To calculate all the features needed for a prediction ProQ2, used many different external programs such as PSI-BLAST (Altschul *et al.*, 1997), PSIPRED (McGuffin *et al.*, 2000) ACCpro (Cheng *et al.*, 2005), Naccess (Hubbard and Thornton, 1993), ProQres (Wallner and Elofsson, 2006), Stride (Frishman and Argos, 1995) and SVM-light (Joachims, 2002). These dependencies made it difficult to distribute the program, to run large batches and to incorporate in novel modeling protocols.

Here, we remove the dependency on Naccess, ProQres, Stride and SVM-light by incorporating ProQ2 as scoring function in the Rosetta modeling suite. We also provide scripts to run the remaining packages (see Availability), and to prepare input files to ProQ2, making the setup as smooth as possible. If you for instance already have a version of Rosetta installed the only step needed to use ProQ2 is to download the scripts that will prepare the input files. A further advantage of the new implementation is that it enables usage of the modeling capabilities of Rosetta, and allows for easy integration with existing Rosetta protocol. Here, demonstrated by the novel method ProQ2-refine, which uses the ability of Rosetta, to rebuild side-chains followed by selection based on ProQ2 score.

2 Methods

2.1 Implementation

ProQ2 (Ray *et al.*, 2012) was implemented as scoring function in the Rosetta modeling software (http:/www.rosettacommons.org). ProQ2 uses two sets of features, one that is calculated from the model sequence and one from the structural model. The sequence-based features are calculated *once* for a given sequence and used as input to Rosetta. The structural features, i.e. contacts, surface areas and secondary structure, and the final prediction using linear SVM weights are all calculated by Rosetta during scoring. There is still some dependency on external programs to calculate the sequence-based features predicted from sequence. For the structural-based features we adapted an already existing implementation of DSSP (Kabsch and Sander, 1983) and Naccess (Hubbard and Thornton, 1993) to assign secondary structure and calculate exposed residue surface. The atom–atom and residue–residue contacts previously calculated by ProQres (Wallner and Elofsson, 2006), were implemented directly in Rosetta as well as the functionality to read and predict SVM models. To account for implementation details, the SVM weights, used previously by ProQ2, were retrained using the original ProQ2 training set.

2.2 Data sets

Data from the Quality Assessment category in CASP11 was downloaded from the CASP11 website. Targets were split into EASY and HARD based on the official CASP definitions as follows: targets in Template Based Modeling (TBM) to EASY and targets in Free Modeling (FM) into HARD, the borderline TBM-hard category was put in EASY if the average model quality as measured by GDT_TS (Zemla *et al.*, 1999) was >40 otherwise in HARD (see Supplementary Table S1). To be able to assess if any given selection is better than random, GDT_TS scores for each model were converted to Z-scores by subtracting the mean GDT_TS and divide by the GDT_TS standard deviation over all models for each target. After filtering cases where any method lacked predictions there were 89 targets, from 13 232 models containing 3 171 799 residues.

CAMEO-QE data for the period 2014.06.06–2015.05.30 was obtained from the Protein Model Portal (Haas *et al.*, 2013). CAMEO is an ongoing community effort in which all newly solved PDB structures are used as targets for structure prediction severs, as in CASP there is also a quality estimation part, CAMEO-QE. All the public methods, i.e. ModFOLD4 (McGuffin *et al.*, 2013), ProQ2 (Ray *et al.*, 2012), QMEAN (Benkert *et al.*, 2008), Verify3D (Eisenberg *et al.*, 1997), Dfire (Yang and Zhou, 2008) and Naïve_PSIBLAST (Haas *et al.*, 2013), that participate in CAMEO-QE were used. After filtering cases where any method lacked predictions there were 395 targets, from 2574 models containing 642 694 residues.

3 Results

The ProQ2 version implemented in Rosetta participated in CASP11 with two methods, *ProQ2* and *ProQ2-refine*. *ProQ2* is only using the ProQ2 score, while *ProQ2-refine* does 10 side-chain repacks to calculate the optimal ProQ2 score given the current backbone. This has previously been shown to improve model selection by finding good backbones with sub-optimal side-chain packing (Wallner, 2014).

Based on the official CASP11 assessment (Kryshtafovych *et al.*, 2015) and from Table 1 it is clear that both ProQ2 versions are among the best if not the best single-model program to estimate global model accuracy, with ProQ2-refine being slightly better overall and on the easy targets, while VoroMQA and MULTICOM-NOVEL (Cao *et al.*, 2014) are better on the hard targets. It is interesting that in terms of model selection the pure clustering methods are much worse than single-model methods. In particular for hard targets where the Z-score for the pure clustering methods Pcons and DAVIS-QAconsensus is almost random ($\Sigma Z_{hard} = 3.9$ and $\Sigma Z_{hard} = 6.5$), compared to the much higher Z-score for the best single-model methods ($\Sigma Z_{hard} > 40$). For easy targets the clustering

Table 1. Model selection in CASP11

Method	ΣZ_{all}	ΣZ_{easy}	ΣZ_{hard}	#
MULTICOM-CONSTRUCT[1]	100.5	59.2	41.3	89
Wallner[2]	99.6	59.5	40.1	89
ProQ2-refine	**91.3**	**51.4**	**40.0**	**89**
Pcons-net[3]	90.8	59.5	31.4	89
VoroMQA	**89.6**	**43.1**	**46.4**	**89**
MULTICOM-NOVEL[1]	**89.1**	**45.2**	**43.9**	**89**
ProQ2	**86.2**	**50.7**	**35.6**	**89**
MULTICOM-CLUSTER[1]	**80.0**	**43.0**	**37.0**	**89**
RFMQA[4]	**74.8**	**37.9**	**36.9**	**88**
myprotein-me	73.3	40.3	33.0	88
nns[4]	63.1	45.4	17.6	89
ModFOLDclust2[5]	62.1	52.3	9.8	89
MUFOLD-Server[6]	60.7	55.4	5.2	89
Wang_SVM	**60.1**	**27.7**	**32.4**	**89**
*Pcons**	58.6	52.2	6.5	89
DAVIS-QAconsensus[7]	54.7	50.8	3.9	89

The summed Z-score for the first ranked model selected by each method, for all, easy and hard targets. Single-model methods in bold, pure clustering in italics, and the others are using a combination of single-model and clustering.

*This method was added after CASP as a reference. Refs: [1](Cao et al., 2014), [2](Wallner and Elofsson, 2005), [3](Wallner et al., 2007), [4](Manavalan et al., 2014), [5](McGuffin and Roche, 2010), [6](Q. Wang et al., 2011), [7](Kryshtafovych et al., 2015).
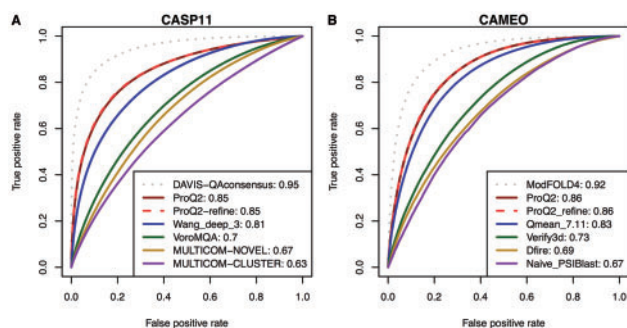
Fig. 1. Local model prediction accuracy for (**A**) CASP11 and (**B**) CAMEO for all single-model methods and one reference consensus method* shown as a dotted line using 3.8 Å as a cutoff for correct residues. Methods are ordered according to the AUC, which is also shown next to the method name. *ModFold4 in CAMEO runs in quasi-single mode, which means that the input is a single model and the consensus ensemble is constructed internally

methods still have an advantage over single-model method, but overall the only methods that are better than the best single-model methods are methods that combine the best single-model methods with clustering. MULTICOM-CONSTRUCT is using the MULTICOM single-model methods and Wallner is using ProQ2 together with Pcons, the latter improved the ΣZ_{all} from 58.6 to 99.6. This clearly shows that single-model methods are very useful in model selection in particular in combination with clustering methods.

The global quality prediction in ProQ2 is actually based on the predicted local error. The local error estimates from ProQ2 was recently combined with Phaser to enable molecular replacement for more targets and using poorer models (Bunkóczi *et al.*, 2015), clearly demonstrating the added value of local model error prediction for solving crystal structure by molecular replacement. The local model prediction accuracy for the CASP11 targets was assessed using ROC curves with a 3.8 Å cutoff for correct residues (Fig. 1A). Here, ProQ2 is clearly much better than all other single-model methods identifying 30% more correct residues compared to the second best, Wang_deep_3, at 10% FPR, and over twice as many for the methods that performed equally well or in some cases even better than ProQ2 in model selection. There is a similar trend for the CAMEO data (Fig. 1B), but the margins up to the reference consensus method ModFOLD4 (McGuffin *et al.*, 2013) and best other single-model method, Qmean_7.11 (Benkert *et al.*, 2008) are smaller. For both CASP11 and CAMEO data there are no noticeable difference between ProQ2 and ProQ2_refine on the local level.

To conclude, there is still a performance gap between the single-model methods and the consensus methods overall. However, the single-model methods are clearly much better in global model selection on hard targets and they are also useful in combination with consensus methods, demonstrated by the performance of the Wallner method, which combines Pcons and ProQ2. Finally, among the methods benchmarked in CASP11 and CAMEO, ProQ2 is currently the single-model method that performs best in both global and local model accuracy prediction.

Conflict of Interest: none declared.

References

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Benkert,P. *et al.* (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins*, **71**, 261–277.

Bunkóczi,G. *et al.* (2015) Local error estimates dramatically improve the utility of homology models for solving crystal structures by molecular replacement. *Structure*, **23**, 397–406.

Cao,R. *et al.* (2014) Designing and evaluating the MULTICOM protein local and global model quality prediction methods in the CASP10 experiment. *BMC Struct. Biol.*, **14**, 13.

Cheng,J. *et al.* (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.*, **33**, W72–W76.

Eisenberg,D. *et al.* (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, **277**, 396–404.

Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.

Haas,J. *et al.* (2013) The Protein Model Portal–a comprehensive resource for protein structure and model information. *Database (Oxford)*, **2013**, bat031.

Hubbard,S.J. and Thornton,J.M. (1993) *NACCESS - Computer Program*.

Joachims,T. (2002) Learning to Classify Text Using Support Vector Machines Kluwer, Massachusetts, USA.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kryshtafovych,A. *et al.* (2015) Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. *Proteins*, doi:10.1002/prot.24919.

Manavalan,B. *et al.* (2014) Random forest-based protein model quality assessment (RFMQA) using structural features and potential energy terms. *PLoS ONE*, **9**, e106542.

McGuffin,L.J. and Roche,D.B. (2010) Rapid model quality assessment for protein structure predictions using the comparison of multiple models without structural alignments. *Bioinformatics*, **26**, 182–188.

McGuffin,L.J. *et al.* (2013) The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res.*, **41**, W368–W372.

McGuffin,L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Ray,A. *et al.* (2012) Improved model quality assessment using ProQ2. *BMC Bioinformatics*, **13**, 224.

Wallner,B. (2014) ProQM-resample: improved model quality assessment for membrane proteins by limited conformational sampling. *Bioinformatics*, **30**, 2221–2223.

Wallner,B. and Elofsson,A. (2003) Can correct protein models be identified? *Protein Sci.*, **12**, 1073–1086.

Wallner,B. and Elofsson,A. (2006) Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.*, **15**, 900–913.

Wallner,B. and Elofsson,A. (2005) Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics*, **21**, 4248–4254.

Wallner,B. *et al.* (2007) Pcons.net: protein structure prediction meta server. *Nucleic Acids Res.*, **35**, W369–W374.

Wang,Q. *et al.* (2011) MUFOLD-WQA: A new selective consensus method for quality assessment in protein structure prediction. *Proteins*, **79 Suppl 10**, 185–195.

Wang,Z. *et al.* (2009) Evaluating the absolute quality of a single protein model using structural features and support vector machines. *Proteins*, **75**, 638–647.

Yang,Y. and Zhou,Y. (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins*, **72**, 793–803.

Zemla,A. *et al.* (1999) Processing and analysis of CASP3 protein structure predictions. *Proteins*, **Suppl 3**, 22–29.