Poster I-100

Meta-MQAP: an SVM-based meta-server for the quality assessment of protein models

Authors:
Marcin Pawlowski (*International Institute of Molecular and Cell Biology in Warsaw*)
Ryszard Matlak (*International Institute of Molecular and Cell Biology in Warsaw*)
Janusz Bujnicki (*International Institute of Molecular and Cell Biology in Warsaw*)

Short Abstract: Presented Meta-MQAP uses the results of the five primary MQAPs to predict the absolute deviation (in Angstrom) of C-alpha atoms of all residues in the model from their counterparts in the native structure, without the knowledge of the actual native structure.

Long Abstract:

Introduction:

Evaluation of model accuracy is one of the most essential steps in protein structure prediction. The existing methods for quality assessment of protein models (MQAPs) are usually based either on the physical effective energy which can be obtained from fundamental analysis of particles forces or on the pseudo energy derived from known protein structures (review: (Lazaridis and Karplus, 2000). So far, the development of MQAPs was focused on the global evaluation of protein structure and most of the existing methods were optimized to discriminate between globally correct and incorrect “decoy” structures rather than detection of correct and incorrect fragments (review: (Gilis, 2004)). Even among MQAPs that are capable of generating independent evaluations for each amino acid in the protein structure, it is usually recommended that scores are averaged over long stretches of residues (e.g. 21 aa in the case of VERIFY3D). Systematic assessment experiments, e.g. Critical Assessment of techniques for protein Structure Prediction (CASP) and LiveBench demonstrated that models with correct fold can be confidently recognized, especially by the fold-recognition meta-servers (Bujnicki et al., 2001; Lundstrom et al., 2001). However, comparative models, especially those based on remotely related templates, often exhibit local inaccuracies that are difficult to identify by global evaluation, in particular misthreadings of short regions (5-10 residues) corresponding to shifted alignments within individual secondary structure elements (Tramontano and Morea, 2003; Tress et al., 2005).

We proposed that inaccuracies due to local alignment shifts can be identified and corrected by identification of variable conformations in alternative homology models, comparison of their VERIFY3D scores averaged over only 5 neighboring residues, and construction of hybrid models comprising the best-scoring fragments (Kosinski et al., 2003). Our method (termed the “FRankenstein’s monster approach”) turned out to consistently produce very accurate models, especially if regions with initially poor scores were systematically varied to generate additional models for evaluation (Kosinski et al., 2005). However, detailed inspection of cases where we failed to identify the most native-like local conformation based on the VERIFY3D score revealed a considerable variation of scores even among models with similar structural features. Therefore, we decided to carry out a systematic evaluation of the capability of VERIFY3D and several other popular MQAPs to identify the best method for

prediction of local accuracy of protein models. However, as the work progressed, we realized that none of the MQAPs we analyzed was sufficiently accurate and robust, and that they all had different strengths and weaknesses. This in turn prompted us to develop a new “meta-predictor” optimized specifically to detect local errors.

Results:
Five MQAPs (VERIFY3D, PROSA, BALA, ANOLEA, PROVE) were tested on over 3000 approximately correct models from the latest CASP-6 modeling assessment experiment. We found that the absolute values of correlation coefficients between the scores returned by these methods and the deviation of individual amino acids were rather modest. Thus, we developed a meta-predictor based on a support vector machine (SVM) approach. Our Meta-MQAP uses the results of the five primary MQAPs to predict the absolute deviation (in Angstrom) of C-alpha atoms of all residues in the model from their counterparts in the native structure, without the knowledge of the actual native structure. Scores returned by our Meta-MQAP predict much better the local accuracy of models than the scores returned by the best of the ‘primary’ MQAPs.

Availability:
We implemented the Meta-MQAP as a web server available for free use by all academic users at the URL https://genesilico.pl/toolkit/