Quality estimates for 3D protein models

Metadata of the chapter that will be visualized online

Chapter Title	Quality Estimates for 3D Protein Models	
Copyright Year	2023	
Copyright Holder	Springer Science+Business Media, LLC, part of Springer Nature	
Author	Family Name	**Maghrabi**
	Particle	
	Given Name	**Ali H. A.**
	Suffix	
	Organization	College of Applied Sciences, Umm Al Qura University
	Address	Mecca, Saudi Arabia
Author	Family Name	**Aldowsari**
	Particle	
	Given Name	**Fahd M. F.**
	Suffix	
	Organization	School of Biological Sciences, University of Reading
	Address	Reading, UK
Corresponding Author	Family Name	**McGuffin**
	Particle	
	Given Name	**Liam J.**
	Suffix	
	Organization	School of Biological Sciences, University of Reading
	Address	Reading, UK
	Email	l.j.mcguffin@reading.ac.uk
Abstract	Protein structure modeling is one of the most advanced and complex processes in computational biology. One of the major problems for the protein structure prediction field has been how to estimate the accuracy of the predicted 3D models, on both a local and global level, in the absence of known structures. We must be able to accurately measure the confidence that we have in the quality predicted 3D models of proteins for them to become widely adopted by the general bioscience community. To address this major issue, it was necessary to develop new model quality assessment (MQA) methods and integrate them into our pipelines for building 3D protein models. Our MQA method, called ModFOLD, has been ranked as one of the most accurate MQA tools in independent blind evaluations. This chapter discusses model quality assessment in the protein modeling field, demonstrating both its strengths and limitations. We also present some of the best methods according to independent benchmarking data, which has been gathered in recent years.	
Keywords (separated by '-')	Protein structure prediction - Model quality assessment - Estimates of model accuracy - Accuracy self-estimates - Template-based modeling - Free modeling	

Chapter 6

Quality Estimates for 3D Protein Models

Ali H. A. Maghrabi, Fahd M. F. Aldowsari, and Liam J. McGuffin

Abstract

Protein structure modeling is one of the most advanced and complex processes in computational biology. One of the major problems for the protein structure prediction field has been how to estimate the accuracy of the predicted 3D models, on both a local and global level, in the absence of known structures. We must be able to accurately measure the confidence that we have in the quality predicted 3D models of proteins for them to become widely adopted by the general bioscience community. To address this major issue, it was necessary to develop new model quality assessment (MQA) methods and integrate them into our pipelines for building 3D protein models. Our MQA method, called ModFOLD, has been ranked as one of the most accurate MQA tools in independent blind evaluations. This chapter discusses model quality assessment in the protein modeling field, demonstrating both its strengths and limitations. We also present some of the best methods according to independent benchmarking data, which has been gathered in recent years.

Key words Protein structure prediction, Model quality assessment, Estimates of model accuracy, Accuracy self-estimates, Template-based modeling, Free modeling

1 Introduction

Understanding protein function is one of the keys for understanding life at the molecular level. Each protein molecule has its own unique sequence, which consist of linear chains of amino acids. These amino acid chains fold to form tertiary structures, which confer the proteins function. In other words, characterizing protein structures leads to the ability to better understand their functions. Experimental methods such as X-ray crystallography and nuclear magnetic resonance have been considered as the methods of choice for 3D structure determination. However, such methods are costly and time-consuming, and some proteins are also problematic or impossible to be characterized using these methods. Consequently, the process of growing protein structure data is relatively slow in comparison to the speed of sequencing genomes and their encoded proteins, which has kept increasing, especially after breakthroughs in the genetic sequencing technology. As a result, a gap has grown

Sławomir Filipek (ed.), *Homology Modeling: Methods and Protocols*, Methods in Molecular Biology, vol. 2627,
https://doi.org/10.1007/978-1-0716-2974-1_6,

between known protein sequences and their resolved structures, and it has been necessary to find another solution.

Computational methods, which predict the structure of proteins directly from their own sequences, have become fast and effective alternatives to experimental methods. Over the past 20 years, there has been an emergence of different types of protein structure prediction methods, the most accurate type being the comparative modeling method, which consists of a number of steps including template recognition, alignment, quality assessment, and ending with refinement. Each of these steps plays an essential part in order to achieve a successful modeling pipeline, but perhaps the most critical step for the wider acceptance of 3D models of proteins has been the protein modeling quality assessment pipeline. In this step, the predicted models are evaluated in terms of their likely accuracy without the need of an experimental structure. Numerous challenges were identified and many approaches to the quality estimation problem have been developed over the years including the use of statistical potentials, stereochemistry checks, and machine learning techniques. Such methods have traditionally been referred to as the model quality assessment (MQA or QA) methods, and they have been evaluated in successive critical assessment of structure prediction (CASP) experiments under the estimates of model accuracy (EMA) and the accuracy self-estimates (ASE) prediction categories.

2 Estimates of Model Accuracy (EMA) Are Essential for Template-Based Modeling (TBM) and Template-Free Modeling (FM)

The fact that evolutionarily related proteins have similar structures has encouraged researchers to develop methods for predicting the structure of proteins from their sequences [1]. One way of modeling a protein structure is by aligning the sequence to those of already experimentally observed protein structures and then using those structures as templates in order to map the 3D coordinates of each aligned residue. This procedure has been termed as homology modeling or comparative modeling [2]. However, sometimes structurally homologous proteins can have a very low sequence identity, and in these cases homology modeling methods fail to identify suitable template structures or produce poor alignments. This issue led to another way of determining protein structure called threading or fold recognition [3]. This modeling method does not use the homologous proteins with known structures but rather uses statistical knowledge of the relationship between the structures, which have been deposited in the PDB database and the targeted sequence. Both approaches have been improved over the years along with the integration of EMA programs, and systematic differences were noticed.

In recent years, fold recognition and homology modeling techniques have somewhat merged with the ability to detect ever more distant evolutionary relationships using profile-profile searching methods and HMM-HMM methods, such as the popular HHpred method [4]. The general concept of modeling based on existing structures is now classified as template-based modeling (TBM), and the success of such methods relies on the availability and accurate detection of suitable templates. As the amount of detectable similarity between target protein and template structures decreases, the accuracy of template-based techniques starts to be insufficient and such methods become unreliable. In this case, another structure prediction technique, traditionally called de novo or ab initio protein structure prediction, is the only remaining option. The technique is based on predicting the structure of proteins without the need of a template and is therefore known as template-free modeling or FM [5]. FM methods are not nearly as accurate as TBM methods when templates are available [6]. However, the concept of such techniques is fairly simpler comprising of only two elements: firstly, an algorithm to search the space of possible protein configurations for cost function minimization; secondly, various restraints, which are the composition of the cost function itself, being either derived from physical laws and structural features predicted by machine learning or other types of statistical systems [7]. FM techniques have been incrementally improving and can provide us with valuable information on how novel domains may fold [8].

Regardless of whether TBM or FM approaches are used to model a protein target, a researcher will often end up with dozens, or even hundreds, of alternative models for the same protein target. The first problem they will then face is how to select the best model from among the alternatives and then, once selected, they will need to know how confident they can be in the model accuracy overall and, more specifically, which local regions of the model can be trusted; EMA methods are critical for answering all of these questions.

3 Methods for Estimates of Model Accuracy

Traditionally, protein structure modeling has been far less trusted in terms of accuracy than deriving protein structures from experiments. Models are typically left unannotated with quality estimates and can span a broad range of the accuracy spectrum, whereas the accuracy of observed protein structures can be estimated from experiments and falls within a narrow range [2]. Therefore, a number of quality evaluation methods have been developed by modelers using techniques such as statistical potentials, molecular mechanics energy-based functions, stereochemistry checks, and

machine learning in order to analyze the correctness of protein structures and models [9]. Examples of the early/simple quality checking tools include WHAT-CHECK [10], PROCHECK [11], and, more recently, MolProbity [12]. These tools use basic stereochemical checks, and they are very useful in identifying unusual geometric features in a model. However, such quality checking tools are not able to produce a single score for ranking alternative models. Other examples of early quality assessment tools that use a variety of different methods include ProSA [13] and DFIRE [14], which have been used along with VERIFY3D [15] in order to provide single scores that relate to the global quality of protein models. Machine learning-based quality assessment programs have also been utilized to provide a higher value of prediction accuracy. ProQ [16], ModFOLD method [17], and QMEAN [18] are examples of early machine learning-based QA methods, which helped programmers to use various combinations of structural features and individual energy potentials in order to predict the accuracy of global model quality.

4 Observed Model Accuracy Scoring

In order to evaluate predicted model quality scores, in the early years of structure prediction, the predicted models were compared with the superposed observed structures simply by using the root-mean-square deviation (RMSD). To overcome some of the RMSD limitations, EMA developers started to use improved similarity scoring measures such as GDT-HA and GDT [19], MaxSub [20], TM-score [21] (which are superposition based), and local Distance Difference Test (lDDT) (which does not require superposition of the model and observed structures) [22]. These scores were used to measure the predicted model quality for each individual model by comparing them to the observed native (solved experimental) structures. The term GDT stands for "global distance test," in both the GDT and GDT-HA scores. These two scores represent the measurement of similarity between two protein structures that both have identical amino acid sequences but may have different tertiary structures, i.e., a predicted model and the observed crystal structure [21]. The difference between GDT and GDT-HA is that GDT-HA is "high accuracy" and uses smaller cutoff distances, which makes it more rigorous and, as a result, is more stringent than GDT [23]. MaxSub is a measure that identifies in a model the largest subset of Cα atoms that superimpose over the experimental structure, producing a single normalized score that represents the quality of that model. The TM-score stands for "template modeling" score. Likewise, this measure is for calculating the similarity between two models with the same sequence but with different tertiary structure. The TM-score is arguably more accurate than

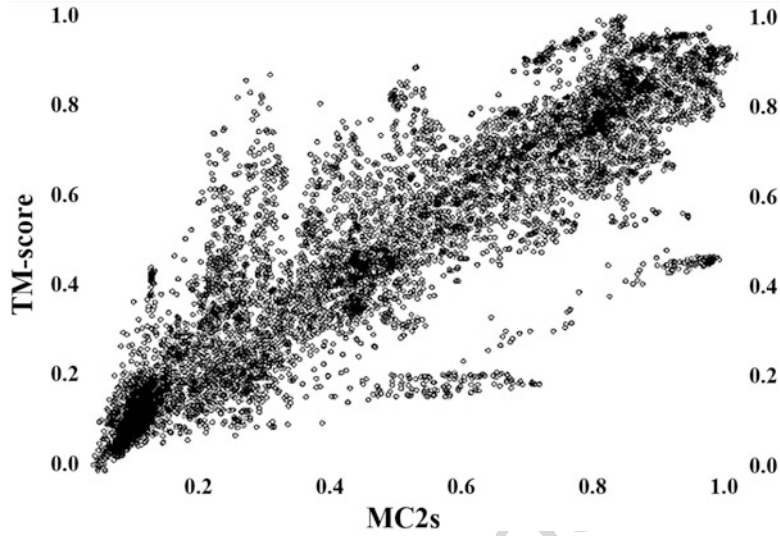

Fig. 1 Predicted model quality scores versus observed model quality scores. The plot compares the predicted scores for one of the top-performing individual EMA methods, ModFOLDclust2 (Mc2s), against TM-score observed scores. (The data set was collected from CASP12)

GDT and GDT-HA in comparing the similarity of structures with full-length protein chains rather than domains [21]. Each of these measures indicate the difference between two protein structures (predicted versus observed) by providing a score between 0 and 1, where 1 is a perfect match between the two compared structures (i.e., identical relative atom coordinates) and 0 is a nonmatched structure [20]. The comparison between the predicted and observed scores of each region of the protein structure is compared using the pairwise correlation technique, an example of this type of correlation can be seen in Fig. 1. Such superposition-based scoring measurement may have some limitations as they are affected by differences in the relative orientation of domains following global superposition in structures with more than a domain. This can lead to, for example, poor scores given for correct small domains because the largest domain will be dominating the global rigid-body superposition. The local Distance Difference Test (lDDT) scoring is independent of superposition, so it does not have the same issues when scoring multiple domains with different relative orientations. A variety of observed model accuracy scoring methods are used as the target functions in order to train and benchmark EMA methods over the years. Practically, the GDT score and the lDDT score have been used more recently due to their adoption as the gold standards for the CASP and CAMEO experiments, respectively [24].

5 EMA Classification

The field of computational protein structure prediction is evolving constantly, following the increase in computational power of machines and the development of intelligent algorithms. Despite the rapid development of methods and fusion of applied approaches, a broad classification and categorization of these methods can be made. Numerous methods have been developed over the years in an attempt to provide users with scores that will give them confidence in their 3D models and allow them to identify any potentially suspect regions. As previously mentioned, the model quality assessment field has its roots in early structure validation tools [10, 11, 25]. While such tools can be used to perform basic stereochemical checks and identify unusual geometric features in a model, they are not able to produce a single global score that can be used for ranking alternative models nor can they be relied upon for discriminating good models from bad (often bad models will still have good stereochemistry). Modern methods for EMA can be classified into three broad categories in terms of input: pure-single-model methods [14, 15, 17, 18, 25–28] which consider only information within an individual model, clustering/consensus approaches [29–33] which can only be used if you have multiple alternative models built for the same protein target, and quasi-single-model methods [34, 35] which can score an individual model against a pool of alternative models generated from the target sequence. Each approach has its advantages and disadvantages. Clustering methods have been far more accurate than pure-single-model methods, but they are more computationally intensive and do not work when very few similar models are available, which is often the case in real-life research scenarios. Pure-single-model methods are less accurate overall, but they are more rapid, they produce consistent scores for single or few models at a time, and they often perform better at model ranking and selection. Quasi-single-model methods attempt to provide comparable accuracy to clustering methods while addressing real-life needs of researchers with few/single models.

Moreover, there are several other factors that EMA methods can be categorized with, such as the predicted property, target function, machine learning method, and other features. Table 1 contains a list of some of the most popular programs and servers for EMA, which have been independently evaluated in the CASP [36] and CAMEO [37] experiments.

Table 1
Examples of different EMA methods used in CASP13

Method	Local/global	Inputs	Structure features	Predicted features	Target function	Machine learning method
FaeNNz [38]	Local (global is avg. local)	Model and full-length target sequence	Statistical potentials of mean force + distance constraints from templates + solvent acc.	Sec. str and surface area	LDDT (local)	Multilayer perceptron
ModFOLD7 [39]	Local (global is sum of local)	Model and full-length target sequence	Pairwise comparisons of generated reference models, residue contacts	Contacts, sec. str and disorder	S-score (local)	Multilayer perceptron
ProQ3 [26]	Local (global is sum of local)	Profile + model + predictions + energies	ProQ2 + energy terms	Sec. str and surface area	S-score (local)	Linear SVM
VoroMQA-A [40]	Local and global	Model	Voronoi tessellation-based contact areas	Not used	Not used	Statistical potential
MULTICOM-CLUSTER [41]	Global	Model and full-length sequence	Secondary structure, solvent accessibility, residue contacts	Contacts, sec. str, surface area, and structural scores	GDT_TS (global)	Deep network + ensemble

The methods have been chosen randomly taking into consideration the differences between them with regard to their measuring method (local/global), inputs, structure features, predicted features, target function, and machine learning method

6 ModFOLD: A Leading EMA Web Server

One of the top leading EMA methods is ModFOLD, which has been developed by Prof. Liam McGuffin and colleagues [17]. Since its inception, ModFOLD has been continuously improved, going through many upgrades until its latest version, ModFOLD8 [38].

6.1 ModFOLD History

In the 2 years following CASP7, performances of protein structural QA servers were observed to be considerably increasing. Model quality assessment programs, or MQAPs, have become the

cornerstone of many protein structure modeling methods. More than a dozen papers were published in the area of QA between CASP7 and CASP8, and 45 methods were submitted for evaluation to CASP8 in that category.

6.1.1 The Initial Construction of ModFOLD

ModFOLD is a machine learning-based QA program which was developed at the University of Reading by the McGuffin group. The original ModFOLD method was developed based on the nFOLD protocol [39], which was a combination of the new GenTHREADER protocol [40] and a number of extra inputs into the underlying neural network, including the SSEA score [41], a new functional site detection score (MetSite) [42], and a simple model quality checking algorithm, MODCHECK [40].

Initially, ModFOLD was developed in two editions: ModFOLD, designed to be fast and used for the global assessment of either single or multiple models, and ModFOLDclust, a more intensive method that carries out clustering of multiple models and provides a per-residue local quality assessment. ModFOLDclust was shown to significantly outperform all of its clustering/multiple MQAP competitors, while ModFOLD has competed well against some of the best "true" single-model MQAP methods [17]. Since CASP ranking relies on the prediction accuracy regardless of the method used, clustering- or consensus-based MQAPs were ranked as the most accurate methods for predicting 3D model quality, outperforming the single-model methods.

6.1.2 ModFOLDclustQ for Speed, Accuracy, and Consistency

Despite their accuracy, it was noticed that a number of advantages of the single-model-based methods were missing in the clustering methods. One missing feature was the speed. Like Pcons and other consensus-based approaches, ModFOLDclust carries out pairwise comparisons of numerous models by using multiple structural alignments, and that makes it often CPU intensive [28]. Another difficulty found in QA programs including ModFOLDclust was the requirement of a large pool of diverse models, and thus, smaller numbers of models can minimize the accuracy. To overcome such problems, McGuffin and Roche designed an upgraded version of the same method, called ModFOLDclustQ [33]. The initial "Q" labeled in the upgraded version name is referred to a score called Q-score; this score was utilized in ModFOLDclustQ while also standing for "Quick." The Q-score is derived from the Q measure that was developed by the Wolynes group [43]. The Q-score has the ability to efficiently estimate structural relations between two proteins based on their residue distances. This method has been suggested by the CASP8 assessors as an alternative to the other scoring methods such as the GDT-TS [44]. By importing Q-score, ModFOLDclustQ was shown to compete with the leading consensus MQAPs. Furthermore, when taking the mean of ModFOLDclustQ score and the older ModFOLDclust score, a significant

increase in prediction accuracy was achieved, with little computational overhead. That led McGuffin and Roche to combine both scoring methods to form a new method named ModFOLDclust2 [35]. There are a number of other MQAPs that also used Q-score to assess each individual residue in a model pertaining to the per-residue accuracy. A successful per-residue consensus-based method was Pcons method, which was superseded then by one of the leading consensus single-model per-residue programs, known as ProQ [45]. The method was then upgraded by updating its structural and predicted features, this upgrade to be as the second top ranking MQAP, ProQ2 [46].

Although upgrading ModFOLDclust to ModFOLDclustQ and combining their scores showed a high improvement in the quality assessment speed and accuracy level, McGuffin also noticed the potential of using ModFOLDclust2 to guide 3D modeling using multiple templates. In the process of modeling, using more than onefold template is helpful in assessing models more accurately. However, it was noticed that such a technique is not preferable in many cases as it may result in poorer model quality. Besides the speed and the accuracy of an MQAP, there has to be consistency as well. To solve such a problem, McGuffin and colleagues have started to investigate the use of local as well as global model quality prediction scores that are produced by ModFOLDclust2. This led to improvements in the selection of target-template alignments for the construction of multiple-template models. After the investigation, it was found that the most accurate and consistent way in improving models is to use accurate local model quality scores to guide alignment selection while using accurate global model quality before selection for re-ranking alignments. Applying this technique has made significant performance improvements to the tertiary structure prediction IntFOLD server [47].

6.1.3 The Quasi-Single-Model Approach

Another important feature that was missing in the clustering-based approaches was addressing the real-life needs of protein researchers, when often only a single or few models for each protein target are available for evaluation. In such cases, clustering methods will provide poor performance. McGuffin's research group was aware of this problem and they found a way to address it. Instead of proceeding with a direct clustering to the submitted model/s, a tertiary structure prediction method [48] was used at the beginning as the first stage of the quality assessment procedure, in order to generate an initial reference set of template-based models. The user-submitted model/s are then pooled with the generated models and clustered using ModFOLDclust2 as the second stage of the process. By integrating this algorithm, if the server received multiple models, then the procedure will go with the full clustering approach, whereas if only single or few models are submitted,

then the pipeline will be diverted to the so-called quasi-single-model approach which operates with comparable accuracy. This method was implemented initially with the ModFOLD v3.0: a server developed using ModFOLDclust2 integrated with the IntFOLD-QA tertiary structure prediction pipeline [33]. The algorithm has since been independently tested for confidence and published as the fourth version of the ModFOLD server [34].

CASP assessments of QA methods were more concerned about the quality scoring results rather than other practical considerations, such as the researcher's accessibility, until the assessment was updated following the eighth and ninth seasons of the experiment [49] (details about CASP in Subheading 7). In CASP10, the criteria were modified to rebalance the quality assessment. This modification was implied by using smaller bespoke data sets rather than allowing large sets of models, which some participants argued had unfairly favored clustering approaches in previous CASPs. Despite this change of criteria in CASP10, ModFOLD4 was ranked among the top-performing methods in the quality assessment category. ModFOLD4 also provided a free service for accurate prediction of global and local QA of 3D protein models. The server had a comparable performance to clustering-based methods but retained the capability of making predictions for a single model at a time [34].

6.2 Latest Versions of ModFOLD

In 2015, the fifth version of ModFOLD was released. This version was integrated with the upgraded tertiary structure prediction IntFOLD3-TS pipeline which gave ModFOLD5 the ability to generate a greater number and variety of reference models [50]. In 2017, ModFOLD was upgraded to its sixth version with a new neural network-based quasi-single-model method that took as its input a sliding window of per-residue scores from six different pure-single and quasi-single scoring methods and a single quality score for each residue in the model [51]. ModFOLD6 was independently evaluated during the CASP12 experiment and it is freely available at https://www.reading.ac.uk/bioinf/ModFOLD/ModFOLD6_form.html (Fig. 2). During the past 2 years, ModFOLD had further improvements and was upgraded to the seventh and eighth versions, which were tested in CASP13 and CASP14, respectively. More details about the ModFOLD server interface and inputs and outputs can be found in Maghrabi and McGuffin 2017 [51].

7 EMA in Community-Wide Experiments

EMA and a few of several other modeling techniques have been developed and utilized through the last decades in order to solve the protein sequence-structure gap dilemma. The methods and servers were included for evaluation as a category in two major

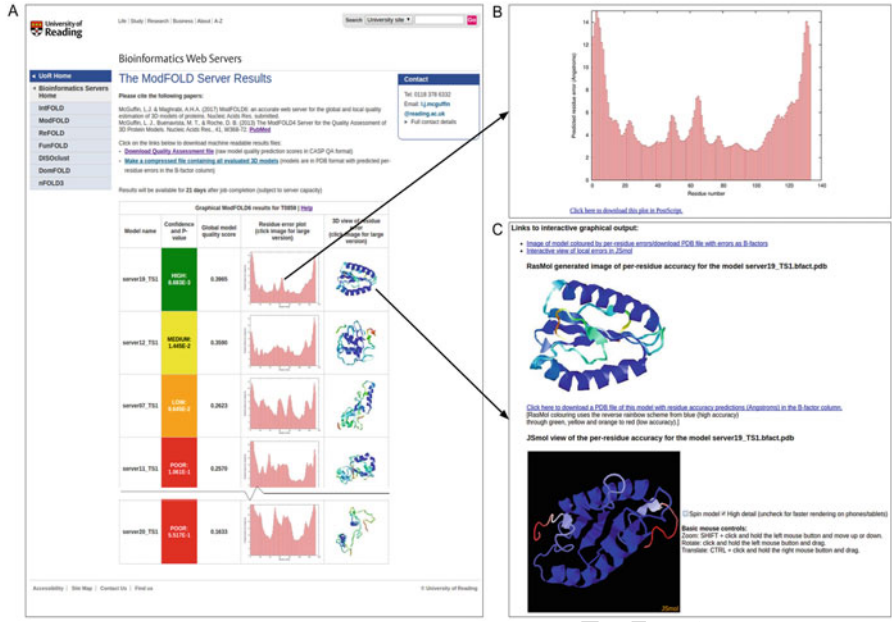

Fig. 2 ModFOLD6 server results for models submitted to CASP12 generated for target T0859 (PDB ID: 5jzr). (**a**) An example of the graphical output from the server showing the main results page with a summary of the results from each method (truncated here to fit page). Clicking on the thumbnail images in the main table allows results to be visualized in more detail. (**b**) A histogram of the local or per-residue errors for the top-ranked model, with the residue number on the x-axis and the predicted residue error (distance of the Cα atom from the native structure in Å) on the y-axis, which may be downloaded. (**c**) Interactive views of models, which can be manipulated in 3D using the JSmol/HTML5 framework and/or downloaded for local viewing. (Adapted from Maghrabi and McGuffin [51])

worldwide organizations that are specialized in the protein structure prediction field. The first organization conducts independent blind testing with the Critical Assessment of Techniques for Protein Structure Prediction (CASP) [36] experiments, which are held every other year. The second organization is the continuously automatic model evaluation project called CAMEO [37]. Both organizations have highlighted the importance of the EMA development for the improvement of protein structure prediction and have helped to encourage progress in the field.

The importance and far-reaching implications of having the ability to predict protein structures from their amino acid is manifested by the ongoing biennial experiment on "Critical Assessment of Structure Prediction" (CASP). The Critical Assessment of Techniques for Protein Structure Prediction or CASP is a global community-wide experiment that has started taking place every other year since 1994 [52]. Protein structure modelers in more than a hundred research centers around the world dedicate their late spring and summer to preparing their methods to be independently tested in this center. CASP is designed as a blind prediction

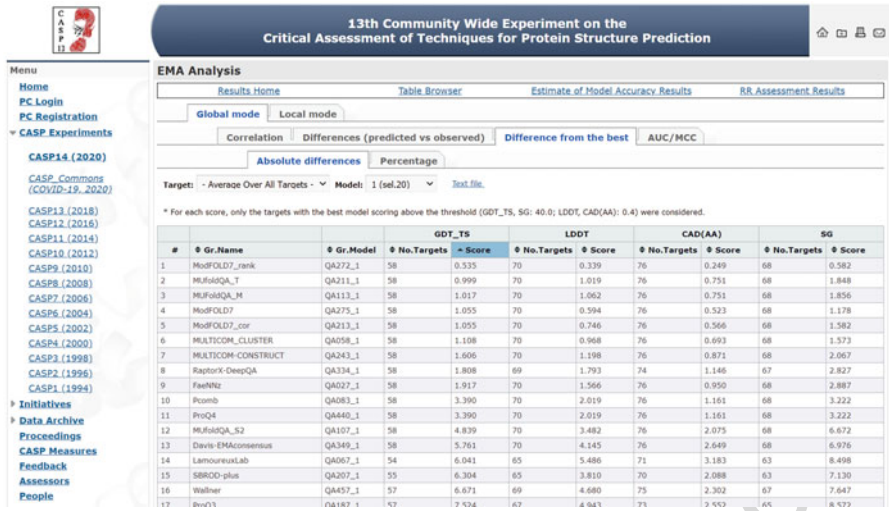

Fig. 3 EMA ranking section in the CASP community web page. Results from CASP13 showing the top ranking EMA based on stage 1 which consists of 20 models, and the scores were ranked against the observed scores from GDT_TS. https:/www.predictioncenter.org/

experiment (Fig. 3). A set of protein sequences are selected by the assessors in order to test the performance of the methods in predicting their protein structures which are already experimentally observed and hidden with the organizers, for an attempt to help advancing these protein prediction methods. In the first CASP, the experiment was quite basic consisting of just three parts: collecting protein targets (which will subsequently be solved experimentally), collecting tertiary structure predictions, and assessing and discussing the results [52]. CASP experiment has since become popular, and its participants and prediction categories have been growing over the years. CASP takes the form of an international competition, which can be thought of as the world championships for protein structure prediction. Fourteen CASP experiments have been performed during the last 25 years, with the last one completed in late 2020. The competition has evolved over the years and is now carried out by dividing its experiments into slightly more complicated subcategories, including the following: tertiary structure prediction; disorder prediction; contact prediction; model quality assessment or QA, which is also called estimates of model accuracy (EMA); binding site prediction; protein-protein interactions; oligomerization state; and protein model refinement [53]. Each category represents an important part of the structure prediction process that needs further improvements in terms of the predictive power of the underlying algorithms. An aim of CASP is to drive new developments, which will lead to higher levels of accuracy and consistency in producing models that are closer in quality to the experimentally derived protein structures.

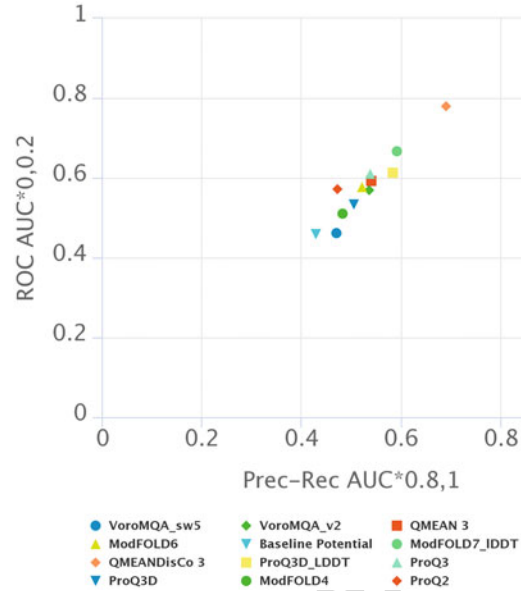

Fig. 4 CAMEO continuous benchmarking for EMA servers. A 6-month result of the top EMA methods being benchmarked continuously in CAMEO servers. https:/www.cameo3d.org/

Another evaluation resource for EMA methods is the CAMEO project, where the methods are continuously automatically evaluated each week, with tables and plots produced that show if there are any significant improvements between competitors (Fig. 4). Every week, CAMEO publishes benchmarking results based on models collected during a 4-day prediction window by assessing an average of a hundred targets during a time frame of 1 week, 1 month, 3 months, 6 months, and 1 year. The server benchmarks the most popular and top-ranked protein prediction methods as well as EMA methods separately.

The benchmarking data is generated consistently for all participants at the same time, enabling them to benchmark and cross-validate the performance of their methods. CAMEO sends emails with submission statistics and low performance warnings weekly in order to facilitate server development and promote shorter release cycles. This server has become a compliment to many participants of CASP and helped them when preparing their methods for upcoming community experiments [54, 55].

8 Recent Advances in EMA Methods

Most recent breakthroughs have arisen with the onset of deep learning. New approaches built using artificial intelligence (AI) have been accelerating the structure prediction field by far. A

method called AlphaFold [56], developed by the DeepMind AI company, has shown significant progress on generating 3D models of proteins in the worldwide protein prediction competition, CASP. The method was placed first in rankings among the teams that entered in protein modeling competitions. The reason behind this success lies in the integration of the deep neural networks (DNNs) approach, which is a system of neural layers trained to accurately predict the distances between residues, and as a result, it generates highly accurate structures. Such a success has drawn the attention from all structural biologists to start studying this field in depth [56].

Unlike standard neural networks (NNs), the multiple layers in the DNNs give it the ability to process more complicated problems [57]. By testing the visual pattern recognition example using DNNs, the neurons in the first layer could recognize edges, and then the neurons in the second layer would learn to recognize more shapes like triangles or rectangles which are built up from edges which already have been learnt in the first layer. The third layer could then recognize static more complex shapes, and the fourth learns animatic shapes, and so on. This reminds us of how children start to learn basic shapes around them when their brains that contain multiple layers of neurons give them a compelling advantage in starting to learn complex patterns. We can expect that having more hidden layers would make our networks more powerful. However, changing a single layer to multiple-layered neural networks could lead to having more complex intermediate layers which can have multiple layers of abstraction [58].

DNNs can compute more advanced problems with several techniques and architectures to be formulated; the multilayer perceptron (MLP) has been the chosen feedforward neural network class that has the ability to map a set of inputs which pass it through hidden layers and send the calculated data to an output unit [59]. MLP networks have been considered as a powerful technique in a large number of applications from different fields of research. The benefits of MLPs come from the appropriateness in dealing with most of the problems involving function approximation, pattern classification, process control, and time series forecasting [60]. MLPs have been used in many successful EMA methods (Table 1) and they have grown in complexity to accommodate the growth in input data.

Recent studies have shown that up until CASP14, there has been a small but significant improvement in EMA methods. It was noted that many of the improved methods have used deep learning but in various ways. However, the indications for such an implementation remain vague and are still under evaluation. The best way to use machine learning for EMA is still not functionally available, and there is plenty of space for developers to work on this area for improvements. We notice that on average the best

EMA methods select models that are better than those provided by the best individual TBM- or FM-based server. However, still, further significant improvements could be achieved if there were possible ways to always select the best model for each target. Finally, we do notice systematic differences when using different model evaluation methods. Single-model methods perform relatively better when using local evaluation methods and appear better at ranking higher-quality models [61].

Assessing the quality of protein structure prediction has been continuously improving over the last decades. Variant types of methods were developed for different tasks in the estimates of model accuracy sector, and the most succeeding ones were the pure-single, quasi-single, and the clustering methods which have shown significant results in controlling the prediction quality in CASP and CAMEO. Recently, around 50 EMA methods participated in CASP13 showing an increase in the number compared to the previous season. The recent concern which was focused on for EMA development was having more FM targets for which high-quality models were generated by the TS servers. Another concern was having higher consensus among high-quality models on average than ever before [62]. There are also some challenges that need to be overcome such as improving the way EMA methods are trained and the integration of deep learning tools for having more accurate prediction checking.

References

1. Kaczanowski S, Zielenkiewicz P (2010) Why similar protein sequences encode similar three-dimensional structures? Theor Chem Accounts 125:643–650. https://doi.org/10.1007/s00214-009-0656-3
2. Martí-Renom MA, Stuart AC, Fiser A et al (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325. https://doi.org/10.1146/annurev.biophys.29.1.291
3. Rost B, Schneider R, Sander C (1997) Protein fold recognition by prediction-based threading. J Mol Biol 270:471–480. https://doi.org/10.1006/jmbi.1997.1101
4. Hildebrand A, Remmert M, Biegert A, Söding J (2009) Fast and accurate automatic structure prediction with HHpred. Proteins 77(Suppl 9):128–132. https://doi.org/10.1002/prot.22499
5. Jothi A (2012) Principles, challenges and advances in ab initio protein structure prediction. Protein Pept Lett 19:1194–1204
6. Moult J, Fidelis K, Rost B et al (2005) Critical assessment of methods of protein structure prediction (CASP)–round 6. Proteins 61 (Suppl 7):3–7. https://doi.org/10.1002/prot.20716
7. Larrañaga P, Calvo B, Santana R et al (2006) Machine learning in bioinformatics. Brief Bioinform 7:86–112
8. Dhingra S, Sowdhamini R, Cadet F, Offmann B (2020) A glance into the evolution of template-free protein structure prediction methodologies. Biochimie 175:85–92. https://doi.org/10.1016/j.biochi.2020.04.026
9. Kryshtafovych A, Fidelis K (2009) Protein structure prediction and model quality assessment. Drug Discov Today 14:386–393. https://doi.org/10.1016/j.drudis.2008.11.010
10. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272–272. https://doi.org/10.1038/381272a0
11. Laskowski RA, Rullmann JAC, MacArthur MW et al (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol

NMR 8:477–486. https://doi.org/10.1007/BF00228148

12. Lovell SC, Davis IW, Arendall WB et al (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450. https://doi.org/10.1002/prot.10286
13. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins Struct Funct Bioinform 17:355–362. https://doi.org/10.1002/prot.340170404
14. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci Publ Protein Soc 11:2714–2726. https://doi.org/10.1110/ps.0217002
15. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404
16. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci Publ Protein Soc 12:1073–1086. https://doi.org/10.1110/ps.0236803
17. McGuffin LJ (2007) Benchmarking consensus model quality assessment for protein fold recognition. BMC Bioinform 8:345. https://doi.org/10.1186/1471-2105-8-345
18. Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. Proteins Struct Funct Bioinform 71:261–277. https://doi.org/10.1002/prot.21715
19. Zemla A, Venclovas Č, Moult J, Fidelis K (1999) Processing and analysis of CASP3 protein structure predictions. Proteins Struct Funct Bioinform 37:22–29. https://doi.org/10.1002/(SICI)1097-0134(1999)37:3+<22::AID-PROT5>3.0.CO;2-W
20. Siew N, Elofsson A, Rychlewski L, Fischer D (2000) MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics (Oxford, England) 16:776–785. https://doi.org/10.1093/bioinformatics/16.9.776
21. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins Struct Funct Bioinform 57:702–710. https://doi.org/10.1002/prot.20264
22. Mariani V, Biasini M, Barbato A, Schwede T (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics (Oxford, England) 29:2722–2728. https://doi.org/10.1093/bioinformatics/btt473
23. Read RJ, Chavali G (2007) Assessment of CASP7 predictions in the high accuracy template-based modeling category. Proteins Struct Funct Bioinform 69:27–37. https://doi.org/10.1002/prot.21662
24. Huang YJ, Mao B, Aramini JM, Montelione GT (2014) Assessment of template-based protein structure predictions in CASP10. Proteins Struct Funct Bioinform 82:43–56. https://doi.org/10.1002/prot.24488
25. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410. https://doi.org/10.1093/nar/gkm290
26. Uziela K, Wallner B (2016) ProQ2: estimation of model accuracy implemented in Rosetta. Bioinformatics (Oxford, England) 32:1411–1413. https://doi.org/10.1093/bioinformatics/btv767
27. Uziela K, Menéndez Hurtado D, Shu N et al (2017) ProQ3D: improved model quality assessments using deep learning. Bioinformatics (Oxford, England) 33:1578–1580. https://doi.org/10.1093/bioinformatics/btw819
28. McGuffin LJ (2008) The ModFOLD server for the quality assessment of protein structural models. Bioinformatics (Oxford, England) 24:586–587. https://doi.org/10.1093/bioinformatics/btn014
29. McGuffin LJ (2009) Prediction of global and local model quality in CASP8 using the ModFOLD server. Proteins 77(Suppl 9):185–190. https://doi.org/10.1002/prot.22491
30. Larsson P, Skwark MJ, Wallner B, Elofsson A (2009) Assessment of global and local model quality in CASP8 using Pcons and ProQ. Proteins 77(Suppl 9):167–172. https://doi.org/10.1002/prot.22476
31. Benkert P, Tosatto SCE, Schwede T (2009) Global and local model quality estimation at CASP8 using the scoring functions QMEAN and QMEANclust. Proteins 77(Suppl 9):173–180. https://doi.org/10.1002/prot.22532
32. Cheng J, Wang Z, Tegge AN, Eickholt J (2009) Prediction of global and local quality of CASP8 models by MULTICOM series. Proteins 77(Suppl 9):181–184. https://doi.org/10.1002/prot.22487
33. McGuffin LJ, Roche DB (2010) Rapid model quality assessment for protein structure predictions using the comparison of multiple models without structural alignments. Bioinformatics (Oxford, England) 26:182–188. https://doi.org/10.1093/bioinformatics/btp629

34. McGuffin LJ, Buenavista MT, Roche DB (2013) The ModFOLD4 server for the quality assessment of 3D protein models. Nucleic Acids Res 41:W368–W372. https://doi.org/10.1093/nar/gkt294
35. Roche DB, Buenavista MT, McGuffin LJ (2014) Assessing the quality of modelled 3D protein structures using the ModFOLD server. Methods Mol Biol (Clifton, NJ) 1137:83–103. https://doi.org/10.1007/978-1-4939-0366-5_7
36. Kryshtafovych A, Schwede T, Topf M et al (2019) Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins Struct Funct Bioinform 87:1011–1020. https://doi.org/10.1002/prot.25823
37. Haas J, Barbato A, Behringer D et al (2018) Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. Proteins 86(Suppl 1):387–398. https://doi.org/10.1002/prot.25431
38. McGuffin LJ, Aldowsari FMF, Alharbi SMA, Adiyaman R (2021) ModFOLD8: accurate global and local quality estimates for 3D protein models. Nucleic Acids Res 49:W425–W430. https://doi.org/10.1093/nar/gkab321
39. Jones DT, Bryson K, Coleman A et al (2005) Prediction of novel and analogous folds using fragment assembly and fold recognition. Proteins 61(Suppl 7):143–151. https://doi.org/10.1002/prot.20731
40. Jones DT, McGuffin LJ (2003) Assembling novel protein folds from super-secondary structural fragments. Proteins Struct Funct Bioinform 53:480–485. https://doi.org/10.1002/prot.10542
41. McGuffin LJ, Jones DT (2003) Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics 19:874–881. https://doi.org/10.1093/bioinformatics/btg097
42. Sodhi JS, Bryson K, McGuffin LJ et al (2004) Predicting metal-binding site residues in low-resolution structural models. J Mol Biol 342:307–320. https://doi.org/10.1016/j.jmb.2004.07.019
43. Eastwood MP, Hardin C, Luthey-Schulten Z, Wolynes PG (2001) Evaluating protein structure-prediction schemes using energy landscape theory. IBM J Res Dev 45:475–497. https://doi.org/10.1147/rd.453.0475
44. Ben-David M, Noivirt-Brik O, Paz A et al (2009) Assessment of CASP8 structure predictions for template-free targets. Proteins 77 (Suppl 9):50–65. https://doi.org/10.1002/prot.22591
45. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086. https://doi.org/10.1110/ps.0236803
46. Wallner B, Elofsson A (2007) Prediction of global and local model quality in CASP7 using Pcons and ProQ. Proteins 69(Suppl 8):184–193. https://doi.org/10.1002/prot.21774
47. Buenavista MT, Roche DB, McGuffin LJ (2012) Improvement of 3D protein models using multiple templates guided by single-template model quality assessment. Bioinformatics (Oxford, England) 28:1851–1857. https://doi.org/10.1093/bioinformatics/bts292
48. Roche DB, Buenavista MT, Tetchner SJ, McGuffin LJ (2011) The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. Nucleic Acids Res 39:W171–W176. https://doi.org/10.1093/nar/gkr184
49. Kryshtafovych A, Fidelis K, Tramontano A (2011) Evaluation of model quality predictions in CASP9. Proteins 79(Suppl 10):91–106. https://doi.org/10.1002/prot.23180
50. McGuffin LJ, Atkins JD, Salehe BR et al (2015) IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Res 43:W169–W173. https://doi.org/10.1093/nar/gkv236
51. Maghrabi AHA, McGuffin LJ (2017) ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. Nucleic Acids Res 45:W416–W421. https://doi.org/10.1093/nar/gkx332
52. Moult J, Pedersen JT, Judson R, Fidelis K (1995) A large-scale experiment to assess protein structure prediction methods. Proteins 23:ii–v. https://doi.org/10.1002/prot.340230303
53. Roche DB, McGuffin LJ (2016) In silico identification and characterization of protein-ligand binding sites. In: Computational design of ligand binding proteins. Springer, pp 1–21
54. Yang J, Anishchenko I, Park H et al (2020) Improved protein structure prediction using predicted inter-residue orientations. Proc Natl Acad Sci 117:1496–1503. https://doi.org/10.1073/pnas.1914677117
55. McGuffin LJ, Adiyaman R, Maghrabi AHA et al (2019) IntFOLD: an integrated web resource for high performance protein

structure and function prediction. Nucleic Acids Res 47:W408–W413. https://doi.org/10.1093/nar/gkz322

56. Senior AW, Evans R, Jumper J et al (2020) Improved protein structure prediction using potentials from deep learning. Nature 577:706–710. https://doi.org/10.1038/s41586-019-1923-7
57. Toth G, Lent CS, Tougaw PD et al (1996) Quantum cellular neural networks. Superlattice Microst 20:473–478. https://doi.org/10.1006/spmi.1996.0104
58. Ba J, Caruana R (2014) Do Deep Nets Really Need to be Deep? In: Ghahramani Z, Welling M, Cortes C et al (eds) Advances in neural information processing systems 27. Curran Associates, Inc, pp 2654–2662
59. Orbach J (1962) Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. Arch Gen Psychiatry 7:218–219. https://doi.org/10.1001/archpsyc.1962.01720030064010
60. Efendigil T, Önüt S, Kahraman C (2009) A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis. Expert Syst Appl 36:6697–6707. https://doi.org/10.1016/j.eswa.2008.08.058
61. Cheng J, Choe M-H, Elofsson A et al (2019) Estimation of model accuracy in CASP13. Proteins Struct Funct Bioinform, vol 87, p 1361
62. Won J, Baek M, Monastyrskyy B et al (2019) Assessment of protein model structure accuracy estimation in CASP13: challenges in the era of deep learning. Proteins Struct Funct Bioinform 87:1351–1360. https://doi.org/10.1002/prot.25804
63. Studer G, Rempfer C, Waterhouse AM et al (2020) QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics 36:1765–1771. https://doi.org/10.1093/bioinformatics/btz828
64. Maghrabi AHA, McGuffin LJ (2020) Estimating the quality of 3D protein models using the ModFOLD7 server. In: Kihara D (ed) Protein structure prediction. Springer US, New York, pp 69–81
65. Olechnovič K, Venclovas Č (2019) VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. Nucleic Acids Res 47:W437–W442. https://doi.org/10.1093/nar/gkz367
66. Wang Z, Eickholt J, Cheng J (2010) MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. Bioinformatics (Oxford, England) 26:882–888. https://doi.org/10.1093/bioinformatics/btq058

Author Queries

Chapter No.: 6 486859_1_En

Query Refs.	Details Required	Author's response
AU1	Please check if edit to sentence starting "Variant types of. . ." is okay.	
AU2	References "[63–66]" were not cited anywhere in the text. Please provide in text citation or delete the reference from the reference list.	