# Quality Assessment of Protein Structure Models

Daisuke Kihara[1,2,3,*], Hao Chen[1] and Yifeng David Yang[1]

[1]*Department of Biological Sciences, College of Science;* [2]*Department of Computer Science, College of Science;* [3]*Markey Center for Structural Biology; 915 West State Street, Purdue University, West Lafayette, IN, 47907, USA*

**Abstract:** Computational protein tertiary structure prediction has made significant progress over the last decade due to the advancement of techniques and the growth of sequence and structure databases. However, it is still not very easy to predict the accuracy of a given predicted structure. Predicting the accuracy, or quality assessment of a prediction model, is crucial for a practical use of the model such as biochemical experimental design and drug design. Recently several model quality assessment programs (MQAPs) have been proposed for assessing global and local accuracy of predicted structures. We will start with reviewing the current status of protein structure prediction methods with an emphasis on the source of errors. Then existing MQAPs are classified into several categories and each is discussed. The categories include methods which evaluate the quality of template-target alignments, those which evaluate stereochemical irregularities of prediction models, and methods which integrate several features into a composite quality assessment score.

## INTRODUCTION

Protein tertiary structure prediction from amino acid sequence has been one of foci of computational biology and biophysics over the past twenty years [1,2,3]. This field has attracted even more researchers after the idea of threading or fold recognition was successfully introduced in 1991 [4]. Moreover, a biennial community-wide structure prediction experiment, the Critical Assessment of Techniques of Protein Structure Prediction (CASP) launched in 1994 [5] also boosted interest in this research field. Steady progress has been made due to advances in techniques as well as the increase in the number of known solved structures which can be used as templates for modeling. Now it is often possible to build accurate, atomic detailed models which can be used for redesigning protein function, *e.g.* altering DNA binding specificity [6]. However, predicting a very accurate model, *e.g.* a model with an average error at each Cα position (the root mean square deviation, RMSD) of less than 1Å to the native structure, is not always possible. Generally speaking, the accuracy of a model strongly depends on the availability of appropriate global or local templates for a target protein to be modeled. In the threading approach, it is common to produce a model with an RMSD of 6Å or 6.5Å to the native, but it is not trivial to further improve it. In the *ab initio* folding approach, usually a pool of thousands of diverse structures are generated where a few structures are selected for final predictions [7-9]. Hence there are cases when a prediction is significantly wrong, *e.g.* with an RMSD of over 10Å.

Current structure prediction methods still need improvement to be able to produce high accuracy models on a regu-

lar basis. However, arguably, the largest problem which hinders practical use of prediction methods by biologists is not necessarily the imperfectness of current prediction methods. Rather, it is the lack of estimated error or the quality of a model. It is important to note that a low resolution model is still useful for certain purposes. A review by Sali and Baker has illustrated applications of prediction models of different qualities [2]: High resolution models with an RMSD of 1 to 1.5Å are useful for almost any application where a tertiary structure of a protein can be useful, including for studying catalytic mechanism of enzymes and in a variety of structure-based protein engineering, such as drug design. A model of an RMSD of around 4Å, where residue positions but not atom positions are mostly correct are still useful, for example, for designing site-directed mutagenesis experiments [10,11] and for performing small ligand docking predictions [12,13]. If the fold of a model is expected to be correct (an RMSD of about 6Å), function of the protein could still be predicted using the predicted tertiary structures [2,14-16]. Therefore, it is important to establish quality estimation methods for predicted models, so that a model can be used wisely by knowing the limitations of the model. However, the practical importance of quality estimation of computational structure models has not been recognized until recently. A similar value, the success rate of a prediction method is routinely reported in papers proposing the new structure prediction method, but it is different from the quality assessment we are mainly interested in this article. A success rate of a method is informative in helping to understand the overall performance of the method but not necessarily useful to estimate the quality of a particular model. A quality assessment method we discuss here is one which takes a computational protein tertiary structure model (and its associated intermediate file used for computation) as input and predicts global and/or local quality information of the model. A commonly used measure of global quality of a model is

*Address correspondence to this author at the Department of Biological Sciences, College of Science; Department of Computer Science, College of Science; Markey Center for Structural Biology; 915 West State Street, Purdue University, West Lafayette, IN, 47907, USA; Tel: (765)496-2284; Fax: (765)496-1189; E-mail: dkihara@purdue.edu

the RMSD value of the model to its native structure. Measures of the local quality of a model include the distance between corresponding residues or atoms between the model and the native structure or agreement of local structure (*e.g.* main-chain or side-chain torsion angle) of a residue with that of the native structure. In what follows, we start by briefly reviewing the procedure of protein structure prediction and discuss where errors of a model could originate. Then we classify types of quality assessment methods for computational models and discuss each type individually.

## PROTEIN STRUCTURE PREDICTION METHODS AND SOURCE OF ERRORS

Protein structure prediction methods can be classified into three categories by the necessity of a template structure used as a scaffold of a model of a target sequence [2,17], namely, homology modeling, threading or fold recognition, and *ab initio* methods. The homology modeling method is based on the observation that proteins with homologous sequences fold to almost identical structures. Therefore, when a highly homologous template structure for the target sequence is available in the Protein Data Bank (PDB) [18], the method can produce an accurate model with an RMSD of 1-2Å to its native structure [19,20]. Conversely, the range of application of homology modeling is relatively narrow because a template structure is necessary for calculation. The second method, threading, seeks a well-fitting known structure for a given target sequence, sometimes from the range of beyond detectable sequence similarity [21-23]. The concept of threading is based on the observation that there are entire groups of proteins which are not evolutionarily related but have similar folds [24-26]. In threading, to find a template structure without apparent sequence similarity to a target sequence, often a combination of several scoring terms which capture compatibility between the target sequence and a template structure are used [27-29]. We call these first two categories template-based methods. Algorithms of template-based methods generally consist of two logical parts: finding a compatible structure(s) for a target sequence in a template database and then aligning the target sequence to that template structure. In homology modeling, a side-chain orientation optimization procedure usually follows. The last category, *ab initio* or *de novo* prediction methods, essentially fold a protein model from scratch, typically using a Monte Carlo optimization on physicochemical or knowledge-based statistical potentials, mimicking a physical protein folding process [7,9,30,31]. An advantage of *ab initio* methods is that theoretically the range of its application is not restricted to proteins which have template structures. However, in practice they can not handle large proteins of more than 200-250 residues in length due to a vast conformation search space. It should also be noted that the majority of current *ab initio* methods start with assembling fragments taken from known structures, and thus heavily rely on threading [9,31-33].

Fig. (**1**) summarizes the protein structure prediction process, showing how either a template-based method or an *ab initio* method is used. The prediction process starts with searching an appropriate template for a target sequence from a database by using a threading method. A conventional homology search tool, such as BLAST, is often used if only

homology modeling is attempted [34]. If a template structure is found which has a satisfactory score, an alignment between the target sequence and the template structure is computed. Usually the alignment computing step is included in the template-finding step, but careful manual refinement of an alignment can often improve the alignment significantly [35]. The resulting sequence-structure alignment only specifies main-chain orientation of aligned regions of the target. Missing regions in the alignment can be filled by a loop modeling procedure. However, loop modeling is known to be difficult, because it is almost equivalent to *ab initio* folding without template structures and also because sequence dependency of structure is relatively weak for short loops [36,37]. Then, based on the predicted main chain orientation, side-chains can be built by using a side-chain rotamer optimization program, such as SCRWL [38]. Alternatively, a homology modeling program, such as MODELLER [39], can handle these two steps for an input target-template alignment. Finally, the whole structure could be refined by using a structure optimization based on atomic detailed physics-based potentials [40-42]. In the case that an appropriate template which covers most of the regions of a target is not found, an *ab initio* method can be used to build a model (the right branch in Fig. **1**). As mentioned above, the majority of existing *ab initio* methods use known structure information retrieved in the initial database searching step, in the form of fragment structures [31,43] or consensus contacts among top threading hits [7]. Some research groups have constructed automatic prediction systems which execute the entire procedure of (Fig. **1**) [44-46].
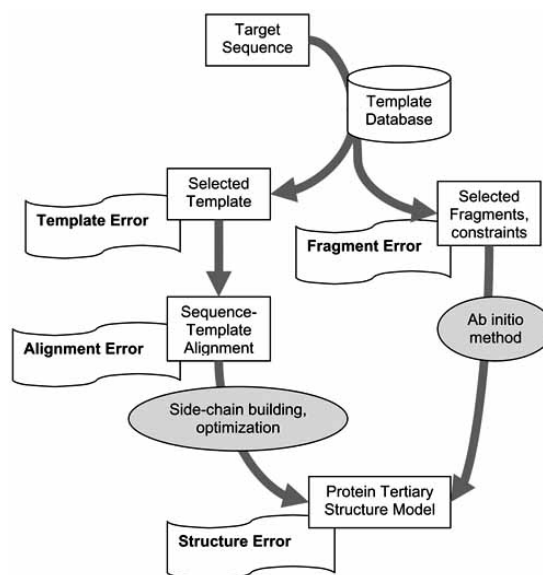


**Fig. (1).** Flowchart of protein structure prediction methods. If an appropriate template structure for a target sequence is found in a template database by a threading method, a structure model will be built on the template structure (the left branch of the chart). If not, an *ab initio* method can be employed (the right branch). Most of the current *ab initio* methods use fragment structures taken from template database. Errors can occur at each step of this prediction procedure.

In principle, errors can occur at each step in the process. In the template recognition step, wrong templates with a different fold but in the correct fold class are often recognized in threading (template recognition level error). A severe template level error can occur when the template database does not contain exactly correct structures. In that case, a threading program still ranks templates in the database according to their scores, and the top ranking structure which has a similar, but not exactly correct fold, may gain a statistically significant score. In template-based structure prediction, it is almost impossible to fix a template level error if the template is considerably different from the correct one. When a recognized template does not share sufficient sequence similarity to the target sequence, it is not easy to align the template and the target correctly [47] (alignment level error). Note that a regular pairwise sequence alignment employing a standard BLOSUM matrix [48] fails to compute a correct alignment by definition if two sequences have virtually no sequence similarity, because BLOSUM matrices are derived from blocks of multiple sequence alignments of a certain level of sequence identity. A small alignment shift error of a few residues may be tolerable for some methods [49] but a severe alignment error cannot be fixed in later stages. Finally, each procedure in the full-atom model construction, *i.e.* loop modeling and side-chain building, and the refinement step will cause errors (tertiary structure level error). It is worthwhile to note that structure optimization is not trivial; indeed simply employing a short molecular dynamics simulation usually results in deterioration of the predicted structure [41].

## ACCURACY OF THE CURRENT METHODS

At this juncture, it would be appropriate to discuss the accuracy of current prediction methods. The most recent CASP results provide us objective data of the performance of state-of-the-art methods. In CASP, participating groups predict the tertiary structure of target proteins which are not publicly released at that time of the experiment. CASP7 held in summer 2006 had three categories, the high accuracy template-based modeling (HA/TBM), the template-based modeling (TBM), and the template-free modeling (FM) (previously called novel fold) category. The number of target protein domains assessed in each category was 28, 108, and 19, respectively. Targets are classified essentially according to the level of their structure similarity to known structures in PDB [50].

Target domains are classified to the TBM category if they have sufficient structure similarity to a known structure in PDB. The sequence identity between the targets and their best template is 23.6% on average [51]. Overall performance of participating predictors was evaluated mainly by two scores, GDT-HA and AL0. The GDT-HA score is the average of the percentage of residues which do not deviate from the target structure by more than 0.5, 1.0, 2.0, and 4.0Å. The AL0 score is the percentage of correctly aligned residues between a prediction and a target structure within 3.8Å in a structure alignment. In the TBM category, the average GDT-HA score of the top five groups among the total of 794 participating groups was 50.3 [51]. Thus, on average about only half of residue position in a prediction model is correctly predicted within the error of 4.0Å or better. When it comes

to the FM category, obviously prediction became more difficult: Very roughly speaking from (Fig. **2B**) of the assessors' report [52], the average GDT-TS score (this is a relaxed score from GDT-HA, which uses 1.0, 2.0, 4.0, and 8.0Å as threshold values) of all the FM targets by all the participants is around 25, and that of top groups is around 30-40. Thus the percentage of correctly predicted residues is about one-third on average. Moreover, comparison of results of past CASP experiments indicates that the improvement of performance has significantly decreased in the last few years [53,54]. Therefore, from a practical point of view, users should be aware that even the state-of-the-art methods are not able to make accurate predictions in many cases. This further emphasizes importance of establishing ways to assess the quality of prediction models to make current methods practically useful.

## TEMPLATE LEVEL ERROR

In what follows, we discuss ways to assess errors at each step of the modeling process (Fig. **1**), namely, the template level, the alignment-level, the selected fragment level, and the structural level error.

A template structure for a given target sequence is identified by considering the significance of the score which indicates fitness of the target to the template. In principle, the significance can be indicated by a raw score, such as the sequence identity, the raw Smith-Waterman alignment score, or a threading score between a target and a template. But more frequently, statistical significance of a raw score is considered, for example, in the form of the E-value (in homology search, *e.g.* BLAST [55]) or the Z-score (used in most of threading algorithms). Simply, the more statistically significant a score of a template is the higher chance that the template is correct. Hence most of the current threading methods provide a recommended threshold Z-score to indicate confident predictions to users.

To go one step further, confidence of the top scoring template can be assessed by checking "consistency" of the selection of the template. There are two ways to employ the idea of examining consistency. First, one can check if several of the top scoring templates with a sufficient Z-score also have a consistent fold, *i.e.* the fold type defined by SCOP [56] or CATH database [57]. For example, SP3, a threading program, considers consistency of the folds in the top two scoring templates in its scoring scheme [58]. Second, consistency of selected templates by different independent programs has been proven to indicate confidence of the prediction. The majority vote of different methods is called the ensemble approach or the meta-server approach and proven to work well in threading [46,59,60] and many other types of bioinformatics predictions [61-63].

## ALIGNMENT LEVEL ERROR

Similar to the error estimation in the template level, in principle, the alignment level error can be estimated from a raw score of a target to template alignment or statistical significance of the raw score, *e.g.* the Z-score or the E-value [64,65]. A Z-score of a raw alignment score is also obtained from a distribution of alignment scores from shuffled sequences (the PRSS program [66]). Tress and his co-workers

reported that a profile alignment score assigned to each residue position is a good predictor of the alignment shift error of that position [67]. An algorithm to extract local regions with a high score in an alignment as reliable regions is proposed by Miller *et al.* [68]. Lee *et al.* predict the quality of global sequence alignments by using a Support Vector Machine (SVM) which takes a position specific scoring matrix score at each position of an input alignment [69]. Similarly, Tondel used a least square regression to predict the RMSD of homology models of kinase proteins, combining amino acid similarity scores of each aligned position, the sequence identity, and the number of gapped regions in a target-template alignment [70].

In our recent work [71], we have compared the correlation coefficient between the global RMSD of models constructed by MODELLER and the sequence identity, the Z-score computed by a threading program, and the Z-score computed by PRSS. We found that the simple sequence identity between a target and a template has the best correlation to the RMSD of models among the three measures when the target and the template belong to the same family, and the PRSS Z-score follows. Interestingly, the threading Z-score does not show a meaningful correlation to the RMSD of models. However, as we will discuss later in this section, using a score derived from suboptimal alignments generally correlates better to RMSD of models to their native structures.

The above methods aim to estimate the accuracy of an alignment, or the tertiary structure built based on the alignment, by considering the significance of alignment scores. Here a single target alignment is examined, although one might note that amino acid similarity scores such as BLOSUM [48] or PAM [72] series are derived from statistics of alignments, thus the target alignment is compared somewhat implicitly with other alignments of different sequences. An alternative strategy to estimate the reliability of an alignment is to compare it explicitly with alternative alignments of the same pair of sequences, *i.e.* to consider suboptimal alignments [73]. There are two pioneering works on computing suboptimal alignments using dynamic programming (DP) algorithm. The first method, proposed by Vingron and his colleagues, computes the best possible (*i.e.,* highest scoring) alignment under the constraint that a certain pair of residues from two sequences should be aligned. Every pair of residues is considered as a constraint, and an alignment is computed for each of the constraints. Finally resulting alignments are sorted by their alignment score [73-75]. The second DP-based method by Sternberg *et al.* computes the optimal alignment as usual, then in subsequent iterations, the previously computed alignments are penalized so that a different alignment can be obtained [76]. Their works showed that one of the suboptimal alignments is often more accurate than the optimal one [76] and consistent regions among optimal and suboptimal alignments are more likely to be correct [75]. A set of alternative alignments can be also obtained by using different parameters for computing alignments [77,78].

Rather than explicitly computing numerous suboptimal alignments to examine consistency, several methods were developed which provide a probability (reliability) to each position in an alignment. One particular type of those methods is inspired by the partition function in statistical mechanics, which is used to express the probability of alternative alignments [79-82]. Alternatively a hidden Markov model can be naturally applied to assign probability of local positions of an alignment [83,84].

Recently, we have also developed a method for estimating the alignment-level error by considering consistency of a target-template alignment with suboptimal alignments [71]. Our method, which computes an index named SPAD (SubOPtimal Alignment Diversity), quantifies how divergent a set of suboptimal alignments are around the optimal alignment on the DP matrix. We showed that the SPAD score has a significant correlation not only to alignment shift level errors but also to global and local structural level errors (*i.e.* RMSD to the native structure and the distance of corresponding residues of a model and its native structure) of structure models built based on optimal alignments. The correlation to the error with SPAD was more significant as compared with the sequence identity, the Z-score by threading, and the Z-score by PRSS. However, as one might expect, alignment-based scores including SPAD lose their correlation to the RMSD of models as the sequence similarity level between a target and a template drops. (Fig. **2**) shows examples of predicted and actual local error of predicted structure models. These models are predictions made by our group in CASP7. The first example, Fig. (**2A**) is an example of a relatively successful prediction, where the overall fold is correctly predicted. The RMSD of the model to the native structure is 3.9Å. The local error of the middle part of the model (residue positions at around 50-100) has a larger error of around 6.0Å, which is well captured by the predicted local error by the SPAD score. (Fig. **2B**) is an example of prediction of low quality, where the overall RMSD is 11.7Å. The model is particularly wrong at the N- and the C-terminus due to the different orientation of the two termini of the template used (1dvoA) and the native structure of the target. Despite the overall low quality of the model, SPAD predicted the local error of the middle part of the model well. The right panel in (Figs. **2A** and **2B**) shows a sausage representation, where the predicted error at each position is represented by the radius of the tube. This representation can be used to visualize the overall conformation of a model and the local error simultaneously. The fact that the structural level error of template-based models can be well predicted by alignment information implies that current template-based structure prediction methods (MODELLER is used in our work) strongly relies of the quality of an input alignment.

To summarize, alignment-level errors can be predicted by considering the strength of the fitness between a target sequence to a template structure, or by considering the stability or consistency of an optimal alignment relative to a set of suboptimal alignments. Alignment-based scores can often predict structure level errors as well, since template-based structure prediction methods rely heavily on the initial target-template alignment. Advantages of examining errors in the alignment level is that inevitable errors are detected in an earlier stage before entering time-consuming steps of tertiary structure building. Also several alternative alignments can be provided in case the optimal alignment is not very reliable.
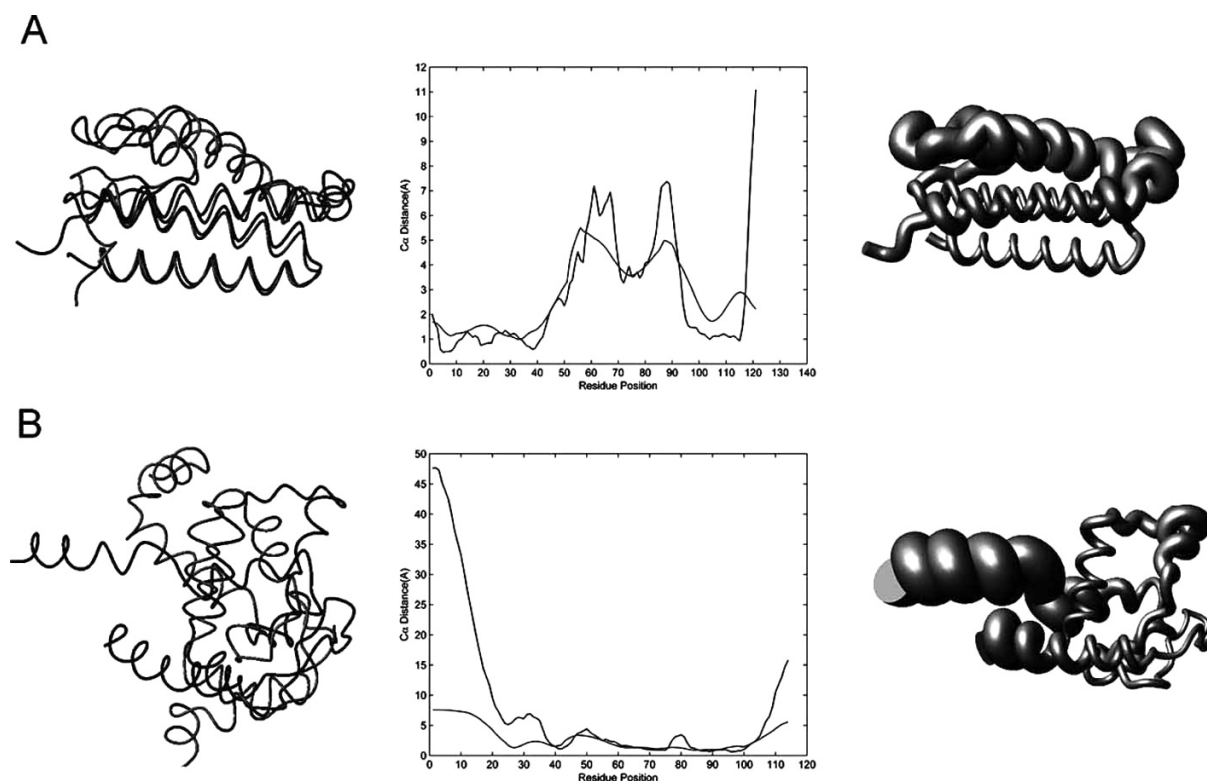
**Fig. (2).** Examples of estimated local errors by the SPAD score, which examines the consistency of optimal and suboptimal alignments. Prediction structures of two CASP7 targets produced by our group, Chen-Tan-Kihara, are shown. **A**, a predicted structure of the target T0367. Left: superimposition of the predicted structure (black) with the native structure (gray); Middle: actual (black) and predicted (gray) local error (Å) at each residue position of the model; Right: the sausage representation of the model, where the radius of the tube is proportional to the predicted local error of that position. The sequence identity (SeqID) between the target protein and the template structure used, 1ufbA, is 14.5%. The RMSD between the predicted structure and its native structure is 3.9 Å. **B**, target T0360. The SeqID between the target and the template, 1dvoA is 13.9%. The RMSD of the predicted model is 11.7Å.

## SELECTED-FRAGMENT LEVEL ERROR

Selecting the right fragments for a target is the basis of *ab initio* folding, which employs fragment assembly. Therefore it would be reasonable to assess the quality of the selected fragments in the course of running an *ab initio* structure prediction method. Since the fragment selection process is essentially threading with sequence fragments, the fragment-level error can be logically decomposed into fragment recognition level and alignment level error. Although we did not find methods which are aimed for predicting quality of the selected fragments in the *ab initio* folding process, we found two methods which use fragments as the unit of assessing quality of structure models, thus could be applied for fragment quality assessment.

Rangwala and Karypis combined two types of information by SVM to predict the RMSD of a fragment of a fixed length from a structure model [85], namely, a profile similarity score between corresponding fragments of a target and a template and a score of agreement between predicted secondary structure of the target with that of the template. TASSER-QA by Zhou and Skolnick [86] consider how well local structures taken from top threading hits agrees (*i.e.* RMSD of fragments) and a residue contact potential o predict the global structure quality of prediction models.

## TERTIARY STRUCTURE LEVEL ERROR

In this section, we review methods for evaluating the quality of structure models per se by examining various structural features. Validation of tertiary structures is also an important step in experimental structural biology. Earlier works on protein structure validation focused on identifying potential errors in crystal structures of proteins. Tools developed for that purpose include PROCHECK [87], MOLPROBITY [88], PROVE [89], and WHATCHECK [90]. PROCHECK and WHATCHECK examine various structural properties, such as the bond length, bond angles, and atom clashes, detecting atoms which have anomalous stereochemical values. Similarly, MOLPROBITY examines atom clashes and plausibility of dihedral angles. PROVE uses the Voronoi procedure to compute the volume occupied by atoms to check if the volume deviates from the standard values. The TAP score [91] measures fitness of local sequence to structure based on torsion angle propensities of amino acids in the form of knowledge-based potential. The TAP score is shown to have a good correlation to the R-free value of protein crystal structures. Morris *et al.* investigated distribution of several conformational parameters including dihedral angles and hydrogen bonds of protein crystal structures of different resolutions [92]. Colovos and Yeates proposed to examine noncovalently bonded interactions be-

tween heavy atoms to find incorrectly determined regions in a protein crystal structure [93]. The deviation of actual Cβ position from an ideal geometry position, which is computed from the backbone atoms, is proposed as a simple measure of geometrical non-ideality around Cα atoms [94].

The basic strategy of these methods is to compare stereochemical properties of a protein structure with their regular values sampled from a set of representative protein structures of good resolution. The same methods can be applied to assess the quality of predicted structures. A homology modeling tool, HOMA, evaluates a model using a combined score which consists of the van der Waals potential, bond length, and bond angle terms computed by X-PLOR, a commonly used structure analysis software for X-ray crystallography [95]. However, it may not always be suitable to use crystal structure validation tools for analyzing predicted structures because those tools concern small deviations of distances or angles, which is the level of the accuracy that may not be meaningful to expect for predicted structures of a moderate accuracy.

Another natural idea of assessing structure level quality is to employ physics-based all atom force field. Previous works along this line include ones which employ the molecular mechanics –Poisson-Boltzmann Surface Area (MM-PBSA) free energy with the AMBER potential [96] or the CHARMM potential [97]. Both of the works showed that the native structure is quite well discriminated from decoy structures but the correlation coefficient of the energy to the RMSD of decoys to the native is not very good, especially when decoys are relatively far from the native (*e.g.* an RMSD of over 5Å). Essentially the same conclusions were drawn by the Explicit Simulation/Implicit Solvent (ES/IS) method, which computes the solvation free energy from short molecular dynamics simulations with explicit solvent [98]. The colony energy approach was combined with the MM-PBSA, which assesses conformational entropy accurately by explicitly sampling the conformational space in the vicinity of a reference structure [99]. This method was applied to loop decoy discrimination. Kmiecik *et al.* showed encouraging results that the AMBER potential was able to show a good correlation to the RMSD of all-atom reconstructed structures based on a coarse-grained structure model, CABS [100].

Wroblewska and Skolnick drew interesting conclusions regarding free energy computed by MM-Generalized Born implicit solvation model with a Surface Area dependent term (MM-GBSA) from a thorough benchmark study using 150 nonhomologous proteins [41]. They showed that the MM-GBSA energy fails to recognize the native structure from decoys when all the structures are sufficiently minimized. It was argued that some of the earlier successes in recognizing native structures by physics-based all-atom force fields were artifact of the decoy preparation procedure; decoy structures contain worse residue packing than the native and unrealistic conformations of side-chains. They later showed that reoptimization of relative weights of energy components of the AMBER force field yielded significant improvement for scoring and refinement of protein models [42]. This is consistent with what Feig and Brooks showed with the CHARMM potential [97].

Alternative to using physics-based force fields, a common way to evaluate predicted structures is to examine atom or residue contacts in the form of knowledge-based contact potentials. A knowledge-based contact potential computes the number of observed contacts of pairs of atoms (or residues) in a set of representative proteins normalized by the expected number of contacts of that pair. Thus, using a knowledge-based contact potential is essentially equivalent to comparing observed stereochemical property (contacts in this case) with its regular distribution. Sippl has done pioneer works on using knowledge-based contact potentials to identify major errors in protein crystal structures [101,102]. Knowledge-based potentials are flexible in choosing definitions of contacts, for example, whether to consider atom contacts or residue contacts (either by considering contacts between Cαs or Cβs), and whether to make it distance-dependent. These choices are to be made according to the targeted application of the potentials. Typically in a knowledge-based atomic detailed potential, each heavy atom in each amino acid residue is handled separately because they are considered to have different structural environments. In principle, atomic detailed knowledge-based potentials are designed to evaluate structures with an atomic level accuracy, but they are shown to have good performance on predicted structures of a moderate accuracy as well. Indeed atomic detailed knowledge-based potentials are routinely used for discriminating near-native structures from non-native decoy structures generated by *ab initio* protein structure prediction programs [103-106]. Melo & Feytmans showed that an atom contact potential combined with an atom-based accessible surface mean force potential performs well in discriminating homology models which are close to the native from incorrect models [107]. Pettitt *et al.* showed that MODCHECK, which combines an atomic detailed knowledge-based potential and a residue-level solvation potential, is successful in the reranking of threading models by building full atoms based on the threading alignments [108]. An interesting scoring method for interatomic contacts and atom-solvent contacts was proposed by McConkey *et al.* [109]. They used a Voronoi tessellation to define atom-atom and atom-solvent contact surfaces and thus integrated the solvent accessible surface and interatomic contacts into one scoring function. The function showed a good performance in discriminate native structures in decoy sets generated by *ab initio* methods.

Considering residue contacts could be advantageous for evaluating predicted structures of a moderate accuracy, where correct main-chain orientation but not accurate atomic positions are expected. Melo and co-workers have developed four types of residue-level potentials; distance-dependent and -independent contact potentials, an accessible surface potential, and a main-chain torsion angle potential and applied them to discriminate correct and incorrect models of a wide range of accuracies [110]. The Verify3D algorithm employs a residue-level score which assesses fitness of a residue to its structural environment described by the secondary structure, burial status and polarity of positions in a structure [111]. Verify3D was used by many successful groups in CASPs [46,112,113].

There are several other interesting ideas for assessing the quality of prediction models using coarse-grained structural

characteristics of models. Holm and Sander use solvation (exposure) preference of atoms computed from a representative protein set [114]. Costantini *et al.* estimate the quality of prediction models by using a score called "globularity", which concerns the accessible surface area, the number of hydrogen bonds, the total volume of empty cavities, and the number of surrounding water molecules relative to the protein's molecular weight [115]. Bartlett and Taylor approached the quality assessment problem from a sequence analysis method, namely, what they call the statistical coupling analysis [116]. This method is also known as the correlated mutation analysis [117]. Physically close residues in a protein structure frequently show correlated mutation behavior in a multiple sequence alignment of the protein family, because they are physically constrained in mutation. The statistical coupling analysis quantifies the degree of the correlated mutations of physically close residues in a predicted model and select models with a high degree of correlated mutations as plausible ones in a pool of decoy structures generated by an *ab initio* folding method.

## COMBINATION OF DIFFERENT FEATURES OF STRUCTURE MODELS

So far we have overviewed various features which capture different aspects of the quality of predicted structures. Those features discussed are summarized in Table **1**. In order to evaluate a predicted structure comprehensively, it is common to construct a composite score which combines several features.

The QMEAN score [118] linearly combines five structural features: a torsion angle potential, an amino acid level solvation potential, a secondary structure specific distance dependent pairwise residue-level potential, amino acid propensity of solvent accessibility, and agreement of predicted secondary structure of a target protein with actual ones of a template used. For each of the features, several different definitions are used and compared. They performed a thorough benchmark study using four decoy sets and compared with five other MQAPs and molecular mechanics (MM) force fields. Several interesting conclusions are observed in their study. First, among several residue contact potentials used, a secondary structure-dependent Cβ atom-based contact potential performed the best. Secondly, their statistical torsion angle potential considering three consecutive residues performed well, which uses a surprisingly coarse bin size to discretize the Ramachandran map. The bin size used for the center residue is 45° and 90° for the adjacent residues. It is not clear if these bin sizes are optimal since other bin sizes are not tested. Using a coarse grid size may be advantageous to distinguish decoys of a moderate accuracy from further worse ones. Thirdly, when compared with MM-based energies (MM energy, MM/GBSA free energy, MM/PBSA free energy), the MM energies showed better performance in the actual quality of the top scoring decoy (the decoy with the lowest energy), while QMEAN outperformed the MM energies in terms of the having more decoys of a good quality in top n ranking decoys. Thus MM-based energies are specifically good at identifying highly accurate structures, while QMEAN showed superior performance in selecting moderately accurate models.

Victor/FRST [119] uses a score which is a linear combination of a residue-specific all-atom distance-dependent contact potential, a solvation potential, a hydrogen bonding potential, and a torsion angle potential. All these terms except for the atom-based contact potential used in Victor/FRST are residue-based which are typically used in threading. Terms used in QMEAN are also residue-based. Thus, practically, these two programs perform threading on a given decoy set and reranking them by their own scores. Victor/FRST performed well in CAFASP4 (Critical Assessment of Fully Automated Structure Prediction) experiment.

Qiu *et al.* constructed a composite score using Support Vector Regression (SVR) [120]. They examined 21 structural features used in ROSETTA *ab initio* prediction program and four model consensus scores in terms of correlation to the GDT-TS score (*i.e.* structure similarity to the native structures) of predicted models. Consequently, twelve terms were selected and combined using SVR. Among the features selected a consensus score which quantifies agreement of predicted models by different methods showed the most significant correlation to the GDT-TS score. The other interesting features used include three scores for hydrogen bonds, ones for long-range backbone hydrogen bonds, one for short-range backbone, and one for side-chain hydrogen bonds, and a score for assessing the contact order of models.

Melo and Sali examined 21 features ranging from alignment-based scores to structure-based scores including statistical potentials, stereochemistry quality descriptors, and measures of protein packing [121]. Among examined, the highest accuracy in discriminating correct models (defined as models with an RMSD of 3.5Å or lower to the native) from incorrect models are achieved by the Z-score of statistical potentials of contacts and accessible surface, a score which captures protein packing, and a target-template alignment Z-score. Interestingly, stereochemical quality measures, which are assessed by the percentage of residues in favored/disfavored regions in the Ramachandran map, show poor accuracy. This is because good stereochemical quality can be achieved irrespective to the overall similarity of a model to its native structure. Finally, they developed two composites scores whose combination of the scoring terms is optimized by linear discriminant analysis and genetic algorithm. The terms used in the two composites scores are model compactness, the percentage of sequence identity, Z-scores for the residue distance and accessibility statistical potentials, and the model length. The resulting score has been implemented in the MODELLER package.

AIDE developed by Mereghetti *et al.* uses neural network (NN) to combine fifteen structural features of a model including solvent accessible surface of hydrophobic and hydrophilic residues, secondary structure content, the number of hydrophobic contacts, and several PROCHECK parameters, such as the percentage of residues in Ramachandran plot allowed/disallowed regions [122]. AIDE predicts the real value of quality, including the RMSD and structural similarity scores to the native (i.e. TM-score, GDT-TS score).

Combining existing quality assessment scores is a promising way to construct a MQAP which practically performs well. Fasnacht *et al.* [123] combined DFIRE [104], a distant

dependent statistical potential, and ProQRes [124] to predict local quality of models. PRoQRes predicts local quality at each residue in a model using a NN which combines atom-atom contacts, residue-residue contacts, solvent accessibility surfaces, and secondary structure information.

ModFold [125] uses a NN to combining outputs of the other MQAPs. The methods combined are ModSSEA, MODCHECK [108], and ProQ [124]. ModSSEA is based on secondary structure element alignments (SSEA) between the DSSP assigned secondary structure of the target model with PSI-PRED predicted secondary structure. MODCHECK combines a residue contact potential and a solvation potential. ProQ is a NN-based method for global quality prediction, which combines scores of residue-contacts, atom-contacts, residue preference of accessibility surfaces, predicted secondary structures, and the overall volume of a model [124].

Eramian *et al.* performed a thorough benchmark of 24 existing assessment scores, including physics-based energy functions, statistical potentials, and machine learning-based scoring functions in terms of the correlation to the RMSD of models [126]. Among examined, DFIRE and DOPE [103], both of which are the statistical atom contact potentials, showed consistently better performance. Matching of predicted (by PSIPRED) and actual secondary structure of a model and the composite score used in ROSETTA also performed well. Finally, they developed a combined score with six scores in the framework of SVM, namely, two secondary structure matching scores, DOPE, residue contact, residue accessible surface, and structural compactness score used in MODPIPE [110]. The combined score outperformed the other 24 individual scores.

Table **1** summarizes all the features for model quality assessment discussed in this review. Moreover, features used by the methods introduced in this section are specified. This table clarifies frequently used features and also features which are not explored yet. It is evident from Table **1** that most of the existing methods concentrate on combining structure-based terms. Since it is shown that alignment-based terms have complementary strength to structure-based terms for predicting quality of template-based models [71], an interesting direction would be to explore a mixed combination of alignment-based and structure-based terms.

## CONCLUDING REMARKS

In this article, we summarized current methods for quality assessment of protein models. We classified sources of errors which may occur along protein structure prediction process and discussed strategies of MQAPs for each error source. Three main purposes of MQAPs are observed. First, in structural biology, stereochemical properties of experimentally solved structures are routinely examined. Second, in structure prediction field, particularly in *ab initio* prediction, reranking models is performed, such that the most native-like structure is selected from a pool of models. The third purpose is to predict the real quality value of a model, such as a RMSD of the model to the native structure. MQAPs of the last type are the most useful for biologists who would like to practically use structure models. MQAPs of the first two types are not necessarily relevant to predict-

ing a real value of quality of a model, since correctness of detailed stereochemical properties does not guarantee overall structural similarity of a model to the native [41,121] and ranking models does not tell the real quality value of the models. Actually, most of the MQAPs introduced in this review are designed to rerank models because their primary purpose is to improve the accuracy of a protein structure prediction method. Therefore, it is desired that more MQAPs of the third type are developed, because they are a key for bridging computational and experimental biology, bringing the structure prediction tools into experimental biology labs.

Structure prediction methods have been already frequently employed by biologists. Most of them employ homology modeling using an obvious homolog as the template and design and verify site-directed mutation experiments [127-132]. Some homology models were further used for ligand docking predictions [133,134] or predicting the quaternary structure [135], which were verified by experiments. There are fewer examples of experimental works combined with threading methods [136]. It would also be appropriate to mention recent notable applications of atomic detailed protein models in designing proteins with a desired function [6,137,138]. Although there are an increasing number of examples of application of structure models, most of them are limited to use of homology models, probably because models are likely to have a high accuracy. However, if coupling with site-directed mutagenesis is intended, threading models would also suffice in many cases, as long as the quality of models is confidently predicted. Thus, establishing the real-value quality assessment methods (the third class of the MQAPs discussed above) would significantly broaden the applicability of protein structure models of a moderate resolution.

At this point, what would be practical advices for biologists who are willing to use structure model for designing and interpreting experiments? Unfortunately, in the current situation when most of prediction methods do not provide a real-value quality assessment, advices would be as general as what would be advised for using sequence-based homology searches: (1) To use models with a score above the recommended threshold value of the structure prediction method; (2) to check the models if they are consistent with biological knowledge of that protein; (3) to use several prediction methods to see if they provide consistent results; and (4) finally, it is recommended to check the original paper of the method to understand the limitation of the methods.

Recognizing importance of model quality assessment, CASP has included quality assessment category [139]. CAFASP has also introduced MQAP category from the fourth round of their experiments [140]. Modbase, a database of homology models, provides several quality assessment scores including the sequence identity between a target and a template and contact potential-based scores they have developed [141]. Recently we have developed EcoliPredict, which is a database of template-based structure models of *E. coli* proteins. EcoliPredict is a component of Ecolihub database, which is a hub site for various types of data on *E. coli* (http://www.ecolihub.org). In EcoliPredict, we provide several quality assessment data computed with several methods to meet requests by users.

**Table 1.    Summary of Features for Quality Assessment of Protein Predicted Structures**

| Category of Features | Features | Methods Which Combine Multiple Features | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HOMA [95] | fRMSD Pred [85] | TASSER-QA [86] | MODCHECK [108] | Globularity score [115] | QMEAN [118] | Qiu *et al.* [120] | Melo & Sali [121] | AIDE [122] | Mod-Fold [125] | ProQ [124] | SVMod [126] | Victor/FRST [119] | Fasnacht *et al.* [123] |
| Alignment-based | Sequence identity between target and template | | | | | | | | x | | | | | | |
| | Profile score | | | | | | | | | | | x | | | |
| | Gap ratio | | | | | | | | | | | | | | |
| | Target-template alignment Z-score | | | | | | | | x | | | | | | |
| | Fraction of target regions modeled | | | | | | | | x | | | | | | |
| | Consistency among alternative alignments | | | | | | | | | | | | | | |
| Fragment-based | Sequence or profile similarity between target and fragments from template | | x | | | | | | | | | | | | |
| | Structural similarity of fragments from threading hits | | | X | | | | | | | | | | | |
| Structure-based | Atom clashes, van der Waals | X | | | | | | | | x | | | | | |
| | Bond length | X | | | | | | | X | x | | | | | |
| | Bond angles | X | | | | | | | | | | | | | |
| | Main chain torsion angles | | | | | | x | X | X | x | | | x | x | |
| | residue-based/atom-based, distance-dependent/independent, contact potential | | | x | x | | x | X | X | | x | x | x | x | x |
| | Solvation potential, burial preference of residues/atoms | | | | x | | x | X | X | x | x | x | x | x | x |
| | Hydrophobic residue contacts | | | | | | | | | x | | | x | | |
| | Residue packing preference | | | | | | | X | | | | | | | |
| | Hydrogen bond | | | | X | | | X | | | | | | x | |
| | hydrogen bond with water molecules | | | | | x | | | | | | | | | |
| | Contact order | | | | | | | X | X | | | | | | |
| | Compactness | | | | | | | X | | | x | X | | | |
| | Accessible surface area | | | | | x | | | | | | | | | |
| | Void | | | | | x | | | | | | | | | |
| | Correlated mutation | | | | | | | | | | | | | | |

(Table 1) contd….

| Category of Features | Features | Methods Which Combine Multiple Features | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HOMA [95] | fRMSD Pred [85] | TASSER-QA [86] | MODCHECK [108] | Globularity score [115] | QMEAN [118] | Qiu et al. [120] | Melo & Sali [121] | AIDE [122] | Mod-Fold [125] | ProQ [124] | SVMod [126] | Victor/FRST [119] | Fasnacht et al. [123] |
| | Secondary structure content | | | | | | | | | x | | | | | |
| | Agreement of predicted secondary structure (target) and actual one (template) | | x | | | | x | | | x | x | X | x | | x |
| | Agreement of predictions with the other methods | | | | | | | X | | | | | | | |
| | Radius of gyration | | | | | | | | x | | | | | | |
| | % of residues in preferred/unpreferred regions in Ramachandran plot | | | | | | | | x | x | | | | | |
| | Molecular mechanics energy | | | | | | | | | | | | | | |

Some methods combine outputs of existing programs. Therefore some features are indirectly taken into account as inputs of component programs.

Not only in experimental structural biology but in any quantitative science error estimation is an indispensable step for using available data with consideration of the range of its accuracy. This is true for the protein structure prediction methods and also for the other bioinformatics tools so that they can enable fruitful research exploration together with experimental methods.

## ACKNOWLEDGEMENTS

## REFERENCES

[1]   Al-Lazikani, B.; Jung, J.; Xiang, Z.; Honig, B. Protein structure prediction. *Curr. Opin. Chem. Biol.*, **2001**, *5*, 51-56.
[2]   Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science*, **2001**, *294*, 93-96.
[3]   Gibson, K.D.; Scheraga, H.A. Minimization of polypeptide energy. I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proc. Natl. Acad. Sci. USA*, **1967**, *58*, 420-427.
[4]   Bowie, J.U.; Luthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three- dimensional structure. *Science*, **1991**, *253*, 164-170.
[5]   Moult, J.; Pedersen, J.T.; Judson, R.; Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins*, **1995**, *23,* 2-4.
[6]   Ashworth, J.; Havranek, J.J.; Duarte, C.M.; Sussman, D.; Monnat, R.J.; Jr., Stoddard, B.L.; Baker D. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **2006**, *441*, 656-659.
[7]   Kihara, D.; Lu, H.; Kolinski, A.; Skolnick, J. Touchstone: an ab initio protein structure prediction method that uses threading-based tertiary restraints. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 10125-10130.
[8]   Borreguero, J.M.; Skolnick, J. Benchmarking of TASSER in the ab initio limit. *Proteins*, **2007**, *68*, 48-56.
[9]   Bonneau, R.; Strauss, C.; Rohl, C.; Chivian, D.; Bradley, P.; Malmstrom, L.; Robertson, T.; Baker, D. De novo prediction of three-dimensional structures for major protein families. *J. Mol. Biol.*, **2002**, *322*, 65-78.
[10]  Wells, G.A.; Birkholtz, L.M.; Joubert, F.; Walter, R.D.; Louw A.I. Novel properties of malarial S-adenosylmethionine decarboxylase as revealed by structural modelling. *J. Mol. Graph. Model.*, **2006**, *24*, 307-318.
[11]  Skowronek, K.J.; Kosinski, J.; Bujnicki, J.M. Theoretical model of restriction endonuclease HpaI in complex with DNA, predicted by fold recognition and validated by site-directed mutagenesis. *Proteins*, **2006**, *63*, 1059-1068.
[12]  Wojciechowski, M.; Skolnick, J. Docking of small ligands to low-resolution and theoretically predicted receptor structures. *J. Comput. Chem.*, **2002**, *23*, 189-197.
[13]  Vakser, I.A. Low-resolution docking: prediction of complexes for underdetermined structures. *Biopolymers*, **1996**, *39*, 455-464.
[14]  Skolnick, J.; Fetrow, J.S.; Kolinski, A. Structural genomics and its importance for gene function analysis. *Nat. Biotechnol.*, **2000**, *18*, 283-287.
[15]  Hawkins, T.; Kihara, D. Function prediction of uncharacterized proteins. *J. Bioinform. Comput. Biol.*, **2007**, *5*, 1-30.
[16]  Kihara, D.; Skolnick, J. Microbial genomes have over 72% structure assignment by the threading algorithm PROSPECTOR_Q. *Proteins*, **2004**, *55*, 464-473.
[17]  Kretsinger, R.H.; Ison, R.E.; Hovmoller, S. Prediction of protein structure. *Methods Enzymol.*, **2004**, *383*, 1-27.
[18]  Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.*, **2000**, *28*, 235-242.
[19]  Sanchez, R.; Pieper, U.; Melo, F.; Eswar, N.; Marti-Renom, M.A.; Madhusudhan, M.S.; Mirkovic, N.; Sali, A. Protein structure modeling for structural genomics. *Nat. Struct. Biol.*, **2000**, (Suppl 7), 986-990.

[20] Marti-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, *29*, 291-325.

[21] Godzik, A. Fold recognition methods. *Methods Biochem. Anal.*, **2003**, *44*, 525-546.

[22] Jones, D.T. Progress in protein structure prediction. *Curr. Opin. Struct. Biol.*, **1997**, *7*, 377-387.

[23] Torda, A.E. Perspectives in protein-fold recognition. *Curr. Opin. Struct. Biol.*, **1997**, *7*, 200-205.

[24] Pascarella, S.; Argos, P. A data bank merging related protein structures and sequences. *Protein Eng.*, **1992**, *5*, 121-137.

[25] Holm, L.; Ouzounis, C.; Sander, C.; Tuparev, G.; Vriend, G. A database of protein structure families with common folding motifs. *Protein Sci.*, **1992**, *1*, 1691-1698.

[26] Orengo, C.A.; Flores, T.P.; Jones, D.T.; Taylor, W.R.; Thornton J.M. Recurring structural motifs in proteins with different functions. *Curr. Biol.*, **1993**, *3*, 131-139.

[27] Liu, S.; Zhang, C.; Liang, S.; Zhou, Y. Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins*, **2007**, *68*, 636-645.

[28] Skolnick, J.; Kihara, D.; Zhang, Y. Development and large scale benchmark testing of the PROSPECTOR 3.0 threading algorithm. *Proteins*, **2004**, *56*, 502-518.

[29] Matsuo, Y.; Nishikawa, K. Protein structural similarities predicted by a sequence-structure compatibility method. *Protein Sci.*, **1994**, *3*, 2055-2063.

[30] Skolnick, J.; Zhang, Y.; Arakaki, A.K.; Kolinski, A.; Boniecki, M.; Szilagyi, A.; Kihara, D. TOUCHSTONE: a unified approach to protein structure prediction. *Proteins*, **2003**, *53*, 469-479.

[31] Jones, D.T. Predicting novel protein folds by using FRAGFOLD. *Proteins*, **2001**, (Suppl 5), 127-132.

[32] Boniecki, M.; Rotkiewicz, P.; Skolnick, J.; Kolinski, A. Protein fragment reconstruction using various modeling techniques. *J. Comput. Aided Mol. Des.*, **2003**, *17*, 725-738.

[33] Zhang, Y.; Skolnick, J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. *Biophys. J.*, **2004**, *87*, 2647-2655.

[34] Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **2003**, *31*, 3381-3385.

[35] Tomii, K.; Hirokawa, T.; Motono, C. Protein structure prediction using a variety of profile libraries and 3D verification. *Proteins*, **2005**, *61* (Suppl 7), 114-121.

[36] Soto, C.S.; Fasnacht, M.; Zhu, J.; Forrest, L.; Honig, B. Loop modeling: Sampling, filtering, and scoring. *Proteins*, **2008**, *70*, 834-843.

[37] Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, **2003**, *19*, 2500-2501.

[38] Canutescu, A.A.; Shelenkov, A.A.; Dunbrack, R.L.; Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.*, **2003**, *12*, 2001-2014.

[39] Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **1993**, *234*, 779-815.

[40] Lu, H.; Skolnick, J. Application of statistical potentials to protein structure refinement from low resolution ab initio models. *Biopolymers*, **2003**, *70*, 575-584.

[41] Wroblewska, L.; Skolnick, J. Can a physics-based, all-atom potential find a protein's native structure among misfolded structures? I. Large scale AMBER benchmarking. *J. Comput. Chem.*, **2007**, *28*, 2059-2066.

[42] Wroblewska, L.; Jagielska, A.; Skolnick, J. Development of a physics-based force field for the scoring and refinement of protein models. *Biophys. J.*, **2008**, *94*, 3227-3240.

[43] Zhou, H.; Skolnick, J. Ab initio protein structure prediction using chunk-TASSER. *Biophys. J.*, **2007**, *93*, 1510-1518.

[44] Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*, **2007**, *69*(Suppl 8), 108-117.

[45] Chivian, D.; Kim, D.E.; Malmstrom, L.; Schonbrun, J.; Rohl, C.A.; Baker, D. Prediction of CASP6 structures using automated Robetta protocols. *Proteins*, **2005**, *61*(Suppl 7), 157-166.

[46] Kolinski, A.; Bujnicki, J.M. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins*, **2005**, *61*(Suppl 7), 84-90.

[47] Tan, Y.H.; Huang, H.; Kihara, D. Statistical potential-based amino acid similarity matrices for aligning distantly related protein sequences. *Proteins*, **2006**, *64*, 587-600.

[48] Henikoff, S.; Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 10915-10919.

[49] Kolinski, A.; Betancourt, M.R.; Kihara, D.; Rotkiewicz, P.; Skolnick, J. Generalized comparative modeling (GENECOMP): a combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. *Proteins*, **2001**, *44*, 133-149.

[50] Clarke, N.D.; Ezkurdia, I.; Kopp, J.; Read, R.J.; Schwede, T.; Tress, M. Domain definition and target classification for CASP7. *Proteins*, **2007**, *69*(Suppl 8), 10-18.

[51] Kopp, J.; Bordoli, L.; Battey, J.N.; Kiefer, F.; Schwede, T. Assessment of CASP7 predictions for template-based modeling targets. *Proteins*, **2007**, *69*(Suppl 8), 38-56.

[52] Jauch, R.; Yeo, H.C.; Kolatkar, P.R.; Clarke, N.D. Assessment of CASP7 structure predictions for template free targets. *Proteins*, **2007**, *69*(Suppl 8), 57-67.

[53] Moult, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 285-289.

[54] Kryshtafovych, A.; Fidelis, K.; Moult, J. Progress from CASP6 to CASP7. *Proteins*, **2007**, *69*(Suppl 8), 194-207.

[55] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.*, **1990**, *215*, 403-410.

[56] Lo Conte, L.; Brenner, S.E.; Hubbard, T.J.; Chothia, C.; Murzin, A.G. SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.*, **2002**, *30*, 264-267.

[57] Orengo, C.A.; Michie, A.D.; Jones, S.; Jones, D.T.; Swindells, M.B.; Thornton, J.M. CATH--a hierarchic classification of protein domain structures. *Structure*, **1997**, *5*, 1093-1108.

[58] Zhou, H.; Zhou, Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins*, **2005**, *58*, 321-328.

[59] Bujnicki, J.M.; Elofsson, A.; Fischer, D.; Rychlewski, L. Structure prediction meta server. *Bioinformatics*, **2001**, *17*, 750-751.

[60] Ginalski, K.; Elofsson, A.; Fischer, D.; Rychlewski, L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **2003**, *19*, 1015-1018.

[61] Hu, J.; Yang, Y.D.; Kihara, D. EMD: an ensemble algorithm for discovering regulatory motifs in DNA sequences. *BMC. Bioinformatics*, **2006**, *7*, 342.

[62] Saini, H.K.; Fischer, D. Meta-DP: domain prediction meta-server. *Bioinformatics*, **2005**, *21*, 2917-2920.

[63] Albrecht, M.; Tosatto, S.C.; Lengauer, T.; Valle, G. Simple consensus procedures are effective and sufficient in secondary structure prediction. *Protein Eng.*, **2003**, *16*, 459-462.

[64] Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **1997**, *25*, 3389-3402.

[65] Sadreyev, R.; Grishin, N. COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. *J. Mol. Biol.*, **2003**, *326*, 317-336.

[66] Pearson, W.R.; Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 2444-2448.

[67] Tress, M.L.; Jones, D.; Valencia, A. Predicting reliable regions in protein alignments from sequence profiles. *J. Mol. Biol.*, **2003**, *330*, 705-718.

[68] Zhang, Z.; Berman, P.; Wiehe, T.; Miller, W. Post-processing long pairwise alignments. *Bioinformatics*, **1999**, *15*, 1012-1019.

[69] Lee, M.; Jeong, C.S.; Kim, D. Predicting and improving the protein sequence alignment quality by support vector regression. *BMC. Bioinformatics*, **2007**, *8*, 471.

[70] Tondel, K. Prediction of homology model quality with multivariate regression. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1540-1551.

[71] Chen, H.; Kihara, D. Estimating quality of template-based protein models by alignment stability. *Proteins*, **2008**, *71*, 1255-1274.

[72] Dayhoff, M.O.; Barker, W.C.; Hunt, L.T. Establishing homologies in protein sequences. *Methods Enzymol.*, **1983**, *91*, 524-545.

[73] Vingron, M. Near-optimal sequence alignment. *Curr. Opin. Struct. Biol.*, **1996**, *6*, 346-352.

[74] Mevissen, H.T.; Vingron, M. Quantifying the local reliability of a sequence alignment. *Protein Eng.*, **1996**, *9*, 127-132.

[75] Vingron, M.; Argos, P. Determination of reliable regions in protein sequence alignments. *Protein Eng.*, **1990**, *3*, 565-569.

[76]    Saqi, M.A.; Sternberg, M.J. A simple method to generate non-trivial alternate alignments of protein sequences. *J. Mol. Biol.*, **1991**, *219*, 727-732.

[77]    Jaroszewski, L.; Li, W.; Godzik, A. In search for more accurate alignments in the twilight zone. *Protein Sci.*, **2002**, *11*, 1702-1713.

[78]    John, B.; Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res.*, **2003**, *31*, 3982-3992.

[79]    Zhang, M.Q.; Marr, T.G. Alignment of molecular sequences seen as random path analysis. *J. Theor. Biol.*, **1995**, *174*, 119-129.

[80]    Schlosshauer, M.; Ohlsson, M. A novel approach to local reliability of sequence alignments. *Bioinformatics*, **2002**, *18*, 847-854.

[81]    Miyazawa, S. A reliable sequence alignment method based on probabilities of residue correspondences. *Protein Eng.*, **1995**, *8*, 999-1009.

[82]    Koike, R.; Kinoshita, K.; Kidera, A. Probabilistic description of protein alignments for sequences and structures. *Proteins*, **2004**, *56*, 157-166.

[83]    Yu, L.; Smith, T.F. Positional statistical significance in sequence alignment. *J. Comput. Biol.*, **1999**, *6*, 253-259.

[84]    Cline, M.; Hughey, R.; Karplus, K. Predicting reliable regions in protein sequence alignments. *Bioinformatics*, **2002**, *18*, 306-314.

[85]    Rangwala, H.; Karypis, G. fRMSDPred: predicting local RMSD between structural fragments using sequence information. *Proteins*, **2008**, *72*, 1005-1018.

[86]    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.

[87]    Laskowski R.A.; MacArthur M.W.; Moss D.S.; Thornton J.M. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **1993**, *26*, 283-291.

[88]    Davis I.W.; Murray L.W.; Richardson J.S.; Richardson D.C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.*, **2004**, *32*, W615-W619.

[89]    Pontius J.; Richelle J.; Wodak S.J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.*, **1996**, *264*, 121-136.

[90]    Hooft R.W.; Vriend G.; Sander C.; Abola E.E. Errors in protein structures. *Nature*, **1996**, *381*, 272.

[91]    Tosatto S.C.; Battistutta R. TAP score: torsion angle propensity normalization applied to local protein structure evaluation. *BMC. Bioinformatics*, **2007**, *8*, 155.

[92]    Morris A.L.; MacArthur M.W.; Hutchinson E.G.; Thornton J.M. Stereochemical quality of protein structure coordinates. *Proteins*, **1992**, *12*, 345-364.

[93]    Colovos C.; Yeates T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, **1993**, *2*, 1511-1519.

[94]    Lovell S.C.; Davis I.W.; Arendall W.B. III.; de Bakker P.I.; Word J.M.; Prisant M.G.; Richardson J.S.; Richardson D.C. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **2003**, *50*, 437-450.

[95]    Bhattacharya A.; Wunderlich Z.; Monleon D.; Tejero R.; Montelione G.T. Assessing model accuracy using the homology modeling automatically software. *Proteins*, **2008**, *70*, 105-118.

[96]    Lee M.R.; Tsai J.; Baker D.; Kollman P.A. Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Biol.*, **2001** , *313*, 417-430.

[97]    Feig M.; Brooks C.L. III. Evaluating CASP4 predictions with physical energy functions. *Proteins*, **2002**, *49*, 232-245.

[98]    Vorobjev Y.N.; Hermans J. Free energies of protein decoys provide insight into determinants of protein stability. *Protein Sci.*, **2001**, *10*, 2498-2506.

[99]    Fogolari F.; Tosatto S.C. Application of MM/PBSA colony free energy to loop decoy discrimination: toward correlation between energy and root mean square deviation. *Protein Sci.*, **2005**, *14*, 889-901.

[100]   Kmiecik S.; Gront D.; Kolinski A. Towards the high-resolution protein structure prediction. Fast refinement of reduced models with all-atom force field. *BMC. Struct. Biol.*, **2007**, *7*, 43.

[101]   Hendlich M.; Lackner P.; Weitckus S.; Floeckner H.; Froschauer R.; Gottsbacher K.; Casari G.; Sippl M.J. Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J. Mol. Biol.*, **1990**, *216*, 167-180.

[102]   Sippl M.J. Recognition of errors in three-dimensional structures of proteins. *Proteins*, **1993**, *17*, 355-362.

[103]   Shen M.Y.; Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, **2006**, *15*, 2507-2524.

[104]   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.

[105]   Lu H.; Skolnick J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins*, **2001**, *44*, 223-232.

[106]   Lu M.; Dousis A.D.; Ma J. OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.*, **2008**, *376*, 288-301.

[107]   Melo F.; Feytmans E. Novel knowledge-based mean force potential at atomic level. *J. Mol. Biol.*, **1997**, *267*, 207-222.

[108]   Pettitt C.S.; McGuffin L.J.; Jones D.T. Improving sequence-based fold recognition by using 3D model quality assessment. *Bioinformatics*, **2005**, *21*, 3509-3515.

[109]   McConkey B.J.; Sobolev V.; Edelman M. Discrimination of native protein structures using atom-atom contact scoring. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 3215-3220.

[110]   Melo F.; Sanchez R.; Sali A. Statistical potentials for fold assessment. *Protein Sci.*, **2002**, *11*, 430-448.

[111]   Eisenberg D.; Luthy R.; Bowie J.U. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, **1997**, *277*, 396-404.

[112]   Terashi G.; Takeda-Shitaka M.; Kanou K.; Iwadate M.; Takaya D.; Hosoi A.; Ohta K.; Umeyama H. Fams-ace: a combined method to select the best model after remodeling all server models. *Proteins*, **2007**, *69*(Suppl 8), 98-107.

[113]   von G.M.; Pas J.; Wyrwicz L.; Ginalski K.; Rychlewski L. Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins*, **2003**, *53*(Suppl 6), 418-423.

[114]   Holm L.; Sander C. Evaluation of protein models by atomic solvation preference. *J. Mol. Biol.*, **1992**, *225*, 93-105.

[115]   Costantini S.; Facchiano A.M.; Colonna G. Evaluation of the structural quality of modeled proteins by using globularity criteria. *BMC. Struct. Biol.*, **2007**, *7*, 9.

[116]   Bartlett G.J.; Taylor W.R. Using scores derived from statistical coupling analysis to distinguish correct and incorrect folds in denovo protein structure prediction. *Proteins*, **2008**, *71*, 950-959.

[117]   Gobel U.; Sander C.; Schneider R.; Valencia A. Correlated mutations and residue contacts in proteins. *Proteins*, **1994**, *18*, 309-317.

[118]   Benkert P.; Tosatto S.C.; Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins*, **2007**, *71*, 261-277.

[119]   Tosatto S.C. The victor/FRST function for model quality estimation. *J. Comput. Biol.*, **2005**, *12*, 1316-1327.

[120]   Qiu J.; Sheffler W.; Baker D.; Noble W.S. Ranking predicted protein structures with support vector regression. *Proteins*, **2008**, *71*, 1175-1182.

[121]   Melo F.; Sali A. Fold assessment for comparative protein structure modeling. *Protein Sci.*, **2007**, *16*, 2412-2426.

[122]   Mereghetti P.; Ganadu M.L.; Papaleo E.; Fantucci P.; De G.L. Validation of protein models by a neural network approach. *BMC. Bioinformatics*, **2008**, *9*, 66.

[123]   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.

[124]   Wallner B.; Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.*, **2006**, *15*, 900-913.

[125]   McGuffin L.J. Benchmarking consensus model quality assessment for protein fold recognition. *BMC Bioinformatics*, **2007**, *8*, 345.

[126]   Eramian D.; Shen M.Y.; Devos D.; Melo F.; Sali A.; Marti-Renom M.A. A composite score for predicting errors in protein structure models. *Protein Sci.*, **2006**, *15*, 1653-1666.

[127]   Soga S.; Shirai H.; Kobori M.; Hirayama N. Identification of the druggable concavity in homology models using the PLB index. *J. Chem. Inf. Model.*, **2007**, *47*, 2287-2292.

[128]   Chang M.H.; DiPiero J.; Sonnichsen F.D.; Romero M.F. Entry to "formula tunnel" revealed by SLC4A4 human mutation and structural model. *J. Biol. Chem.*, **2008**, *283*, 18402-18410.

[129]   Banerjee R.; Pennington M.W.; Garza A.; Owens I.S. Mapping the UDP-glucuronic acid binding site in UDP-glucuronosyltransferase-

1A10 by homology-based modeling: confirmation with biochemical evidence. *Biochemistry,* **2008**, *47,* 7385-7392.

[130]  Balla A.; Tuymetova G.; Toth B.; Szentpetery Z.; Zhao X.; Knight Z.A.; Shokat K.; Steinbach P.J.; Balla T. Design of drug-resistant alleles of type-III phosphatidylinositol 4-kinases using mutagenesis and molecular modeling. *Biochemistry,* **2008**, *47,* 1599-1607.

[131]  Duan J.; Dahlback B.; Villoutreix B.O. Proposed lipocalin fold for apolipoprotein M based on bioinformatics and site-directed mutagenesis. *FEBS Lett.,* **2001**, *499,* 127-132.

[132]  Rotkiewicz P.; Sicinska W.; Kolinski A.; DeLuca H.F. Model of three-dimensional structure of vitamin D receptor and its binding mechanism with 1alpha,25-dihydroxyvitamin D(3). *Proteins,* **2001**, *44,* 188-199.

[133]  Grintsevich E.E.; Benchaar S.A.; Warshaviak D.; Boontheung P.; Halgand F.; Whitelegge J.P.; Faull K.F.; Ogorzalek Loo R.R.; Sept D.; Loo J.A. Mapping the cofilin binding site on yeast G-actin by chemical cross-linking. *J. Mol. Biol.,* **2008**, *377,* 395-409.

[134]  Costanzi S. On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J. Med. Chem.,* **2008**, *51,* 2907-2914.

[135]  Occhipinti E.; Martelli P.L.; Spinozzi F.; Corsi F.; Formantici C.; Molteni L.; Amenitsch H.; Mariani P.; Tortora P.; Casadio R. 3D structure of Sulfolobus solfataricus carboxypeptidase developed by

molecular modeling is confirmed by site-directed mutagenesis and small angle X-ray scattering. *Biophys. J.,* **2003**, *85,* 1165-1175.

[136]  Arastu-Kapur S.; Arendt C.S.; Purnat T.; Carter N.S.; Ullman B. Second-site suppression of a nonfunctional mutation within the Leishmania donovani inosine-guanosine transporter. *J. Biol. Chem.,* **2005**, *280,* 2213-2219.

[137]  Jiang L.; Althoff E.A.; Clemente F.R.; Doyle L.; Rothlisberger D.; Zanghellini A.; Gallaher J.L.; Betker J.L.; Tanaka F.; Barbas C.F. III. De novo computational design of retro-aldol enzymes. *Science,* **2008**, *319,* 1387-1391.

[138]  Rothlisberger D.; Khersonsky O.; Wollacott A.M.; Jiang L.; DeChancie J.; Betker J.; Gallaher J.L.; Althoff E.A.; Zanghellini A.; Dym O. Kemp elimination catalysts by computational enzyme design. *Nature,* **2008**, *453,* 190-195.

[139]  Cozzetto D.; Kryshtafovych A.; Ceriani M.; Tramontano A. Assessment of predictions in the model quality assessment category. *Proteins,* **2007**, *69*(Suppl 8), 175-183.

[140]  Fischer D. Servers for protein structure prediction. *Curr. Opin. Struct. Biol.,* **2006**, *16,* 178-182.

[141]  Pieper U.; Eswar N.; Davis F.P.; Braberg H.; Madhusudhan M.S.; Rossi A.; Marti-Renom M.; Karchin R.; Webb B.M.; Eramian D.; Shen, M.Y.; Kelly, L.; Melo, F.; Sali, A. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.,* **2006**, *34,* D291-D295.