Chapter 4

Assessment of Protein Structure Predictions

E. Capriotti and M. A. Marti-Renom*

4.1 Introduction

Since the beginning of the 1980s, protein structure prediction and simulation have been one of the most challenging tasks for computational structure biology. Although progress has been made, one can openly say that reliably predicting the fold of all known protein sequences is still far from reach. Current approaches can predict a three-dimensional (3D) protein structure for parts of ~60% of the sequences of an average genome.[1–3] Recently, automatic large-scale predictions of 3D structure models are being made available on the web. For example, the ModBase database[2] currently stores more than 4.2 million models, the SwissModel repository[1] stores ~1.3 million models, and the PMDB database[3] stores ~75 000 models. Therefore, comparative protein structure modeling is filling the gap between the known sequence and structure spaces.

In the post-genomic era, a more difficult task lies ahead in annotating, understanding, and modifying the function of proteins. This task is greatly aided by the knowledge of the protein structures, as the biochemical function of a protein is determined by

*Structural Genomics Unit, Bioinformatics Department, Prince Felipe Research Center, Avda. Autopista del Saler, 16 , 46013 Valencia, Spain. Corresponding author Email: mmarti@cipf.es.

its structure and dynamics. In the absence of an experimentally determined structure, 3D models are often valuable for rationalizing existing evidence and guiding new experiments.[4] However, the accuracy of a model determines its utility (Chapter 5), making a means of reliably determining the accuracy of a model an important problem in protein structure prediction.[4,5] Model assessment aims to predict the likely accuracy of a protein structure model in the absence of its known 3D structure.

Model assessment has been previously applied to: (i) determine whether or not a model has the correct fold,[6–9] (ii) discriminate between the native and near-native states,[10–19] and (iii) select the most near-native model in a set of decoys that does not contain the native structure.[16–18,20–23] Several scoring schemes have been developed for these tasks, including physics-based energies, knowledge-based potentials, combined scoring functions, and clustering approaches. Physics-based energy functions are true energy functions describing the interactions acting upon all atoms in a protein structure and are typically developed for and used in molecular dynamic simulations. Statistical or knowledge-based potentials are derived from known protein structures by applying the inverse of the Boltzmann's equation and comparing a system in the thermodynamic equilibrium with the database of folded protein structures. Combined scoring functions usually integrate several different scores with the aim of extracting the most informative features from each of the individual input scores. Finally, the so-called clustering approaches use consensus information from an ensemble of protein structure models provided by one or more methods.

We begin this chapter by introducing the problem of protein structure prediction (Chapter 1). Next, we describe the four main approaches to model assessment. Details describing some of the most widely used scoring function are also provided together with a table of Internet resources for model assessment (Table 4.1). Finally, some of the results from the recent evaluation of model assessment methods carried out at the seventh Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiment are introduced before a final outlook of the future of model assessment.

Table 4.1 A List of URLs to Some Relevant Internet Resources

Title	Ref.	URL
DECOY SETS		
Decoys 'R' Us	(72)	http://dd.compbio.washington.edu
RAPPER	(73)	http://mordred.bioc.cam.ac.uk/~rapper/decoys.php
Skolnick lab	(17)	http://cssb.biology.gatech.edu/skolnick/files/all-atom/
ROSETTA	N/A	http://www.bakerlab.org
Sali lab	N/A	http://www.salilab.org
PHYSICS-BASED ENERGIES		
CHARMM	(26)	http://www.charmm.org
AMBER	(25)	http://amber.scripps.edu
GROMOS	(74)	http://www.igc.ethz.ch/gromos/
KNOWLEDGE-BASED POTENTIALS		
VERIFY3D	(75)	http://nihserver.mbi.ucla.edu/Verify_3D/
TAP	(58)	http://protein.cribi.unipd.it/tap/
FRST	(76)	http://protein.cribi.unipd.it/frst/
ANOLEA	(77)	http://protein.bio.puc.cl/cardex/servers/anolea/
DFIRE	(41)	http://sparks.informatics.iupui.edu/hzhou/dfire.html
PROSA-Web	(78)	https://prosa.services.came.sbg.ac.at/prosa.php
PROQ	(21)	http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi
SIFT	(79)	http://sift.cchmc.org
HOPPscore	(80)	http://hoppscore.lbl.gov/run.html
HARMONY	(81)	http://caps.ncbs.res.in/harmony/

4.2 Protein Structure Prediction

The aim of protein structure prediction is to build a 3D model for a protein of unknown structure (target) either using *ab-initio* methods (i.e. template-free approaches) or on the basis of sequence similarity to proteins of known structure (i.e. template-based approaches such as comparative modeling or threading). Chapter 1 in this book provides a comprehensive introduction to protein structure prediction.

Since the accuracy of a protein structure model determines its usability,[4] two basic conditions must be met to build a useful 3D model. First, an accurate model needs to be built based on the correct template and approximate correct alignment. Second, a reliable score for the model has to be computed to assess its accuracy. Thus, the aim of the second step is to predict errors in models produced in the first step. Next, we outline some of the typical errors in protein structure models. The first two types or errors are specific of template-based approaches while the rest also apply to template-free approaches:

Template selection. The initial step in template-based protein structure prediction is the selection of a template structure. Although selecting the incorrect template is a major error affecting models based on very low sequence identity to their templates (i.e. under ~25% sequence identity), current model assessment methods are usually able to reliably detect it.

Misalignments. One of the largest sources of errors in models from template-based approaches is the incorrect alignment between the target and the template sequences. Such errors affect models based on ~40% or less sequence identity to the closest template(s). The use of multiple sequence alignments, multiple templates and iterative model-building and target-template alignment modification may alleviate such errors.

Template-free modeling. Segments of the target sequence that have no equivalent region in the template structure (i.e. whole protein for template-free modeling or insertions in template-based modeling) are the most difficult regions to model.

Rigid body shifts. As a consequence of sequence divergence there is a natural diversity between two homologous sequences. One type of structural diversity is the rigid distortion of parts of the models. The use of multiple templates may reduce such error.

Side chain packing. The correct packing of side-chain atoms is essential for high-resolution modeling where the resulting models may be

used for docking of small molecules. Therefore, methods for predicting the detailed accuracy of a model are becoming even more important in the advent of a large number of determined structures and the use of models for the docking of small molecules.[24]

Fortunately, during the last decade, the development of more accurate fold assignment and target-template alignment methods together with the use of multiple sources or structural information are mitigating the errors in protein structure models.

4.3 Model Assessment

Protein structure model assessment addresses the general question of how accurate a model is. More specialized questions include: (i) evaluating whether or not the model has the correct fold, (ii) selecting the most accurate model from a set of decoys or alternative solutions, (iii) estimating the overall accuracy of a model (i.e. defining a score that correlates with the RMSD after superimposing a model and its native structure), and (iv) estimating the accuracy of different regions in a model. In the next sections we introduce the four types of available approaches for model assessment, which are used to address some or all of the problems mentioned above: physics-based energies, knowledge-based potentials, combined scoring functions, and clustering approaches. Table 4.1 provides a list of relevant accessible Internet resources.

4.3.1 *Physics-based energies*

Molecular mechanics energy functions with solvation models are the usual components of physics-based energies. Generally speaking, chemical force fields are functional forms encoding a set of parameters for describing the energy of a system of particles. The function and the parameters describing a force field are usually derived both from experimental observations and quantum mechanical calculations. A basic representation of a force field energy function depends on two main contributions: a term describing the energy from chemical bonds between the atoms in the system and a term describing the

interactions between non-bonded atoms in the system. The first term depends on the distances, angles, and dihedral angles between bonded atoms in the molecule. The second term depends on the electrostatic and van der Waals interactions between non-bonded atoms in the molecule. Such energy scores have been classically developed as part of molecular mechanics simulation packages such as AMBER,[25] CHARMM,[26] MM-PBSA,[27] or GROMOS.[28] However, some physics-based approaches, which are outlined next, have also been used for ranking structural decoys.[10,17,19,29–32]

Lazaridis and Karplus applied an effective energy function (EEF1) combining the CHARMM 19 force field with a Gaussian model for solvation free energy to discriminate native structure on a dataset of 650 decoys for six proteins.[10] The results showed that the native state was always more stable than any of the misfolded structures and molecular dynamics simulation reduced the free energy gap between near-native and misfolded structures.

The all-atom version of the Optimized Potential for Liquid Simulations (OLPS),[33] combined with the Surface Generalized Born (SGB) method, was used to discriminate near-native conformations in a set of 49 000 minimized decoy structures for 32 proteins.[18] This energy function was able to correctly identify the native structure within the decoy set in 70% of the tested proteins. The analysis also highlighted the contribution of the solvation free energy in the detection of the native-like structure.

A Molecular Mechanics-Poisson Boltzmann Solvent Accessible Surface Area (MM-PBSA) model was recently used to calculate the free energy of a protein loop structure as a surrogate of the similarity of the decoy to its native structure.[27] The results from such simulations indicated that the MM-PBSA free energy estimator was able to detect native-like structures for 81% of the decoy sets. Moreover the use of the colony energy approach[34] reduce the MM/energy dependency on minor conformational changes. Thus, the authors were able to correlate free energy scores with the root mean square deviation (RMSD) of a decoy set with respect to the native structure.

Recently, Maupetit and colleagues[22] proposed a coarse-grained optimized potential for efficient structure prediction (OPEP). Their

method was able to detect native conformations in 83% of the 29 test proteins with more than 28 000 decoy structures.

Finally, an updated version of the AMBER force field,[17] with terms representing the solvation contribution, was tested for its ability in identifying near-native structures for 150 target proteins within a set of 14 000 decoy structures. The authors concluded that the ability of the method for identifying near-native structures in the decoy set decreased with the time of molecular simulation of the decoys. This version of AMBER was able to detect 100% of near-native structures after only minimizing the structural decoys (i.e. with no molecular mechanics simulation). However, the accuracy decreased to ~70% after a small simulation of 200 picoseconds and to ~30% for 2 nanoseconds simulations. Therefore, such results indicate that molecular mechanics force fields are able to identify near native structures but cannot drive the simulation towards the native conformation of the protein. The authors of the study also concluded that the native structure often does not appear to be in the lowest free energy state.[17] As of today, the refinement problem (i.e. the ability to move the coordinates of a protein structure prediction towards its native conformation) has no generally applicable solution.

In summary, physics-based scoring functions provide good means for selecting near-native structure models in a set of predicted decoys. The introduction of solvation terms clearly improved the ability of such force fields to discriminate between near-native and no-native conformations. However, a universal energy function for model refinement is still far from reach and the relative weight for each energy term contribution may need to be optimized for each decoy set under consideration.

4.3.2 *Knowledge-based potentials*

Statistical potentials, also called potentials of mean force, constitute the main implementation of the knowledge-based potentials for model assessment. In general terms, such potentials encode the statistical preferences of different residues or atom types to be exposed to the solvent, or to interact with each other in a pair-wise or higher

order fashion. Such preferences are normally extracted from a set of selected structures, which represent the known structural space for globular proteins. The basic hypothesis is that protein crystal structures contain a large amount of information describing the stabilizing forces of protein folding, which can be extracted by using the following three assumptions of statistical mechanics: (i) protein folding can be described by a free energy function, (ii) the conformation of a system can be approximated by two-body interactions, and (iii) high frequency conformations should correspond to low free energy structures. If such assumptions are true, it is then possible to derive an atomic energy function for which the global minimum corresponds to the observed native crystal structure.

Since the end of the 1970s, several authors have used such approximations to derive statistical rules from known protein structures.[8,35–45] The main characteristic shared by most knowledge-based potentials is the use of the inverse Boltzmann distribution to derive pseudo-energies from a non-redundant set of protein structures, which states that the probability ($p(x)$) of state x with energy $\varepsilon(x)$ is:

$$p(x) = \frac{1}{Z} e^{-\varepsilon(x)/kT} \tag{1.1}$$

where k is the Boltzmann's constant and T is the absolute temperature. The partition function Z, which can be considered the ground state energy, is defined as:

$$Z = \sum_{x} e^{-\varepsilon(x)/kT} \tag{1.2}$$

Thus, a general representation of the energy function is:

$$\varepsilon(x) = -kT \log\left(\frac{p(x)_{obs}}{p(x)_{\exp}}\right) \tag{1.3}$$

where $p(x)_{obs}$ and $p(x)_{exp}$ are the observed and expected occurrences of the state x respectively. The inverse of the equation (1.3) is then the

pseudo-energy score of the knowledge-based potential, which calculates the energy relative to state x ($\varepsilon(x)$) using the distribution function $p(x)$:

$$\varepsilon(x) = -kT \log p(x) - kT \log Z \tag{1.4}$$

Although there has been debate about the physics basis of statistical potentials,[46–48] it is assumed that the database of protein structures represents the conformational space of globular proteins in thermodynamic equilibrium.

Several types of statistical potentials have been derived which assess different structural features of models. Such potentials include contact,[8,23,37] distance,[16,40,41,45,49] solvent accessibility,[8,42,50] and a combination of solvent accessibility and pair-wise interaction.[16,41,44,45,49,51] Next, we summarize a few particular implementations and applications of knowledge-based potentials for model assessment. Our list is not exhaustive nor complete, but it highlights different approaches for model assessment using knowledge-based potentials. For a recent evaluation and reviews of such methods, see Sec. 4.5 within this chapter and references.[52,53]

Although significant work was done beforehand, knowledge-based potentials became more widely used after the work of Sippl in the beginning of the 1990s.[42,54] Sippl's PROSA, a Cα/Cβ distance-dependence potential that used a poly-protein of 230 different folds for calculating the final Z-score of a model, was originally benchmarked using a set of 163 protein structures. The author concluded that such potentials were accurate for detecting the native structure for most available globular proteins. A new atomic-level statistical potential based on atom-type definitions was later developed by Melo and Feytmans.[49] Using such an approach, it was possible to obtain average frequencies of pair-wise contacts about 15 times higher than the ones obtained using reduced representations for each amino acid. Similarly, Samudrala and Moult[45] developed a residue specific all-atom probability discriminatory ratio (RAPDF), which resulted in a better discrimination of native models compared to other simplified protein representations and illustrated the importance of using a detailed

atomic description of the system. Similar conclusions were obtained by Lu and Skolnick[16] using their heavy-atom potential for discriminating native from near-native structures in a set of decoys. The authors pointed out that their atomic potential tended to pick lower RMSD structures being able to discriminate the native structure in 87% of 119 protein decoy sets. A significant improvement of such atomic-based potentials was later obtained by using a mathematical programming approach.[55] Qiu and Elber compared the performances of their potential with other existing methods, concluding that their potential reached similar accuracy using a much smaller number of parameters.

More recently, the development of alternative reference states (i.e. $p(x)_{exp}$ in equation 1.6) for assessing random interactions has lead to significant improvements in the final accuracy for assessing a protein structure model. For example, a distance-scaled finite ideal-gas reference state was used to derive the DFIRE potential.[41] On average, DFIRE all-atom potential identified the native structure for 84% of the 32 decoys sets used in its benchmark. In a subsequent work, Zhou and co-workers showed that a reduced description of the original DFIRE potential resulted in a similar success rate as its all-atom potential for ranking native structures in a benchmark of 96 decoy sets.[56] Similarly, the DOPE potential,[44] which uses a reference state based on non-interacting atoms in a homogeneous sphere with the radius dependent on a sample native structures, resulted in a higher accuracy than DFIRE, recognizing 87% of the native structures in the 32 decoy sets.

Finally, a new kind of knowledge-based potentials considering the relative orientation of different residue atoms has recently flourished.[57–60] Although the orientation-dependence potentials have not yet been extensively tested, their large-scale application together with coarse-grained protein representations could be very promising.

In summary, knowledge-based potentials, which use empirical observations of proteins of known structures, have proved useful for assessing the accuracy of protein structure models. The use of different reference states together with multi-body representations of protein structures may finally meet the needed accuracy for large-scale protein structure assessment.

4.3.3 *Combined scoring functions*

To improve the accuracy of methods for assessing the accuracy of protein structure models, several scoring functions have been developed using a weighted combination of individual scores from physics and/or knowledge-based approaches.[21,55,60–64] Such scores have been shown to increase the ability to discriminate incorrect models from correct models compared to their individual input scores.[64,65] However, combined scoring functions require the optimization of weights and parameters for each individual input score. As a result, the optimized scoring functions are very dependent on the training set of models used for their derivation. Next, we outline some such approaches developed in the last few years.

The ProQ program implements a neural-network that combines several structural features calculated from the assessed model.[21] Such features include: atom- and residue-based contact potentials, predicted and model secondary structure agreement, solvent accessible surface, fraction of modeled protein, Cα-Cα distance discrepancy between the model and the used template, and protein shape. ProQ was able to detect the correct protein structure model for 62% to 77% of several LiveBench decoy sets.[66]

A Support Vector Machine learning approach was implemented in the SVMod score.[64] SVMod was trained in regression mode taking into account different individual input scores including: three MODPIPE scores, two secondary structure agreement scores, and the DOPE all heavy atom score. The optimal SVMod score was able to select protein structure models on average ~0.45Å apart from the closest model to the native structure in a set of 300 protein structure decoys from 20 target proteins.

Similar to the work by Eramian and co-workers, Benkert and co-workers recently developed a linear combination of individual scores in the QMEAN program.[60] QMEAN combined a coarse-grained torsion angle potential, a secondary structure specific distance-dependent pairwise potential, a solvation potential and two terms accounting for the agreement between the model and predicted solvent accessibility and secondary structure from sequence. The QMEAN score was tested on

a large set of 22 420 models of 95 target proteins from CASP.[67] QMEAN was favorably compared with other existing methods showing a statistically significant improvement in the detection of the native structure and in discriminating between correct and incorrect protein structure models.

In summary, combined scoring functions are able to capture particular structural features from models that may have been detected by each individual score. Therefore, such approaches leverage the input information towards the final goal of detecting the most accurate model in a pool of possible solutions.

4.3.4 *Clustering approaches*

One of the most challenging tasks in protein structure model assessment is to devise a score that correlates with the actual accuracy of the model. One would hope that a perfect scoring function would assign favorable scores to models that are structurally similar to their native structure. Unfortunately, this is not usually the case, and current scoring functions, either physics-based, knowledge-based, or a combination of both, do not always favorably score models close to the native structure. However, when some correlation between the score and the model accuracy exists, structurally comparing all models from independent structure predictions of the same sequence may help in selecting the most accurate model in a set of possible solutions. In other words, an accurate scoring function should more often produce a structural conformation near the native structure than a misfolded structure. This hypothesis has recently been exploited in different implementations of the so-called clustering approaches.[20,68–70]

Shortle and co-workers first applied a clustering approach for predicting the accuracy of models from 10 small proteins in sets of 500 to 1000 ensemble models of low-energy.[20] The authors demonstrated that the conformation with the largest number of models within 4Å RMSD was closer to the native structure than were the majority of models from other clusters in the ensemble. The same approach was later efficiently used in the 4th CASP experiment.[69]

More recently, a new cluster-density method, which weighted the final score of a model using the mean RMSD of its conformation respect to the other ones in the decoy set, was implemented in the self-RAPDF method.[68] The results demonstrated that the use of the density scoring function increased the number of selected near-native conformations from 75% to 92% with respect to the RAPDF method.

A large-scale benchmarking of a clustering-based approach was recently carried out using the SPICKER strategy[70] over a 1489 decoy sets of up to 280 000 models generated by the TASSER program.[62] The results indicated that the top five identified conformations had RMSD values in the top 1.4% of all decoys. The results also indicated that for 78% of the 1489 target proteins, the difference in RMSD from their native conformation to the selected model and RMSD from native to the absolutely best individual model in the decoy was below 1Å.

In summary, the information from an ensemble of decoy conformations can be used to derive statistical probabilities, which facilitate the identification of near-native structures in a set of possible solutions. However, as it is evident from the conceptual implementation of clustering approaches, their final accuracy depends on the quality of the scoring functions used to generate the ensemble of conformations. In other words, an inaccurate scoring function will result in an inaccurate conformation selection by clustering. Another limitation of clustering approaches is their inability to assess the quality of a model on its own.

4.4 Evaluation of Model Quality Assessment Methods

In the seventh edition of the CASP experiment, a new category was introduced with the aim of blindly evaluate methods for model assessment.[52] The new category, named Model Quality Assessment, introduced two measures, evaluating assessment methods at the whole model and at the residue level.

A total of 23 864 models from 95 different target sequences were assessed by 28 different model assessment methods. The model quality

assessors concluded that the Pcons program (65 and CASP7 special issue), which uses the ProQ method for model assessment, was able to constantly select models with near-native conformation. Although not with statistical significance, due to the limited number of groups participating in the residue-by-residue assessment category, the same method seemed able to identify reliable regions of models in a reasonable number of cases. However, as mentioned above, consensus-based methods, such as Pcons, rely in an ensemble of solutions to score individual models. Lee and co-workers (see CASP7 special issue) were able to identify good quality models based only on a non-relative score. Their strategy consisted of blindly relying on their own protein structure predictions. All assessed models were structurally compared with the model of the same target produced by their own method. The final assessment score corresponded to the structural distance of the assessed model to their model.

The strategy by Lee and co-workers resulted in good assessments because their models were consistently accurate for most of the targets in CASP7. Current methods can effectively select accurate models from a set of decoys or ensemble conformations. However, a substantial improvement of methods for evaluating regions of a model as well as assessing the absolute quality of a model on its own is still needed.

4.5 Future Outlook

Despite the large amount of sequence and structure information available and the ever increasing interest of the protein structure community in a reliable 3D structure assessment method, the quest for a perfect scoring function is still open.[53] Currently, the most reliable methods can reasonably select near-native conformations from a decoy set of possible conformations. However, three unsolved tasks lie ahead, to: (i) refine a protein structure model towards its native structure, (ii) reliably predict the absolute accuracy of a model, and (iii) identify regions or residues in a model most likely to contain errors.

During the past years, the most successful approaches for model assessment have relied on either combining individual scores (Sec. 4.3.3)

and/or clustering by structure similarity the resulting ensemble of predicted models (Sec. 4.3.4). However, such approaches do not add to our basic knowledge of the molecular mechanisms by which proteins adopt their native conformations. One can expect that imperfect individual score functions will hamper both, combined and clustering-based scoring functions. Therefore, a more reliable scoring function based on physics (Sec. 4.3.1) and/or empirical observations (Sec. 4.3.2) is clearly needed. Such a scoring function will have to address outstanding open problems such as:

Solvation. One of the major forces towards the folding of a protein is the initial hydrophobic collapse. However, none of the existing methods for protein structure simulation (including those for model assessment) are able to accurately model the effect of water on the protein structure.

Topological determinants. Characterizing the properties of short-to-medium range interactions is needed for elucidating the topological determinants of a protein fold.

Side-chain packing. Although successful methods for protein structure prediction may use a reduced representation of protein structures, which usually simplifies side-chains as single pseudo-atoms, the correct modeling of interaction in the core of proteins will be required for high-resolution protein structure prediction. Such level of details will also be needed for assessing the accuracy of models for protein-protein and/or protein-ligand interactions.

Protein structure flexibility and disorder. Current methods for model assessment do not include data about unstructured parts of proteins Therefore, the use of such information will likely result in more accurate scoring functions for model assessment.

Small, multi-domain, and non-globular proteins. Most of the methods introduced here were developed to assess single domain globular protein models. Therefore, the average accuracy of such methods

significantly drops when applied to models very small or multi-domain proteins and to disordered or transmembrane proteins.

Detailed knowledge-based potentials. The rapid increase of structures deposited in the Protein Data Bank,[71] due in part to the Structural Genomics initiatives, allows the inclusion of multi-body terms in statistical potentials. Such detailed knowledge-based potentials are likely to result in more accurate methods for model assessment.

The increasing interest of the computational structural biologist in addressing such problems and the opportunity to blindly test their methods in an automatic, large-scale, and (hopefully) continuous manner will push the fields of model assessment and protein structure prediction towards very interesting and challenging times.

Acknowledgements

We are grateful to Pascal Benkert, Francisco Melo, and Anna Tramontano for providing access to their work prior to publication. Our research is supported by the European Union (Marie Curie Program FP6-039722), the Spanish Ministerio de Educación y Ciencia (BIO2007/66670), and the Generalitat Valenciana (GV/2007/065).

References

1. Kopp J, Schwede T. (2006) The SWISS-MODEL Repository: New features and functionalities. *Nucl Acids Res* **34**: D315–318.
2. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, Shen MY, Kelly L, Melo F, Sali A. (2006) MODBASE: A database of annotated comparative protein structure models and associated resources. *Nucl Acids Res* **34**: D291–295.
3. Castrignano T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. (2006) The PMDB Protein Model Database. *Nucl Acids Res* **34**: D306–309.
4. Baker D, Sali A. (2001) Protein structure prediction and structural genomics. *Science* **294**: 93–96.

5. Ginalski K, Grishin NV, Godzik A, Rychlewski L. (2005) Practical lessons from protein structure prediction. *Nucl Acids Res* **33**: 1874–1891.
6. Miyazawa S, Jernigan RL. (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* **256**: 623–644.
7. Domingues FS, Koppensteiner WA, Jaritz M, Prlic A, Weichenberger C, Wiederstein M, Floeckner H, Lackner P, Sippl MJ. (1999) Sustained performance of knowledge-based potentials in fold recognition. *Proteins* **Suppl 3**: 112–120.
8. Melo F, Sanchez R, Sali A. (2002) Statistical potentials for fold assessment. *Protein Sci* **11**: 430–448.
9. McGuffin LJ, Jones DT. (2003) Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**: 874–881.
10. Lazaridis T, Karplus M. (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J Mol Biol* **288**: 477–487.
11. Gatchell DW, Dennis S, Vajda S. (2000) Discrimination of near-native protein structures from misfolded models by empirical free energy functions. *Proteins* **41**: 518–534.
12. Vorobjev YN, Hermans J. (2001) Free energies of protein decoys provide insight into determinants of protein stability. *Protein Sci* **10**: 2498–2506.
13. Seok C, Rosen JB, Chodera JD, Dill KA. (2003) MOPED: Method for optimizing physical energy parameters using decoys. *J Comput Chem* **24**: 89–97.
14. Tsai J, Bonneau R, Morozov AV, Kuhlman B, Rohl CA, Baker D. (2003) An improved protein decoy set for testing energy functions for protein structure prediction. *Proteins* **53**: 76–87.
15. Zhu J, Zhu Q, Shi Y, Liu H. (2003) How well can we predict native contacts in proteins based on decoy structures and their energies? *Proteins* **52**: 598–608.
16. Lu H, Skolnick J. (2001) A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins* **44**: 223–232.
17. Wroblewska L, Skolnick J. (2007) Can a physics-based, all-atom potential find a protein's native structure among misfolded structures? I. Large scale AMBER benchmarking. *J Comput Chem* **28**: 2059–2066.
18. Felts AK, Gallicchio E, Wallqvist A, Levy RM. (2002) Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the Surface Generalized Born solvent model. *Proteins* **48**: 404–422.
19. Dominy BN, Brooks CL. (2002) Identifying native-like protein structures using physics-based potentials. *J Comput Chem* **23**: 147–160.
20. Shortle D, Simons KT, Baker D. (1998) Clustering of low-energy conformations near the native structures of small proteins. *Proc Natl Acad Sci USA* **95**: 11158–11162.

21. Wallner B, Elofsson A. (2003) Can correct protein models be identified? *Protein Sci* **12**: 1073–1086.
22. Maupetit J, Tuffery P, Derreumaux P. (2007) A coarse-grained protein force field for folding and structure prediction. *Proteins* **69**: 394–408.
23. Park B, Levitt M. (1996) Energy functions that discriminate X-ray and near native folds from well-constructed decoys. *J Mol Biol* **258**: 367–392.
24. Allen KN. (2007) Form finds function. *Nat Chem Biol* **3**: 452–453.
25. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ. (2005) The Amber biomolecular simulation programs. *J Comput Chem* **26**: 1668–1688.
26. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* **4**: 187–217.
27. Fogolari F, Brigo A, Molinari H. (2003) Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys J* **85**: 159–166.
28. Soares TA, Daura X, Oostenbrink C, Smith LJ, van Gunsteren WF. (2004) Validation of the GROMOS force-field parameter set 45Alpha3 against nuclear magnetic resonance data of hen egg lysozyme. *J Biomol NMR* **30**: 407–422.
29. Lee MC, Duan Y. (2004) Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins* **55**: 620–634.
30. Hsieh MJ, Luo R. (2004) Physical scoring function based on AMBER force field and Poisson-Boltzmann implicit solvent for protein structure prediction. *Proteins* **56**: 475–486.
31. Lee MR, Tsai J, Baker D, Kollman PA. (2001) Molecular dynamics in the endgame of protein structure prediction. *J Mol Biol* **313**: 417–430.
32. Fogolari F, Tosatto SC. (2005) Application of MM/PBSA colony free energy to loop decoy discrimination: toward correlation between energy and root mean square deviation. *Protein Sci* **14**: 889–901.
33. Jorgensen WL, Maxwell DS, Tirado-Rives J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* **118**: 11225–11236.
34. Xiang Z, Soto CS, Honig B. (2002) Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proc Natl Acad Sci USA* **99**: 7432–7437.
35. Tanaka S, Scheraga HA. (1976) Statistical mechanical treatment of protein conformation. I. Conformational properties of amino acids in proteins. *Macromolecules* **9**: 142–159.
36. Bowie JU, Luthy R, Eisenberg D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**: 164–170.

37. Miyazawa S, Jernigan RL. (1985) Estimation of Effective Interresidue Contact Energies from Protein Crystal-Structures — Quasi-Chemical Approximation. *Macromolecules* **18**: 534–552.
38. Levitt M, Chothia C. (1976) Structural patterns in globular proteins. *Nature* **261**: 552–558.
39. Eisenberg D, McLachlan AD. (1986) Solvation energy in protein folding and binding. *Nature* **319**: 199–203.
40. Sippl MJ. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol* **213**: 859–883.
41. 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* **11**: 2714–2726.
42. Sippl MJ. (1993) Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Comput Aided Mol Des* **7**: 473–501.
43. Melo F, Feytmans E. (1998) Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol* **277**: 1141–1152.
44. Shen MY, Sali A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**: 2507–2524.
45. Samudrala R, Moult J. (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J Mol Biol* **275**: 895–916.
46. Finkelstein AV, Badretdinov A, Gutin AM. (1995) Why do protein architectures have Boltzmann-like statistics? *Proteins* **23**: 142–150.
47. Thomas PD, Dill KA. (1996) Statistical potentials extracted from protein structures: How accurate are they? *J Mol Biol* **257**: 457–469.
48. Rooman MJ, Wodak SJ. (1995) Are database-derived potentials valid for scoring both forward and inverted protein folding? *Protein Eng* **8**: 849–858.
49. Melo F, Feytmans E. (1997) Novel knowledge-based mean force potential at atomic level. *J Mol Biol* **267**: 207–222.
50. Jones DT, Taylor WR, Thornton JM. (1992) A new approach to protein fold recognition. *Nature* **358**: 86–89.
51. Zhang C, Liu S, Zhou H, Zhou Y. (2004) The dependence of all-atom statistical potentials on structural training database. *Biophys J* **86**: 3349–3358.
52. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. (2007) Assessment of predictions in the model quality assessment category. *Proteins* **in press.**
53. Skolnick J. (2006) In quest of an empirical potential for protein structure prediction. *Curr Opin Struct Biol* **16**: 166–171.
54. Sippl MJ. (1995) Knowledge-based potentials for proteins. *Curr Opin Struct Biol* **5**: 229–235.

55. Qiu J, Elber R. (2005) Atomically detailed potentials to recognize native and approximate protein structures. *Proteins* **61**: 44–55.
56. Zhang C, Liu S, Zhou H, Zhou Y. (2004) An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state. *Protein Sci* **13**: 400–411.
57. Buchete NV, Straub JE, Thirumalai D. (2004) Orientational potentials extracted from protein structures improve native fold recognition. *Protein Sci* **13**: 862–874.
58. Tosatto SC, Battistutta R. (2007) TAP score: Torsion angle propensity normalization applied to local protein structure evaluation. *BMC Bioinformatics* **8**: 155.
59. Betancourt MR, Skolnick J. (2004) Local propensities and statistical potentials of backbone dihedral angles in proteins. *J Mol Biol* **342**: 635–649.
60. Benkert P, Tosatto SCE, Schomburg D. (2007) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* **in press**.
61. Bradley P, Misura KM, Baker D. (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* **309**: 1868–1871.
62. Zhang Y, Arakaki AK, Skolnick J. (2005) TASSER: An automated method for the prediction of protein tertiary structures in CASP6. *Proteins* **61 Suppl** 7: 91–98.
63. Zhang Y, Skolnick J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* **57**: 702–710.
64. Eramian D, Shen MY, Devos D, Melo F, Sali A, Marti-Renom MA. (2006) A composite score for predicting errors in protein structure models. *Protein Sci* **15**: 1653–1666.
65. Wallner B, Fang H, Elofsson A. (2003) Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller. *Proteins* **53 Suppl 6**: 534–541.
66. Rychlewski L, Fischer D. (2005) LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. *Protein Sci* **14**: 240–245.
67. Kryshtafovych A, Venclovas C, Fidelis K, Moult J. (2005) Progress over the first decade of CASP experiments. *Proteins* **61 Suppl** 7: 225–236.
68. Wang K, Fain B, Levitt M, Samudrala R. (2004) Improved protein structure selection using decoy-dependent discriminatory functions. *BMC Struct Biol* **4**: 8.
69. Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D. (2001) Rosetta in CASP4: Progress in *ab initio* protein structure prediction. *Proteins* **Suppl 5**: 119–126.
70. Zhang Y, Skolnick J. (2004) SPICKER: A clustering approach to identify near-native protein folds. *J Comput Chem* **25**: 865–871.
71. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) The Protein Data Bank. *Nucl Acids Res* **28**: 235–242.
72. Samudrala R, Levitt M. (2000) Decoys 'R' Us: A database of incorrect conformations to improve protein structure prediction. *Protein Sci* **9**: 1399–1401.

73. DePristo MA, de Bakker PI, Lovell SC, Blundell TL. (2003) *Ab initio* construction of polypeptide fragments: Efficient generation of accurate, representative ensembles. *Proteins* **51**: 41–55.
74. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NF, Yu HB. (2006) Biomolecular modeling: Goals, problems, perspectives. *Angew Chem Int Ed Engl* **45**: 4064–4092.
75. Bowie JU, Zhang K, Wilmanns M, Eisenberg D. (1996) Three-dimensional profiles for measuring compatibility of amino acid sequence with three-dimensional structure. *Methods Enzymol* **266**: 598–616.
76. Tosatto SC. (2005) The victor/FRST function for model quality estimation. *J Comput Biol* **12**: 1316–1327.
77. Melo F, Devos D, Depiereux E, Feytmans E. (1997) ANOLEA: A www server to assess protein structures. *Ismb* **5**: 187–190.
78. Wiederstein M, Sippl MJ. (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucl Acids Res* **35**: W407–410.
79. Adamczak R, Porollo A, Meller J. (2004) Accurate prediction of solvent accessibility using neural networks-based regression. *Proteins* **56**: 753–767.
80. Sims GE, Kim SH. (2006) A method for evaluating the structural quality of protein models by using higher-order phi-psi pairs scoring. *Proc Natl Acad Sci USA* **103**: 4428–4432.
81. Pugalenthi G, Shameer K, Srinivasan N, Sowdhamini R. (2006) HARMONY: A server for the assessment of protein structures. *Nucl Acids Res* **34**: W231–234.