Protein Structure Prediction by Comparative Modeling: An Analysis of Methodology

Jennifer Wang, Biochemistry 218
Submitted December 11, 2009

---

1. Introduction

Protein structure determination has become an important area of research in molecular biology and structural genomics. Understanding the tertiary structure of proteins can shed light on protein function and active sites, facilitating site-directed mutagenesis studies and drug design. However, experimental determination of protein structure through X-ray crystallography or NMR spectroscopy remains a difficult and costly process.[1,2] The development of computational methods for protein structure prediction from primary structure information has thus enabled this work to advance more quickly.

Two main approaches to computational structure determination exist: *ab initio* prediction and comparative modeling. *Ab initio* or *de novo* prediction uses physical and chemical first principles to calculate the most favorable protein conformation. Whereas this method is limited by computational power and accuracy, comparative modeling relies on information from previously solved 3D structures in the Brookhaven Protein Data Bank (PDB) and other databases. Also known as homology or template-based modeling, comparative modeling is based on the biological observation that proteins of similar amino acid sequences, usually evolutionarily related, fold into similar 3D structures. Given a protein sequence, it should then be possible to find homologous sequences and, using the homologue structures as templates, predict the folding pattern of the protein in question.

Indeed, current homology modeling techniques have achieved structure predictions of such accuracy that they have been successfully used in drug design, virtual screening, and site-directed mutagenesis applications.[2] Depending on the degree of similarity between target and template sequences, predicted structures are generally within 3.5 Å, sometimes even 1 Å, of experimentally determined structures.[2] Nonetheless, several limitations in the methodology

continue to hinder the acquisition of even more precise structure predictions, particularly for distantly related proteins, making homology modeling a challenging and rapidly evolving field.

The present review will examine the current methods and approaches utilized in comparative protein modeling. An analysis of four representative modeling programs will serve to further demonstrate these methodologies and their performance as accurate structure predictors. Finally, major challenges and obstacles currently facing template-based modeling will be discussed.

2. The Comparative Modeling Procedure

Generally speaking, given a protein sequence of interest, the comparative modeling procedure requires identification of homologous sequences with known structures, alignment of the query sequence to the selected template structure, 3D model construction, and refinement of the predicted model. The actual modeling process is of course much more complex, and the methods employed by various prediction servers to identify suitable templates and structures may widely differ.

Because comparative modeling relies on the presumed similarity in 3D structure between proteins of similar sequences, accurate identification of homologous sequences is crucial to the generation of accurate structural models. Most commonly, this process occurs through comparison of protein sequence profiles or Hidden Markov Models (HMMs). Although simple BLAST searches for homologues by pairwise sequence alignment are also employed, the reliability of results in the "twilight zone" (less than 30% identity with query) begins to dwindle, limiting the usefulness of homologues identified by BLAST. Profile methods, on the other hand, allow for identification of more distantly related proteins, broadening the range of proteins able to be modeled by homology methods. Additionally, use of more advanced methods of sequence comparison, such as those comparing sequences to profiles (PSI-BLAST), profiles to profiles (FFAS), or HMMs to HMMs, contributes greatly to the accuracy of homologue identification. In a benchmark study of the SCOP structure database, PSI-BLAST demonstrated twice the accuracy of BLAST, while FFAS improved upon PSI-BLAST by another 20%. These increases in search specificity eventually translated to greater alignment of the model with the experimental structure, lowering C_α root mean square deviation (rmsd) from 7.8 Å to 4.4 Å.[1]

Once a set of homologues is selected for use as templates, alignment of the query

sequence to the template structures can proceed, often followed by modifications to optimize placement of insertions and deletions outside tight secondary structure elements. Quite often, multiple structures are superposed according to the multiple sequence alignments of the target and template sequences, and the "average" template structure that results forms the basis of the query structure. Whether the use of averaged multiple templates versus a single best template results in higher quality models has been a subject of debate. If utilized properly, multi-template modeling may produce a structure incorporating the best features of each template; alternatively, an average model with intermediate quality will be generated. Use of a single best template will avoid the latter possibility but present the dilemma of determining which single structure is in fact the "best." One study indicated that multi-template modeling tends to be superior to single-template modeling, but the degree of improvement achieved through multiple templates appeared to be low, inconsistent, and difficult to predict.[3] Depending on the modeling program, the manner in which template structures are used may vary widely.

Finally, after alignment of the query sequence with the structural template, construction and refinement of the 3D model of the query protein can occur. With the template structure providing the backbone for the model, model building programs must predict side-chain conformation from specific residue information and the structure of any regions, such as extended loops, that failed to align with the template. While it is possible to model these features by searching structural libraries, loops longer than six or seven residues pose modeling challenges due to their greater range of possible conformations.[4] Consequently, loop prediction algorithms must often employ *ab initio* methods to determine the most energetically favorable conformation. Determination of side-chain packing by database search tends to be a more accurate process, especially if query and template backbone conformations are highly similar. While modeling side-chain conformation exactly is not entirely necessary for an accurate overall structure prediction, its reliability becomes critical if the models are to be used in ligand-binding or active site studies, and energy optimization algorithms may also be employed.[1]

Numerous techniques for comparative model building have been developed, including rigid body assembly, segment matching, satisfaction of spatial restraints, and artificial evolution.[5] Whereas rigid body assembly and segment matching rely more heavily on actual structural elements from the selected templates, the spatial restraints technique represents a more probabilistic approach to modeling. Artificial evolution, the newest method, incorporates energy

minimization principles from *ab initio* prediction to gradually modify the template structure. When used optimally, no single method appears to outperform the others, though some are notably stronger at certain modeling tasks.[6] To better illustrate these different homology modeling approaches, four of the most commonly used programs—SWISS-MODEL, SegMOD, MODELLER, and Nest—will be examined in the following section.

3. Comparative Modeling Programs

3.1 SWISS-MODEL

Developed by Peitsch *et al.* in 1993, SWISS-MODEL (http://swissmodel.expasy.org) is one of the most widely used web-based servers for automated structure prediction, having computed over 120,000 user requests in 2002.[7] The SWISS-MODEL server is complemented by the SWISS-pdb-Viewer (DeepView), a graphical user interface for structure modeling, and the SWISS-MODEL Repository, a database of structures generated by the server and available for download. In addition to a fully automated mode requiring minimum user input (ie. protein sequence only), SWISS-MODEL offers two more advanced user modes in which users can submit their own multiple sequence alignment or manually adjust the modeling parameters.[7,8]

To generate a 3D structure from a provided sequence, SWISS-MODEL utilizes rigid body assembly, the oldest homology modeling technique in which sections from aligned regions of the template are connected together by separately constructed non-conserved regions to form the model backbone. Suitable template structures, those with similar sequences to the query, are first identified by a gapped BLAST search of the SWISS-MODEL template library ExPDB, a subset of PDB. The selected templates are then superposed using an iterative least squares algorithm, the backbone atom positions averaged, and the query sequence fitted to the template to optimize placement of insertion and deletion regions. Fragments that cannot be modeled by homology to the template are computed based on energy considerations or, if the region cannot be solved, searched against a library of loop structures to find an appropriate match. Finally, side chain conformations and intermolecular interactions are adjusted to minimize conformational energy and correct any irregularities in overall 3D structure that resulted from the "cut-and-paste" assembly process.[7-9]

Assessments of the SWISS-MODEL prediction algorithm reveal variable accuracy dependent on the degree of query-template sequence similarity. In a study of 1200 predicted

protein models, the majority of structures demonstrating greater than 40% sequence identity between query and template showed less than 3 Å rmsd from their experimental structures.[10] However, proteins in the "twilight zone" failed to exhibit such modeling accuracy, though this is true of several other prediction programs. Assessment by the server evaluation project EVA-CM (EValuation of Automated protein structure prediction, Comparative Modeling division) showed that on average, SWISS-MODEL predictions demonstrated the lowest overall deviation (2 Å C_α rmsd) from experimental structures when compared to other servers' models. However, this apparently greater accuracy in modeling may be due to the relatively shorter regions that the program modeled in cases of low homology.[9]

Jennifer Wang 12-15-09 6:00 PM
Deleted: (Jennifer: Wht is EVA-CM?)

A more recent benchmark study by Wallner and Elofsson[6] of several automated homology prediction servers, including Modeller, Nest, 3D-Jigsaw, and SegMOD/ENCAD, showed that SWISS-MODEL was relatively poor at producing reliable models. Compared to a <1% failure rate in other programs, SWISS-MODEL was unable to generate predictions for 10% of the provided alignments due to difficulties in loop modeling that crashed the program. SWISS-MODEL also produced more models with poor stereochemistry for difficult query proteins and relatively higher numbers of 3D models that failed to converge (ie. >3 Å rmsd) with the backbone structure, reducing the fidelity of the final model to its original template.

3.2 SegMOD

SegMOD, developed in 1992, is one of the first fully automated homology modeling programs to employ segment match modeling.[11] It is routinely used in conjunction with the energy minimization program ENCAD to build highly accurate models from minimal input data. Unlike the web-based SWISS-MODEL, the SegMOD/ENCAD programs are not publicly available but can be requested for use from its developer Michael Levitt at Stanford University.[6]

The modeling technique behind SegMOD was conceived years before its implementation in a fully automated structure prediction program. Used initially for fitting structures to electron density maps, segment matching is based on the observation that the structure of most penta- or hexapeptide regions of a given protein can be found within other protein structures. Given a vast library of solved structures, such as the PDB, it is possible to build a model of a novel protein, segment by segment, from template fragments presumably similar in conformation. Segment fit is determined by matching inter-C_α distances between query and template, which are calculated

from C_α coordinates derived from highly resolved structures in the PDB. The calculated root mean square deviations and van der Waals energy of the template segments are also measured, and out of several qualifying matches, one is selected for use in the model. The process is repeated using segments chosen at random from the query protein until all residue positions have been solved. Gaps in coordinate information can be solved if there is enough information from flanking regions to identify a plausible fragment from the database.[11]

Two elements in particular appear to contribute to the accuracy and robustness of the segment matching method as executed by SegMOD: averaging of randomly selected segments and energy minimization of the final model. Levitt argues that randomization avoids systematic errors that may accumulate from decisions lacking scientific rationale, such as where to begin modeling.[11] But because SegMOD selects at random which segment to model and, to some extent, which template fragment to use, a number of plausible models are independently generated. Rather than choosing one arbitrary "best" model, the mean coordinates of the structures are used, as in SWISS-MODEL. This method proved to result in even lower root mean square deviations from the known structure. Moreover, the final energy refinement step performed by ENCAD, a method adapted from *ab initio* modeling, corrects any suboptimal side-chain conformations resulting from the piecewise assembly of the model, yielding structures with excellent stereochemistry.

Extensive testing of SegMOD confirms its ability to generate highly reliable and robust models. The author's own analysis of structure predictions for eight different proteins revealed an average overall backbone rmsd from the crystal structure of approximately 1.65 Å and a considerably low rmsd of 0.97 Å for side-chain torsion angles. In comparison, the overall rmsd for the mean models of each protein, the result of averaging ten structures together, was only 1.19 Å, illustrating the significant contribution of the coordinate averaging step to improving model quality. It was also found that the accuracy of the final model was not highly dependent on the accuracy of the coordinates of the input models, also likely a consequence of the averaging step.[11] The Wallner and Elofsson benchmark study concluded that SegMOD was among the more highly performing predictors because of its speed, accurate stereochemistry, and ability to generate the most "acceptable" structures (<10% bad residues). The program's primary criticism in this study was faulty backbone conformation, though its shortcomings in this area were still less than those of other programs.[6]

Jennifer Wang 12-15-09 6:30 PM
Deleted: SegMOD accomplishes this task by first reading in the C_α coordinates of highly resolved structures from PDB and then searching this coordinate database for fragments that fit the spatial arrangement of the segment currently being modeled (Jennifer, this is not clear. You don't know the structure of your protein so how can you compare the coordinate database for fragments that fit the spatial arrangement of the segment being modeled if you don't know the structure of the segment being modeled???) . Segment fit is determined by matching inter-C_α distances between query and template and selecting one database segment from among several with the lowest calculated root mean square deviations and van der Waals energy (ahh now it is clear. You should rewrite previous two sentences to eliminate my perceived problem after sentence one). Once a match is found,

Jennifer Wang 12-15-09 6:34 PM
Deleted: another

Jennifer Wang 12-15-09 6:34 PM
Deleted: is

Jennifer Wang 12-15-09 6:33 PM
Deleted: and the process repeated

3.3 MODELLER

Perhaps the most widely implemented structure prediction program is MODELLER.[12] Though not a web-based homology modeling server, the MODELLER software package is available for download from the developers' website (http://salilab.org/modeller), enabling its use with other structure prediction servers. Indeed, MODELLER has become the model-building program of choice for several homology modeling servers because of its relative speed and reliability. Several of the strongest performing prediction servers in the CASP8 experiment, such as HHpred, incorporate MODELLER in their methodology.[13,14]

Whereas SWISS-MODEL employs rigid body assembly in its construction of the protein backbone, MODELLER builds 3D structures that must satisfy certain spatial restraints, including C_α-C_α bond length, main-chain and side-chain dihedral angles, and van der Waals interactions. These restraints are expressed as probability density functions representing the probability of occurrence of a certain conformation and are calculated from the structures of homologous template sequences. Compilation of the probability density functions from individual template structures into a single representative probability density function, followed by optimization of this function, then describes the overall spatial restraints for the query protein. As a result, the predicted model represents the most probable structure of the protein in question, based on structural information derived from known homologues.[12]

Like SWISS-MODEL and SegMOD, MODELLER is fully automated and can generate a 3D model starting from a user-provided protein sequence. The sequence is converted into a profile, which is searched against a MODELLER-specific database of nonredundant PDB sequences, each of which is representative of a closely related (>95% identity) set of PDB sequences. Homologous sequences returned by the profile search are submitted to a multiple sequence alignment, which in turn is used to construct a multiple structure alignment to identify the optimal template structures. These template structures are finally used to calculate the spatial restraints from which the predicted model is constructed.[12,15,16]

When compared against other homology modeling programs in the Wallner and Elofsson study, MODELLER was considered one of the better performing structure predictors. MODELLER was highly successful in generating reliable structures with accurate backbone dihedral angles and produced a relatively constant proportion of residues with poor stereochemistry across all levels of sequence identity. The greatest shortcoming of the program

lay in its ability to accurately model side-chain torsion angles, producing only 30% correct rotamers in sequences demonstrating <50% identity. Like SWISS-MODEL, MODELLER also failed to generate convergent models for a much higher proportion of the tested sequences than other programs. However, it was found that simply rerunning MODELLER with different parameters could overcome the convergence problem.[6]

3.4 NEST

Implementation of yet another homology modeling technique, artificial evolution, was accomplished in 2001 at Columbia University with the program NEST.[17] Now a part of the JACKAL software package, NEST combines template-based methods with *ab initio*-like energy minimization principles to generate highly accurate 3D models. The loop and side-chain modeling programs LOOPY and SCAP have been integrated into NEST, enabling NEST to fully generate and refine structural predictions starting with user-provided sequence alignments. Like MODELLER, the JACKAL package, which includes the core program NEST, a graphical environment, and several supporting prediction tools, is available for free download from the developer's website.[18]

The goal of the NEST algorithm is to model the gradual process of molecular evolution from a homologous "ancestral" protein to the current protein of interest. Whereas SWISS-MODEL and SegMOD consider the template and query proteins in terms of regional similarity, and MODELLER considers the proteins more in their entirety, NEST considers each difference from the template sequence to be an independent evolutionary event—either a point mutation, an insertion, or a deletion. Assuming the template structure to be the ancestral state, a single change in alignment is introduced and the structure recalculated using energy minimization and molecular dynamics principles. This process is repeated until all differences between query and template have been accounted for. Since the actual order of these evolutionary events cannot be determined, NEST begins with the least energetically costly changes, usually point mutations on the protein surface, to facilitate the stepwise optimization process. Once all necessary "mutations" have been made, the program proceeds with model refinement.[17]

In addition to this new approach to homology modeling, the NEST method introduces the idea of colony energy in exploring the conformational space of residue side-chains and loops. Colony energy takes into account the force field energy, such as local van der Waals interactions,

as well as the entropic favorability or likelihood of a given conformation,[19] and its use in determining loop and side-chain structure represents a departure from the loop library search used in SWISS-MODEL. Although the loop modeling program employed by NEST is considered an *ab initio* method, NEST itself still relies on template sequences—either a single one, multiple ones, or a composite[18]—to begin the model building process.

Assessment of NEST-generated structures by Wallner and Elofsson[6] demonstrated that NEST performed equally well as Modeller and SegMOD and better than the other tested methods. Faster than Modeller, NEST did not crash or encounter convergence problems as frequently and also generated the most models with a near-optimal score based on MaxSub assessment criteria. Notably, NEST was unlikely to deteriorate the model compared to the template, doing so in only 5% of cases, half as often as Modeller. The main flaw identified in the study was poor model stereochemistry, though the fraction of bad residues was constant across all levels of sequence identity and was comparable to that observed in other homology modeling programs.

4. Current Challenges and Limitations

Major strides have been made in the homology modeling field over the past two decades. From CASP1 in 1994 to the most recent CASP8, it is clear that template-based methods of structure prediction have gained popularity over *ab initio* modeling because of their greater success in achieving models that approach the quality of high-resolution crystal structures.[20] However, despite this progress, researchers continue to confront several challenges and shortcomings in the comparative modeling methodology that limit their ability to further increase model quality at all levels of homology.

The first of such challenges lies in accurate sequence-structure alignment and optimal template selection, especially in distant evolutionary cases. It is generally agreed that in the "twilight zone" of sequence identity, the inability of current methods to identify proper homologues and align relatively dissimilar sequences leads to highly unreliable structures. While improvements in alignment programs like PSI-BLAST have led to a gradual improvement in remote homology modeling over the lifetime of CASP, there are still intrinsic limitations in gleaning structural information from the relationship between distant homologues. Though evolutionarily related structures may share similar features and equivalent regions, a decreasing

degree of close identity will introduce more differences in structure and increase the likelihood that a large proportion of target residues, up to 50% for the most distant homologues, cannot be mapped to the template.[20] These difficult cases are not impossible to model well, especially if human intervention is used,[21] but developing automated methods for remote homology modeling will certainly be an important area of future focus.

The next major obstacle lies in the refinement or "defrosting" of predicted models, considered the bottleneck step in comparative modeling.[20] This is particularly true for proteins at the medium and high identity levels, for which template identification and sequence-structure alignment are not nearly as problematic. Oftentimes, refinement attempts lead to deterioration in model quality compared to the original template, but adjustments to backbone conformation and side-chain stereochemistry are essential to generating near-native models. Like SegMOD and NEST have shown, energy minimization and molecular dynamics techniques present the most promise in this area. Nonetheless, utilization of these energy-based methods requires surmounting of the computational obstacles faced by *de novo* modelers, such as the size of the conformational sampling space and insufficient information about interatomic forces.[20,22]

Another factor suggested to contribute to suboptimal modeling of high-identity sequences is the amount of structural information available. Even when template selection and alignment were optimized, one study found that most of the CASP5 targets failed to achieve CASP assessment scores above 70%.[23] The authors attributed this shortcoming to the lack of information necessary to build a reliable model, which is certainly the case for remote homology modeling. Current estimates place the number of protein sequences in databases at two orders of magnitude greater than the number of distinct protein structures in the PDB.[2] Expansion of the structure database to include more diverse templates that may be closely related to the query protein would help to remedy this information shortage, although one assessment of CASP6 results revealed that this was not the case.[24]

A different perspective is that the ability to recognize distinct tertiary structures from available information, rather than the ability to acquire suitable templates from a more comprehensive pool, is in need of greater attention. In contrast to the remote homology modeling situation, where the challenge is determining which sequences are most related, certain related proteins may exhibit similar sequences but have divergent structures, a setback encountered in CASP6. Changes such as β-strand inversions and hairpin flips, circular

permutations, and fusion of duplicated domains may go undetected by current modeling methodology, which rely primarily on sequence alignment to infer structure.[22] Though there may not be a straightforward solution to this problem, ongoing research in the fold recognition community may point to more sensitive methods of obtaining subtle structural information from sequences.

5. Future Directions

Although comparative modeling has been highly successful in advancing protein structure analysis *in silico*, it is clear that there is still much work to be done before the goal of experimental-quality structure predictions can be realized. Given the limitations of template-based modeling, combining traditional homology modeling techniques with other computational methods for protein structure determination will most enable the field to progress.

Indeed, the power of new, more integrative approaches is apparent in the top performers of the CASP8 experiment. While many of the best prediction servers continue to use basic techniques employed by homology modeling, including profile-profile searches and the MODELLER program, they often rely heavily on protein threading, fold recognition, fragment assembly, and other novel techniques to achieve high-quality results. Most notable is the recent emergence of meta-servers such as 3D-Jury and Pcons, which compile the results of other homology prediction servers to generate their own structure predictions. The ability of meta-servers to outperform individual autonomous servers in recent trials[13,25] is indicative of their success and future potential, though it has been argued that their use has not contributed significantly to our understanding of protein structure and prediction but rather represents a more practical means of achieving a desired end.[5]

With the next installment of CASP quickly approaching in 2010, it remains to be seen how far template-based modeling has progressed since CASP8. The insights to be gained from CASP9 and other future community-wide assessments will be important indicators of how much further we must go to see computationally determined protein structures one day rivaling their experimentally determined counterparts.

References

1. Jaroszewski, L. (2009) Protein structure prediction based on sequence similarity. *Methods Mol Biol*, **569**, 129-56.
2. Kopp, J., and Schwede, T. (2004) Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, **5**(4), 405-16.
3. Larsson, P., Wallner, B., Lindahl, E., Elofsson, A. (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci*, **17**, 990-1002.
4. Xiang, Z. (2007) Homology-based modeling of protein structure. In Y. Xu, D. Xu, and J. Liang (Eds.), *Computational Methods for Protein Structure Prediction and Modeling: Basic Characterization* (pp.319-57). New York: Springer.
5. Guo, J., Ellrott, K., Xu, Y. (2008) A historical perspective of template-based protein structure prediction. *Methods Mol Biol*, **413**, 3-42.
6. Wallner, B., and Elofsson, A. (2005) All are not equal: a benchmark of different homology modeling programs. *Protein Sci*, **14**, 1315-27.
7. Guex, N., and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-23.
8. Schwede, T., Kopp, J., Guez, N., and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, **31**(13), 3381-5.
9. Arnold, K., Bordoli, L., Kopp, J., Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics*, **22**(2), 195-201.
10. Peitsch, M.C., Schwede, T., Guex, N. (2001) Automated protein modeling – the proteome in 3D. *Pharmacogenomics*, **1**(3), 257-66.
11. Levitt, M. (1992) Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol*, **226**, 507-33.
12. Sali, A., and Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol*, **234**, 779-815.
13. Cozzetto, D., Kryshtafovych, A., Fidelis, K., Moult, J., Rost, B., Tramontano, A. "Evaluation of template-based models in CASP8 with standard measures." (2009) *Proteins*, **77**(Suppl 9), 18-28.

14. Hildebrand, A., Remmert, M., Biegert, A., Soding, J. (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins*, **77**(Suppl 9), 128-32.
15. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins*, **23**, 318-26.
16. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A. (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, Chapter 2:Unit 2.9.
17. Petrey, D., Xiang, Z., Tang, C.L., Xie, L., Gimpelev, M., Mitros, T., Soto, C.S., Goldsmith-Fischman, S., Kernytsky, A., Schlessinger, A., Koh, I.Y.Y., Alexov, E., Honig, B. (2003) Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins*, **53**, 430-5.
18. Xiang, J.Z. (2002) JACKAL: a protein structure modeling package. Retrieved 6 Dec 2009 from http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal_Manual
19. Xiang, Z., Soto, C.S., Honig, B. (2002) Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *PNAS*, **99**(11), 7432-37.
20. Moult, J. (2005) A decade of CASP: progress, bottlenecks, and prognosis in protein structure prediction. *Curr Opin Struc Biol*, **15**, 285-89.
21. Venclovas, C., and Margelevicius, M. (2005) Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment." *Proteins*, **61**(Suppl 7), 99-105.
22. Ginalski, K. (2006) Comparative modeling for protein structure prediction. *Curr Opin Struc Biol*, **16**, 172-77.
23. Contreras-Moreira, B., Ezkurdia, I., Tress, M.L., Valencia, A. (2005) Empirical limits for template-based protein structure prediction: the CASP5 example. *FEBS Letters*, **579**: 1203-1207.
24. Tress, M., Ezkurdia, I., Grana, O., Lopez, G., Valencia, A. (2005) Assessment of predictions submitted for the CASP6 comparative modeling category. *Proteins*, **61**(Suppl 7), 27-45.
25. Fischer, D. (2006) Servers for protein structure prediction. *Curr Opin Struc Biol*, **16**, 178-82.