PERSPECTIVE

BIOTECHNOLOGY and BIOENGINEERING

Computational Methods in Protein Structure Prediction

C.A. Floudas

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263; telephone: +1-609-258-4595; fax: +1-609-258-0211; e-mail: floudas@titan.princeton.edu

Published online 23 April 2007 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21411

ABSTRACT: This review presents the advances in protein structure prediction from the computational methods perspective. The approaches are classified into four major categories: comparative modeling, fold recognition, first principles methods that employ database information, and first principles methods without database information. Important advances along with current limitations and challenges are presented.

Biotechnol. Bioeng. 2007;97: 207–213.

KEYWORDS: protein structure prediction; protein folding; computational methods

Introduction

Protein structure prediction from amino acid sequence is a fundamental scientific problem and it is regarded as a grand challenge in computational biology and chemistry. Given an amino acid sequence (i.e., the primary structure) which represents a monomeric globular protein in aqueous solution and at physiological temperatures, the objectives are to determine (i) all helical segments and all beta-strands, (ii) all pairs of beta-strands which form beta-sheets (i.e., the beta-sheet topology), (iii) all disulfide bridges if cysteines are present, (iv) all loops that connect secondary structure elements, and (v) the three-dimensional folded protein structure.

The protein structure prediction problem has attracted the interest of many researchers across several disciplines. Several viewpoints provide competing explanations to the protein folding question. The classical viewpoint regards folding as a hierarchical process, implying that the process is initiated by rapid formation of secondary structural elements, followed by the slower arrangement of the actual three dimensional structure of the tertiary fold (e.g., Baldwin and Rose, 1999). An opposing perspective is based on the idea of a hydrophobic collapse, and suggests that the tertiary and secondary features form concurrently. Another perspective has also emerged that combines components of the aforementioned two viewpoints, that is (a) local interactions are responsible for the formation of helices and beta strands, (b) hydrophobic long range interactions are responsible for the formation of beta-sheets and their topologies, and (c) the combination of induced restraints from (a) and (b) drive the protein into its folded structure (e.g., Floudas et al., 2006; Klepeis and Floudas, 2003c).

According to Anfinsen's (1973) thermodynamic hypothesis, proteins are not assembled into their native structures by a biological process, but folding is a purely physical process that depends only on the specific amino acid sequence of the protein and the surrounding solvent. Many approaches to computational protein structure prediction using first principles have been developed that are based on Anfinsen's thermodynamic hypothesis.

Progress for all variants of computational protein structure prediction methods is assessed in the biannual, community-wide Critical Assessment of Protein Structure Prediction (CASP) experiments (Moult et al., 1997, 2001, 2003, 2005; Moult, 1999). In the CASP experiments, research groups apply their prediction methods to amino acid sequences for which the native structure is not known but to be determined and to be published soon. Even though the number of amino acid sequences provided by the CASP experiments is small, these competitions provide a good measure to benchmark methods and progress in the field in an arguably unbiased manner (Murzin, 2004). A review on the progress and challenges, based on a decade of CASP 1–5 events, can be found in Moult (2005).

Correspondence to: C.A. Floudas

Classification of Protein Structure Prediction Methods

Computational methods for protein structure prediction can be classified into four groups: (1) comparative modeling, (2) fold recognition, (3) first principles methods with database information, and (4) first principles methods without database information (e.g., see recent review (Floudas et al., 2006)). In the sequel, we will discuss key contributions and advances in the aforementioned four classes.

Comparative Modeling

Comparative modeling relies on the principle that sequences which are related evolutionarily exhibit similar three dimensional folded structures, that is sequence similarity suggests structural similarity. With this as a guiding principle, comparative modeling consists of five main stages: (a) identification of related sequences of known structure; (b) aligning of the target sequence to the template structures; (c) modeling of structurally conserved regions using the known templates; (d) modeling side chains and loops which are different than the templates; (e) refining and evaluating the quality of the model through conformational sampling.

The accuracy of predictions by comparative modeling depends on the degree of sequence similarity. If the target and the template sequence have more than 50% of their sequences, predictions are of very good to high quality and have been shown to be as accurate as low-resolution X-ray predictions (Kopp and Schwede, 2004).

For 30–50% sequence identity, more than 80% of the Cα-atoms can be expected to be within 3.5 Å of their true positions (Kopp and Schwede, 2004), while for less than 30% sequence identity, the prediction is likely to contain significant errors (Kopp and Schwede, 2004; Vitkup et al., 2001).

Recent methods for comparative modeling have departed from the traditional domain of (i) sequence–sequence comparison to (ii) profile–sequence comparison, to (iii) sequence–profile comparison, as well as to (iv) profile–profile comparison. In (ii) and (iii), profiles are generated using position-specific substitution matrices and the main hypothesis is that the alignment of conserved motifs is prevalent. In (iv), profiles are compared via direct alignment.

A number of recent reviews that focus primarily on comparative modeling approaches include (Dunbrack, 2006; Ginalski et al., 2005; Ginalski, 2006; Petrey and Honig, 2005), and are also discussed in Floudas et al. (2006).

Assessors of homology methods in CASP5 pointed out that the general approach to structure prediction by comparative modeling has not changed over the last two decades (Tramontano and Morea, 2003). Tress et al. (2005), in their assessment of comparative modeling in CASP6, pointed out that even though there is little improvement with respect to CASP5, there were a few groups which approached the quality of the best templates for easy and difficult targets.

This was also evidenced by the very good performance of a few groups during CASP7. A recent advance for automated comparative modeling is the TASSER-Lite tool, which is based on an extension of the TASSER approach, discussed in Section "First Principles Prediction With Database Information," for homologous sequences (Pandit et al., 2006). This development takes advantage of the homologous sequences and is based on custom-made parameter optimization, which leads into significant reduction of the computational time while it maintains high quality predictions.

Fold Recognition and Threading

Fold recognition methods are based on the principle that the number of different folded protein structures is significantly more limited than the vast number of different sequences generated out of the genome projects. The number of different folds has been estimated based on clustering the structures deposited in the protein data bank (Berman et al., 2000) into families. A recent study and assessment revealed that the protein data bank already contains enough structures to cover small protein structures up to a length of about 100 residues (Kihara and Skolnick, 2003). Approaches for fold recognition include (a) advanced sequence similarity/comparison methods, and (b) secondary structure prediction and comparison of sequences. Secondary structure information is frequently used in combination with other one-dimensional descriptors in fold recognition methods. Przybylski and Rost (2004) showed that an approach, which uses secondary structure information and solvent accessibility can outperform methods that utilize three-dimensional structure data.

Threading methods aim at fitting a target sequence to a known structure in a library of folds. Skolnick and coworkers developed and successfully applied threading methods in the CASP5–7 experiments (Skolnick et al., 2003; Zhang et al., 2005). Their recent advance (Skolnick et al., 2004) is an iterative approach that first aligns target and known structures ignoring pairwise residue interactions. In subsequent alignments, information from previous alignments is then used to evaluate pairwise interaction energies. The method combines three different pair potentials to account for the fact that different scoring functions are capable of assigning different target sequences to the same template. By identifying structurally similar regions in multiple templates, accurate regions of structure prediction can be distinguished from less accurate ones. Skolnick et al. (2004) found that accurate fragments can be identified even if no template is convincing as a whole. This observation led to the development of a fragment assembly method based on their threading approach (Zhang and Skolnick, 2004a,b). Xu and Xu (2000) developed a threading

DOI 10.1002/bit

algorithm that considers pair contacts between α-helices and β-strands and allows for alignment gaps in loop regions. The method furthermore allows to incorporate constraints about a target protein such as known disulfide bonds or distance restraints (Xu and Xu, 2000). Another approach poses the fold recognition problem as a global optimization of an energy function (Xu et al., 2003), and it is formulated as an integer programming problem, and solved approximately as a linear programming relaxation. The review by Floudas et al. (2006) provides a more detailed account of the fold recognition and threading contribution up to early 2005.

Wang et al. (2005), in their assessment of fold recognition methods in CASP6, pointed out that there has been notable improvement in the fold recognition methods. The top-ranked methods for the FR/Homologous targets used servers and consensus-based metrics, while the top-ranked methods for FR/Analogous targets used fragment-building methods in addition to fold recognition meta servers.

First Principles Prediction With Database Information

First principles methods that utilize database information can be further classified as: (a) fragment-based recombination methods; (b) hybrid methods that combine multiple sequence comparison, threading, MC optimization with scoring functions, and clustering; and (c) methods that combine information from secondary structure and selected tertiary restraints with MC optimization or deterministic global optimization.

In (a), fragment-based recombination methods, the fundamental principle is that sequence-dependent local interactions direct the chain to sample specific sets of local conformers, while non-local interactions prefer low free energy conformers which are compatible with the biased local conformers.

Baker and co-workers studied the distributions of local structures based on short sequence segments of up to 10 residues based on the protein database, and developed effective approaches that compare fragments of a target to fragments of known structures. Once appropriate fragments have been identified, they are assembled to a structure, often with the aid of scoring functions and optimization algorithms. Compact structures can be assembled by randomly combining fragments using simulated annealing (Rohl et al., 2004; Simons et al., 1997). Subsequently, the fitness of a conformation can be assessed with scoring functions derived from conformational statistics of known proteins. Incorporation of information from independently conducted secondary structure predictions has resulted in improved scoring functions (Simons et al., 1999). Recent reviews that focus on the fragment-based recombination methods include Baker (2006) and Bujnicki (2006), and are also discussed in Floudas et al. (2006).

In (b), hybrid methods, Skolnick, Kolinski, and coworkers (Skolnick et al., 2001, 2003; Zhang and Skolnick, 2004a,b) developed approaches that combine multiple sequence comparison, threading, optimization with scoring functions, and clustering. The method uses a united atom lattice model with three or fewer atoms per residue (Zhang et al., 2003). Threading is used to provide information on long-range interactions by identifying contacts between distant side-chains. Clustering and selecting centroids of most populated clusters results in conformers closer to native than the lowest energy conformers (Zhang and Skolnick, 2004c). The hierarchical approach TASSER (Zhang et al., 2005) that combines template identification through threading, parallel hyperbolic MC sampling structure assembly via rearranging continuous template fragments, clustering using SPICKER (Zhang and Skolnick, 2004c), and post-analysis using the TM scores, was introduced and applied in CASP6 and CASP7. Kolinski, Bujnicki, and coworkers (Kosinski et al., 2003, 2005; Kolinski and Bujnicki, 2005) also developed an alternative hybrid method which generates initial models using the reduced lattice approach CABS (Kolinski, 2004), scores them using Verify3D to identify good quality folded fragments and uses them to derive tertiary restraints that enhance the MC sampling of decoys which are subsequently clustered for the final structure selection.

In (c), Friesner and co-workers developed a method that introduces information on secondary structure and selected tertiary restraints and uses the principles of the deterministic αBB global optimization method (Adjiman et al., 1996, 1997, 1998a,b; Androulakis et al., 1995; Floudas, 2000, 2005; Floudas et al., 2005) in combination with a reduced force field model (Eyrich et al., 1999a,b).

First Principles Prediction Without Database Information

First principles protein structure prediction methods that do not utilize database information attempt to identify the lowest free energy structure of the protein in its environment using only physics laws and the amino acid sequence. This class of methods can be applied to any given target sequence using only physically meaningful potentials and atomic level representations.

With such a broad range of targets and the inability to directly or indirectly apply database information, these methods are the most difficult of the protein structure prediction methods.

Rose and co-workers introduced a hierarchical approach to structure prediction (LINUS) that emphasizes the important role of local steric effects and conformational entropy (Srinivasan and Rose, 1995, 2002). Using a Monte Carlo algorithm, the approach identifies protein conformational biases through a discrete set of moves and a simplified physics-based force field.

Scheraga and co-workers introduced a hierarchical approach to this problem that uses a simplified united-residue force field for initial calculations and then

subsequent refinement of the coarse model using an all-atom potential (Lee et al., 2001; Liwo et al., 1997a,b, 2002, 2001; Pillardy et al., 2001). This united-residue force field, which reduces the representation of each amino acid residue to just two interaction sites, allows the stochastic conformational space annealing algorithm to more efficiently identify low energy structures (Lee et al., 1997, 1998, 2000; Lee and Scheraga, 1999; Ripoll et al., 1998). Recent work has focused on improved algorithms to handle β-strands (Czaplewski et al., 2004a), detailed analysis of the role of disulfide bonds in protein structure (Czaplewski et al., 2004b), and the introduction of replica-exchange MC with minimization for the united residue force field (Nanias et al., 2005).

Floudas and co-workers introduced a first principles physics-based method, ASTRO-FOLD (Klepeis and Floudas, 2003c), which combines the classical and new views of protein folding. This approach identifies first helical regions through detailed free energy calculations (Klepeis and Floudas, 1999), and the application of global optimization methodologies (Klepeis and Floudas, 2002). The prediction of β-strands and β-sheet topologies are addressed via a novel mixed-integer linear optimization formulation to maximize hydrophobic interactions (Klepeis and Floudas, 2003a). To predict the ensemble of loop conformers, free energy calculations are used (Klepeis and Floudas, 2005). The structure prediction of loops with flexible stems has been recently enhanced via extensive sampling and clustering (Monnigmann and Floudas, 2005).

Using the secondary structure predictions and the loop structure predictions to develop restraints, the tertiary structure is identified using a novel class of hybrid global optimization algorithms (Klepeis et al., 2003a,b) in the spirit of an NMR structure refinement protocol using predicted restraints (Klepeis et al., 1999). The article by Klepeis et al. (2002), and the recent review by Floudas et al. (2006) provides a more detailed account of the first principles-based approaches including the ASTRO-FOLD framework. A challenging double blind structure prediction for a four-helical bundle protein of 102 amino acids, denoted as S824, with rmsd of 5.18 angstroms for the lowest energy conformer compared to the NMR structure was reported in Klepeis et al. (2005).

Recent enhancements of the ASTRO-FOLD framework include the development of a predictive approach for the interhelical distances in α helical proteins (McAllister et al., 2006), and the development of a distance dependent Ca–Ca force field that can discriminate effectively the folded structure form high-resolution decoys (Rajgaria et al., 2006). The enhanced ASTRO-FOLD approach, which includes the recent advances on the prediction of interhelical restraints and the distance dependent force fields, is shown in Figure 1. A recent successful blind prediction of a designed protein sequence of 102 amino acids, denoted as S836, provided by the Hecht group is shown in Figure 2. This four-helical

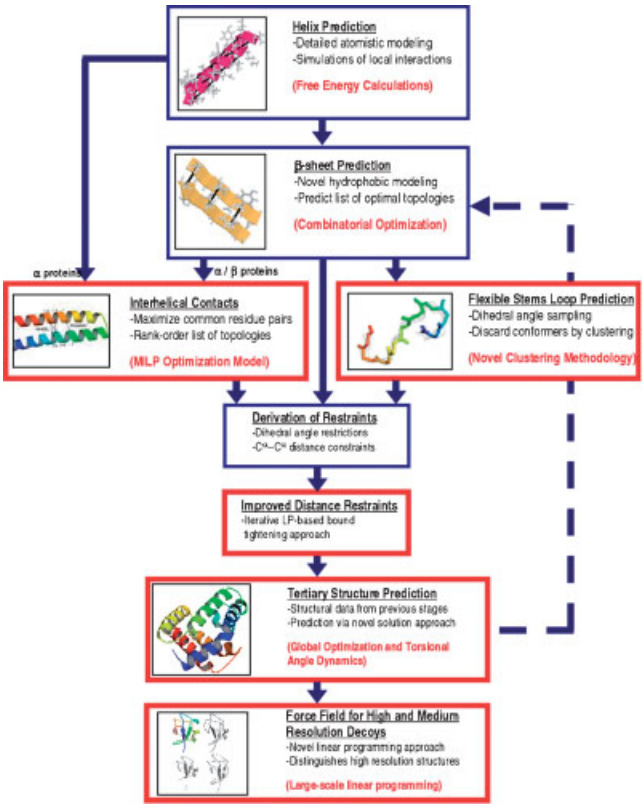

Figure 1. The enhanced ASTRO-FOLD protein structure prediction framework.

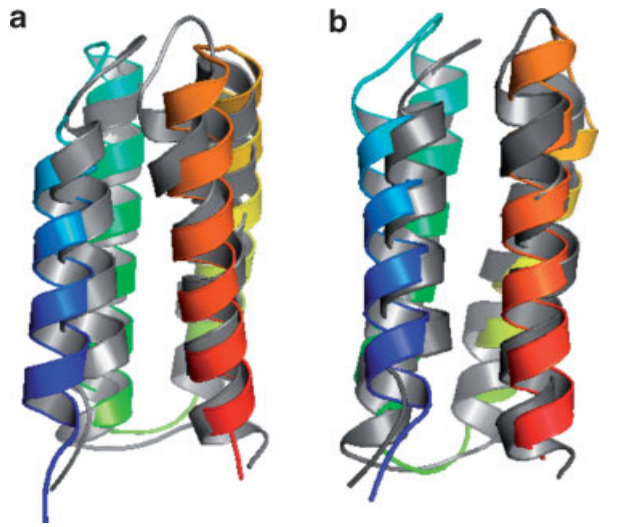

Figure 2. The predicted structures of the 4-helical bundle S836. The 102 amino acid sequence is as follows: MYGKLNDLLEDLQEVLKHVNQHWQGGQKNMNKVDHH-LQNVIEDIHDFMQGGGSGGKLQEMMKEFQQVLDEIKQQLQGGDNSLHNVHENIKEIFHH-LEELVHR. (**a**) The lowest energy predicted structure of S836 (color) versus the native S836 structure (gray). This structure has an energy of −1,740.11 kcal/mol and a C^{α} rmsd to the native model of 2.84 Å. (**b**) The predicted structure of S836 with the lowest C^{α} rmsd (color) compared to the predicted structure the native S836 structure (gray). This structure has an energy of −1,697.88 kcal/mol and a C^{α} rmsd to the native model of 2.39 Å.

DOI 10.1002/bit

bundle protein structure prediction was based on the enhanced ASTRO-FOLD framework and resulted in rmsd of 2.84 angstroms for the lowest energy conformer and lowest rmsd of 2.39 angstroms for the ensemble of generated conformers compared to the NMR structure (McAllister and Floudas, 2007).

For both aforementioned categories of first principles methods applied to new folds (NF) targets and fold recognition with analogous proteins (FR/A) targets, the evaluators of the progress during a decade of CASP experiments pointed out that the quantitative metric of GDT-TS revealed only marginal progress even though visual inspection for five targets revealed that more progress was achieved (Kryshtafovych et al., 2005; Vincent et al., 2005).

Final Comments and Personal Views

Based on recent CASP events (CASP 5, 6, and 7), it becomes evident that the first principles methods that utilize database information, more specifically, the fragment-based methods (Baker and co-workers) and the hybrid methods (Skolnick and co-workers; Zhang and co-workers; Kolinski and co-workers)(see Section ''First Principles Methods With Database Information'') are at present leading in consistency for successful predictions primarily for medium resolution structures and for a few high resolution structures. The success of the fragment-based approaches can be attributed to the creation of libraries of fragments which when they contain correctly biased conformers, they restrict the sampling search within the correct and easily identifiable domain from the computational perspective. The success of the hybrid methods primarily stems from an effective comparative modeling tool, the correct template identification, the use of database information for the generation of spatial restraints that guide the MC sampling search, and an iterative approach, which is enhanced from clustering and metrics for the assessment and selection of predicted structures. Both aforementioned classes of methods benefit from the recent findings that the existing PDB is close to being complete for low to medium resolution single domain protein structures (Zhang and Skolnick, 2005). The selective reduction and focus of the search space for conformational sampling based on database information extracted from protein fragments and/or lattice-based simulations, results in improved computational performance especially compared to the first principles methods which do not utilize database information and hence do not guide the search based on evolutionary sequence and known structural information. From this perspective, the first principles methods without database information will benefit from technological advances in large-scale distributed computing environments which allow for extensive unbiased sampling of protein conformational space.

Despite several successful medium resolution blind predictions, it is also apparent that significant advances are needed for consistent medium resolution predictions particularly in the difficult domain of free modeling (i.e., when no structurally similar templates can be successfully identified) as recently classified in CASP7. In a way, the free modeling domain defines the boundary of success for comparative modeling, fold recognition, fragment-based, and hybrid methods. The challenging free modeling domain is also more aligned with the goals of first principles methods, which do not use database information. As a consequence, it is expected that novel approaches using first principles only may play an important role in advancing the quality of protein structure predictions in this domain. Furthermore, the development and validation of protein structure prediction methods for high resolution is currently at an early stage and represents a formidable challenge.

Major challenges in comparative modeling and fold recognition include the optimal template selection, the quality of sequence to structure alignment especially for low sequence similarity, modeling of the core such as β-bulges, the development of improved force fields and refinement methods, improved modeling of side chains and structurally divergent regions, and high resolution refinement of comparative models.

Significant challenges for both categories of the first principles methods include: (a) the protein structure prediction for new folds (NF) targets or targets in the free modeling domain, and the assessment of prediction methods that do not use templates (Vincent et al., 2005; Moult et al., 2005); (b) the improvement of loop structure prediction (Kryshtafovych et al., 2005); (c) the derivation of scoring functions that will reliably select the most accurate models from a set of candidate structures (Moult et al., 2005; Kryshtafovych et al., 2005); (d) the correct identification of domains for large targets, and (e) the high resolution protein structure prediction for which the recent work of Bradley et al. (2005) for small proteins represents a promising and exciting direction.

CAF gratefully acknowledges financial support from the National Science Foundation and the National Institutes of Health (R01 GM52032; R24 GM069736) and the U.S. Environmental Protection Agency, EPA Grant R 832721-010.

References

Adjiman CS, Androulakis IP, Maranas CD, Floudas CA. 1996. A global optimization method, αBB, for process design. Comp Chem Eng 20:S419–S424.

Adjiman CS, Androulakis IP, Floudas CA. 1997. Global optimization of MINLP problems in process synthesis and design. Comp Chem Eng 21:S445–S450.

Adjiman CS, Dallwig S, Floudas CA, Neumaier A. 1998a. A global optimisation method for general twice-differentiable NLPs-I. Theoretical advances. Comp Chem Eng 22:1137–1158.

Adjiman CS, Androulakis IP, Floudas CA. 1998b. A global optimization method for general twice-differentiable NLPs-II. Implementation and computational results. Comp Chem Eng 22:1159–1179.

Androulakis IP, Maranas CD, Floudas CA. 1995. αBB: A global optimisation method for general constrained nonconvex problems. J Global Optim 7:337–363.
Anfinsen CB. 1973. Principles that govern the folding of protein chains. Science 181(4096):223–230.
Baker D.Prediction and design of macromolecular structures and interactions. 2006. Phil Trans R Soc B 361:459–463.
Baldwin RL, Rose. GD. 1999. Is protein folding hierarchic?: Local structure and protein folding. Trends Biochem Sci 24:26–33.
Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. Nucleic Acids Res 28:235–242.
Bradley P, Misura KMS, Baker D. 2005. Toward high resolution de novo structure prediction for small proteins. Science 309:1868–1871.
Bujnicki JM. 2006. Protein structure prediction by recombination of fragments. Chem Bio Chem 7:19–27.
Czaplewski C, Liwo A, Pillardy J, Oldziej S, Scheraga HA. 2004a. Improved conformational space annealing method to treat beta-structure with the UNRES force-field and to enhance scalability of parallel implementation. Polymer 45:677–686.
Czaplewski C, Oldziej S, Liwo A, Scheraga HA. 2004b. Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. Protein Eng, Design Sel 17:29–36.
Dunbrack RL. 2006. Sequence comparison and protein structure prediction. Curr Opin Struct Biol 16:374–384.
Eyrich VA, Standley DM, Felts AK, Friesner RA. 1999a. Protein tertiary structure prediction using a branch and bound algorithm. Proteins 35:41–57.
Eyrich VA, Standley DM, Friesner RA. 1999b. Prediction of protein tertiary structure to low resolution: Performance for a large and structurally diverse test set. J Mol Biol 288:725–742.
Floudas CA. 2000. Deterministic global optimization: theory, methods and applications. Nonconvex optimization and its applications. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Floudas CA. 2005. Research challenges, opportunities, and synergism in systems engineering and computational biology. AIChE J 51:1872–1884.
Floudas CA, Akrotirianakis IG, Caratzoulas S, Meyer CA, Kallrath J. 2005. Global optimization in the 21st century: Advances and challenges. Comp Chem Eng 29:1185–1202.
Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. 2006. Advances in protein structure prediction and de novo protein design: A review. Chem Eng Sci 61:966–988.
Ginalski K. 2006. Comparative modeling for protein structure prediction. Curr Opin Struct Biol 16:172–177.
Ginalski K, Grishin NV, Godzik A, Rychlewski L. 2005. Practical lessons from protein structure prediction. Nucl Acids Res 33:1874–1891.
Kihara D, Skolnick J. 2003. The PDB is a covering set of small protein structures. J Mol Biol 334:793–802.
Klepeis JL, Floudas CA. 1999. Free energy calculations for peptides via deterministic global optimization. J Chem Phys 110:7491–7512.
Klepeis JL, Floudas CA. 2002. Ab initio prediction of helical segments in polypeptides. J Comput Chem 23:245–266.
Klepeis JL, Floudas CA. 2003a. Prediction of beta-sheet topology and disulfide bridges in polypeptides. J Comput Chem 24:191–208.
Klepeis JL, Floudas CA. 2003c. ASTRO-FOLD: A combinatorial and global optimization framework for ab initio prediction of three-dimensional structures of proteins from the amino acid sequence. Biophys J 85:2119–2146.
Klepeis JL, Floudas CA. 2005. Analysis and prediction of loop segments in protein structure. Comp Chem Eng 29:423A–436A
Klepeis JL, Floudas CA, Morikis D, Lambris JD. 1999. Predicting peptide structures using nmr data and deterministic global optimization. J Comput Chem 20:1354–1370.
Klepeis JL, Schafroth HD, Westerberg KM, Floudas CA. 2002. Deterministic global optimization and ab initio approaches for the structure prediction of polypeptides, dynamics of protein folding and protein-protein interaction. In: Friesner RA, editor. Advances in Chemical Physics, Vol. 120. New York: Wiley. 254–457.
Klepeis JL, Pieja MT, Floudas CA. 2003a. A new class of hybrid global optimization algorithms for peptide structure prediction: Integrated hybrids. Comp Phys Commun 151:121–140.
Klepeis JL, Pieja MT, Floudas CA. 2003b. Hybrid global optimization algorithms for protein structure prediction: Alternating hybrids. Biophys J 84:869–882.
Klepeis JL, Wei YN, Hecht MH, Floudas CA. 2005. Ab initio prediction of the three-dimensional structure of a de novo designed protein: A double-blind case study. Proteins 58:560–570.
Kolinski A. 2004. Protein modeling and structure prediction with a reduced representation. Acta Biochimica Polonica 51:349–371.
Kolinski A, Bujnicki JM. 2005. Generalized protein structure prediction based on combination of fold recognition with de novo folding and evaluation of models. Proteins 61(Suppl 7): 84–90.
Kopp J, Schwede T. 2004. Automated protein structure homology modeling: A progress report. Pharmacogenomics J 5(4):405–416.
Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. 2003. A ''Frankestein's monster'' approach to comparative modelling: Merging the finest fragments of fold-recognition models and iterative model refinement aided by 3D structure evaluation. Proteins 53:369–379.
Kosinski J, Gajda MJ, Cymerman IA, Kurowski MA, Pawlowski M, Boniecki M, Obarska A, Papaj G, Sroczynska-Obuchowicz P, Tkaczuk KL, Sniezynska P, Sasin JM, Augustyn A, Bujnicki JM, Feder MJ. 2005. The frankenstein becomes a cyborg: The automatic recombination and realignment of fold recognition models in casp6. Proteins 61(Suppl 7): 106–113.
Kryshtafovych A, Venclovas C, Fidelis K, Moult J. 2005. Progress over the first decade of casp experiments. Proteins 61(S7):225–236.
Lee J, Scheraga HA. 1999. Conformational space annealing by parallel computations: Extensive conformational search of met-enkephalin and the 20-residue membrane-bound portion of melittin. Int J Quant Chem 75:255–265.
Lee J, Scheraga HA, Rackovsky S. 1997. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. J Computat Chem 18:1222–1232.
Lee J, Scheraga HA, Rackovsky S. 1998. Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing. Biopolymers 46:103–115.
Lee J, Pillardy J, Czaplewski C, Arnautova Y, Ripoll DR, Liwo A, Gibson KD, Wawak RJ, Scheraga HA. 2000. Effecient parallel algorithms in global optimisation of potential energy functions for peptides, proteins and crystals. Comp Phys Commun 128:399–411.
Lee J, Ripoll DR, Czaplewski C, Pillardy J, Wedemeyer WJ, Scheraga HA. 2001. Optimization of parameters in macromolecular potential energy functions by conformational space annealing. J Phys Chem B 105:7291–7298.
Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. 1997a. A united-residue force field for off-lattice protein structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data J Comput Chem 18:849–873.
Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. 1997b. A united-residue force field for off-lattice protein structure simulations. II. Parameterization of short-range interactions and determination of weights of energy terms by z-score optimization. J Comput Chem 18:874–887.
Liwo A, Czaplewski C, Pillardy J, Scheraga HA. 2001. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. J Chem Phys 115:2323–2347.
Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. 2002. A method for optimizing potential-energy functions by hieracrchical design of the potential-energy landscape: Application to the UNRES force field. Proc Natl Acad Sci U S A 99:1937–1942.

DOI 10.1002/bit

McAllister SR, Floudas CA. 2007. Enhanced astro-fold framework for protein structure prediction. In preparation.
McAllister SR, Mickus BE, Klepeis JL, Floudas CA. 2006. Novel approach for alpha-helical topology prediction in globular proteins: Generation of interhelical restraints. Proteins 65:930–952.
Monnigmann M, Floudas CA. 2005. Protein loop structure prediction with flexible stem geometries. Proteins 61:748–762.
Moult J. 1999. Predicting protein three-dimensional structure. Curr Opin Biotechnol 10:583–588.
Moult J. 2005. A decade of casp: Progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol 15:285–289.
Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. 1997. Critical assessment of methods of protein structure prediction (CASP)—Round II. Proteins S1:2–6.
Moult J, Fidelis K, Zemla A, Hubbard T. 2001. Critical assessment of methods of protein structure prediction CASP—Round 4. Proteins S5:2–7.
Moult J, Fidelis K, Zemla A, Hubbard T. 2003. Critical assessment of methods of protein structure prediction (CASP)—Round V. Proteins 53:334–339.
Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. 2005. Critical assessment of methods of protein structure prediction (casp)—Round 6. Proteins 61(S7):3–7.
Murzin AG. 2004. Protein structure watch: Making ''predictions'' easy-www.forcasp.org.
Nanias M, Chinchio M, Oldziej S, Czaplewski C. 2005. Protein structure prediction with the unres force field using replica-exchange monte carlo with minimization: Comparison with mcm, csa and cfmc. J Comput Chem 26:1472–1486.
Pandit SB, Zhang Y, Skolnick J.Tasser-lite: An automated tool for protein comparative modeling. Biophysical J 2006. 91(11):4180–4190.
Petrey D, Honig B. 2005. Protein structure prediction: Inroads to biology. Mol Cell 20:811–819.
Pillardy J, Czaplewski C, Liwo A, Wedemeyer WJ, Lee J, Ripoll DR, Arlukowicz P, Oldziej S, Arnautova EA, Scheraga HA. 2001. Development of physics-based energy functions that predict medium resolution structure for proteins of α, β and α/β structural classes. J Phys Chem B 105:7299–7311.
Przybylski D, Rost. B. 2004. Improving fold recognition without folds. J Mol Biol 341:255–269.
Rajgaria R, McAllister SR, Floudas CA. 2006. A novel high resolution ca-ca distance dependent force field based on a high resolution decoy set. Proteins 65:726–741.
Ripoll D, Liwo A, Scheraga HA. 1998. New developments of the electrostatically driven monte carlo method: Tests on the membrane-bound portion of melittin. Biopolymers 46:117–126.
Rohl CA, Strauss CEM, Chivian D, Baker D. 2004. Modeling structurally variable regions in homologous proteins with Rosetta. Proteins 55:656–677.
Simons KT, Kooperberg C, Huang C, Baker D. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268:209.
Simons KT, Ruczinki I, Kooperberg C, Fox BA, Bystroff C, Baker D. 1999. Improved recognition of native-like structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 34:82.
Skolnick J, Kolinski A, Kihara D, Betancourt M, Rotkiewicz P, Boniecki M. 2001Ab initio protein structure prediction via a combination of threading, lattice folding, clustering and structure refinement. Proteins (Suppl 5): 149–156.
Skolnick J, Zhang Y, Arakaki AK, Kolinski A, Boniecki M, Szilágyi A, Kihara D. 2003. TOUCHSTONE: A unified approach to protein structure prediction. Proteins 53:469–479.
Skolnick J, Kihara D, Zhang Y. 2004. Development and large scale benchmark testing of the PROSPECTOR 3 threading algorithm. Proteins 56:502–518.
Srinivasan R, Rose. GD. 1995. LINUS: A hierarchic procedure to predict the fold of a protein. Proteins 22:81–89.
Srinivasan R, Rose. GD. 2002. Ab initio prediction of protein structure using LINUS. Proteins 47:489–495.
Tramontano A, Morea V. 2003. Assessment of homology-based predictions in CA SP5. Proteins 53:352–368.
Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A. 2005. Assessment of predictions submitted for the casp6 comparative modeling category. Proteins 61(S7):27–45.
Vincent JJ, Tsien CH, Sathyanarayana BK, Lee. B. 2005. Assessment of casp6 predictions for new and nearly new fold targets. Proteins 61(S7):67–83.
Vitkup D, Melomud E, Moult J, Sander C. 2001. Completeness in structural genomics. Nature Struct Biol 8(6):559–566.
Wang G, Jin Y, Dunbrack RL. 2005. Assessment of fold recognition predictions in casp6. Proteins 61(S7):46–66.
Xu Y, Xu. D. 2000. Protein threading using PROSPECT: Design and evolution. Proteins 40:343–354.
Xu J, Li M, Kim D, Xu. Y. 2003. RAPTOR: Optimial protein threading by linear programming. J Bioinformat Comput Biol 1:95–117.
Zhang Y, Skolnick J. 2004a. Tertiary structure predictions on a comprehensive bench-mark of medium to large size proteins. Biophys J 87:2647–2655.
Zhang Y, Skolnick J. 2004b. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci U S A 101(20):7594–7599.
Zhang Y, Skolnick J. 2004c. SPICKER: A clustering approach to identify near-native protein folds. J Comput Chem 25:865–871.
Zhang Y, Skolnick J. 2005. The protein structure prediction problem could be solved using the current pdb library. Proc Natl Acad Sci U S A 102:1029–1034.
Zhang Y, Kolinski A, Skolnick J. 2003. TOUCHSTONE II: A new approach to abinitio protein structure prediction. Biophys J 85:1145–1164.
Zhang Y, Arakaki AK, Skolnick J. 2005. Tasser: An automated method for the prediction of protein tertiary structures in casp6. Proteins 61(S7):91–98.