Comparative Modeling: The State of the Art and Protein Drug Target Structure Prediction

Tianyun Liu[1], Grace W. Tang[2] and Emidio Capriotti[*,2,3]

[1]*Department of Genetics,* [2]*Department of Bioengineering, Stanford University, Stanford, CA 94305, USA*

[3]*Department of Mathematics and Computer Science, University of Balearic Islands, Palma de Mallorca, Spain*

Abstract: The goal of computational protein structure prediction is to provide three-dimensional (3D) structures with resolution comparable to experimental results. Comparative modeling, which predicts the 3D structure of a protein based on its sequence similarity to homologous structures, is the most accurate computational method for structure prediction. In the last two decades, significant progress has been made on comparative modeling methods. Using the large number of protein structures deposited in the Protein Data Bank (~65,000), automatic prediction pipelines are generating a tremendous number of models (~1.9 million) for sequences whose structures have not been experimentally determined. Accurate models are suitable for a wide range of applications, such as prediction of protein binding sites, prediction of the effect of protein mutations, and structure-guided virtual screening. In particular, comparative modeling has enabled structure-based drug design against protein targets with unknown structures. In this review, we describe the theoretical basis of comparative modeling, the available automatic methods and databases, and the algorithms to evaluate the accuracy of predicted structures. Finally, we discuss relevant applications in the prediction of important drug target proteins, focusing on the G protein-coupled receptor (GPCR) and protein kinase families.

Keywords: Protein structure prediction, comparative modeling, sequence alignment, homology, drug target, drug design.

INTRODUCTION

Protein three-dimensional (3D) structure is essential for functional annotation and rational drug design [1]. Experimental techniques to crystallize and characterize protein structures are difficult, resulting in an increasing gap between the number of available protein sequences and known structures (see Fig. **1**). Hence, the main goal of computational prediction methods is to link the protein sequence to its 3D structure and finally to its functionally relevant features.

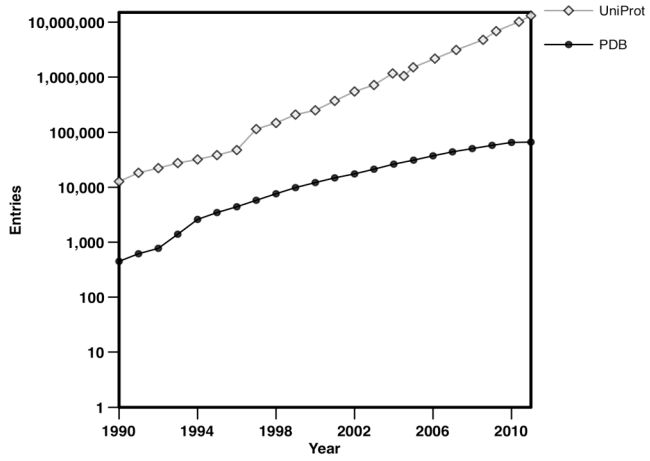

Fig. (1). Number of entries in the UniProt [164] and PDB [11] databases.

*Address correspondence to this author at the Department of Bioengineering, Stanford University, 318 Campus Dr, Room S240 Mail code: 5448, Stanford, CA 94305, USA;
Tel: +1 (650) 725 4484; Fax: +1 (650) 724 4021;
E-mail: emidio@stanford.edu

The primary sequence of a protein determines its 3D structure, but the mechanism of the transition from unfolded to folded state is not completely understood. Protein folding is a complex problem that has never been solved by analytical approaches, mainly because there is no theoretical model to describe the atomic interactions within the protein and the effect of the solvent. Early computational studies on limited protein structure data revealed that within a protein family, members from different species conserved structure more effectively than sequence [1-3]. Hence, the classical concept of homology, which referred to structural correspondence between traits [4] derived from a common ancestor, was expanded to molecular biology. For proteins, homology indicates derivation from a common "ancestor" [5] and is often inferred using sequence similarity. Given that protein structure is more conserved than sequence, sequence similarity suggests structural similarity, enabling structure prediction of proteins with similar sequences. Even though analytical solutions to the folding problem are not available, empirical methods based on sequence similarity have shown good performance.

Of the dominant approaches to protein 3D structure prediction, comparative modeling and fold recognition rely on the assumption of structural similarity based on sequence similarity. Both approaches require a minimal level of sequence similarity between the unknown protein (target) and at least one protein with known structure (template). For this reason, they are also known as template-based approaches. In contrast, new fold methods do not explicitly require any template structure and therefore have broad applicability. The new fold prediction algorithms can be classified as *ab initio* methods when they only rely on physicochemical principles or *de novo* methods when they include information from known protein structures [6]. *Ab initio* methods based on all atom simulation with empirical

1386-2073/11 $58.00+.00

force fields have been successfully used to predict the folding of short peptides [7, 8]. Alternatively, the more accurate *de novo* algorithms use a library of short protein fragments extracted from the PDB [9] or restraints from threading results to reduce the space of possible conformations [10]. However, predictions from new fold methods are still generally lower quality than those from template-based approaches. Template-based methods are highly accurate and can predict structures very similar to the native structure. Their application, however, is limited to cases for which a homologous protein with known structure is available.

To expand the application of template-based prediction methods to proteins with unknown fold, a better coverage of protein structural space is required. As evidenced by the continued discovery of new protein structures corresponding to new folds, our knowledge of protein three-dimensional space is incomplete. The current Protein Data Bank (PDB) [11] contains ~65,000 protein structures corresponding to ~1,400 unique folds from the SCOP database [12]. To augment this, worldwide structural genomics initiatives (SGIs) have been launched to explore different regions of protein structural space by selecting targets from novel, structurally uncharacterized protein families [13]. To date, the SGIs have deposited more than 9,600 new structures in the PDB. Analysis of structures deposited through 2006 showed that SGI structures compared to non-SGI ones have a higher rate of new SCOP folds and superfamilies [14]. Recent studies suggest that full coverage of protein structural space is achievable by selecting targets from large and diverse superfamilies with varied structures and functions [15-17].

Comparative modeling methods have also benefited from the increase in computational power. Algorithms are now faster and more accurate. Biannually since 1994, the assessors of the Critical Assessment of techniques for protein Structure Prediction (CASP) gauge the advancements in protein structure prediction. Results from the eighth edition show that template-based methods generally produce the most accurate predictions. A positive correlation also exists between prediction accuracy and target and template sequence similarity [18]. Structural alignments of homologs likewise improve with increasing sequence similarity [1]. The minimum level of sequence similarity to infer structural similarity has been well characterized [19]. It is now largely accepted that template-based prediction methods can be safely applied if target and template have more than 35% sequence identity for alignments of ~100 residues. High quality models from comparative modeling can be suitable for a wide range of applications, including functional annotation, ligand-binding site prediction, virtual screening, docking of small molecules, and molecular replacement.

In this review, we focus on comparative modeling structure prediction methodology. We first discuss the theoretical basis of comparative modeling, including method development and predicted structure evaluation. We then present applications of comparative modeling, focusing on the prediction of drug target protein structures from the G-protein-coupled receptor (GPCR) and protein kinase families. Finally, we discuss future applications of comparative modeling and its impact on drug design.

PROTEIN STRUCTURE PREDICTION BY COMPARATIVE MODELING

The prediction of protein three-dimensional (3D) structure is still an unsolved problem. Comparative modeling approaches can be successful when there is detectable sequence similarity between the unknown protein (target) and the structure of another protein (template). The basis of this empiric observation is that amino acid substitution operates within the constraints of structure and function. Therefore, structure and function tend to be more conserved than sequence.

Theoretical Basis of Comparative Modeling

The application of comparative modeling approaches is possible because small changes in protein sequence usually result in small changes in 3D structure [1]. Mutations accumulated during evolution are constrained to conserve protein intramolecular and intermolecular interactions that mediate designated functions in protein families and superfamilies [20]. Structural comparison of 25 proteins from eight families revealed the existence of highly conserved structural regions [1]. For 32 pairs of homologous proteins, segments with greater than 50% sequence identity showed more than 90% of Cα atoms to be structurally superimposed, while less conserved regions (~20% sequence identity) showed less than 42% structural similarity. The calculated Root Mean Square Deviation (RMSD) for the high and low sequence conservation segments measured ~1 Å and ~3 Å, respectively. This analysis [1] quantitatively defined the expected degree of success in the prediction of a target protein structure from a homologous template structure as a function of sequence similarity.

Sequence similarity not only establishes the accuracy but also the applicability of comparative modeling. As the number of solved protein structures increased, a more accurate and exhaustive analysis of sequence versus structural similarity was performed [19]. A large set of exhaustive pairwise alignments between 792 proteins with less than 25% sequence identity was used to define the "twilight zone". This heterogeneous region of sequence alignment space corresponds to pairs of non-homologs (true negatives) and remote homologs (false negatives). The curve separating the "twilight zone" from the region of confident homology detection, which is populated by alignments between homologs (true positives), has been estimated using alignments from structurally related proteins from the FSSP database [21]. This curve [19] outperformed a previous curve in discriminating between true positive and false positive alignments [22]. Specifically, the new separation curve is more conservative for shorter alignments and less conservative for longer alignments to decrease the false positive rate and increase the true positive rate, respectively (see Fig. **2**). Comparative modeling is generally constrained to the region of confident homology detection. For some targets, however, no structural neighbors fall in this region even with the growth of structural data. Incomplete knowledge of structural space therefore implies that the problem of the "twilight zone" is limiting the usage of comparative modeling.

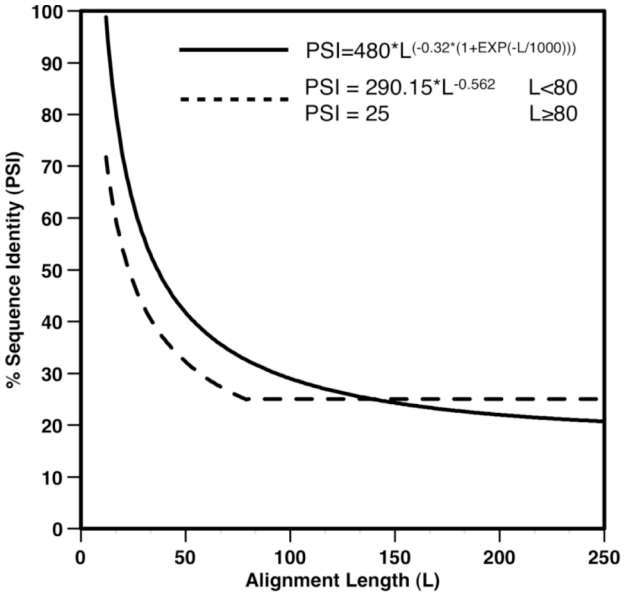

Fig. (2). Twilight zone curves from Rost's (continuous line) and Sander's (dashed line) works [19, 22].

Comparative Modeling Method

Conservation of protein 3D structure between two proteins with similar sequences allows the prediction of the structure of one protein using the structural features of the other (see Fig. **3**). Accordingly, comparative modeling procedures can be divided into four sequential steps:

1. Fold assignment and template selection
2. Target-template sequence alignment
3. Model building and refinement
4. Prediction evaluation

Fold Assignment and Template Selection

Fold assignment and template selection is the first step to all comparative modeling methods. It encompasses the comparison of the target protein to a set of proteins with known structural features, searching for homologous proteins that are likely to have a similar structure. Template protein structures used for this step are collected from the PDB [11], though the SCOP [12], DALI [23] and CATH [24] databases are frequently used to narrow the search. The simplest searching methods are based on a pairwise sequence comparison of target and template using BLAST [25] or FASTA [26]. The PSI-BLAST algorithm [27] was later developed to improve the detection of homologous proteins with a low level of sequence identity. It is based on an iterative BLAST search that for a given sequence, performs a run of BLAST to select a set of homologs from a sequence database, calculates a position-specific scoring matrix from the derived multiple sequence alignment, and uses the resulting matrix to scan the database for new homologs. Profile-based algorithms, which implement an alignment procedure that includes information from related proteins, perform much better than standard methods based on pairwise alignments [28-31]. Among profile-based methods, hidden Markov models (HMMs) perform the best [32]. The most successful profile HMM procedures for detection of remote protein homologs are SAM [33], HMMER [34], and HHPred [32]. More details about remote homology search methods are discussed in a previously published review [35].

In situations where more than one template is found, the custom is to select the template protein with highest sequence identity to the target. This increases the likelihood of a high quality prediction. However, there are exceptions to template selection depending on the final purpose of the predicted structure. If the aim is to study interactions between the target and a small ligand or another protein,

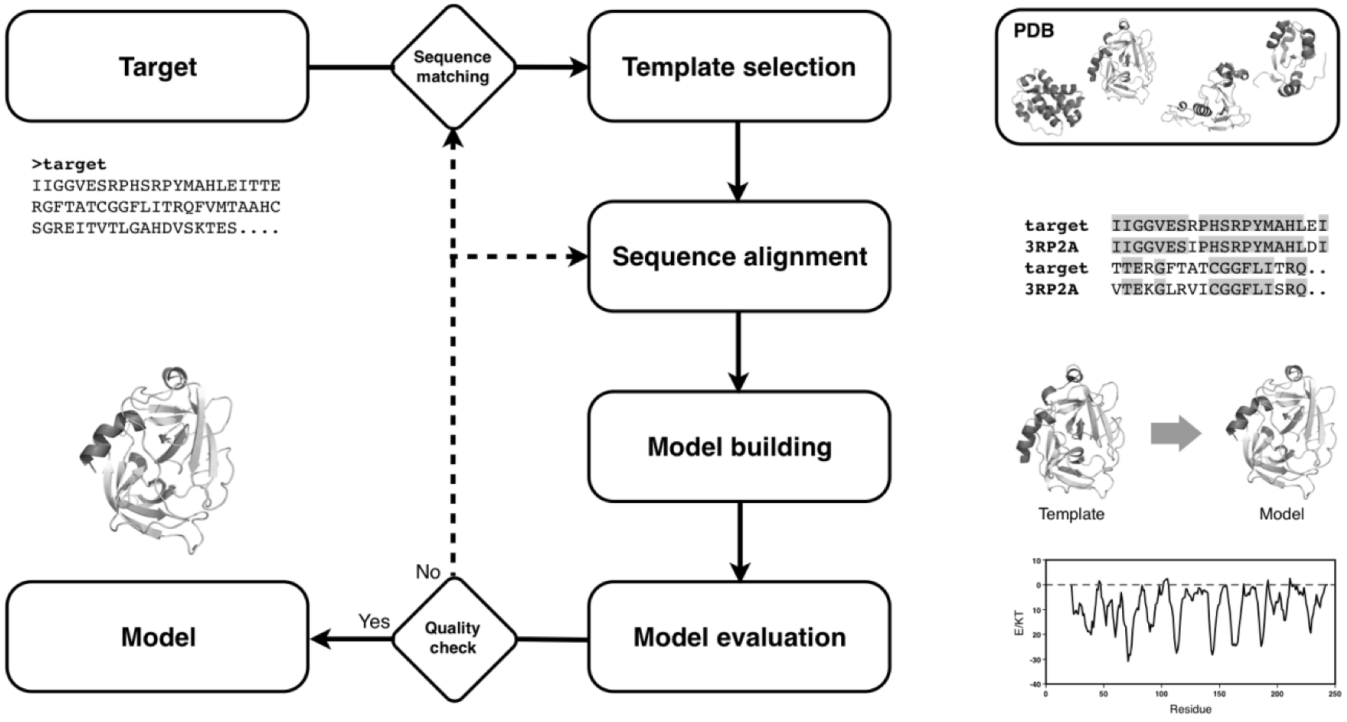

Fig. (3). Flowchart of comparative modeling method.

templates that contain similar types of interactions are ideal. On the other hand, if the aim is to study the conformation of an active site, high-resolution templates are preferable. Selection of the best template is therefore mainly driven by problem-specific considerations. Alternatively, the use of information from multiple templates can increase the quality of the predicted structure [36], though the standard comparative modeling procedure uses only one template.

Target-Template Sequence Alignment

Target-template sequence alignment aims to establish the correct correspondences between the residues of the target and template. Standard sequence alignment methods such as Needleman–Wunsch [37] and Smith–Waterman [38] are based on dynamic-programming algorithms and can be suitable for this task. They calculate an alignment score using substitution scoring matrices such as BLOSUM [39] and PAM [40]. Most template selection and fold assignment methods will also return a default sequence alignment between the target and template. When sequence identity between target and template is high, standard methods produce similar alignments. However, when sequence identity is low (less than 40%), more sophisticated alignment procedures are needed. In these cases, methods relying on multiple sequence alignments and/or structure information derived from homologous proteins have achieved better results [28, 30, 41, 42]. These resources are used to build sequence profiles for the target and template proteins for subsequent profile alignment. Although automated methods for sequence alignment have reached a good level of accuracy, manual inspection of the calculated alignment is still strongly suggested when sequence similarity is lower than 40%.

Model Building and Refinement

Model building involves the explicit prediction of the target structure atomic coordinates using residue equivalences defined in the sequence alignment. The methods for model building are based on three main approaches: rigid body assembly [43], segment matching [44], and spatial restraints satisfaction [45]. Modeling based on rigid body assembly uses small rigid bodies derived from structurally aligned proteins. First, the atomic coordinates of the conserved regions are used to reconstruct the main chain of the conserved residues, forming the protein core. Loops are then built in by scanning a database of structural peptide fragments, searching for those that fit the conformation of the core. When the backbone of the structure is complete, side chain atom coordinates are predicted, taking into account the preferred intrinsic residue conformations and those of the equivalent residues in the template. Modeling by segment matching is based on the observation that approximately 100 six-residue peptides can account for 76% of conformational space [46]. These peptides can serve as the building blocks for structure prediction. Similar to rigid body assembly, the target's protein core is built using select Cα atoms from conserved residues. Appropriate peptide building blocks are then fitted into the structure to produce an all atom model. Modeling by satisfaction of spatial restraints uses the idea that structural features of conserved residues are similar. For this class of methods, evolutionary conservation is a criteria to select and generate homology-based restraints for the target structure using distances and angles between equivalent residues in the template. These restraints are usually supplemented with generic stereochemical restraints from molecular mechanics force fields. Finally, an optimization procedure is performed to search for global low energy conformations that minimize the number of restraint violations. Restraint-based modeling approaches are the most flexible because they can easily incorporate many different types of restraints and constraints. These can be derived from different template structures or from various experimental data sources like NMR experiments, fluorescence spectroscopy, and site-direct mutagenesis.

In general, the most difficult tasks in model building are the prediction of loop regions and side-chain conformations. The frequent insertions and deletions in the exposed loop regions are the main reason for loop structural variability between members of the same protein family. As a result, dedicated methods have been developed to predict loop regions either by hand using molecular graphics [47], through database searching [48], or by *ab initio* methods [49].

For side chain atoms, the large number of possible conformations reflected in the distribution of χ_1 and χ_2 dihedral angles makes predicting them difficult. For building side chains, three major approaches have been developed [50-52]. First is the Minimum Perturbation protocol in which each substitution is followed by a rotation about the side chain's torsion angles to relieve clashes [51]. Second is the Coupled Perturbation protocol in which the side chain torsion angles of structurally adjacent residues are also rotated [52]. In recent years, major improvement in building side chains has been achieved using rotamer libraries. The best representation of this method is the program SCWRL [50], which makes use of backbone-dependent rotamer libraries through a search method based on graph theory.

The refinement of initial comparative modeling predictions generally involves simultaneous optimization of the predicted conformations for the non-conserved regions along with the physically adjacent regions [53-55]. In principle, molecular dynamics (MD) techniques should be able to achieve this goal. Given a sufficiently accurate interatomic force field, MD simulations performed in an appropriate environment should lead a protein model to its native conformation, reflective of the global free energy minimum at the given temperature and environment. However, previous CASP experiments indicate the conformational space explored is fairly local to the initial conformation [55]. Four reasons have been suggested: the ruggedness of the potential energy landscape, inadequate sampling of alternative conformations, insufficient accuracy in the description of the interatomic forces, and too short simulation times [53, 55, 56]. To solve these problems, present optimization attempts with molecular dynamics are normally based on limited conformational sampling with a detailed force field, or more extensive sampling with a simplified force field [56]. However, these approaches have proved to be ineffective, and thus using MD simulation as a refinement method is only useful for limited structural arrangements [57]. An analysis of the performances of six

comparative modeling methods has been recently published [58].

Prediction Evaluation

The final step in comparative modeling is the evaluation of predicted structures. Evaluation focuses on the model as an entity or as specific segments. Both the target-template alignment and the geometry and stereochemistry of the predicted model are assessed. Computational methods such as PROCHECK, AQUA, SFCHECK, Squid, and MolProbity [59-62] check the stereochemistry of the predicted structure, including bond lengths and angles, peptide bond and side-chain ring planarities, chirality, main-chain and side-chain torsion angles, and clashes between non-bonded pairs of atoms. Another class of programs evaluates the predicted structure using statistical potentials of mean force [63-66]. This approach calculates the structural environment of each atom in the model and compares it with the mean expected environment in the high-resolution template structure. Although there is debate about the theoretical basis of statistical potential-based methods [67-69], they have been successfully applied to model assessment and selection of near native conformations from decoy predicted structures [70]. In the recent CASP8 experiments, QMEAN algorithm [71], which combines different structure-based scoring functions, was ranked among the best methods for global and local model quality assessment [72]. Alternative evaluation methods use physics-based energy functions to score the energy of the predicted model, taking into account bonded and non-bonded interactions between the atoms in the system. Therefore, classical physics-based energies for molecular dynamics simulations, such as AMBER [73] and CHARMM [74], have been used to assess the quality of predicted models [75,76]. A more complete description of the methods for model assessment has been recently published [70].

After evaluation, depending on the quality of the target structure, it is possible to go back to the first or second step to select a better template or to improve the quality of the sequence alignment. In these cases, the prediction procedure is iterated until the final structure reaches an acceptable level of accuracy.

Accuracy and Limitations of Comparative Modeling

Comparative modeling structure prediction relies on an evolutionary relationship between target and template. Incorrect assumption of this relationship will affect template selection and propagate to the quality of the target-template sequence alignment and the identification of the target-template structural and functional divergences [1]. Application of comparative modeling can therefore be limited by the absence of an appropriate template. This can occur when the target is from a new fold class not represented in the PDB. Identification of such targets typically involves a Z-score, which evaluates the quality of the selected template. It compares the score of the target and template pair with the distribution of scores from all possible target and template pairs. In this case, the templates are a representative set of all protein folds, where a high quality template corresponds to a high Z-score [77-79].

Comparative modeling can also suffer from weak evolutionary relationships. Weak sequence similarity between the target and template leads to incorrect selection of template structure or incorrect alignment of target and template. Thus, a lower level of sequence similarity in the alignment generally corresponds to a less accurate predicted structure [80]. Comparative modeling is particularly difficult in the "twilight zone," where correct templates and equivalences between target and template residues are hard to detect. For the identification of appropriate templates, powerful methods for remote homolog detection and statistical potentials for scoring the compatibility of the target sequence to the template fold can improve results. With respect to uncertain residue equivalences between target and template, prediction methods can usually return an ensemble of alternative solutions. Therefore, the selection of a near native structure is a key issue in protein structure prediction. Clustering algorithms, built on the assumption that well-populated structural clusters are more reflective of native conformations than low energy structures, have been developed to identify high quality predictions [72, 81, 82].

The accuracy of comparative modeling can further be improved using multiple template structures. However, additional data is only beneficial in certain scenarios. Generally, the templates under consideration should share a similar low level of sequence identity (less than 30%) with the target as well as with each other [83]. The benefit derived from this structural complementarity is dependent on the accuracy of the modeling alignment. Therefore, the development of optimized methods for the combination of multiple templates is still one of the major bottlenecks in comparative modeling [84].

Comparative Modeling at CASP

The Critical Assessment of techniques for protein Structure Prediction (CASP) evaluates the progress of methods in protein structure prediction. It was observed that the use of multiple templates or template fragments improved predictions over the single best template in approximately 1/3 of 64 template-based models (TBMs) [85]. Surprisingly, the best predictions from automated tools were not significantly less accurate than the manually curated multiple template models. In fact, the best automated prediction server ranked fifth in the prediction of the 64 TBMs with accuracy comparable to the human curated models. These results support the idea that automatic tools enable large-scale structural predictions, providing models with accuracy on par with manually generated models.

The CASP assessors use several measures to evaluate the quality of predicted structures [86]. Model accuracy and quality are measured on the overall structure or on a per residue basis. A specific category in CASP tests the available methods for estimating model accuracy, using Pearson's correlation coefficient between observed and predicted correctness as the score. Algorithms based on the clustering of alternative models are leaders in this category of CASP [72, 81]. An analysis of the performance of quality assessment methods in CASP8 has been recently published [87].

Resources and Tools for Comparative Modeling

In recent years, many different tools and resources for comparative modeling have been made available online. Although a complete description of available methods for comparative modeling is not the aim of this review, we provide a select list of representative tools (see Table **1**). Depending on the user's familiarity with computational tools, three types of end users are possible. ModBase [88], Swiss-Model [89], and Protein Model Portal [90] give access to large repositories of pre-computed predicted structures associated to millions of protein targets. These resources are particularly useful for those without strong experience in the use of computational tools and need to retrieve generic models for target proteins. Intermediate users can predict target structures using automated servers like LOMETS, ModWeb, or Robetta and evaluate the results using statistical potential-based tools like ANOLEA [91] or PROSA [92]. Finally, expert users can make use of all classes of tools reported in Table **1** by following the comparative modeling methodology outlined above. Depending on the final usage of the predicted structure, more complicated procedures such as using multiple templates or enforcing restraints could be performed with the listed tools.

APPLICATIONS OF COMPARATIVE MODELING

The first modeling experiment in 1975 predicted the structure of the calcium binding component of troponin using its homology to the calcium binding motif of parvalbumin [93]. Although it was the first published model, it was not deposited in the PDB until 1980, two years after other models were deposited [94, 95]. Nowadays, the large number of available structural data and user-friendly tools make large-scale comparative modeling practical. In this section, we summarize the applications of comparative modeling, focusing on the available computational tools and resources for drug design. In particular, we describe the application of comparative modeling to the prediction of G protein-coupled receptor (GPCR) and protein kinase families for virtual screening experiments.

Model Accuracy and Applications

Model accuracy should always be estimated prior to release and usage as it determines the predicted structure's suitability for biological experiments. As comparative modeling relies on homology between target and template, which is inferred by sequence similarity, the level of sequence identity between target and template defines three kinds of applications. Low quality models with sequence identity lower than 30% are in the "twilight zone" and are expected to have less than 50% of Cα atoms within 3.5 Å of their correct positions. These models can be used to confirm or reject the hypothesis that two remotely related proteins belong to the same fold class [96, 97]. Differing folds between target and template will lead to poor geometry and stereochemistry scores during prediction evaluation. Medium quality models, predicted using templates with sequence identity between 30% and 50%, have approximately 85% of Cα atoms within 3.5 Å of the native three-dimensional structure. They are suitable for refinement of functional predictions [98, 99] and prediction of binding affinity changes due to site-directed mutagenesis. High quality models with alignment sequence identity higher than 50% have high alignment Z-score and are expected to have average accuracy comparable to low resolution X-ray structures (3 Å resolution) or medium resolution NMR structures [100]. They can be used for docking small ligands [101] or for predicting the structure of interacting proteins [102, 103].

Models with a moderate level of accuracy have also seen additional applications. In molecular replacement, models enable recovery of the information lost when reducing from the three-dimensional space of the native structure to the bi-dimensional space of the X-ray diffraction experiment [104]. An automatic method for molecular replacement [105] has been developed for the reconstruction of new protein structures from X-ray data. Rational protein design has also benefited from comparative modeling. Antibodies represent a special case [106] for which the relationship between the sequence and structure of the functional site is well defined. This has allowed the rational design of new antibodies with different biochemical properties, which have had broad application in therapy and research [107-111]. More complex is the prediction of enzyme structures to be subsequently used in the design of enzyme inhibitors. Because enzymes can exist in different conformations, template selection must be driven by problem-specific considerations that account for the end use of the predicted model. These changes in enzyme structures are difficult to predict and compute [112, 113], but high quality models have been built and used for docking small ligands and for designing new inhibitors [114-116].

Computational Modeling and Drug Design

Structure libraries of drug target proteins play a key role in drug design, enabling computational methods to score and rank the predicted affinity between drugs and targets. This process is known as virtual screening [117] and has reduced the costs of experimental high-throughput assays. Virtual screening is knowledge-based and requires structural information of the target and ligand (ligand-based screening) or of the target alone (target-based screening) [118]. The process begins with identification of the protein binding-site and determination of the residues that interact with the ligand. This is the target binding-site into which different drugs can be docked and scored for binding affinity. Docking algorithms predict the relative orientation of the target and ligand with respect to one another. Rigid targets and ligands dominated in the past, but the increase in computational power has produced a shift towards docking flexible molecules [119]. Although a description of computational methods for drug design is not the aim of this review, we have included a selection of available resources and tools for predicting the binding affinity between a model structure and ligand (see Table **2**). This list contains databases of chemical compounds and target proteins as well as web available tools for binding site prediction, small molecule docking, virtual screening, and drug design. An exhaustive list of resources for drug design is provided by the Click2drug website (http://www.click2drug.org) at the

Table 1. Resources and Tools for Comparative Modeling

Name	URL	Ref.
Comparative Modeling Repositories		
MODBASE	http://modbase.compbio.ucsf.edu	[88]
Protein Model Portal	http://www.proteinmodelportal.org/	[90]
SWISS-MODEL Repository	http://swissmodel.expasy.org/repository	[89]
Structure and Classification Databases		
CATH	http://www.cathdb.info	[24]
PDB	http://www.pdb.org	[11]
PFAM	http://pfam.sanger.ac.uk	[165]
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop	[12]
Template Selection		
BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi	[25]
FASTA	http://www.ebi.ac.uk/Tools/fasta	[26]
FFAS03	http://ffas.ljcrf.edu	[166]
HHPred	http://toolkit.tuebingen.mpg.de/hhpred	[167]
Phyre	http://www.sbg.bio.ic.ac.uk/~phyre	[168]
SAM-T08	http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html	[169]
SP5	http://sparks.informatics.iupui.edu/SP5	[170]
Threader	http://bioinf.cs.ucl.ac.uk/threader	[171]
Alignment Tools		
CLUSTALW	http://www.ebi.ac.uk/clustalw	[172]
MAFFT	http://mafft.cbrc.jp/alignment/server	[173]
MUSCLE	http://www.drive5.com/muscle	[174]
T-Coffee	http://www.tcoffee.org	[175]
Automatic and Manual Modeling		
3D-JIGSAW	http://bmm.cancerresearchuk.org/~3djigsaw	[176]
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER	[177]
LOMETS	http://zhanglab.ccmb.med.umich.edu/LOMETS	[178]
MODELLER	http://www.salilab.org/modeller	[179]
MODWEB	https://modbase.compbio.ucsf.edu/scgi/modweb.cgi	[180]
ROBETTA	http://robetta.bakerlab.org	[181]
SWISS-MODEL	http://swissmodel.expasy.org	[182]
Model Evaluation		
ANOLEA	http://protein.bio.puc.cl/cardex/servers/anolea	[91]
DFIRE	http://sparks.informatics.iupui.edu/yueyang/DFIRE	[183]
FRST	http://protein.bio.unipd.it/frst	[184]
HARMONY	http://caps.ncbs.res.in/harmony	[185]
ModFOLD	http://www.reading.ac.uk/bioinf/ModFOLD/	[81]
MolProbity	http://molprobity.biochem.duke.edu	[62]
PROCHECK	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK	[186]
ProQ	http://www.sbc.su.se/~bjornw/ProQ	[187]
PROSA-web	https://prosa.services.came.sbg.ac.at	[92]
QMEAN	http://swissmodel.expasy.org/qmean	[188]
VERIFY3D	http://nihserver.mbi.ucla.edu/Verify_3D	[189]

Swiss Institute of Bioinformatics. In literature, there are many examples of target protein structural models that have been successfully used for the discovery and optimization of new compounds [120-123]. In this review, we focus on G-protein-coupled receptors and protein kinases, two of the most targeted protein families in drug discovery.

Modeling the G-Protein-Coupled Receptors

Transmembrane proteins (TM) participate in a wide range of important biological processes such as transportation of small molecules, signal transduction, and cell recognition and communication. They comprise approximately 30% of genes in the human genome, but are significantly under-represented in structural databases. In the PDB, TM proteins comprise only about 1% of total deposited structures [124], mainly because TM proteins are difficult to crystallize. The G-protein-coupled receptors (GPCRs) constitute the largest family of TM proteins and make up roughly 3% of genes in the human genome (over 800 genes encode a GPCR according to human genome research data) [125]. GPCRs can be divided into six classes with no shared sequence homology between classes [126]. The Rhodopsin family (also known as class A GPCRs) is the largest family and consists of 672 proteins, accounting for nearly 85% of the GPCR genes [127]. Over half of the GPCRs in class A are predicted to encode olfactory receptors, while the remaining receptors are bound with known endogenous compounds or are classified as orphan receptors [127]. Class A GPCRs are characterized by a small extracellular N-terminal domain, a canonical seven transmembrane domain, and a long intracellular C-terminal domain.

Knowledge of the 3D structure of GPCRs gives insight to the molecular mechanisms underlying diseases, syndromes caused by mutations in these receptors, and structure-based drug design. GPCRs are currently the single largest drug target family, representing 20-50% of marketed drugs [128]. Although structural understanding of GPCRs has benefited from a number of recent breakthroughs [129-133], coverage of the superfamily's phylogenetic tree is still incomplete. Relatively few high resolution GPCR structures are known [134]. Therefore, it is important to build 3D structural models of GPCRs. Such models can be achieved via homology modeling or new fold methods [135]. In this section, we review the application of homology modeling to GPCRs, focusing on template selection.

The key step in homology modeling is deciding the template GPCR structure that will maximize the predicted model's accuracy. Current structural data includes five class A GPCRs (rhodopsin, beta-1 adrenergic receptor, beta-2 adrenergic receptor, adenosine A_{2A} receptor, and CXCR4 chemokine receptor) (see Table **3**) and one class B GPCR [136]. The other GPCR families possess no structural representatives. The lack of structural data for these classes makes them unsuitable for comparative modeling as discussed earlier. We therefore focus on homology modeling of class A GPCRs. The most studied structures available for class A GPCR modeling have been those of bovine rhodopsin. Since the year 2000, when the first X-ray structure of bovine rhodopsin was solved, an additional eighteen rhodopsin structures in different activation states have been made available (see Table **3**). However, rhodopsin is a light activated GPCR and is distant in sequence homology to other class A GPCRs. Given that sequence similarity has a strong bearing on resulting model accuracy, there is a degree of uncertainty in using rhodopsin X-ray structures as templates for homology modeling of other GPCR targets. Potential complications include target-template alignment error due to low homology and uncertainty of whether other GPCR proteins would adopt the same binding site geometry. The many conformations of the binding site depend on the nature and function of the ligands. Therefore, modeling the conformational changes resulting from GPCR activation is a challenging task. Most class A GPCR structures were crystallized with inverse agonists or antagonists and therefore represent inactive conformations. However, the recent publication of rhodopsin and rhodopsin bound to a G-protein derived synthetic peptide may represent inactivated and activated states, respectively, providing important insight to the structural changes associated with GPCR activation [137]. Such examples of different activation states will facilitate the creation of active and inactive GPCR homology models, as it is ideal to use templates with similar types of interactions.

Despite the uncertainties and difficulties in constructing rhodopsin-based GPCR homology models, successful outcomes have been reported, particularly in ligand-oriented homology modeling [138, 139]. One scenario involved a community-wide modeling and docking experiment prior to the publication of the human adenosine A_{2A} receptor crystal structure [140]. Evaluation and analysis of the resulting predictions suggested the importance of using additional biochemical insight such as disulphide bridges in the extracellular loops. Other reported examples also suggest that cautious incorporation of knowledge-based constraints (site-direct mutagenesis and ligand binding data) can improve the quality of models and ligand docking [139, 140]. Such sources of information are suitable for inclusion through model building by spatial restraints satisfaction.

Additional insight to modeling class A GPCRs comes from recently published ligand-bound GPCR structures. These structures help thaw the frozen picture of proteins in this class. Comparative modeling using these different templates provides the first opportunity to examine changes in the predicted structures and possible consequences. Already, comparisons between available structures are being studied. The four structures of beta-2 adrenergic receptor bound with antagonists or partial inverse agonists are very similar to each other [129, 132]. Their overall architecture resembles that of rhodopsin, but with changes in the tertiary structure and position of helices, as well as a more open binding pocket. Because beta-2 and beta-1 adrenergic receptors are very closely related, it is not surprising that their ligand binding sites are similar, with expected differences due to different bound ligands [141]. Recently [142], the human beta-2 adrenergic receptor was also solved in an agonist-bound active form. Comparison to a crystal structure of the inactive state revealed subtle changes in the binding pocket. However, these small changes are associated with movement and rearrangement of the transmembrane helices, similar to those observed in opsin, an active form of rhodopsin. This new structure provides further insight into the process of agonist binding and activation. Intermediate in

Table 2. Resources and Tools for Drug Design

Name	URL	Ref.
Databases of Proteins and Ligands		
ChEMBL	https://www.ebi.ac.uk/chembl	
DrugBank	http://www.drugbank.ca	[190]
PDTD	http://www.dddc.ac.cn/pdtd	[191]
PubChem	http://pubchem.ncbi.nlm.nih.gov	[192]
sc-PDB	http://bioinfo-pharma.u-strasbg.fr/scPDB	[193]
Binding Site Prediction		
3DLigandSite	http://www.sbg.bio.ic.ac.uk/~3dligandsite	[194]
MetaPocket	http://metapocket.eml.org	[195]
PocketDepth	http://proline.physics.iisc.ernet.in/pocketdepth	[196]
Docking Tools		
AutoDock	http://autodock.scripps.edu	[197]
FINDSITE^{LHM}	http://cssb.biology.gatech.edu/findsitelhm	[198]
FLIPDock	http://flipdock.scripps.edu	[199]
MEDock	http://bioinfo.mc.ntu.edu.tw/medock	[200]
PATCHDOCK	http://bioinfo3d.cs.tau.ac.il/PatchDock	[201]
SwissDock	http://swissdock.vital-it.ch	[202]
Virtual Screening and Drug Design		
ANCHOR	http://structure.pitt.edu/anchor	[203]
e-LEA3D	http://bioinfo.ipmc.cnrs.fr/lea.html	[204]
PharmMapper	http://59.78.96.61/pharmmapper	[205]
SuperPred	http://bioinformatics.charite.de/superpred	[206]
SPROUT	http://www.simbiosys.ca/sprout	[206]

similarity to the beta-2 adrenergic receptors and beta-1 adrenergic receptors is the structure of the A_{2A} adenosine receptor in complex with a high-affinity antagonist. This structure sheds light on structural divergence as a function of sequence divergence. Understanding this relationship is crucial because comparative modeling uses sequence changes to inform structural changes.

Worth *et al*. performed a systematic analysis of sequence-structure relationships for known GPCRs and concluded that available structures represented only a subset of all possible class A GPCR structural variations [143]. For a given GPCR, some structural features may be represented in a subset of crystal structures or not at all, suggesting that models should be built using multiple templates. Multiple template comparative modeling (using all currently available GPCR structures) provides an improvement over single template modeling, as evaluated by the accuracy of rigid protein-flexible ligand docking on these models [143, 144]. Yarnitzky *et al*. also investigated the choice of experimental templates [145]. They similarly concluded that multiple template or fragment-based modeling could produce better models than single template modeling. They also suggested that molecular dynamics simulation be used to sample structural features not observed in X-ray structures to improve refinement. The use of varied template selection has also seen success with other comparative modeling applications discussed earlier.

Despite recent breakthroughs in GPCR structural biology, the structures discussed above cover only three of the nineteen class A subfamilies. These structures also exhibit high conformational similarity in their transmembrane regions, indicating more GPCR crystal structures in various conformational states are needed to resolve GPCR structure-function relationships. In 2009, Mobarec *et al.* investigated the 22 available structures and suggested that better templates were required to generate models with sufficient accuracy for structure-based drug discovery [144]. However, this analysis was based on information prior to the 2010 breakthroughs, namely the new structures of CXCR4 chemokine receptor bound to two different drug-like antagonists [134]. The overall structure of CXCR4 has moderate differences from rhodopsin, beta-2 adrenergic, beta-1 adrenergic, and alpha-2A adrenergic receptors in terms of the length of helix V, VI, and VII. Compared to rhodopsin and beta-2 adrenergic receptor, the binding site in CXCR4 is closer to the extracellular surface and is thus larger and more open. The fact that CXCR4's ligand binding site is bound to two structurally dissimilar antagonists suggests a degree of structural plasticity to GPCR binding sites. This poses a problem to comparative

Table 3. Class A GPCRs with Experimentally Determined 3D Structures

Protein Name	PDB Code
Rhodopsin (opsin)	1f88, 1hzx, 1l9h, 1gzm, 1u19, 2hpy, 2g87, 2i35, 2i36, 2i37, 2j4y, 2ped, 2ziy, 2z73, 3cap, 3c9l, 3c9m, 3dqb
Adenosine-A_{2A} receptor	3eml
Beta-1 adrenergic receptor	2vt4
Beta-2 adrenergic receptor	2rh1, 2r4r, 2r4s, 3d4s, 3nya, 3ny8, 3ny9, 3p0g
CXCR4 chemokine receptor	3oe0, 3oe5, 3oe6, 3odu

modeling as it complicates the sequence-structure relationship and may partially account for the moderate performance of existing GPCR templates. Nonetheless, these newly solved structures of CXCR4 introduce valuable information to GPCR modeling, which is heavily reliant on existing structural knowledge of GPCRs. Recently, the GPCRRD database (http://zhanglab.ccmb.med.umich.edu/GPCRRD/) collecting experimental restraints available from the literature has been implemented to assist GPCR structure modeling and function annotation [146]. In summary, with increasing availability of experimental structures, exciting possibilities for advancements in GPCR comparative modeling are expected.

Modeling Protein Kinases

Protein kinases are involved in numerous cellular processes including cell proliferation, differentiation, inflammation, and apoptosis. Abnormal activity has been linked to a variety of diseases including cancer, Alzheimer's, and inflammation [147-150]. Regulation of protein kinase activity is therefore an important therapeutic strategy. Protein kinase is one of the largest enzyme families, covering approximately 2% of the human proteome. The Protein Kinase Resource (http://pkr.genomics.purdue.edu) [151] provides a comprehensive coverage of all kinase-related data and derived information, including 457 kinase structures of which 296 are from human. It has been estimated that 518 different human protein kinases exist. According to the sequence similarity of their catalytic domains, 409 kinases have been grouped into eight major kinase families (AGC, CAMK, CK1, CMGC, RGC, STE, TK, and TKL) and the remaining ones into the "others" and "atypical" groups [152]. Human protein kinases can also be classified using various other rules. An alternative and simple classification scheme uses substrate preferences to divide them into serine/threonine, tyrosine, histidine, and aspartic/glutamic kinases. Serine/threonine kinases have Enzyme Classification (EC) number 2.7.11.1 while tyrosine kinases have EC number 2.7.10.1 or 2.7.10.2. More recently, a large-scale analysis of ATP binding sites helped define a new classification of protein kinase families based on structural similarity [153]. The high sequence and structural similarity within the families make unsolved kinases ideal targets for comparative modeling and subsequent virtual screening. We focus this section on prediction of the two largest families, serine/threonine and tyrosine kinases.

Kinases phosphorylate substrates through a conserved catalytic domain. The N-terminal lobe is composed of β-strands in anti-parallel conformation and one α-helix (αC), while the C-terminal lobe is composed of multiple α-helices (see Fig. **4**). A large loop between the first and second β-strands in the N-terminal lobe interacts with the phosphate groups of ATP. The two lobes are connected by the linker or hinge region. Interactions in the interface between the N-terminal lobe, C-terminal lobe, ATP, and other ligands have been previously described [154, 155].

At the interface between the N and C-terminal lobes, protein kinases have an activation loop whose phosphorylation state determines its activation state. Phosphorylated and unphosphorylated loops correspond to active and inactive states, respectively. The catalytic site undergoes significant structural rearrangement when the kinase switches between states (see Fig. **4**). This distinction is important for comparative modeling because the state of the template will affect the state of the modeled target. Upon activation, the loop is released, making the binding site accessible. The position of the αC helix also moves closer to the binding site. In the particular case of human CDK2, the distance between Lys33 and Asp86 (Fig. **4C**) and Asp127 and Asp145 (Fig. **4B**) reduces by ~4 Å each. A complete description of structural changes between active and inactive conformations and conservation of regions involved in the catalytic process across kinase families has been previously described [156, 157].

Characterization of the active site has largely focused on the active conformation rather than the inactive conformation. While knowledge of the active state has provided important insight to the design of new compounds, a large number of inhibitors bind to the inactive conformation of these proteins. It has been observed that catalytic domains of dissimilar kinases adopt similar active conformations but highly variable inactive conformations [158]. This structural plasticity is a significant barrier to the use of comparative modeling in virtual screening. This is because building inactive kinase models using templates in the active state produces models indistinguishable from the templates. Template selection is not an issue; rather, alternative approaches focused on the prediction of loop regions in the catalytic pocket of inactive kinases are needed.

Although structure prediction of the inactive state of protein kinases is difficult, during recent years, several structures predicted by comparative modeling have been successfully used for non-virtual and virtual screening [123]. The first success came from virtual screening against the G protein kinase GRK2. The model was built from a cAPK template and tested against a library of 13,000 compounds [159]. Results showed that only one of four high scoring ligands displayed any inhibition activity against GRK2.

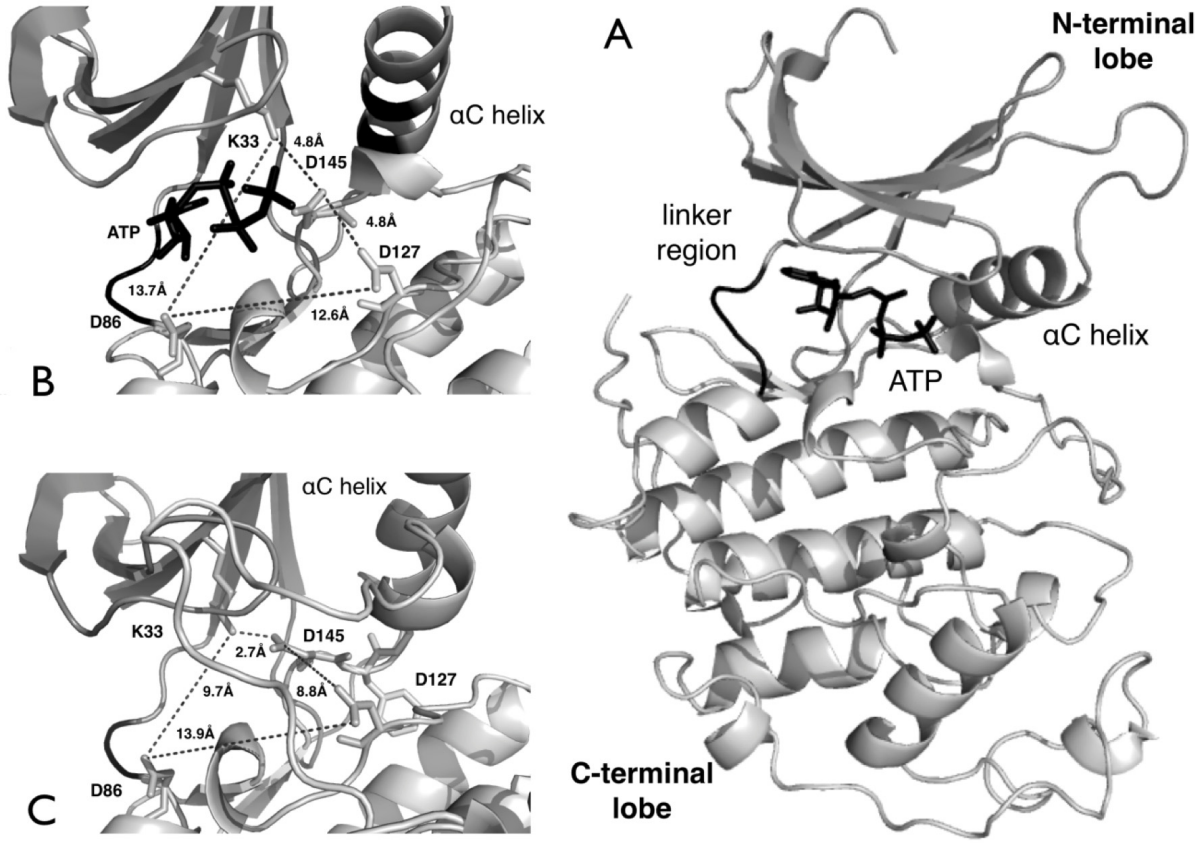

Fig. (4). Structural comparison of human CDK2 in active (PDB: 1FIN) and inactive states (PDB: 1HCL).

Later, a highly potent inhibitor of human CK2 was found using a model built from the structure of ANP-bound CK2 from *Zea mays* [116]. A docking procedure selected a subset of ~1,600 high scoring ligands from an initial library of 400,000 compounds. After visual inspection, this was reduced to twelve potential inhibitors, one of which inhibited human CK2 in experimental assays. In this particular case, sequence identity between target and template was 92%, suggesting a high quality model. Crystallographic experiments confirmed this, showing ATP binding site Root Mean Square Deviation (RMSD) of 0.64 Å between predicted and experimental structures. Additionally, in a 2005 paper [160], authors described the screening of ~6,500 compounds using a model of JAK2 created from a FGFR1 template. Extensive experimental tests demonstrated that one of seven high scoring compounds inhibited JAK2 tyrosine autophosphorylation.

In the cases presented, template selection took into account the nature of the co-crystallized ligand. This is motivated by the link between ligand structure and loop conformation, meaning loop conformation and the desired interactions are better represented when a similar compound is bound. This suggests that the best template for virtual screening has a similar sequence to the target as well as a similar bound ligand to the compounds being tested, as discussed earlier. Although an exhaustive study of the effect of ligation state on docking efficiency to the template has not been performed, it is evident that ligation state affects the structure of the predicted model [161]. This is also demonstrated in a recent large-scale structure modeling and virtual screening study of the entire human kinome [162]. Sequence profile-based alignments with ligand-bound and ligand-free templates were used to predict the structures of all human kinases and their ATP-binding sites. Virtually screening of more than two million compounds led to docking refinement of five million ligand-kinase complexes that were evaluated using different scoring functions. Modeling accuracy was validated by experimental data and all predicted structures, ligand conformations, and ligand ranking lists were made available online (http://cssb.biology.gatech.edu/kinomelhm/). The authors tested their structural predictions by comparing them against the experimental structures of 57 ligand-bound (holo) and 48 ligand-free (apo) human kinases. The predicted structures gave average RMSDs of 2.75 Å and 3.13 Å, respectively. The lower RMSD obtained for holo versus apo conformations reflects the fact that a larger number of templates used were in a ligand-bound active state. Similar analysis on the kinase binding sites showed average RMSDs of 1.27 Å and 2.36 Å for C_α and all atoms, respectively. This difference confirms that modeling of side chain atoms still needs further improvement. Nevertheless, the results are compatible with the estimated binding site plasticity that allows two protein kinases in the same family to bind the same class of ligands.

FUTURE OUTLOOK

During the last decade, comparative modeling techniques have become routinely used in many practical applications. There has been continuous improvement to the overall accuracy of predicted models due to better methods for

template selection, sequence alignment, and evaluation. Small advances have also been made in model refinement. Many applications, in particular virtual screening and drug design, are critically dependent on the accuracy achieved by comparative modeling methods. We have presented successful examples on both GPCRs and protein kinases. However, the speed at which sequence data is produced and the large number of models that we can obtain from them will require fast and accurate computational tools to evaluate the quality of the predicted structures. Additionally, continued computational speedups and more accurate scoring functions will be required to achieve an exhaustive sampling of binding site conformational space and selection of the best target-ligand complexes.

Another important challenge will be the development of high quality repositories for experimental and computational data. In the specific case of virtual screening and drug design, we believe that a better integration of different sources of data is needed. These include broad and continuously updated databases that collect the structures of possible drug targets and compounds as well as available binding assays. This will facilitate characterization of ligand binding sites and testing of new data-driven hypotheses. We also expect increased development of protein family specific databases, where researchers interested in a particular class of proteins can deposit and retrieve highly curated data. In this direction are the SARfari databases that integrate chemogenomics information for GPCRs and kinases [163]. These portals contain sequence, alignment, structure, and screening data in a user-friendly web interface that allows exploration across public and private data. Computational data from modeling and docking simulations should be similarly collected and made available to the scientific community. Most important is reporting all the information needed to reproduce the results. A centralized database collecting predicted structures from comparative modeling in concert with virtual and non-virtual screening data will be particularly useful to restrict the search of new drugs to a smaller set of plausible compounds. Although the improvement of comparative modeling methods will be a key factor in obtaining high quality predictions, the collection of well-curated and integrated data will allow resource optimization, making the selection of new potential drug target interactions more efficient.

ACKNOWLEDGMENTS

The authors would like to thank Professor Russ B. Altman for his comments and suggestions. TL and GWT acknowledge support from the SIMBIOS National Center for Simulation of Biological Structures (GM-61374) and NIH (LM-05652). GWT also acknowledges support from the Stanford Bio-X Bioengineering Graduate Fellowship. EC acknowledges support from the Marie Curie International Outgoing Fellowship program (PIOF-GA-2009-237225) funded by the European Commission's FP7 grant.

REFERENCES

[1] Chothia, C.; Lesk, A.M. The relation between the divergence of sequence and structure in proteins. *EMBO J.,* **1986**, *5*(4), 823-826.

[2] Chothia, C.; Lesk, A.M. The evolution of protein structures. *Cold. Spring. Harb. Symp. Quant. Biol.,* **1987**, *52*, 399-405.

[3] Bajaj, M.; Blundell, T. Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng.,* **1984**, *13*, 453-492.

[4] Owen, R. *Lectures on the comparative anatomy and physiology of the invertebrate animals,* delivered at the royal college of surgeons, Longman, Brown, Green, and Longmans: London, **1843**.

[5] Reeck, G.R.; de Haen, C.; Teller, D.C.; Doolittle, R.F.; Fitch, W.M.; Dickerson, R.E.; Chambon, P.; McLachlan, A.D.; Margoliash, E.; Jukes, T.H.; *et al*. "Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell,* **1987**, *50*(5), 667.

[6] Dill, K.A.; Ozkan, S.B.; Weikl, T.R.; Chodera, J.D.; Voelz, V.A. The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.,* **2007**, *17*(3), 342-346.

[7] Klepeis, J.L.; Wei, Y.; Hecht, M.H.; Floudas, C.A. Ab initio prediction of the three-dimensional structure of a de novo designed protein: a double-blind case study. *Proteins,* **2005**, *58*(3), 560-570.

[8] Muff, S.; Caflisch, A. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein. *Proteins,* **2008**, *70*(4), 1185-1195.

[9] Rohl, C.A.; Strauss, C.E.; Misura, K.M.; Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.,* **2004**, *383*, 66-93.

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

[11] Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.,* **2007**, *35*(Database issue), D301-303.

[12] Andreeva, A.; Howorth, D.; Chandonia, J.M.; Brenner, S.E.; Hubbard, T.J.; Chothia, C.; Murzin, A.G. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.,* **2008**, *36*(Database issue), D419-425.

[13] Dessailly, B.H.; Nair, R.; Jaroszewski, L.; Fajardo, J.E.; Kouranov, A.; Lee, D.; Fiser, A.; Godzik, A.; Rost, B.; Orengo, C. PSI-2: structural genomics to cover protein domain family space. *Structure,* **2009**, *17*(6), 869-881.

[14] Chandonia, J.M.; Brenner, S.E. The impact of structural genomics: expectations and outcomes. *Science,* **2006**, *311*(5759), 347-351.

[15] Marsden, R.L.; Orengo, C.A. Target selection for structural genomics: an overview. *Methods Mol. Biol.,* **2008**, *426*, 3-25.

[16] Reeves, G.A.; Dallman, T.J.; Redfern, O.C.; Akpor, A.; Orengo, C.A. Structural diversity of domain superfamilies in the CATH database. *J. Mol. Biol.,* **2006**, *360*(3), 725-741.

[17] Todd, A.E.; Orengo, C.A.; Thornton, J.M. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.,* **2001**, *307*(4), 1113-1143.

[18] Kryshtafovych, A.; Fidelis, K.; Moult, J. CASP8 results in context of previous experiments. *Proteins.* **2009**, *77* (Suppl 9), 217-228.

[19] Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.,* **1999**, *12*(2), 85-94.

[20] Worth, C.L.; Gong, S.; Blundell, T.L. Structural and functional constraints in the evolution of protein families. *Nat. Rev. Mol. Cell Biol.,* **2009**, *10*(10), 709-720.

[21] Holm, L.; Sander, C. The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.,* **1996**, *24*(1), 206-209.

[22] Sander, C.; Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins,* **1991**, *9*(1), 56-68.

[23] Holm, L.; Sander, C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.,* **1998**, *26*(1), 316-319.

[24] Cuff, A.L.; Sillitoe, I.; Lewis, T.; Redfern, O.C.; Garratt, R.; Thornton, J.; Orengo, C.A. The CATH classification revisited--architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D310-314.

[25] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.,* **1990**, *215*(3), 403-410.

[26] Pearson, W.R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.,* **1990**, *183*, 63-98.

[27] Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* **1997**, *25*(17), 3389-3402.

[28] Capriotti, E.; Fariselli, P.; Rossi, I.; Casadio, R. A Shannon entropy-based filter detects high- quality profile-profile alignments in searches for remote homologues. *Proteins,* **2004**, *54*(2), 351-360.

[29] Marti-Renom, M.A.; Madhusudhan, M.S.; Sali, A. Alignment of protein sequences by their profiles. *Protein Sci.,* **2004**, *13*(4), 1071-1087.

[30] Rychlewski, L.; Jaroszewski, L.; Li, W.; Godzik, A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.,* **2000**, *9*(2), 232-241.

[31] Sadreyev, R.I.; Baker, D.; Grishin, N.V. Profile-profile comparisons by COMPASS predict intricate homologies between protein families. *Protein Sci.,* **2003**, *12*(10), 2262-2272.

[32] Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics,* **2005**, *21*(7), 951-960.

[33] Karplus, K.; Barrett, C.; Hughey, R. Hidden Markov models for detecting remote protein homologies. *Bioinformatics,* **1998**, *14*(10), 846-856.

[34] Eddy, S.R. Profile hidden Markov models. *Bioinformatics,* **1998**, *14*(9), 755-763.

[35] Fariselli, P.; Rossi, I.; Capriotti, E.; Casadio, R. The WWWH of remote homolog detection: the state of the art. *Brief. Bioinform.,* **2007**, *8*(2), 78-87.

[36] Larsson, P.; Wallner, B.; Lindahl, E.; Elofsson, A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.,* **2008**, *17*(6), 990-1002.

[37] Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.,* **1970**, *48*(3), 443-453.

[38] Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.,* **1981**, *147*(1), 195-197.

[39] Henikoff, S.; Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA,* **1992**, *89*(22), 10915-10919.

[40] Dayhoff, M.O.; Schwartz, R.; Orcutt, B.C. In Atlas of protein sequence and structure; Nat. Biomed. Res. Found.; 1978, pp 345-358.

[41] Jones, D.T.; Tress, M.; Bryson, K.; Hadley, C. Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins,* **1999**, *Suppl 3*, 104-111.

[42] Marsden, R.L.; McGuffin, L.J.; Jones, D.T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.,* **2002**, *11*(12), 2814-2824.

[43] Sutcliffe, M.J.; Haneef, I.; Carney, D.; Blundell, T.L. Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.,* **1987**, *1*(5), 377-384.

[44] Levitt, M. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.,* **1992**, *226*(2), 507-533.

[45] Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.,* **1993**, *234*(3), 779-815.

[46] Unger, R.; Harel, D.; Wherland, S.; Sussman, J.L. A 3D building blocks approach to analyzing and predicting structure of proteins. *Proteins,* **1989**, *5*(4), 355-373.

[47] de la Paz, P.; Sutton, B.J.; Darsley, M.J.; Rees, A.R. Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.,* **1986**, *5*(2), 415-425.

[48] Jones, T.A.; Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.,* **1986**, *5*(4), 819-822.

[49] Bruccoleri, R.E.; Karplus, M. Conformational sampling using high-temperature molecular dynamics. *Biopolymers,* **1990**, *29*(14), 1847-1862.

[50] Bower, M.J.; Cohen, F.E.; Dunbrack, R.L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.,* **1997**, *267*(5), 1268-1282.

[51] Shih, H.H.; Brady, J.; Karplus, M. Structure of proteins with single-site mutations: a minimum perturbation approach. *Proc. Natl. Acad. Sci. USA,* **1985**, *82*(6), 1697-1700.

[52] Snow, M.E.; Amzel, L.M. Calculating three-dimensional changes in protein structure due to amino-acid substitutions: the variable region of immunoglobulins. *Proteins,* **1986**, *1*(3), 267-279.

[53] Fan, H.; Mark, A.E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci.,* **2004**, *13*(1), 211-220.

[54] Flohil, J.A.; Vriend, G.; Berendsen, H.J. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins,* **2002**, *48*(4), 593-604.

[55] Linge, J.P.; Williams, M.A.; Spronk, C.A.; Bonvin, A.M.; Nilges, M. Refinement of protein structures in explicit solvent. *Proteins,* **2003**, *50*(3), 496-506.

[56] Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Free energy landscapes of peptides by enhanced conformational sampling. *J. Mol. Biol.,* **2000**, *296*(1), 197-216.

[57] Lee, M.R.; Tsai, J.; Baker, D.; Kollman, P.A. Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Biol.,* **2001**, *313*(2), 417-430.

[58] Wallner, B.; Elofsson, A. All are not equal: a benchmark of different homology modeling programs. *Protein Sci.,* **2005**, *14*(5), 1315-1327.

[59] Laskowski, R.A.; Rullmannn, J.A.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR,* **1996**, *8*(4), 477-486.

[60] Oldfield, T.J. SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. *J. Mol. Graph.,* **1992**, *10*(4), 247-252.

[61] Vaguine, A.A.; Richelle, J.; Wodak, S.J. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D Biol. Crystallogr.,* **1999**, *55*(Pt 1), 191-205.

[62] Chen, V.B.; Arendall, W.B., 3rd; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr., 66*(Pt 1), 12-21.

[63] Luthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature,* **1992**, *356*(6364), 83-85.

[64] Melo, F.; Feytmans, E. Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.,* **1998**, *277*(5), 1141-1152.

[65] Sippl, M.J. Recognition of errors in three-dimensional structures of proteins. *Proteins,* **1993**, *17*(4), 355-362.

[66] Topham, C.M.; Srinivasan, N.; Thorpe, C.J.; Overington, J.P.; Kalsheker, N.A. Comparative modelling of major house dust mite allergen Der p I: structure validation using an extended environmental amino acid propensity table. *Protein Eng.,* **1994**, *7*(7), 869-894.

[67] Finkelstein, A.V.; Badretdinov, A.; Gutin, A.M. Why do protein architectures have Boltzmann-like statistics? *Proteins,* **1995**, *23*(2), 142-150.

[68] Rooman, M.J.; Wodak, S.J. Are database-derived potentials valid for scoring both forward and inverted protein folding? *Protein Eng.,* **1995**, *8*(9), 849-858.

[69] Thomas, P.D.; Dill, K.A. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.,* **1996**, *257*(2), 457-469.

[70] Capriotti, E.; Marti-Renom, M.A. *In Computational Structural Biology: Methods and Applications*; Schwede, Peitsch, Eds.; World Scientific Publishing Company, 2008; pp 89-109.

[71] Benkert, P.; Tosatto, S.C.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins,* **2008**, *71*(1), 261-277.

[72] Benkert, P.; Tosatto, S.C.; Schwede, T. Global and local model quality estimation at CASP8 using the scoring functions QMEAN and QMEANclust. *Proteins,* **2009**, *77*(Suppl 9), 173-180.

[73] Case, D.A.; Cheatham, T.E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.,* **2005**, *26*(16), 1668-1688.

[74] Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.,* **1983**, *4*(2), 187-217.

[75] Jagielska, A.; Wroblewska, L.; Skolnick, J. Protein model refinement using an optimized physics-based all-atom force field. *Proc. Natl. Acad. Sci. USA,* **2008**, *105*(24), 8268-8273.

[76] Lazaridis, T.; Karplus, M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.,* **1999**, *288*(3), 477-487.

[77] McGuffin, L.J.; Jones, D.T. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics,* **2003**, *19*(7), 874-881.

[78] Skolnick, J.; Kihara, D.; Zhang, Y. Development and large scale benchmark testing of the PROSPECTOR_3 threading algorithm. *Proteins,* **2004**, *56*(3), 502-518.

[79] Sommer, I.; Zien, A.; von Ohsen, N.; Zimmer, R.; Lengauer, T. Confidence measures for protein fold recognition. *Bioinformatics,* **2002**, *18*(6), 802-812.

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

[81] McGuffin, L.J.; Roche, D.B. Rapid model quality assessment for protein structure predictions using the comparison of multiple models without structural alignments. *Bioinformatics,* **2010**, *26*(2), 182-188.

[82] Zhang, Y.; Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *J. Comput. Chem.,* **2004**, *25*(6), 865-871.

[83] Chakravarty, S.; Godbole, S.; Zhang, B.; Berger, S.; Sanchez, R. Systematic analysis of the effect of multiple templates on the accuracy of comparative models of protein structure. *BMC Struct. Biol.,* **2008**, *8*, 31.

[84] Fernandez-Fuentes, N.; Rai, B.K.; Madrid-Aliste, C.J.; Fajardo, J.E.; Fiser, A. Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics,* **2007**, *23*(19), 2558-2565.

[85] Venclovas, C.; Margelevicius, M. The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins. *Proteins,* **2009**, *77 Suppl 9*, 81-88.

[86] Cozzetto, D.; Kryshtafovych, A.; Fidelis, K.; Moult, J.; Rost, B.; Tramontano, A. Evaluation of template-based models in CASP8 with standard measures. *Proteins,* **2009**, *77*(Suppl 9), 18-28.

[87] Cozzetto, D.; Kryshtafovych, A.; Tramontano, A. Evaluation of CASP8 model quality predictions. *Proteins,* **2009**, *77*(Suppl 9), 157-166.

[88] Pieper, U.; Eswar, N.; Webb, B.M.; Eramian, D.; Kelly, L.; Barkan, D.T.; Carter, H.; Mankoo, P.; Karchin, R.; Marti-Renom, M.A.; Davis, F.P.; Sali, A. MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D347-354.

[89] Kiefer, F.; Arnold, K.; Kunzli, M.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D387-392.

[90] Arnold, K.; Kiefer, F.; Kopp, J.; Battey, J.N.; Podvinec, M.; Westbrook, J.D.; Berman, H.M.; Bordoli, L.; Schwede, T. The Protein Model Portal. *J. Struct. Funct. Genomics,* **2009**, *10*(1), 1-8.

[91] Melo, F.; Devos, D.; Depiereux, E.; Feytmans, E. ANOLEA: a www server to assess protein structures. *Proc. Int. Conf. Intell. Syst. Mol. Biol.,* **1997**, *5*, 187-190.

[92] Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.,* **2007**, *35*(Web Server issue), W407-410.

[93] Kretsinger, R.H.; Barry, C.D. The predicted structure of the calcium-binding component of troponin. *Biochim. Biophys. Acta,* **1975**, *405*(1), 40-52.

[94] McLachlan, A.D. The double helix coiled coil structure of murein lipoprotein from Escherichia coli. *J. Mol. Biol.,* **1978**, *121*(4), 493-506.

[95] Isaacs, N.; James, R.; Niall, H.; Bryant-Greenwood, G.; Dodson, G.; Evans, A.; North, A.C. Relaxin and its structural relationship to insulin. *Nature,* **1978**, *271*(5642), 278-281.

[96] Sanchez, R.; Sali, A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins,* **1997**, *Suppl 1*, 50-58.

[97] Sanchez, R.; Sali, A. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proc. Natl. Acad. Sci. USA,* **1998**, *95*(23), 13597-13602.

[98] Matsumoto, R.; Sali, A.; Ghildyal, N.; Karplus, M.; Stevens, R.L. Packaging of proteases and proteoglycans in the granules of mast cells and other hematopoietic cells. A cluster of histidines on mouse mast cell protease 7 regulates its binding to heparin serglycin proteoglycans. *J. Biol. Chem.,* **1995**, *270*(33), 19524-19531.

[99] Xu, L.Z.; Sanchez, R.; Sali, A.; Heintz, N. Ligand specificity of brain lipid-binding protein. *J. Biol. Chem.,* **1996**, *271*(40), 24711-24719.

[100] Sánchez, R.; Sali, A. Advances in comparative protein-structure modelling. *Curr. Opin. Struct. Biol.,* **1997**, *7*(2), 206: 214.

[101] Ring, C.S.; Sun, E.; McKerrow, J.H.; Lee, G.K.; Rosenthal, P.J.; Kuntz, I.D.; Cohen, F.E. Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. USA,* **1993**, *90*(8), 3583-3587.

[102] Totrov, M.; Abagyan, R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. *Nat. Struct. Biol.,* **1994**, *1*(4), 259-263.

[103] Vakser, I.A. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins,* **1997**, *Suppl 1*, 226-230.

[104] Giorgetti, A.; Raimondo, D.; Miele, A.E.; Tramontano, A. Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics,* **2005**, *21*(Suppl 2), ii72-76.

[105] Raimondo, D.; Giorgetti, A.; Bosi, S.; Tramontano, A. Automatic procedure for using models of proteins in molecular replacement. *Proteins,* **2007**, *66*(3), 689-696.

[106] Morea, V.; Tramontano, A.; Rustici, M.; Chothia, C.; Lesk, A.M. Antibody structure, prediction and redesign. *Biophys. Chem.,* **1997**, *68*(1-3), 9-16.

[107] Donini, M.; Morea, V.; Desiderio, A.; Pashkoulov, D.; Villani, M.E.; Tramontano, A.; Benvenuto, E. Engineering stable cytoplasmic intrabodies with designed specificity. *J. Mol. Biol.,* **2003**, *330*(2), 323-332.

[108] Kim, S.J.; Park, Y.; Hong, H.J. Antibody engineering for the development of therapeutic antibodies. *Mol. Cells,* **2005**, *20*(1), 17-29.

[109] Sanz, L.; Cuesta, A.M.; Compte, M.; Alvarez-Vallina, L. Antibody engineering: facing new challenges in cancer therapy. *Acta Pharmacol. Sin.,* **2005**, *26*(6), 641-648.

[110] Teillaud, J.L. Engineering of monoclonal antibodies and antibody-based fusion proteins: successes and challenges. *Expert Opin. Biol. Ther.,* **2005**, *5* (Suppl 1), S15-27.

[111] Wang, H.W.; Cole, D.; Jiang, W.Z.; Jin, H.T.; Fu, N.; Chen, Z.L.; Jin, N.Y. Engineering and functional evaluation of a single-chain antibody against HIV-1 external glycoprotein gp120. *Clin. Exp. Immunol.,* **2005**, *141*(1), 72-80.

[112] Goodsell, D.S.; Morris, G.M.; Olson, A.J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.,* **1996**, *9*(1), 1-5.

[113] Ma, B.; Shatsky, M.; Wolfson, H.J.; Nussinov, R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.,* **2002**, *11*(2), 184-197.

[114] Ooms, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.,* **2000**, *7*(2), 141-158.

[115] Ragno, R.; Simeoni, S.; Castellano, S.; Vicidomini, C.; Mai, A.; Caroli, A.; Tramontano, A.; Bonaccini, C.; Trojer, P.; Bauer, I.; Brosch, G.; Sbardella, G. Small molecule inhibitors of histone arginine methyltransferases: homology modeling, molecular docking, binding mode analysis, and biological evaluations. *J. Med. Chem.,* **2007**, *50*(6), 1241-1253.

[116] Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.,* **2003**, *46*(13), 2656-2662.

[117] Oprea, T.I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.,* **2004**, *8*(4), 349-358.

[118] Grant, M.A. Protein structure prediction in structure-based ligand design and virtual screening. *Comb. Chem. High Throughput Screen.,* **2009**, *12*(10), 940-960.

[119] Corbeil, C.R.; Therrien, E.; Moitessie, N. Modeling Reality for Optimal Docking of Small Molecules to Biological Targets. *Current Computer-Aided Drug Design,* **2009**, *5*, 241-263.

[120] Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins,* **2003**, *50*(1), 5-25.

[121] Diller, D.J.; Li, R. Kinases, homology models, and high throughput docking. *J. Med. Chem.,* **2003**, *46*(22), 4638-4647.

[122] Muegge, I.; Enyedy, I.J. Virtual screening for kinase targets. *Curr. Med. Chem.,* **2004**, *11*(6), 693-707.

[123] Rockey, W.M.; Elcock, A.H. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? *Curr. Protein Pept. Sci.,* **2006**, *7*(5), 437-457.

[124] White, S.H. The progress of membrane protein structure determination. *Protein Sci.,* **2004**, *13*(7), 1948-1949.

[125] Fredriksson, R.; Schioth, H.B. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.,* **2005**, *67*(5), 1414-1425.

[126] Attwood, T.K.; Findlay, J.B. Fingerprinting G-protein-coupled receptors. *Protein Eng.,* **1994**, *7*(2), 195-203.

[127] Millar, R.P.; Newton, C.L. The year in G protein-coupled receptor research. *Mol. Endocrinol.,* **2010**, *24*(1), 261-274.

[128] Schwalbe, H.; Wess, G. Dissecting G-protein-coupled receptors: structure, function, and ligand interaction. *Chembiochem,* **2002**, *3*(10), 915-919.

[129] Ak, T.; Gulcin, I. Antioxidant and radical scavenging properties of curcumin. *Chem Biol Interact.,* **2008**, *174*(1), 27-37.

[130] Cherezov, V.; Rosenbaum, D.M.; Hanson, M.A.; Rasmussen, S.G.; Thian, F.S.; Kobilka, T.S.; Choi, H.J.; Kuhn, P.; Weis, W.I.; Kobilka, B.K.; Stevens, R.C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science,* **2007**, *318*(5854), 1258-1265.

[131] Scheerer, P.; Park, J.H.; Hildebrand, P.W.; Kim, Y.J.; Krauss, N.; Choe, H.W.; Hofmann, K.P.; Ernst, O.P. Crystal structure of opsin in its G-protein-interacting conformation. *Nature,* **2008**, *455*(7212), 497-502.

[132] Wacker, D.; Fenalti, G.; Brown, M.A.; Katritch, V.; Abagyan, R.; Cherezov, V.; Stevens, R.C. Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.,* **2010**, *132*(33), 11443-11445.

[133] Wu, B.; Chien, E.Y.; Mol, C.D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F.C.; Hamel, D.J.; Kuhn, P.; Handel, T.M.; Cherezov, V.; Stevens, R.C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science,* **2010**, *330*(6007), 1066-1071.

[134] Cherezov, V.; Abola, E.; Stevens, R.C. Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods Mol. Biol.,* **2010**, *654*, 141-168.

[135] Becker, O.M.; Marantz, Y.; Shacham, S.; Inbal, B.; Heifetz, A.; Kalid, O.; Bar-Haim, S.; Warshaviak, D.; Fichman, M.; Noiman, S. G protein-coupled receptors: in silico drug discovery in 3D. *Proc. Natl. Acad. Sci. USA,* **2004**, *101*(31), 11304-11309.

[136] ter Haar, E.; Koth, C.M.; Abdul-Manan, N.; Swenson, L.; Coll, J.T.; Lippke, J.A.; Lepre, C.A.; Garcia-Guzman, M.; Moore, J.M. Crystal structure of the ectodomain complex of the CGRP receptor, a class-B GPCR, reveals the site of drug antagonism. *Structure,* **2010**, *18*(9), 1083-1093.

[137] Rasmussen, S.G.; Choi, H.J.; Rosenbaum, D.M.; Kobilka, T.S.; Thian, F.S.; Edwards, P.C.; Burghammer, M.; Ratnala, V.R.; Sanishvili, R.; Fischetti, R.F.; Schertler, G.F.; Weis, W.I.; Kobilka, B.K. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature,* **2007**, *450*(7168), 383-387.

[138] Katritch, V.; Rueda, M.; Lam, P.C.; Yeager, M.; Abagyan, R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins,* **2010**, *78*(1), 197-211.

[139] Radestock, S.; Weil, T.; Renner, S. Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring. *J. Chem. Inf. Model.,* **2008**, *48*(5), 1104-1117.

[140] Michino, M.; Abola, E.; Brooks, C.L., 3rd; Dixon, J.S.; Moult, J.; Stevens, R.C. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.,* **2009**, *8*(6), 455-463.

[141] Warne, T.; Serrano-Vega, M.J.; Baker, J.G.; Moukhametzianov, R.; Edwards, P.C.; Henderson, R.; Leslie, A.G.; Tate, C.G.; Schertler, G.F. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature,* **2008**, *454*(7203), 486-491.

[142] Rasmussen, S.G.; Choi, H.J.; Fung, J.J.; Pardon, E.; Casarosa, P.; Chae, P.S.; Devree, B.T.; Rosenbaum, D.M.; Thian, F.S.; Kobilka, T.S.; Schnapp, A.; Konetzki, I.; Sunahara, R.K.; Gellman, S.H.; Pautsch, A.; Steyaert, J.; Weis, W.I.; Kobilka, B.K. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature, 469*(7329), 175-180.

[143] Worth, C.L.; Kleinau, G.; Krause, G. Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One,* **2009**, *4*(9), e7011.

[144] Mobarec, J.C.; Sanchez, R.; Filizola, M. Modern homology modeling of G-protein coupled receptors: which structural template to use? *J. Med. Chem.,* **2009**, *52*(16), 5207-5216.

[145] Yarnitzky, T.; Levit, A.; Niv, M.Y. Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. *Curr. Opin. Drug Discov. Devel.,* **2010**, *13*(3), 317-325.

[146] Zhang, J.; Zhang, Y. GPCRRD: G protein-coupled receptor spatial restraint database for 3D structure modeling and function annotation. *Bioinformatics, 26*(23), 3004-3005.

[147] Adcock, I.M.; Chung, K.F.; Caramori, G.; Ito, K. Kinase inhibitors and airway inflammation. *Eur. J. Pharmacol.,* **2006**, *533*(1-3), 118-132.

[148] Basu, A. The potential of protein kinase C as a target for anticancer treatment. *Pharmacol. Ther.,* **1993**, *59*(3), 257-280.

[149] Bradshaw, D.; Hill, C.H.; Nixon, J.S.; Wilkinson, S.E. Therapeutic potential of protein kinase C inhibitors. *Agents Actions,* **1993**, *38*(1-2), 135-147.

[150] Leclerc, S.; Garnier, M.; Hoessel, R.; Marko, D.; Bibb, J.A.; Snyder, G.L.; Greengard, P.; Biernat, J.; Wu, Y.Z.; Mandelkow, E.M.; Eisenbrand, G.; Meijer, L. Indirubins inhibit glycogen synthase kinase-3 beta and CDK5/p25, two protein kinases involved in abnormal tau phosphorylation in Alzheimer's disease. A property common to most cyclin-dependent kinase inhibitors? *J. Biol. Chem.,* **2001**, *276*(1), 251-260.

[151] Scheeff, E.D.; Bourne, P.E. Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.,* **2005**, *1*(5), e49.

[152] Manning, G.; Whyte, D.B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science,* **2002**, *298*(5600), 1912-1934.

[153] Kinnings, S.L.; Jackson, R.M. Binding site similarity analysis for the functional classification of the protein kinase family. *J. Chem. Inf. Model.,* **2009**, *49*(2), 318-329.

[154] Mao, L.; Wang, Y.; Liu, Y.; Hu, X. Molecular determinants for ATP-binding in proteins: a data mining and quantum chemical analysis. *J. Mol. Biol.*, **2004**, *336*(3), 787-807.

[155] Traxler, P.; Furet, P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol. Ther.,* **1999**, *82*(2-3), 195-206.

[156] Kornev, A.P.; Haste, N.M.; Taylor, S.S.; Eyck, L.F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.,* **2006**, *103*(47), 17783-17788.

[157] Thompson, E.E.; Kornev, A.P.; Kannan, N.; Kim, C.; Ten Eyck, L.F.; Taylor, S.S. Comparative surface geometry of the protein kinase family. *Protein Sci.,* **2009**, *18*(10), 2016-2026.

[158] Huse, M.; Kuriyan, J. The conformational plasticity of protein kinases. *Cell,* **2002**, *109*(3), 275-282.

[159] Kassack, M.U.; Hogger, P.; Gschwend, D.A.; Kameyama, K.; Haga, T.; Graul, R.C.; Sadee, W. Molecular modeling of G-protein coupled receptor kinase 2: docking and biochemical evaluation of inhibitors. *AAPS PharmSci,* **2000**, *2*(1), E2.

[160] Sandberg, E.M.; Ma, X.; He, K.; Frank, S.J.; Ostrov, D.A.; Sayeski, P.P. Identification of 1,2,3,4,5,6-hexabromocyclohexane as a small molecule inhibitor of jak2 tyrosine kinase autophosphorylation [correction of autophophorylation]. *J. Med. Chem.,* **2005**, *48*(7), 2526-2533.

[161] Jacobson, M.; Sali, A. In Annu. Rep. Med. Chem.; Elsevier Academic Press Inc: San Diego, CA., 2004; pp 259-276.

[162] Brylinski, M.; Skolnick, J. Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening. *J. Chem. Inf. Model,* **2010**, *50*(10), 1839-1854.

[163] Overington, J. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. *J. Comput. Aided Mol. Des,* **2009**, *23*(4), 195-198.

[164] UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.,* **2010**, *38*(Database issue), D142-148.

[165] Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J.E.; Gavin, O.L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E.L.; Eddy, S.R.; Bateman, A. The Pfam protein families database. *Nucleic Acids Res.,* **2010**, *38*(Database issue), D211-222.

[166] Jaroszewski, L.; Rychlewski, L.; Li, Z.; Li, W.; Godzik, A. FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res.,* **2005**, *33*, W284-288.

[167] Soding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.,* **2005**, *33*(Web Server issue), W244-248.

[168] Kelley, L.A.; Sternberg, M.J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.,* **2009**, *4*(3), 363-371.

[169] Karplus, K. SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Res.,* **2009**, *37*(Web Server issue), W492-497.

[170] Zhang, W.; Liu, S.; Zhou, Y. SP5: improving protein fold recognition by using torsion angle profiles and profile-based gap penalty model. *PLoS One,* **2008**, *3*(6), e2325.

[171] Jones, D.T. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.,* **1999**, *287*(4), 797-815.

[172] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* **1994**, *22*(22), 4673-4680.

[173] Katoh, K.; Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.,* **2008**, *9*(4), 286-298.

[174] Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.,* **2004**, *32*(5), 1792-1797.

[175] Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.,* **2000**, *302*(1), 205-217.

[176] Bates, P.A.; Kelley, L.A.; MacCallum, R.M.; Sternberg, M.J. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins,* **2001**, *Suppl 5*, 39-46.

[177] Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.,* **2010**, *5*(4), 725-738.

[178] Wu, S.; Zhang, Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.,* **2007**, *35*(10), 3375-3382.

[179] Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein. Sci.,* **2007**, *Chapter 2*, Unit 2 9.

[180] Eswar, N.; John, B.; Mirkovic, N.; Fiser, A.; Ilyin, V.A.; Pieper, U.; Stuart, A.C.; Marti-Renom, M.A.; Madhusudhan, M.S.; Yerkovich, B.; Sali, A. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.,* **2003**, *31*(13), 3375-3380.

[181] Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.,* **2004**, *32*(Web Server issue), W526-531.

[182] Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics,* **2006**, *22*(2), 195-201.

[183] Yang, Y.; Zhou, Y. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci.,* **2008**, *17*(7), 1212-1219.

[184] Tosatto, S.C. The victor/FRST function for model quality estimation. *J. Comput. Biol.,* **2005**, *12*(10), 1316-1327.

[185] Pugalenthi, G.; Shameer, K.; Srinivasan, N.; Sowdhamini, R. HARMONY: a server for the assessment of protein structures. *Nucleic Acids Res.,* **2006**, *34*, W231-234.

[186] Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Crys.,* **1993**, *28*, 283-291.

[187] Wallner, B.; Elofsson, A. Can correct protein models be identified? *Protein Sci.,* **2003**, *12*(5), 1073-1086.

[188] Benkert, P.; Kunzli, M.; Schwede, T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.,* **2009**, *37*(Web Server issue), W510-514.

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

[190] Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.,* **2008**, *36*(Database issue), D901-906.

[191] Gao, Z.; Li, H.; Zhang, H.; Liu, X.; Kang, L.; Luo, X.; Zhu, W.; Chen, K.; Wang, X.; Jiang, H. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics,* **2008**, *9*, 104.

[192] Li, Q.; Cheng, T.; Wang, Y.; Bryant, S.H. PubChem as a public resource for drug discovery. *Drug Discov. Today,* **2010**, *15*(23-24), 1052-1057.

[193] Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J. Chem. Inf. Model,* **2006**, *46*(2), 717-727.

[194] Wass, M.N.; Kelley, L.A.; Sternberg, M.J. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.,* **2010**, *38*, W469-473.

[195] Huang, B. MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS,* **2009**, *13*(4), 325-330.

[196] Kalidas, Y.; Chandra, N. PocketDepth: a new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.,* **2008**, *161*(1), 31-42.

[197] Morris, G.M.; Huey, R.; Olson, A.J. Using AutoDock for ligand-receptor docking. *Curr. Protoc. Bioinformatics,* **2008**, *Chapter 8*, Unit 8 14.

[198] Brylinski, M.; Skolnick, J. FINDSITE: a threading-based approach to ligand homology modeling. *PLoS Comput. Biol.,* **2009**, *5*(6), e1000405.

[199] Zhao, Y.; Sanner, M.F. FLIPDock: docking flexible ligands into flexible receptors. *Proteins,* **2007**, *68*(3), 726-737.

[200] Chang, D.T.; Oyang, Y.J.; Lin, J.H. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.,* **2005**, *33*, W233-238.

[201] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.,* **2005**, *33*, W363-367.

[202] Grosdidier, A.; Zoete, V.; Michielin, O. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins,* **2007**, *67*(4), 1010-1025.

[203] Meireles, L.M.; Domling, A.S.; Camacho, C.J. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res.,* **2010**, *38*(Web Server issue), W407-411.

[204] Douguet, D. e-LEA3D: a computational-aided drug design web server. *Nucleic Acids Res.,* **2010**, *38*, W615-621.

[205] Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.,* **2010**, *38 Suppl*, W609-614.

[206] Dunkel, M.; Gunther, S.; Ahmed, J.; Wittig, B.; Preissner, R. SuperPred: drug classification and target prediction. *Nucleic Acids Res.,* **2008**, *36*, W55-59.

Received: August 6, 2010 **Revised: January 23, 2011** **Accepted: March 1, 2011**