CellPress

Beware of docking!

Yu-Chian Chen[1,2,3]

[1] Human Genetic Center, Department of Medical Research, China Medical University Hospital, Taichung, Taiwan
[2] Research Center for Chinese Medicine and Acupuncture, China Medical University, Taichung 40402, Taiwan
[3] Department of Biomedical Informatics, Asia University, Taichung, 41354, Taiwan

Docking is now routine in virtual screening or lead optimization for drug screening and design. The number of papers related to docking has dramatically increased over the past decade. However, there are many issues to consider when undertaking a docking study. Frequent problems or issues arise, such as the wrong binding site of the target protein, screening using an unsuitable small-molecule database, the choice of docking pose, high dock score but failed in molecular dynamics (MD) simulation, and lack of clarity over whether the compound is an inhibitor or agonist. These problems should be cause for caution and concern before performing docking. Some papers show comprehensive biochemistry experiments but only a simple docking figure. This review presents some evidence to show that the docking might be questionable, despite a high score. In some cases, the accuracy of docking can even change from 0% to 92.66%. Thus, please beware of docking!

Docking for structure-based drug design

Since its beginnings in the 1960s, docking, along with the tremendous developments in physics, chemistry, informational technology, biochemistry, and computers, has become a powerful tool and an essential technique, not only in drug screening but also in protein–protein interactions and the behavior of nanomaterials. The current field of computer-aided drug design (CADD) is dominated by technologies used to dock small molecules into macromolecules, particularly protein targets, and its use is increasing year by year. In modern CADD, structure-based drug design is essential [1–4] and most big pharmaceutical companies have this department. Many commercial drugs are directly designed from CADD method [5]. Undoubtedly, docking techniques are very important scientific advances for understanding of chemical compounds, as noted particularly when three top computational scientists won the 2013 Nobel Prize in chemistry.

Protein–ligand or protein–protein docking is a computational technology to predict the orientation of a ligand when it is bound to a protein receptor or enzyme. In most cases, one can choose the best 'binding affinity' to be the potent ligand for further biochemistry experiments and development. Because docking is simple and the equipment requirement is low (it even works well on a personal computer), docking-related papers have sharply increased over the past decade (Figure 1). However, can we or should we trust the results of these docking studies? In this paper I provide a critical survey of the field, pointing out the strengths and weaknesses of the current family of docking protocols.

Careful evaluation shows that accuracy is a major problem with docking studies, because if the docking is not approached with precision then these papers will be of little value [6–8]. Questionable docking results can be found, even in high-profile journals. There are frequent problems such as an inaccurate binding site of the target protein, screening using an unsuitable small-molecule database, the choice of docking pose, high dock score (binding affinity) but failed in MD simulation, lack of clarity over whether the compound is an inhibitor or agonist, or the docking results are inconsistent with bioassays. The worst case is often found in some very high profile journals, which show an excellent bioassay but only with a simple docking figure. These problems in the interpretation of docking should be cause for caution and concern. Although some papers declare docking results with a high accuracy by comparing the ligand pose before and after docking, here I present some evidence that the docking might be still questionable. In some cases the accuracy of docking can even change from 0% to 92.66%.

Docking algorithms and programs

The original concept of docking comes from the concept of 'lock and key' of rational drug design, but the precise algorithms used to fit the 'key' (the ligand) into the 'lock' (the receptor protein) vary across programs. The latest developments in docking programs, the docking web server, screen software, and screen webserver are listed in Table 1, from which we can see that the number of new algorithms has been increasing in recent years. If we further analyze all the docking papers (Figure 2), we can see that the most commonly used docking programs are Autodock [9] and GOLD [10]. This does not mean that Autodock or GOLD are more accurate than other docking programs, they are merely more popular and well known. It is possible that their high citation rate is due to these programs being free and being created earlier than the other recent docking programs. Nowadays, a new algorithm to predict protein structure for docking, Rosetta (http://boinc.bakerlab.org/), has also been highly evaluated.

Although they vary, the different algorithms of each docking program must strike a balance between speed and accuracy. The algorithms for docking also vary by differences in scoring functions. Binding affinity is usually

Corresponding author: Chen, Y-C. (ycc929@MIT.edu).
Keywords: docking; computer-aided drug design (CADD); molecular dynamics (MD); ligand-based drug design; residence time; occupation time.

0165-6147/
 http://dx.doi.org/10.1016/j.tips.2014.12.001

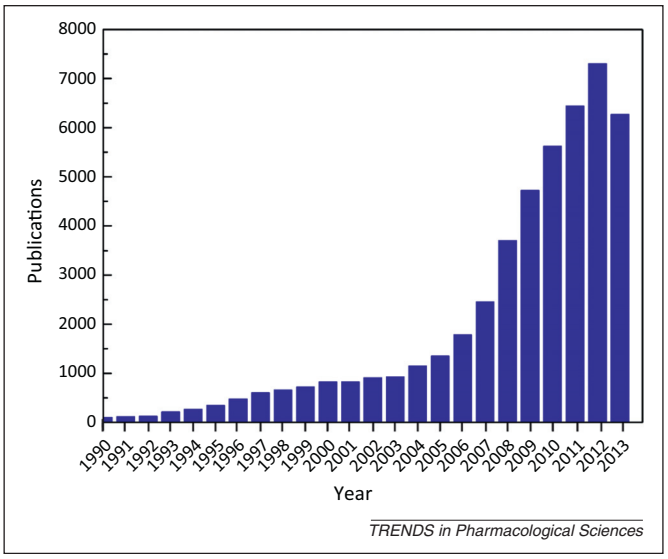

Figure 1. The increase in the number of papers, from 1990 to 2013, retrieved from the PubMed Central (PMC)-NCBI database (http://www.ncbi.nlm.nih.gov/pmc/). Keywords were 'docking' or 'dock' shown in the abstract or title.

considered to be a priority in the evaluation of the best candidate for virtual screening. There are several docking programs for a user to choose from based on his or her particular requirements. At present, docking algorithms emphasize different aspects of structure-based drug design (SBDD), such as fragment-based drug design [11–13], flexible docking [14], docking in water, solvation, and specific pH [15,16]. For example, if we need to screen more than 10 000 compounds from a database, then flexible docking maybe not a good choice unless we have a very powerful and high-speed computer. By contrast, if we need to dock only a few compounds in the specific protein binding site, at a specific pH, water, or solvation, then the flexible docking program might be a good choice. Choice of docking program therefore depends on what type of hardware you have and how large a database you are screening. For drug screening, the traditional Chinese medicine (TCM) database at Taiwan contains more than 61 000 compounds [17]; using the iScreen webserver for screening specific TCM and customized docking, multiple docking operations including standard, in-water, pH

Table 1. Recent software and webservers for docking and virtual screening as compiled and categorized by "http://www.Click2Drug.org" of the SIB Swiss Institute of Bioinformatics, which provides a comprehensive list of computer-aided drug design software and web services for structure-based and ligand-based calculations

Program name	Novel features	Refs
Docking Software		
Autodock	Free open-source EA-based docking software. Flexible ligand. Flexible protein side chains. Maintained by the Molecular Graphics Laboratory, Scripps Research Institute, La Jolla.	[82]
DOCK	Anchor-and-grow based docking program. Free for academic use. Flexible ligand. Flexible protein. Maintained by the Soichet group at the University of California San Francisco (UCSF).	[83]
GOLD	GA-based docking program. Flexible ligand. Partial flexibility for protein. Product of a collaboration between the University of Sheffield, GlaxoSmithKline, and the Cambridge Crystallographic Data Centre (CCDC).	[84]
Glide	Exhaustive search-based docking program. Exists in extra precision (XP), standard precision (SP) and virtual high-throughput screening modes. Ligand and protein flexible. Provided by Schrödinger.	[6]
SCIGRESS	Desktop/server molecular modeling software suite employing linear scaling semi-empirical quantum methods for protein optimization and ligand docking. Developed and distributed by Fujitsu.	[85]
GlamDock	Docking program based on a Monte-Carlo with minimization (basin-hopping) search in a hybrid interaction matching/internal coordinate search space. Part of the Chil^2 suite. Open for general research.	[185]
GEMDOCK (generic evolutionary method for molecular docking)	Program for computing a ligand conformation and orientation relative to the active site of target protein.	[87]
iGEMDOCK	Graphic environment for the docking, virtual screening, and post-screening analysis. Free for non-commercial researches. For Windows and Linux.	[87]
HomDock	Program for similarity-based docking, based on a combination of the ligand-based GMA molecular alignment tool and the docking tool GlamDock. Part of the Chil^2 suite. Open for general research.	[88]
ICM	Docking program based on pseudo-Brownian sampling and local minimization. Ligand and protein flexible. Provided by MolSoft.	[89]
FlexX, Flex-Ensemble (FlexE)	Incremental build-based docking program. Flexible ligand. Protein flexibility through ensemble of protein structure. Provided by BioSolveIT.	[86]
Fleksy	Program for flexible and induced fit docking using receptor ensemble (constructed using backbone-dependent rotamer library) to describe protein flexibility. Provided by the Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen.	[90]
FITTED (flexibility induced through targeted evolutionary description)	Suite of programs to dock flexible ligands into flexible proteins. This software relies on a genetic algorithm to account for flexibility of the two molecules and location of water molecules, and on a novel application of a switching function to retain or displace water molecules and to form potential covalent bonds (covalent docking) with the protein side chains. Part of the Molecular FORECASTER package and FITTED Suite. Free for an academic site license (excluding cluster).	[91]
VLifeDock	Multiple approaches for protein–ligand docking. Provides three docking approaches: grid-based docking, GA docking, and VLife's own GRIP docking program. Several scoring functions can be used: PLP score, XCscore, and Steric + Electrostatic score. Available for Linux and Windows. Provided by VLife (http://www.vlifesciences.com/)	
ParaDockS (parallel docking suite)	Free open-source program for docking small, drug-like molecules to a rigid receptor employing either the knowledge-based potential PMF04 or the empirical energy function p-Score.	[92]

Table 1 (*Continued*)

Program name	Novel features	Refs
Molegro Virtual Docker	Protein–ligand docking program with support for displaceable waters, induced-fit-docking, user-defined constraints, molecular alignment, ligand-based screening, and KNIME workflow integration. Provided by Molegro (http://www.scientificsoftware-solutions.com/product.php?productid=17625)	
eHiTS	Exhaustive search-based docking program. Provided by SimBioSys.	[93]
DAIM–SEED–FFLD	Free open-source fragment-based docking suite. The docking is realized in three steps. DAIM (decomposition and identification of molecules) decomposes the molecules into molecular fragments that are docked using SEED (program for docking libraries of fragments with solvation energy evaluation). Finally, the molecules are reconstructed '*in situ*' from the docked fragments using the FFLD program (program for fragment-based flexible ligand docking). Developed and maintained by the Computational Structural Biology of ETH, Zurich, Switzerland.	[94–96]
Autodock Vina	MC-based docking software. Free for academic use. Flexible ligand. Flexible protein side chains. Maintained by the Molecular Graphics Laboratory, The Scripps Research Institute, La Jolla.	[97]
VinaMPI	Massively parallel message-passing interface (MPI) program based on the multithreaded virtual docking program AutodockVina. Free and open-source. Provided by the University of Tennessee.	[98]
FlipDock	GA-based docking program using FlexTree data structures to represent a protein–ligand complex. Free for academic use. Flexible ligand. Flexible protein. Developed by the Department of Molecular Biology at the Scripps Research Institute, La Jolla.	[99]
FRED	FRED performs a systematic, exhaustive, nonstochastic examination of all possible poses within the protein active site, filters for shape complementarity and pharmacophore features before selecting and optimizing poses using the Chemgauss4 scoring function. Provided by OpenEye scientific software.	[100]
HYBRID	Docking program similar to FRED, except that it uses the chemical Gaussian overlay (CGO) ligand-based scoring function. Provided by OpenEye scientific software.	[44]
POSIT	Ligand-guided pose prediction. POSIT uses bound ligand information to improve pose prediction. Using a combination of several approaches, including structure generation, shape alignment, and flexible fitting, it produces a predicted pose whose accuracy depends on similarity measures to known ligand poses. As such, it produces a reliability estimate for each predicted pose. In addition, if provided with a selection of receptors from a crystallographic series, POSIT will automatically determine which receptor is best suited for pose prediction. Provided by OpenEye scientific software.	[100]
Rosetta Ligand	Monte Carlo minimization procedure in which the rigid body position and orientation of the small molecule and the protein side chain conformations are optimized simultaneously. Free for academic and non-profit users.	[101]
Surflex-Dock	Docking program based on an idealized active site ligand (a protomol), used as a target to generate putative poses of molecules or molecular fragments, which are scored using the Hammerhead scoring function. Distributed by Tripos.	[102]
CDocker	CHARMm-based docking program. Random ligand conformations are generated by MD and the positions of the ligands are optimized in the binding site using rigid body rotations followed by simulated annealing. Provided by Accelrys.	[103]
LigandFit	CHARMm-based docking program. Ligand conformations generated using Monte Carlo techniques are initially docked into an active site based on shape, followed by further CHARMm minimization. Provided by Accelrys.	[35]
MOE	Suite of medicinal chemistry tools including ligand–receptor docking, protein/ligand interaction diagrams, contact statistics, electrostatic and interaction maps, LigX (ligand optimization in pocket), ligand and structure-based scaffold replacement, multiple molecule flexible alignment, conformation generation, analysis, and clustering, active site detection and visualization, multi-fragment search, ligand- and structure-based query editor, high-throughput conformation generation, pharmacophore search. Distributed by Chemical Computing Group.	[104]
Lead Finder	Program for molecular docking, virtual screening, and quantitative evaluation of ligand binding and biological activity. Distributed by Moltech. For Windows and linux.	[105]
YASARA Structure	Adds support for small-molecule docking to YASARA View/Model/Dynamics using Autodock and Fleksy. Provided by YASARA (http://www.yasara.org/).	[106]
GalaxyDock	Protein–ligand docking program that allows flexibility of pre-selected side chains of ligand. Developed by the Computational Biology Lab, Department of Chemistry, Seoul National University.	[107,108]
MS-Dock	Free multiple-conformation generator and rigid docking protocol for multi-step virtual ligand screening.	[109]
FINDSITE-LHM	Homology modeling approach to flexible ligand docking. It uses a collection of common molecule substructures derived from evolutionarily related templates as the reference compounds in similarity-based ligand-binding pose prediction. It also provides a simple scoring function to rank the docked compounds. Freely available to all academic users and not-for-profit institutions. Provided by the Skolnick Research Group.	[110]
BetaDock	Molecular docking simulation software based on the theory of Beta-complex.	[111]
ADAM	Automated docking tool. Can be used for vHTS. Distributed by IMMD (http://www.immd.co.jp/en/product_2.html).	[112]
hint! (hydropathic interactions).	Estimates LogP for modeled molecules or data files; numerically and graphically evaluates binding of drugs or inhibitors into protein structures and scores DOCK orientations, constructs hydropathic	[113]

Table 1 (*Continued*)

Program name	Novel features	Refs
	(LOCK and KEY) complementarity maps that can be used to predict a substrate from a known receptor or protein structure or to propose the hydropathic structure from known agonists or antagonists, and evaluates/predicts the effects of site-directed mutagenesis on protein structure and stability.	
DockVision	Docking package including Monte Carlo, genetic algorithm, and database screening docking algorithms.	[114]
PLANTS (protein-ligand ANT system)	Docking algorithm based on a class of stochastic optimization algorithms termed ant colony optimization (ACO). In the case of protein–ligand docking, an artificial ant colony is employed to find a minimum energy conformation of the ligand in the binding site. These ants are used to mimic the behavior of real ants and mark low-energy ligand conformations with pheromone trails. The artificial pheromone trail information is modified in subsequent iterations to generate low-energy conformations with a higher probability. Developed by Konstanz University.	[115]
EADock	Hybrid evolutionary docking algorithm with two fitness functions, in combination with a sophisticated management of the diversity. EADock is interfaced with the CHARMM package for energy calculations and coordinate handling.	[116]
EUDOC	Program for the identification of drug interaction sites in macromolecules and drug leads from chemical databases.	[117]
FLOG	Rigid body docking program using databases of pre-generated conformations. Developed by the Merck Research Laboratories.	[118]
Hammerhead	Automatic, fast fragment-based docking procedure for flexible ligands, with an empirically tuned scoring function and an automatic method for identifying and characterizing the binding site on a protein.	[119]
ISE-Dock	Docking program based on the iterative stochastic elimination (ISE) algorithm.	[120]
ASEDock	Docking program based on a shape-similarity assessment between a concave portion (i.e., concavity) on a protein and the ligand. Developed by Yoka Systems.	[121]
HADDOCK (high ambiguity driven biomolecular docking)	An approach that makes use of biochemical and/or biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments, mutagenesis data or bioinformatic predictions. First developed from protein–protein docking, it can also be applied to protein–ligand docking. Developed and maintained by the Bijvoet Center for Biomolecular Research, The Netherlands.	[122]
Computer-aided drug-design platform using PyMOL	PyMOL plugins provide a graphical user interface incorporating individual academic packages designed for protein preparation (AMBER package and Reduce), molecular mechanics applications (AMBER package), and docking and scoring (AutoDock Vina and SLIDE).	[123]
Autodock Vina plugin for PyMOL	Allows defining binding sites and export to Autodock and VINA input files, performing receptor and ligand preparation automatically, starting docking runs with Autodock or VINA from within the plugin, viewing grid maps generated by autogrid in PyMOL, handling multiple ligands and setting up virtual screenings, and setting up docking runs with flexible side chains.	[124]
GriDock	Virtual screening front-end for AutoDock 4. GriDock was designed to perform the molecular dockings of a large number of ligands stored in a single database (SDF or Zip format) in the lowest possible time. It takes full advantage of all local and remote CPUs through the MPICH2 technology, balancing the computational load between processors/grid nodes. Provided by the Drug Design Laboratory of the University of Milano.	[125]
DockoMatic	GUI application that is intended to ease and automate the creation and management of AutoDock jobs for high-throughput screening of ligand/receptor interactions.	[126]
BDT	Graphic front-end application which controls the conditions of AutoGrid and AutoDock runs. Maintained by the Universitat Rovira i Virgili.	[127]
Docking Web services		
SwissDock	A web service to predict the molecular interactions that may occur between a target protein and a small molecule.	[128]
DockingServer	Offers a web-based, easy to use interface that handles all aspects of molecular docking from ligand and protein set-up.	[129]
1-Click Docking	Free online molecular docking solution. Solutions can be visualized online in 3D using the WebGL/Javascript-based molecule viewer of GLmol. Provided by Mcule (https://mcule.com/apps/1-click-docking/)	
Blaster	Public access service for structure-based ligand discovery. Uses DOCK as the docking program and various ZINC Database subsets as the database. Provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at UCSF.	[130]
Docking At UTMB	Web-driven interface for performing structure-based virtual screening with AutoDock Vina. Maintained by the Watowich lab at the University of Texas Medical Branch (http://docking.utmb.edu/).	
Pardock	All-atom energy-based Monte Carlo, rigid protein–ligand docking, implemented in a fully automated, parallel processing mode which predicts the binding mode of the ligand in receptor target site. Maintained by the Supercomputing Facility for Bioinformatics and Computational Biology, IIT Delhi.	[131]
FlexPepDock	High-resolution peptide docking (refinement) protocol, implemented within the Rosetta framework. The input for this server is a PDB file of a complex between a protein receptor and an estimated conformation for a peptide.	[132]
PatchDock	Web server for structure prediction of protein–protein and protein–small-molecule complexes based on shape complementarity principles.	[133]

Table 1 (*Continued*)

Program name	Novel features	Refs
BSP-SLIM	Web service for blind molecular docking on low-resolution protein structures. The method first identifies putative ligand-binding sites by structurally matching the target to the template holo-structures. The ligand–protein docking conformation is then constructed by local shape and chemical feature complementarities between ligand and the negative image of binding pockets. Provided by the University of Michigan.	[134]
BioDrugScreen	Computational drug design and discovery resource and server. The portal contains the DOPIN (Docked Proteome Interaction Network) database constituted by millions of pre-docked and pre-scored complexes from thousands of targets from the human proteome and thousands of drug-like small molecules from the NCI diversity set and other sources. The portal is also a server that can be used to (i) customize scoring functions and apply them to rank molecules and targets in DOPIN; (ii) dock against pre-processed targets of the PDB; and (iii) search for off-targets. Maintained by the laboratory of Samy Meroueh at the Center for Computational Biology and Bioinformatics at Indiana University School of Medicine.	[135]
GPCRautomodel	Web service that automates the homology modeling of mammalian olfactory receptors (ORs) based on the six 3D structures of G protein-coupled receptors (GPCRs) available so far, and (ii) performs the docking of odorants on these models, using the concept of colony energy to score the complexes. Provided by INRA.	[136]
kinDOCK	Allows comparative docking of ligands into the ATP-binding site of a protein kinase (target). A sequence alignment of the target and a protein kinase profile is performed using HMMER. It uses protein–protein superposition (automatically restricted to the ligand binding pocket) of the target 3D structure with those of known complexes of protein kinases/ligands.	[137]
iScreen	Web service for docking and screening the small-molecule database of traditional Chinese medicine (TCM) to a target protein. iScreen is also implemented with the *de novo* evolution function for the selected TCM compounds using the LEA3D genetic algorithm.	[14]
idTarget	Web server for identifying biomolecular targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. Maintained by the National Taiwan University.	[138]
Score	Allows calculation of different dock scores of a ligand–receptor complex that can be submitted as a whole file containing both interaction partners or as two separated files. The calculation phase is provided by VEGA. Provided by the Drug Design Laboratory of the University of Milano (http://159.149.85.2/score.htm).	
Pose & Rank	Web server for scoring protein–ligand complexes. Provided by the laboratory of Andrej Sali.	[139]
PLATINUM	Calculates hydrophobic properties of molecules and their match or mismatch in receptor–ligand complexes. These properties may help to analyze results of molecular docking.	[140]
LPCCSU	Analysis of interatomic contacts in ligand–protein complexes.	[141]
Others (docking)		
COPICAT (comprehensive predictor of interactions between chemical compounds and target proteins)	System for predicting interactions between chemical compounds and proteins by using Support Vector Machine (SVM). COPICAT realizes comprehensive prediction of protein–chemical interactions by utilizing very general, or the most easily available, data (i.e., amino acid sequences and chemical structures). Maintained by the Sakakibara Laboratory, Department of Biosciences and Informatics, Keio University.	[142]
Screening Software		
Pharmer	Free open-source pharmacophore search technology that can search millions of chemical structures in seconds.	[143]
Catalyst	Pharmacophore modeling and analysis; 3D database building and searching; ligand conformer generation and analysis tools; geometric, descriptor-based querying; shape-based screening. Distributed by Accelrys as part of Discovery Studio.	[144]
PharmaGist	Freely available web server for pharmacophore detection. The download version includes virtual screening capability.	[145]
LigandScout	Fully integrated platform for virtual screening based on 3D chemical feature pharmacophore models. Developed by inte:ligand.	[146]
CoLibri	Assembles huge compound collections from multiple sources and various input formats into a virtual screening library, removes duplicates, assesses the distribution of physicochemical properties of the compounds and makes selections/filters based on any property-threshold, molecule name-pattern, or presence/absence of a particular substructure motif. Generates a fragment library. Modifies molecules or fragments for generating, transforming and general handling of virtual screening libraries. Distributed by BioSolveIT (http://www.biosolveit.com/newsletter/archive/issue11.html).	
Corina	Generates 3D structures for small and medium-sized drug-like molecules. Distributed by Molecular Networks (http://www.molecular-networks.com/products/corina).	
DecoyFinder	Graphical tool which helps in finding sets of decoy molecules for a given group of active ligands. It does so by finding molecules which have a similar number of rotational bonds, hydrogen bond acceptors, hydrogen bond donors, logP value, and molecular weight, but are chemically different, defined by a maximum Tanimoto value threshold between active ligand and decoy molecule MACCS fingerprints. Optionally, a maximum Tanimoto value threshold can be set between decoys to assure chemical diversity in the decoy set.	[147]

Table 1 (*Continued*)

Program name	Novel features	Refs
DOVIS (docking-based virtual screening)	Tool for virtual screening of chemical databases containing up to millions of small, drug-like compounds. The designed docking-based virtual screening pipeline uses the AutoDock 4.0 program as its docking engine and is integrated into an HPC environment. Its purpose is to remove many technical and administrative complexities involved in employing AutoDock for large-scale virtual screening. Developed by the Biotechnology High Performance Computing Software Applications Institute.	[148]
SHAFTS (shape-feature similarity)	A program for 3D molecular similarity calculation and ligand-based virtual screening. SHAFTS adopts a hybrid similarity metric combined with molecular shape and colored (labeled) chemistry groups annotated by pharmacophore features for 3D similarity calculation and ranking, which is designed to integrated the strength of pharmacophore matching and volumetric overlay approaches. Provided by the East China University of Science and Technology.	[149]
PyRX	Virtual screening software for computational drug discovery that can be used to screen libraries of compounds against potential drug targets. PyRx includes docking wizard with easy-to-use user interface which makes it a valuable tool for CADD. PyRx also includes chemical spreadsheet-like functionality and a visualization engine that are essential for rational drug design. AutoDock 4 and AutoDock Vina are used as docking software. Free and open-source for Windows, Linux, and Mac OSX.	[150]
MOLA	Free software for virtual screening using AutoDock4/Vina in a computer cluster using non-dedicated multi-platform computers. MOLA is integrated on a customized Live-CD GNU/LINUX operating system and is distributed as a MOLA.iso file. Distributed by BioChemCore.	[151]
NNScore (neural network-based scoring function for the characterization of protein–ligand complexes)	Reads PDBQT files as input. Developed at UCSF.	[152]
WinDock	Program for structure-based drug discovery tasks under a uniform, user-friendly graphical interface for Windows-based PCs. Combines existing small-molecule searchable 3D libraries, homology modeling tools, and ligand–protein docking programs in a semi-automatic, interactive manner, which guides the user through the use of each integrated software component. Developed by the Howard University College of Medicine.	[153]
DockoMatic	GUI application that is intended to ease and automate the creation and management of AutoDock jobs for HTS of ligand/receptor interactions.	[154,155]
Screen Suite	Ligand-based screening tool using fingerprints, 2D and 3D descriptors. Distributed by ChemAxon (http://www.chemaxon.com/products/screen/).	
MolSign	Program for pharmacophore identification and modeling. Can be used for querying databases as a pharmacophore-based search. Provided by VLife (http://www.vlifesciences.com/)	
eHiTS_LASSO	Similarity searching tool that uses LASSO descriptors (interacting surface point types) to find molecules with diverse chemical scaffolds but similar surface properties. Distributed by SimBioSys.	[156]
Spectrophores	Converts 3D molecular property data (electrostatic potentials, molecular shape, lipophilicity, hardness and softness potentials) into one-dimensional spectra independent of the position and orientation of the molecule. It can be used to search for similar molecules and screen databases of small molecules. Open-source software developed by Silicos (http://openbabel.org/docs/dev/Fingerprints/spectrophore.html).	
Shape-it	Free open-source shape-based alignment tool by representing molecules as a set of atomic Gaussians. Open-source software developed by Silicos (http://silicos-it.com/software/software.html).	
Align-it (formerly Pharao)	Pharmacophore-based tool to align small molecules. The tool is based on the concept of modeling pharmacophoric features by Gaussian 3D volumes instead of the more common point or sphere representations. The smooth nature of these continuous functions has a beneficial effect for the optimization problem introduced during alignment. Open-source software developed by Silicos (http://silicos-it.com/software/software.html).	
Open3DALIGN	Command-line tool aimed at unsupervised molecular alignment. Alignments are computed in an atom-based fashion (by means of a novel algorithm inspired by the LAMDA algorithm of Richmond and coworkers), in a pharmacophore-based fashion using Pharao as the alignment engine, or finally using a combination of the latter two methods. Free open-source software for Windows, Linux, and Mac (http://open3dalign.sourceforge.net/?Description).	
Molegro virtual docker	The built-in Docking Template tool makes it possible to perform ligand-based screening by flexibly aligning a number of ligands (and determine a score for their similarity) and to perform hybrid docking by guiding the docking simulation by combining the template similarity score with a receptor-based docking scoring function. Distributed by Molegro.	[157]
GMA (graph-based molecular alignment)	Combined 2D/3D approach for the fast superposition of flexible chemical structures. Part of the Chil^2 suite. Open for general research.	[158]
Fuzzee	Allows the identification of functionally similar molecules, based upon functional and structural groups or fragments. Part of the Chil^2 suite. Open for general research (http://www.chil2.de/Fuzzee.html).	
REDUCE (formerly FILTER)	Tool to filter compounds from libraries using descriptors and functional groups. Part of the Molecular FORECASTER package, from Molecular Forecaster (http://fitted.ca).	

Table 1 (*Continued*)

Program name	Novel features	Refs
SELECT (selection and extraction of libraries employing clustering techniques)	Creates subset of libraries by diversity or similarity using clustering techniques. Part of the Molecular FORECASTER package (http://fitted.ca).	
AutoclickChem	Computer program capable of performing click-chemistry reactions *in silico*. AutoClickChem can be used to produce large combinatorial libraries of compounds for use in virtual screens. Because the compounds of these libraries are constructed according to the reactions of click chemistry, they can be easily synthesized for subsequent testing in biochemical assays. Exists as a web server. Distributed by the National Biomedical Computation Resource.	[159]
REACTOR (rapid enumeration by automated combinatorial tool and organic reactions)	Creates a library of molecules by combining fragment libraries from a defined reaction, or from a generic attachment point on the fragments. Part of the Molecular FORECASTER package (http://fitted.ca).	
QUASI	Generates pharmacophores and performs pharmacophore-based virtual screening. Distributed by De Novo Pharmaceuticals.	[160]
GASP (genetic algorithm similarity program)	Generates pharmacophores using a genetic algorithm. Distributed by Tripos.	[161]
Tuplets	Pharmacophore-based virtual screening. Distributed by Tripos.	[162]
KeyRecep	Estimates the characteristics of the binding site of the target protein by superposing multiple active compounds in 3D space so that the physicochemical properties of the compounds match maximally with each other. Can be used to estimate activities and vHTS. Distributed by IMMD (http://www.immd.co.jp/en/product_2.html).	
CATS (chemically advanced template search)	Topological pharmacophore descriptor for scaffold-hopping, screening compound selection, and focused library profiling. Developed by the Swiss Federal Institute of Technology Zurich (ETHZ).	[163]
LigPrep	2D to 3D structure conversions, including tautomeric, stereochemical, and ionization variations, as well as energy minimization and flexible filters to generate ligand libraries that are optimized for further computational analyses. Distributed by Schrodinger (http://www.schrodinger.com/).	[164]
Balloon	Free command-line program that creates 3D atomic coordinates from molecular connectivity via distance geometry and conformer ensembles using a multi-objective genetic algorithm. The input can be SMILES, SDF, or MOL2 format. Output is SDF or MOL2.	[165]
Epik	Enumerates ligand protonation states and tautomers in biological conditions. Distributed by Schrodinger.	[166]
Bluto	Performs energy minimization and energy analysis of protein or protein–ligand complexes by using force field, for structural optimization of docking models of multiple ligands onto a protein. Provides tabular reports of the energy analysis such as the interaction energy. Suitable for vHTS. Distributed by IMMD (http://www.immd.co.jp/en/product_2.html).	
VSDMIP (virtual screening data management on an integrated platform)	Comes with a PyMOL graphical user interface. Developed by the Centro de Biologia Molecular Severo Ochoa.	[167]
Web services		
Blaster	Public access service for structure-based ligand discovery. Uses DOCK as the docking program and various ZINC Database subsets as the database. Provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at UCSF.	[133]
AnchorQuery	Specialized pharmacophore search for targeting protein–protein interactions. Interactively searches more than 20 million readily synthesizable compounds all of which contain an analog of a specific amino acid. Provided by the University of Pittsburgh (http://anchorquery.ccbb.pitt.edu).	
GFscore	A rank-based consensus scoring function based on five scoring functions: FlexX Score, G_Score, D_Score, ChemScore, and PMF Score available in the TRIPOS Cscore module. The aim is to eliminate as many molecules as possible from proprietary in-house database after a virtual library screening (VLS) using TRIPOS FlexX for docking and the TRIPOS Cscore module for scoring. Developed and maintained by the Institute for Structural Biology and Microbiology, Marseille, France.	[168]
FINDSITECOMB	Web service for large-scale virtual ligand screening using a threading/structure-based FINDSITE-based approach. It offers the advantage that comparable results are obtained when predicted or experimental structures are used. The user can either provide a protein structure in PDB format or a protein sequence whose structure will first be predicted before its use in virtual ligand screening. Freely available to all academic users and not-for-profit institutions. Provided by the Skolnick Research Group.	[169]
FINDSITE-LHM	Homology modeling approach to flexible ligand docking. It uses a collection of common molecule substructures derived from evolutionarily related templates as the reference compounds in similarity-based ligand-binding pose prediction. It also provides a simple scoring function to rank the docked compounds. Freely available to all academic users and not-for-profit institutions. Provided by the Skolnick Research Group.	[110]
New Human GPCR modeling and virtual screening database	Web server to use the new human GPCR modeling and virtual screening database as well as a new function similarity detection algorithm to screen all human GPCRs against the ZINC8 non-redundant (TC<0.7) ligand set combined with ligands from the GLIDA database (a total of 88 949 compounds). Freely available to all academic users and not-for-profit institutions. Provided by the Skolnick Research Group.	[170]

Table 1 (*Continued*)

Program name	Novel features	Refs
Aggregator Advisor	Free web service to suggest molecules that aggregate or may aggregate under biochemical assay conditions. The approach is based on chemical similarity to known aggregators and physical properties. Provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at UCSF (http://advisor.bkslab.org/).	
e-LEA3D	Searches the FDA-approved drugs either by keyword or by substructure. Also builds combinatorial library of molecules.	[171]
Combinatorial library design	Web server providing a click chemistry engine to connect one or more reactants on a central core (scaffold) (http://chemoinfo.ipmc.cnrs.fr/eDESIGN/reagent.html).	[172]
eDesign	Web server providing a *de novo* drug design engine to create new molecules either from scratch (lead-hopping) or based on a user-defined scaffold on which R-groups have to be optimized. Alternatively, the same tool can be used to screen a library of molecules. The structure-based function is based on the program PLANTS. Maintained by the Institut de Pharmacologie Moleculaire et Cellulaire, France (http://chemoinfo.ipmc.cnrs.fr/eDESIGN/edesign.html).	
GFscore	Web server to discriminate true negatives from false negatives in a dataset of diverse chemical compounds using a consensus scoring in a non-linear neural network manner. The global scoring function is a combination of the five scoring functions found in the Cscore package from Tripos.	[168]
ZincPharmer	Free online interactive pharmacophore search software for screening the ZINC database. ZINCPharmer can import LigandScout and MOE pharmacophore definitions as well as perform structure-based pharmacophore elucidation.	[173]
SimDOCK	Allows rapid selection of ligands fitting the active site of the submitted protein by superposition of its 3D structure with those of known complexes of protein/ligands of the family (http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/simdock.php).	[137]
pep:MMs:MIMIC	Web-oriented tool that, given a peptide 3D structure, is able to automate a multiconformers 3D similarity search among 17 million of conformers calculated from 3.9 million commercially available chemicals collected in the MMsINC database.	[174]
wwLig-CSRre	Online tool to enrich a bank of small compounds with compounds similar to a query.	[175]
Superimpose	Superimpose is a framework for superposition to discover known drug-active compounds similar to a given molecule.	[176]
AURAmol	Web service taking a candidate 2D or 3D molecular shape and using it to search for similarly shaped molecules in large databases. Provided by the University of York (http://www.cs.york.ac.uk/auramol/index.html).	
Feature Trees	Web service to perform small-molecule similarity searches using the Feature Trees descriptor against a pre-defined set of large compound libraries (http://public.zbh.uni-hamburg.de/ftwi/resources//ftrees_help.html#overview).	
SiMMap	Web server statistically deriving site-moiety map with several anchors, based on the target structure and several docked compounds. Each anchor includes three elements: a binding pocket with conserved interacting residues, moiety composition of query compounds, and pocket–moiety interaction type (electrostatic, hydrogen bonding, or van der Waals). Compounds highly agreeing with anchors of the site–moiety map are expected to activate or inhibit the target protein.	[177]
ShaEP	Free program to superimpose two rigid 3D molecular structure models, based on shape and electrostatic potentials, and computes a similarity index for the overlay. It can be used for the virtual screening of libraries of chemical structures against a known active molecule, or as a preparative step for 3D QSAR methods.	[178]
PharmMapper	Freely accessible web-server designed to identify potential target candidates for a given probe small molecule (drugs, natural products, or other newly discovered compounds with binding targets unidentified) using a pharmacophore mapping approach.	[179]

environment, and flexible docking can be performed [15]. The iSMART services, including the TCM drug network, TCM–pathway–disease network, and ligand-based drug design, were brought online to seal the gaps between TCM studies and computational systems biology and network analysis [16], and some new algorithms were developed for drug design and screening [18–21]. In our experience, translating drug discovery from CADD to clinical trials is difficult. For example, some top potent hits screened by docking from the TCM database failed in the final bioactivity test [20,22–27].

However, if a hit comes first from a bioassay – for example, if it was screened from high-throughput screening (HTS) by 'automation-friendly' robotics, data processing, and control software *in vitro* – then it is easier to move to *in vivo*. Explaining the underlying biological mechanism is easier and more reasonable [23]. However, given the rise of docking studies, it is important to conduct them with care. The following considers some important considerations that developers should keep in mind when undertaking docking analyses.

Is the protein structure available, and reliable?

A great deal of preparatory work is needed before an attempt at docking can be made. This includes determining the stable 3D structure using energy minimization for ligand and protein. This preparatory work can be seen in the form of a flow chart in Figure 3. As shown, the target protein sequence must be available from the start, and then the protein structure must be available, either from homology modeling, *ab initio* methods, or downloaded from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do). If the protein and ligand are prepared well, then docking can be performed. A major problem at

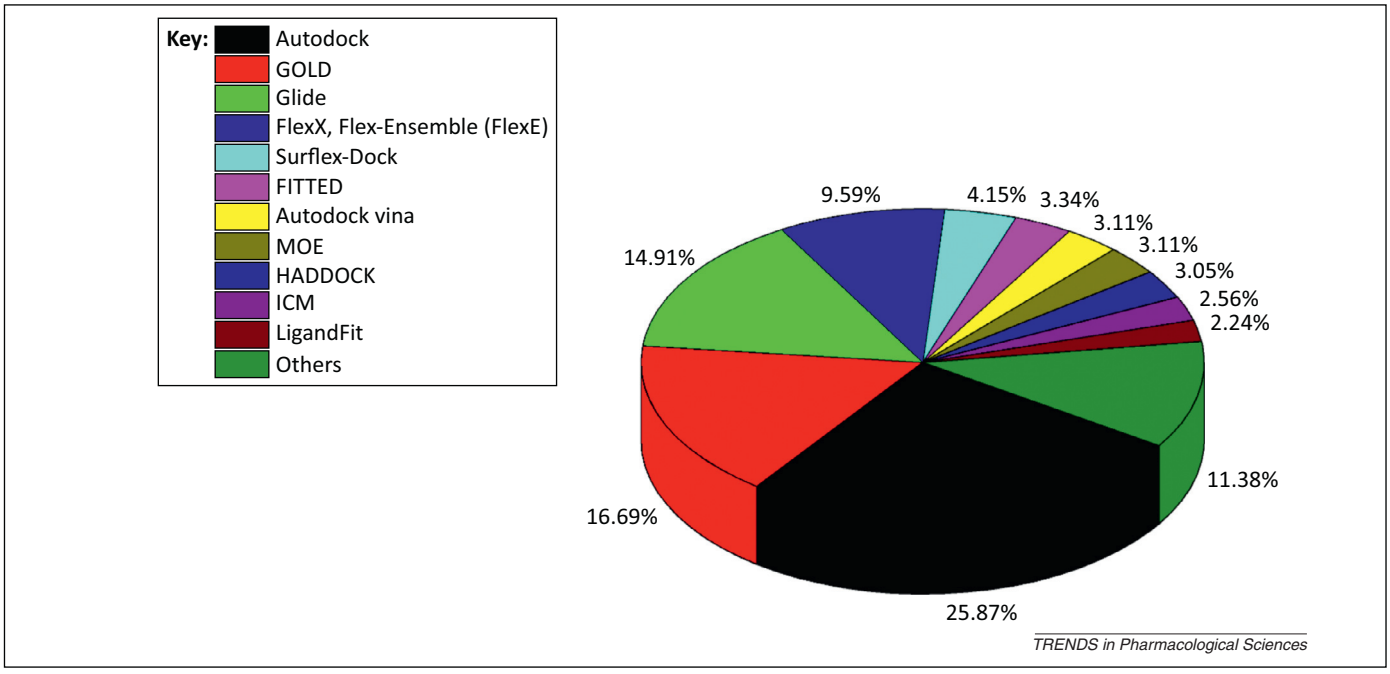

Figure 2. All docking publications from 1990 to 2013. Autodock, GOLD, and Glide are the most-used docking programs.

this stage is whether the protein structure is available and if the reported structure is reliable. If the protein structure is available for download from PDB, one must question the X-ray resolution and the conditions under which the protein crystal structure was obtained. Depending on the temperature regime employed, different structures can be obtained. Can we directly utilize these protein structures without further energy minimization or MD and clustering [28]? The accuracy of the protein structure obtained from MD simulation will be even influenced by the force-field accuracy [29]. If the protein structure is not available in PDB, then the difficulty will increase sharply. There are many methods for protein structure prediction. In addition, many different types of protein structure validation methods have been utilized to improve the accuracy of prediction, such as the famous Ramachandran plot [30].

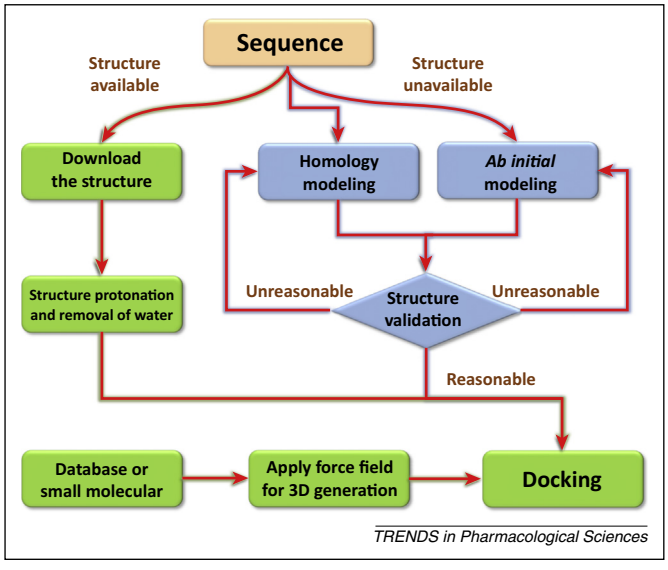

Figure 3. Flow chart of the docking procedure.

Flexible docking or rigid docking

Rigid docking refers to the protein and ligand being fixed so that bond angles or lengths are not changeable. Flexible docking, which allows for conformational shifts, is widely used today, although it requires much more time and computational throughput. Other docking methods include varying the solvation, varying the pH, and docking in the presence or absence of water [14]. A user may choose different docking modes based on the ability of their computational hardware and the character of the target protein. The docking mode can very roughly be dived into four models (Figure 4): lock and key, conformation isomerism, induced-fit, and conformation selection. From Figure 4, it seems that only the induced-fit mode will change the shape of the binding pocket. However, the entry into the binding pocket will sometimes change the shape of the binding site, according to simulation from MD experiments [16,19,20,23,31]. When type of shift occurs with ligand binding, rigid docking programs will not provide useful answers. This seems to suggest that there may be errors from docking programs that utilize a fixed protein structure.

How to choose the best docking scoring function?

There are many scoring functions, for example, LigScore1, LigScore2 [32], PLP1, PLP2 [33], PMF [34], Dock Score [35], Jain [36], Ludi [37], and binding energy [38]. How can we choose the proper scoring function to obtain the correct binding pose and possible ligand? This is a difficult question to answer. Theoretically, a lower Gibbs free energy of a complex indicates that the protein–ligand complex is more stable. However, different algorithms of these scoring functions are designed for different requirements on the part of the user. It is indeed hard to decide which scoring function

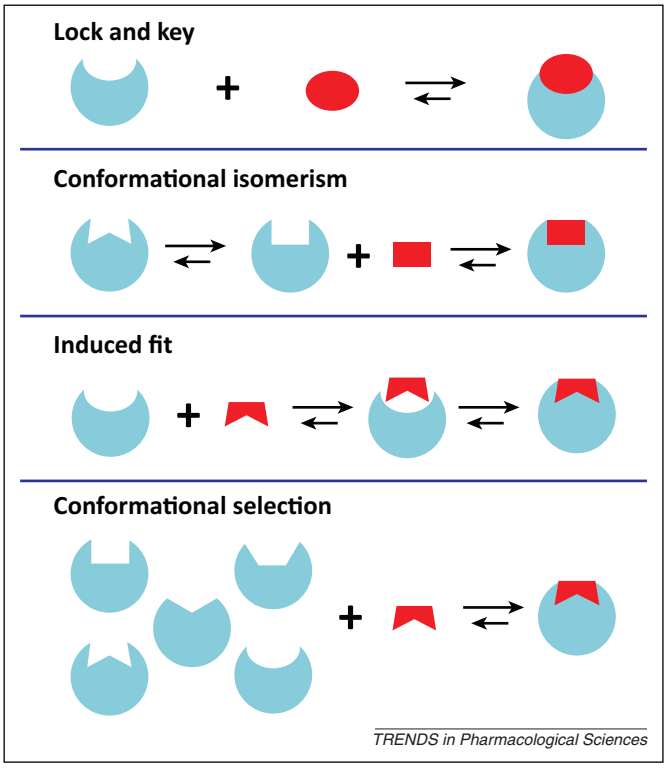

Figure 4. The docking mode can be roughly divided into four types of docking mode.

is suitable for different target proteins. For example, we have previously discussed the accuracy of each scoring function that results from using regression of IC_{50} with each scoring function [21]. We also propose the weighted-score function and show that its accuracy is better than the consensus score (Box 1). It is not a surprise that a weighted score is more accurate than the consensus score. Consensus score allows the user to pick up several or all the scoring functions to evaluate the overall score. Each score function is of equal weight in the consensus scoring function. By contrast, the weighted-scoring algorithm is based on their coefficients of regression. A higher coefficient of regression is given a higher weight in this algorithm. Thus, the weighted score is more accurate than the consensus score [39].

However, the weighted-scoring function must have enough information about the compound IC_{50} for regression to obtain a formula for predicting IC_{50} from the scoring function. It is more precise than the consensus score, but not quite so practical because in most screening situations it is difficult to obtain enough IC_{50} data. If a target protein has hundreds of different ligands with varying IC_{50}, then I will suggest using the weighted score. However, in our experience it is very hard to collect this much bioactivity data from one single paper. If bioactivity data are collected from different papers, they may vary due to differences in laboratory protocols. The consensus score avoids this problem because it does not require bioactivity data. To choose the consensus score or weighted score involves a trade-off for the user. Obviously, it is hard to obtain such a huge bioactivity data and then use it to employ a docking program. The optimal weight score function would be obtained by all the regression results of bioactivity versus dock score value, and this entails a huge amount of work. That is why we developed the weighted score as a compromise.

Box 1. The weighted-score function

Equation is known as a weighted equation in our previous paper [24].

$$WS = \sum_{i=1}^{n} W_i \times \frac{SF_i}{SF_c} = a \times pIC_{50} + b \quad [I]$$

where i is the number of scoring functions (12 in this study); $pIC_{50} = -\log(IC_{50})$; SF_c is the value of the control scoring function; SF_i is the value of the scoring function; W_i is the weight. SF_c in Equation I was employed to normalize the weighted equation [24].

Accuracy of docking?

Deciding what kind of docking program is the best for a specific target protein is problematic. Basically, if one can do the regression of bioactivity versus scoring function, then we can choose the best regression of scoring function as the evaluation criteria. To try to improve accuracy of screening and determine viable compounds for clinical trials, we have developed a website which integrates TCM and a computational systems biology tool (iSMART) [16]; including the docking webserver iScreen [15] to screen the TCM database [17]. The user can upload their protein structure downloaded from the PDB, and the webserver contains some useful modes such as customized docking, standard, aqueous, pH environment, and flexible docking. We have applied a re-docking protocol for proteins and compared the RMSDs between the results and the crystallized structures. The results show that the accuracy is higher than 70%, no matter whether we use the faster speed of 4× docking (speed fourfold normal docking speed) or 2× (twofold normal speed) docking [10]. Because the accuracy seems to be acceptable, why is it so difficult to find a potent lead for a commercial drug, given that most leads seem to have a very good binding affinity *in silico, in vitro*, or *in vivo*? First, the protein structure might be questionable if we merely download the protein–ligand complex from the PDB website and delete the ligand from the complex without further isolating the pure protein structure. Second, the environment of the binding site is a big issue. For example, the binding site may be located in the inner surface of the cell membrane, and therefore even if the ligand has a high dock score, it still must be transferred into the hydrophobic lipid bilayer to reach the binding domain. Third, the target protein maybe located at different pH within the context of the human body.

How do we know whether the highest dock score indicates that the ligand is an agonist or inhibitor?

Determining whether a good dock score indicates a good inhibitor or a good agonist is difficult. Many studies merely use a docking program, and then claim that they have found an agonist or inhibitor. But how do we know the real mechanism of the complex? The docking program only provides the calculated binding affinity. That is all, and nothing else. Some papers claim that a well-docked ligand indicates that it is a strong agonist or inhibitor [40–43]. Of course we will not know if it is an agonist or inhibitor until we perform further validation of bioassays. The only thing

Table 2. High predicted activities (pEC_{50}) with low scoring function for TCM compounds docking with peroxisome proliferator-activated receptor γ (PPAR-γ) protein [48]

Name	Dock score	MLR	SVM	BNT
Swerchirin	0.047	8.92	7.77	7.96
Homoferreirin	0.232	7.96	8.05	7.92
Alpinetin	0.340	7.35	7.30	7.09
Gentisein	0.443	7.47	6.86	7.16
Trichodermin	0.464	8.96	7.79	7.81
2-α-Hydroxyhelioscopinolide B	0.492	8.61	8.59	8.26
14-Deoxy-12-hydroxyandrographolide	0.741	8.07	8.66	8.67
Naphthoquinone VI	1.174	6.89	7.77	7.64
Medicagol	1.484	8.97	7.97	8.08
Isozedoarondiol	1.507	8.14	7.01	7.23
T2384[a]	77.618	7.52	7.06	8.50

[a]Control

that we sure of from the docking result is that the ligand binds well in the binding site. Therefore, it not advisable to overinterpret the docking results in a paper unless other validations have been performed.

Docking results are not consistent with ligand-based studies

Sometimes, from ligand-based investigation, the model is perfect with a good R square (indicating the model is reliable for prediction). We can then use this ligand-based model to predict and obtain the predicted activities of potent candidates. Unfortunately, according to the majority of our studies, most structure-based results do not seem to be consistent with ligand-based results. McGaughey and colleagues suggested that the difference might be because all virtual screening methods are database-dependent and can vary greatly for particular targets [44]. For example, in Table 2 the control compound (T2384) has a very high dock score (77.615) and thus has a very high predictive activity from the different ligand-based methods, including multiple linear regression (MLR) [45], support vector machine (SVM) [46], and Bayesian network toolbox (BNT) [47]. However, the other compounds also have a very high predicted activity but a very low dock score [48]. From Table 2, the docking results show that the control compound (D71904) with the lowest binding energy has a low predicted activity from the ligand-based prediction (MLR and SVM). By contrast, other compounds have a very low dock binding affinity (high binding energy), but obtain a very high predicted activity from the ligand-based prediction (MLR and SVM). In Table 3, the poor dock score also shows a high predicted activity from MLR, SVM, and BNT. However, ironically, the control compound (SAHA) has a very high dock score but a very poor ligand-based prediction by MLR and BNT. From Table 4, the predicted activity from ligand-based studies (MLR, SVM, and BNT) is very high, thus indicating that these compounds should be potent. However, they also show a very poor dock score. Table 5 also shows the high predicted activities (pIC_{50}) but with a very low dock score (Box 2).

The aforementioned issues are not rare in the process of CADD, but they seldom show up in published papers.

Table 3. High predicted activities (pIC50) with high binding energies for top TCM compounds docking with cAMP-specific 3′,5′-cyclic phosphodiesterase 4D (PDE4D) protein [180]

Name	Binding energy	MLR	SVM
D71904[a]	−121.826	6.45	5.43
Dracorhodin perchlorate	−80.976	7.56	6.53
Berberine	−50.722	7.65	6.88
Groenlandicine	−48.981	7.78	7.07
Berberrubine	−47.007	7.60	6.82
α-Asarone	−45.687	8.64	6.35
Coptisine	−38.006	7.34	6.24

[a] Control

Table 4. The high predicted activities (pIC_{50}) with low scoring function for top TCM compounds docking with histone deacetylase 2 (HDAC2) protein [181]

Name	Dock score	MLR	SVM	BNT
Irisflorentin	0.132	8.79	7.15	6.82
Methyl diacetoxy-[6]-gingerdiol	0.366	7.56	7.15	6.41
Diffractaic acid	0.475	8.83	7.15	6.48
Neogitogenin	0.537	8.73	7.33	6.93
Emetine	0.548	6.93	7.15	8.04
Bassianin	0.595	6.86	7.15	6.68
15-α-Hydroxysoladulcidine	0.783	7.23	7.15	7.01
Verticine	0.973	6.98	7.15	7.25
Rubschisantherin	1.186	8.90	7.99	7.53
Spiropachysine	1.267	6.57	7.15	7.82
SAHA[a]	45.997	4.65	7.15	4.50

[a]Control

Table 5. High predicted activities (pIC_{50}) with low scoring function for top TCM compounds docking with HIV integrase (IN) protein [182]

Name	Dock score	SVM	MLR
5-Hydroxy-7,8,2′,5′-tetramethoxyflavone	2.514	6.60	6.72
3′-O-angeloylhamaudol	2.617	6.37	6.87
2,4-Dihydroxyacetophenone	2.919	7.58	6.60
3′-O-angeloylhamaudol	3.137	6.37	6.23
Medicarpin	3.971	6.41	7.33
1,2,3,7-Tetramethoxyanthone	5.790	6.49	7.41
(−)-Maackiain	6.433	6.50	8.60
2,4-Heptanedione	7.140	7.77	6.90
Dimethoxytoluenel	8.202	7.32	7.14
Nerol oxide	8.406	6.41	6.50

Box 2. Multiple linear regression (MLR), support vector machine (SVM), and Bayesian network toolbox (BNT) algorithms

The MLR equation is shown in Equation II

$$Y = b_0 + b_1 \times X_1 + b_2 \times X_2 + \ldots + b_7 \times X_7 \quad [II]$$

The regression of SVM is shown in Equation III

$$r^2 = \frac{\left(I\sum_{i=1}^{I} f\left(x_i\right) y_i - \sum_{i=1}^{I} f\left(x_i\right) \sum_{i=1}^{I} y_i\right)^2}{\left(I\sum_{i=1}^{I} f\left(x_i\right)^2 - \left(\sum_{i=1}^{I} f\left(x_i\right)\right)^2\right)\left(I\sum_{i=1}^{I} y_i^2 - \left(\sum_{i=1}^{I} y_i\right)^2\right)} \quad [III]$$

The definition of Bayesian network is shown in Equation IV

$$P_B(x_1, \ldots, x_n) = \Pi_{i=1}^{n} P_B((x_i)|\Pi x_i) = \Pi_{i=1}^{n} \theta_{x_i|\Pi x_i} \quad [IV]$$

Authors put the best data, showing a high dock score, in their papers. However, the best lead may not be found from the highest dock score. From our past studies, the top 10 most-potent ligands obtained from such screening usually fail in the real bioassay test. However, the top 100 ligands often provide the potent hits *in vitro*. This shows that, although the highest dock scores might not provide the best leads, overall the prediction of docking seems not so bad. These optimized leads always show significance in bioactivity experiments. Of course, even with a potent lead that has been computationally optimized and validated with bioassays, I admit that it is still a long road to push this lead to a commercial drug. However, this protocol is based on theoretical computation algorithms and is definitely better than random biochemistry experimental tests.

Docking results are not consistent with MD simulation?

Most of the docking papers retrieved from PubMed or Web of Science database lack the further validation of MD simulation. Docking is a tool that can calculate the binding affinity of a protein–protein or protein–ligand complex. It often takes only 1–10 s to analyze a 300 kDa protein and 50–300 Da ligand complex. In other words, docking is an ideal tool for virtual screening the hits from the ligand database. MD simulation is also a technique for the calculation of 'movement' of the complex. The biggest different between docking and MD is the variable, time. Docking only considers the binding energy or binding affinity. By contrast, the MD emphasizes the complex shape that changes with time, although this increased complexity reduces the calculation capacity and increases the time per screening as compared to docking experiments. In Figure 5, the ligand (Angeliferula) is docked well into the binding site of protein phosphatase 2A (PP2A), and with the highest dock score. However, there is a huge difference between the pose suggested by docking and the final ligand pose after a MD simulation of 20 ns. The same problem can be seen in Figures 6–8. Further details of are shown in the videos in the supplementary material online. MD simulations in Videos S1–S5 also illustrate the

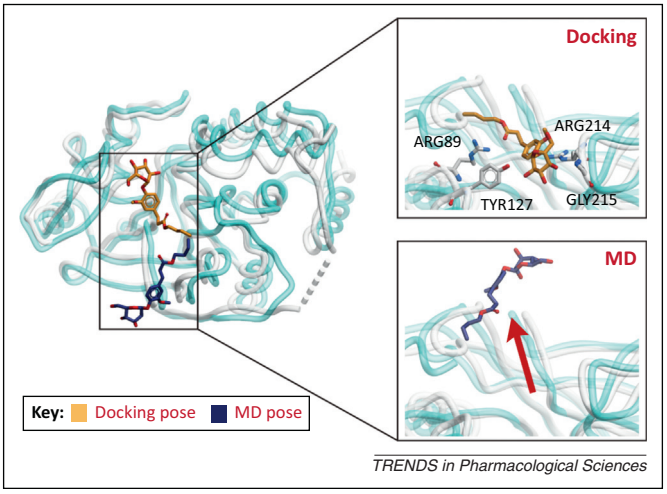

Figure 6. The ligand, dichotomoside E, is docked well into the binding site of protein phosphatase 2A (PP2A), and has the highest dock score. There is also a clear difference between the docking pose and the final ligand pose after a molecular dynamics (MD) simulation of 20 ns [183].

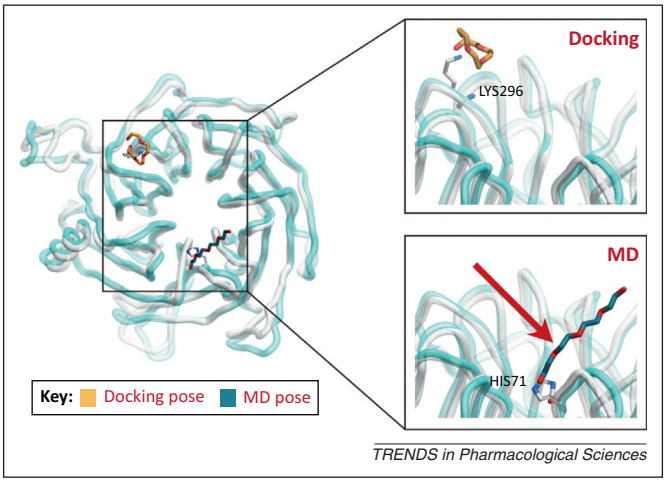

Figure 7. The ligand, FOG1, is docked well into the binding site of retinoblastoma-associated protein (RbAp48), and with the highest dock score. Again, there is a clear difference between the docking pose and the final ligand pose after a molecular dynamics (MD) simulation of 20 ns [184].

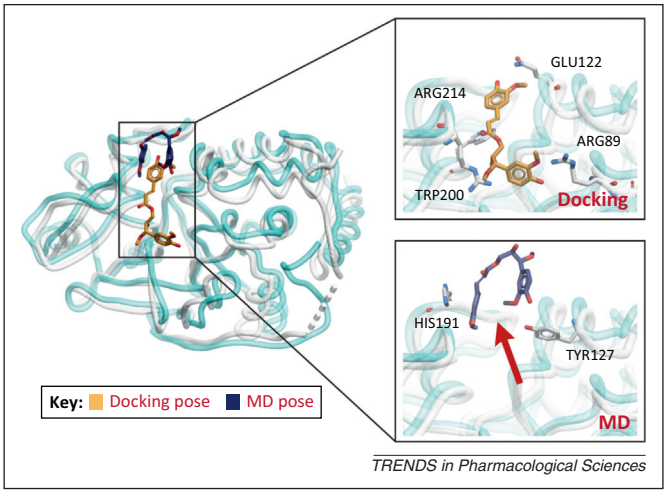

Figure 5. The ligand, angeliferula, is docked well into the binding site of protein phosphatase 2A (PP2A), and with the highest dock score. There is a clear difference between the docking pose and the final ligand pose after a molecular dynamics (MD) simulation of 20 ns [183].

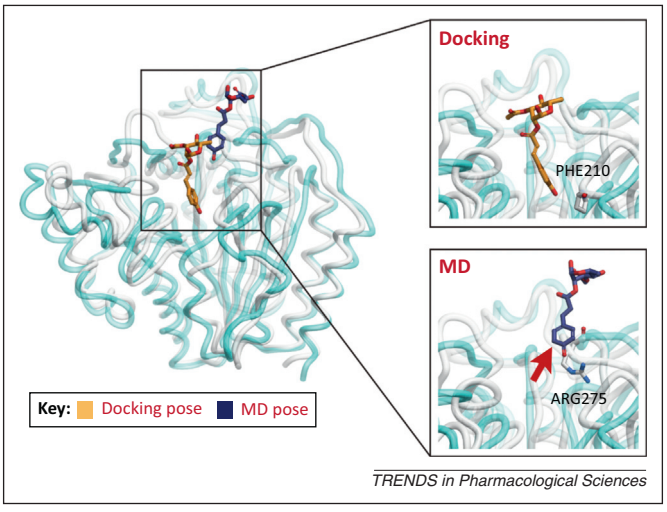

Figure 8. The ligand, ningposides C, is docked well into the binding site of histone deacetylase 2 (HDAC2), and with the highest dock score. As before, there is a major difference between docking pose and the final ligand pose after a molecular dynamics (MD) simulation of 20 ns [181].

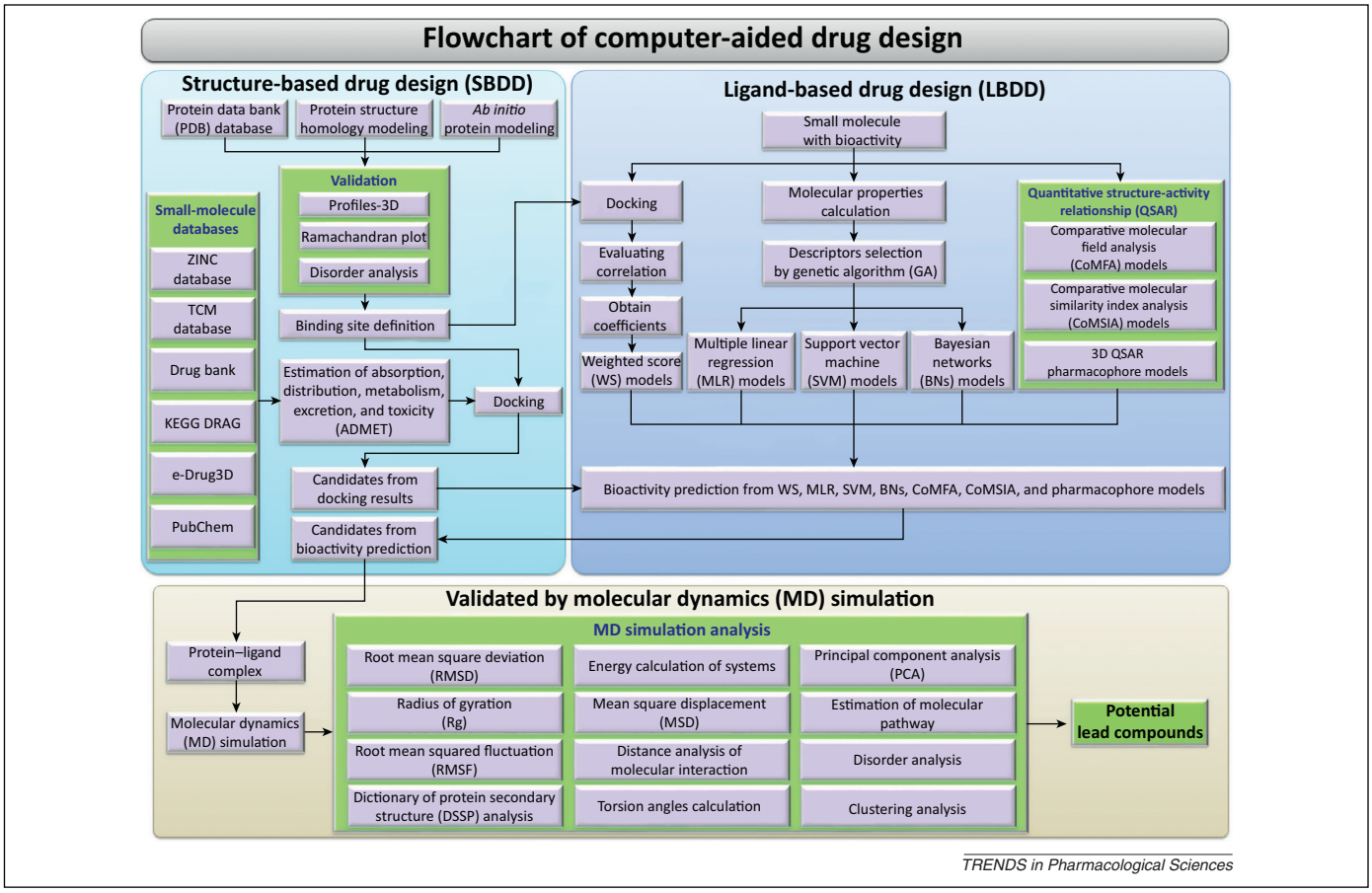

Figure 9. Flowchart of computer-aided drug design (CADD). The protocol is divided into three parts: structure-based drug design, ligand-based drug design, and validation by molecular dynamics simulation.

process in which the ligand may 'fly' away from the binding pocket, demonstrating that MD simulation is probably an essential validation before we can draw any conclusions from the docking results. However, the simulation time of MD is usually not very long (typically less than 1 ms because of hardware limitations). Hence, MD is definitely not definitive or essential in CADD, but can provide a further validation of docking results.

Concluding remarks

More and more studies now rely on advanced computational tools, and the issue of questionable docking results is therefore becoming increasingly important. Many expert biochemists publish significant results in high-profile journals, but with only a simple figure to validate the key residue binding with the ligand by docking [49–84]. Indubitably these papers are all very significant; however, they all lack any validation of MD studies. I am not saying these types of studies must always be accompanied by validation by MD simulation. However, experience has shown that docking analyses alone can create an inaccurate picture of ligand binding. Indeed, sometimes the ligand will fly away during the MD simulation, despite a high dock score. Hence, I strongly suggest scientists should draw their conclusions carefully, especially if one only has the docking or virtual screening results and no further validation. We have proposed a novel integrated framework and improved methodology of CADD previous paper [18]. A modified protocol is proposed in Figure 9 and Box 3.

Box 3. Integrated framework and improved methodology of computer-aided drug design (CADD)

The complete flowchart of the CADD protocol includes structure-based drug design (SBDD), ligand-based drug design (LBDD), lead optimization, and finally validation by molecular dynamics (MD) simulation (Figure 9). For SBDD, the protein structure may be available from protein databank (http://www.rcsb.org/pdb/home/home.do) or modeled by homology modeling or *ab initio* methods. The predicted protein structure should be validated by the software of profiles-3D, Ramachandran plot, and disorder protein analysis. If the structure of protein and the binding site meet the necessary standards, we can then perform docking. For screening, a suitable small-molecule database should be available. Researchers can also screen peptide databases (we have constructed our own small-peptide database for screening).

The SBDD usually combines with LBDD to evaluate the lead. The methodology of LBDD is to build the model based on bioactivity or molecular descriptors, and then predict the bioactivity of the potent lead screened by SBDD. For example, a weight score (WS) model is built base on the correlation and regression of bioactivity and dock score. Multiple linear regression (MLR), support vector machine (SVM), and Bayesian networks (BNs) models are built base on the descriptors of ligands. The best-known ligand-based model, QSAR, is based on the bioactivity and the pharmacophore features. Even though these models are very different, they all have the same goal: to predict the bioactivity without the target protein structure.

Finally, the complex of optimal lead docking in the protein binding site employs MD simulation for further validation. Many validation and analysis methods based on the results of MD are available to predict the entry and exit of ligand in the complex.

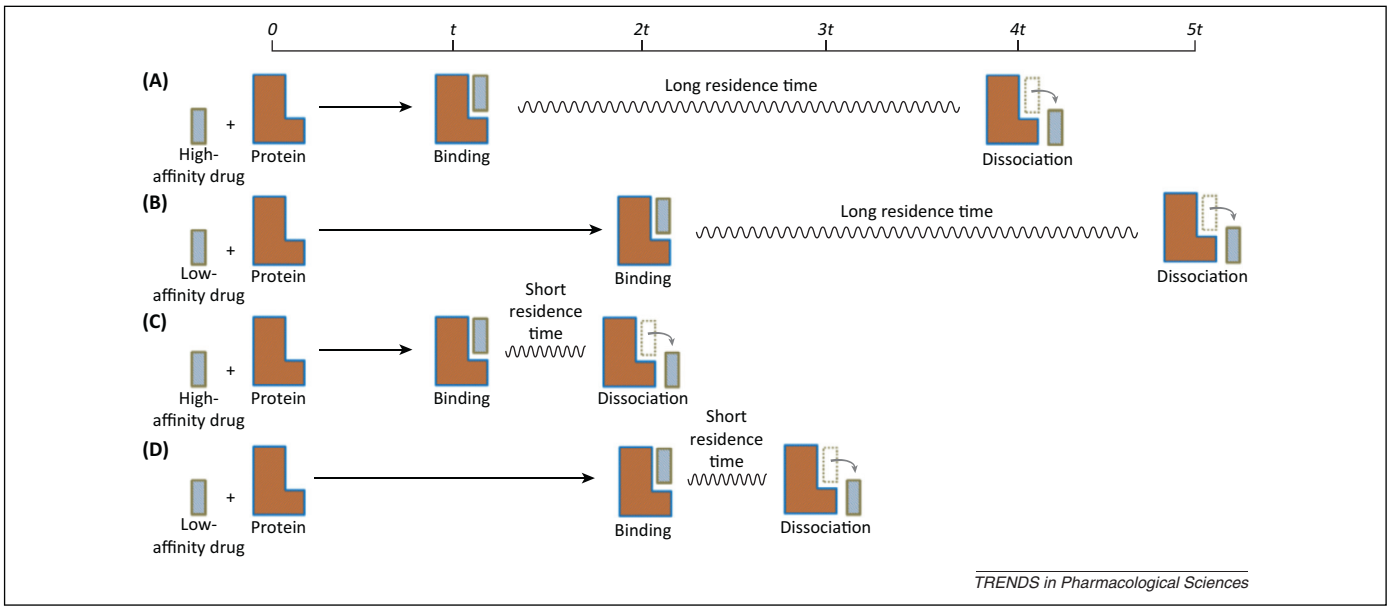

Figure 10. The possible four modes for drug design *in silico*. Case **(A)** shows a ligand with rapid docking in the binding site and also a long residence time in molecular dynamics (MD) simulation. Thus, it might exhibit superiority as a potent lead. In case **(B)** the ligand has a lower dock score but a long residence time. In **(C)** it shows rapid binding with a high dock score but with a short residence time. The total occupation time is short; however, this might not be disadvantageous because the ligand can bind again quickly. Theoretically, case **(D)** is not a good candidate because it shows a lower docking speed and short residence time.

In conclusion, the drug design principle can be divided into four modes (Figure 10), (A) fast binding and long residence time ($1/K_{off}$), (B) slow binding but with long residence time, (C) fast binding but with short residence time, and (D) slow binding and short residence time. Because of hardware and software limitations, it is currently impossible to simulate 4 h or 8 h of MD. In the future, progress in either hardware or software might allow residence time to be simulated by MD. However, for now, MD cannot model residence time because the simulation time is too short (only a few milli- or microseconds). In this paper, the fly-away ligand simulated by MD is only one piece of evidence to show that a ligand with the highest dock score does not mean that the ligand is a potent lead. That is the frequent mistake in many of the published papers. Theoretically, case (A) might be the best choice in drug design. However, in our experience, we have found that more than 300 commercial drugs and control compounds are located in cases (B) and (C). The debate continues.

Acknowledgments

The research was supported by grants from the National Science Council of Taiwan (NSC102-2325-B039-001 and NSC102-2221-E-468-027) and China Medical University Hospital (DMR-103-096). This study is also supported in part by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH102-TD-B-111-004), Taiwan Department of Health Cancer Research Center of Excellence (MOHW103-TD-B-111-03), and CMU under the Aim for Top University Plan of the Ministry of Education, Taiwan. Tim Williams is gratefully acknowledged for help in language editing.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tips.2014.12.001.

References

1 Warren, G.L. *et al.* (2012) Essential considerations for using protein-ligand structures in drug discovery. *Drug Discov. Today* 17 (23–24), 1270–1281
2 Merz, K.M., Jr *et al.* (2010) *Drug Design: Structure- and Ligand-Based Approaches*, Cambridge University Press
3 Zheng, M. *et al.* (2013) Computational methods for drug design and discovery: focus on China. *Trends Pharmacol. Sci.* 34, 549–559
4 Pei, J. *et al.* (2014) Systems biology brings new dimensions for structure-based drug design. *J. Am. Chem. Soc.* 136, 11556–11565
5 Sliwoski, G. *et al.* (2014) Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395
6 Friesner, R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749
7 Hou, X. *et al.* (2013) How to improve docking accuracy of AutoDock4.2: a case study using different electrostatic potentials. *J. Chem. Inf. Model.* 53, 188–200
8 Lee, M.R. and Sun, Y.X. (2007) Improving docking accuracy through molecular mechanics generalized born optimization and scoring. *J. Chem. Theory Comput.* 3, 1106–1119
9 Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins Struct. Funct. Bioinform.* 8, 195–202
10 Jones, G. *et al.* (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 245, 43–53
11 Chen, Y. and Shoichet, B.K. (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* 5, 358–364
12 Schmidt, T. *et al.* (2013) Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today* 19, 890–897
13 Perot, S. *et al.* (2010) Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* 15, 656–667
14 Rosenfeld, R. *et al.* (1995) Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.* 24, 677–700
15 Tsai, T.Y. *et al.* (2011) iScreen: world's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *J. Comput. Aided Mol. Des.* 25, 525–531
16 Chang, K.W. *et al.* (2011) iSMART: an integrated cloud computing web server for traditional Chinese medicine for online virtual screening, de novo evolution and drug design. *J. Biomol. Struct. Dyn.* 29, 243–250
17 Chen, C.Y.C. (2011) TCM database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* 6, e15939

18 Chen, C.Y.C. (2013) A novel integrated framework and improved methodology of computer-aided drug design. *Curr. Top. Med. Chem.* 13, 965–988
19 Chen, C.Y.C. and Tou, W.L. (2013) How to design a drug for the disordered proteins. *Drug Discov. Today* 18, 910–915
20 Tou, W.I. and Chen, C.Y. (2014) May disordered protein cause serious drug side effect? *Drug Discov. Today* 19, 347–372
21 Chen, C.Y. (2009) Weighted equation and rules – a novel concept for evaluating protein-ligand interaction. *J. Biomol. Struct. Dyn.* 27, 271–282
22 Li, M. *et al.* (2014) Developing hypothetical inhibition mechanism of novel urea transporter B inhibitor. *Sci. Rep.* 4, 5775
23 Tou, W.I. *et al.* (2013) Drug design for neuropathic pain regulation from traditional Chinese medicine. *Sci. Rep.* 3, 844
24 Yang, S.C. *et al.* (2011) Identification of potent EGFR inhibitors from TCM Database@Taiwan. *PLoS Comput. Biol.* 7, e1002189
25 Chang, S.S. *et al.* (2011) Two birds with one stone? Possible dual-targeting H1N1 inhibitors from traditional Chinese medicine. *PLoS Comput. Biol.* 7, e1002315
26 Tsou, Y.A. *et al.* (2012) Uroporphyrinogen decarboxylase as a potential target for specific components of traditional Chinese medicine: a virtual screening and molecular dynamics study. *PLoS ONE* 7, e50087
27 Chen, K.Y. *et al.* (2012) In silico identification of potent pancreatic triacylglycerol lipase inhibitors from traditional Chinese medicine. *PLoS ONE* 7, e43932
28 Zhang, J. *et al.* (2011) Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure* 19, 1784–1795
29 Raval, A. *et al.* (2012) Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins* 80, 2071–2079
30 Ramachandran, G.N. *et al.* (1963) Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99
31 Chen, H-Y. *et al.* (2014) Discovery of novel insomnia leads from screening traditional Chinese medicine database. *J. Biomol. Struct. Dyn.* 32, 776–791
32 Krammer, A. *et al.* (2005) LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.* 23, 395–407
33 Gehlhaar, D.K. *et al.* (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* 2, 317–324
34 Muegge, I. and Martin, Y.C. (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* 42, 791–804
35 Venkatachalam, C.M. *et al.* (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* 21, 289–307
36 Jain, A. (1996) Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided Mol. Des.* 10, 427–440
37 Böhm, H-J. (1998) Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* 12, 309
38 Tirado-Rives, J. and Jorgensen, W.L. (2006) Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J. Med. Chem.* 49, 5880–5884
39 Feher, M. (2006) Consensus scoring for protein-ligand interactions. *Drug Discov. Today* 11, 421–428
40 Alves, C.N. *et al.* (2007) Calculation of binding energy using BLYP/MM for the HIV-1 integrase complexed with the S-1360 and two analogues. *Bioorganic Med. Chem.* 15, 3818–4382
41 Chen, P.Y. *et al.* (2013) Old drug new use: searching for mitogen-activated protein kinase kinase 1 (MEK1) inhibitor by the computer aided drug design. *J. Life Sci.* 7, 453–458
42 Chen, P.Y. *et al.* (2012) Computational analysis of novel drugs designed for use as acetylcholinesterase inhibitors and histamine H3 receptor antagonists for Alzheimer's disease by docking, scoring and de novo evolution. *Mol. Med. Rep.* 5, 1043–1048
43 Hong, H.J. *et al.* (2012) Computational pharmaceutical analysis of anti-Alzheimer's Chinese medicine *Coptidis Rhizoma* alkaloids. *Mol. Med. Rep.* 5, 142–1474
44 McGaughey, G.B. *et al.* (2007) Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* 47, 1504–1519
45 Chatterjee, S. and Hadi, A.S. (1986) Influential observations, high leverage points, and outliers in linear regression. *Stat. Sci.* 1, 379–393
46 Chang, C-C. and Lin, C-J. (2011) LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27
47 Friedman, N. *et al.* (1997) Bayesian network classifiers. *Mach. Learn.* 29, 131–163
48 Chen, K-C. and Chen, C.Y-C. (2014) In Silico identification of potent PPAR-agonists from traditional Chinese medicine: a bioactivity prediction, virtual screening, and molecular dynamics study. *Evid. Based Complement. Altern. Med.* 2014, 192452
49 Zhu, J. *et al.* (2013) A novel small molecule RAD51 inactivator overcomes imatinib-resistance in chronic myeloid leukaemia. *EMBO Mol. Med.* 5, 353–365
50 Shie, J.J. *et al.* (2011) A practical synthesis of zanamivir phosphonate congeners with potent anti-influenza activity. *J. Am. Chem. Soc.* 133, 17959–17965
51 Wen, W.H. *et al.* (2010) Analogs of zanamivir with modified C4-substituents as the inhibitors against the group-1 neuraminidases of influenza viruses. *Bioorganic Med. Chem.* 18, 4074–4084
52 Ho, C.W. *et al.* (2010) Development of GlcNAc-inspired iminocyclitiols as potent and selective N-acetyl-beta-hexosaminidase inhibitors. *ACS Chem. Biol.* 5, 489–497
53 Wang, S.Y. *et al.* (2009) Bioactivity-guided screening identifies pheophytin a as a potent anti-hepatitis C virus compound from *Lonicera hypoglauca* Miq. *Biochem. Biophys. Res. Commun.* 385, 230–235
54 Hsu, H.J. *et al.* (2008) Factor Xa active site substrate specificity with substrate phage display and computational molecular modeling. *J. Biol. Chem.* 283, 12343–12353
55 Shao, Y.M. *et al.* (2008) Design, synthesis, and evaluation of trifluoromethyl ketones as inhibitors of SARS-CoV 3CL protease. *Bioorganic Med. Chem.* 16, 4652–4660
56 Shie, J.J. *et al.* (2007) Synthesis of tamiflu and its phosphonate congeners possessing potent anti-influenza activity. *J. Am. Chem. Soc.* 129, 11892–11893
57 Shao, Y.M. *et al.* (2007) Structure-based design and synthesis of highly potent SARS-CoV 3CL protease inhibitors. *Chembiochem* 8, 1654–1657
58 Lu, I.L. *et al.* (2006) Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* 49, 5154–5161
59 Tsai, K.C. *et al.* (2006) Discovery of a novel family of SARS-CoV protease inhibitors by virtual screening and 3D-QSAR studies. *J. Med. Chem.* 49, 3485–3495
60 Parai, M.K. *et al.* (2012) Design, synthesis, and biological and structural evaluations of novel HIV-1 protease inhibitors to combat drug resistance. *J. Med. Chem.* 55, 6328–6341
61 Bahl, N. *et al.* (2011) Delineation of lipopolysaccharide (LPS)-binding sites on hemoglobin: from in silico predictions to biophysical characterization. *J. Biol. Chem.* 286, 37793–37803
62 Nalam, M.N. *et al.* (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.* 84, 5368–5378
63 Jorissen, R.N. *et al.* (2009) Additivity in the analysis and design of HIV protease inhibitors. *J. Med. Chem.* 52, 737–754
64 Sazinsky, S.L. *et al.* (2008) Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20167–20172
65 Altman, M.D. *et al.* (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *J. Am. Chem. Soc.* 130, 6099–6113
66 Sherman, W. and Tidor, B. (2008) Novel method for probing the specificity binding profile of ligands: applications to HIV protease. *Chem. Biol. Drug Des.* 71, 387–407
67 Altman, M.D. *et al.* (2008) Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* 70, 678–694

68 Lippow, S.M. *et al.* (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* 25, 1171–1176
69 Chellappan, S. *et al.* (2007) Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis. *Chem. Biol. Drug Des.* 69, 298–313
70 Green, D.F. *et al.* (2006) Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* 45, 12547–12559
71 Armstrong, K.A. *et al.* (2006) Optimal charges in lead progression: a structure-based neuraminidase case study. *J. Med. Chem.* 49, 2470–2477
72 Green, D.F. and Tidor, B. (2005) Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface. *Proteins* 60, 644–657
73 Ahn, J.S. *et al.* (2005) Defining Cdk5 ligand chemical space with small molecule inhibitors of tau phosphorylation. *Chem. Biol.* 12, 811–823
74 Zhao, S. *et al.* (2013) Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature* 502, 698–702
75 Datta, D. *et al.* (2013) Substrate and inhibitor-induced dimerization and cooperativity in caspase-1 but not caspase-3. *J. Biol. Chem.* 288, 9971–9981
76 Wallrapp, F.H. *et al.* (2013) Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1196–E1202
77 Ingram, J.R. *et al.* (2012) Investigation of the proteolytic functions of an expanded cercarial elastase gene family in Schistosoma mansoni. *PLoS Negl. Trop. Dis.* 6, e1589
78 Sadowsky, J.D. *et al.* (2011) Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6056–6061
79 Tan, E.S. *et al.* (2009) The molecular basis of species-specific ligand activation of trace amine-associated receptor 1 (TAAR1). *ACS Chem. Biol.* 4, 209–220
80 Tan, E.S. *et al.* (2008) Toward deciphering the code to aminergic G protein-coupled receptor drug design. *Chem. Biol.* 15, 343–353
81 Song, L. *et al.* (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat. Chem. Biol.* 3, 486–491
82 Morris, G.M. *et al.* (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791
83 Lang, P.T. *et al.* (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* 15, 1219–1230
84 Verdonk, M.L. *et al.* (2003) Improved protein–ligand docking using GOLD. *Proteins* 52, 609–623
85 Li, J. and Reynolds, C.H. (2004) Calculation of the binding affinities for stromelysin-1 (MMP-3) inhibitors using a linear scaling semi-empirical quantum chemistry method. *Abstr. Pap. Am. Chem. Soc.* 228, U506–U507
86 Kramer, B. *et al.* (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins Struct. Funct. Genet.* 37, 228–241
87 Yang, J.M. and Chen, C.C. (2004) GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* 55, 288–304
88 Marialke, J. *et al.* (2008) Similarity based docking. *J. Chem. Inf. Model.* 48, 186–196
89 Abagyan, R. *et al.* (1994) ICM – a new method for protein modeling and design – applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15, 488–506
90 Nabuurs, S.B. *et al.* (2007) A flexible approach to induced fit docking. *J. Med. Chem.* 50, 6507–6518
91 Corbeil, C.R. *et al.* (2007) Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* 47, 435–449
92 Meier, R. *et al.* (2010) ParaDockS: a framework for molecular docking with population-based metaheuristics. *J. Chem. Inf. Model.* 50, 879–889
93 Zsoldos, Z. *et al.* (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. *J. Mol. Graph. Model.* 26, 198–221
94 Kolb, P. and Caflisch, A. (2006) Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* 49, 7384–7392
95 Majeux, N. *et al.* (2001) Efficient electrostatic solvation model for protein-fragment docking. *Proteins Struct. Funct. Genet.* 42, 256–268
96 Budin, N. *et al.* (2001) Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* 382, 1365–13722
97 Trott, O. and Olson, A.J. (2010) Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461
98 Ellingson, S.R. *et al.* (2013) VinaMPI: Facilitating multiple receptor high-throughput virtual docking on high-performance computers. *J. Comput. Chem.* 34, 2212–2221
99 Zhao, Y. and Sanner, M.F. (2007) FLIPDock: Docking flexible ligands into flexible receptors. *Proteins Struct. Funct. Bioinform.* 68, 726–773
100 McGann, M. (2011) FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 51, 578–596
101 Lemmon, G. and Meiler, J. (2012) Rosetta ligand docking with flexible XML protocols. *Methods Mol. Biol.* 819, 143–155
102 Spitzer, R. and Jain, A.N. (2012) Surflex-Dock: docking benchmarks and real-world application. *J. Comput. Aided Mol. Des.* 26, 687–6997
103 Walters, W.P. and Murcko, M.A. (2002) Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.* 54, 255–271
104 Vilar, S. *et al.* (2008) Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem.* 8, 1555–1572
105 Stroganov, O.V. *et al.* (2008) Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. *J. Chem. Inf. Model.* 48, 2371–2385
106 Krieger, E. *et al.* (2002) Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. *Proteins* 47, 393–402
107 Shin, W.H. and Seok, C. (2012) GalaxyDock: protein–ligand docking with flexible protein side-chains. *J. Chem. Inf. Model.* 52, 3225–3232
108 Shin, W.H. *et al.* (2013) GalaxyDock2: protein–ligand docking using beta-complex and global optimization. *J. Comput. Chem.* 34, 2647–2656
109 Sauton, N. *et al.* (2008) MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics* 9
110 Brylinski, M. and Skolnick, J. (2009) FINDSITE: a threading-based approach to ligand homology modeling. *PLoS Comput. Biol.* 5, e1000405
111 Kim, D.S. *et al.* (2011) BetaDock: shape-priority docking method based on beta-complex. *J. Biomol Struct. Dyn.* 29, 219–242
112 Mizutani, M.Y. *et al.* (1994) Rational automatic search method for stable docking models of protein and ligand. *J. Mol. Biol.* 243, 310–326
113 Cozzini, P. *et al.* (2002) Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem.* 45, 2469–2483
114 Hart, T.N. and Read, R.J. (1992) A multiple-start Monte Carlo docking method. *Proteins Struct. Funct. Bioinform.* 13, 206–222
115 Korb, O. *et al.* (2009) Empirical scoring functions for advanced protein–ligand docking with PLANTS. *J. Chem. Inf. Model.* 49, 84–96
116 Grosdidier, A. *et al.* (2007) EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins* 67, 1010–1025
117 Pang, Y.P. *et al.* (2001) EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* 22, 1750–1771
118 Miller, M.D. *et al.* (1994) FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 8, 153–174
119 Welch, W. *et al.* (1996) Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* 3, 449–462
120 Gorelik, B. and Goldblum, A. (2008) High quality binding modes in docking ligands to proteins. *Proteins* 71, 1373–1386
121 Goto, J. *et al.* (2008) ASEDock-docking based on alpha spheres and excluded volumes. *J. Chem. Inf. Model.* 48, 583–590
122 Dominguez, C. *et al.* (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737
123 Lill, M.A. and Danielson, M.L. (2011) Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des.* 25, 13–19

124 Seeliger, D. and de Groot, B.L. (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* 24, 417–422
125 Vistoli, G. *et al.* (2010) Homology modeling and metabolism prediction of human carboxylesterase-2 using docking analyses by GriDock: a parallelized tool based on AutoDock 4.0. *J. Comput. Aided Mol. Des.* 24, 771–787
126 Bullock, C.W. *et al.* (2010) Dockomatic – automated ligand creation and docking. *BMC Res. Notes* 3, 289
127 Vaque, M. *et al.* (2006) BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock. *Bioinformatics* 22, 1803–1804
128 Grosdidier, A. *et al.* (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* 39, W270–W277
129 Bikadi, Z. and Hazai, E. (2009) Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.* 1, 15
130 Irwin, J.J. *et al.* (2009) Automated docking screens: a feasibility study. *J. Med. Chem.* 52, 5712–5720
131 Gupta, A. *et al.* (2007) ParDOCK: an all atom energy based Monte Carlo docking protocol for protein–ligand complexes. *Protein Pept. Lett.* 14, 632–646
132 London, N. *et al.* (2011) Rosetta FlexPepDock web server – high resolution modeling of peptide–protein interactions. *Nucleic Acids Res.* 39, W249–W253 (Web Server issue)
133 Schneidman-Duhovny, D. *et al.* (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367
134 Lee, H.S. and Zhang, Y. (2012) BSP-SLIM: a blind low-resolution ligand–protein docking approach using predicted protein structures. *Proteins* 80, 93–110
135 Li, L. *et al.* (2010) BioDrugScreen: a computational drug design resource for ranking molecules docked to the human proteome. *Nucleic Acids Res.* 38, D765–D773
136 Launay, G. *et al.* (2012) Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng. Des. Sel.* 25, 377–386
137 Martin, L. *et al.* (2006) kinDOCK: a tool for comparative docking of protein kinase ligands. *Nucleic Acids Res.* 34, W325–W333
138 Wang, J.C. *et al.* (2012) idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res.* 40, W393–W399
139 Fan, H. *et al.* (2011) Statistical potential for modeling and ranking of protein-ligand interactions. *J. Chem. Inf. Model.* 51, 3078–3092
140 Pyrkov, T.V. *et al.* (2009) PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics* 25, 1201–1202
141 Sobolev, V. *et al.* (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15, 327–332
142 Sakakibara, Y. *et al.* (2012) COPICAT: a software system for predicting interactions between proteins and chemical compounds. *Bioinformatics* 28, 745–746
143 Koes, D.R. and Camacho, C.J. (2011) Pharmer: efficient and exact pharmacophore search. *J. Chem. Inf. Model.* 51, 1307–1314
144 Desaphy, J. *et al.* (2012) Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J. Chem. Inf. Model.* 52, 2287–229929
145 Dror, O. *et al.* (2009) Novel approach for efficient pharmacophore-based virtual screening: method and applications. *J. Chem. Inf. Model.* 49, 2333–2343
146 Wolber, G. *et al.* (2006) Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aided Mol. Des.* 20, 773–788
147 Cereto-Massague, A. *et al.* (2012) DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* 28, 1661–1662
148 Zhang, S. *et al.* (2008) DOVIS: an implementation for high-throughput virtual screening using AutoDock. *BMC Bioinformatics* 9, 126
149 Liu, X. *et al.* (2011) SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *J. Chem. Inf. Model.* 51, 2372–2385
150 Wolf, K. (2009) Digital briefs: new software and websites for the chemical enterprise. *Chem. Eng. News* 87, 32
151 Abreu, R.M. *et al.* (2010) MOLA: a bootable, self-configuring system for virtual screening using AutoDock4/Vina on computer clusters. *J. Cheminform.* 2, 10
152 Durrant, J.D. and McCammon, J.A. (2011) NNScore 2.0: a neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.* 51, 2897–2903
153 Hu, Z. and Southerland, W. (2007) WinDock: structure-based drug discovery on Windows-based PCs. *J. Comput. Chem.* 28, 2347–2351
154 Jacob, R.B. *et al.* (2011) DockoMatic: automated peptide analog creation for high throughput virtual screening. *J. Comput. Chem.* 32, 2936–2941
155 Bullock, C. *et al.* (2013) DockoMatic 2.0: high throughput inverse virtual screening and homology modeling. *J. Chem. Inf. Model.* 53, 2161–2170
156 Reid, D. *et al.* (2008) LASSO-ligand activity by surface similarity order: a new tool for ligand based virtual screening. *J. Comput. Aided Mol. Des.* 22, 479–487
157 Thomsen, R. and Christensen, M.H. (2006) MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* 49, 3315–3321
158 Marialke, J. *et al.* (2007) Graph-based molecular alignment (GMA). *J. Chem. Inf. Model.* 47, 591–601
159 Durrant, J.D. and McCammon, J.A. (2012) AutoClickChem: click chemistry in silico. *PLoS Comput. Biol.* 8, e1002397
160 Todorov, N.P. *et al.* (2007) QUASI: a novel method for simultaneous superposition of multiple flexible ligands and virtual screening using partial similarity. *J. Chem. Inf. Model.* 47, 1007–1020
161 Jones, G. *et al.* (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.* 9, 532–549
162 Abrahamian, E. *et al.* (2003) Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J. Chem. Inf. Comput. Sci.* 43, 458–468
163 Reutlinger, M. *et al.* (2013) Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for 'orphan' molecules. *Mol. Inform.* 32, 133–138
164 Meraj, K. *et al.* (2012) Molecular modeling, docking and ADMET studies towards development of novel Disopyramide analogs for potential inhibition of human voltage gated sodium channel proteins. *Bioinformation* 8, 1139–1146
165 Puranen, J.S. *et al.* (2010) Accurate conformation-dependent molecular electrostatic potentials for high-throughput in silico drug discovery. *J. Comput. Chem.* 31, 1722–1732
166 Shelley, J.C. *et al.* (2007) Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21, 681–691
167 Cabrera, A.C. *et al.* (2011) VSDMIP 1.5: an automated structure- and ligand-based virtual screening platform with a PyMOL graphical user interface. *J. Comput. Aided Mol. Des.* 25, 813–824
168 Betzi, S. *et al.* (2006) GFscore: a general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.* 46, 1704–1712
169 Zhou, H. and Skolnick, J. (2013) FINDSITE(comb): a threading/structure-based, proteomic-scale virtual ligand screening approach. *J. Chem. Inf. Model.* 53, 230–240
170 Zhou, H. and Skolnick, J. (2012) FINDSITE(X): a structure-based, small molecule virtual screening approach with application to all identified human GPCRs. *Mol. Pharm.* 9, 1775–1784
171 Douguet, D. (2010) e-LEA3D: a computational-aided drug design web server. *Nucleic Acids Res.* 38, W615–W621
172 Gillet, V.J. *et al.* (2002) Combinatorial library design using a multiobjective genetic algorithm. *J. Chem. Inf. Comput. Sci.* 42, 375–385
173 Koes, D.R. and Camacho, C.J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40, W409–W414
174 Floris, M. *et al.* (2011) Swimming into peptidomimetic chemical space using pepMMsMIMIC. *Nucleic Acids Res.* 39, W261–W269
175 Sperandio, O. *et al.* (2009) wwLigCSRre: a 3D ligand-based server for hit identification and optimization. *Nucleic Acids Res.* 37, W504–W509
176 Bauer, R.A. *et al.* (2008) Superimpose: a 3D structural superposition server. *Nucleic Acids Res.* 36, W47–W54

177 Chen, Y.F. *et al.* (2010) SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. *Nucleic Acids Res.* 38, W424–W430
178 Vainio, M.J. *et al.* (2009) ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* 49, 492–502
179 Liu, X. *et al.* (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* 38, W609–W614
180 Chang, T.T. *et al.* (2011) In silico pharmacology suggests ginger extracts may reduce stroke risks. *Mol. Biosyst.* 7, 2702–2710
181 Hung, T-C. *et al.* (2014) In silico investigation of traditional Chinese medicine compounds to inhibit human histone deacetylase 2 for patients with Alzheimer's disease. *Biomed. Res. Int.* 2014, 769867
182 Hung, T-C. *et al.* (2014) Lead screening for HIV-1 integrase (IN) inhibited by traditional Chinese medicine. *Biomed. Res. Int.* 2014, 479367
183 Chen, K-C. *et al.* (2014) Potential protein phosphatase 2A agents from traditional Chinese medicine against cancer. *Evid. Based Complement. Altern. Med.* 2014, 436863
184 Huang, H-J. *et al.* (2014) Lead discovery for Alzheimer's disease related target protein RbAp48 from traditional Chinese medicine. *Biomed. Res. Int.* 2014, 764946
185 Tietze, S. and Apostolakis, J. (2007) GlamDock: development and validation of a new docking tool on several thousand protein–ligand complexes. *J. Chem. Inf. Model.* 47, 1657–1672