

Free Resources to Assist Structure-Based Virtual Ligand Screening Experiments

Bruno O. Villoutreix*, Nicolas Renault, David Lagorce, Olivier Sperandio, Matthieu Montes and Maria A. Miteva

INSERM U648, University Paris V, 45 rue des Sts Peres, 75006 Paris, France

Abstract: In today's research environment, a wealth of experimental/theoretical structural data is available and the number of therapeutically relevant macromolecular structures is growing rapidly. This, coupled with the huge number of small non-peptide potential drug candidates easily available (over 7 million compounds), highlight the need of using computer-aided techniques for the efficient identification and optimization of novel hit compounds. Virtual (or in silico) ligand screening based on the three-dimensional structure of macromolecular targets (SB-VLS) is firmly established as an important approach to identify chemical entities that have a high likelihood of binding to a target molecule to elicit desired biological responses. A myriad of free applications and services facilitating the drug discovery process have been posted on the Web. In this review, we cite over 350 URLs that are useful for SB-VLS projects and essentially free for academic groups. We attempt to provide links for in silico ADME/tox prediction tools, compound collections, some ligand-based methods, characterization/simulation of 3D targets and homology modeling tools, druggable pocket predictions, active site comparisons, analysis of macromolecular interfaces, protein docking tools to help identify binding pockets and protein-ligand docking/scoring methods. As such, we aim at providing both, methods pertaining to the field of Structural Bioinformatics (defined here as tools to study macromolecules) and methods pertaining to the field of Chemoinformatics (defined here as tools to make better decisions faster in the arena of drug/lead identification and optimization). We also report several recent success stories using these free computer methods. This review should help readers finding free computer tools useful for their projects. Overall, we are confident that these tools will facilitate rapid and cost-effective identification of new hit compounds. The URLs presented in this review will be updated regularly at www.vls3d.com in the coming months, "Links" section.

Keywords: Virtual ligand screening, structure-based drug design, docking, scoring, databases, structural bioinformatics.

INTRODUCTION

Researchers including protein scientists are under ever increasing pressure to discover new therapeutic compounds, as there are urgent needs to identify new molecules able to fight against life-threatening diseases. Drug discovery is however a complex and expensive endeavor that requires the use of various techniques such as experimental high-(or medium)-throughput screening (HTS), NMR, X-ray crystallography, experimental ADME (absorption, distribution, metabolism and excretion)/tox assays and combinatorial chemistry among many others [1-4]. Yet, to assist this highly complex process (Fig. 1a), it is now well accepted that computer methods such as homology modeling, protein docking and virtual screening/de novo design can help [3, 5-31].

With regard to small ligand-receptor *in silico* screening methods, one can usually distinguish two main virtual screening strategies/concepts: ligand-based screening and structure-based screening [32, 33]. For ligand-based methods (similarity and substructure search, clustering, QSAR, pharmacophore matching or three-dimensional shape matching), the concept is to use information provided by compounds

that are known to bind to the desired target and to use these data to identify other molecules in the databases with similar properties. For structure-based methods (SB-VLS) (Fig. 1b), it is assumed that the three-dimensional (3D) structure of the target is known either by X-ray crystallography or NMR experiments or predicted by homology. The principle here is to dock all the ligands present in a database into the binding pocket of the selected target and evaluate the fit between the molecules. The quality of the fit is then used to rank the small molecules. To perform SB-VLS computations, many tools are usually needed: protein structure prediction, structural analysis, compound collections, in silico ADME/tox prediction, definition of ligand binding site (selection of targets) and in silico docking and scoring methods [9, 28, 32, 34-39].

In the present review, we introduce and summarize many applications usually free to academic groups to prepare the targets and the compound collections and to perform SB-VLS experiments. In addition, we give several examples of successful hit identification using these tools. The different programs and online tools are listed in 8 Tables and in some cases can appear in several Tables since they may match several keywords. Some tools dealing with macromolecules (e.g., sequence alignment and related) are not extensively listed here since the goal of this review is to focus on virtual screening while still providing a significant number of links

*Address correspondence to this author at the INSERM U648, University Paris 5, 45 rue des Sts Peres, 75006 Paris, France;
E-mail: bruno.villoutreix@univ-paris5.fr

to programs useful to prepare and visualize macromolecules for SB-VLS experiments. Additional data about “free” structural bioinformatics tools can be found in [40, 41]. Along the same line, we do not list many free tools able to perform quantum mechanics computations, interested users can find valuable information in [42]. Thus, the present review should provide many useful links to researchers working in the field of Structural Bioinformatics and Drug Design.

I. PATENT SEARCH AND CHEMISTRY TOOLKITS TO DEAL WITH BOTH MACRO- AND SMALL MOLECULES

Prior to starting a SB-VLS project, it is usually important to carry out patent search on the target of interest and on small molecules binding this target [11]. Online tools to perform such inquiries are presented in Table 1. Then, it is important to install one or several chemistry toolkits and molecular graphics systems that are able to read and manipulate small/large molecule file formats, such as PDB, Mol2, SDF or SMILES [43, 44]. Several tools or tutorials dealing with small molecule and macromolecule file formats are listed in Table 2.

Table 1. Patent Search

http://www.plutarque.com/plutarque
http://ep.espacenet.com/
http://www.european-patent-office.org/inpadoc/
http://www.lib.duke.edu/chem/patents.htm
http://www.uspto.gov/patft/index.html
http://www.ercim.org/publication/Ercim_News/enw60/zimmermann.html A project has been initiated to extract chemical information from depictions of molecules in the public literature

II. IN SILICO ADME/TOX FILTERING

About 30 million compounds have been synthesized (found) during these last 20-40 years, yet, many of them do not have satisfactory ADME (absorption, distribution, metabolism and excretion)/tox properties to be considered as promising drug candidates (Fig. 2). One of the primary reasons for ADME/tox prediction is due to the fact that unacceptable ADME/tox properties account for about 40% of drug candidate failures in the clinical phase of drug development [45]. Experimental ADME/tox evaluation methods usually involve solubility and metabolism assays [46, 47], among others, but it is also possible to predict some of these properties (e.g., absorption prediction via computation of logP, molecular weight, polar surface area (e.g., if greater than 140 Å², poor intestinal absorption), H-bond donor acceptor) using in silico methods (see for instance [38, 48-54]). With regard to in silico ADME/tox prediction, it is commonly accepted that absorption, distribution and excretion are dependent on similar descriptors while metabolism and toxicity are of a quite different nature and depend of numerous factors. Clearly all in silico ADME prediction methods (this observation also applies to *in vitro* and in animal models) should be considered with care because the human or-

ganism is an immensely complicated system and it is not possible at present to simulate all the different events in a computer [53, 55, 56]. Yet, in silico tools can provide valuable information in many cases [57].



Fig. (2). ADME/tox: properties that make a compound “drug-like”

For a compound to become a successful drug, it needs not only to perform the right molecular interactions with its target, but also to be processed correctly by the body. For most projects the ideal compound is soluble, able to pass through the gut wall, not broken down too quickly by liver enzymes and has a minimal affinity to fatty tissues and blood proteins. Collectively these properties are known as the ADME properties (for absorption, distribution, metabolism and excretion). In addition, a compound should exist in the body long enough to have a therapeutic effect, this is often referred to as its pharmacokinetics properties.

Small molecules usually go through one or several in silico ADME/tox filtering steps in an attempt to generate a database of molecules that have physical properties and chemical functionality consistent with known drugs/leads/hits (a molecule having characteristics, topological descriptors, physicochemical descriptors, similar to those of known marketed drugs can be classified as “druglike” [58-60]). Several methods are available to compute some of these properties (Table 3). Common filtering protocols are variations of Lipinski’s rule-of-five (potential for oral bioavailability), they can include a limit on the number of rotatable bonds, on the polar surface area, on calculated logP among others [50, 55, 57, 61-78]. For instance, the lipophilicity (traditionally expressed as logP) of a drug is the most used single physicochemical property to predict its permeation in biological systems. Many different approaches have been developed to try to compute this value, they all involve simplification and are not yet able to fully reproduce experimental results. Several “Lipinski-like” approaches are available online, the input is usually a molecule in SDF, Mol2 or SMILES format. The so-called “rule-of-five” [78] states that if a compound satisfies any two of the following

Table 2. Chemistry Toolkits, Graphics (Small and Macro-Molecules) and Utilities

URLs	Short summary	Keywords
http://www.cheminformatics.org/	Links to cheminformatics programs and QSAR datasets. These include, diversity and similarity searches. Many compounds designed for specific targets (e.g., coagulation factor Xa...).	Compound searching
http://wwmm-svc.ch.cam.ac.uk/wwmm/html/?cid=observer	World wide molecular matrix, provide several services, including OpenBabel online	Chemistry tools
http://www.tripos.com/mol2/mol2_format2.html	Mol2 file format (2D or 3D)	Mol2 format (Chemistry)
http://www.rcsb.org/pdb	The Protein Data Bank, see section describing the PDB format [106]	Macro-molecule 3D structures
http://pubs3.acs.org/acs/journals/toc.page?incoden=jcisid8&indecade=1&involume=32&inissue=3	Information about different chemical structure file formats including SDF [44]	Chemistry tools
http://openbabel.sourceforge.net/babel.shtml http://www.es.embnet.org/Services/MolBio/babel/ http://vcclab.org/lab/babel/	OpenBabel: File format conversion	Chemistry tools
http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html	FAF-Drugs: OpenBabel online [89]	Chemistry tools
http://www.macinchem.fsnet.co.uk/applescripts.htm	iBabel: File format conversion and other chemistry tools (essentially for Mac or Linux/Unix)	Chemistry tools
http://www.daylight.com/cheminformatics/tutorials/index.html	Tutorial for SMILES and chemistry toolkit	SMILES format (Chemistry)
http://www.ccl.net/cca/data/MMFF94s/	MMFF validation suite	Simulation tools
http://sourceforge.net/projects/perlmol	Chemistry toolkit	Chemistry tools
http://www.iupac.org/dhtml_home.html	IUPAC International Chemical Identifier project	Chemistry tools
http://www2.chemie.uni-erlangen.de/services/gifcreator/	GIF/PNG-creator with SMILES input	Compound drawing
http://www-ra.informatik.uni-tuebingen.de/software/joelib/ http://sourceforge.net/projects/joelib/	Computational chemistry package	Chemistry tools
http://www.uku.fi/~thassine/ghemical/ http://www.uiowa.edu/~ghemical/osx.shtml	Computational chemistry package [194]	Chemistry tools
http://www.molinspiration.com/jme/	JME: Java Molecular Editor by Dr. P. Ertl, draws small molecules and get SMILES	Compound drawing
http://www.chemaxon.com/products.html	Computer tools for chemistry Marvin is a suite of Java based chemistry software that have different forms: Marvin Applets, Marvin Beans, MarvinSketch	Chemistry tools
http://sourceforge.net/projects/cdk/	CDK: Chemistry development kit [195]	Chemistry tools
http://sourceforge.net/projects/frowns	Chemistry toolkit	Chemistry tools
http://chemcpp.sourceforge.net/html/index.html	a C++ toolbox for chemoinformatics [196]	Chemistry tools
http://sourceforge.net/projects/sketchel	SketchEI: Chemical structure sketching tool	Compound drawing
http://cactus.nci.nih.gov/services/translate/	Online SMILES translator and structure generator from F. Oellien and M.C. Nicklaus	Compound drawing

(Table 2). Contd....

URLs	Short summary	Keywords
http://cactus.cit.nih.gov/SDF_toolkit/	The SDF toolkit (in Perl) essentially for small molecules	Chemistry tools
http://www.umass.edu/microbio/chime/index.html	To display and rotate macromolecules and small molecules in a single internet browser	Molecular graphics
http://hugin.ethz.ch/wuthrich/software/molmol/index.html	MOLMOL [197]: molecular graphics program for the structure of biological macromolecules	Molecular graphics
http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml	Cn3D {Hogue, 1997 #2595: displays structures of macromolecules and performs sequence alignments	Molecular graphics
http://cobra.mih.unibas.ch/dino/intro.php	DINO : 3D viewer essentially for macromolecules	Molecular graphics
http://www.rahul.net/pcm/dismol/	DisMol: Java applet viewer for macromolecules and small molecules	Molecular graphics
http://www.avatar.se/molscript/doc/molscript.html http://www.strubi.ox.ac.uk/bobscript/	MolScript: creates molecular graphics image of macromolecules and small molecules [198]	Molecular graphics
http://asia.genesilico.pl/colorado3d/	Colorado3D [199]: web server for the visual analysis of protein structures	Molecular graphics Structural analysis
http://pymol.sourceforge.net/	Pymol : molecular graphics system to look at macromolecules and small molecules	Molecular graphics
http://www.ks.uiuc.edu/Research/vmd/	VMD [200]: molecular visualization program for displaying, animating, and analyzing large systems	Molecular graphics
http://www.molworks.com/en/	MolWorks: graphic tool for drawing and sketching molecules	Chemistry tools
http://geoffhutchison.net/projects/chem/	ChemSpotlight: metadata importer plugging for Mac OS X, which reads common chemical file formats (PDB, Mol2, SDF...)	Chemistry tools
http://www-ibmc.u-strasbg.fr/upr9002/westhof/	DRAWNA [201]: program for drawing schematic views of nucleic acids	Molecular graphics
http://www.molsoft.com/icm_browser.html	ICM browser [202]: biomolecular modeling package (can read many different file formats)	Molecular graphics
http://www.eyesopen.com/products/applications/vida.html	OpenEye Vida : molecular modeling package for macromolecules and small molecules (can read many different file formats)	Molecular graphics
http://www.chemicalgraphics.com/PovChem/	PovChem is a chemical visualization and illustration program, it can calculate and display hydrogen bonds	Molecular graphics
http://www.csc.fi/gopenmol/	With gOpenMol allows visualization and analysis of small molecules, and to lesser extent protein structures, of chemical properties, total electron densities and molecule orbitals [203]	Molecular graphics
http://mitglied.lycos.de/PageOfMH/	KMovisto is a 3D molecule viewer essentially for Linux. It can import and export OpenBabel files	Molecular graphics
http://www.yasara.org/	YASARA is a molecular-graphics, -modeling and -simulation program [204, 205]	Molecular graphics & modeling

(Table 2). Contd....

URLs	Short summary	Keywords
http://gdis.seul.org/ http://gdis.sourceforge.net/	GDIS is a program for the display and manipulation of isolated molecules and periodic systems	Molecular graphics
http://kemistry.sourceforge.net/kdrawchem.php http://xdrawchem.sourceforge.net/	KDrawChem and XDrawChem are molecular structure drawing programs	Compound drawing
http://bkchem.zirael.org/index.html	BKchem is a chemical drawing program	Compound drawing
http://www.cgl.ucsf.edu/chimera/index.html	UCSF Chimera [206]: biomolecular modeling package	Molecular graphics & modeling
http://sourceforge.net/projects/tautgen	Tautomer generator is a program that generates a set of molecules (tautomers) from a molecular core and number of hydrogen atoms	Simulation tools
http://www.moloc.ch/about.html	Moloc: Roche Biostructural modeling package for small and large molecules	Chemistry tools
http://www.jmol.sourceforge.net	Open source molecule viewer	Molecular graphics
http://sourceforge.net/projects/jchempaint	JChemPaint is a program for drawing 2D chemical structures	Compound drawing
http://www.mayachemtools.org/index.html	MayaChemTools is a growing collection of Perl scripts to support day-to-day computational discovery needs	Chemistry tools
https://sourceforge.net/projects/mcdl	A Java Chemical Structure Editor [207]	Compound drawing
http://molvis.sdsc.edu/vis-res/molvisfw/author_no_descriptions.jsp	Links to Free Molecular Visualization and Modeling Software	Molecular graphics & modeling
http://bioclipse.net/index.php?option=com_content&task=view&id=2&Itemid=4	Bioclipse is a Java-based visual platform for chemo- and bioinformatics	Molecular graphics
http://www.epa.gov/ncct/dssto/MoreonSDF.html	Information about SDF format	SDF format

Table 3. ADME/tox Prediction and Databases

URLs	Short summary	Keywords
http://vcclab.org/lab/alogps http://146.107.217.178/lab/alogps/	AlogP: Tools to predict logP (with several methods) [208]	ADME/tox
http://zinc.docking.org	ZINC [90]: ADME/tox online	ADME/tox
http://edetox.ncl.ac.uk/	Find compound properties	Chemistry
http://blue.chem.psu.edu/~rajarshi/code/java/	ADME/tox based on CDK [195]	ADME/tox
http://sourceforge.net/projects/solubility/	ADME/tox computations	ADME/tox
http://www.eyesopen.com	ADME/tox online	ADME/tox
http://www.molinspiration.com/	ADME/tox online	ADME/tox
http://www.molsoft.com/mprop	ADME/tox online	ADME/tox
http://www.chemaxon.com/products.html	ADME/tox online	ADME/tox
http://www.mol-net.de/	ADME/tox online	ADME/tox
http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html	FAF-Drugs: ADME/tox online [89]	ADME/tox

(Table 3) contd...

URLs	Short summary	Keywords
http://www.syrres.com/esc/est_soft.htm http://www.syrres.com/esc/est_kowdemo.htm	Compute logP, retrieves experimental logP for over 13,000 compounds	ADME/tox
http://home.pchome.com.tw/team/gentamicin/mol/mol.htm	DBFILTER can check the mol2 format of compounds in a database and pick out problematic structures for the docking package DOCK. It can also compute ADME/tox properties (12 kinds of filters)	ADME/tox
http://sw16.im.med.umich.edu/software/xtool/	Xscore : logP computation tool [209]	ADME/tox
http://www.logp.com/	Compute logP	ADME/tox
http://www.syrres.com/esc/physdemo.htm	PHYSPROP database, contains chemical structures, names and physical properties for over 25,000 compounds	Chemistry Database
http://www.moldiscovery.com/soft_metasite.php	The MetaSite has been developed to predict the site of metabolism (i.e., the place in a molecule where the metabolic reaction occurs) for substrates of 2C9, 2D6, 3A4, 1A2 and 2C19 cytochromes [210]	ADME/tox
http://intro.bio.umb.edu/downloads/jlogp/	logP prediction	ADME/tox
http://www.compudrug.com/	Artificial neural network based approach, using atomic fragmental descriptors, to predict logP on a wide range of organic compounds and other ADME/tox tools (Demo)	ADME/tox
http://preadme.bmdrc.org/preadme/index.php	PreADMET is a web-based application for predicting ADME data	ADME/tox
http://www.niehs.nih.gov/nct/cebs.htm	Chemical Effects in Biological Systems (CEBS) knowledge base (application of systems biology to ADME/Tox)	ADME/tox
http://www.fda.gov/nctr/science/centers/toxicoinformatics/index.htm	Toxicoinformatics database at the FDA (application of systems biology to ADME/Tox)	ADME/tox Database
http://edge.oncology.wisc.edu/	EDGE [211]: scientific resource for toxicology-related gene expression information (application of systems biology to ADME/Tox)	ADME/tox
http://cgl.imim.es/biochemoinformatic.htm	cMoIP: compute molecular properties	ADME/tox
http://bidd.nus.edu.sg/group/bidd.htm	DART [212], ADME-AP [213], TRMP [214], TTD [215]: databases for facilitating the search for drug Absorption, Distribution, Metabolism, Excretion associated proteins	Databases for ADME/tox
http://www.predictive-toxicology.org/lazar/form.php	Lazar [216]: tool for the prediction of toxic activities of chemical structures	ADME/tox
http://www.compumine.se/adme/adme.jsp	Compumime: tool for the prediction of ADME properties	ADME/tox
http://umbbd.msi.umn.edu/predict/	Biocatalysis/biodegradation database [217]: tool for prediction of microbial catabolic reactions involving chemical structures	ADME/tox
http://www.molworks.com/en/	MolWorks: many tools including methods for estimating the properties of small molecules	ADME/tox
http://www.epa.gov/ncct/dsstox/index.html	Distributed Structure-Searchable Toxicity (DSSTox) Database Network [218, 219]	Databases & ADME/tox
http://toxnet.nlm.nih.gov/	Databases on toxicology, hazardous chemicals, environmental health, and toxic releases	Databases & ADME/tox

rules, it is likely to exhibit poor intestinal absorption: (a) molecular weight > 500; (b) the number of hydrogen bond

donors > 5 (O-H or N-H groups); (c) the number of hydrogen bond acceptors > 10 (any N or O atom, including do-

nors); (d) calculated $\log P > 5.0$. Other rules involve removing compounds containing specific chemical substructures associated with poor chemical stability, reactivity or toxicity (reactive groups include epoxides, anhydrides...), frequent hitters and promiscuous inhibitors, while methods to predict drug metabolism (e.g. cytochrome-mediated metabolism, Pgp efflux...) are under developments and are usually integrated only in commercial packages. The selected molecules after applying Lipinski's rule-of-five or related filters are erroneously called "drug-like" since in fact many organic chemicals conform to the above listed rules while they are by no means drug-like [1]. In other words, these rules define only some necessary conditions for a drug candidate (such as likely solubility, bio-availability), not sufficient ones. In all cases, it is important to note that these *in silico* filters should not be applied blindly but rather tailored to a specific project.

III. COMPOUND COLLECTIONS

Databases of small molecules are needed for basically all virtual screening projects. These compound collections (purchasable molecules, about 7 millions in total in 2006) can be obtained in 2D and in general in SDF format from Chemical vendors. Often, as mentioned above, these molecules need to go through ADME/tox filtering tools. Many collections are listed in (Table 4), some libraries contain only drugs that are already in use today while some others contain only natural products [10, 79, 80]. Several suggestions have been made to facilitate the design of a database suitable for *in silico* screening experiments, interested users can find recommendations in the following articles [9, 81-90]. In some situations, scientists may need a database of small fragments and here also, suggestions have been reported about how to prepare such a library [91, 92]. Collections of small molecules co-crystallized in a binding pocket with or without information about experimental K_d or K_i or IC_{50} can be important to probe new docking-scoring methods. These databases are also listed in (Table 4). For SB-VLS experiments, the small molecules have to be in 3D. Several tools able to generate 3D structures from SMILES or SDF input are listed in Table 5. However, it is important to note that computing charges, multiple protonation states, stereo-chemistries (e.g., racemic mixtures) and regio-isomeric forms (E/Z isomerism), tautomeric, and conformational states (multiple conformers) for these databases are challenging and obviously not always possible [9]. Some data have been published to help users generate the bioactive conformations of the ligands (see for example [93]) but these computations are obviously challenging. A few databases are ready to download (molecules in 3D with ADME/tox filtering, multiple conformers or single conformation in Mol2 format) and can be found on the Web (see for instance [89, 90]; Table 4). Several recent studies suggest to focus the chemical space of compound libraries according to restraints dictated by the receptor or receptor family [94]. Thus, it is possible to build target-specific compound libraries starting from available compound collections and via the use of SB-VLS or via ligand-based VLS experiments.

IV. TARGET PREPARATION – BINDING SITE DEFINITION

SB-VLS is the process of identifying chemical compounds that may bind to a given receptor by computationally

screening a selected 3D receptor-binding pocket against a compound library by automated docking. Thus, the 3D structures of the receptors are required to perform this task. They are typically obtained by X-ray crystallography, NMR, but homology models can also be used [9, 16, 95-100]. Creating a homology model is generally carried out in several steps: finding a known template structure (or structures) that has a high sequence similarity/identity with the target sequence (experimental target structures are deposited at the Protein Data Bank, see below); aligning the target sequence and the template sequence taking into account structural data; building the model (the core and loop segments if possible), refinement of the model structure and assessing the quality of the resulting 3D model [18]. Packages to build/refine models are listed in (Tables 2, 6a and 8). Many tools are nowadays available to build a model and to evaluate its quality, several of them have been reviewed recently by Nayeem *et al.* [101]. In some cases, the target is not a protein but can be RNA or DNA [102]. Several tools dealing with these macromolecules are also listed in (Tables 2, 6, 7 to 8). For some projects, it can be important to investigate the overall and/or local flexibility of the targets. Many approaches can be used for this purpose, some of them are based on experimental data from NMR, crystallography or biochemistry, other are based on theoretical calculations like molecular dynamics (MD) simulation, enhanced MD, Monte Carlo simulation, normal mode analysis or principal component analysis [103]. These calculations may be performed on a protein or more generally on key members of a protein family. Other methods to investigate the flexibility of the receptor include the graph-theory algorithm FIRST [104] and computing packing density [105]. Several tools able to investigate receptor flexibility are reported in (Tables 2 and 8).

The target 3D structures usually need to be prepared for SB-VLS experiments (these structures can be found at the Protein Data Bank [106] and in some other databases, see Table 6a). This step often involves the addition of hydrogen atoms and the prediction of the correct protonation state for the titratable residues (pKa prediction [107], electrostatic computations), adding or removing tightly-bound water molecules, counterions, metal ions, cofactors, sugar molecules, removing subunits not involved in ligand binding or far from the binding site, introducing corrections for the tautomeric states of histidine residues and re-orientations of hydroxyl groups. Several programs performing some of these tasks are reported in Tables 2, 6 and 8.

Next, it is important to identify binding pockets and "hot spots" on the selected targets [9, 34, 108-114]. In some situations, the region that needs to be probed is well known, like a catalytic site, but in some other cases, tools to predict binding pockets and hot spots are needed. Programs to predict possible binding sites for SB-VLS experiments are listed in (Table 6b), these methods are usually geometry-based and/or energy based. Because numerous projects now aim at inhibiting macromolecular interactions (e.g., protein-protein interaction)[35, 115], information about macromolecular interfaces may need to be predicted. Such data can be obtained via macromolecular docking combined with site directed mutagenesis experiments [116, 117]. Several free macromolecular docking tools are listed in (Table 6b). In some studies, it can be beneficial to analyze in depth macromolecular

Table 4. Free Compound Collections, Target-Ligand Databases and Utilities

URLs	Short summary	Keywords
http://chembank.med.harvard.edu	ChemBank: Free collections and utilities, known drugs, many annotated molecules, molecules with druglike and non-druglike properties	Compound Database, Compound searching
http://pubchem.ncbi.nlm.nih.gov/	PubChem: An information resource linking chemistry and biology	Compound Database
http://www.chemthes.com/index.html	Chemical thesaurus : database including chemical entities, interactions, reactions, processes	Compound Database
http://www.chemnet.com/	Chemical suppliers and collections	Compound Database
http://www.rxlist.com/02top.htm	Top 200 prescriptions in 2002 (structure and name of the compounds)	Compound Database
http://www.bioscreening.com/compound_libraries.htm	Web directory about compound collections and many related links, database search	Compound Database, Compound searching
http://www.cermn.unicaen.fr/chimiotheque	Free collections	Compound Database
http://www.genome.ad.jp/dbget/ligand.html	Free collections	Compound Database
http://www.nmrshiftdb.org http://sourceforge.net/projects/nmrshiftdb	NMRShiftDB - Free collection, some molecules are in 3D [220-222]	Compound Database
http://Ligand.info	Utilities such as ligand clustering and ligand similarity search [223]	Compound searching
http://cdb.ics.uci.edu/CHEM/Web/	ChemDB: Free collections and utilities such as similarity search	Compound Database, Compound searching
http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html (see also: http://www.vls3d.com/)	FAF-Drugs: Free collections (and ADME/tox) and utilities [89]	Compound Database, ADME/Tox
http://zinc.docking.org	ZINC: Free collections [90] (and link to commercial vendors)	Compound Database, ADME/Tox, Compound searching
http://bioweb.ucr.edu/ChemMine	ChemMine: Free collections and similarity search utilities [224, 225]	Compound Database, Compound searching
http://www.mdli.com	Available Chemicals Directory (essentially commercial)	Compound Database, Compound searching
http://www.chemnavigator.com	Commercial collection (in part commercial)	Compound Database, Compound searching

(Table 4) contd....

URLs	Short summary	Keywords
http://www.ebi.ac.uk/chebi/	Dictionary of small molecules	Compound Database
http://www.bindingdb.org/	BindingDB: Measured binding affinities, macromolecule-ligand complexes [226]	Compound Database & Macromolecules
http://www.pdbbind.org/	PDBbind: macromolecules with co-crystallized ligands and experimental binding affinities [227]	Compound Database & Macromolecules
http://kibank.iis.u-tokyo.ac.jp/	KiBank: Proteins with co-crystallized ligands and experimental binding affinities [228]	Compound Database & Macromolecules
http://relibase.ebi.ac.uk	RELIBASE: Proteins with co-crystallized ligands [229-231]	Compound Database & Macromolecules
http://www.ccdc.cam.ac.uk/products/life_sciences/validate/astex/	CCDC/Astex validation test set: 305 protein-ligand complexes to calibrate docking and scoring tools [232]	Compound Database & Macromolecules
http://redpoll.pharmacy.ualberta.ca/drugbank/	DrugBank: Numerous data about drugs and targets including drugs already in use [233]	Compound Database & Macromolecules
http://www.agklebe.de/affinity	AffinDB: Proteins with co-crystallized ligands and experimental binding affinities [234]	Compound Database & Macromolecules
http://www.inteligand.com/	Ilib Diverse: tool to create virtual drug-like libraries	Virtual Database generator
http://dtp.nci.nih.gov/index.html	The US National Cancer Institute collections including natural products	Compound Database
http://cactus.nci.nih.gov/ncidb2/chem_www.html http://cactus.nci.nih.gov/ncidb2/	Compilation of web sites that offer chemistry databases/search services, data about toxic molecules, hazardous substances...database browser	Compound Database & ADME/Tox
http://solvdb.ncms.org/solvdb.htm	A free database of commercially available solvents searchable by many properties	Solvent Database
http://www.molecularmodels.ca/ http://www.molecularmodels.ca/molecule/molecule_index.html	Models of main functional groups, courses in organic chemistry...	Chemistry courses
http://www.inf.uni-konstanz.de/bioml/research/index.html	MoFa: Molecular fragment miner	Compound searching
http://sourceforge.net/projects/chem-file/	Main chemical-structures	Compound Database
http://chemfinder.cambridgesoft.com/	View properties, purchase compounds	Compound Database

(Table 4) contd....

URLs	Short summary	Keywords
http://nci.chemfinder.com	View structures and data of Open NCI DB compounds	Compound Database
http://edetox.ncl.ac.uk/	Find compound properties	Compound searching
http://www.genome.ad.jp/dbget/ligand.html	Similarity search and other tools	Compound searching
http://www.emolecules.com/	With eMolecules you can draw your chemical structure and instantly search millions of molecules from across the Web and from chemical suppliers worldwide	Compound searching
http://bioinf.charite.de/superligands/	SuperLigands [235]: a database of ligand structures derived from the Protein Data Bank with similarity searches and other tools	Compound Database from the PDB, Compound searching
http://www.qspr.pe.kr/chemdb.html	Chemistry and biology database: numerous links (databases, tools) valuable for drug design projects	Compound Database, macromolecules & links to programs
http://alpha2.bmc.uu.se/hicup/	Small molecules from the PDB	Compound Database from the PDB
http://chem.sis.nlm.nih.gov/chemidplus/	ChemID Plus: chemical name, physical and toxicological properties	Compound Database, ADME/Tox
http://llama.med.harvard.edu/~jklekota/QueryChem.html	QueryChem [236]: searches public databases using text and structure	Compound Database, Compound searching
http://www.zelinsky.ru/	Compound collection and building blocks	Compound Database

interfaces and several programs and online tools can help in this process, they are also listed in (Table 6b). If binding sites (or docking results) are probed by site directed mutagenesis or if the receptor is present in a micro-organism in which drug resistance is likely, it can be important to use computer tools to predict the potential structural impact of a mutation on a given target. Several programs and online methods are available to carry out this analysis (Table 7). Many tools for comparing target binding sites and for identifying co-crystallized ligands present at the PDB able to fit into a target structure (e.g., detection of cross-reacting targets or of ligands that are already known to fit a given pocket) have been reported and are generally freely accessible online [118-121] (Table 7). A database of targets for drug design projects has recently been prepared for such purpose and for inverse docking experiments [122, 123].

V. DOCKING AND SCORING SMALL MOLECULES

Assuming that the receptor 3D structure is available, a primary challenge in lead discovery via SB-VLS is to predict

both ligand orientation and binding affinity; the former is often referred to as 'molecular docking' while the latter is referred to as 'scoring'. Docking protocols can be described as a combination of two components; a search strategy and a simple scoring/fitting function to assess the poses prior to the real scoring step. The search algorithm should generate an optimum number of conformations for the ligands and of poses that should include the experimentally determined binding mode. The complexity of molecular docking implies several approximations, from rigid body docking (where both partners are treated as rigid but "flexibility" can be generated prior to docking), to (pseudo)-flexible ligand docking (where the receptor is held rigid and the ligand is partially flexible) to flexible docking (where both receptor and ligand flexibility are considered). Algorithms dealing with flexibility can be divided in three types, namely systematic (e.g., incremental construction algorithms; conformational search methods, database methods with libraries containing pre-generated conformations for each molecules, stochastic or random algorithms (e.g. Monte Carlo methods, Tabu

Table 5. Small Molecules 2D-to-3D, 2D or 3D Search and De Novo Ligand Builder

URLs	Short summary	Keywords
http://blue.chem.psu.edu/~rajarshi/code/java/	2D to 3D based on CDK [195]	Small molecules 2D-to-3D
http://iris12.colby.edu/~www/jme/smiledg.html http://iris12.colby.edu/~www/jme/dg.html	2D to 3D conversion and other tools	Small molecules 2D-to-3D
http://relibase.ebi.ac.uk	2D to 3D conversion	Small molecules 2D-to-3D
http://www.molecular-networks.com/online_demos/corina_demo.html	Corina: 2D to 3D conversion	Small molecules 2D-to-3D
http://www.eyesopen.com	Omega: 2D to 3D conversion	Small molecules 2D-to-3D
http://www.molsoft.com	ICM: 2D to 3D conversion	Small molecules 2D-to-3D
http://xdrawchem.sourceforge.net/	XDrawChem: Possible 2D to 3D conversion with BUILD3D	Small molecules 2D-to-3D
http://davapc1.bioch.dundee.ac.uk/programs/prodrg/	Possible 2D to 3D [237, 238]	Small molecules 2D-to-3D
http://bioserv.cbs.cnrs.fr/HTML_BIO/APPLET_ACD/create_molecule.html	2D to 3D with Corina	Small molecules 2D-to-3D
http://sourceforge.net/projects/easymol	EasyMol: A Java tool to design 2D molecules and render them in 3D	Small molecules 2D-to-3D
http://projects.villa-bosch.de/mcm/people/wang/3dfs_body.html	3DFS is a program to search 3D databases for compounds matching a pharmacophore query [239]	Compound searching
ftp2.ipc.pku.edu.cn	LigBuilder [240]: Based on the three-dimensional structure of the target protein, it can automatically build ligand molecules within the binding pocket	Compound searching
http://www.biopharmics.com/downloads.html	Surflex-Sim [241]: ligand-based method	Compound searching
http://www.cs.york.ac.uk/auramol/index.html	AURAmol allows a user to take a candidate 2D or 3D molecular shape and use it to search for similarly shaped molecules in large databases	Compound searching

Table 6a. Protein Data Bank, Receptor 3D Structures, Homology Modeling, 2D/3D Structure Prediction of the Receptor and Macromolecular Interaction Databases

URLs	Short summary	Keywords
http://www.rcsb.org/pdb	The Protein Data Bank [106]	Macromolecule database
http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml	The Entrez Structure Database [242]	Macromolecule database
http://www.ks.uiuc.edu/Development/MDTools/pdbcat/	PDBCat can be used to manipulate and process PDB files using commonly available tools such as Perl, awk, etc.	Tools to manipulate PDB files
http://www.biochem.ucl.ac.uk/bsm/sidechains/	Atlas of protein side-chain interactions within known protein structures and interactions with DNA	Macromolecule interaction database

(Table 6a) contd...

URLs	Short summary	Keywords
http://www.bind.ca/Action	BIND: The Biomolecular Interaction Network Database [243]	Macromolecule interaction database
http://dip.doe-mbi.ucla.edu/hold/main.html	DIP [244]: Database of interacting proteins	Macromolecule interaction database
http://mint.bio.uniroma2.it/mint/Welcome.do	MINT database [245] stores data on functional interactions between proteins	Macromolecule interaction database
http://www.ces.clemson.edu/compbio/protcom/	ProtCom [246]: a collection of protein-protein transient complexes	Macromolecule interaction database
http://www.expasy.org/	Many structural bioinformatics tools	Tools to analyze macromolecules
http://bioserv.rpbs.jussieu.fr/	Many tools for structural bioinformatics, loop prediction, simulation, small molecules [247]	Macromolecule interaction database
http://nihserver.mbi.ucla.edu/Verify_3D/	The UCLA-DOE Structure Evaluation server is a tool designed to help in the refinement of crystallographic structures and models [248]	Validate protein structure
http://nihserver.mbi.ucla.edu/ERRATv2/	ERRAT is a protein structure verification algorithm [249]	Validate protein structure
http://dunbrack.fccc.edu/SCWRL3.php	SCWRL3.0 is the most recent version of the SCWRL program for prediction of protein side-chain conformations [250]	Predict protein side chain conformation
http://bioserv.rpbs.jussieu.fr/SCit	SCit is a web server providing services for protein side chain conformation analysis and side chain positioning [251]	Predict protein side chain conformation
http://www.abo.fi/fak/mnf/bkf/research/johnson/bodil/about.php	Bodil [252]: biomolecular modeling package	Molecular modeling
http://www.sali.org/modeller/modeller.html	Modeller [253]: software package for homology or comparative modeling of protein 3D structures	Homology modeling
http://trantor.bioc.columbia.edu/programs/jackal/index.html http://honiglab.cpmc.columbia.edu/cgi-bin/jackal/nest.cgi	Jackal: protein structure modeling package [254]	Homology modeling & related tools
http://www.zbh.uni-hamburg.de/wurst/	Wurst [255]: web server for protein structure prediction with a structural scoring function, sequence profiles and optimized substitutions matrices	Protein structure prediction
http://bioinf.cs.ucl.ac.uk/psipred/psiform.html	PSIPRED [256]: web servers performing secondary structure prediction, transmembrane topology prediction or protein fold recognition	Protein structure prediction
http://www.predictprotein.org/newwebsite/	PredictProtein [257]: webserver for homology modeling and protein function prediction	Homology modeling & related tools
http://csbl.bmb.uga.edu/protein_pipeline/login.php	Prospect-PSPP [258]: automatic computational pipeline for protein structure prediction	Protein structure prediction

(Table 6a) contd...

URLs	Short summary	Keywords
http://robeta.bakerlab.org/	Robetta [259]: web server for protein structure prediction and analysis	Protein structure prediction
http://ligin.weizmann.ac.il/space/	SPACE [260]: suite of tools for protein structure prediction and analysis based on complementarity and environment	Protein structure prediction
http://trantor.bioc.columbia.edu/SMS/index_s.html	Java Protein Dossier [261]: web-based visualization tool including large collections of physicochemical parameters describing proteins' structure, stability, function and interaction with other macromolecules	Protein structure analysis
http://www.bch.msu.edu/labs/kuhn/web/software.html	Consolv: Tool to analyze protein-water interaction [262]	Protein structure analysis
http://schubert.bio.uniroma1.it/SCR_FIND/ http://schubert.bio.uniroma1.it/CHC_FIND/ http://schubert.bio.uniroma1.it/CAMPO/	CAMPO, SCR_FIND, CHC_FIND [263]: web tools to analyze evolutionary conserved residues, structurally conserved regions and conserved hydrophobic contacts	Protein structure analysis
http://bioinformatics.biol.uoa.gr/PRED-TMBB/	Pred-TMBB [264]: web server for predicting the topology of beta-barrel outer membrane proteins	Protein structure prediction
http://proteindbs.mnet.missouri.edu/PDBS_V2.php	ProteinDBS [265]: web server for detection of similar protein tertiary structures	Protein structure analysis, Structural similarity search
http://fatcat.burnham.org/	FATCAT [266]: web server for flexible structure comparison and structure similarity searching	Protein structure analysis, Structural similarity search
http://pre-s.protein.osaka-u.ac.jp/~prebi/	PreBi [267]: Server for predicting biological homo protein-protein interfaces in crystal protein structures	Protein structure analysis, Interface search
http://www.igs.cnrs-mrs.fr/Caspr2/index.cgi	CaspR [268]: web server for automated molecular replacement, method of choice for X-ray crystallography structure determination when structural homologues are available in PDB, using homology modeling	Protein structure determination
http://wishart.biology.ualberta.ca/shiftor/cgi-bin/predictor_current.py	PREDITOR [269]: Program for predicting dihedral angles from chemical shifts and/or sequential homology	Protein structure determination
http://www.ebgm.jussieu.fr/~gelly/index.html	Protein Peeling [270]: Tool for splitting a 3D protein structure into protein units which are an intermediate level of protein structure description between protein domains and secondary structures	Protein structure analysis
http://opaas.ibms.sinica.edu.tw	OPAAS [271]: web server for optimal, permuted and other alternative alignments of protein structures	Structural similarity search
http://localizome.org/	Localizome [272]: server for identifying transmembrane topologies and TM helices of eukaryotic proteins using domain information	Protein structure prediction
http://bioinformatics.univ-reunion.fr/PBE/	PBE [273]: platform for protein structure analysis using well defined library of short structural motifs (SSMs) known as structural alphabets	Protein structure prediction
http://bioserv.rpbs.jussieu.fr/SABBAC.html	SABBAC [274]: online structural alphabet-based protein backbone reconstruction from alpha-carbon trace	Protein structure prediction

(Table 6a) contd...

URLs	Short summary	Keywords
http://ps2.life.nctu.edu.tw/	(PS)2 [275]: automated homology modeling server using a consensus strategy between psi-blast, impala and T-coffee with a final 3D structure modeled with Modeller	Homology modeling & related tools
http://www.fiserlab.org/servers/archpred	ArchPRED [276]: template based loop structure prediction server	Loop structure prediction
http://bioinformatics.bc.edu/clotelab/transFold/	TransFold [277]: web server for predicting the structure and residue contacts of transmembrane beta-barrels	Protein structure prediction
http://caps.ncbs.res.in/harmony/	Harmony [278]: web server for the assessment of protein structures	Protein structure validation
http://rosettadesign.med.unc.edu/	RosettaDesign [279]: Server for identifying low energy amino acid sequences from the backbone coordinates of the target structure	Protein structure prediction
http://www-cryst.bioc.cam.ac.uk/coda/search_coda.html	CODA [280]: combined algorithm for predicting loops	Loop structure prediction
http://protiminer.csie.ntu.edu.tw/	ProteMiner [281]: software package that searches the PDB for proteins containing a substructure similar to the one specified by the user	Structural similarity search
http://molprobrity.biochem.duke.edu/	MolProbrity [282]: web server for structure validation and all-atoms contact analysis for nucleic acids and their complexes	Protein structure validation
http://astral.berkeley.edu/	ASTRAL [283]: web server providing databases and tools useful for analyzing protein structures and their sequences	Protein structure analysis
http://dd.stanford.edu/	Decoys 'R' us [284]: database of computer generated conformations of proteins sequences that possess some characteristics of native proteins	Database of side-chain, protein structure validation
http://www.sgc.ox.ac.uk/iSee/	iSee [285]: software package including the structural genomics workflow into one file, from DNA to protein structure, using the Molsoft ICM-browser technology	Protein structure analysis, annotation
http://bioserv.cbs.cnrs.fr/	Automatic threading, optimization modeling and evaluation, homology modeling [286]	Protein structure prediction

Table 6b. Pocket Prediction and Search for Functional Regions on Targets, Analysis of Interfaces

URLs	Short summary	Keywords
http://www.bioinformatics.leeds.ac.uk/qsitefinder http://www.bioinformatics.leeds.ac.uk/pocketfinder	Server to predict binding sites, Q-site and pocketfinder [287]	Pocket detection
http://sts.bioengr.uic.edu/castp/	CASTp [288]: Server to predict binding site - Binding sites and active sites of proteins and DNAs are often associated with structural pockets and cavities	Binding site prediction
http://interface.bioc.columbia.edu/screen	SCREEN: Server to predict binding site [289]	Binding site prediction
http://medock.csie.ntu.edu.tw/	MEDock: Online tool to define binding site [290]	Binding site prediction
http://consurf.tau.ac.il	ConSurf [291] identifies functional regions in proteins	Binding site prediction

(Table 6b) contd....

URLs	Short summary	Keywords
http://www.tau.ac.il/~itaymay/cp/rate4site.html	Rate4Site [292]: an algorithmic tool for the identification of functional regions on proteins by surface mapping of evolutionary determinants within their homologues	Binding site prediction
http://www.ccl.net/cca/software/UNIX/pass/overview.shtml	PASS [293]: Pocket detection method based upon the size, shape of buried volumes	Pocket detection
http://xray.bmc.uu.se/usf/voidoo.html	Voidoo [294]: Pocket detection tool	Pocket detection
http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html	SurfNet [295]: Pocket detection tool	Pocket detection
http://vakser.bioinformatics.ku.edu/resources/gramm/gramm1/ http://vakser.bioinformatics.ku.edu/resources/gramm/grammx	Gramm: Tools for protein-protein docking GrammX [296]: web interface of Gramm	Macromolecular docking
http://bombyx.inria.fr/Intervor/intervor.html	Intervor: Tools to analyze interfaces	Interface analysis
http://projects.villa-bosch.de/mcm/software/molsurfer	MolSurfer [297]: a macromolecular interface navigator	Interface analysis
http://swift.cmbi.kun.nl/swift/ligin/	LIGIN: Molecular docking using surface complementarity. The LIGIN program [298] is also available as part of the WHATIF software package.	Macromolecular docking
http://bioinfo3d.cs.tau.ac.il/	Protein docking tools (PatchDock) and related [299, 300]. PatchDock, webserver for macromolecules and small molecules docking based on shape complementarity criteria	Macromolecular docking
http://mmb.pcb.ub.es/PyDock/	PyDock: tool for protein-protein docking [301]	Macromolecular docking
http://www-cryst.bioc.cam.ac.uk/~viji/docking/	Crescendo [301]: functional site prediction by the detection of protein-protein interaction sites	Binding site prediction
http://i.moltalk.org/	iMolTalk [302]: On-line tools including detection of the interface between two chains of a structure	Interface analysis
http://www.bmm.icnet.uk/docking/	3D-Dock [303]: software package for rigid-body protein-protein docking using shape complementarity, electrostatics, biochemical information, residue level pair potential score and a refinement tool	Macromolecular docking
http://nrc.bu.edu/cluster/	ClusPro [304]: protein-protein docking webserver using 3 docking programs - DOT [305], ZDOCK [306], GRAMM [296]	Macromolecular docking
http://www.sdsc.edu/CCMS/DOT/	DOT [305]: protein-protein docking software with a surface grid applied to the fixed and the moved protein. Further refinement is performed via energy minimization and molecular dynamics	Macromolecular docking
http://zlab.bu.edu/zdock/index.shtml	ZDOCK [306]: protein-protein docking software evaluating based on shape complementarity, desolvation energy and electrostatics. Best predictions from ZDOCK are given to RDOCK where they are minimized by CHARMM	Macromolecular docking
http://graylab.jhu.edu/docking/rosetta/	Protein-protein docking package [307]	Macromolecular docking
http://www.biotec.tu-dresden.de/~bhuang/bdock/bdock.html	BDOCK: protein-protein docking software integrating the degree of burial of surface residues into protein-protein docking	Macromolecular docking
http://www.weizmann.ac.il/Chemical_Research_Support/molfit/home.html	MolFit [308]: protein-protein docking software estimating the extent of geometric and chemical surface complementarity	Macromolecular docking

(Table 6b) contd....

URLs	Short summary	Keywords
http://www.csd.abdn.ac.uk/hex/	Hex [309]: protein-protein docking and molecular superposition program	Macromolecular docking
http://www.ddl.unimi.it/escherng/index.htm	ESCHER-NG [310]: protein-protein and DNA-protein docking software	Macromolecular docking
http://www.bmm.icnet.uk/docking/	FTDock (Fourier Transform Dock) performs rigid-body docking on two biomolecules in order to predict their correct binding geometry[308]	Macromolecular docking
http://www.moldiscovery.com/soft_grid.php	GRID [311]: Tool for analysis of binding sites	Binding site prediction & analysis
http://www.ccdc.cam.ac.uk/products/life_sciences/superstar	SuperStar [312]: Tool for analysis of binding sites	Binding site prediction & analysis
http://www.bioinformatics.leeds.ac.uk/sb/	SitesBase [120]: Tool for analysis of binding sites [119]	Binding site analysis
http://gordion.hpc.eng.ku.edu.tr/prism	Prism [313] predicts and analyzes putative protein-protein interaction sites.	Binding site prediction & analysis
http://bioinfo3d.cs.tau.ac.il/SiteEngine/ http://bioinfo3d.cs.tau.ac.il/I2I-SiteEngine/	SiteEngines [314]: recognition and comparison of binding sites and protein-protein interfaces	Binding site prediction & analysis
http://structure.pitt.edu/servers/fastcontact/	FastContact [315]: a free energy scoring tool for protein-protein complex structures	Macromolecular scoring

Table 7. Comparison of Binding Sites/Protein Functional Sites – Protein Function Prediction (see also Table 6b)

URLs	Short summary	Keywords
http://smid.blueprint.org	Annotated binding sites	Binding site analysis
http://www.mgs.bionet.nsc.ru/mgs/gnw/pdbsitescan/	PDBSiteScan automatically performs the best superposition of sites from PDBSite with the 3D structure of a protein under study	Binding site comparison
http://www.bioinformatics.leeds.ac.uk/sb/	SiteBase [120]: compare nucleotide and ligand binding site	Binding site comparison
http://www.russell.embl.de/pints/	PINTS [316]: detection of similarities between protein structures consisting of amino acids that are close in space	Structural similarity search
http://sumo-pbil.ibcp.fr/	Tools for analysis of binding sites and to find similar motifs based on a search in the PDB [121]	Structural similarity search
http://portray.bmc.uu.se/cgi-bin/spasm/scripts/spasm.pl	SPASM [317]: detection of similar motifs based on a search in the PDB	Structural similarity search
http://www.ebi.ac.uk/thornton-srv/databases/CSA/	Catalytic site atlas [318]: find similar catalytic sites	Structural similarity search
http://pvsoar.bioengr.uic.edu/	pvSoar [319]: detection of a protein surface pattern derived from a pocket or a void against all known surface patterns from the CASTp database	Structural similarity search, Binding site comparison

(Table 7) contd...

URLs	Short summary	Keywords
http://bioinfo-pharma.u-strasbg.fr/scPDB/	Sc-PDB: tool for analysis of binding sites [123]	Binding site analysis
http://pdbfun.uniroma2.it/	Pdbfun [320]: web server for the identification of local structural similarities between annotated residues in proteins	Structural similarity search
http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/simdock.php	Ligand transposition server [321]	Binding site comparison
http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/fragamar.php	Docking with substructure query in a protein family	Binding site comparison
http://bioinfo3d.cs.tau.ac.il/SiteEngine/	SiteEngine [314] recognizes regions on the surface of one protein that resemble a specific binding site of another	Structural similarity search, Binding site comparison
http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/	The ProFunc server [322] had been developed to help identify the likely biochemical function of a protein from its three-dimensional structure	Binding site prediction
http://bionmr-c1.unl.edu/	CPASS [314]: website for active site comparison	Binding site comparison
http://ef-site.hgc.jp/eF-site	eF-Site [323]: Electrostatic surface of functional sites	Electrostatic computations, binding site search
http://projects.villa-bosch.de/dbase/ps2/	proSAT2 [324]: Features for visualizing SwissProt and PROSITE functional annotations by mapping of information on variants and mutations from the UniProt KnowledgeBase and the BRENDA enzyme information system onto protein structures	Functional site visualization
http://www.cbs.dtu.dk/services/FeatureMap3D/	FeatureMap3D [325]: tool mapping protein features such as post-translational modifications, protease cleavage sites or exonic structure onto 3D structures of homologous proteins	Functional site visualization
http://prokware.mbc.nctu.edu.tw/	ProKware [326]: Integrated system containing interactive graphic interface and abundant protein property annotations at the structural level and domain-domain interaction in protein 3D structures	Functional site visualization
http://protomot.csie.ntu.edu.tw/step1.cgi	Protomot [327]: server that carries out prediction of protein binding sites based on the structural templates automatically extracted from the PDB crystals	Binding site prediction

Table 8. Target Analysis: Flexibility, Energy Minimization, Normal Modes, Molecular Dynamics, Water Molecules in Targets, Ions, pKa and Electrostatics, Point Mutations and Related Utilities

URLs	Short summary	Keywords
http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi	I-Mutant2.0 : predicting stability changes upon mutation from the protein sequence or structure	Analysis of mutations
http://sride.enzim.hu	SRide [328]: server for identifying stabilizing residues in proteins	Analysis of mutations
http://www.mutdb.org/	MutDB services [329]: interactive structural analysis of mutation data	Analysis of mutations

(Table 8) contd....

URLs	Short summary	Keywords
http://foldx.embl.de/ http://fold-x.embl-heidelberg.de:1100/cgi-bin/main.cgi	FoldX: empirical force field that was developed for the rapid evaluation of the effect of mutations on the stability, folding and dynamics of proteins and nucleic acids [330, 331]	Analysis of mutations
http://www.es.embnet.org/Services/MolBio/tinker/	Tinker online	Molecular modeling & graphics, simulations
http://www.cs.gsu.edu/~cscrwh/amp/amp.html	AMMP is a modern full-featured molecular mechanics, dynamics and modeling program. It can manipulate both small molecules and macromolecules including proteins, nucleic acids and other polymers	Molecular modeling & graphics, simulations
http://homepages.nyu.edu/~mt33/PINY_MD/PINY.html	PINY_MD is capable of performing a wide variety of molecular dynamics, electronic structure, and geometry optimization calculations. Such capabilities include force-field based simulations on system ranging in complexity from simple molecular liquids and crystals to large biomolecular systems	Molecular modeling & graphics, simulations
http://dasher.wustl.edu/tinker/	Macromolecules and small molecules modeling	Molecular modeling & graphics, simulations
http://blocks.fhrc.org/sift/SIFT.html	SIFT [332]: analyzes protein point mutations	Analysis of mutations
www.bork.embl-heidelberg.de/PolyPhen/	PolyPhen [333]: analyzes protein point mutations	Analysis of mutations
http://www.bioinfo.no/tools/normalmodes	WEBnm@: a web application for normal mode analysis of proteins [334] to investigate for large amplitude movements. Normal modes are performed with MMTK [335].	Molecular simulations
http://bioserv.rpbs.jussieu.fr/PCE	Online tools to compute protein electrostatics based on MEAD [336]: PCE (Protein Continuum Electrostatics) [337]	Electrostatics
http://biophysics.cs.vt.edu/H++	H++ [338]: a server for estimating pKas and adding missing hydrogens to macromolecules	Electrostatics
http://propka.chem.uiowa.edu/	PropKa [339]: fast empirical method to predict pKas in proteins	Electrostatics
http://sourceforge.net/projects/apbs	APBS [340]: software package for the numerical solution of the Poisson-Boltzmann equation	Electrostatics
http://sourceforge.net/projects/oops-pl	Open Protein Simulator (OOPS) is a program designed to serve as a test bed for different algorithms for protein folding, dynamics and structure prediction	Molecular simulations
http://somosierra.cnb.uam.es/wwwPDG/Software/software.php	Tools to analyze protein structures	Molecular modeling, structural analysis
http://www.ytbl.york.ac.uk/~tom/structure.html	Web resources for protein structure analysis, links to numerous web sites	Molecular modeling, structural analysis
http://swift.cmbi.ru.nl/whatif/	WhatIf [341]: versatile molecular modeling package	Molecular modeling, structural analysis

(Table 8) contd....

URLs	Short summary	Keywords
http://igs-server.cnrs-mrs.fr/elnemo/	elNémo is the Web-interface to The Elastic Network Model [342], a fast and simple tool to compute the low frequency normal modes of a protein	Molecular simulations
http://honiglab.cpmc.columbia.edu/cgi-bin/GRASS/surfserv_enter.cgi	GRASS [343], Graphical Representation and Analysis of Structure Server	Molecular surface analysis
http://cubic.bioc.columbia.edu/services/profbval/	PROFBval [344]: method for predicting residue mobility based on amino-acid sequence. Identification of extremely rigid or flexible residues on the protein surface is helpful for identifying functionally important residues in proteins. A common measure of atom mobility in proteins is B-value data from x-ray crystallography structures. PROFbval is the first web server to predict normalized backbone B-values from amino-acid sequence	Protein flexibility prediction
http://www.umass.edu/microbio/chime/find-ncb/	Non covalent bond finder	Structural analysis
http://kinemage.biochem.duke.edu/~jsr/html/anatax.2a.html	Structural analysis of proteins	Structural analysis
http://www.bioinfo.weizmann.ac.il/mutaprot	MutaProt [345]: Tool for structural analysis of point mutations: MutaProt.	Analysis of mutations
http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi	SVMProt [346]: protein functional family prediction	Structural analysis
http://www.ks.uiuc.edu/Development/MDTools/sodium/	This program places the required number of sodium ions around a system of electric charges, e.g., the atoms of a biological macromolecule (protein, DNA, protein/DNA complex)	Tools to prepare molecules for simulations
http://www.ivec.org/GULP/	GULP is a program for performing a variety of types of simulation on materials using boundary conditions of 0-D (molecules and clusters), 1-D (polymers), 2-D (surfaces, slabs and grain boundaries), or 3-D (periodic solids) [347]	Molecular simulations
http://sourceforge.net/projects/btk	The Biomolecule Toolkit is a library for modeling biological macromolecules such as proteins, DNA and RNA. It provides a C++ interface for common tasks in structural biology to facilitate the development of molecular modeling, design and analysis tools	Molecular modeling
http://monster.northwestern.edu/monster.jsp	Monster [348]: web application for inferring potentially stabilizing non-bonding interactions in macromolecular structures	Structural analysis, mutations
http://dis.embl.de/	DisEMBL [349]: computational tool for prediction of disordered/unstructured regions within a protein sequence.	Structural analysis, mutations
http://bioinf.cs.ucl.ac.uk/disopred/disopred.html	Disopred2 [350]: Prediction and functional analysis of native disorder in proteins	Structural analysis, mutations
http://www.bioinf.org.uk/software/avp/index.html	AVP (Another Void Program) is a new method for the analysis of voids in proteins and packing quality in a single united program	Cavity search
http://rumour.biology.gatech.edu/YamppWeb/	YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models (Python)	Molecular simulations
http://ignm.cccb.pitt.edu/GNM_Online_Calculation.htm	oGNM [351]: calculating the equilibrium dynamics of any structure submitted in PDB format, using the Gaussian network Model (GNM)	Molecular simulations, flexibility

(Table 8) contd....

URLs	Short summary	Keywords
http://sourceforge.net/projects/protomol	ProtoMol: object-oriented component based framework for molecular dynamics simulations	Molecular simulations
http://lorenz.immstr.pasteur.fr/pdb_hydro.php	PDB_Hydro [352]: Tools for mutating and solvating protein structures	Tools to prepare molecules for simulations
http://pheps.orgchm.bas.bg/home.html	PHEPS [353]: fast pH-dependent electrostatic calculations for proteins	Electrostatics
http://www.jenner.ac.uk/PPD	PPD [354]: an integrated, web-accessible database of experimentally determined protein pKa values	Electrostatics
http://polymerase.ucd.ie/cgi-bin/pKa_Design/server_start.cgi	pKD [355] : re-designing protein pKa values for a set of point mutations	Electrostatics, mutations
http://lorenz.immstr.pasteur.fr/nomad-ref.php	NOMAD-Ref [356]: Tools using normal modes for structural refinement of large proteins	Molecular simulations
http://biomechanics.ecs.umass.edu/umms.html	UMMS [357]: Tool using normal mode to analyze the harmonic behaviors (fluctuations) of a macromolecule around its equilibrium and elastic network interpolation to generate the anharmonic pathways for conformational transitions of two metastable conformations of the same macromolecule	Molecular simulations, flexibility
http://gibk26.bse.kyutech.ac.jp/jouhou/readout/	Readout [358]: structure-based calculation of direct and indirect readout energies and specificities for protein-DNA recognition	Analysis of protein-DNA complex and DNA structure
http://cupsat.uni-koeln.de/	CUPSAT [359]: server for prediction of protein stability upon point mutations by assessment of the difference in free energy of unfolding between wild-type and mutant proteins using structural environment specific atom potentials and torsion angle potential	Analysis of mutations
http://mmb.pcb.ub.es/FSolv/	FSolv: fast method for the determination of fractional contributions to solvation in proteins	Solvation
http://mmb.pcb.ub.es/PMut/	PMut [360]: predicts the pathologic character of a punctual mutation in a protein	Analysis of mutations
http://www.netasa.org/qgrid/index.html	Qgrid [361]: webserver for detection of charged and hydrophobic clusters in proteins	Structural analysis
http://cic.cs.wustl.edu/RNA/	ILM [362]: web server combining two algorithms, iterated loop matching and maximum weighted matching, for predicting RNA secondary structures	RNA structure prediction
http://rna.cbi.pku.edu.cn	RDfolder [363]: webserver for prediction of RNA secondary structure from two methods, random stacking of helical regions and helical regions distribution	RNA structure prediction
http://www.ks.uiuc.edu/Research/namd/	NAMD [364]: molecular dynamics package for simulation of large biomolecular systems	Molecular simulations
http://www.ibpc.fr/UPR9080/Curindex.html	CURVES: software package for calculating a helical parameter description for any irregular nucleic acid segment with respect to an optimal, global helical axis	Nucleic acid analysis
http://www-ibmc.u-strasbg.fr/upr9002/westhof/	MANIP [365]: interactive tool for modeling of RNA structure	Nucleic acid analysis

(Table 8) contd....

URLs	Short summary	Keywords
http://www.scsb.utmb.edu/fantom/	FANTOM (Fast Newton-Raphson Torsion Angle Minimizer) calculates low-energy conformations of polypeptides and proteins, compatible with distance and dihedral angle constraints obtained typically from NMR experiments. Protein-solvent interaction is included with a fast routine GETAREA for the calculation of accessible surface areas of individual atoms and their gradients. FANTOM is suited for the exploration of low energy conformations of cyclic peptides and of flexible loops in proteins as well. In addition to the above uses, with the newly added program EXDIS, FANTOM is an efficient tool for homology modeling of proteins [366, 367]	Molecular simulations
http://www.charmm.org/	CHARMM (Chemistry at HARvard Molecular Mechanics) is a program for macromolecular simulations	Molecular simulations
http://www.scripps.edu/mb/case/Biomer/	B is a Java-based, on-line biomolecular modeling package	Molecular modeling
http://www.gromacs.org/	GROMACS [368] is a package for performing standard MD simulations, energy minimizations, NMR refinement...	Molecular simulations
http://mmb.pcb.ub.es/MODEL/	Molecular Dynamics Extended Library (Database of Molecular Dynamics Trajectories)	Molecular simulation database
http://eds.bmc.uu.se/	The Uppsala Electron Density Server [369]	Protein X-ray structural analysis
http://www.biocomp.chem.uw.edu.pl/services/BioShell/about.html	BioShell is a suite of programs designed for pre- and post-processing in protein structure modeling protocols	Tools to prepare molecules for simulations

search/evolutionary algorithms, they make random changes on some variables and usually require multiple independent runs), and deterministic searches or simulation methods (e.g., energy minimization and molecular dynamics) [124]. Some VLS packages use more than one of these approaches. Generating a broad range of binding modes is ineffective without a model to rank each conformation that is both accurate and efficient/fast [125]. The scoring functions commonly used are said to be: force-field based, empirical and knowledge-based [124]. Because the process of ranking is highly complex, a myriad of scoring functions has been developed during the past few years, many of them have been compared and analyzed in depths [6, 9, 126-131]. Some authors suggest the use of ligand efficiency to select hits instead of relying only on scores [132, 133]. The concepts of consensus scoring (see for instance [134, 135]) or consensus docking-consensus scoring [136] have also been discussed. A few scoring methods are listed in (Table 9). In this table, several SB-VLS engines are also reported. Many VLS tools have been compared and interested users can find additional data about the packages in the following articles [36, 124, 137-145]. The main methods freely available for non-profit organizations are listed below. Yet, prior to briefly describe these programs, it is important to add that finding hits is always difficult and that in most cases it is important to combine several methods such as in silico screening (ligand-

based and/or structure-based) and experimental HTS (Fig. 1). Several tools to perform ligand-base/similarity searches have been added in this review (Table 5) although they do not exactly belong to SB-VLS experiments as defined in the introduction (see reviews [30, 96, 146-148] for further details about de novo ligand design and ligand-based VLS). Also, it should be emphasized that multi-step procedures are often used to facilitate the in silico screening process, they can combine different packages, use molecular dynamics after docking, among others [136, 149-154]. More complex methods to evaluate protein-ligand interaction energy are under development as recently reviewed by Gresh [155].

DOCK [27, 156, 157]

DOCK was one of the first docking programs to be developed for structure-based drug design. The general approach of this method is divided into three main steps. First, the determination of a set of overlapping spheres in contact with the surface of the receptor site. These spheres fill the molecular surface of the binding site and represent a negative image of the target site. Second, the center of these spheres is matched with the ligand atoms via the use of a graph-matching algorithm. Third, a scoring function is used to evaluate the pertinence of the docking poses by approximating the protein/ligand binding energy. The evaluation of the ligand orientation uses a grid-based procedure in which steric

Table 9. Docking and/or Scoring Engines for Small Molecule-Macromolecule Interactions

URLs	Short summary	Keywords
http://www.scripps.edu/mb/olson/doc/autodock	AutoDock [370], small molecule docking	Small molecule docking
http://www.quimica.urv.cat/~pujadas/BDT/	BDT (is an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock [160])	Graphics interface for small molecule docking
http://fulcrum.physbio.mssm.edu/~mezei/dockres/	Dockres (reads the log file of docking runs performed by Autodock (version 3.0.5) and extracts the top scoring poses. The extraction can be subject to various filters (e.g., the residue nearest to the ligand...). The program also calculates distributions of various properties of the ligand set (e.g., molecular weight, number of hydrogen bond donors) and the distribution of docking sites as well as the distribution of docking free energies per target residue)	AutoDock post-docking processing
http://www.simbiosys.ca/ehits/index.html	eHits [165]: Small molecule docking	Small molecule docking
http://www.eyesopen.com	FRED [162]: small molecule docking. See the online demos to try many OpenEye applications	Small molecule docking
http://dock.compbio.ucsf.edu/	DOCK [27, 156, 157]: Small molecule docking.	Small molecule docking
http://www.cgl.ucsf.edu/chimera/docs/UsersGuide/index.html	ViewDock to analyze Dock data and tools for post-processing DOCK results [371]	Small molecule docking post-docking processing
http://www.biopharmics.com	Surflex [163]: Small molecule docking	Small molecule docking
http://www.tcd.uni-konstanz.de/research/plants.php	Plants [158]: Small molecule docking	Small molecule docking
ftp://ftp2.ipc.pku.edu.cn/pub/software	PSI-DOCK [166]: Small molecule docking.	Small molecule docking
http://ang.cz3.nus.edu.sg/cgi-bin/prog/rune.pl	PEARLS [372]: Program for Energetic Analysis of Receptor-Ligand System	Scoring
http://gfscore.cnrs-mrs.fr/index.htm	GFscore [373]: A General Non-Linear Consensus Scoring Function for High-Throughput Docking	Scoring
ftp://ftp2.ipc.pku.edu.cn/	SCORE [167] is an empirical method developed for estimating the binding affinity of protein-ligand complex with known three-dimensional structure	Scoring
http://pc1664.pharmazie.uni-marburg.de/drugscore/	DrugScore [374]: evaluate ligand-receptor interaction energy	Scoring
http://fold-x.embl-heidelberg.de:1100/cgi-bin/main.cgi	FOLD-X can compute binding energy [331]	Scoring
http://bidd.nus.edu.sg/group/CLiBE/CLiBE.asp	CLiBE [375]: Computed ligand binding energy	Help for scoring functions
http://sw16.im.med.umich.edu/software/xtool/	Xscore [376]: tool to predict binding energy between ligand and receptor	Scoring
http://www.cgal.org/index.html	The goal of the CGAL Open Source Project is to provide easy access to efficient and reliable geometric algorithms to users in industry and academia in the form of a C++ library	Help for docking

(Table 9) contd....

URLs	Short summary	Keywords
http://www.openmolgrid.org	OpenMolGrid: tools to speed-up computations	Tools to speed-up computations
http://www.dddc.ac.cn/tarfisdock/	tarFisDock [377] docks ligands into the proteins targets in PDTD (Potential Drug Target Database), and outputs the top 2%, 5% or 10% candidates ranked by the energy score, including their binding conformations and a table of the related target information	Small molecule docking
http://abcis.cbs.cnrs.fr/kindock/ http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/simdock.php	Kindock and Simdock [321]: Tool for comparative docking of protein kinase ligands and ligand transposition server	Small molecule positioning

and electrostatic interactions between the putative ligand and the receptor are pre-computed at each grid point. Several approaches can be actually used at this stage: contact score, energy score (Lennard-Jones van der Waals potential and Coulombic electrostatics with distance-dependent dielectric constant), Generalized Born/Surface Area (GB/SA) score (implemented in DOCK5) and chemical score.

Plants [158]

The docking algorithm PLANTS is based on a class of stochastic optimization algorithms called ant colony optimization (ACO). ACO is inspired by the behavior of real ants finding a shortest path between their nest and a food source. In Plants, the ligand is flexible (automatic rotatable bond identification, atom typing...) and the flexibility of the protein is partially considered by the optimization of the positions of hydrogen atoms (or some other groups) that could be involved in hydrogen bonding. Several scoring functions have been implemented and the package has also a rescoring capability. Several new functionalities are presently implemented, including rigid docking and flexibility of amino acid side chains.

AutoDock [159]

This program employs a Lamarckian genetic algorithm (LGA) but also encompasses a Monte Carlo simulated annealing procedure and a traditional genetic algorithm, although it seems that the LGA approach is more efficient and reliable. The scoring function of AutoDock3.0 is modeled after the AMBER force-field, and uses a pairwise sum of energetic terms with parameters for van der Waals, hydrogen bonding and distance-dependent dielectric electrostatics, as well as conformational torsional restriction entropy and empirical solvation terms. An easy-to-use front-end application to run Autodock has been recently reported and should indeed help the use of this program for screening large databases [160]. The new version of AutoDock (4.0) also includes receptor side chain flexibility. An automated version of AutoDock has been assessed for virtual ligand screening and has been found very efficient as compared to other tools [161].

FRED [162]

This method performs docking of rigid ligands into a rigid receptor site. It exhaustively docks compounds into the binding pocket (by rigidly rotating and translating each conformer, this strategy completely avoids the sampling issues associated with stochastic methods). Then FRED filters the pose ensemble by rejecting the ones that clash with the protein using a negative image of the active site. All remaining poses are ranked with one or several scoring functions. Optionally, FRED can perform a systematic solid body optimization of the top ranked poses. A full atom optimization of the pose via the MMFF force-field can be performed. The refined poses can then be scored using one or more scoring functions, including MASC (Multiple Active Site Correction) corrected scoring functions [130]. Among the many features of FRED, it is important to note the use of an extremely fast Gaussian based scoring function that evaluates the surface complementarity between the receptor and the ligands

Surflex [163]

Surflex is based on a previously developed program named Hammerhead [164]. It uses the same concept of pocket finder and binding site-probing definition (protomols) but it is characterized by an innovative incremental construction of the ligand and recently refined scoring function [131]. The following describes the overall procedure and the two main phases of the algorithm. The program first creates an idealized binding site (called protomol) that serves as a target to which putative ligands or ligand fragments are aligned on the basis of molecular similarity. Ligands are docked into the protein to optimize the value of the scoring function. Each putative ligand is fragmented, each of which may have some rotatable bonds. Each fragment is then conformationally searched and each conformation of each fragment is aligned to the protomol to yield poses that maximize molecular similarity to the protomol. The aligned fragments are scored and pruned on the basis of the scoring function and the degree of protein interpenetration. Two procedures can be used to construct the full ligand from the aligned fragments (incremental or whole molecule approach). The best scoring poses are subjected to gradient-based optimization of conformation and alignment, and the top scoring poses are returned along

with their scores. The poses can be post-processed at a later stage using user-defined parameters. The scoring function terms involve, in rough order of significance, hydrophobic complementarity, polar complementarity, entropic terms, and solvation terms.

eHits [165]

eHITS performs fast flexible ligand docking. A systematic algorithm is used in eHITS with no random, stochastic or evolutionary element. eHITS provides a comprehensive search space coverage. The eHITS system generates all major docking modes that are compatible with the steric and chemistry constraints of the target cavity for each candidate structure. The output consists of multiple sets of 3D coordinates per structure with coarse fitness scores that are highly configurable. A user-defined binding pocket or the entire receptor surface can be searched. The docking program eHITS takes a unique approach to the scoring problem. eHITS uses eHITS_Score, a statistically derived empirical scoring function, with many novel features, including consideration of the temperature factors of crystal structures. In addition to a generic, default weight set, eHITS_Score applies an automated receptor family clustering method which is used to create sets of family-specific scoring function weights which better represents the investigated receptor. eHITS 6.1 now has a very efficient and accurate pre-screening tool for virtual high throughput screening of large databases. In addition, eHITS can evaluate all possible protonation states of the ligand and receptor in a single run.

PSI-DOCK [166]

PSI-DOCK is a new VLS package that makes use of a unique two-step searching/docking strategy. In the first step, it quickly explores the possible binding poses of the ligands using a Tabu-enhanced genetic algorithm with a rapid shape complementarity scoring function. Then, the computations continue with further refinements of the poses obtained in step one using competition genetic algorithm with an accurate scoring function. This one is a modified version of SCORE 3.0 that better accounts for solvation free energy (SCORE is an empirical scoring function to estimate binding free energy of a protein–ligand complex with known 3D structure [167]).

VI. SUCCESS STORIES WITH FREE SB-VLS PACKAGES

VLS methods based on the 3D structure of the receptor have been shown to be useful in prioritizing large libraries and influential for drug design projects despite present technical and theoretical limitations (see recent reviews [1, 2, 14, 16, 25, 39, 99, 168]). In fact, in silico screening approaches can be much more effective than HTS (see for instance [169]) in spite of some present limitations. Many hits have been identified using free (for non-profit institutions) VLS packages on various targets. The receptor structure was in general obtained by X-ray crystallography or predicted by comparative model building, the binding pocket was either a catalytic site or a pocket located at a protein-protein interaction site, the active compounds were identified after screening compound collections of different kinds/sizes.

Many new hits were successfully identified with DOCK. For example, novel L-xylulose reductase inhibitors with improved activity and specificity were recently discovered by screening 249071 compounds of the NCI Database with DOCK [170]. Novel and potent inhibitors of SARS-CoV proteinase were found after DOCK VLS experiments [171, 172] (and AutoDock)[173]. The 3D structure of the toxin adenylyl cyclase edema factor (EF) served to dock 205,226 ACD compounds into the catalytic site using DOCK. From 24 tested compounds two pyrazoloquinazolines could be identified as selective inhibitors of the toxins EF and CyaA [174]. Recently, a series of novel small molecular inhibitors of Cyclophilin A have been discovered with DOCK and the SPECS database (280,000 small molecules) [175]. In the search for novel inhibitors of Plasmodium falciparum dihydrofolate reductase (PfDHFR), several filters were first used to reduce the number of potential candidates in a database of 230,000 ACD compounds to 4,061 molecules. Docking of this “focused” library was carried out with DOCK and the authors could identify 12 compounds that are structurally unrelated to known anti-folates [176]. A database of 200,000 compounds (ChemDiv) was docked into the binding site of bcr-abl tyrosine kinase using DOCK. 15 compounds were selected for biological testing; eight of these compounds inhibited K562 tumor cell growth with IC50 values between 10 and 200 microM [177]. About 250,000 molecules were screened with DOCK against the sorbitol dehydrogenase binding site and 7 compounds were found to inhibit the enzyme with affinities in the low micromolar range, thus opening new avenues for the treatment of diabetes complications [178]. VLS applications of DOCK to find new ligands inhibiting protein-protein interactions have also been published. Screening with DOCK and with a large database was applied to identify small nonpeptidic compounds that block lymphoid T cell tyrosine kinase SH2 domain-dependent interactions [179]. A small nonpeptidic ligand of a Bcl-2 surface pocket was discovered by using a screening strategy by DOCK based on the predicted structure of Bcl-2 protein [180]. Highly specific ligands of fibrin were designed by screening of a wide virtual library of oligo-peptides and analogues by means of a docking/scoring approaches with DOCK [181]. Many success stories of homology modeling and VLS by DOCK have been published (see above). The study of Vangrevelinghe *et al.* further illustrates this point. They discovered novel and selective inhibitors for protein Casein Kinase II by screening a subset of 400,000 molecules of the Novartis database with DOCK on a homology model of CK2 [182]. The best one had an IC50 of 80 nM. A 3D homology model of the eukaryotic Shaker K⁺ channel was used to dock more than 50,000 compounds of the China Natural Product Database with DOCK into the extracellular TEA binding site. Extracellular application of four compounds inhibited the delayed rectifier current (IK) at micromolar concentration [183]. Small molecule inhibitors of integrin alpha-v-beta3 have recently been identified using DOCK (compound collection of about 88,695 molecules and IC50 values ranging from 30 to 200 microM) [184].

Successful applications of AutoDock for identification of new hits on different targets have also been reported in the literature. Recently Rogers *et al.* [185] discovered several novel inhibitors of protein phosphatase 2C activity using

VLS on the NCI Diversity Set with AutoDock. Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase against the NCI Diversity Set with AutoDock was performed to identify novel nonfolate inhibitors [186]. The 19 found inhibitors (the best one had $K_i = 154$ nM) served as novel templates/scaffolds for development of more potent and specific non-folate-based AICAR transformylase inhibitors.

Numerous examples of drug design and protein-ligand interaction studies guided by docking experiments can be found in the literature. Recently docking studies by AutoDock revealed nine hexose-binding clusters on the glucose transporter protein GLUT1 [187]. Binding and reactivity of peroxidases to indole derivatives have also been studied by molecular docking with AutoDock [188]. Docking experiments with AutoDock on a series of small molecules blocking protein-protein association of p53 and CREB binding protein (CBP) have been performed in order to provide insights into the structural basis of protein/ligand recognition [189]. These novel inhibitors have been discovered in target structure-guided NMR spectroscopy screening of a small focused chemical library constructed based on the structural knowledge of CBP-p53 binding. Recently, the newly developed program eHiTS has been applied to dock brequinar (quinoline carboxylic acid) derivatives for identification of species-specific inhibitors of dihydroorotate dehydrogenase [190]. In our group, we successfully identified new hits on several different targets using DOCK or FRED and/or Surflex (manuscripts in preparation).

CONCLUSION

Computational methods pervade all aspects of drug discovery research. We have listed and commented many free tools that facilitate the drug discovery process, ranging from homology modeling to protein docking and virtual screening. With regard to SB-VLS methods, it is however important to note that many free tools are still missing in the field. Concerning compound collection preparations, free tools to predict the 3D structure of the small molecules starting from SDF files or SMILES strings are needed. Free computer programs to generate multi-conformers starting from a single 3D structure would also be valuable. ADME/Tox methods to better assess toxicity would also be important. Finally, it would be valuable to have more open-source packages for docking and scoring (see comments about potential pitfalls with open-source software in [191]). National and International projects toward these aims should be actively supported. There is indeed a real need to further assess SB-VLS techniques, to develop new protocols and to invent new approaches/concepts. In fact, as with many other approaches in Science, SB-VLS methods have strengths and weaknesses (i.e., problems with docking, scoring, generation of correct ligand 3D structures, flexibility of ligands and receptors...) that need to be considered (see for instance the following articles [1, 6, 9, 13, 17, 19, 155, 192, 193]). Yet, present SB-VLS methods can really contribute to the discovery of new hits and many success stories have already been reported. SB-VLS and ligand-based approaches can also be used for other applications, for instance, small molecules can stabilize some proteins and thus increase the likelihood of crystallization, as such, virtual screening can assist Structural Genom-

ics initiatives [95]. Along the same line of reasoning, small molecules could be designed via SB-VLS methods, not to develop new drugs but as a probe to understand better molecular functions (i.e., macromolecular interactions, allostery...).

ACKNOWLEDGMENTS

Support from the Inserm institute is greatly appreciated.

REFERENCES

- [1] Kubinyi, H. (2003) *Nat. Rev. Drug Discov.*, 2, 665-668.
- [2] Kubinyi, H. (2002) *Drug Discov. Today*, 7, 707-709.
- [3] Rupasinghe, C. N. and Spaller, M. R. (2006) *Curr. Opin. Chem. Biol.*, 10, 188-193.
- [4] Harris, N. V. and Clark, D. E. (2004) *GOR*, 6, 27-31.
- [5] Stoermer, M. J. (2006) *Med. Chem.*, 2, 89-112.
- [6] Krovat, E. M., Steindl, T. and Langer, T. (2005) *Curr. Comput.-Aided Drug Des.*, 1, 93-102.
- [7] Ghosh, S., Nie, A., An, J. and Huang, Z. (2006) *Curr. Opin. Chem. Biol.*, 10, 194-202.
- [8] Xu, H. (2002) *Curr. Top Med. Chem.*, 2, 1305-1320.
- [9] Klebe, G. (2006) *Drug Discov. Today*, 11, 580-594.
- [10] Rollinger, J. M., Langer, T. and Stuppner, H. (2006) *Curr. Med. Chem.*, 13, 1491-1507.
- [11] Stahl, M., Guba, W. and Kansy, M. (2006) *Drug Discov. Today*, 11, 326-333.
- [12] Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P. and Desjarlais, R. L. (2005) *Curr. Pharm. Des.*, 11, 323-333.
- [13] Kitchen, D. B., Decornez, H., Furr, J. R. and Bajorath, J. (2004) *Nat. Rev. Drug Discov.*, 3, 935-949.
- [14] Shoichet, B. K. (2004) *Nature*, 432, 862-865.
- [15] Oprea, T. I. and Matter, H. (2004) *Curr. Opin. Chem. Biol.*, 8, 349-358.
- [16] Alvarez, J. C. (2004) *Curr. Opin. Chem. Biol.*, 8, 365-370.
- [17] Jain, A. N. (2004) *Curr. Opin. Drug Discov. Devel.*, 7, 396-403.
- [18] Villoutreix, B. O. (2002) *Curr. Protein Pept. Sci.*, 3, 341-364.
- [19] Abagyan, R. and Totrov, M. (2001) *Curr. Opin. Chem. Biol.*, 5, 375-382.
- [20] Langer, T. and Hoffmann, R. D. (2001) *Curr. Pharm. Des.*, 7, 509-527.
- [21] Waszkowycz, B. (2002) *Curr. Opin. Drug Discov. Devel.*, 5, 407-413.
- [22] Taylor, R. D., Jewsbury, P. J. and Essex, J. W. (2002) *J. Comput. Aided Mol. Des.*, 16, 151-166.
- [23] Davis, A. M., Teague, S. J. and Kleywegt, G. J. (2003) *Angew. Chem. Int. Ed. Engl.*, 42, 2718-2736.
- [24] Jorgensen, W. L. (2004) *Science*, 303, 1813-1818.
- [25] Fradera, X. and Mestres, J. (2004) *Curr. Top. Med. Chem.*, 4, 687-700.
- [26] Brooijmans, N. and Kuntz, I. D. (2003) *Annu. Rev. Biophys. Biomol. Struct.*, 32, 335-373.
- [27] Kuntz, I. D. (1992) *Science*, 257, 1078-1082.
- [28] Bajorath, J. (2001) *Drug Discov. Today*, 6, 989-995.
- [29] Hopkins, A. L., Mason, J. S. and Overington, J. P. (2006) *Curr. Opin. Struct. Biol.*, 16, 127-136.
- [30] Schneider, G. and Fechner, U. (2005) *Nature Rev.*, 4, 649-663.
- [31] Sperandio, O., Miteva, M. A., Delfaud, F. and Villoutreix, B. O. (2006) *Curr. Protein Pept. Sci.*, 369-393.
- [32] Schneidman-Duhovny, D., Nussinov, R. and Wolfson, H. J. (2004) *Curr. Med. Chem.*, 11, 91-107.
- [33] Dror, O., Shulman-Peleg, A., Nussinov, R. and Wolfson, H. J. (2004) *Curr. Med. Chem.*, 11, 71-90.
- [34] Hajduk, P. J., Huth, J. R. and Tse, C. (2005) *Drug Discov. Today*, 10, 1675-1682.
- [35] Whitty, A. and Kumaravel, G. (2006) *Nat. Chem. Biol.*, 2, 112-118.
- [36] Chen, H., Lyne, P. D., Giordanetto, F., Lovell, T. and Li, J. (2006) *J. Chem. Inf. Model.*, 46, 401-415.
- [37] Anderson, A. C. (2003) *Chem. Biol.*, 10, 787-797.
- [38] Xu, J. and Hagler, A. (2002) *Molecules*, 7, 566-600.
- [39] Shoichet, B. K., McGovern, S. L., Wei, B. and Irwin, J. J. (2002) *Curr. Opin. Chem. Biol.*, 6, 439-446.

- [40] Carpy, A. J. and Marchand-Geneste, N. (2006) *SAR QSAR Environ. Res.*, 17, 1-10.
- [41] Xu, D., Xu, Y. and Uberbacher, E. C. (2000) *Curr. Protein Pept. Sci.*, 1, 1-21.
- [42] Marchand-Geneste, N. and Carpy, A. J. (2004) *SAR QSAR Environ. Res.*, 15, 43-54.
- [43] Weininger, D. (1988) *J. Chem. Inf. Comput. Sci.*, 28, 31-36.
- [44] Dalby, A., Nourse, J. G., Hounshell, W. D., Gushurst, A. K. I., Grier, D. L., Leland, B. A. and Laufer, J. (1992) *J. Chem. Inf. Comput. Sci.*, 32, 244-255.
- [45] Kennedy, T. (1997) *Drug Discov. Today*, 2, 436-444.
- [46] van de Waterbeemd, H. (2005) *Basic Clin. Pharmacol. Toxicol.*, 96, 162-166.
- [47] Van de Waterbeemd, H. (2005) *Expert Opin. Drug Metab. Toxicol.*, 1, 1-4.
- [48] Martin, Y. C. (2005) *J. Med. Chem.*, 48, 3164-3170.
- [49] Ertl, P., Rohde, B. and Selzer, P. (2000) *J. Med. Chem.*, 43, 3714-3717.
- [50] Butina, D., Segall, M. D. and Frankcombe, K. (2002) *Drug Discov. Today*, 7, S83-88.
- [51] Malkia, A., Murtomaki, L., Urtti, A. and Kontturi, K. (2004) *Eur. J. Pharm. Sci.*, 23, 13-47.
- [52] Gunturi, S. B., Narayanan, R. and Khandelwal, A. (2006) *Bioorg. Med. Chem.*, 14, 4118-4129.
- [53] Krejsa, C. M., Horvath, D., Rogalski, S. L., Penzotti, J. E., Mao, B., Barbosa, F. and Migeon, J. C. (2003) *Curr. Opin. Drug Discov. Devel.*, 6, 470-480.
- [54] Vedani, A., Dobler, M. and Lill, M. A. (2006) *Basic Clin. Pharmacol. Toxicol.*, 99, 187-194.
- [55] Ekins, S., Nikolsky, Y. and Nikolskaya, T. (2005) *Trends Pharmacol. Sci.*, 26, 202-209.
- [56] Tetko, I. V., Bruneau, P., Mewes, H. W., Rohrer, D. C. and Poda, G. I. (2006) *Drug Discov. Today*, 11, 700-707.
- [57] Yamashita, F. and Hashida, M. (2004) *Drug Metab. Pharmacokin.*, 19, 327-338.
- [58] Biswas, D., Roy, S. and Sen, S. (2006) *J. Chem. Inf. Model.*, 46, 1394-1401.
- [59] Walters, W. P., Ajay and Murcko, M. A. (1999) *Curr. Opin. Chem. Biol.*, 3, 384-387.
- [60] Ajay, A., Walters, W. P. and Murcko, M. A. (1998) *J. Med. Chem.*, 41, 3314-3324.
- [61] Mannhold, R. and van de Waterbeemd, H. (2001) *J. Comput. Aided Mol. Des.*, 15, 337-354.
- [62] Jonsdottir, S. O., Jorgensen, F. S. and Brunak, S. (2005) *Bioinformatics*, 21, 2145-2160.
- [63] van de Waterbeemd, H. and Gifford, E. (2003) *Nat. Rev. Drug Discov.*, 2, 192-204.
- [64] van de Waterbeemd, H. (2002) *Curr. Opin. Drug Discov. Devel.*, 5, 33-43.
- [65] Muegge, I. (2003) *Med. Res. Rev.*, 23, 302-321.
- [66] Muegge, I., Heald, S. L. and Britelli, D. (2001) *J. Med. Chem.*, 44, 1841-1846.
- [67] Beresford, A. P., Segall, M. and Tarbit, M. H. (2004) *Curr. Opin. Drug Discov. Devel.*, 7, 36-42.
- [68] Olah, M. M., Bologa, C. G. and Oprea, T. I. (2004) *Curr. Drug Discov. Technol.*, 1, 211-220.
- [69] Li, A. P. (2001) *Drug Discov. Today*, 6, 357-366.
- [70] Rishton, G. M. (2003) *Drug Discov. Today*, 8, 86-96.
- [71] Gasteiger, J. (2003) *Mini Rev. Med. Chem.*, 3, 789-796.
- [72] Ekins, S. (2006) *J. Pharmacol. Toxicol. Methods*, 53, 38-66.
- [73] Ekins, S., Boulanger, B., Swaan, P. W. and Hupcey, M. A. (2002) *Mol. Divers.*, 5, 255-275.
- [74] Ekins, S., Ring, B. J., Grace, J., McRobie-Belle, D. J. and Wrighton, S. A. (2000) *J. Pharmacol. Toxicol. Methods*, 44, 313-324.
- [75] Ekins, S., Waller, C. L., Swaan, P. W., Cruciani, G., Wrighton, S. A. and Wikel, J. H. (2000) *J. Pharmacol. Toxicol. Methods*, 44, 251-272.
- [76] de Groot, M. J. (2006) *Drug Discov. Today*, 11, 601-606.
- [77] Keseru, G. M. (2001) *J. Comput. Aided Mol. Des.*, 15, 649-657.
- [78] Lipinski, C. A., Lombardo, F., Dominy, B. W. and Feeney, P. J. (1997) *Adv. Drug Del. Rev.*, 23, 3-25.
- [79] Shen, J., Xu, X., Cheng, F., Liu, H., Luo, X., Shen, J., Chen, K., Zhao, W., Shen, X. and Jiang, H. (2003) *Curr. Med. Chem.*, 10, 2327-2342.
- [80] Baurin, N., Baker, R., Richardson, C., Chen, I., Foloppe, N., Potter, A., Jordan, A., Roughley, S., Parratt, M., Greaney, P., Morley, D. and Hubbard, R. E. (2004) *J. Chem. Inf. Comput. Sci.*, 44, 643-651.
- [81] Clark, D. E. and Pickett, S. D. (2000) *Drug Discov. Today*, 5, 49-58.
- [82] Blake, J. F. (2005) *Med. Chem.*, 1, 649-655.
- [83] Blake, J. F. (2004) *Curr. Opin. Chem. Biol.*, 8, 407-411.
- [84] Knox, A. J., Meegan, M. J., Carta, G. and Lloyd, D. G. (2005) *J. Chem. Inf. Model.*, 45, 1908-1919.
- [85] Verdonk, M. L., Berdini, V., Hartshorn, M. J., Mooij, W. T., Murray, C. W., Taylor, R. D. and Watson, P. (2004) *J. Chem. Inf. Comput. Sci.*, 44, 793-806.
- [86] Bologa, C. G., Olah, M. M. and Oprea, T. I. (2006) *Methods Mol. Biol.*, 316, 375-388.
- [87] Krier, M., Bret, G. and Rognan, D. (2006) *J. Chem. Inf. Model.*, 46, 512-524.
- [88] Saxena, A. K. and Prathipati, P. (2006) *SAR QSAR Environ. Res.*, 17, 371-392.
- [89] Miteva, M. A., Violas, S., Montes, M., Gomez, D., Tuffery, P. and Villoutreix, B. O. (2006) *Nucleic Acids Res.*, 34, W738-744.
- [90] Irwin, J. J. and Shoichet, B. K. (2005) *J. Chem. Inf. Model.*, 45, 177-182.
- [91] Congreve, M., Carr, R., Murray, C. and Jhoti, H. (2003) *Drug Discov. Today*, 8, 876-877.
- [92] Rees, D. C., Congreve, M., Murray, C. W. and Carr, R. (2004) *Nat. Rev. Drug Discov.*, 3, 660-672.
- [93] Bostrom, J. (2001) *J. Comput. Aided Mol. Des.*, 15, 1137-1152.
- [94] Orry, A. J., Abagyan, R. A. and Cavasotto, C. N. (2006) *Drug Discov. Today*, 11, 261-266.
- [95] Marsden, B. D., Sundstrom, M. and Knapp, S. (2006) *Expert Opin. Drug Discov.*, 1, 123-136.
- [96] Wieman, H., Tondel, K., Anderssen, E. and Drablos, F. (2004) *Mini Rev. Med. Chem.*, 4, 793-804.
- [97] Rockey, W. M. and Elcock, A. H. (2002) *Proteins*, 48, 664-671.
- [98] Mestres, J. (2005) *Drug Discov. Today*, 10, 1629-1637.
- [99] Congreve, M., Murray, C. W. and Blundell, T. L. (2005) *Drug Discov. Today*, 10, 895-907.
- [100] Rockey, W. M. and Elcock, A. H. (2006) *Curr. Protein Pept. Sci.*, 437-457.
- [101] Nayeem, A., Sitkoff, D. and Krystek, S., Jr. (2006) *Protein Sci.*, 15, 808-824.
- [102] Franceschi, F. and Duffy, E. M. (2006) *Biochem. Pharmacol.*, 71, 1016-25.
- [103] Kovacs, J. A., Chacon, P. and Abagyan, R. (2004) *Proteins*, 56, 661-668.
- [104] Jacobs, D. J., Rader, A. J., Kuhn, L. A. and Thorpe, M. F. (2001) *Proteins*, 44, 150-165.
- [105] Halle, B. (2002) *Proc. Natl. Acad. Sci. USA*, 99, 1274-1279.
- [106] Deshpande, N., Address, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R. K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M. and Bourne, P. E. (2005) *Nucleic Acids Res.*, 33, D233-237.
- [107] Davies, M. N., Toseland, C. P., Moss, D. S. and Flower, D. R. (2006) *BMC Biochem.*, 7, 18.
- [108] Hattotuwa, C. K., Davies, M. N. and Flower, D. R. (2006) *Curr. Med. Chem.*, 13, 1283-1304.
- [109] An, J., Totrov, M. and Abagyan, R. (2004) *Genome Inform. Ser. Workshop Genome Inform.*, 15, 31-41.
- [110] Hajduk, P. J., Huth, J. R. and Fesik, S. W. (2005) *J. Med. Chem.*, 48, 2518-2525.
- [111] Sottriffer, C. and Klebe, G. (2002) *Farmaco*, 57, 243-251.
- [112] Elcock, A. H. (2001) *J. Mol. Biol.*, 312, 885-896.
- [113] Carpy, A. J. and Marchand-Geneste, N. (2003) *SAR QSAR Environ. Res.*, 14, 329-337.
- [114] Laurie, A. T. and Jackson, R. M. (2006) *Curr. Protein Pept. Sci.*, 395-406.
- [115] Fairlie, D. P. (2004) *Aust. J. Chem.*, 57, 855-857.
- [116] Autin, L., Miteva, M. A., Lee, W. H., Mertens, K., Radtke, K. P. and Villoutreix, B. O. (2005) *J. Thromb. Haemost.*, 3, 2044-2056.
- [117] Autin, L., Steen, M., Dahlback, B. and Villoutreix, B. O. (2006) *Proteins*, 63, 440-450.
- [118] Powers, R., Copeland, J. C., Germer, K., Mercier, K. A., Ramathan, V. and Revesz, P. (2006) *Proteins*, 65, 124-135.
- [119] Gold, N. D. and Jackson, R. M. (2006) *J. Chem. Inf. Model.*, 46, 736-742.

- [120] Gold, N. D. and Jackson, R. M. (2006) *Nucleic Acids Res.*, *34*, D231-234.
- [121] Jambon, M., Imberty, A., Deleage, G. and Geourjon, C. (2003) *Proteins*, *52*, 137-145.
- [122] Rognan, D. (2006) *J. Physiol. Paris*.
- [123] Kellenberger, E., Muller, P., Schalon, C., Bret, G., Foata, N. and Rognan, D. (2006) *J. Chem. Inf. Model.*, *46*, 717-727.
- [124] Sousa, S. F., Fernandes, P. A. and Ramos, M. J. (2006) *Proteins*, *65*, 15-26.
- [125] Jain, A. N. (2006) *Curr. Protein Pept. Sci.*, 407-420.
- [126] Stahl, M. and Rarey, M. (2001) *J. Med. Chem.*, *44*, 1035-1042.
- [127] Wang, R., Lu, Y., Fang, X. and Wang, S. (2004) *J. Chem. Inf. Comput. Sci.*, *44*, 2114-2125.
- [128] Wang, R., Lu, Y. and Wang, S. (2003) *J. Med. Chem.*, *46*, 2287-2303.
- [129] Ferrara, P., Gohlke, H., Price, D. J., Klebe, G. and Brooks, C. L., 3rd. (2004) *J. Med. Chem.*, *47*, 3032-3047.
- [130] Vigers, G. P. and Rizzi, J. P. (2004) *J. Med. Chem.*, *47*, 80-89.
- [131] Pham, T. A. and Jain, A. N. (2005) *J. Med. Chem.*, 5856-5868.
- [132] Kuntz, I. D., Chen, K., Sharp, K. A. and Kollman, P. A. (1999) *Proc. Natl. Acad. Sci. USA*, *96*, 9997-10002.
- [133] Hopkins, A. L., Groom, C. R. and Alex, A. (2004) *Drug Discov. Today*, *9*, 430-431.
- [134] Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N. and Hirono, S. (2006) *J. Chem. Inf. Model.*, *46*, 380-391.
- [135] Feher, M. (2006) *Drug Discov. Today*, *11*, 421-428.
- [136] Miteva, M. A., Lee, W. H., Montes, M. O. and Villoutreix, B. O. (2005) *J. Med. Chem.*, *48*, 6012-6022.
- [137] Cole, J. C., Murray, C. W., Nissink, J. W., Taylor, R. D. and Taylor, R. (2005) *Proteins*, *60*, 325-332.
- [138] Ha, S., Andreani, R., Robbins, A. and Muegge, I. (2000) *J. Comput. Aided Mol. Des.*, *14*, 435-448.
- [139] Bursulaya, B. D., Totrov, M., Abagyan, R. and Brooks, C. L., 3rd. (2003) *J. Comput. Aided Mol. Des.*, *17*, 755-763.
- [140] Cummings, M. D., DesJarlais, R. L., Gibbs, A. C., Mohan, V. and Jaeger, E. P. (2005) *J. Med. Chem.*, *48*, 962-976.
- [141] Kellenberger, E., Rodrigo, J., Muller, P. and Rognan, D. (2004) *Proteins*, *57*, 225-242.
- [142] Perola, E., Walters, W. P. and Charifson, P. S. (2004) *Proteins*, *56*, 235-249.
- [143] Lang, P. T., Kuntz, I. D., Maggiora, G. M. and Bajorath, J. (2005) *J. Biomol. Screen.*, *10*, 649-652.
- [144] Kontoyianni, M., McClellan, L. M. and Sokol, G. S. (2004) *J. Med. Chem.*, *47*, 558-565.
- [145] Kontoyianni, M., Sokol, G. S. and McClellan, L. M. (2005) *J. Comput. Chem.*, *26*, 11-22.
- [146] Bajorath, J. (2004) *Drug Discov. Today*, *9*, 13-14.
- [147] Xue, L., Stahura, F. L. and Bajorath, J. (2004) *J. Chem. Inf. Comput. Sci.*, *44*, 2032-2039.
- [148] Stahura, F. L. and Bajorath, J. (2004) *Comb. Chem. High Throughput Screen.*, *7*, 259-269.
- [149] Frembgen-Kesner, T. and Elcock, A. H. (2006) *J. Mol. Biol.*, *359*, 202-214.
- [150] Lorber, D. M. and Shoichet, B. K. (2005) *Curr. Top. Med. Chem.*, *5*, 739-749.
- [151] Wang, J., Kollman, P. A. and Kuntz, I. D. (1999) *Proteins*, *36*, 1-19.
- [152] Wang, J., Kang, X., Kuntz, I. D. and Kollman, P. A. (2005) *J. Med. Chem.*, *48*, 2432-2444.
- [153] Bissantz, C., Folkers, G. and Rognan, D. (2000) *J. Med. Chem.*, *43*, 4759-4767.
- [154] Maiorov, V. and Sheridan, R. P. (2005) *J. Chem. Inf. Model.*, *45*, 1017-1023.
- [155] Gresh, N. (2006) *Curr. Pharm. Des.*, *12*, 2121-2158.
- [156] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E. (1982) *J. Mol. Biol.*, *161*, 269-288.
- [157] DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R. (1988) *J. Med. Chem.*, *31*, 722-729.
- [158] Korb, O., Stutzle, T. and Exner, T. E. (2006) Ant Colony Optimization and Swarm Intelligence, 5th International Workshop, 247-258.
- [159] Morris, G., Goodsell, D., Halliday, R., Huey, R., Hart, W., Belew, R. and Olson, A. (1998) *J. Comput. Chem.*, *19*, 1639-1662.
- [160] Vague, M., Arola, A., Aliagas, C. and Pujadas, G. (2006) *Bioinformatics*, *22*, 1803-1804.
- [161] Park, H., Lee, J. and Lee, S. (2006) *Proteins*, *65*, 549-554.
- [162] McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A. and Brown, F. K. (2003) *Biopolymers*, *68*, 76-90.
- [163] Jain, A. N. (2003) *J. Med. Chem.*, *46*, 499-511.
- [164] Welch, W., Ruppert, J. and Jain, A. N. (1996) *Chem. Biol.*, *3*, 449-462.
- [165] Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S. and Johnson, A. P. (2006) *Curr. Protein Pept. Sci.*, *7*, 421-435.
- [166] Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K. and Lai, L. (2006) *Proteins*, *62*, 934-946.
- [167] Wang, R., Liu, L., Lai, L. and Tang, Y. (1998) *J. Mol. Model.*, *4*, 379-394.
- [168] Hardy, L. W. and Malikayil, A. (2003) *Curr. Drug. Discov.*, *15*, 15-20.
- [169] Polgar, T., Baki, A., Szendrei, G. I. and Keseru, G. M. (2005) *J. Med. Chem.*, *48*, 7946-7959.
- [170] Carbone, V., Ishikura, S., Hara, A. and El-Kabbani, O. (2005) *Bioorg. Med. Chem.*, *13*, 301-312.
- [171] Tsai, K. C., Chen, S. Y., Liang, P. H., Lu, I. L., Mahindroo, N., Hsieh, H. P., Chao, Y. S., Liu, L., Liu, D., Lien, W., Lin, T. H. and Wu, S. Y. (2006) *J. Med. Chem.*, *49*, 3485-3495.
- [172] Liu, Z., Huang, C., Fan, K., Wei, P., Chen, H., Liu, S., Pei, J., Shi, L., Li, B., Yang, K., Liu, Y. and Lai, L. (2005) *J. Chem. Inf. Model.*, *45*, 10-17.
- [173] Liu, B. and Zhou, J. (2005) *J. Comput. Chem.*, *26*, 484-490.
- [174] Soelaiman, S., Wei, B. Q., Bergson, P., Lee, Y. S., Shen, Y., Mrksich, M., Shoichet, B. K. and Tang, W. J. (2003) *J. Biol. Chem.*, *278*, 25990-25997.
- [175] Li, J., Chen, J., Gui, C., Zhang, L., Qin, Y., Xu, Q., Zhang, J., Liu, H., Shen, X. and Jiang, H. (2006) *Bioorg. Med. Chem.*, *14*, 2209-2224.
- [176] Rastelli, G., Pacchioni, S., Sirawaraporn, W., Sirawaraporn, R., Parenti, M. D. and Ferrari, A. M. (2003) *J. Med. Chem.*, *46*, 2834-2845.
- [177] Peng, H., Huang, N., Qi, J., Xie, P., Xu, C., Wang, J. and Yang, C. (2003) *Bioorg. Med. Chem. Lett.*, *13*, 3693-3699.
- [178] Darmanin, C., Iwata, T., Carper, D. A. and El-Kabbani, O. (2006) *Med. Chem.*, *2*, 239-242.
- [179] Huang, N., Nagarsekar, A., Xia, G., Hayashi, J. and MacKerell, A. D., Jr. (2004) *J. Med. Chem.*, *47*, 3502-3511.
- [180] Wang, J. L., Liu, D., Zhang, Z. J., Shan, S., Han, X., Srinivasula, S. M., Croce, C. M., Alnemri, E. S. and Huang, Z. (2000) *Proc. Natl. Acad. Sci. USA*, *97*, 7124-7129.
- [181] Bianucci, A. M., Massarelli, I., Chiellini, F., Eidelman, C. and Chiellini, E. (2004) *J. Biomater. Sci. Polym. Ed.*, *15*, 1203-1222.
- [182] Vangrevelinghe, E., Zimmermann, K., Schoepfer, J., Portmann, R., Fabbro, D. and Furet, P. (2003) *J. Med. Chem.*, *46*, 2656-2662.
- [183] Liu, H., Li, Y., Song, M., Tan, X., Cheng, F., Zheng, S., Shen, J., Luo, X., Ji, R., Yue, J., Hu, G., Jiang, H. and Chen, K. (2003) *Chem. Biol.*, *10*, 1103-1113.
- [184] Zhou, Y., Peng, H., Ji, Q., Qi, J., Zhu, Z. and Yang, C. (2006) *Bioorg. Med. Chem. Lett.*, *16*, 5878-5882.
- [185] Rogers, J. P., Beuscher, A. E. t., Flajolet, M., McAvoy, T., Nairn, A. C., Olson, A. J. and Greengard, P. (2006) *J. Med. Chem.*, *49*, 1658-1667.
- [186] Li, C., Xu, L., Wolan, D. W., Wilson, I. A. and Olson, A. J. (2004) *J. Med. Chem.*, *47*, 6681-6690.
- [187] Cunningham, P., Afzal-Ahmed, I. and Naftalin, R. J. (2006) *J. Biol. Chem.*, *281*, 5797-5803.
- [188] Hallingback, H. R., Gabdoulline, R. R. and Wade, R. C. (2006) *Biochemistry*, *45*, 2940-2950.
- [189] Sachchidanand, Resnick-Silverman, L., Yan, S., Mutjaba, S., Liu, W. J., Zeng, L., Manfredi, J. J. and Zhou, M. M. (2006) *Chem. Biol.*, *13*, 81-90.
- [190] Hurt, D. E., Sutton, A. E. and Clardy, J. (2006) *Bioorg. Med. Chem. Lett.*, *16*, 1610-1615.
- [191] Stahl, M. T. (2005) *Drug Discov. Today*, *10*, 219-222.
- [192] Jalaie, M. and Shanmugasundaram, V. (2006) *Mini Rev. Med. Chem.*, *6*, 1159-1167.
- [193] Tirado-Rives, J. and Jorgensen, W. L. (2006) *J. Med. Chem.*, *49*, 5880-5884.
- [194] Hassinen, T. and Perakyla, M. (2001) *J. Comp. Chem.*, *22*, 1229-1242.
- [195] Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R. and Willighagen, E. L. (2006) *Curr. Pharm. Des.*, *12*, 2111-2120.
- [196] Mahe, P., Ueda, N., Akutsu, T., Perret, J. L. and Vert, J. P. (2005) *J. Chem. Inf. Model.*, *45*, 939-951.

- [197] Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graph.*, 14, 51-55, 29-32.
- [198] Kraulis, P. J. (1991) *J. Appl. Crystallogr.*, 24, 946-950.
- [199] Sasin, J. M. and Bujnicki, J. M. (2004) *Nucleic Acids Res.*, 32, W586-589.
- [200] Humphrey, W., Dalke, A. and Schulten, K. (1996) *J. Mol. Graph.*, 14, 33-38, 27-38.
- [201] Massire, C., Gaspin, C. and Westhof, E. (1994) *J. Mol. Graph.*, 12, 201-206, 196.
- [202] Abagyan, R. A., Totrov, M. and Kuznetsov, D. (1994) *J. Comput. Chem.*, 15, 488-506.
- [203] Laaksonen, L. (1992) *J. Mol. Graph.*, 10, 33-34.
- [204] Krieger, E., Nielsen, J. E., Spronk, C. A. and Vriend, G. (2006) *J. Mol. Graph. Model.*, 25, 481-6.
- [205] Krieger, E., Koraimann, G. and Vriend, G. (2002) *Proteins*, 47, 393-402.
- [206] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) *J. Comput. Chem.*, 25, 1605-1612.
- [207] Trepalin, S. V., Yarkov, A. V., Pletnev, I. V. and Gakh, A. A. (2006) *Molecules*, 11, 219-231.
- [208] Tetko, I. V. and Tanchuk, V. Y. (2002) *J. Chem. Inf. Comput. Sci.*, 42, 1136-1145.
- [209] Wang, R., Gao, Y. and Lai, L. (2000) *Perspect. Drug Discov. Des.*, 19, 47-66.
- [210] Cruciani, G., Carosati, E., De Boeck, B., Ethirajulu, K., Mackie, C., Howe, T. and Vianello, R. (2005) *J. Med. Chem.*, 48, 6970-6979.
- [211] Thomas, R. S., Rank, D. R., Penn, S. G., Craven, M. W., Drinkwater, N. R. and Bradfield, C. A. (2002) *Meth. Enzymol.*, 357, 198-205.
- [212] Ji, Z. L., Han, L. Y., Yap, C. W., Sun, L. Z., Chen, X. and Chen, Y. Z. (2003) *Drug Saf.*, 26, 685-690.
- [213] Sun, L. Z., Ji, Z. L., Chen, X., Wang, J. F. and Chen, Y. Z. (2002) *Bioinformatics*, 18, 1699-1700.
- [214] Zheng, C. J., Zhou, H., Xie, B., Han, L. Y., Yap, C. W. and Chen, Y. Z. (2004) *Bioinformatics*, 20, 2236-2241.
- [215] Chen, X., Ji, Z. L. and Chen, Y. Z. (2002) *Nucleic Acids Res.*, 30, 412-415.
- [216] Helma, C. (2006) *Mol Divers.*, 10, 147-58.
- [217] Ellis, L. B., Roe, D. and Wackett, L. P. (2006) *Nucleic Acids Res.*, 34, D517-521.
- [218] Richard, A. M., Williams, C. R. and Cariello, N. F. (2002) *Curr. Opin. Drug Discov. Devel.*, 5, 136-143.
- [219] Richard, A. M. and Williams, C. R. (2002) *Mutat. Res.*, 499, 27-52.
- [220] Steinbeck, C. and Kuhn, S. (2004) *Phytochemistry*, 65, 2711-2717.
- [221] Steinbeck, C., Krause, S. and Kuhn, S. (2003) *J. Chem. Inf. Comput. Sci.*, 43, 1733-1739.
- [222] Steinbeck, C. (2001) *Curr. Opin. Drug Discov. Devel.*, 4, 338-342.
- [223] Grothuss, v. M., Pas, J. and Rychlewski, L. (2003) *Bioinformatics*, 19, 1041-1042.
- [224] Girke, T., Cheng, L. C. and Raikhel, N. (2005) *Plant Physiol.*, 138, 573-577.
- [225] Chen, X. and Reynolds, C. H. (2002) *J. Chem. Inf. Comput. Sci.*, 42, 1407-1414.
- [226] Chen, X., Lin, Y. and Gilson, M. K. (2002) *Biopolymers Nucleic Acid Sci.*, 61, 127-142.
- [227] Wang, R., Fang, X., Lu, Y., Yang, C. Y. and Wang, S. (2005) *J. Med. Chem.*, 48, 4111-4119.
- [228] Zhang, J., Aizawa, M., Amari, S., Iwasawa, Y., Nakano, T. and Nakata, K. (2004) *Comput. Biol. Chem.*, 28, 401-407.
- [229] Bergner, A., Gunther, J., Hendlich, M., Klebe, G. and Verdonk, M. (2001) *Biopolymers*, 61, 99-110.
- [230] Hendlich, M. (1998) *Acta Crystallogr. D Biol. Crystallogr.*, 54, 1178-1182.
- [231] Hendlich, M., Bergner, A., Gunther, J. and Klebe, G. (2003) *J. Mol. Biol.*, 326, 607-620.
- [232] Nissink, J. W., Murray, C., Hartshorn, M., Verdonk, M. L., Cole, J. C. and Taylor, R. (2002) *Proteins*, 49, 457-471.
- [233] Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) *Nucleic Acids Res.*, 34, D668-672.
- [234] Block, P., Sotriffer, C. A., Dramburg, I. and Klebe, G. (2006) *Nucleic Acids Res.*, 34, D522-526.
- [235] Michalsky, E., Dunkel, M., Goede, A. and Preissner, R. (2005) *BMC Bioinformatics*, 6, 122.
- [236] Klekota, J., Roth, F. P. and Schreiber, S. L. (2006) *Bioinformatics*, 22, 1670-1673.
- [237] Schuttelkopf, A. W. and van Aalten, D. M. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, 60, 1355-1363.
- [238] van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W. and Vriend, G. (1996) *J. Comput. Aided Mol. Des.*, 10, 255-262.
- [239] Wang, T. and Zhou, J. (1999) *J. Mol. Mod.*, 5, 231-251.
- [240] Wang, R., Gao, Y. and Lai, L. (2000) *J. Mol. Mod.*, 6, 498-516.
- [241] Jain, A. N. (2004) *J. Med. Chem.*, 47, 947-961.
- [242] Wang, Y., Anderson, J. B., Chen, J., Geer, L. Y., He, S., Hurwitz, D. I., Liebert, C. A., Madej, T., Marchler, G. H., Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Song, J. S., Thiessen, P. A., Yamashita, R. A. and Bryant, S. H. (2002) *Nucleic Acids Res.*, 30, 249-252.
- [243] Bader, G. D., Betel, D. and Hogue, C. W. (2003) *Nucleic Acids Res.*, 31, 248-250.
- [244] Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M. and Eisenberg, D. (2002) *Nucleic Acids Res.*, 30, 303-305.
- [245] Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. and Cesareni, G. (2002) *FEBS Lett.*, 513, 135-140.
- [246] Kundrotas, P. J. and Alexov, E. (2006) *Biophys. J.*, 91, 1724-1736.
- [247] Alland, C., Moreews, F., Boens, D., Carpentier, M., Chiusa, S., Lonquety, M., Renault, N., Wong, Y., Cantalloube, H., Chomilier, J., Hochez, J., Pothier, J., Villoutreix, B. O., Zagury, J. F. and Tuffery, P. (2005) *Nucleic Acids Res.*, 33, W44-49.
- [248] Bowie, J. U., Luthy, R. and Eisenberg, D. (1991) *Science*, 253, 164-170.
- [249] Colovos, C. and Yeates, T. O. (1993) *Protein Sci.*, 2, 1511-1519.
- [250] Canutescu, A. A., Shelenkov, A. A. and Dunbrack, R. L., Jr. (2003) *Protein Sci.*, 12, 2001-2014.
- [251] Gautier, R., Camproux, A. C. and Tuffery, P. (2004) *Nucleic Acids Res.*, 32, W508-511.
- [252] Lehtonen, J. V., Still, D. J., Rantanen, V. V., Ekholm, J., Bjorklund, D., Ifiikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J., Pentikainen, O., Nyronen, T., Salminen, T., Gyllenberg, M. and Johnson, M. S. (2004) *J. Comput. Aided Mol. Des.*, 18, 401-419.
- [253] Sali, A. and Blundell, T. L. (1993) *J. Mol. Biol.*, 234, 779-815.
- [254] Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., Soto, C. S., Goldsmith-Fischman, S., Kernysky, A., Schlessinger, A., Koh, I. Y., Alexov, E. and Honig, B. (2003) *Proteins*, 53 (Suppl 6), 430-435.
- [255] Torda, A. E., Procter, J. B. and Huber, T. (2004) *Nucleic Acids Res.*, 32, W532-535.
- [256] McGuffin, L. J., Bryson, K. and Jones, D. T. (2000) *Bioinformatics*, 16, 404-405.
- [257] Rost, B., Yachdav, G. and Liu, J. (2004) *Nucleic Acids Res.*, 32, W321-326.
- [258] Guo, J. T., Ellrott, K., Chung, W. J., Xu, D., Passovets, S. and Xu, Y. (2004) *Nucleic Acids Res.*, 32, W522-525.
- [259] Kim, D. E., Chivian, D. and Baker, D. (2004) *Nucleic Acids Res.*, 32, W526-531.
- [260] Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J. and Edelman, M. (2005) *Nucleic Acids Res.*, 33, W39-43.
- [261] Neshich, G., Rocchia, W., Mancini, A. L., Yamagishi, M. E., Kuser, P. R., Fileto, R., Baudet, C., Pinto, I. P., Montagner, A. J., Palandrani, J. F., Krauchenco, J. N., Torres, R. C., Souza, S., Togawa, R. C. and Higa, R. H. (2004) *Nucleic Acids Res.*, 32, W595-601.
- [262] Raymer, M. L., Sanschagrin, P. C., Punch, W. F., Venkataraman, S., Goodman, E. D. and Kuhn, L. A. (1997) *J. Mol. Biol.*, 265, 445-464.
- [263] Paiardini, A., Bossa, F. and Pascarella, S. (2005) *Nucleic Acids Res.*, 33, W50-55.
- [264] Bagos, P. G., Liakopoulos, T. D., Spyropoulos, I. C. and Hamodrakas, S. J. (2004) *Nucleic Acids Res.*, 32, W400-404.
- [265] Shyu, C. R., Chi, P. H., Scott, G. and Xu, D. (2004) *Nucleic Acids Res.*, 32, W572-575.
- [266] Ye, Y. and Godzik, A. (2004) *Nucleic Acids Res.*, 32, W582-585.
- [267] Tsuchiya, Y., Kinoshita, K., Ito, N. and Nakamura, H. (2006) *Nucleic Acids Res.*, 34, W20-24.
- [268] Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M. and Abergel, C. (2004) *Nucleic Acids Res.*, 32, W606-609.
- [269] Berjanskii, M. V., Neal, S. and Wishart, D. S. (2006) *Nucleic Acids Res.*, 34, W63-69.

- [270] Gelly, J. C., Etchebest, C., Hazout, S. and de Brevern, A. G. (2006) *Nucleic Acids Res.*, *34*, W75-78.
- [271] Shih, E. S., Gan, R. C. and Hwang, M. J. (2006) *Nucleic Acids Res.*, *34*, W95-98.
- [272] Lee, S., Lee, B., Jang, I., Kim, S. and Bhak, J. (2006) *Nucleic Acids Res.*, *34*, W99-W103.
- [273] Tyagi, M., Sharma, P., Swamy, C. S., Cadet, F., Srinivasan, N., de Brevern, A. G. and Offmann, B. (2006) *Nucleic Acids Res.*, *34*, W119-123.
- [274] Maupetit, J., Gautier, R. and Tuffery, P. (2006) *Nucleic Acids Res.*, *34*, W147-151.
- [275] Chen, C. C., Hwang, J. K. and Yang, J. M. (2006) *Nucleic Acids Res.*, *34*, W152-157.
- [276] Fernandez-Fuentes, N., Zhai, J. and Fiser, A. (2006) *Nucleic Acids Res.*, *34*, W173-176.
- [277] Waldispühl, J., Berger, B., Clote, P. and Steyaert, J. M. (2006) *Nucleic Acids Res.*, *34*, W189-193.
- [278] Pugalenti, G., Shameer, K., Srinivasan, N. and Sowdhamini, R. (2006) *Nucleic Acids Res.*, *34*, W231-234.
- [279] Liu, Y. and Kuhlman, B. (2006) *Nucleic Acids Res.*, *34*, W235-238.
- [280] Deane, C. M. and Blundell, T. L. (2001) *Protein Sci.*, *10*, 599-612.
- [281] Chang, D. T., Chen, C. Y., Chung, W. C., Oyang, Y. J., Juan, H. F. and Huang, H. C. (2004) *Nucleic Acids Res.*, *32*, W76-82.
- [282] Davis, I. W., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2004) *Nucleic Acids Res.*, *32*, W615-619.
- [283] Chandonia, J. M., Hon, G., Walker, N. S., Lo Conte, L., Koehl, P., Levitt, M. and Brenner, S. E. (2004) *Nucleic Acids Res.*, *32*, D189-192.
- [284] Samudrala, R. and Levitt, M. (2000) *Protein Sci.*, *9*, 1399-1401.
- [285] Abagyan, R., Lee, W. H., Raush, E., Budagyan, L., Totrov, M., Sundstrom, M. and Marsden, B. D. (2006) *Trends Biochem. Sci.*, *31*, 76-78.
- [286] Douguet, D. and Labesse, G. (2001) *Bioinformatics*, *17*, 752-753.
- [287] Laurie, A. T. and Jackson, R. M. (2005) *Bioinformatics*, *21*, 1908-1916.
- [288] Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) *Nucleic Acids Res.*, *34*, W116-118.
- [289] Nayal, M. and Honig, B. (2006) *Proteins*, *63*, 892-906.
- [290] Chang, D. T., Oyang, Y. J. and Lin, J. H. (2005) *Nucleic Acids Res.*, *33*, W233-238.
- [291] Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. and Ben-Tal, N. (2003) *Bioinformatics*, *19*, 163-164.
- [292] Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. and Ben-Tal, N. (2002) *Bioinformatics*, *18* (Suppl 1), S71-77.
- [293] Brady, G. P. and Stouten, P. F. W. (2000) *J. Comput.-Aided Mol. Des.*, *14*, 383-401.
- [294] Kleywegt, G. J. and Jones, T. A. (1994) *Acta Crystallogr. D Biol. Crystallogr.*, *50*, 178-185.
- [295] Laskowski, R. A. (1995) *J. Mol. Graph.*, *13*, 323-330, 307-328.
- [296] Tovchigrechko, A. and Vakser, I. A. (2006) *Nucleic Acids Res.*, *34*, W310-314.
- [297] Gabdoulline, R. R., Wade, R. C. and Walther, D. (1999) *Trends Biochem. Sci.*, *24*, 285-287.
- [298] Sobolev, V., Moallem, T. M., Wade, R. C., Vriend, G. and Edelman, M. (1997) *Proteins*, (Suppl. 1), 210-214.
- [299] Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H. J. (2005) *Nucleic Acids Res.*, *33*, W363-367.
- [300] Shatsky, M., Dror, O., Schneidman-Duhovny, D., Nussinov, R. and Wolfson, H. J. (2004) *Nucleic Acids Res.*, *32*, W503-507.
- [301] Chelliah, V., Blundell, T. L. and Fernandez-Recio, J. (2006) *J. Mol. Biol.*, *357*, 1669-1682.
- [302] Diemand, A. V. and Scheib, H. (2004) *Nucleic Acids Res.*, *32*, W512-516.
- [303] Gabb, H. A., Jackson, R. M. and Sternberg, M. J. (1997) *J. Mol. Biol.*, *272*, 106-120.
- [304] Comeau, S. R., Gatchell, D. W., Vajda, S. and Camacho, C. J. (2004) *Nucleic Acids Res.*, *32*, W96-99.
- [305] Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyyi, V., Mitchell, J. C., Nelson, E., Tsigelny, I. and Ten Eyck, L. F. (2001) *Protein Eng.*, *14*, 105-113.
- [306] Chen, R., Li, L. and Weng, Z. (2003) *Proteins*, *52*, 80-87.
- [307] Daily, M. D., Masica, D., Sivasubramanian, A., Somarouthu, S. and Gray, J. J. (2005) *Proteins*, *60*, 181-186.
- [308] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. and Vakser, I. A. (1992) *Proc. Natl. Acad. Sci. USA*, *89*, 2195-2199.
- [309] Ritchie, D. W. and Kemp, G. J. (2000) *Proteins*, *39*, 178-194.
- [310] Ausiello, G., Cesareni, G. and Helmer-Citterich, M. (1997) *Proteins*, *28*, 556-567.
- [311] Kastenholtz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E. and Fox, T. (2000) *J. Med. Chem.*, *43*, 3033-3044.
- [312] Verdonk, M. L., Cole, J. C., Watson, P., Gillet, V. and Willett, P. (2001) *J. Mol. Biol.*, *307*, 841-859.
- [313] Ogmen, U., Keskin, O., Aytuna, A. S., Nussinov, R. and Gursoy, A. (2005) *Nucleic Acids Res.*, *33*, W331-336.
- [314] Shulman-Peleg, A., Nussinov, R. and Wolfson, H. J. (2005) *Nucleic Acids Res.*, *33*, W337-341.
- [315] Camacho, C. J. and Zhang, C. (2005) *Bioinformatics*, *21*, 2534-2536.
- [316] Stark, A., Sunyaev, S. and Russell, R. B. (2003) *J. Mol. Biol.*, *326*, 1307-1316.
- [317] Kleywegt, G. J. (1999) *J. Mol. Biol.*, *285*, 1887-1897.
- [318] Porter, C. T., Bartlett, G. J. and Thornton, J. M. (2004) *Nucleic Acids Res.*, *32*, D129-133.
- [319] Binkowski, T. A., Adamian, L. and Liang, J. (2003) *J. Mol. Biol.*, *332*, 505-526.
- [320] Ausiello, G., Zanzoni, A., Peluso, D., Via, A. and Helmer-Citterich, M. (2005) *Nucleic Acids Res.*, *33*, W133-137.
- [321] Martin, L., Catherinot, V. and Labesse, G. (2006) *Nucleic Acids Res.*, *34*, W325-329.
- [322] Laskowski, R. A., Watson, J. D. and Thornton, J. M. (2005) *Nucleic Acids Res.*, *33*, W89-93.
- [323] Kinoshita, K., Furui, J. and Nakamura, H. (2002) *J. Struct. Funct. Genomics*, *2*, 9-22.
- [324] Gabdoulline, R. R., Ulbrich, S., Richter, S. and Wade, R. C. (2006) *Nucleic Acids Res.*, *34*, W79-83.
- [325] Wernersson, R., Rapacki, K., Staerfeldt, H. H., Sackett, P. W. and Molgaard, A. (2006) *Nucleic Acids Res.*, *34*, W84-88.
- [326] Hung, J. H., Huang, H. D. and Lee, T. Y. (2006) *Nucleic Acids Res.*, *34*, W89-94.
- [327] Chang, D. T., Weng, Y. Z., Lin, J. H., Hwang, M. J. and Oyang, Y. J. (2006) *Nucleic Acids Res.*, *34*, W303-309.
- [328] Magyar, C., Gromiha, M. M., Pujadas, G., Tusnady, G. E. and Simon, I. (2005) *Nucleic Acids Res.*, *33*, W303-305.
- [329] Dantzer, J., Moad, C., Heiland, R. and Mooney, S. (2005) *Nucleic Acids Res.*, *33*, W311-314.
- [330] Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) *Nucleic Acids Res.*, *33*, W382-388.
- [331] Guerois, R., Nielsen, J. E. and Serrano, L. (2002) *J. Mol. Biol.*, *320*, 369-387.
- [332] Ng, P. C. and Henikoff, S. (2003) *Nucleic Acids Res.*, *31*, 3812-3814.
- [333] Ramensky, V., Bork, P. and Sunyaev, S. (2002) *Nucleic Acids Res.*, *30*, 3894-3900.
- [334] Hollup, S. M., Salensminde, G. and Reuter, N. (2005) *BMC Bioinformatics*, *6*, 52.
- [335] Hinsin, K. (2000) *J. Comput. Chem.*, *21*, 79-95.
- [336] Bashford, D. and Karplus, M. (1990) *Biochemistry*, *29*, 10219-10225.
- [337] Miteva, M. A., Tuffery, P. and Villoutreix, B. O. (2005) *Nucleic Acids Res.*, *33*, W372-375.
- [338] Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. and Onufriev, A. (2005) *Nucleic Acids Res.*, *33*, W368-371.
- [339] Li, H., Robertson, A. D. and Jensen, J. H. (2005) *Proteins*, *61*, 704-721.
- [340] Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. USA*, *98*, 10037-10041.
- [341] Vriend, G. (1990) *J. Mol. Graph.*, *8*, 52-56, 29.
- [342] Delarue, M. and Sanejouand, Y. H. (2002) *J. Mol. Biol.*, *320*, 1011-1024.
- [343] Nayal, M., Hitz, B. C. and Honig, B. (1999) *Protein Sci.*, *8*, 676-679.
- [344] Schlessinger, A., Yachdav, G. and Rost, B. (2006) *Bioinformatics*, *22*, 891-893.
- [345] Eyal, E., Najmanovich, R., Sobolev, V. and Edelman, M. (2001) *Bioinformatics*, *17*, 381-382.
- [346] Cai, C. Z., Han, L. Y., Ji, Z. L., Chen, X. and Chen, Y. Z. (2003) *Nucleic Acids Res.*, *31*, 3692-3697.
- [347] Gale, J. D. (1997) *JCS Faraday Trans.*, *93*, 629-637.
- [348] Salerno, W. J., Seaver, S. M., Armstrong, B. R. and Radhakrishnan, I. (2004) *Nucleic Acids Res.*, *32*, W566-568.

- [349] Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. and Russell, R. B. (2003) *Structure*, 11, 1453-1459.
- [350] Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004) *J. Mol. Biol.*, 337, 635-645.
- [351] Yang, L. W., Rader, A. J., Liu, X., Jursa, C. J., Chen, S. C., Karimi, H. A. and Bahar, I. (2006) *Nucleic Acids Res.*, 34, W24-31.
- [352] Azuara, C., Lindahl, E., Koehl, P., Orland, H. and Delarue, M. (2006) *Nucleic Acids Res.*, 34, W38-42.
- [353] Kantardjiev, A. A. and Atanasov, B. P. (2006) *Nucleic Acids Res.*, 34, W43-47.
- [354] Toseland, C. P., McSparron, H., Davies, M. N. and Flower, D. R. (2006) *Nucleic Acids Res.*, 34, D199-203.
- [355] Tynan-Connolly, B. M. and Nielsen, J. E. (2006) *Nucleic Acids Res.*, 34, W48-51.
- [356] Lindahl, E., Azuara, C., Koehl, P. and Delarue, M. (2006) *Nucleic Acids Res.*, 34, W52-56.
- [357] Jang, Y., Jeong, J. I. and Kim, M. K. (2006) *Nucleic Acids Res.*, 34, W57-62.
- [358] Ahmad, S., Kono, H., Arauzo-Bravo, M. J. and Sarai, A. (2006) *Nucleic Acids Res.*, 34, W124-127.
- [359] Parthiban, V., Gromiha, M. M. and Schomburg, D. (2006) *Nucleic Acids Res.*, 34, W239-242.
- [360] Ferrer-Costa, C., Gelpi, J. L., Zamakola, L., Parraga, I., de la Cruz, X. and Orozco, M. (2005) *Bioinformatics*, 21, 3176-3178.
- [361] Ahmad, S. and Sarai, A. (2004) *Nucleic Acids Res.*, 32, W104-107.
- [362] Ruan, J., Stormo, G. D. and Zhang, W. (2004) *Nucleic Acids Res.*, 32, W146-149.
- [363] Ying, X., Luo, H., Luo, J. and Li, W. (2004) *Nucleic Acids Res.*, 32, W150-153.
- [364] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Schulten, K. (2005) *J. Comput. Chem.*, 26, 1781-1802.
- [365] Massire, C. and Westhof, E. (1998) *J. Mol. Graph. Model.*, 16, 197-205, 255-197.
- [366] Soman, K. V., Midoro-Horiuti, T., Ferreone, J. C., Goldblum, R. M., Brooks, E. G., Kurosky, A., Braun, W. and Schein, C. H. (2000) *Biophys. J.*, 79, 1601-1609.
- [367] von Freyberg, B., Richmond, T. J. and Braun, W. (1993) *J. Mol. Biol.*, 233, 275-292.
- [368] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005) *J. Comput. Chem.*, 26, 1701-1718.
- [369] Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A. and Jones, T. A. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2240-2249.
- [370] Goodsell, D. S., Morris, G. M. and Olson, A. J. (1996) *J. Mol. Recognit.*, 9, 1-5.
- [371] Springer, C., Adalsteinsson, H., Young, M. M., Kegelmeyer, P. W. and Roe, D. C. (2005) *J. Med. Chem.*, 48, 6821-6831.
- [372] Han, L. Y., Lin, H. H., Li, Z. R., Zheng, C. J., Cao, Z. W., Xie, B. and Chen, Y. Z. (2006) *J. Chem. Inf. Model.*, 46, 445-450.
- [373] Betzi, S., Suhre, K., Chetrit, B., Guerlesquin, F. and Morelli, X. (2006) *J. Chem. Inf. Model.*, 46, 1704-1712.
- [374] Sottriffer, C. A., Gohlke, H. and Klebe, G. (2002) *J. Med. Chem.*, 45, 1967-1970.
- [375] Chen, X., Ji, Z. L., Zhi, D. G. and Chen, Y. Z. (2002) *Comput. Chem.*, 26, 661-666.
- [376] Wang, R., Lai, L. and Wang, S. (2002) *J. Comput. Aided Mol. Des.*, 16, 11-26.
- [377] Li, H., Gao, Z., Kang, L., Zhang, H., Yang, K., Yu, K., Luo, X., Zhu, W., Chen, K., Shen, J., Wang, X. and Jiang, H. (2006) *Nucleic Acids Res.*, 34, W219-224.