Topical Perspective

Open source molecular modeling

Somayeh Pirhadi [b], Jocelyn Sunseri [a], David Ryan Koes [a],*

[a] *Department of Computational and Systems Biology, University of Pittsburgh, United States*
[b] *Young Researchers and Elites Club, Science and Research Branch, Islamic Azad University, Tehran, Iran*

ARTICLE INFO

ABSTRACT

The success of molecular modeling and computational chemistry efforts are, by definition, dependent on quality software applications. Open source software development provides many advantages to users of modeling applications, not the least of which is that the software is free and completely extendable. In this review we categorize, enumerate, and describe available open source software packages for molecular modeling and computational chemistry. An updated online version of this catalog can be found at https://opensourcemolecularmodeling.github.io.

© 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. What is open source?

Free and open source software (FOSS) is software that is both considered "free software," as defined by the Free Software Foundation (http://fsf.org) and "open source," as defined by the Open Source Initiative (http://opensource.org). The distinctions between free and open source software are largely philosophical – the free software movement is primary motivated by user freedoms ("free as in speech, not free as in beer") while the open source movement is more concerned with promoting an open development model to enhance software quality. However, as a practical matter, especially with regards to scientific software, such distinctions remain philosophical rather than practical as the most popular software licenses are both free and open source.

The unifying theme of open source software licenses is that they allow users to use, modify, and distribute software without significant restrictions. This is achieved by making the full source code of the software available to users. Broadly speaking, open source licenses fall into two categories: permissive and copyleft. Permissive licenses, such as the Apache, BSD, MIT, and Python licenses, place minimal restrictions on how modified code may be distributed, such as requiring attribution and limiting liability. They specifically do not require that redistributions of modified source code be licensed under the same license as the original source code.

This enables source code licensed under a permissive license to be incorporated into commercial, proprietary programs that are not open source. In contrast, copyleft licenses, such as the different versions of the GNU Public License (GPL), require that public redistributions of licensed software remain licensed under a GPL license. That is, the source code must remain publicly and freely available. The GNU Lesser General Public License (LGPL) is slightly less restrictive version of the GPL used primarily for libraries as it does not require software that uses LGPL licensed software as a library to be licensed under the LGPL. Although copyleft licenses do not prohibit selling software, since the full source code must remain freely available, in practice vendors of copylefted software must commercialize the support of the product, rather than the product itself. Finally, we note there are other software licenses that make source code available, but are not open source licenses. These licenses typically prohibit the redistribution of the source code. Such licenses, which we will refer to as "source-available" licenses, have some popularity in academia as they allow source code to be distributed to other researchers in non-profit institutions, but allow the code to be sold to commercial entities.

1.2. Advantages and disadvantages of open source

The value of open source software in cheminformatics and molecular modeling is somewhat controversial. Unsurprisingly, those affiliated with commercial scientific software argue that traditional commercial development, with its associated support and continuous development, provides a superior value [1], while open source advocates feel the benefits outweigh the burdens [2,3]. Our goal is not to revisit these arguments. Instead, we assert that open

---

* Corresponding author.
  E-mail address: dkoes@pitt.edu (D.R. Koes).

source scientific software is a *de facto* part of the scientific community, and so in this review we catalog those open source packages that fall within the domain of cheminformatics and molecular modeling.

There are a few aspects of the open source software debate that we find particularly relevant. First, opponents are right to point out that free software is not free – users of open source software generally take on a much greater burden in supporting the software than with commercial software. This is one reason why it is important, when possible, to seek open source software that is under active development and supported by a broad community. Therefore, in this review we attempt to quantify the current level of development and usage of each package as an indirect measure of quality and usability. Second, the primary advantage of open source software is the ability to redistribute code without restriction. This inherently enables reproducibility and lets scientists "stand on the shoulders of giants" instead of reinventing the wheel. Consequently, in this review we limit ourselves to a survey of true open source software and exclude source-available software that may place restrictions on the publication of reproducible research results.

2. Methods

We organize software packages into seven categories: cheminformatics, visualization, QSAR (quantitative structure–activity relationship) and ADMET modeling, quantum chemistry, ligand dynamics and free energy calculations, and virtual screening (including ligand design). We identified open source software packages by browsing the relevant categories (Molecular Science, Chemistry, Bio-Informatics, Medical Science) on the popular SourceForge (http://sourceforge.net) repository, searching for categories on GitHub (http://github.com), searching for categories on OpenHub (https://www.openhub.net), searching for categories together with "open source software" on Google, and browsing the Click2Drug (http://click2drug.org) and VLS3D [4] directories. Finally, a draft document was tweeted (@david_koes) to solicit suggestions for additional packages from the community.

For every identified software package, we report its primary URL and software license and assign it an activity code. For simplicity, BSD-like licenses (e.g., NCSA) are reported as BSD. Activity codes consist of a development activity level (alphabetical) and usage activity level (numerical). Activity codes were assigned as follows:

2.1. Development activity

A Substantial development (e.g., a new major release, the addition of new features, or substantial refinements of existing features) within the last 18 months. Note this includes all projects that were created in the last 18 months.
B Evidence of some development within the last 18 months such as a minor release or bug fixes to a development branch.
C No evidence of development (changes to the source code or documentation) within the last 18 months. Note that in cases where a package does not follow an open development model (i.e., source is only released with official releases) the estimate of development activity will be overly conservative.

2.2. Usage activity

1. Substantial user usage within the last 18 months (more than 20 downloads a month on average from SourceForge, more than 20 stars or forks on GitHub, more than 10 citations a year, and/or a clearly active user community as indicated by traffic on mailing lists or discussion boards).
2. Moderate user usage within the last 18 months.

3. Minimal or no identifiable user usage within the last 18 months (fewer than 50 downloads total on SourceForge, three or fewer stars and/or forks on GitHub, or fewer than one citation a year).

We omit some packages with extended periods of inactivity (e.g., more than 10 years) where there is little evidence of any usage or packages that are referenced in the literature but for which we could not find a extant source code repository. We also omit packages that provide common and/or trivial functionality (e.g., molecular weight calculators) and those that require non-open source packages in order to function.

3. Cheminformatics

Cheminformatics involves the representation, manipulation and analysis of molecular data [5,6]. Cheminformatic toolkits, although they may contain standalone utility programs, are primarily designed to function as libraries for other programs so that common functions, such as parsing molecular data, need not be reimplemented. As libraries, the native programming language of a toolkit is particularly relevant as it influences the language programs that integrate with the toolkit can be written in. In some cases, alternative language bindings, which essentially translate between programming languages, may be available, but due to the use of different idioms by different languages (e.g., object-oriented vs. functional, manual vs. automatic memory management) use of these non-native bindings may be cumbersome. In addition to toolkits, we catalog standalone programs, including conformer generators for converted 2D information into 3D molecular structures (some of which have been critically evaluated [7]), and graphical environments for creating and managing workflows.

3.1. Toolkits (*Table 1*)

The Biochemical Algorithms Library (BALL) [8] provides an object-oriented C++ library for structural bioinformatics, and its capabilities include molecular mechanics, support for reading and writing a variety of file formats, protein–ligand scoring, docking, and QSAR modeling.

The Chemistry Development Kit (CDK) [9] is a cheminformatics toolkit written in Java. Its capabilities include support for reading and writing a variety of chemical formats, descriptor and fingerprint calculation, force field calculations, substructure search, and structure generation.

Chem^f [10] is a minimal cheminformatics toolkit written in the functional language Scala.

chemfp [11] is a high-performance library with a Python interface for generating and searching for molecular fingerprints. chemkit is a C++ cheminformatics toolkit that includes support for visualization with the Qt framework and molecular modeling.

ChemmineR [12] is a cheminformatics package for the R statistical programming language that is built using Open Babel. Its capabilities include property calculations, similarity search, and classification and clustering of compounds.

Cinfony [13] provides a single, simple standardized interface to other cheminformatics toolkits, including Open Babel, RDKit, the CDK, Indigo, JChem, OPSIN, and several web services.

CurlySMILES [14] provides parsing functionality for an extension of the SMILES format that supports the description of complex molecular systems.

DisCuS (Database System for Compound Selection) [15] provides support for analyzing the results of a high-throughput screen.

Fafoom (flexible algorithm for optimization of molecules) [16] is a Python library for identifying low energy conformers using a genetic algorithm.

Table 1
Open source cheminformatics toolkits.

Name	URL	License	Activity	Citation
BALL	http://www.ball-project.org/	LGPL	A2	[8]
CDK	https://sourceforge.net/projects/cdk	LGPL	A1	[9]
Chem^f	https://github.com/stefan-hoeck/chemf	GPL	C2	[10]
chemfp	http://chemfp.com	MIT	C3	[11]
chemkit	https://github.com/kylelutz/chemkit	BSD	B1	
ChemmineR	https://www.bioconductor.org/packages/release/bioc/html/ChemmineR.html	Artistic	A1	[12]
Cinfony	https://github.com/cinfony/cinfony	BSD/GPL	B1	[13]
CurlySMILES	http://www.axeleratio.com/csm/proj/main.htm	GPL	C2	[14]
DisCuS	https://github.com/mwojcikowski/discus	GPL	B3	[15]
Fafoom	https://github.com/adrianasupady/fafoom	LGPL	A2	[16]
fmcsR	http://www.bioconductor.org/packages/fmcsR	Artistic	A1	[17]
frowns	http://frowns.sourceforge.net	Python	C2	
Helium	http://www.moldb.net/helium.html	BSD	B2	
Indigo	http://lifescience.opensource.epam.com/indigo	GPL	A1	[18]
JoeLib	http://sourceforge.net/projects/joelib	GPL	C1	
LICSS	https://github.com/KevinLawson/excel-cdk	GPL	A2	[19]
MayaChemTools	http://www.mayachemtools.org	LGPL	A2	[20]
Mychem	http://mychem.sourceforge.net	GPL	B2	
ODDT	https://github.com/oddt/oddt	BSD	A2	[21]
Open Babel	http://openbabel.org	GPL	A1	[22]
OPSIN	http://opsin.ch.cam.ac.uk	Artistic	A1	[23]
OrChem	http://orchem.sourceforge.net	LGPL	C2	[24]
osra	http://sourceforge.net/projects/osra	GPL	A1	[25]
OUCH	https://github.com/odj/Ouch	GPL	C2	
pybel	http://openbabel.org/docs/dev/UseTheLibrary/Python_Pybel.html	GPL	A1	[26]
rcdk	https://cran.r-project.org/web/packages/rcdk	LGPL	B2	[27]
RDKit	http://www.rdkit.org	BSD	A1	
RInChI	http://www-rinchi.ch.cam.ac.uk	Apache	A3	
rpubchem	https://r-forge.r-project.org/projects/rpubchem	GPL	C3	
rubabel	https://github.com/princelab/rubabel	MIT	C2	[28]
SMSD	http://www.ebi.ac.uk/thornton-srv/software/SMSD	CCAL	B2	[29]
Som-it	http://silicos-it.be	LGPL	C3	
webchem	https://github.com/ropensci/webchem	MIT	A2	

fmcsR [41] is an R package that efficiently performs flexible maximum common substructure matching that allows minor mismatches between atoms and bonds in the common substructure.

Frowns is a cheminformatics toolkit mostly written in Python that provides basic support for SMILES and SD files, SMARTS search, fingerprint generation, and property perception.

Helium is a cheminformatics toolkit written using modern C++ idioms that provides support for SMILES files, fingerprints generation, and SMARTS and SMIRKS.

Indigo [18] is a cheminformatics toolkit written in C++ with C, Python, Java (including a KNIME node), and C# bindings. Its capabilities include general support for manipulating molecules, property calculation, combinatorial chemistry, scaffold detection and R-group decomposition, reaction processing, substructure matching and similarity search.

JOELib is a cheminformatics toolkit written in Java. Its capabilities include SMARTS substructure search, descriptor calculation, and processing/filtering pipes.

LICSS [19] integrates with the CDK to provide representations and analysis of chemical data embedded within Microsoft Excel.

MayChemTools is a collection of Perl scripts for manipulating chemical data, interfacing with databases, generating fingerprints, performing similarity search, and computing molecular properties.

Mychem is built using OpenBabel and provides an extension to the MySQL database package that adds the ability to search, analyze, and convert chemical data within a MySQL database.

The Open Drug Discovery Toolkit (ODDT) [21] is entirely written in Python, is built on top of RDKit and Open Babel, and is focused on providing enhanced functionality for managing and implementing drug discovery workflows, such as making it easy to implement a docking pipeline.

Open Babel [22] is substantial cheminformatics toolkit written in C++ with Python, Perl, Java, Ruby, R, PHP, and Scala

bindings. Its capabilities include support for more than 100 chemical file formats, fingerprint generation, property determination, similarity and substructure search, structure generation, and molecular force fields. It has absorbed the Confab [47] conformer generator which produces 3D structures through the systematic enumeration of torsions and energy minimization.

OPSIN [23], the Open Parser for Systematic IUPAC nomenclature, converts plain-text chemical nomenclature to machine readable CML or InChi formats.

OrChem is built using the CDK and provides an extension to Oracle databases that adds the ability to incorporate and search chemical data.

OSRA [25] provides optical structure recognition. It takes as input an image and generates a SMILES string.

Ouch (Ouch Uses Chemical Haskell) is a minimal cheminformatics toolkit written in the functional language Haskell.

Pybel [26] provides the full functionality of Open Babel, but common routines are provided in a simplified, more 'pythonic' interface.

rcdk [27] provides an R interface to the CDK and working with fingerprints.

RDKit is a substantial cheminformatics toolkit written in C++ with Python, Java and C# bindings. Its capabilities include file handling, manipulation of molecular data, chemical reactions, substantial support for fingerprinting, substructure and similarity search, 3D conformer generation, property determination, force field support, shape-based alignment and screening, and integration with PyMOL, KNIME, and PostgreSQL.

RInChI provides tools for creating and manipulating reaction InChIs, a unique string for describing a reaction.

rpubchem is an R package for interfacing with the PubChem database.

Table 2
Standalone open source cheminformatics software.

Name	URL	License	Activity	Citation
cApp	http://www.structuralchemistry.org/pcsb	GPL	A3	[30]
checkmol/matchmol	http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm.html	GPL	C3	[31]
ConvertMAS	http://sourceforge.net/projects/convertmas	GPL	A3	
Filter-it	http://silicos-it.be	LGPL	C3	
Frog2	https://github.com/tuffery/Frog2	GPL	B2	[32]
LMR	https://github.com/IanAWatson/Lilly-Medchem-Rules	GPL	B2	[33]
Molpher	https://www.assembla.com/spaces/molpher/wiki	GPL	C2	[34]
MoSS	http://www.borgelt.net/moss.html	MIT	A2	[35]
OMG	http://sourceforge.net/projects/openmg	GPL	C1	[36]
sdf2xyz2sdf	http://sdf2xyz2sdf.sourceforge.net	GPL	C2	[37]
sdsorter	https://sourceforge.net/projects/sdsorter	GPL	B3	
Shape	http://sourceforge.net/projects/shapega	GPL	C3	[38]
Strip-it	http://silicos-it.be	LGPL	C3	

Table 3
Open source graphical development environments for cheminformatics.

Name	URL	License	Activity	Citation
AMBIT	http://ambit.sourceforge.net	GPL	A1	[39]
Bioclipse	http://www.bioclipse.net	Eclipse	B1	[40]
Galaxy Tool	https://github.com/bgruening/galaxytools	Academic	A1	[41]
KNIME	https://www.knime.org	GPL	A1	[42]
Orange	orange.biolab.si	BSD	A1	[43]
SA2	http://sa2.sourceforge.net	GPL	A1	[44]
Taverna	http://www.taverna.org.uk	LGPL	A1	[45]
Weka	https://sourceforge.net/projects/weka	GPL	A1	[46]

rubabel [28] is similar to Pybel in that it provides a native Ruby interface to Open Babel.

The Small Molecule Subgraph Detector (SMSD) [29] is a Java library for calculating the maximum common subgraph between small molecules.

Som-it is an R package for creating and visualizing self-organizing maps from large datasets.

webchem is an R package for interfacing with a dozen different on-line resources for chemical data.

3.2. Standalone programs (Table 2)

cApp [30] is a Java application that provides tools for evaluating physico-chemical properties, performing similarity searches, and querying the PubChem database.

The utilities checkmol and matchmol [31] decompose and compare functional groups of input molecules.

ConvertMAS is a utility for converting between formats and merging and splitting multi-molecule files.

Filter-it filters a set of molecules based on their properties such as physicochemical parameters and graph-based properties.

Frog2 [32] uses a two stage Monte Carlo approach coupled with energy minimization to rapidly generate 3D conformers.

The Lilly MedChem Rules (LMR) [33] apply filters to avoid reactive and promiscuous compounds.

Molpher [34] generates a virtual chemical library that represents the chemical space between two input molecules as it consists of the path found by morphing one molecule to another.

MoSS (Molecular Substructure miner) [35] finds common molecular substructures and discriminative fragments within a compound library.

The Open Molecule Generator (OMG) [36] enumerates all possible chemical structures given constraints on their composition.

sdf2xyz2sdf [37] converts between SDF and TINKER XYZ files.

sdsorter provides convenient routines for manipulating, sorting, and filtering the contents of sdf molecular data files based on the embedded sd data tags.

Shape [38] employs a genetic algorithm to generate conformations of carbohydrates.

Strip-it is built using Open Babel and extracts molecular scaffolds.

3.3. Graphical development environments (Table 3)

Ambit [39] integrates with the CDK to provide web-based applications for chemical search and analysis and includes a tautomer generation algorithm [48].

Bioclipse [40] is a workbench, based on the Eclipse framework, for manipulating and analyzing biochemical data and databases. It integrates with the CDK and Jmol to provide cheminformatic functionality and also has modules for bioinformatics (primarily sequence analysis) and QSAR modeling.

Galaxy [41] is a web platform for exploring biomedical data and includes as a component a Chemical Toolbox that integrates a number of other cheminformatics tools to offer an array of molecular search, property calculation, clustering, and manipulation capabilities.

The Konstanz Information Miner (KNIME) is a general workflow environment that includes a number of plugins for cheminformatics, such as CDK [49] and RDKit modules, as well as bioinformatics and machine learning modules.

Orange [43] is a graphical interface for construction interactive workflows and performing data analysis and visualization.

Screening Assistant 2 (SA2) [44] is a GUI written in Java that integrates with other toolkits to help manage, analyze, and visualize libraries of compounds.

Taverna [45] is a graphical workflow editor that includes support for integrating with web services and the CDK [50].

Weka [46] is a platform for data mining and machine learning that can be adapted for cheminformatics.

4. Visualization

An essential component of any molecular modeling exercise is the ability to visualize and, sometimes, edit molecular data.

Visualization software usually either deals with exclusively 2D or 3D molecular data and may be primarily intended for desktop usage (native 'fat' clients) or as a component embedded in a web browser. Packages designed for visualizing entire datasets are cataloged in the QSAR section (Table 10) and those with a particular emphasize on visualizing the output of quantum chemistry packages are in the Quantum Chemistry section (Table 13).

4.1. 2D desktop applications (Table 4)

BKChem is a 2D molecular editor written in python that uses the Tk GUI toolkit.

chemfig is a tool for embedding chemical drawings in LaTeX documents.

Chemtool is a 2D molecular editor for Linux systems that uses the GTK toolkit.

JChemPaint [51] is a Java-based 2D molecular editor built using the CDK toolkit.

LeView [52] generates 2D representations of ligand-protein interactions that highlight features such as hydrogen bonds.

mol2chemfig [53] converts SMILES files into LaTeX source code.

Molsketch is a 2D molecular editor written in C++ with the Qt toolkit that includes support for the Windows and Android operating systems.

SketchEl is a Java-based 2D molecular editor that includes support for a datasheet view for handling multi-molecule files.

4.2. 3D desktop applications (Table 5)

Avogadro [54] is a 3D molecular viewer and editor with a modular plugin architecture with both Python and C++ bindings that includes interactive structure optimization for real-time editing.

BALLView [55] provides interactive 3D visualizations as part of the BALL [8] cheminformatics toolkit.

gMol provides basic interactive 3D visualizations of molecular data readable by Open Babel.

Jamberoo provides a basic Java-based 3D molecular viewer and editor.

LP Molecular Viewer is an ActiveX/ATL control for embedding interactive 3D representations of molecular data in Microsoft products.

Luscus [56] is a 3D viewer and editor that is designed with a focus on electronic structure information.

Molecular Rift [57] integrates with the Oculus Rift virtual reality headset to provide immersive visualization of 3D molecular data.

OpenStructure [58] is a computational structural biology framework that provides a 3D viewer for manipulating structural information and includes an interactive Python shell.

PLIP (Protein–Ligand Interaction Profiler) [59] runs as a web application and analyzes and visualizes protein–ligand interactions in 3D.

PyMOL is a substantial 3D molecular viewer that includes a full Python interface to support scripting and plugin development.

RasTop and OpenRasMol are based off the venerable RasMol software and provide basic 3D visualization.

SPADE (Structural Proteomics Application Development Environment) [60] is a graphical Python interface for structural informatics.

QuteMol [61] provides high-quality, visually engaging renderings of 3D molecular data.

4.3. Web-based visualization (Table 6)

3Dmol.js [62] is a JavaScript library that provides WebGL-accelerated interactive 3D visualizations of molecular structures and surfaces.

CH5M3D [63] uses JavaScript and HTML5 to provide visualization and editing of 3D structures of small molecules.

Chemozart [64] is a WebGL-based web application for 3D editing of small molecules.

CWC (ChemDoodle Web Components) [65] provides a suite of web-based visualizers and editors for 2D and 3D molecular data.

JSME [66] is a pure JavaScript 2D molecular editor that can export and import SMILES data.

Jmol [67] is a Java applet for interactive 3D visualization that provides significant cheminformatics support and a custom scripting language.

JSmol [68] is the JavaScript port of Jmol and does not require the Java plugin to run.

NGL [69] is a WebGL-accelerated viewer and JavaScript library for interactive 3D visualization of macromolecules.

PV (Protein Viewer) [70] is a WebGL-accelerated viewer for interactive 3D visualization of macromolecules with a functional-style API.

5. QSAR/ADMET modeling

Quantitative structure–activity relationship (QSAR) approaches find relationships between the chemical structures of a series of compounds (or structural-related properties) and a biological activity [71], including ADMET properties (absorption, distribution, metabolism, excretion, and toxicity). QSAR methods calculate relevant molecular descriptors, build informative models using these descriptors and then apply the models. Models and datasets may also be visualized to aid in model development and understanding of compound libraries. Note that most of the cheminformatics toolkits shown in Table 1 also are capable of generating descriptors (and are often used as libraries by QSAR modeling software).

5.1. Descriptor calculators (Table 7)

4D Flexible Atom-Pair Kernel (4D FAP) computes a '4D' similarity measure from the molecular graphs of an ensemble of conformations which can be incorporated into QSAR models.

The BlueDesc descriptor calculator is a command-line tool that converts an MDL SD file into ARFF and LIBSVM format using CDK and JOELib2 for machine learning and data mining purposes. It computes 174 descriptors taken from both libraries.

MolSig [73] computes molecular graph descriptors that include stereochemistry information.

PaDEL-Descriptor [74] calculates molecular descriptors and fingerprints. It computes 1875 descriptors (1444 1D, 2D descriptors and 431 3D descriptors) and 12 types of fingerprints.

Topological maximum cross correlation descriptors (TMACC) [75] generates 2D autocorrelation descriptors that are low dimensional and interpretable and appropriate for QSAR modeling.

5.2. Model building (Table 8)

AZOrange [76] is a machine learning package that supports QSAR model building in a full work flow from descriptor computation to automated model building, validation and selection. It promotes model accuracy by using several high performance machine learning algorithms for efficient data set specific selection of the statistical approach.

Bioalerts [77] uses RDKit fingerprints to create models from discrete (e.g., toxic/non-toxic) and continuous data. It includes the capability to visualize problematic functional groups.

Chemistry aware model builder (camb) [78] is an R package for the generation of quantitative models. Its capabilities

Table 4
Open source tools for 2D molecular visualization and editing on the desktop.

ame	URL	License	Activity	Citation
BKchem	http://bkchem.zirael.org	GPL	C3	
chemfig	https://www.ctan.org/pkg/chemfig	LaTeX	A2	
Chemtool	http://ruby.chemie.uni-freiburg.de/~martin/chemtool	GPL	B3	
JChemPaint	http://jchempaint.github.io	LGPL	B1	[51]
LeView	http://www.pegase-biosciences.com/leview-ligand-environment-viewer	GPL	B3	[52]
mol2chemfig	http://chimpsky.uwaterloo.ca/mol2chemfig	LaTeX	C3	[53]
Molsketch	http://sourceforge.net/projects/molsketch	GPL	A1	
SketchEl	http://sketchel.sourceforge.net	GPL	A1	

Table 5
Open source tools for 3D molecular visualization and editing on the desktop.

Name	URL	License	Activity	Citation
Avogadro	http://avogadro.cc	GPL	A1	[54]
BALLView	http://www.ball-project.org/ballview	LPGL	A2	[55]
gMol	https://github.com/tjod/gMol/wiki	GPL	A3	
Jamberoo	https://sourceforge.net/projects/jbonzer	LGPL	A3	
LPMV	https://sourceforge.net/projects/lpmolecularviewer	LGPL	B3	
Luscus	https://sourceforge.net/projects/luscus	Academic	A1	[56]
Molecular Rift	https://github.com/Magnusnorrby/MolecularRift	GPL	A3	[57]
OpenStructure	http://www.openstructure.org	LGPL	A2	[58]
PLIP	https://github.com/ssalentin/plip	Apache	A2	[59]
PyMOL	https://sourceforge.net/projects/pymol	Python	A1	
RasTop	https://sourceforge.net/projects/rastop	GPL	C1	
OpenRasMol	https://sourceforge.net/projects/openrasmol	GPL	C1	
SPADE	http://www.spadeweb.org	BSD	C3	[60]
QuteMol	http://qutemol.sourceforge.net	GPL	C1	[61]

Table 6
Open source tools for web-based molecular visualization.

Name	URL	License	Activity	Citation
3Dmol.js	3dmol.csb.pitt.edu	BSD	A1	[62]
CH5M3D	https://sourceforge.net/projects/ch5m3d	GPL	C1	[63]
Chemozart	https://chemozart.com	Apache	A1	[64]
CWC	https://web.chemdoodle.com	GPL	A1	[65]
JSME	http://peter-ertl.com/jsme	BSD	A1	[66]
Jmol	http://jmol.sourceforge.net	LGPL	A1	[67]
JSmol	https://sourceforge.net/projects/jsmol	LGPL	A1	[68]
NGL	http://proteinformatics.charite.de/ngl	MIT	A1	[69]
PV	https://biasmv.github.io/pv	MIT	A1	[70]

Table 7
Open source software for computing molecular descriptors.

Name	URL	License	Activity	Citation
4D-FAP	http://www.ra.cs.uni-tuebingen.de/software/4DFAP	LGPL	C2	[72]
BlueDesc	http://www.ra.cs.uni-tuebingen.de/software/bluedesc	GPL	C3	
MolSig	http://molsig.sourceforge.net	GPL	C2	[73]
PaDEL-descriptor	http://www.yapcwsoft.com/dd/padeldescriptor	Public Domain	C1	[74]
TMACC	http://comp.chem.nottingham.ac.uk/download/tmacc	GPL	C2	[75]

Table 8
Open source software for building QSAR models.

Name	URL	License	Activity	Citation
AZOrange	https://github.com/AZCompTox/AZOrange	LGPL	C2	[76]
Bioalerts	https://github.com/isidroc/bioalerts	GPL	A3	[77]
camb	https://github.com/cambDI	GPL	B2	[78]
eTOXlab	https://github.com/manuelpastor/eTOXlab	GPL	B3	[79]
Open3DGRID	http://open3dgrid.sourceforge.net	GPL	B1	
Open3DQSAR	http://open3dqsar.sourceforge.net	GPL	B1	[80]
QSAR-tools	https://github.com/dkoes/qsar-tools	BSD	A3	

include descriptor calculation (including 905 two-dimensional and 14 fingerprint type descriptors for small molecules, 13 whole protein sequence descriptors, and 8 types of amino acid descriptors), model generation, ensemble modeling, and visualization.

eTOXLab [79] provides a portable modeling framework embedded in a self-contained virtual machine for easy deployment.

Open3DGrid and Open3DQSAR [80] are a suite of related tools that build 3D QSAR models. Open3DGrid generates molecular interaction fields (MIFs) in a variety of formats, and Open3DQSAR builds

predictive models from the MIFs of aligned molecules. Calculations can be visualized in real time in PyMOL.

QSAR-tools is a set of Python scripts that use RDKit to build linear QSAR models from 2D chemical data.

5.3. Model application (Table 9)

SMARTCyp [92] is a QSAR model that predicts the sites of cytochrome P450-mediated metabolism of drug-like molecules directly from the 2D structure of a molecule using fragment-based energy rules.

Toxtree [82] is a Java GUI application for estimating the "toxic hazard" of molecules using a variety of toxicity prediction modules, such as oral toxicity, skin and eye irritation prediction, covalent protein binding and DNA binding, Cytochrome P450-mediated drug metabolism (using SMARTCyp) and more.

UG-RNN/AquaSol [83] is an undirected graph recursive neural network that has been trained to predict aqueous solubility from molecular graphs.

5.4. Visualization (Table 10)

CheS-Mapper (chemical space mapper) [93,84]. is a 3D-viewer for small compounds in chemical datasets. It embeds a dataset into 3D space by performing dimensionality reduction on the properties of the compounds.

DataWarrior [85] is a data visualization and analysis tool for chemical data with a rich set of available property calculations, similarity metrics, modeling capabilities, and data set representations.

DecoyFinder [86] provides a GUI for selecting a set of decoy compounds from a large library that are appropriate matches to a given set of actives.

Scaffold Hunter [87] provides a Java-based GUI for visualizing the relationship between compounds in a dataset.

Synergy Maps [88] visualizes synergistic activity extracted from screens of drug combinations.

VIDEAN (visual and interactive descriptor analysis) [89] is a visual tool for iteratively choosing a subset of descriptors appropriate for predicting a target property with the aid of statistical methods.

WCSE (Wikipedia chemical structure explorer) [90] runs as a web application and provides a 2D interface for visualizing and searching for 2D molecules.

WebChemViewer [91] is an online viewer for viewing and interacting with lists of compounds and their associated data.

6. Quantum chemistry

Quantum mechanics (QM) is increasingly accessible and relevant to molecular modeling and *in silico* drug design [94]. Quantum chemistry models molecules at the level of their wave function and offers enhanced accuracy and detail compared to classical models at a substantial increase in computation time. Corresponding to the large diversity of solution methods for Schrödinger's equation, there are many software packages for *ab initio* quantum chemistry, as well as programs for interpreting and visualizing the results of these calculations.

6.1. Ab initio calculation (Table 11)

ABINIT [95] can calculate the total energy, charge density and electronic structure of molecules and periodic solids with density functional theory (DFT) and Many-Body Perturbation Theory (MBPT), using pseudopotentials and a planewave or wavelet basis. ABINIT also can optimize the geometry, perform molecular dynamics simulations, or generate dynamical matrices, Born effective charges, and dielectric tensors and many more properties.

ACES [96] performs calculations such as single point energy calculations, analytical gradients, and analytical Hessians, and is highly parallelized, including support for GPU computing. A focus of ACES is the use of MBPT and the coupled-cluster approximation to reliable treat electron correlation.

BigDFT [119,120,97] performs *ab initio* calculations using Daubechies wavelets and has the capability to use a linear scaling method. Periodic systems, surfaces and isolated systems can be simulated with the proper boundary conditions. It is included as part of ABINIT.

CP2K [98] performs simulations of solid state, liquid, molecular and biological systems. Its particular focus is massively parallel and linear scaling electronic structure methods and state-of-the-art *ab initio* molecular dynamics (AIMD) simulations. It is optimized for the mixed Gaussian and Plane-Waves method using pseudopotentials and can run on parallel and on GPUs.

Dacapo is a total energy program that uses density functional theory. It can do molecular dynamics/structural relaxation while solving the Schrödinger equations. It has support for parallel execution and is used through the Atomic Simulation Environment (ASE) [99].

ErgoSCM [100] is a quantum chemistry program for large-scale self-consistent field calculations. It performs electronic structure calculations using Hartree–Fock and Kohn–Sham density functional theory and achieves linear scaling for both CPU usage and memory utilization.

ERKALE [101] is designed to compute X-ray properties, such as ground-state electron momentum densities and Compton profiles, and core (X-ray absorption and X-ray Raman scattering) and valence electron excitation spectra of atoms and molecules.

GPAW [102] is a DFT code that uses the projector-augmented wave (PAW) technique [121,122] and integrates with the atomic simulation environment (ASE) [99].

HORTON (Helpful Open-source Research TOol for N-fermion systems) has as a primary design goal ease of extensibility for researching new methods in *ab initio* electronic structure theory.

JANPA [103] computes natural atomic orbitals from a reduced one-particle density matrix.

MPQC (massively parallel quantum chemistry program) [104] offers many features including closed shell, unrestricted and general restricted open shell Hartree–Fock energies and gradients, closed shell, unrestricted and general restricted open shell density functional theory energies and gradients, second order open shell perturbation theory and Z-averaged perturbation theory energies.

NWChem [105] provides a full suite of methods for modeling both classical and QM systems. Its capabilities include molecular electronic structure, QM/MM, pseudopotential plane-wave electronic structure, and molecular dynamics and is designed to scale across hundreds of processors.

Octopus pervorms *ab initio* calculations using time-dependent DFT (TDDFT) and pseudopotentials. Included in the project is libxc [123] which is a standalone library of exchange-correlation functionals for DFT (released under the LGPL).

OpenMX (Open source package for material eXplorer) [107] is designed for nano-scale material simulations based on DFT, norm-conserving pseudopotentials, and pseudo-atomic localized basis functions. OpenMX is capable of performing calculations of physical properties such as magnetic, dielectric, and electric transport properties and is optimized for large-scale parallelism.

Psi4 [108] is a suite of *ab initio* quantum chemistry programs that supports a wide range of computations (e.g., Hartree–Fock, MP2, coupled-cluster) and general procedures such as geometry optimization and vibrational frequency analysis with more than 2500 basis functions.

Table 9
Open source software that applies a QSAR model.

Name	URL	License	Activity	Citation
SMARTCyp	http://www.farma.ku.dk/smartcyp	LGPL	C1	[81]
Toxtree	http://toxtree.sourceforge.net	GPL	A1	[82]
UG-RNN	http://cdb.ics.uci.edu/cgibin/tools/AquaSolWeb.py	Apache	C1	[83]

Table 10
Open source software for visualizing QSAR models and compound datasets.

Name	URL	License	Activity	Citation
CheS-Mapper	http://ches-mapper.org	GPL	A2	[84]
DataWarrior	http://www.openmolecules.org/datawarrior	GPL	A1	[85]
DecoyFinder	http://urvnutrigenomica-ctns.github.io/DecoyFinder	GPL	A1	[86]
Scaffold Hunter	http://scaffoldhunter.sf.net	GPL	A1	[87]
Synergy Maps	https://github.com/richlewis42/synergy-maps	MIT	A2	[88]
VIDEAN	https://github.com/jimenamartinez/VIDEAN	BSD	A3	[89]
WCSE	http://www.cheminfo.org/wikipedia	BSD	A2	[90]
WebChemViewer	http://sourceforge.net/projects/webchemviewer	BSD	C3	[91]

PyQuante is a collection of modules, mostly written in Python, for performing Hartree–Fock and DFT calculations with a focus on providing a well-engineered set of tools. A new version is under development (https://github.com/rpmuller/pyquante2).

PySCF is also written primarily in Python and supports several popular methods such as Hartree–Fock, DFT, and MP2. It also has easy of use and extension as primary design goals.

QMCPACK [109] is a many-body *ab initio* quantum Monte Carlo implementation for computing electronic structure properties of molecular, quasi-2D and solid-state systems. The standard file formats utilized for input and output are in XML and HDF5.

QUANTUM ESPRESSO [110] is designed for modeling at the nanoscale using DFT, plane waves, and pseudopotentials and its capabilities include ground-state calculations, structural optimization, transition states and minimum energy paths, *ab initio* molecular dynamics, DFT perturbation theory, spectroscopic properties, and quantum transport.

RMG [111] is a DFT code that uses real space grids to provide high scalability across thousands of processors and GPU acceleration for both structural relaxation and molecular dynamics.

Siam Quantum (SQ) is optimized for parallel computation and its capabilities include the calculation of Hartree–Fock and MP2 energies, minimum energy crossing point calculations, geometry optimization, population analysis, and quantum molecular dynamics.

6.2. Helper applications (Table 12)

FragIt [112] generates fragments of large molecules to use as input files in quantum chemistry programs that support fragment based methods.

cclib [113] provides a consistent interface for parsing and interpreting the results of a number of quantum chemistry packages.

GaussSum [113] uses cclib to extract useful information from the results of quantum chemistry programs (ADF, GAMESS, Gaussian, Jaguar) including monitoring the progress of geometry optimization, the UV/IR/Raman spectra, molecular orbital (MO) levels and MO contributions.

Geac (Gaussian ESI Automated Creator) extracts data from Gaussian log files.

Nancy_EX [114] post-processes Gaussian output and analyzes excited states including natural transition orbitals, detachment and attachment density matrices, and charge-transfer descriptors.

orbkit [115] is a post-processing tool for the results of quantum chemistry programs. It has native support for a number of programs (MOLPRO, TURBOMOLE, GAMESS-US, PROAIMS/AIMPAC, Gaussian) and additionally interfaces with cclib for additional file format support. It can extract grid-based quantities such as molecular orbitals and electron density, as well as Muliken population charges and other properties.

Table 11
Open source quantum chemistry software for performing *ab initio* calculations.

Name	URL	License	Activity	Citation
ABINIT	http://www.abinit.org	GPL	A1	[95]
ACES	http://www.qtp.ufl.edu/aces	GPL	A1	[96]
BigDFT	http://bigdft.org	GPL	A1	[97]
CP2K	http://www.cp2k.org	GPL	A1	[98]
DACAPO	https://wiki.fysik.dtu.dk/dacapo	GPL	C1	[99]
ErgoSCF	http://www.ergoscf.org	GPL	C2	[100]
ERKALE	https://github.com/susilehtola/erkale	GPL	B2	[101]
GPAW	https://wiki.fysik.dtu.dk/gpaw	GPL	A1	[102]
HORTON	http://theochem.github.io/horton	GPL	A1	
JANPA	http://janpa.sourceforge.net	BSD	A1	[103]
MPQC	http://www.mpqc.org	LGPL	B1	[104]
NWChem	http://www.nwchem-sw.org	ECL	A1	[105]
Octopus	http://www.tddft.org/programs/octopus	GPL	A1	[106]
OpenMX	http://www.openmx-square.org	GPL	A1	[107]
Psi4	http://www.psicode.org	GPL	A1	[108]
pyquante	http://sourceforge.net/projects/pyquante	BSD	A1	
PySCF	https://github.com/sunqm/pyscf	BSD	A1	
QMCPACK	http://qmcpack.org	BSD	A2	[109]
Quantum espresso	http://www.quantum-espresso.org	GPL	A1	[110]
RMG	http://rmgdft.sourceforge.net	BSD/GPL	A1	[111]
SQ	https://sites.google.com/site/siamquantum	GPL	A2	

Table 12
Open source software for analyzing the results of quantum chemistry calculations.

Name	URL	License	Activity	Citation
FragIt	https://github.com/FragIt	GPL	A2	[112]
cclib	https://github.com/cclib/cclib	LGPL	A1	[113]
GaussSum	http://sourceforge.net/projects/gausssum	GPL	A1	[113]
Geac	https://github.com/LaTruelle/Geac	GPL	A3	
Nancy_EX	http://sourceforge.net/projects/nancyex	GPL	A2	[114]
orbkit	http://orbkit.sourceforge.net	GPL	A2	[115]

Table 13
Open source software for quantum chemistry visualization.

Name	URL	License	Activity	Citation
CCP1GUI	http://www.scd.stfc.ac.uk/research/app/40501.aspx	GPL	C3	
ccwatcher	http://sourceforge.net/projects/ccwatcher	GPL	B2	
Gabedit	http://gabedit.sourceforge.net	BSD	C1	[116]
J-ICE	http://j-ice.sourceforge.net	GPL	A1	[117]
QMForge	http://qmforge.sourceforge.net	GPL	A1	
wxMacMolPlt	http://brettbode.github.io/wxmacmolplt	GPL	A1	[118]

6.3. Visualization (Table 13)

CCP1GUI provides a graphical user interface to various computational chemistry codes with an emphasis on integration with the GAMESS-UK quantum chemistry program.

ccwatcher provides a graphical interface for the monitoring of computational chemistry programs.

Gabedit [116] is a graphical user interface to a large number of quantum chemistry packages. It can create input files and graphically visualize calculation results.

J-ICE [117] is a Jmol-based viewer for crystallographic and electronic properties that can be deployed as a Java applet embedded in a web browser.

QMForge provides a graphical user interface for analyzing and visualizing results of quantum chemistry DFT calculations (Gaussian, ADF, GAMESS, Jaguar, NWChem, ORCA, QChem). Analyses include a number of population analyses, Mayer's bond order, charge decomposition, and fragment analysis.

wxMacMolPlt [118] is a multi-platform GUI for setting up and visualizing input and output files for the GAMESS quantum chemistry software.

7. Ligand dynamics and free energy calculations

Simulation based approaches for analyzing ligand-protein interactions have the potential to fully account for the dynamics of the ligand, protein, and water system. Molecular dynamics simulation is a powerful tool for computing free energies as long as the system can be properly parameterized and the simulation code adapted and analyzed to implement the desired free energy method.

7.1. Simulation software (Table 14)

Campari [124] conducts flexible Monte Carlo sampling of biopolymers in internal coordinate space, with built-in analysis routines to estimate structural properties and support for replica exchange and Wang–Landau sampling.

DL_POLY Classic [125] is a general purpose molecular dynamics simulation package that can run in parallel and includes a Java graphical user interface.

GALAMOST (GPU accelerated large-scale molecular simulation toolkit) [126] uses GPU computing to perform traditional molecular dynamics with a special focus on polymeric systems at mesoscopic scales.

Gromacs [127] is a complete and well-established package for molecular dynamics simulations that provides high performance on both CPUs and GPUs. It can be used for free energy and QM/MM calculations and includes a comprehensive set of analysis tools.

Iphigenie [128] is a molecular mechanics program that features polarizable force fields, the HADES reaction field, and QM/(P)MM hybrid simulations.

LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) [129] is a highly modular classical molecular dynamics simulator that includes a diverse array of energy potentials and integrators.

MDynaMix [130] is a basic general purpose molecular dynamics package.

MMTK (Molecular Modelling Toolkit) [131] is a library written in Python (with some time critical parts written in C) for constructing and simulating molecular systems. Its capabilities include molecular dynamics, energy minimization, and normal mode analysis and it is well-suited for methods development.

OpenMM [132] is a substantial toolkit for high performance molecular dynamics simulations that includes support for GPU acceleration.

ProtoMol [133], and the associated MDLab Python bindings [151], provides an object-oriented framework for prototyping algorithms for molecular dynamics simulations and includes an interface to OpenMM.

ProtoMS [134] is a Monte Carlo biomolecular simulation program which can be used to calculate relative and absolute free energies and water placement with the GCMC and JAWS methodologies.

Sire is a collection of modular libraries intended to facilitate fast prototyping and the development of new algorithms for molecular simulation and molecular design. It has apps for system setup, simulation, and analysis.

WESTPA (The Weighted Ensemble Simulation Toolkit with Parallelization and Analysis) [135] is a library for performing weighted ensemble simulations to sample rare events and compute rigorous kinetics.

yank is built off of OpenMM and provides a Python interface for performing alchemical free energy calculations.

7.2. Simulation setup and analysis (Table 15)

AmberTools [136] is an open source component of the non-open source Amber package and provides a large suite of analysis programs. As of Amber15, AmberTools includes the lower performance, but readily extendable, sander molecular dynamics code.

Table 14
Open source software for performing molecular simulations.

Name	URL	License	Activity	Citation
Campari	http://campari.sourceforge.net	GPL	B1	[124]
DL_POLY Classic	http://www.ccp5.ac.uk/DL_POLY_CLASSIC/	BSD	C3	[125]
GALAMOST	http://galamost.ciac.jl.cn	GPL	A2	[126]
Gromacs	http://www.gromacs.org	LGPL	A1	[127]
Iphigenie	https://sourceforge.net/projects/iphigenie	GPL	A2	[128]
LAMMPS	http://lammps.sandia.gov	GPL	A1	[129]
MDynaMix	http://www.fos.su.se/~sasha/mdynamix	GPL	A1	[130]
MMTK	http://dirac.cnrs-orleans.fr/MMTK	CeCILL	C1	[131]
OpenMM	https://simtk.org/home/openmm	GPL/MIT	A1	[132]
ProtoMol	http://protomol.sourceforge.net	GPL	C1	[133]
ProtoMS	http://www.essexgroup.soton.ac.uk/ProtoMS	GPL	A2	[134]
Sire	http://siremol.org	GPL	C2	
WESTPA	https://westpa.github.io/westpa	GPL	A2	[135]
yank	http://getyank.org	LGPL	A2	

LOOS (Lightweight Object-Oriented Structure library) [137] is a C++ library (with Python bindings) for reading and analyzing molecular dynamics trajectories that also includes a number of standalone programs.

lsfitpar [138] derives bonded parameters for Class I force fields by performing a robust fit to potential energy scans provided by the user.

MDAnalysis [139] is a Python library for reading and analyzing molecular dynamics simulations with some time critical sections written in C.

MDTraj [140] provides high-performance reading, writing, and analysis of molecular dynamics trajectories in a diversity of formats from a Python interface.

MEMBPLUGIN [141] analyzes molecular dynamics simulations of lipid bilayers and is most commonly used as a VMD plugin.

MEPSA (Minimum Energy Pathway Analysis) [142] provides tools for analyzing energy landscapes and pathways.

MSMBuilder [143] is an application and Python library for building Markov models of high-dimensional trajectory data.

packmol [144] is a utility for setting molecular systems by realistically packing molecules to obey a variety of constraints and can create solvent mixtures and lipid bilayers.

PDB2PQR [145] prepares structures for electrostatics calculations by adding hydrogens, calculating sidechain pK_a, adding missing heavy atoms, and assigning force field-dependent parameters; users can specify an ambient pH.

PLUMED [146] interfaces with an assortment of molecular dynamics software packages to provide a unified interface for performing free energy calculations using methods such as metadynamics, umbrella sampling and steered MD (Jarzynski).

ProDy [147] is a Python toolkit for analyzing proteins and includes facilities for trajectory analysis and druggability predictions using simulations of molecular probes [152].

Pteros [148] is a C++ library (with Python bindings) for reading and analyzing molecular dynamics trajectories.

PyEMMA [149] is a Python library for performing kinetic and thermodynamic analyses of molecular dynamics simulations using Markov models.

PyRED [150] generates RESP and ESP charges for the AMBER, CHARMM, OPLS, and Glycam and force fields.

PYTRAJ is a Python interface to the `cpptraj` tool of AmberTools.

simpletraj is a lightweight Python library for parsing molecular dynamics trajectories.

WHAM (Weighted Histogram Analysis Method) calculates the potential of mean force (PMF) from umbrella sampling simulations.

8. Virtual screening and ligand design

The goal of virtual, or *in silico*, screening is to computationally identify small molecules in a compound library that are active against a given target. Virtual screening methods usually adopt either a ligand-based approach, where properties of known active compounds are used to identify additional compounds, or a structure-based approach, where the interactions between putative ligands and the receptor structure are used. Tools for chemical similarity, which can be used for ligand-based screening, are cataloged in the Cheminformatics section.

In contrast to virtual screening, which evaluates predetermined compounds, *de novo* ligand design attempts to create a molecule

Table 15
Open source software for setting up and analyzing molecular simulations.

Name	URL	License	Activity	Citation
AmberTools	http://ambermd.org	GPL	A1	[136]
LOOS	http://loos.sourceforge.net	GPL	A1	[137]
lsfitpar	http://mackerell.umaryland.edu/~kenno/lsfitpar	GPL	A2	[138]
MDAnalysis	http://mdanalysis.org	GPL	A1	[139]
MDTraj	mdtraj.org	LGPL	A1	[140]
MEMBPLUGIN	https://sourceforge.net/projects/membplugin	GPL	C1	[141]
MEPSA	http://bioweb.cbm.uam.es/software/MEPSA	GPL	A3	[142]
MSMBuilder	http://msmbuilder.org	LGPL	A1	[143]
packmol	http://www.ime.unicamp.br/~martinez/packmol	GPL	A1	[144]
PDB2PQR	http://www.poissonboltzmann.org	BSD	A1	[145]
PLUMED	http://www.plumed.org	LGPL	A1	[146]
ProDy	http://prody.csb.pitt.edu	MIT	A1	[147]
Pteros	http://pteros.sourceforge.net	Artistic	B2	[148]
PyEMMA	http://www.emma-project.org	LGPL	A1	[149]
PyRED	http://upjv.q4md-forcefieldtools.org	GPL	C1	[150]
PYTRAJ	https://github.com/Amber-MD/pytraj	GPL	A1	
simpletraj	https://github.com/arose/simpletraj	GPL	A2	
WHAM	http://membrane.urmc.rochester.edu/content/wham	BSD	C1	

Table 16
Open source software for ligand-based virtual screening.

Name	URL	License	Activity	Citation
ACPC	https://github.com/UnixJunkie/ACPC	BSD	B2	[153]
Align-it	http://silicos-it.be	LGPL	C3	
Open3DALIGN	http://open3dalign.sourceforge.net	GPL	B2	[37]
PAPER	https://simtk.org/home/paper	GPL	C2	[154]
Pharmer	http://pharmer.sf.net	GPL	B1	[155]
Pharmit	http://pharmit.sf.net	GPL	A3	[156]
Shape-it	http://silicos-it.be	LGPL	C3	
USRCAT	https://bitbucket.org/aschreyer/usrcat	MIT	C2	[157]

'from scratch' that binds to a protein. Methods differ in how they specify the objective to optimize (e.g., docking score to a protein) and how candidate molecules are created, where a key challenge is maintaining synthetic accessibility.

8.1. Ligand-based (Table 16)

ACPC (AutoCorrelation of Partial Charges) [153] computes ligand similarity based on a rotation and translation invariant electrostatic descriptor.

Align-it is a successor of Pharao [158] and aligns and scores 3D representations of molecules based on their pharmacophore features. It includes a plugin for integration with PyMOL.

Open3DALIGN [37] performs unsupervised rigid-body molecular alignment.

PAPER [154] performs GPU accelerated alignment of molecular shapes using Gaussian overlays.

Pharmer [155] uses efficient data structures to rapidly screen large libraries for ligand conformations that match a pharmacophore.

Pharmit [156] is a successor of Pharmer that also incorporates shape matching and energy minimization (if a receptor structure is available) as part of the screen. It is primarily intended to be used as a backend to a web service.

Shape-it uses Gaussian volumes to align and score molecular shapes.

USRCAT [157] performs "ultra-fast shape recognition" with the addition of pharmacophoric information to rapidly screen compound libraries for similar molecules.

8.2. Docking and scoring (Table 17)

ADplugin is a plugin for PyMOL for interfacing with AutoDock and AutoDock Vina.

APBS [159] performs solvation free energy calculations using the Poisson–Boltzmann implicit solvent method.

AutoDock [160] is an automated docking program that uses a physics-based semiempirical scoring function [182] mapped to atom type grids to evaluate poses and a genetic algorithm to explore the conformational space. It includes the ability to incorporate sidechain flexibility and covalent docking.

AutoDock Vina [161] is an entirely separate code base and approach from Autodock that was developed with a focus on run-time performance and ease of system setup. It uses a fully empirical scoring function and an iterated local search global optimizer to produce docked poses. It includes support for multi-threading and flexible sidechains.

Clusterizer-DockAccessor [162] are tools for accessing the quality of docking protocols. It interfaces with a number of open source and free tools.

DockoMatic [163] provides a graphical user interface for setting up and analyzing AutoDock and AutoDock Vina docking jobs, including when run on a cluster. It also includes the ability to run

inverse virtual screens (find proteins that bind a given ligand) and support for homology model construction.

DOVIS [164] is an extension of AutoDock 4.0 that provides more efficient parallelization of large virtual screening jobs.

idock [165] is a multi-threaded docking program that includes support for the AutoDock Vina scoring function and a random forest scoring function. I can output per-atom free energy information for hotspot detection.

MOLA [166] is a pre-packaged distribution of AutoDock and AutoDock Vina for deployment on multi-platform computing clusters.

NNScore [167] uses a neural network model to score protein–ligand poses.

Paradocks [168] is a parallelized docking program that includes a number of population-based metaheuristics, such as particle swarm optimization, for exploring the space of potential poses.

PyRx [169] is a visual interface for AutoDock and AutoDock Vina that simplifies setting up and analyzing docking workflows. Its future as an open-source solution is in doubt due to a recent shift to commercialization.

rDock [170] is designed for docking against proteins or nucleic acids and can incorporate user-specified constraints. It uses an empirical scoring function that includes solvent accessible surface area terms. A combination of genetic algorithms, Monte Carlo, and simplex minimization is used to explore the conformational space. Distinct scoring functions are provided for docking to proteins and nucleic acids.

RF-Score [171,183] uses a random forest classifier to score protein–ligand poses.

smina [172] is a fork of AutoDock Vina designed to better support energy minimization and custom scoring function development (scoring function terms and atom type properties can be specified using a run-time configuration file). It also simplifies the process of setting up a docking run with flexible sidechains.

VHELIBS (Validation HELper for LIgands and Binding Sites) [173] assists the non-crystallographer in validating ligand geometries with respect to electron density maps.

VinaLC [174] is a fork of AutoDock Vina designed to run on a cluster of multiprocessor machines.

VinaMPI [175] is a wrapper for AutoDock Vina that uses OpenMPI to run large-scale virtual screens on a computing cluster.

Zodiac [176] is a visual interface for structure-based drug design that includes support for haptic feedback.

8.3. Pocket detection (Table 18)

eFindSite [177] using homology modeling and machine learning predicts ligand binding sites in a protein structure.

fpocket [178] detects and delineates protein cavities using Voronoi tessellation and is able to process molecular dynamics simulations.

KVFinder [179] is a PyMOL plugin for identifying and characterizing pockets.

Table 17
Open source software for molecular docking and scoring.

Name	URL	License	Activity	Citation
ADplugin	https://github.com/ADplugin	LGPL	A2	
APBS	http://www.poissonboltzmann.org	BSD	A1	[159]
AutoDock	http://autodock.scripps.edu	GPL	C1	[160]
AutoDock Vina	http://vina.scripps.edu	Apache	C1	[161]
Clusterizer-DockAccessor	http://cheminf.com/software/clusterizer_dockaccessor	GPL	A3	[162]
DockoMatic	https://sourceforge.net/projects/dockomatic	LGPL	B1	[163]
DOVIS	http://bhsai.org/software	GPL	C2	[164]
idock	https://github.com/HongjianLi/idock	Apache	A2	[165]
MOLA	http://www.esa.ipb.pt/~ruiabreu/mola	GPL	C2	[166]
NNScore	http://nbcr.ucsd.edu/data/sw/hosted/nnscore	GPL	C1	[167]
Paradocks	https://github.com/cbaldauf/paradocks	GPL	A2	[168]
PyRx	http://pyrx.sourceforge.net	BSD	A1	[169]
rDock	http://rdock.sourceforge.net	LGPL	C1	[170]
RF-Score	https://github.com/HongjianLi/RF-Score	Apache	A2	[171]
smina	https://sourceforge.net/projects/smina	GPL	A1	[172]
VHELIBS	http://urvnutrigenomica-ctns.github.io/VHELIBS	GPL	A2	[173]
VinaLC	http://mvirdb1.llnl.gov/static_catsid/vina	Apache	C2	[174]
VinaMPI	http://cmb.ornl.gov/~sek	Apache	C2	[175]
Zodiac	https://sourceforge.net/projects/zodiac-zeden	GPL	C1	[176]

Table 18
Open source software for pocket detection.

Name	URL	License	Activity	Citation
eFindSite	http://brylinski.cct.lsu.edu/efindsite	GPL	C2	[177]
fpocket	http://fpocket.sourceforge.net	GPL	C1	[178]
KVFinder	http://lnbio.cnpem.br/facilities/bioinformatics/software-2	GPL	B1	[179]
mcvol	http://www.bisb.uni-bayreuth.de/data/mcvol/mcvol.html	GPL	C2	[180]
PAPCA	https://sourceforge.net/projects/papca	BSD	C2	
PCS	https://sourceforge.net/projects/cavity-search	GPL	C2	
PocketPicker	http://gecco.org.chemie.uni-frankfurt.de/pocketpicker	BSD	C1	[181]
POVME	https://sourceforge.net/projects/povme	GPL	C1	[91]

mcvol [180] calculates protein volumes and identifying cavities using a Monte Carlo algorithm.

PAPCA (PocketAnalyzerPCA) is a pocket detection utility designed to analyze ensembles of protein conformations.

PCS (Pocket Cavity Search) measures the volume of internal cavities and evaluates the environment of ionizable residues.

PocketPicker [181] is a PyMOL plugin that automatically identifies potential ligand binding sites using a grid-based shape descriptor.

POVME (POcket Volume MEasurer) [91] is a tool for measuring and characterizing pocket volumes that includes a graphical user interface.

8.4. Ligand design

AutoClickChem [184] performs *in silico* click chemistry to generate large libraries of synthetically accessible compounds.

AutoGrow [185] uses a genetic algorithm to explore the space of reactants and reactions accessible via AutoClickChem and identifies compounds that dock well using AutoDock Vina.

igrow, like AutoGrow, uses a genetic algorithm but transforms ligands using branch exchange and uses idock as the underlying docking evaluation protocol.

OpenGrowth [186] assembles candidate ligands by connecting small organic fragments in the active site of proteins. It includes a graphical user interface.

9. Discussion

We have cataloged over 200 open-source packages for molecular modeling that provide a wide range of capabilities. As shown in Fig. 1, the most popular license (55%) is some variant of the copyleft GNU Public License, which ensures that derivative works remain
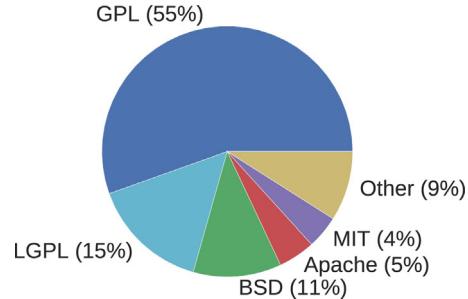
**Fig. 1.** Distribution of open source licenses used in cataloged software packages.

open source. Interestingly 78% of the packages cataloged have a corresponding citeable publication which suggests that much of the software originates from academia. The distribution of average citations generated a year (as reported by Google Scholar) for the citeable publications is shown in Fig. 3. A significant majority (84%) of publications generate at least one citation a year, 29% generate at least 10 citations, and 8% generate more than 100 citations a year on average.

A substantial portion of the packages cataloged are under active development and see significant usage, as shown in Fig. 2. We rated 56% of the packages as 'A' level development, meaning major features or releases were made within the last 18 months, and 51% see substantial usage (rank 1). There are a number of projects (30%) where development has apparently ceased (no changes within the last 18 months). Note that our methodology for identifying packages ignores cases where software is no longer available, this is an underestimate. However, although we did find instances where an open source package was referenced in a paper but was no longer available, we did not find this to be a common occurrence. Most packages, even those that have remained unchanged for a

Table 19
Open source software for ligand design.

Name	URL	License	Activity	Citation
AutoClickChem	https://sourceforge.net/projects/autoclickchem	GPL	C2	[184]
AutoGrow	http://autogrow.ucsd.edu	GPL	A1	[185]
igrow	https://github.com/HongjianLi/igrow	Apache	A3	
OpenGrowth	http://opengrowth.sf.net	GPL	A1	[186]
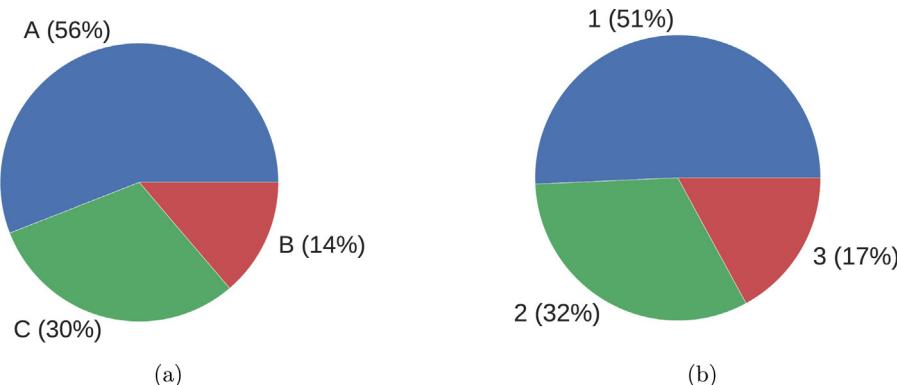
Fig. 2. Activity distributions of cataloged software packages. (a) Distribution of development activity. (b) Distribution of user activity.
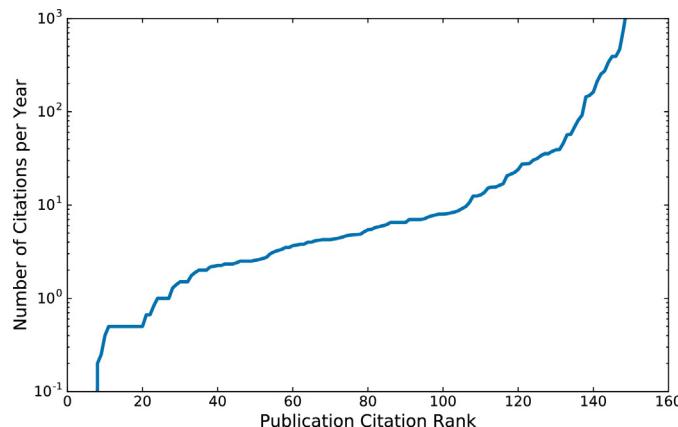
Fig. 3. Distribution of citations as reported by Google Scholar generated on average every year by those software packages with citeable publications.

decade, see some usage. In fact, a number of packages (23), still see significant usage despite having received no development for the past 18 months. This underlies the importance of releasing source code through a third-party site such as SourceForge or GitHub as it ensures the continued existence of a project.

A major advantage of open source is that in cases where a popular project is not being actively developed (e.g., AutoDock Vina [161]) new projects can fork the source code and continue development (e.g., smina [172]). However, a potential problem area with open source development is the lack of central coordination and efficient pooling of resources. For example, there are several forks of AutoDock Vina that improve it's performance on computing clusters and there are an array of tools in several categories that effectively perform the same task. This underscores the importance of efforts like Blue Obelisk [187,188] and Open Chemistry (http://www.openchemistry.org) which foster collaboration among open source cheminformatics projects.

It is clear that open source software plays an important role in the scientific community and is a vibrant sub-community of its own with a wide assortment of projects under development and in widespread use. The open source software packages cataloged here provide launching points for the development of new tools for enabling further scientific discovery. As open source software is in a constant state of change, as a supplement to this article we provide a website, https://opensourcemolecularmodeling.github.io with an updated and community-editable catalog of open source molecular modeling software.

Acknowledgements

References

[1] A.I. Krylov, J.M. Herbert, F. Furche, M. Head-Gordon, P.J. Knowles, R. Lindh, F.R. Manby, P. Pulay, C.-K. Skylaris, H.-J. Werner, What is the price of open-source software? J. Phys. Chem. Lett. 6 (14) (2015) 2751–2754, http://dx.doi.org/10.1021/acs.jpclett.5b01258.

[2] J.D. Gezelter, Open source and open data should be standard practices, J. Phys. Chem. Lett. 6 (7) (2015) 1168–1169, http://dx.doi.org/10.1021/acs.jpclett.5b00285.

[3] C.R. Jacob, How open is commercial scientific software? J. Phys. Chem. Lett. 7 (2) (2016) 351–353, http://dx.doi.org/10.1021/acs.jpclett.5b02609.

[4] B.O. Villoutreix, D. Lagorce, C.M. Labbé, O. Sperandio, M.A. Miteva, One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade, Drug Discov. Today 18 (21–22) (2013) 1081–1089, http://dx.doi.org/10.1016/j.drudis.2013.06.013.

[5] A.R. Leach, V.J. Gillet, An Introduction to Chemoinformatics, Springer Science + Business Media, 2007, http://dx.doi.org/10.1007/978-1-4020-6291-9.

[6] J. Gasteiger, Introduction, in: Chemoinformatics, Wiley-Blackwell, 2003, pp. 1–13, http://dx.doi.org/10.1002/3527601643.ch1.

[7] J.-P. Ebejer, G.M. Morris, C.M. Deane, Freely available conformer generation methods: how good are they? J. Chem. Inf. Model. 52 (5) (2012) 1146–1158, http://dx.doi.org/10.1021/ci2004658.

[8] A. Hildebrandt, A. Dehof, A. Rurainski, A. Bertsch, M. Schumann, N.C. Toussaint, A. Moll, D. Stöckel, S. Nickels, S.C. Mueller, H.-P. Lenhof, O. Kohlbacher, BALL – biochemical algorithms library 1.3, BMC Bioinform. 11 (1) (2010) 531, http://dx.doi.org/10.1186/1471-2105-11-531.

[9] C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, E. Willighagen, Recent developments of the chemistry development kit (CDK) – an open-source Java library for chemo- and bioinformatics, CPD 12 (17) (2006) 2111–2120, http://dx.doi.org/10.2174/138161206777585274.

[10] S. Höck, R. Riedl, chemf: a purely functional chemistry toolkit, J. Cheminform. 4 (1) (2012) 38, http://dx.doi.org/10.1186/1758-2946-4-38.

[11] A. Dalke, chemfp-fast and portable fingerprint formats and tools., J. Cheminform. 3 (S-1) (2011) 12.

[12] Y. Cao, A. Charisi, L.-C. Cheng, T. Jiang, T. Girke, ChemmineR: a compound mining framework for R, Bioinformatics 24 (15) (2008) 1733–1734, http://dx.doi.org/10.1093/bioinformatics/btn307.

[13] N.M. O'Boyle, G.R. Hutchison, Cinfony – combining open source cheminformatics toolkits behind a common interface, Chem. Cent. J. 2 (1) (2008) 24, http://dx.doi.org/10.1186/1752-153x-2-24.

[14] A. Drefahl, CurlySMILES: a chemical language to customize and annotate encodings of molecular and nanodevice structures, J. Cheminform. 3 (1) (2011) 1, http://dx.doi.org/10.1186/1758-2946-3-1.

[15] M. Wójcikowski, P. Zielenkiewicz, P. Siedlecki, DiSCuS: an open platform for (not only) virtual screening results management, J. Chem. Inf. Model. 54 (1) (2014) 347–354, http://dx.doi.org/10.1021/ci400587f.

[16] A. Supady, V. Blum, C. Baldauf, First-principles molecular structure search with a genetic algorithm, J. Chem. Inf. Model. 55 (11) (2015) 2338–2348, http://dx.doi.org/10.1021/acs.jcim.5b00243.

[17] Y. Wang, T.W.H. Backman, K. Horan, T. Girke, fmcsR: mismatch tolerant maximum common substructure searching in R, Bioinformatics 29 (21) (2013) 2792–2794, http://dx.doi.org/10.1093/bioinformatics/btt475.

[18] D. Pavlov, M. Rybalkin, B. Karulin, M. Kozhevnikov, A. Savelyev, A. Churinov, Indigo: universal cheminformatics API, J. Cheminform. 3 (Suppl. 1) (2011) P4, http://dx.doi.org/10.1186/1758-2946-3-s1-p4.

[19] K.R. Lawson, J. Lawson, LICSS – a chemical spreadsheet in microsoft excel, J. Cheminform. 4 (1) (2012) 3, http://dx.doi.org/10.1186/1758-2946-4-3.

[20] M. Sud, MayaChemTools: an open source package for computational discovery, in: 243rd ACS National Meeting & Exposition, San Diego, CA, 2012.

[21] M. Wójcikowski, P. Zielenkiewicz, P. Siedlecki, Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0078-2.

[22] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminform. 3 (1) (2011) 33, http://dx.doi.org/10.1186/1758-2946-3-33.

[23] D.M. Lowe, P.T. Corbett, P. Murray-Rust, R.C. Glen, Chemical name to structure: OPSIN an open source solution, J. Chem. Inf. Model. 51 (3) (2011) 739–753, http://dx.doi.org/10.1021/ci100384d.

[24] M. Rijnbeek, C. Steinbeck, OrChem – an open source chemistry search engine for Oracle®, J. Cheminform. 1 (1) (2009) 17, http://dx.doi.org/10.1186/1758-2946-1-17.

[25] I.V. Filippov, M.C. Nicklaus, Optical structure recognition software to recover chemical information: OSRA an open source solution, J. Chem. Inf. Model. 49 (3) (2009) 740–743, http://dx.doi.org/10.1021/ci800067r.

[26] N.M. O'Boyle, C. Morley, G.R. Hutchison, Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit, Chem. Cent. J. 2 (1) (2008) 5, http://dx.doi.org/10.1186/1752-153x-2-5.

[27] R. Guha, et al., Chemical informatics functionality in R, J. Stat. Softw. 18 (5) (2007) 1–16.

[28] R. Smith, R. Williamson, D. Ventura, J.T. Prince, Rubabel: wrapping open Babel with Ruby, J. Cheminform. 5 (1) (2013) 35, http://dx.doi.org/10.1186/1758-2946-5-35.

[29] S. Rahman, M. Bashton, G.L. Holliday, R. Schrader, J.M. Thornton, Small Molecule Subgraph Detector (SMSD) toolkit, J. Cheminform. 1 (1) (2009) 12, http://dx.doi.org/10.1186/1758-2946-1-12.

[30] P. Amani, T. Sneyd, S. Preston, N.D. Young, L. Mason, U.-M. Bailey, J. Baell, D. Camp, R.B. Gasser, A.-D. Gorse, P. Taylor, A. Hofmann, A practical Java tool for small-molecule compound appraisal, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0079-1.

[31] N. Haider, Functionality pattern matching as an efficient complementary structure/reaction search tool: an open-source approach, Molecules 15 (8) (2010) 5079–5092, http://dx.doi.org/10.3390/molecules15085079.

[32] M.A. Miteva, F. Guyon, P. Tuffery, Frog2: efficient 3D conformation ensemble generator for small compounds, Nucleic Acids Res. 38 (Web Server) (2010) W622–W627, http://dx.doi.org/10.1093/nar/gkq325.

[33] R.F. Bruns, I.A. Watson, Rules for identifying potentially reactive or promiscuous compounds, J. Med. Chem. 55 (22) (2012) 9763–9772, http://dx.doi.org/10.1021/jm301008n.

[34] D. Hoksza, P. Škoda, M. Voršilák, D. Svozil, Molpher: a software framework for systematic chemical space exploration, J. Cheminform. 6 (1) (2014) 7, http://dx.doi.org/10.1186/1758-2946-6-7.

[35] C. Borgelt, T. Meinl, M. Berthold, MoSS, in: Proceedings of the 1st International Workshop on Open Source Data Mining Frequent Pattern Mining Implementations – OSDM '05, Association for Computing Machinery (ACM), 2005, http://dx.doi.org/10.1145/1133905.1133908.

[36] J.E. Peironcely, M. Rojas-Chertó, D. Fichera, T. Reijmers, L. Coulier, J.-L. Faulon, T. Hankemeier, OMG: open molecule generator, J. Cheminform. 4 (1) (2012) 21, http://dx.doi.org/10.1186/1758-2946-4-21.

[37] P. Tosco, T. Balle, F. Shiri, SDF2XYZ2SDF: how to exploit TINKER power in cheminformatics projects, J. Mol. Model. 17 (11) (2011) 3021–3023, http://dx.doi.org/10.1007/s00894-011-1111-7.

[38] J. Rosen, L. Miguet, S. Pérez, Shape: automatic conformation prediction of carbohydrates using a genetic algorithm, J. Cheminform. 1 (1) (2009) 16, http://dx.doi.org/10.1186/1758-2946-1-16.

[39] N. Jeliazkova, V. Jeliazkov, AMBIT RESTful web services: an implementation of the OpenTox application programming interface, J. Cheminform. 3 (1) (2011) 18, http://dx.doi.org/10.1186/1758-2946-3-18.

[40] O. Spjuth, J. Alvarsson, A. Berg, M. Eklund, S. Kuhn, C. Mäsak, G. Torrance, J. Wagener, E.L. Willighagen, C. Steinbeck, J.E. Wikberg, Bioclipse 2: a scriptable integration platform for the life sciences, BMC Bioinform. 10 (1) (2009) 397, http://dx.doi.org/10.1186/1471-2105-10-397.

[41] J. Goecks, A. Nekrutenko, J. Taylor, T.G. Team, Galaxy: a comprehensive approach for supporting accessible reproducible, and transparent computational research in the life sciences, Genome Biol. 11 (8) (2010) R86, http://dx.doi.org/10.1186/gb-2010-11-8-r86.

[42] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel, KNIME – the Konstanz information miner, SIGKDD Explor. Newsl. 11 (1) (2009) 26, http://dx.doi.org/10.1145/1656274.1656280.

[43] J. Demšar, T. Curk, A. Erjavec, Črt Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, B. Zupan, Orange: data mining toolbox in python, J. Mach. Learn. Res. 14 (2013) 2349–2353 http://jmlr.org/papers/v14/demsar13a.html.

[44] V. Guilloux, A. Arrault, L. Colliandre, S. Bourg, P. Vayer, L. Morin-Allory, Mining collections of compounds with Screening Assistant 2, J. Cheminform. 4 (1) (2012) 20, http://dx.doi.org/10.1186/1758-2946-4-20.

[45] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J. Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A.N. de la Hidalga, M.P.B. Vargas, S. Sufi, C. Goble, The Taverna workflow suite: designing and executing workflows of Web Services on the desktop web or in the cloud, Nucleic Acids Res. 41 (W1) (2013) W557–W561, http://dx.doi.org/10.1093/nar/gkt328.

[46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software, SIGKDD Explor. Newsl. 11 (1) (2009) 10, http://dx.doi.org/10.1145/1656274.1656278.

[47] N.M. O'Boyle, T. Vandermeersch, C.J. Flynn, A.R. Maguire, G.R. Hutchison, Confab – systematic generation of diverse low-energy conformers, J. Cheminform. 3 (1) (2011) 8, http://dx.doi.org/10.1186/1758-2946-3-8.

[48] N.T. Kochev, V.H. Paskaleva, N. Jeliazkova, Ambit-Tautomer: an open source tool for tautomer generation, Mol. Inf. 32 (5–6) (2013) 481–504, http://dx.doi.org/10.1002/minf.201200133.

[49] S. Beisken, T. Meinl, B. Wiswedel, L.F. de Figueiredo, M. Berthold, C. Steinbeck, KNIME-CDK: workflow-driven cheminformatics, BMC Bioinform. 14 (1) (2013) 257, http://dx.doi.org/10.1186/1471-2105-14-257.

[50] T. Kuhn, E.L. Willighagen, A. Zielesny, C. Steinbeck, CDK-Taverna: an open workflow environment for cheminformatics, BMC Bioinform. 11 (1) (2010) 159, http://dx.doi.org/10.1186/1471-2105-11-159.

[51] S. Krause, E. Willighagen, C. Steinbeck, JChemPaint – using the collaborative forces of the internet to develop a free editor for 2D chemical structures, Molecules 5 (1) (2000) 93–98, http://dx.doi.org/10.3390/50100093.

[52] S. Caboche, LeView: automatic and interactive generation of 2D diagrams for biomacromolecule/ligand interactions, J. Cheminform. 5 (1) (2013) 40, http://dx.doi.org/10.1186/1758-2946-5-40.

[53] E.K. Brefo-Mensah, M. Palmer, mol2chemfig a tool for rendering chemical structures from molfile or SMILES format to LATE X code, J. Cheminform. 4 (1) (2012) 24, http://dx.doi.org/10.1186/1758-2946-4-24.

[54] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor visualization, and analysis platform, J. Cheminform. 4 (1) (2012) 17, http://dx.doi.org/10.1186/1758-2946-4-17.

[55] A. Moll, A. Hildebrandt, H.-P. Lenhof, O. Kohlbacher, BALLView: a tool for research and education in molecular modeling, Bioinformatics 22 (3) (2005) 365–366, http://dx.doi.org/10.1093/bioinformatics/bti318.

[56] G. Kovačević, V. Veryazov, Luscus: molecular viewer and editor for MOLCAS, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0060-z.

[57] M. Norrby, C. Grebner, J. Eriksson, J. Boström, Molecular rift: virtual reality for drug designers, J. Chem. Inf. Model. 55 (11) (2015) 2475–2484, http://dx.doi.org/10.1021/acs.jcim.5b00544.

[58] M. Biasini, T. Schmidt, S. Bienert, V. Mariani, G. Studer, J. Haas, N. Johner, A.D. Schenk, A. Philippsen, T. Schwede, OpenStructure: an integrated software framework for computational structural biology, Acta Crystallogr. D Biol. Cryst. 69 (5) (2013) 701–709, http://dx.doi.org/10.1107/s0907444913007051.

[59] S. Salentin, S. Schreiber, V.J. Haupt, M.F. Adasme, M. Schroeder, PLIP: fully automated protein–ligand interaction profiler, Nucleic Acids Res. 43 (W1) (2015) W443–W447, http://dx.doi.org/10.1093/nar/gkv315.

[60] D.J. Sweeney, A Computational Tool for Biomolecular Structure Analysis Based on Chemical and Enzymatic Modification of Native Proteins (Ph.D. thesis), Wright State University, 2011.

[61] M. Tarini, P. Cignoni, C. Montani, Ambient occlusion and edge cueing for enhancing real time molecular visualization, IEEE Trans. Visual. Comput. Graphics 12 (5) (2006) 1237–1244, http://dx.doi.org/10.1109/tvcg.2006.115.

[62] N. Rego, D. Koes, 3Dmol.js: molecular visualization with WebGL, Bioinformatics 31 (8) (2014) 1322–1324, http://dx.doi.org/10.1093/bioinformatics/btu829.

[63] C.W. Earley, CH5M3D: an HTML5 program for creating 3D molecular structures, J. Cheminform. 5 (1) (2013) 46, http://dx.doi.org/10.1186/1758-2946-5-46.

[64] M. Mohebifar, F. Sajadi, Chemozart: a web-based 3D molecular structure editor and visualizer platform, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0101-7.

[65] M.C. Burger, ChemDoodle Web Components: HTML5 toolkit for chemical graphics interfaces, and informatics, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0085-3.

[66] B. Bienfait, P. Ertl, JSME: a free molecule editor in JavaScript, J. Cheminform. 5 (1) (2013) 24, http://dx.doi.org/10.1186/1758-2946-5-24.

[67] R.M. Hanson, Jmol – a paradigm shift in crystallographic visualization, J. Appl. Cryst. 43 (5) (2010) 1250–1260, http://dx.doi.org/10.1107/s0021889810030256.

[68] R.M. Hanson, J. Prilusky, Z. Renjian, T. Nakane, J.L. Sussman, JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia, Israel J. Chem. 53 (3–4) (2013) 207–216, http://dx.doi.org/10.1002/ijch.201300024.

[69] A.S. Rose, P.W. Hildebrand, NGL Viewer: a web application for molecular visualization, Nucleic Acids Res. 43 (W1) (2015) W576–W579, http://dx.doi.org/10.1093/nar/gkv402.

[70] M. Biasini, pv: v1.8.1, 2015, http://dx.doi.org/10.5281/zenodo.20980.

[71] , in: Exploring QSAR, Anal. Chem. 67 (17) (1995) 563A, http://dx.doi.org/10.1021/ac00113a734.

[72] A. Jahn, G. Hinselmann, N. Fechner, C. Henneges, A. Zell, in: Probabilistic modeling of conformational space for 3D machine learning approaches, Mol. Inf. 29 (5) (2010) 441–455, http://dx.doi.org/10.1002/minf.201000036.

[73] P. Carbonell, L. Carlsson, J.-L. Faulon, in: Stereo signature molecular descriptor, J. Chem. Inform. Model. 53 (4) (2013) 887–897, http://dx.doi.org/10.1021/ci300584r.

[74] C.W. Yap, in: PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (7) (2010) 1466–1474, http://dx.doi.org/10.1002/jcc.21707.

[75] J.L. Melville, J.D. Hirst, in: TMACC: interpretable correlation descriptors for quantitative structure–activity relationships, J. Chem. Inf. Model. 47 (2) (2007) 626–634, http://dx.doi.org/10.1021/ci6004178.

[76] J.C. Stålring, L.A. Carlsson, P. Almeida, S. Boyer, in: AZOrange – high performance open source machine learning for QSAR modeling in a graphical programming environment, J. Cheminform. 3 (1) (2011) 28, http://dx.doi.org/10.1186/1758-2946-3-28.

[77] I. Cortes-Ciriano, in: Bioalerts: a python library for the derivation of structural alerts from bioactivity and toxicity data sets, J. Cheminform. 8 (1) (2016), http://dx.doi.org/10.1186/s13321-016-0125-7.

[78] D.S. Murrell, I. Cortes-Ciriano, G.J.P. van Westen, I.P. Stott, A. Bender, T.E. Malliavin, R.C. Glen, in: Chemically Aware Model Builder (camb): an R package for property and bioactivity modelling of small molecules, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0086-2.

[79] P. Carrió, O. López, F. Sanz, M. Pastor, in: eTOXlab an open source modeling framework for implementing predictive models in production environments, J. Cheminform. 7 (1) (2015) 8, http://dx.doi.org/10.1186/s13321-015-0058-6.

[80] P. Tosco, T. Balle, in: Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields, J. Mol. Model. 17 (1) (2010) 201–208, http://dx.doi.org/10.1007/s00894-010-0684-x.

[81] P. Rydberg, M. Rostkowski, D.E. Gloriam, L. Olsen, in: The contribution of atom accessibility to site of metabolism models for cytochromes P450, Mol. Pharm. 10 (4) (2013) 1216–1223, http://dx.doi.org/10.1021/mp3005116.

[82] G. Patlewicz, N. Jeliazkova, R. Safford, A. Worth, B. Aleksiev, in: An evaluation of the implementation of the Cramer classification scheme in the Toxtree software, SAR QSAR Environ. Res. 19 (5–6) (2008) 495–524, http://dx.doi.org/10.1080/10629360802083871.

[83] A. Lusci, G. Pollastri, P. Baldi, in: Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules, J. Chem. Inf. Model. 53 (7) (2013) 1563–1575, http://dx.doi.org/10.1021/ci400187y.

[84] M. Gütlein, A. Karwath, S. Kramer, in: CheS-Mapper 2.0 for visual validation of (Q)SAR models, J. Cheminform. 6 (1) (2014), http://dx.doi.org/10.1186/s13321-014-0041-7.

[85] T. Sander, J. Freyss, M. von Korff, C. Rufener, in: DataWarrior: an open-source program for chemistry aware data visualization and analysis, J. Chem. Inf. Model. 55 (2) (2015) 460–473, http://dx.doi.org/10.1021/ci500588j.

[86] A. Cereto-Massague, L. Guasch, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallve, in: DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets, Bioinformatics 28 (12) (2012) 1661–1662, http://dx.doi.org/10.1093/bioinformatics/bts249.

[87] S. Wetzel, K. Klein, S. Renner, D. Rauh, T.I. Oprea, P. Mutzel, H. Waldmann, in: Interactive exploration of chemical space with Scaffold Hunter, Nat. Chem. Biol. 5 (8) (2009) 581–583, http://dx.doi.org/10.1038/nchembio.187.

[88] R. Lewis, R. Guha, T. Korcsmaros, A. Bender, in: Synergy Maps: exploring compound combinations using network-based visualization, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0090-6.

[89] M.J. Martínez, I. Ponzoni, M.F. Díaz, G.E. Vazquez, A.J. Soto, in: Visual analytics in cheminformatics: user-supervised descriptor selection for QSAR methods, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0092-4.

[90] P. Ertl, L. Patiny, T. Sander, C. Rufener, M. Zasso, in: Wikipedia Chemical Structure Explorer: substructure and similarity searching of molecules from Wikipedia, J. Cheminform. 7 (1) (2015), http://dx.doi.org/10.1186/s13321-015-0061-y.

[91] J.D. Durrant, L. Votapka, J. Sørensen, R.E. Amaro, in: POVME 2.0: an enhanced tool for determining pocket shape and volume characteristics, J. Chem.

[92] P. Rydberg, D.E. Gloriam, J. Zaretzki, C. Breneman, L. Olsen, in: SMARTCyp: a 2D method for prediction of cytochrome p450-mediated drug metabolism, ACS Med. Chem. Lett. 1 (3) (2010) 96–100, http://dx.doi.org/10.1021/ml100016x.

[93] M. Gütlein, A. Karwath, S. Kramer, in: CheS-Mapper – chemical space mapping and visualization in 3D, J. Cheminform. 4 (1) (2012) 7, http://dx.doi.org/10.1186/1758-2946-4-7.

[94] M.D. Vivo, in: Bridging quantum mechanics and structure-based drug design, Front. Biosci. 16 (1) (2011) 1619, http://dx.doi.org/10.2741/3809.

[95] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, P. Ghosez, M. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah, J. Zwanziger, in: ABINIT: first-principles approach to material and nanosystem properties, Comput. Phys. Commun. 180 (12) (2009) 2582–2615, http://dx.doi.org/10.1016/j.cpc.2009.07.007.

[96] V. Lotrich, N. Flocke, M. Ponton, A.D. Yau, A. Perera, E. Deumens, R.J. Bartlett, in: Parallel implementation of electronic structure energy gradient, and Hessian calculations, J. Chem. Phys. 128 (19) (2008) 194104, http://dx.doi.org/10.1063/1.2920482.

[97] S. Mohr, L.E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker, T. Deutsch, in: Accurate and efficient linear scaling DFT calculations with universal applicability, Phys. Chem. Chem. Phys. 17 (47) (2015) 31360–31370, http://dx.doi.org/10.1039/c5cp00437c.

[98] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, in: cp2k: atomistic simulations of condensed matter systems, Wiley Interdisc. Rev.: Comput. Mol. Sci. 4 (1) (2013) 15–25, http://dx.doi.org/10.1002/wcms.1159.

[99] S. Bahn, K. Jacobsen, in: An object-oriented scripting interface to a legacy electronic structure code, Comput. Sci. Eng. 4 (3) (2002) 56–66, http://dx.doi.org/10.1109/5992.998641.

[100] E. Rudberg, E.H. Rubensson, P. Sałek, in: Kohn-Sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage, J. Chem. Theory Comput. 7 (2) (2011) 340–350, http://dx.doi.org/10.1021/ct100611z.

[101] J. Lehtola, M. Hakala, A. Sakko, K. Hämäläinen, in: ERKALE – a flexible program package for X-ray properties of atoms and molecules, J. Comput. Chem. 33 (18) (2012) 1572–1585, http://dx.doi.org/10.1002/jcc.22987.

[102] J. Enkovaara, C. Rostgaard, J.J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H.A. Hansen, H.H. Kristoffersen, M. Kuisma, A.H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P.G. Moses, J. Ojanen, T. Olsen, V. Petzold, N.A. Romero, J. Stausholm-Møller, M. Strange, G.A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G.K.H. Madsen, R.M. Nieminen, J.K. Nørskov, M. Puska, T.T. Rantala, J. Schiøtz, K.S. Thygesen, K.W. Jacobsen, in: Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method, J. Phys.: Condens. Matter 22 (25) (2010) 253202 http://stacks.iop.org/0953-8984/22/i=25/a=253202.

[103] T.Y. Nikolaienko, L.A. Bulavin, D.M. Hovorun, in: JANPA: an open source cross-platform implementation of the Natural Population Analysis on the Java platform, Comput. Theoret. Chem. 1050 (2014) 15–22, http://dx.doi.org/10.1016/j.comptc.2014.10.002.

[104] C. Janssen, E. Seidl, M. Colvin, Object-oriented implementation of parallel ab initio programs, in: ACS Symposium Series, Parallel Computing in Computational Chemistry, vol. 592, 1995.

[105] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H.V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, W. de Jong, in: NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations, Comput. Phys. Commun. 181 (9) (2010) 1477–1489, http://dx.doi.org/10.1016/j.cpc.2010.04.018.

[106] X. Andrade, D. Strubbe, U.D. Giovannini, A.H. Larsen, M.J.T. Oliveira, J. Alberdi-Rodriguez, A. Varas, I. Theophilou, N. Helbig, M.J. Verstraete, L. Stella, F. Nogueira, A. Aspuru-Guzik, A. Castro, M.A.L. Marques, A. Rubio, in: Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems, Phys. Chem. Chem. Phys. 17 (47) (2015) 31371–31396, http://dx.doi.org/10.1039/c5cp00351b.

[107] T. Ozaki, H. Kino, in: Efficient projector expansion for the ab initio LCAO method, Phys. Rev. B 72 (4) (2005), http://dx.doi.org/10.1103/physrevb.72.045121.

[108] J.M. Turney, A.C. Simmonett, R.M. Parrish, E.G. Hohenstein, F.A. Evangelista, J.T. Fermann, B.J. Mintz, L.A. Burns, J.J. Wilke, M.L. Abrams, N.J. Russ, M.L. Leininger, C.L. Janssen, E.T. Seidl, W.D. Allen, H.F. Schaefer, R.A. King, E.F. Valeev, C.D. Sherrill, T.D. Crawford, in: Psi4: an open-source ab initio electronic structure program, Wiley Interdisc. Rev. Comput. Mol. Sci. 2 (4) (2011) 556–565, http://dx.doi.org/10.1002/wcms.93.

[109] J. Kim, K. Esler, J. McMinis, D.M. Ceperley, in: Quantum monte carlo algorithms: making most of large-scale multi/many-core clusters, Conf. Ser. (Chattanooga, TN), Sci. Discov. Adv. Comput. (SciDac), J. Physics (2010).

[110] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, A.D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A.P. Seitsonen, A. Smogunov, P. Umari, R.M. Wentzcovitch, in: QUANTUM ESPRESSO: a

modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter 21 (39) (2009) 395502, http://dx.doi.org/10.1088/0953-8984/21/39/395502.

[111] S. Moore, E. Briggs, M. Hodak, W. Lu, J. Bernholc, C.-W. Lee, Scaling the RMG quantum mechanics code, in: Proceedings of the Extreme Scaling Workshop, University of Illinois at Urbana-Champaign, 2012, p. 8.

[112] C. Steinmann, M.W. Ibsen, A.S. Hansen, J.H. Jensen, in: FragIt: a tool to prepare input files for fragment based quantum chemical calculations, PLoS ONE 7 (9) (2012) e44480, http://dx.doi.org/10.1371/journal.pone.0044480.

[113] N.M. O'boyle, A.L. Tenderholt, K.M. Langner, in: cclib. A library for package-independent computational chemistry algorithms, J. Comput. Chem 29 (5) (2008) 839–845, http://dx.doi.org/10.1002/jcc.20823.

[114] T. Etienne, X. Assfeld, A. Monari, in: New insight into the topology of excited states through detachment/attachment density matrices-based centroids of charge, J. Chem. Theory Comput. 10 (9) (2014) 3906–3914, http://dx.doi.org/10.1021/ct500400s.

[115] G. Hermann, V. Pohl, J.C. Tremblay, B. Paulus, H.-C. Hege, A. Schild, ORBKIT-A Modular Python Toolbox for Cross-Platform Post-Processing of Quantum Chemical Wavefunction Data, 2016 arXiv:1601.03069.

[116] A.-R. Allouche, in: Gabedit – a graphical user interface for computational chemistry softwares, J. Comput. Chem. 32 (1) (2010) 174–182, http://dx.doi.org/10.1002/jcc.21600.

[117] P. Canepa, R.M. Hanson, P. Ugliengo, M. Alfredsson, in: J-ICE: a new Jmol interface for handling and visualizing crystallographic and electronic properties, J. Appl. Cryst. 44 (1) (2010) 225–229, http://dx.doi.org/10.1107/s0021889810049411.

[118] B.M. Bode, M.S. Gordon, in: Macmolplt: a graphical user interface for GAMESS, J. Mol. Graph. Modell. 16 (3) (1998) 133–138, http://dx.doi.org/10.1016/s1093-3263(99)00002-9.

[119] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S.A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, R. Schneider, in: Daubechies wavelets as a basis set for density functional pseudopotential calculations, J. Chem. Phys. 129 (1) (2008) 014109, http://dx.doi.org/10.1063/1.2949547.

[120] S. Mohr, L.E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, S. Goedecker, in: Daubechies wavelets for linear scaling density functional theory, J. Chem. Phys. 140 (20) (2014) 204110, http://dx.doi.org/10.1063/1.4871876.

[121] P.E. Blöchl, in: Projector augmented-wave method, Phys. Rev. B 50 (24) (1994) 17953–17979, doi:10.1103/physrevb.50.17953. URL http://dx.doi.org/10.1103/physrevb.50.17953.

[122] G. Kresse, D. Joubert, in: From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59 (3) (1999) 1758–1775, http://dx.doi.org/10.1103/physrevb.59.1758.

[123] M.A. Marques, M.J. Oliveira, T. Burnus, in: Libxc: a library of exchange and correlation functionals for density functional theory, Comput. Phys. Commun. 183 (10) (2012) 2272–2281, http://dx.doi.org/10.1016/j.cpc.2012.05.007.

[124] A. Vitalis, R.V. Pappu, in: Methods for Monte Carlo simulations of biomacromolecules, Annu. Rep. Comput. Chem. 5 (2009) 49–76, http://dx.doi.org/10.1016/S1574-1400(09)00503-9.

[125] W. Smith, C. Yong, P. Rodger, in: DL_POLY: application to molecular simulation, Mol. Simul. 28 (5) (2002) 385–471, http://dx.doi.org/10.1080/08927020290018769.

[126] Y.-L. Zhu, H. Liu, Z.-W. Li, H.-J. Qian, G. Milano, Z.-Y. Lu, in: GALAMOST: GPU-accelerated large-scale molecular simulation toolkit, J. Comput. Chem. 34 (25) (2013) 2197–2211, http://dx.doi.org/10.1002/jcc.23365.

[127] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, in: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (7) (2013) 845–854, http://dx.doi.org/10.1093/bioinformatics/btt055.

[128] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, P. Tavan, in: Optimizing the accuracy and efficiency of fast hierarchical multipole expansions for MD simulations, J. Chem. Theory Comput. 8 (10) (2012) 3628–3636, http://dx.doi.org/10.1021/ct300080n.

[129] S. Plimpton, in: Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19, http://dx.doi.org/10.1006/jcph.1995.1039.

[130] A.P. Lyubartsev, A. Laaksonen, in: M. DynaMix – a scalable portable parallel MD simulation package for arbitrary molecular mixtures, Comput. Phys. Commun. 128 (3) (2000) 565–589, http://dx.doi.org/10.1016/s0010-4655(99)00529-9.

[131] K. Hinsen, in: The molecular modeling toolkit: a new approach to molecular simulations, J. Comput. Chem. 21 (2) (2000) 79–85, http://dx.doi.org/10.1002/(sici)1096-987x(20000130)21:2<79::aid-jcc1>3.0.co;2-b.

[132] P. Eastman, M.S. Friedrichs, J.D. Chodera, R.J. Radmer, C.M. Bruns, J.P. Ku, K.A. Beauchamp, T.J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M.R. Shirts, V.S. Pande, in: OpenMM 4: a reusable extensible, hardware independent library for high performance molecular simulation, J. Chem. Theory Comput. 9 (1) (2013) 461–469, http://dx.doi.org/10.1021/ct300857j.

[133] T. Matthey, T. Cickovski, S. Hampton, A. Ko, Q. Ma, M. Nyerges, T. Raeder, T. Slabach, J.A. Izaguirre, in: ProtoMol an object-oriented framework for prototyping novel algorithms for molecular dynamics, ACM Trans. Math. Softw. 30 (3) (2004) 237–265, http://dx.doi.org/10.1145/1024074.1024075.

[134] J. Michel, R.D. Taylor, J.W. Essex, in: Efficient generalized Born models for Monte Carlo simulations, J. Chem. Theory Comput. 2 (3) (2006) 732–739, http://dx.doi.org/10.1021/ct600069r.

[135] M.C. Zwier, J.L. Adelman, J.W. Kaus, A.J. Pratt, K.F. Wong, N.B. Rego, E. Suárez, S. Lettieri, D.W. Wang, M. Grabe, D.M. Zuckerman, L.T. Chong, in: WESTPA: an interoperable highly scalable software package for weighted ensemble simulation and analysis, J. Chem. Theory Comput. 11 (2) (2015) 800–809, http://dx.doi.org/10.1021/ct5010615.

[136] R. Salomon-Ferrer, D.A. Case, R.C. Walker, in: An overview of the Amber biomolecular simulation package, Wiley Interdisc. Rev.: Comput. Mol. Sci. 3 (2) (2012) 198–210, http://dx.doi.org/10.1002/wcms.1121.

[137] T.D. Romo, N. Leioatts, A. Grossfield, in: Lightweight object oriented structure analysis: tools for building tools to analyze molecular dynamics simulations, J. Comput. Chem. 35 (32) (2014) 2305–2318, http://dx.doi.org/10.1002/jcc.23753.

[138] K. Vanommeslaeghe, M.A.D.M. Yang Jr., in: Robustness in the fitting of molecular mechanics parameters, J. Comput. Chem. 36 (14) (2015) 1083–1101, http://dx.doi.org/10.1002/jcc.23897.

[139] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, in: MDAnalysis: a toolkit for the analysis of molecular dynamics simulations, J. Comput. Chem. 32 (10) (2011) 2319–2327, http://dx.doi.org/10.1002/jcc.21787.

[140] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C.X. Hernández, C.R. Schwantes, L.-P. Wang, T.J. Lane, V.S. Pande, in: MDTraj: a modern open library for the analysis of molecular dynamics trajectories, Biophys. J. 109 (8) (2015) 1528–1532, http://dx.doi.org/10.1016/j.bpj.2015.08.015.

[141] R. Guixà-González, I. Rodriguez-Espigares, J.M. Ramírez-Anguita, P. Carrió-Gaspar, H. Martinez-Seara, T. Giorgino3, J. Selent, in: MEMBPLUGIN: studying membrane complexity in VMD, Bioinformatics 30 (10) (2014) 1478–1480, http://dx.doi.org/10.1093/bioinformatics/btu037.

[142] I. Marcos-Alcalde, J. Setoain, J.I. Mendieta-Moreno, J. Mendieta, P. Gómez-Puertas, in: MEPSA: minimum energy pathway analysis for energy landscapes: Fig. 1, Bioinformatics (2015) btv453, http://dx.doi.org/10.1093/bioinformatics/btv453.

[143] K.A. Beauchamp, G.R. Bowman, T.J. Lane, L. Maibaum, I.S. Haque, V.S. Pande, in: MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale, J. Chem. Theory Comput. 7 (10) (2011) 3412–3419, http://dx.doi.org/10.1021/ct200463m.

[144] L. Martínez, R. Andrade, E.G. Birgin, J.M. Martínez, in: PACKMOL: a package for building initial configurations for molecular dynamics simulations, J. Comput. Chem. 30 (13) (2009) 2157–2164, http://dx.doi.org/10.1002/jcc.21224.

[145] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, in: PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res. 35 (Web Server Issue) (2007) W522–W525, http://dx.doi.org/10.1093/nar/gkm276.

[146] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, in: PLUMED 2: new feathers for an old bird, Comput. Phys. Commun. 185 (2) (2014) 604–613, http://dx.doi.org/10.1016/j.cpc.2013.09.018.

[147] A. Bakan, L.M. Meireles, I. Bahar, in: ProDy: protein dynamics inferred from theory and experiments, Bioinformatics 27 (11) (2011) 1575–1577, http://dx.doi.org/10.1093/bioinformatics/btr168.

[148] S.O. Yesylevskyy, in: Pteros 2.0: evolution of the fast parallel molecular analysis library for C++ and python, J. Comput. Chem. 36 (19) (2015) 1480–1488, http://dx.doi.org/10.1002/jcc.23943.

[149] M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, in: PyEMMA 2: a software package for estimation validation, and analysis of Markov models, J. Chem. Theory Comput. 11 (11) (2015) 5525–5542, http://dx.doi.org/10.1021/acs.jctc.5b00743.

[150] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanskia, P. Cieplak, in: The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building, Phys. Chem. Chem. Phys. 12 (28) (2010), http://dx.doi.org/10.1039/c0cp00111b.

[151] T. Cickovski, S. Chatterjee, J. Wenger, C.R. Sweet, J.A. Izaguirre, in: MDLab: a molecular dynamics simulation prototyping environment, J. Comput. Chem. (2009), http://dx.doi.org/10.1002/jcc.21418.

[152] A. Bakan, N. Nevins, A.S. Lakdawala, I. Bahar, in: Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules, J. Chem. Theory Comput. 8 (7) (2012) 2435–2447, http://dx.doi.org/10.1021/ct300117j.

[153] F. Berenger, A. Voet, X. Lee, K.Y. Zhang, in: A rotation-translation invariant molecular descriptor of partial charges and its use in ligand-based virtual screening, J. Cheminform. 6 (1) (2014) 23, http://dx.doi.org/10.1186/1758-2946-6-23.

[154] I.S. Haque, V.S. Pande, in: PAPER: accelerating parallel evaluations of ROCS, J. Comput. Chem. 31 (1) (2010) 117–132, http://dx.doi.org/10.1002/jcc.21307.

[155] D.R. Koes, C.J. Camacho, in: Pharmer: efficient and exact pharmacophore search, J. Chem. Inform. Model. 51 (6) (2011) 1307–1314, http://dx.doi.org/10.1021/ci200097m.

[156] J. Sunseri, D.R. Koes, in: Pharmit: interactive exploration of chemical space, Nucleic Acids Res. (2016) gkw287, http://dx.doi.org/10.1093/nar/gkw287.

[157] A.M. Schreyer, T. Blundell, in: USRCAT: real-time ultrafast shape recognition with pharmacophoric constraints, J. Cheminform. 4 (1) (2012) 27, http://dx.doi.org/10.1186/1758-2946-4-27.

[158] J. Taminau, G. Thijs, H.D. Winter, in: Pharao: pharmacophore alignment and optimization, J. Mol. Graph. Modell. 27 (2) (2008) 161–169, http://dx.doi.org/10.1016/j.jmgm.2008.04.003.

[159] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, in: Electrostatics of nanosystems: Application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U.S.A. 98 (18) (2001) 10037–10041, http://dx.doi.org/10.1073/pnas.181342398.

[160] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, in: AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009) 2785–2791, http://dx.doi.org/10.1002/jcc.21256.

[161] O. Trott, A.J. Olson, in: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function efficient optimization, and multithreading, J. Comput. Chem. (2009), http://dx.doi.org/10.1002/jcc.21334.

[162] F. Ballante, G.R. Marshall, in: An automated strategy for binding-pose selection and docking assessment in structure-based drug design, J. Chem. Inform. Model. 56 (1) (2016) 54–72, http://dx.doi.org/10.1021/acs.jcim.5b00603.

[163] C. Bullock, N. Cornia, R. Jacob, A. Remm, T. Peavey, K. Weekes, C. Mallory, J.T. Oxford, O.M. McDougal, T.L. Andersen, in: DockoMatic 2.0: high throughput inverse virtual screening and homology modeling, J. Chem. Inform. Model. 53 (8) (2013) 2161–2170, http://dx.doi.org/10.1021/ci400047w.

[164] X. Jiang, K. Kumar, X. Hu, A. Wallqvist, J. Reifman, in: DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0, Chem. Central J. 2 (1) (2008) 18, http://dx.doi.org/10.1186/1752-153x-2-18.

[165] H. Li, K.-S. Leung, M.-H. Wong, idock. A multithreaded virtual screening tool for flexible ligand docking, in: 2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Institute of Electrical & Electronics Engineers (IEEE), 2012, http://dx.doi.org/10.1109/cibcb.2012.6217214.

[166] R.M. Abreu, H.J. Froufe, M. Queiroz, I.C. Ferreira, in: MOLA: a bootable self-configuring system for virtual screening using AutoDock4/Vina on computer clusters, J. Cheminform. 2 (1) (2010) 10, http://dx.doi.org/10.1186/1758-2946-2-10.

[167] J.D. Durrant, J.A. McCammon, in: NNScore 2.0: a neural-network receptor–ligand scoring function, J. Chem. Inform. Model. 51 (11) (2011) 2897–2903, http://dx.doi.org/10.1021/ci2003889.

[168] R. Meier, M. Pippel, F. Brandt, W. Sippl, C. Baldauf, in: ParaDockS: a framework for molecular docking with population-based metaheuristics, J. Chem. Inform. Model. 50 (5) (2010) 879–889, http://dx.doi.org/10.1021/ci900467x.

[169] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, in: Methods in Molecular Biology, Springer Science + Business Media, 2014, pp. 243–250, http://dx.doi.org/10.1007/978-1-4939-2269-7_19.

[170] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A.B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R.E. Hubbard, S.D. Morley, in: rDock: a fast versatile and open source program for docking ligands to proteins and nucleic acids, PLoS Comput. Biol. 10 (4) (2014) e1003571, http://dx.doi.org/10.1371/journal.pcbi.1003571.

[171] H. Li, K.-S. Leung, M.-H. Wong, P.J. Ballester, in: Improving AutoDock Vina using random forest: the growing accuracy of binding affinity prediction by the effective exploitation of larger data sets, Mol. Inf. 34 (2–3) (2015) 115–126, http://dx.doi.org/10.1002/minf.201400132.

[172] D.R. Koes, M.P. Baumgartner, C.J. Camacho, in: Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise, J. Chem. Inform. Model. 53 (8) (2013) 1893–1904, http://dx.doi.org/10.1021/ci300604z.

[173] A. Cereto-Massagué, M. Ojeda, R.P. Joosten, C. Valls, M. Mulero, M. Salvado, A. Arola-Arnal, L. Arola, S. Garcia-Vallvé, G. Pujadas, in: The good the bad and the dubious: VHELIBS a validation helper for ligands and binding sites, J. Cheminform. 5 (1) (2013) 36, http://dx.doi.org/10.1186/1758-2946-5-36.

[174] X. Zhang, S.E. Wong, F.C. Lightstone, in: Message passing interface and multithreading hybrid for parallel molecular docking of large databases on petascale high performance computing machines, J. Comput. Chem. 34 (11) (2013) 915–927, http://dx.doi.org/10.1002/jcc.23214.

[175] S.R. Ellingson, J.C. Smith, J. Baudry, in: VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers, J. Comput. Chem. 34 (25) (2013) 2212–2221, http://dx.doi.org/10.1002/jcc.23367.

[176] N. Zonta, I.J. Grimstead, N.J. Avis, A. Brancale, in: Accessible haptic technology for drug design applications, J. Mol. Model. 15 (2) (2008) 193–196, http://dx.doi.org/10.1007/s00894-008-0387-8.

[177] M. Brylinski, W.P. Feinstein, in: eFindSite: improved prediction of ligand binding sites in protein models using meta-threading machine learning and auxiliary ligands, J. Comput.-Aided Mol. Des. 27 (6) (2013) 551–567, http://dx.doi.org/10.1007/s10822-013-9663-5.

[178] P. Schmidtke, A. Bidon-Chanal, F.J. Luque, X. Barril, in: MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories, Bioinformatics 27 (23) (2011) 3276–3285, http://dx.doi.org/10.1093/bioinformatics/btr550.

[179] S.H. Oliveira, F.A. Ferraz, R.V. Honorato, J. Xavier-Neto, T.J. Sobreira, P.S. de Oliveira, in: KVFinder: steered identification of protein cavities as a PyMOL plugin, BMC Bioinform.s 15 (1) (2014) 197, http://dx.doi.org/10.1186/1471-2105-15-197.

[180] M.S. Till, G.M. Ullmann, in: McVol – a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm, J. Mol. Model. 16 (3) (2009) 419–429, http://dx.doi.org/10.1007/s00894-009-0541-y.

[181] M. Weisel, E. Proschak, G. Schneider, in: PocketPicker: analysis of ligand binding-sites with shape descriptors, Chem. Central J. 1 (1) (2007) 7, http://dx.doi.org/10.1186/1752-153x-1-7.

[182] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, in: A semiempirical free energy force field with charge-based desolvation, J. Comput. Chem. 28 (6) (2007) 1145–1152, http://dx.doi.org/10.1002/jcc.20634.

[183] P.J. Ballester, J.B.O. Mitchell, in: A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking, Bioinformatics 26 (9) (2010) 1169–1175, http://dx.doi.org/10.1093/bioinformatics/btq112.

[184] J.D. Durrant, J.A. McCammon, in: AutoClickChem: click chemistry in silico, PLoS Comput. Biol. 8 (3) (2012) e1002397, http://dx.doi.org/10.1371/journal.pcbi.1002397.

[185] J.D. Durrant, S. Lindert, J.A. McCammon, in: AutoGrow 3.0: an improved algorithm for chemically tractable semi-automated protein inhibitor design, J. Mol. Graph. Modell. 44 (2013) 104–112, http://dx.doi.org/10.1016/j.jmgm.2013.05.006.

[186] N. Chéron, N. Jasty, E.I. Shakhnovich, in: OpenGrowth: an automated and rational algorithm for finding new protein ligands, J. Med. Chem. (2015), http://dx.doi.org/10.1021/acs.jmedchem.5b00886, 150923135558008.

[187] R. Guha, M.T. Howard, G.R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, E.L. Willighagen, in: The Blue Obelisk – interoperability in chemical informatics, J. Chem. Inform. Model. 46 (3) (2006) 991–998, http://dx.doi.org/10.1021/ci050400b.

[188] N.M. O'Boyle, R. Guha, E.L. Willighagen, S.E. Adams, J. Alvarsson, J.-C. Bradley, I.V. Filippov, R.M. Hanson, M.D. Hanwell, G.R. Hutchison, C.A. James, N. Jeliazkova, A.S. Lang, K.M. Langner, D.C. Lonie, D.M. Lowe, J. Pansanel, D. Pavlov, O. Spjuth, C. Steinbeck, A.L. Tenderholt, K.J. Theisen, P. Murray-Rust, in: Open data open source and open standards in chemistry: the Blue Obelisk five years on, J. Cheminform. 3 (1) (2011) 37, http://dx.doi.org/10.1186/1758-2946-3-37.