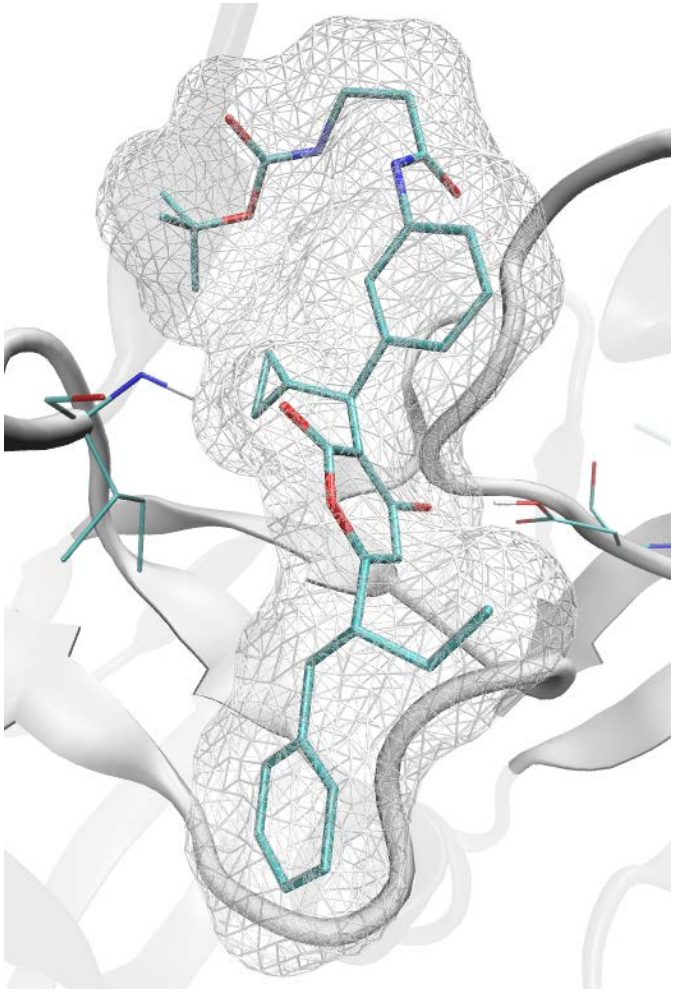
BioMolTech

In relentless pursuit of accuracy



Modeling the HIV-1 protease complex with U100313 is a particularly tough case for docking software. Lead-Finder successfully predicts correct ligand position (rendered in licorice) with 1.2Å RMSD as evidenced from X-ray structure (electron density rendered in wireframe).

Lead Finder

v.1.1.16

User Guide

www.biomoltech.com

BioMolTech Lead Finder User Guide
Revision 20150120

www.biomoltech.com

If you have comments about this User Guide, submit your feedback to support@biomoltech.com.

Table of Contents

Getting Started with Lead Finder

Lead Finder software is a computational chemistry application for modeling protein-ligand interaction. Lead Finder can be used in molecular docking studies and for the quantitative evaluation of ligand binding and biological activity.

Lead Finder introduces three scoring functions optimized for the accurate prediction of 3D docked ligand poses, protein-ligand binding energy and rank-ordering active and inactive compounds in virtual screening experiments. Lead Finder is designed to satisfy needs of computational and medicinal chemists involved in the drug discovery process, pharmacologists and toxicologists involved in the modeling and evaluation of ADMET properties in silico, and biochemists and enzymologists working on enzyme specificity and rational enzyme design.

This User Guide is intended for experienced computational chemists who are familiar with the modern molecular docking and virtual screening approaches.

Refer to the following paper for detailed description of Lead Finder's molecular docking algorithm and its benchmarking data: Oleg V. Stroganov, Fedor N. Novikov, Viktor S. Stroylov, Val Kulkov and Ghermes G. Chilov. Lead Finder: An Approach To Improve Accuracy of Protein−Ligand Docking, Binding Energy Estimation, and Virtual Screening. *J. Chem. Inf. Model.*, **2008,** *48,* 2371-2385. DOI: 10.1021/ci800166p, http://pubs.acs.org/doi/abs/10.1021/ci800166p.

Data Flow in Lead Finder

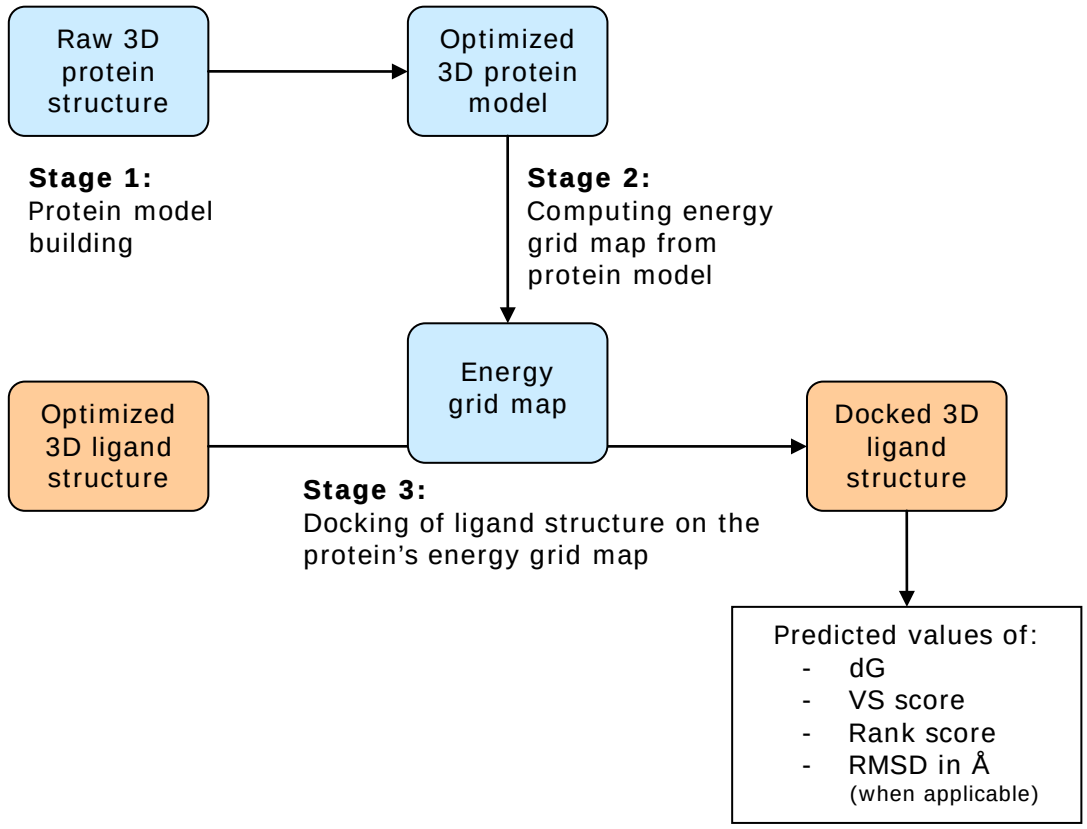
Lead Finder takes a 3D protein model and one or more 3D ligand structures as input and generates one or more predicted 3D docked ligand poses or calculates predicted scores for a fixed ligand pose. Lead Finder assumes that the protein and ligand structures are rigid; however, it analyses possible conformations of ligand structures by rotating functional groups along each freely rotatable bond (FRB).

For each ligand pose Lead Finder determines values of the free energy of binding, the virtual screening score (VS score) and if applicable, the pose ranking score using its three built-in scoring functions. These values can be exported into a tab-separated file for spreadsheet analysis, in a text report file with full details on the contributing energy components, and if applicable, into an SD file.

Docking process with Lead Finder starts with preparation of an optimized 3D protein model and then calculating an energy grid map for the protein model. This energy grid map is then used in docking ligand structures. Lead Finder can save the calculated energy grid map to a disk file and retrieve it in subsequent docking experiments. When a pre-computed energy grid map is used, Lead Finder does not require a protein structure on the input.

Lead Finder's typical data flow is presented in Figure 1.

Figure 1. Data flow in Lead Finder.



Lead Finder can import and export protein and ligand structures from and to a variety of file formats, including PDB, GRO, MOL, MOL2 and SDF.

Lead Finder is a command-line application for Windows 32-bit and Linux 32/64-bit platforms. The minimum system requirements are:

- CPU: at least 1GHz Intel or AMD-compatible
- RAM: at least 512 Mb
- Disk space: at least 200 Mb + enough space to store the output SD files

- If applicable, one available USB port for a HASP key

You must have administrative privileges to install the system drivers for HASP key.

Preparation of a Protein Structure for Docking

A protein structure must be prepared for docking. You should consider the following issues before using your protein structure model in a docking study:

- Protein structures, especially the ones taken from public sources such as the Protein Data Bank (PDB), will often contain no explicit hydrogen atoms. Lead Finder requires explicit and correct placement of hydrogen atoms in the protein structure.

- A protein structure must be properly protonated. At a minimum, the following issues must be considered: (1) the protonation state of a protein depends on pH; (2) the protonation of His should be reviewed as a special case; (3) protons should be attached to the most appropriate atom when alternatives exist, such as for the chemically equivalent atoms in His, Glu, Asp; (4) the proton orientation may have to be further optimized.

- The accuracy of the crystallographic 3D protein structure is crucially important. As a rule, the higher resolution X-ray crystal structure, the better. We recommend using structures with at least 2.5 Å or better resolution.

- It is highly recommended that any defects in the experimentally resolved protein structures, such as missed atoms or residues, incorrect bond lengths, angles etc., especially in the proximity to the ligand's binding site, are corrected. Lead Finder can check a protein structure for errors and inconsistencies when invoked with the --check-only parameter (see the *Configuration Parameters* section for more details). Also, the quality of a protein structure may be assessed by a number of Internet services available at the PDB site www.rcsb.org.

- Protein cofactors that are significant for ligand binding should be retained within the protein structure.

- Any non-intrinsic parts of a protein such as ligands, water molecules, buffer ions, etc. must be removed before docking calculations. Cofactors and structurally or catalytically

important metal ions bound to protein should be retained; sometimes, conservative structural water molecule(s) known to play a crucial role in ligand binding may be retained as well.

- Lead Finder accepts protein structure input in one of the following file formats: PDB, GRO or MOL2. To convert your structure to one of these formats, you may use free Open Babel software available at http://sourceforge.net/projects/openbabel/.

Lead Finder can take an already prepared protein model or automatically build a protein model from a given protein structure. The pre-docking preparation of a protein structure with Lead Finder involves:

- Extraction of a docked ligand from a protein structure and saving it into MOL and PDB file formats. The extracted ligand in PDB format can be used as a reference ligand to define the size and position of the energy grid box (refer to the *Creating Energy Grid Maps* section for more details);

- Removal of ligands from a protein structure;

- Addition of missing hydrogen atoms to a protein structure and optimization of positions of the hydrogen atoms with respect to the electrostatic, Van der Waals and hydrogen bond interactions;

- Assignment of ionization states of amino acids;

- Repairs to some of the common defects in PDB structures, like the incorrect aminoacid labels, missing or unresolved aminoacid side chains;

- Addition of missing hydrogen atoms to cofactors.

Lead Finder saves the prepared protein model in the MOL2 file format. Refer to the *Configuration Parameters* section on page 13 for detailed information about the protein model preparation parameters and *Example 1: Preparation of Protein* on page 27 for an example of using the protein preparation function.

Preparation of Ligand Structures for Docking

Lead Finder requires 3D ligand structures as input. If your ligands are only available as SMILES or InChI strings, or as two-dimensional structures in MOL or SDF format, you have to

convert them into 3D structures before running docking experiments with Lead Finder. A number of molecular modeling techniques is available for 3D-optimization, including such tools as Corina (http://www.molecular-networks.com/software/corina/) or ACD/ChemSketch (http://www.acdlabs.com/).

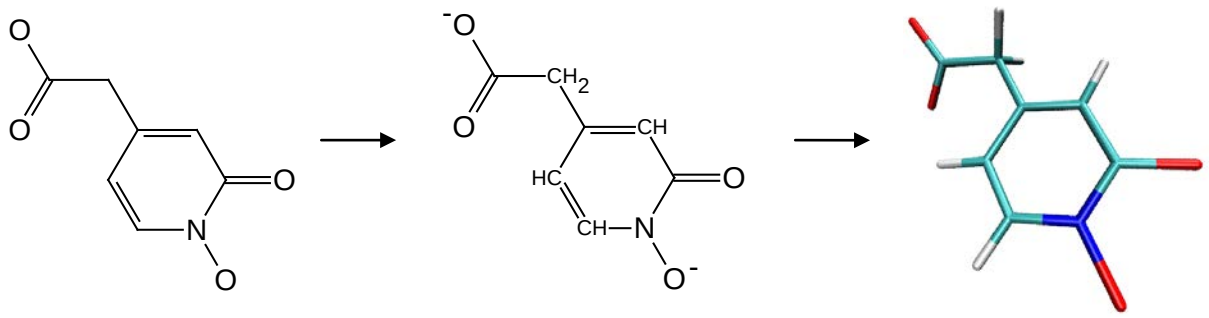
Explicit assignment of all appropriate hydrogen atoms in a ligand is mandatory. Missing hydrogen atoms in ligand structures will result in incorrect docking predictions!

Ligand protonation at a specific pH is recommended. Numerous software tools are available for predicting pKa and protonation states of ionogenic groups, including such tools as ACD/pKa (www.acdlabs.com).

Lead Finder accepts ligand structure input in one of the following file formats: PDB, GRO, MOL2, MOL or SDF. To convert your structure to one of these formats, you may use free Open Babel software available at http://sourceforge.net/projects/openbabel/.

A typical workflow for preparing a ligand structure for docking is presented in Figure 2.

Figure 2. A typical workflow for preparing a ligand structure for docking.



An initial 2D-structure of a ligand

2D-structure with all H-atoms and ionization states adjusted to pH 7. Note that the hydroxyl group bound to the nitrogen atom is ionized!

A complete 3D-structure, ready for processing by Lead Finder

Creating Energy Grid Maps

Every docking experiment with Lead Finder begins with the computation of energy maps for the site of interest in a protein structure (the “"energy grid box”) where a ligand is expected to dock. In most cases, the energy grid box is built around a protein’s putative active site. The dimensions of the energy grid box set the boundaries of the search space for docking a ligand. Therefore, choosing a correct position and dimensions of the energy grid box has crucial importance in docking experiments.

The amount of computations in search of the optimal docked poses increases in proportion to the volume of the energy grid box. Therefore, the size of the energy grid box must be kept at a minimum to include the site of interest but not more. Lead Finder incorporates a cavity detection algorithm that determines an optimal orientation of the energy grid box to minimize its size while maximizing the overlap with protein cavity. This algorithm is enabled by default.

Lead Finder offers two ways to define position and dimensions of an energy grid box:

- by using a reference ligand, and
- by setting coordinates of the center and dimensions of the energy grid box.

Using a reference ligand is perhaps the easiest way to define position and dimensions of an energy grid box. Lead Finder sets the center of the energy grid box at the geometrical center of the reference ligand and creates a minimum bounding box around the reference ligand. The sides of the minimum bounding box are then moved away from the center by the default value of 6 Å to set boundaries of the energy grid box. Once this is done, the cavity detection algorithm is employed to find an optimal orientation of the energy grid box that maximizes its overlap with protein cavity.

The size and orientation of the energy grid box can be customized through the use of --grid-size and --no-grid-reorientation parameters.

Another way to define geometry of the search space for docking a ligand is by setting coordinates of the center and dimensions of the energy grid box. The --grid-center parameter sets the center of the energy grid box and the --grid-size parameter defines lengths of its edges in Å. The cavity detection algorithm is disabled when this method of defining dimensions of the energy grid box is used. Refer to the *Configuration Parameters* section for more information.

In some cases, especially when docking small ligand structures, manual adjustment of the --min-cav-vol parameter may be necessary to obtain better docking predictions. This parameter defines the minimum threshold for the volume of inner cavities (enclosures) for inclusion into the energy grid computation. Cavities with the volume that is less than the threshold volume are treated as if they did not exist, i.e. they were part of protein interior. The default value of this parameter is 40.5 $Å^3$, which is equal to the volume of a benzene molecule.

Lead Finder requires an energy grid map to be created and loaded into memory before it proceeds with docking optimizations for a ligand. The computation of an energy grid map is a CPU-intensive task that can take several minutes to complete, especially when working with large energy grid boxes. For this reason, we recommend that you save the computed energy grid map to a disk file for future reuse by specifying a file name with the --save-grid parameter. Saving energy grid maps to a disk file can be done either as part of a regular docking or screening experiment or by starting Lead Finder in the energy grid generation mode through the use of the --grid-only parameter. Loading energy grid maps into memory can be done with the --load-grid parameter. When you load an energy grid map, you do not need to provide a protein structure on input, except when a ligand is covalently bound to the protein structure.

Lead Finder saves energy grid maps in either *bin* or AutoDock *map* format. AutoDock *map* files can be visualized with software like VMD[1]. However, *map* files consume considerably more disk space than *bin* files.

Energy grid maps written in the *bin* format are more compact and faster to load than *map* files. Therefore, the *bin* format is recommended when you have no immediate need to visualize energy grid maps. Note that *bin* files can be converted into *map* files by invoking Lead Finder with --grid-only, --load-grid and --save-grid parameters.

---

[1] VMD (Visual Molecular Dynamics) is free cross-platform molecular visualization software for displaying, animating, and analyzing large biomolecular systems using 3D graphics and built-in scripting, published by the Theoretical and Computational Biophysics Group in the Beckman Institute at the University of Illinois. Visit http://www.ks.uiuc.edu/Research/vmd/ for complete information and documentation.

Installation

Depending on the type of software license you acquired, your Lead Finder installation may require the use of a HASP key. The HASP key is a USB device that contains information about your software license. While running, Lead Finder will periodically check your license information and will continue to run as long as your software license is valid and unexpired. Therefore, your HASP key must remain connected to a USB port of the computer where Lead Finder is running. In a network installation, the HASP key must remain connected to a host computer, and all nodes of the network running Lead Finder must have network access to the host computer.

The HASP key requires additional system software that is installed as part of the Lead Finder installation. To successfully install the HASP system driver, you must have appropriate access permissions, “Administrator” under Windows or “root” under Linux on the system where Lead Finder is to be installed and used. For more information about the HASP key driver, visit the manufacturer’s website at www.aladdin.com.

Installation on a Windows system

To start Lead Finder installation on a Windows system, run “leadfinder-setup.exe”. Note you have to install the HASP HL driver during the Lead Finder installation procedure if your license requires the use of a HASP key.

During the installation process, you will be given an option to install example datasets with instructions and an option to add Lead Finder installation directory to the system path. Since Lead Finder is a command line application, adding its installation directory to the path will allow you to invoke Lead Finder from any location by simply typing “leadfinder.exe” at a command prompt and therefore selecting this option is highly recommended. Note you will need to restart your system for the HASP driver and the modified system path to take effect.

Installation on a Linux system

To start Lead Finder installation on a Linux system, execute the following command:

“tar xzvf leadfinder-setup.tar.gz“

This command will unpack Lead Finder installation files into a new directory named “lead-finder” under your current directory. After unpacking the installation files, execute:

"lead-finder/install.pl"

This command will complete the installation of Lead Finder. During the installation process, you will be given choices where to install the HASP key driver (if applicable), Lead Finder executable files and the example files.

Updating software license on your HASP key

When the software license on your HASP key has to be renewed or upgraded or otherwise replaced, you can obtain a *v2c* file with a new license from BioMolTech electronically and apply it to your HASP key using the following procedure.

Under Windows, invoke "HASP key update" from the Lead Finder program group, then click on the "Apply License Update" tab. Specify location of the *v2c* license update file you obtained from us and click on "Apply Update". Your HASP key will be updated with your new license.

Under Linux, execute "update_key_linux". You will be prompted to either retrieve HASP key information (*i*) or to update your key (*u*). Choose *u* to update your key and enter the full path and name of the *v2c* file with your new license. Your HASP key will be updated with your new license.

In rare cases, additional information about your HASP key may be required before a new license can be issued. If we ask you to provide your HASP key information to us, choose "Collect Information" under Windows or (*i*) option under Linux to save your HASP key information to a disk file, then email it to us in a file attachment.

Configuration Parameters

The configuration parameters can be provided to Lead Finder in a command line or in a parameter file. In a command line, parameter names are prefixed by one or two dashes ("-"). In a parameter file, names of parameters are not prefixed by dashes. If the same parameter is specified both in a command line and in a parameter file, the command line value takes precedence over the value specified in a parameter file. Therefore, you can create a parameter file to store docking parameters as a template for your common use scenario and then customize it further on a case-by-case basis by specifying particular parameters in the command line.

Single-line comments are allowed in a parameter file after a semi-colon character. Comments are not allowed in the command line.

The appearance order of parameters has no significance.

General Settings

-f FILENAME, --parameters=FILENAME

Read configuration parameters from FILENAME, a text file specifying one configuration parameter per one line of text. If the same parameter is specified both in the parameters file and in the command line, the value from the command line takes precedence.

-h [PARAMETER], --help[=PARAMETER]

Print help screen with a full list of configuration parameters if no PARAMETER is specified, or print help screen on a specific PARAMETER.

-v, --verbose

Print information about calculations in progress to the standard output.

-q, --quiet

Suppress all console output (quiet mode), with the exception of fatal errors and the program initialization message.

-debug [PATH], --debug[=PATH]

Save technical information from a docking experiment into directory PATH that can be sent to the development team for debugging. If PATH is not specified, the debug files will be saved into a subdirectory *debug* in your current directory. If a directory with such name already exists, *debug.1* will be used instead (or *debug.2*, and so on).

Job Options

-check [FILENAME]
--check-only [FILENAME]

Check the protein structure and the ligands for any problems and save an analysis report into FILENAME. If FILENAME is not specified, the report is saved into *X_check.log*, where X is the name of a ligand input file or a protein input file if the ligand input file is not provided.

-build [TEMPLATE], --model-build-only[=TEMPLATE]

Remove ligand from protein structure and save it into *TEMPLATE_ligand.mol* and *TEMPLATE_reference.pdb*; add missing hydrogen atoms, assign ionization states and bond types, save the prepared protein structure model into *TEMPLATE_protein.mol2* and exit. When TEMPLATE specification is omitted, the name of the source protein structure file is used as *TEMPLATE*. When this parameter is used in combination with the --load-grid parameter, a protein model is generated from the energy grid file and saved into *TEMPLATE_protein.mol2*.

When a protein structure contains multiple docked ligands, you can use --extract-reference parameter to specify the ligand to be removed and used as a reference ligand.

-addH=[auto|on|off], --add-hydrogens=[auto|on|off]

Setting this parameter to "on" forces the addition of missing hydrogen atoms and assignment of ionization states to the given protein structure at a certain pH value. When set to "off", the protein structure is taken "as is". When this parameter is not specified or set to "auto", Lead Finder tries to guess if the protein model needs to be prepared and if necessary, forces the addition of missing hydrogen atoms and assignment of ionization states to the protein structure.

-grid, --grid-only

Compute energy grid maps, save them into a *bin* or *map* files and exit. The output file name must be specified. See --save-grid parameter in INPUT/OUTPUT section for more information.

-score, --score-only

Calculate dG and VS scores for a fixed position of a ligand and its locally optimized configuration. The ligand must be provided in a file format that retains information about the bond order, protonation states etc, such as MOL or MOL2. Note that in many software packages the original PDB coordinate system is lost when ligand structures are saved in MOL or MOL2 file format. To position your MOL or MOL2 ligand in the original coordinate system, add a reference to the original ligand in --ligand-reference parameter. The calculated scores for a ligand are saved into *ligand_report.log* and both the original and the locally optimized ligand poses are output into *ligand_docked.pdb*.

-vs, --virtual-screening

When this parameter is specified, Lead Finder will use a faster but slightly less precise docking algorithm. Use of this parameter is recommended for screening of large ligand libraries where the processing speed is very important.

-xp, --extra-precision

When this parameter is specified, Lead Finder will use the most rigorous sampling and scoring algorithms to increase accuracy and reliability of predictions at the cost of slower speed of processing.

-dfn, --default-filenames

Specify this parameter to instruct Lead Finder to read protein structure from *protein.pdb*, ligand structure from *ligand.mol*, and reference ligand from *reference.pdb*. Specification of this parameter has the same effect as specification of the following three parameters:
--protein=protein.pdb --ligand=ligand.mol --ligand-reference=reference.pdb

-np N, --number-of-processes=N

Specify a number of instances of the executable code to be run in parallel to process input data. This option is useful in multi-processor or computing cluster environment. Note 1: this option is available only when your Lead Finder software license allows parallel processing. Note 2: this parameter can only be specified in the command line.

-js N, --mpi-job-size=N

Specify how many ligands should be processed by a remote node in the cluster environment before waiting for the master node to pick up the calculation results. The master node polls remote nodes each time it processes a ligand from its own job list. The default value of this parameter is 100. Note: this parameter can be specified only in the command line.

Input / Output

-mm FILENAME, --protein=FILENAME

Import the protein structure from FILENAME. The following file formats are automatically detected and imported: PDB, GRO, MOL2. Note this parameter is not required when a pre-computed grid map is loaded via the --load-grid parameter, except when a ligand is covalently bound to protein (-cov or --covalent-bond parameter).

-li FILENAME, --ligand=FILENAME

Import a ligand structure or structures from FILENAME for docking. The following file formats are automatically detected and imported: MOL, SDF, MOL2, PDB and GRO.

-og FILENAME.[map|bin], --save-grid=FILENAME.[map|bin]

Compute and save energy grid maps into FILENAME in *bin* or *map* (AutoDock) format. Energy grid maps in *map* format can be visualized in visualization software like VMD, but they are larger in size than *bin* files.

When *map* is specified as the output format, Lead Finder will create a *map* file for each individual energy component so they can be visualized separately. In case of the *bin* format, all energy components are written into a single file.

-g FILENAME.[map|bin], --load-grid=FILENAME.[map|bin]

Load energy grid map from FILENAME.bin or FILENAME.map in *bin* or *map* format. By reusing previously calculated energy grid maps for a protein structure you avoid the

energy grid map generation step that may be time consuming, especially in case of large protein structures.

-o FILENAME.[pdb|sdf], --output-poses=FILENAME.[pdb|sdf]

Output docked poses into FILENAME.pdb or FILENAME.sdf. If this parameter is not specified, the docked poses will be saved into *X_docked.pdb* or *X_docked.sdf* in the current directory, where X is the ligand input file specified in the --ligand= parameter.

-os FILENAME.[csv|txt], --output-tabular=FILENAME.[csv|txt]

Output ligand id, predicted dG and VS scores, individual energy components and other information into a comma-separated (.csv) or tab-separated (.txt) values file FILENAME for importing into spreadsheet processing software such as Microsoft Excel. If this parameter is not specified and more than one ligand is to be processed, *energy.log* file will be written by default.

-l FILENAME, --text-report=FILENAME

Output a report in text format into FILENAME with the predicted dG and VSscore values of the top ranked pose for each of the processed ligands in a multi-ligand experiment with details on the breakdown of dG by energy component. In case of docking a single ligand, output the predicted dG, VSScore and Ranking values for all found poses with dG < 0 up to the number of poses specified in the --max-poses parameter. If this parameter is not specified, the report will be written into *X_report.log* file, where X is the name of the ligand source file.

Protein Model Building

-cryst X, --crystallographic-distance=X

Assemble crystallographic subunits located within X Å from the reference ligand. The following conditions must be met: a reference ligand is found and extracted from the protein structure and information about dimensions of the crystallographic unit cell and the symmetry group are present in the originating protein file. When the parameter is not specified, the default value of 7 Å is used. Set the parameter value to 0 to switch off the crystallographic unit assembly.

-pH X, --pH=X

Specify pH value at which the protein model must be ionized.

Energy Grid Maps

-lr FILENAME, --ligand-reference=FILENAME

This parameter has three uses depending on the context:

1. Define position and dimensions of the energy grid box: this parameter specifies the center and dimensions of energy grid maps in the absence of --grid-center and --grid-size parameters. Specify name of a file containing a reference ligand from the PDB or other source that is positioned in the same binding site of the same protein and in the same coordinate system. The following file formats are automatically detected and imported: MOL, MOL2, PDB and GRO. The energy grid maps are created and calculated for the area surrounding the reference ligand. When no resolved protein-ligand structure is available and therefore a reference ligand does not exist, you can still define the search space by positioning an arbitrary molecule in the docking site and specifying it as a reference ligand.

If a reference ligand is not specified and the position and dimensions of energy grid maps are not otherwise defined, Lead Finder will try to automatically detect ligand in the protein structure and use it as a reference ligand to define position and dimensions of energy grid maps.

2. Calculation of RMSD value between the predicted pose and the reference ligand pose. When the chemical structure of a ligand matches that of a reference ligand, which may be the case when Lead Finder's performance is evaluated on an experimentally resolved structure, an RMSD value between the predicted pose and the reference pose will be calculated.

3. Specification of a fixed ligand position for calculation of dG and VS scores. Specify name of a file containing a reference ligand whose atom coordinates must be used instead of coordinates of a prepared ligand. This option is useful when the prepared ligand saved in MOL or MOL2 format does not retain the original coordinate system.

-xref [resname NAME | resnum X [Y] | chain Z]
--extract-reference=[resname NAME | resnum X [Y] | chain Z]

The --extract-reference parameter provides an alternative method of specifying position and dimensions of the active center when a reference ligand or coordinates of the active center are not immediately available. This parameter can be used in one of the following ways:

--extract-reference
Automatically detect ligand in the protein structure. This is the default mode of operation when a reference ligand is not provided. The ligand must be a small molecule and not one of the following types: nucleotide, standard aminoacid or a buffer component.

--extract-reference=resname NAME
Extract a named ligand from the protein structure.

--extract-reference=resname NAME chain Z
Extract a named ligand from chain Z of the protein structure.

--extract-reference=resnum X
Extract residue X as a ligand.

--extract-reference=resnum X Y
Extract residues X to Y as a ligand.

--extract-reference=resnum X chain Z
Extract residue X of chain Z as a ligand.

--extract-reference=resnum X Y chain Z
Extract residues X to Y of chain Z as a ligand.

--extract-reference=chain Z
Extract the whole chain Z as a ligand, provided that it has no more than 500 non-hydrogen atoms.

-ar FILENAME, --additional-reference=FILENAME

Specify additional reference ligands as needed. The following file formats are automatically detected and imported: MOL, MOL2, PDB and GRO. The additional reference ligands are used exclusively for the RMSD calculation of a docked ligand pose when a number of symmetric ligand binding positions exist. In such case, the RMSD will be calculated with respect to the nearest reference ligand. Specification of additional reference ligands does not affect generation of the energy grid maps, docking process and the resulting accuracy of predictions in any way.

-gc X,Y,Z
--grid-center=X,Y,Z

Specify center of the energy grid box in Å in the coordinate system of the protein structure. When --grid-center and --grid-size parameters are not specified, Lead Finder uses the center of a reference ligand as the center of the energy grid box.

-gsp X, --grid-spacing=X

Specify spacing (the distance between the closest nodes) of the grid in Å. The default value is 0.375 Å.

-gsz X|X,Y,Z
--grid-size=X|X,Y,Z

Specify size of the energy grid box. When a reference ligand is specified or automatically extracted via --extract-reference, create an energy grid box by moving planes of the reference ligand's minimum bounding box away from its center by X Å each (6 Å by default). The dimensions of a reference ligand are ignored when X,Y,Z are specified and a box with edges of X, Y and Z Å is created instead. When a reference ligand is not specified, create an energy grid either as a cube with edges of X Å (30 Å by default), or as a box with edges of X, Y and Z Å. You can specify size of the energy grid box using an alternative method through the --grid-npoints parameter.

Specification of the --grid-size parameter automatically turns off Lead Finder's cavity detection algorithm. For more information, refer to the --no-grid-reorientation parameter.

-gnp A,B,C
--grid-npoints=A,B,C

Specify size of the energy grid box by a number of nodes per each of the three Cartesian dimensions, with nodes separated from each other by the distance specified by the value of --grid-size=X parameter (in Å). The --grid-npoints parameter is mutually exclusive with the --grid-size parameter.

-nog, --no-grid-reorientation

By default, Lead Finder will apply its cavity detection algorithm to reorient the energy grid box in order to maximize its overlap with the binding site. Specify this parameter to turn off Lead Finder's cavity detection algorithm. When this parameter is specified, X, Y and Z values from the --grid-size parameter will apply to the coordinate system of the protein structure without regard to the location or orientation of the binding site.

Specification of the --grid-size parameter automatically turns off the cavity detection algorithm.

-gat <comma-separated list>|all
--grid-atom-types=<comma-separated list>|all

Specify types of ligand atoms to be included in the computation of energy grid maps. *All* means C, A, CF, N, NX, O, S, H, P, F, Cl, Br, I, B (where A is an aromatic carbon atom, CF – a carbon atom covalently bonded with a fluorine atom, NX is a nitrogen atom that does not form hydrogen bonds – for example, an amide nitrogen atom). You can specify your own list of atoms in a comma-separated list. The default value of this parameter is *all*.

-xcav V, --min-cavity-vol=V

Set minimum threshold for the volume of inner cavities (enclosures) in $Å^3$ inside the body of a protein structure for inclusion into the energy grid computation. Cavities with the volume less than V will be treated as if they were part of protein body. This parameter only applies to inner cavities. It does not apply to pockets, hollows, etc. Note this parameter may be significant in docking studies of small ligand molecules. The default value of this parameter is 40.5 $Å^3$, which is equal to the volume of a benzene molecule.

Genetic Algorithm Optimization

--poolsize=X

Increase the number of individuals in the initial pool by the factor of X. Lead Finder automatically determines the appropriate number of individuals in the initial pool based on the number of the conformational degrees of freedom, the energy grid size and other factors; setting X to a value other than one will alter the automatically determined number of individuals in the initial pool.

--popsize=X

Increase the number of individuals in a population by the factor of X. Lead Finder automatically determines the appropriate number of individuals in a population based on a number of factors; setting X to a value other than one will alter the automatically determined number of individuals in the population.

Other Parameters

-w, --retain-water

When this parameter is specified, water molecules from the protein structure file are not removed prior to the energy grid calculation and/or docking. By default, the water molecules are removed. This option can be important when structurally conserved water molecules are essential for ligand binding.

-env solution|membrane, --environment=solution|membrane

Specify type of environment surrounding the ligand binding site. *membrane* applies to membrane-buried protein sites; s*olution* applies to protein sites surrounded by aqueous media. The default value is *solution*.

-metal A,B
--metal-coord=A,B

When this parameter is not specified, Lead Finder will attempt to guess the metal atom's coordination number. By specifying this parameter, you can explicitly define the metal atom's coordination number. Use this parameter when the metal atom's coordination number has unusual value or can have multiple values. A is the metal

atom's sequential number (index) from a protein PDB file, B is the metal atom's coordination number. You can have as many --metal-coord parameters as there are metal atoms in a protein structure.

-cov A,B[,X]
--covalent-bond=A,B[,X]

When ligand is covalently bound to a protein, indexes of the bonded atoms have to be provided. Specify A as the sequential number (index) of ligand's bonded atom from its PDB file, and B as the sequential number (index) of protein's bonded atom from the protein's PDB file. X specifies the covalent bond's length in Å. While the specification of covalent bond's length is optional, it is highly recommended when the bond length is known. The specification of covalent bond's length will improve docking predictions as the ligand will be correctly positioned with respect to the protein. You do not have to specify the value of X when you know that the ligand is already correctly placed so that it does not violate the covalent bond length rules.

-rta, --rotate-amide

Specify this parameter to turn on the rotation of amide bond during docking. By default, the rotation of amide bond is turned off.

-rtc, --rotate-conjugated-bond

Specify this parameter to turn on the rotation of conjugated double or aromatic bonds during docking. By default, the rotation of such bonds is turned off.

-mp N, --max-poses N

Specify the maximum number of poses with dG<0 to be output when docking a single ligand. The default value of this parameter is 20.

Common Operations

Checking Protein and Ligand Structures for Potential Problems

Checking your protein and ligand structure files for problems that Lead Finder cannot repair is always a good idea since inconsistencies in the source structure files may fail your docking experiment.

To check a protein structure in *protein.pdb* and ligand structures in *ligands.sdf* for potential problems, execute:

leadfinder --check-only --protein=protein.pdb --ligand=ligands.sdf

Lead Finder will produce a text report for any problems it will encounter. In most cases, error reports and warnings will be self-explanatory.

The total charge on the protein structure ("total charge on macromolecule") is a good indicator of the absence of the explicitly assigned hydrogen atoms. The normal value of the total charge is a single-digit value. If you obtain a double digit or even higher value for the total charge, the most likely your protein model has missing hydrogen atoms. In such case, you need to build a protein model with Lead Finder or other software.

Protein Model Preparation

A protein structure must be prepared for docking by assignment of explicit hydrogen atoms and specifying ionization states of aminoacids. The docked ligands must be removed from the protein structure. The following command initiates the preparation of a protein model by Lead Finder at pH=5.5 using a protein structure from 1SRE.pdb:

leadfinder --model-build-only --pH=5.5 --protein=1SRE.pdb

The prepared protein model is saved into *1SRE_protein.mol2*. The removed ligand is saved into *1SRE_reference.pdb* to be used as a reference ligand for the definition of size and position of the energy grid, and *1SRE_ligand.mol* as a MOL file.

Creating and Saving Energy Grid Maps

Creating and saving energy grid maps for future reuse saves time when you perform multiple docking experiments with the same protein structure. When you load a pre-computed energy

grid map from a disk file, the energy grid map calculation step is skipped and Lead Finder proceeds directly to docking optimizations.

To calculate an energy grid map and save it into *gridmap.bin* for a protein structure from *protein.pdb* using *ligandref.pdb* as a reference ligand, execute:

```
leadfinder --grid-only --save-grid=gridmap.bin --protein=protein.pdb
    --ligand-reference=ligandref.pdb
```

To load a pre-computed energy grid map and dock a ligand from *ligand.mol*, execute:

```
leadfinder --load-grid=gridmap.bin --ligand=ligand.mol
```

If you save a grid map into an AutoDock *map* file such as *gridmap.map*, you can visualize such energy grid maps with specialized visualization software like VMD. Note that *map* files consume more disk space than files in the *bin* format.

Docking a Ligand

To dock a ligand from *ligand.mol* using a pre-computed energy grid map from *gridmap.bin* and write docked poses into *liganddocked.pdb* and a resulting report into *report.txt*, execute:

```
leadfinder --load-grid=gridmap.bin --ligand=ligand.mol
    --output-poses=liganddocked.pdb --text-report=report.txt
```

Virtual Screening of a Compound Library

To perform virtual screening of a compound library from *lib123.sdf* using a pre-computed energy grid map from *gridmap.bin* and to write predicted score values into a tab-separated file named *lib123scores.csv* for spreadsheet processing and predicted top-scoring docked poses into *lib123processed.sdf*, execute:

```
leadfinder --virtual-screening --load-grid=gridmap.bin --ligand=lib123.sdf
    --output-tabular=lib123scores.csv --output-poses=lib123processed.sdf
```

In the virtual screening mode, Lead Finder uses a significantly faster but slightly less precise docking algorithm. The virtual screening mode is recommended for use on large compound libraries when the processing speed is very important.

Examples

The standard Lead Finder installation includes three examples that can be used for training or as a starting point or a template to create your own docking and screening experiments.

Example 1: Preparation of Protein Model

Lead Finder prepares a protein structure for docking by automatically removing docked ligand from the active center (if present), adding missing hydrogen atoms and optimizing their positions with respect to the electrostatic, Van der Waals and hydrogen bond interactions, assigning the ionization states of amino acids at a given pH, repairing some of the common defects in PDB structures, like the incorrect aminoacid labels, missing or unresolved aminoacid side chains, adding crystallographic subunits if necessary, and adding missing hydrogen atoms to cofactors.

To prepare a protein model of streptavidin (PDB: 1SRE), downloaded from the Protein Data Bank as *1sre.pdb*, execute:

leadfinder --model-build-only --pH=5.5 --protein=1sre.pdb

Lead Finder will generate a prepared protein model and save it into *1sre_protein.mol2*. Also, the removed ligand will be saved into *1sre_ligand.mol* and *1sre_reference.pdb*. The latter file can be used to specify the position and dimensions of active center during the energy grip computation step.

To switch off the assembly of crystallographic units, set --crystallographic-distance (-cryst) parameter to zero value:

leadfinder --model-build-only --pH=5.5 --protein=1sre.pdb
--crystallographic-distance=0

When multiple ligands are present in a protein structure, you can control the removal of a ligand by specifying ligand's name or residues and/or chains forming the ligand:

leadfinder --model-build-only --pH=5.5 --protein=1sre.pdb
--extract-reference=resname HAB chain B

The above command will extract residue HAB, chain B, and save it into *1SRE_ligand.mol* and *1SRE_reference.pdb*.

Example 2: Calculating Energy Grid Maps

Lead Finder uses energy grid maps at certain stages of docking optimizations. The energy grid maps must be either computed or loaded from a disk file. If you plan to run several docking experiments on your protein model, pre-computing and saving energy grid maps will save you time in subsequent docking experiments.

Lead Finder calculates and saves energy grid maps into disk files in either *bin* (the internal binary format) or AutoDock *map* format. To calculate and save energy grid maps without proceeding further to docking computations, use --grid-only parameter.

Note that the energy grid maps in AutoDock *map* format can be visualized with software like VMD. However, *map* files are generated for every individual energy component, therefore producing many *map* files in your work directory for a single protein target.

Lead Finder's internal *bin* format is recommended when you have no immediate need to visualize energy grid maps. Energy grid map generation will create only one *bin* file in your work directory for a given protein target. Note that *bin* files can be converted into *map* files at any point of time by invoking Lead Finder with --grid-only, --load-grid and --save-grid parameters.

To automatically prepare a structure of streptavidin (PDB: 1SRE) at pH=5.5 for docking, calculate energy grid maps and save them in *bin* format, execute the following command:

```
leadfinder --grid-only --protein=1sre.pdb --pH=5.5 --save-grid=1sre.bin
```

Since no reference ligand is provided in this example, Lead Finder will automatically locate a ligand in the protein structure and use it as a reference ligand to define position and dimensions of the energy grid maps. Once the execution of this command is complete, the energy grid maps will be written into *1sre.bin* in your current directory. The prepared protein model will be saved into *1sre_protein.mol2*.

If you want to use an already prepared protein model, you can take it straight to the generation of energy grid maps:

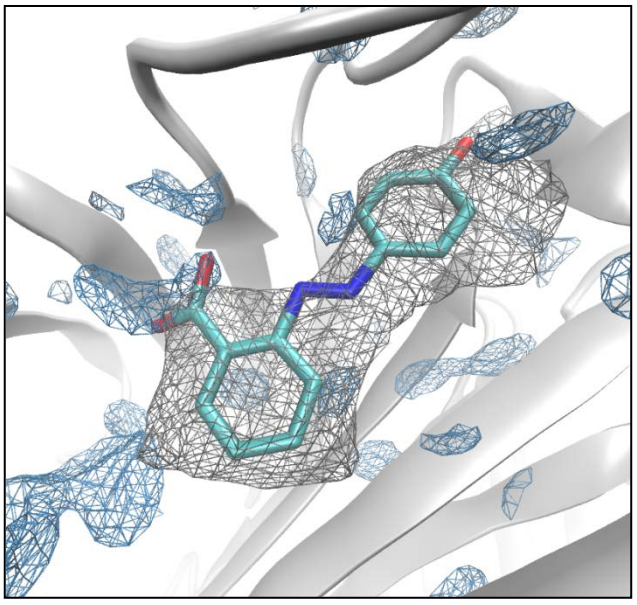
```
leadfinder --grid-only --protein=1sre_protein.mol2 --save-grid=1sre.bin
        --parameters=1sre.par
```

Note that this command provides that additional parameters must be taken from *1sre.par* (a parameter file). *1sre.par* contains parameters specifying position and dimensions of the active center for the generation of energy grid maps.

If a reference ligand is available (Lead Finder can extract reference ligand from the protein structure through --model-build-only parameter), use the following command line to calculate energy grid maps for the area around the reference ligand:

leadfinder --grid-only --protein=1sre_protein.mol2 --ligand-reference=1sre_reference.pdb --save-grid=1sre.bin

Note that the position and dimensions of the energy grid box generated using this alternative method will be different from the ones obtained by the previous command, since the reference ligand has different dimensions than the energy grid box specified in *1sre.par* file.



This picture is an example of visualization of grid maps (saved in the *map* format) in VMD software. The structure of streptavidin, 1SRE, is rendered in new cartoon, reference ligand (2-((4'-hydroxyphenyl)-azo)benzoic acid) - in licorice, VdW grid - in gray wireframe, H-bond donor grid – in blue wireframe. The reference ligand's shape seems to fit well into the protein's binding site and the ligand's carboxylic group is placed correctly for H-bonding interactions.

Other grid maps (electrostatic, etc.) may be loaded to VMD and visualized as well.

Example 3: Ligand Docking

This example uses a prepared 3D-structure of streptavidin (PDB 1SRE) as a receptor and a 3D-optimized structure of 2-(4-hydroxyphenyl)azobenzoic acid as a ligand.

Initially, Lead Finder computes energy grid maps for the protein structure. Then, the computed energy grid maps are saved into *1sre_prep.bin* file for future reuse by specifying

the --save-grid parameter. Once the energy grid computation and file saving is complete, Lead Finder proceeds to docking optimizations of a ligand structure imported from *lig_hpaba.mol* file. The docked poses are written into *lig_hpaba_docked.pdb* and a detailed text report is written into *lig_hpaba_report.log*. To start the process, execute:

```
leadfinder --save-grid=1sre_prep.bin --parameters=1sre.par
        --protein=1sre_prep.mol2    --ligand=lig_hpaba.mol
        --output-tabular=lig_hpaba.csv --verbose
```

Once docking is complete, *lig_hpaba_docked.pdb* will contain predicted docked ligand poses ranked by their Rank Score from the most favorable (pose 0) one to the least favorable one. *lig_hpaba.csv* will contain information on the predicted values of dG, VS Score and individual energy components for each of the docked poses in a spreadsheet format. Additionally, a text report in *lig_hpaba_report.log* will be generated.

When the chemical structure of a ligand matches that of a specified reference ligand, Lead Finder calculates RMSD values between predicted poses and the reference pose. Use the reference ligand extracted from the original 1SRE structure, saved into *1sre_lig_ref.pdb*, to evaluate performance of Lead Finder in predicting 3D poses of an experimentally resolved ligand structure:

```
leadfinder --load-grid=1sre_prep.bin --ligand-reference=1sre_lig_ref.pdb
        --ligand=lig_hpaba.mol --output-tabular=lig_hpaba.csv --verbose
```

Note that both *lig_hpaba.mol* and *1sre_lig_ref.pdb* can be used as a ligand structure for docking. However, only *1sre_lig_ref.pdb* is suitable for use as a reference ligand since it employs the same 3D coordinate system as the protein structure. The structure in *lig_hpaba.mol* is not suitable for use as a reference ligand since the *mol* format does not retain the PDB coordinate system.

Example 4: Virtual Screening

Like in previous examples, the same prepared 3D-structure of streptavidin (PDB 1SRE) is featured in this virtual screening example as a receptor. Create an energy grid map from the prepared protein structure and save it to a disk file as a first step of the process:

```
leadfinder --parameters=1sre_vs.par --grid-only --save-grid=1sre_prep.bin
```

ligands.sdf is a set of 32 arbitrary compounds, prepared for docking by assigning explicit hydrogen atoms, 3D optimization and protonation/deprotonation of appropriate ionazable groups. The compounds in this library possess 0 to 15 freely rotatable bonds (FRB). The average number of FRBs is 6.6. To start the virtual screening process, execute the following command:

```
leadfinder --parameters=1sre_vs.par --load-grid=1sre_prep.bin
        --ligand=ligands.sdf
```

Once processing is complete, *ligands_docked.sdf* will contain one top-scoring docked pose for each ligand with its predicted VS and dG scores. *ligands_scores.csv* will contain predicted scores and individual energy component information for each ligand in a spreadsheet format. A detailed text report about each docked ligand will be written into *ligands_report.txt*.

Note that in this example Lead Finder takes most of its configuration parameters from *1sre_vs.par*. The command line parameters are used along with the parameters from *1sre_vs.par.* Note that when the same parameter is specified both in the parameter file and in the command line, the command line value overrides the value from the parameter file.

Glossary

Binding affinity – the free energy of protein-ligand binding (kcal/mol).

dG score – Lead Finder's quantitative estimate of the ligand binding affinity.

Docking – positioning of a ligand in the protein binding site that minimizes free energy of the protein-ligand interaction.

Energy grid map – a 3D function mapping nodes in Cartesian space to the energy of a particular type of interaction that atoms of a particular type would possess when placed at the position of such node. The energy grid maps are calculated for each type of interactions considered by Lead Finder and for each atom type.

HASP key – the USB device that contains information about your software license.

Virtual screening – rank-ordering of compounds by their estimated binding potency with respect to a particular protein. Lead Finder docks each compound and calculates VS Score for the docked ligand. The value of VS Score is a quantitative estimate of ligand activity (binding potency).

VS score - Lead Finder's estimation of ligand activity (binding potency) for rank-ordering ligands in virtual screening. For details, see the Science section of BioMolTech Lead Finder's website at www.biomoltech.com.