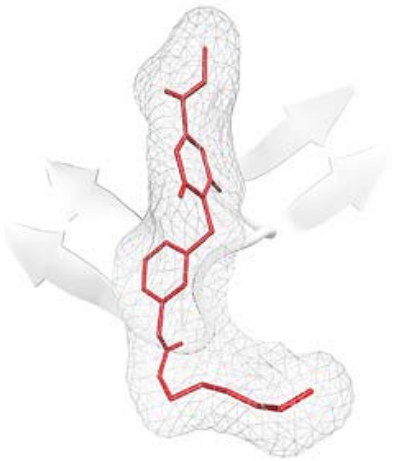
Lead-Finder

Software for Drug Discovery

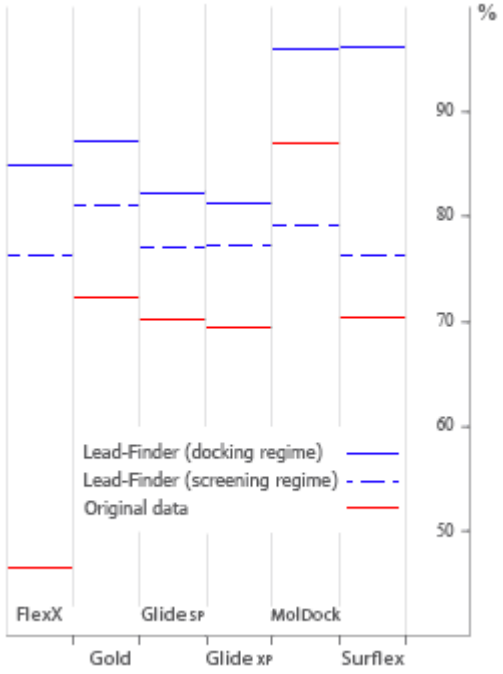


Lead-Finder software is an integrated solution for simulating structure and affinity of protein-ligand complexes. The software combines automatic processing of protein structures, extra precision protein-ligand docking and calculation of free energy of ligand binding.

Lead-Finder is intended to meet the requirements of computational and medicinal chemists involved in drug discovery, pharmacologists and toxicologists modeling ADMET properties *in silico*, and biochemists and enzymologists working on modeling protein-ligand interactions, enzyme specificity and rational enzyme design.

Lead-Finder combines extra precise protein-ligand docking and binding energy estimation with a high speed of calculations providing efficient solutions for the following tasks:

- **Ligand docking**
  high-accuracy prediction of the structure of non-covalent or covalently bound protein-ligand complexes (validated on the set of 407 structures)
- **Virtual screening**
  massive libraries of chemical compounds can be screened against a therapeutic target of interest to find potent binders with high fidelity (validated on 34 protein targets) and high speed of calculations (~5000 compounds per processor per day)
- **Binding energy estimation**
  extra precise estimation of the free energy of protein-ligand binding validated on the set of experimentally measured binding energies of 330 diverse protein-ligand complexes
- **Protein structure preparation**
  automatic preparation of fully functional protein structure models by adding hydrogen atoms to crude PDB coordinates according to optimal ionization state at a given pH



Docking success rate (%) of different programs obtained on their native test sets and current Lead-Finder benchmarks (in the default docking and fast screening regimes).

Unsurpassed Accuracy of Lead-Finder in Ligand Docking

Accuracy of protein-ligand docking was assessed on the set of 407 structures, which combines almost all published test sets of such programs as FlexX, Glide SP, Glide XP, Gold, LigandFit, MolDock, Surflex. As can be seen from the table, Lead-Finder outperforms all docking programs with docking success rate[1] ranging from 80.0% (for GlideXP and FlexX test sets) to 96.0% (for Surflex and MolDock test sets).

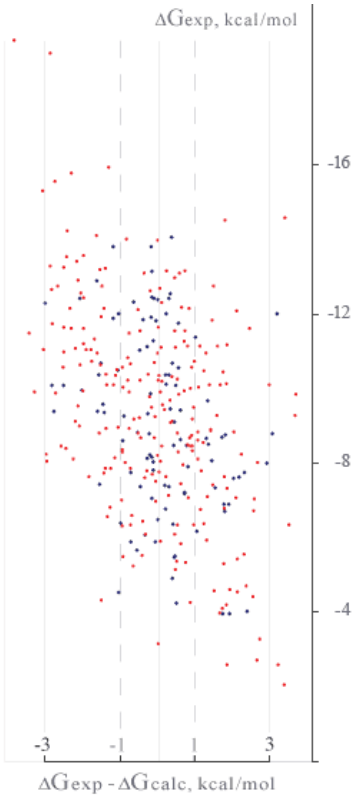
Table Docking success rate (%) of different programs obtained on their native test sets and current Lead-Finder benchmarks in default docking and screening regimes.

	FlexX [1]	Glide SP [2]	Glide XP [3]	Gold [4]	Gold [5]	Gold [6]	LigandFit [7]	MolDock [8]	Surflex [9]	**All test sets**
Original data	46.5	70.2	69.4	72.4	76.5	n/a	n/a	87.0	70.4	**n/a**
Lead-Finder (docking regime)	85.0	82.3	81.3	87.3	90.6	92.4	87.3	96.1	96.3	**85.0**
Lead-Finder (screening regime)	76.5	77.3	77.2	81.3	78.8	83.7	82.3	79.2	76.5	**79.0**
Number of structures	200	282	268	134	85	92	84	77	81	**407**

Accuracy of Binding Energy Estimations with Lead-Finder

Ability of Lead-Finder to estimate free energy of protein-ligand binding was benchmarked against the set of 330 diverse protein-ligand complexes, which is currently the most extensive benchmarking study of such kind. Lead-Finder demonstrated unique precision of binding energy prediction (RMSD = 1.5 kcal/mol) combined with high speed of calculations (less than one second per compound on average). These data illustrate that Lead-Finder is the most accurate in binding energy calculations compared to other docking programs [3, 10-11].

To make current parameterization and benchmarking studies more robust, protein-ligand complexes with known 3 D structure and experimentally measured binding constants for were chosen on the basis of maximal diversity of ligand's physicochemical properties (molecular weight (Mw), clog P, number of hydrogen bond donors (dHB) and acceptors (aHB), net charge) and range of protein – ligand binding energies.



Deviation of the predicted free energy of ligand binding plotted against the experimentally measured binding energy for the set of 330 protein-ligand complexes (blue dots correspond to the training set, red dots — to the test set).

[1] M. Rarey, B. Kramer, T. Lengauer, J Comp Aid Mol Des, 1997, 11, 369–384.
[2] R.A. Friesner, R.B. Murphy, M.P. Repasky, et al, J Med Chem 2004, 47, 1739-1749.
[3] R.A. Friesner, J.L. Banks, R.B. Murphy, et al, J Med Chem, 2006, 49, 6177-6196.
[4] G. Jones, P. Willett, R.C. Glen, et al, J Mol Biol, 1997, 267, 727-748.
[5] M.J. Hartshorn, M.L. Verdonk, G. Chessari, et al, J. Med Chem, 2007, 50, 726-741.
[6] J.W.M. Nissink, C. Murray, M. Hartshorn, et al, PROTEINS: Structure, Function, and Genetics, 2002, 49, 457–471.
[7] C.M. Venkatachalam, X. Jiang, et al, J Mol Graph Model, 2003,21, 289–307.
[8] R. Thomsen, M.H. Christensen, J Med Chem, 2006, 49, 3315–3321.
[9] A.N. Jain, J Med Chem, 2003, 46, 499-511.
[10] G. M. Morris, D. S. Goodsell, , et al, J Comput Chem, 1998, 19, 1639–1662
[11] R. Huey, G. M. Morris, A. J. Olson, et al, J Comput Chem, 2007, 28, 1145–1152

[1] Docking success rate is determined as a percentage of correctly docked ligands (for which top-scored pose was within 2 Å RMSD from the reference ligand coordinates) for a set of protein-ligand complexes extracted from PDB.

Lead-Finder Performance in Virtual Ligand Screening

Ability of Lead-Finder to distinguish active compounds from inactive during virtual screening experiments was benchmarked against 34 therapeutically relevant protein targets[2]. For quantitative characterization of virtual screening efficiency two well recognized parameters were used: area under the so-called receiver operating curve (ROC[3]) and enrichment factor (EF[4]). Results of virtual screening studies are presented in the Table below:

Protein target	PDB code	ROC	EF20	EF40	EF70	Number of ligands
Beta-secretase	1m4h	0.98	15.2	16.9	16.3	40
HIV-1 protease	1pro	0.98	16.5	13.2	13.4	50
Factor Xa	1fjs	0.98	14.4	12.8	11.4	50
Estrogen receptor antagonists	3ert	0.97	37.73	23.83	15.24	30
Ribonuclease A	1qhc	0.95	45.3	12.9	8.9	30
Epidermal growth factor receptor kinase	1m17	0.95	5.4	7.3	8.1	50
cAMP-dependent protein kinase	1fmo	0.94	7.9	6.2	6.6	50
Urokinase-type plasminogen activator	1gj7	0.94	5.8	6.9	7.3	20
p38 MAP kinase	1kv2	0.92	3.1	4.2	5.4	50
Acetylcholinesterase	1e66/1eve[4]	0.91	3.8	4.3	5.1	30
HSP90	1uy6	0.89	2.9	3.9	4.5	30
Lck kinase	1qpe	0.87	5.4	4.6	3.8	40
Estrogen receptor agonists	1l2i	0.86	1.65	2.29	2.66	30
Vascular endothelial growth factor receptor kinase 2	2oh4	0.86	7.2	4.0	3.7	50
Thermolysin	4tmn	0.86	22.4	16.6	3.7	20
Neuraminidase	2qwg	0.84	8.4	3.7	2.6	30
Thymidylate synthase	1f4g	0.77	4.5	3.2	2.3	15
Progesteron receptor	1sr7	0.76	2.7	2.1	2.0	20
Oligopeptide-binding protein	1b5j	1.00	111.8	89.4	78.3	16
Orotidine-5'-P decarboxylase	1eix	0.99	56.0	64.0	26.1	18
Protein tyrosine phosphatase 1B	1c84	0.99	73.8	55.4	11.7	20
Peroxisome proliferator activated receptor gamma	1fm9	0.98	9.6	11.8	11.7	50
Ribonuclease T1	1rnt	0.97	74.1	74.1	35.4	10
Thrombin	1c4v	0.96	12.0	8.6	10.8	40
Tyrosine kinase C-SRC	2src	0.96	11.0	12.1	7.7	50
Trypsin	1qbo	0.95	9.0	9.4	9.5	20
Thymidine kinase	1kim	0.94	31.8	27.8	20.5	10
Mineralocorticoid receptor	2aa2	0.94	10.1	5.8	10.4	10
Poly(ADP-ribose) polymerase	1efy	0.92	3.5	5.7	7.6	10
Penicillopepsin	1bxo	0.91	6.0	8.2	5.5	6
Cyclooxygenase-2	1cx2	0.91	2.8	4.0	5.1	50
Fibroblast growth factor receptor kinase (FGFR)	1fgi	0.86	2.7	3.0	3.6	50

For almost all targets Lead-Finder demonstrated impressive enrichment indicators, as can be seen from the Table, confirming its applicability for *in silico* lead compound discovery.

[2] Benchmarking experiments consisted of creating a test library of compounds (obtained by mixing active ligands with the set of 1904 decoy compounds freely available from Surflex developers at http://www.jainlab.org/downloads.html), docking each compound from the library to each particular target, and rank-ordering compounds according to calculated VS-score.

[3] Area under ROC plot (or simply ROC) is an integral parameter of virtual screening performance and corresponds to the area under the curve built according the rule: for a given fraction of screened library the Y-coordinate denotes the fraction of active compounds found, the X-coordinate represents the fraction of the inactive ligands (decoys) compounds. An ideal curve reflects 100% of true actives found, and 0% of decoys; this ideal curve returns ROC=1

[4] Enrichment factor is calculated for a certain percentage of active compounds, and is defined as the fraction of active compounds found divided by the fraction of the screened library. For example, EF70 denotes enrichment factor at 70% of active ligands found, EF100 — enrichment at 100% of actives found.

Speed of Calculations

Lead-Finder allows screening massive libraries of chemical compounds in a reasonable time preserving accuracy of ligand docking and energy estimations due to unique docking algorithm and extra precision scoring function.

Speed of ligand docking was benchmarked in a number of real-life virtual screening studies, in which commercial library of 300000 compounds (STK library by Vitas-M Laboratory http://www.vitasmlab.com/compound-libraries-2.htm) was screened against a number of protein targets. Using a cluster of 16 computational nodes (dual Xeon 5150 2.6 GHz) we were able to screen the whole library in 8-20 hours depending on the protein target. Average time for docking on compound from the library for particular protein is provided in the Table below.

Protein Target	Time per ligand, seconds	Total screening time, hours
Monoamine oxydase A	9.00	7.50
HIV-1 reverse transcriptase	10.44	8.70
Thyroid hormone receptor beta1	11.02	9.19
Vitamin D receptor	13.42	11.18
Dihydrofolate reductase	14.98	12.48
HMG-CoA reductase	15.20	12.67
Leukotriene A4 hydrolase	15.79	13.16
Peroxisome proliferator-activated receptor gamma	18.43	15.36

Unique Docking Algorithm

To tackle computationally challenging problem of ligand docking Lead-Finder applies unique approach combining genetic algorithm search, local optimization procedures, smart exploitation of the knowledge generated during the search run. Rational combination of different optimization strategies makes Lead-Finder efficient in terms of coarse sampling of ligand's phase space and refinement of promising solutions.

Entire docking algorithm has several tenths of settings, each of which has certain impact on the algorithm performance, robustness and speed of calculations. These relations were studied by us and for the sake of user comfort all settings were zipped in to two regimes of ligand docking algorithm with carefully balanced speed/accuracy ratio. First one is the default docking regime, which settings were adjusted to achieve maximum docking accuracy at a reasonable time of calculations. Second is the so-called screening regime, which was designed to be maximally fast at the cost of small (~5%) decrease in docking accuracy (for details see section Technology at www.moltech.ru).

Extra Precision Scoring Function

Extra precise representation of protein-ligand interactions implemented in Lead-Finder scoring function is the second (in addition to docking algorithm) component of successful ligand docking. Lead-Finder scoring function is based on a semi-empiric molecular mechanical functional, which explicitly accounts for different types of molecular interactions. Individual energy contributions are scaled with empiric coefficients to fit particular purposes: accurate binding energy predictions, correct energy-ranking of docked ligand poses, correct rank-ordering of active and inactive compounds during virtual screening experiments. For these reasons three distinct types of scoring functions based on the same set of energy contributions but different sets of energy-scaling coefficients are used by Lead-Finder. More details about the Lead-Finder scoring function can be found in Technology section at www.moltech.ru.

Contact Info

Address: 119992, Russian Federation, Moscow, Leninskie gory, 1/75A
Tel/Fax: +7(495)9394653
e-mail: info@moltech.ru URL www.moltech.ru
Contact person: Ghermes Chilov, CEO