

Surflex-Dock™

LIGAND-RECEPTOR DOCKING AND VIRTUAL SCREENING



Surflex-Dock offers unparalleled enrichments in virtual high-throughput screening¹ combined with state-of-the-art speed, accuracy and usability. It uses an empirical scoring function (based on the Hammerhead docking system) that has been updated and re-parameterized with additional negative training data⁵, along with a search engine that relies on a surface-based molecular similarity method^{3,4}.

Virtual High Throughput Screening

Screening large compound libraries remains an expensive and time-consuming task. Computational high-throughput screening can enrich the fraction of suitable compounds in a screening collection and thereby reduce the cost of lead discovery.

Two aspects determine the quality of a docking method: docking accuracy and screening enrichment. Docking accuracy measures the likelihood that a method will correctly identify and recognize the true binding mode of a ligand bound to a target protein. Screening enrichment measures the relative improvement in the identification of true binding ligands using a docking method versus random screening.

Surflex-Dock addresses the protein ligand docking problem, demonstrating a novel approach that is faster, more accurate, and is five to ten-fold more specific than competing methods in detecting true binding ligands from non-binding ligands in computational screening applications^{1,5}. The docking search engine makes use of a patented³ technology.

The Algorithm

Surflex-Dock employs an idealized active site ligand called a protomol⁷, as a target to generate putative poses of molecules or molecular fragments. These putative poses are scored using the Hammerhead scoring function⁸, which also serves as an objective function for local optimization of poses. Flexible docking proceeds by a crossover procedure that combines pieces of poses from intact molecules⁶. Molecular alignment is done through a patented similarity-based algorithm^{3,4}.

Protein-Ligand Interaction

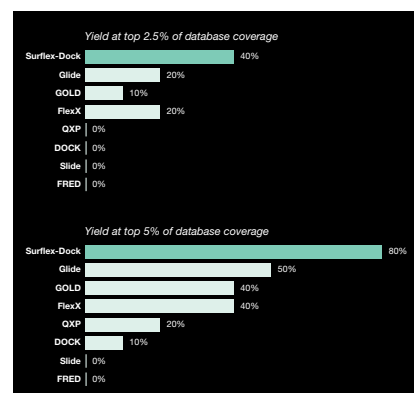
The Surflex-Dock scoring function is a linear combination of non-linear functions of protein-ligand atomic surface distances. Protein-ligand interactions include steric, polar, entropic and solvation terms⁶.

Selected Features

- **Accurate Scoring** – Scoring function is derived from known binding affinity data and negative training data to reduce false positive binding scores.
- **Speed** – On average 17s per ligand (~3s per rotatable bond). Faster than GOLD and GLIDE and comparable with FlexX.
- **Easier Protein Preparation** – Protein structure preparation is simple and part of the docking workflow. No special treatment of cofactors/metals/waters is required.
- **Easier Docking Preparation** – Docking gives good results with default settings; users will typically only need to tweak a few parameters to optimize docking for a specific target.
- **Protomol guided docking** – Docking is guided by a 'protomol' which can be automatically generated and/or user-defined. The protomol is an idealized representation of a ligand that makes every potential interaction with the binding site. Stored in mol2 format, it can be edited easily and on the fly.
- **Ring Flexing** – Generic ring conformations are applied to each flexible ring system (regardless of atom types), and the resulting conformations are minimized. The method is robust and fast.
- **Parallelization** – A simple parallelization protocol exists to enable the use of multiple processors.

Enrichment Factors

Recovery of 10 Thymidine Kinase inhibitors from a database of 1000 drug-like molecules: after docking and ranking Surflex-Dock returns 80% of the true hits in the top 5% of the database. This performance is typical in the majority of cases⁵.



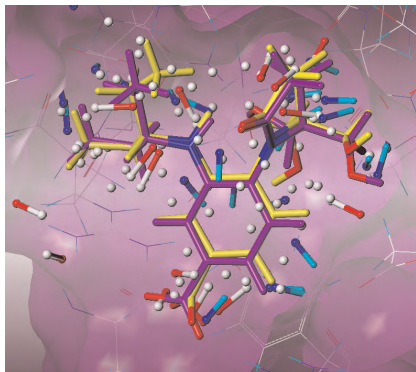
Hit Rate: (actives recovered/total hits recovered) x 100
Yield: (actives recovered/total actives) x 100
 Enrichment in thymidine kinase inhibitors. The percentage of known actives recovered by virtual screening within the 2.5% and 5% top scoring fraction of the database¹.

	Top 2.5%		Top 5%	
	Hit Rate (%)	Yield (%)	Hit Rate (%)	Yield (%)
Surflex-Dock	16	40	16	80
Glite	8	20	10	50
GOLD	4	10	8	40
FlexX	8	20	8	40
QXP	0	0	4	20
DOCK	0	0	2	10
Slide	0	0	0	0
FRED	0	0	0	0

Description of hit lists generated by 8 docking tools on the thymidine kinase inhibitors. Hit lists are generated from the top 2.5% and 5% scoring compounds¹. Reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

Complementary Software

- Concord[®] for generating accurate 3D conformations
- CScore[™] for ranking the affinity of compounds bound to a target with consensus scoring
- Biopolymer[™] for macromolecular 3D structure modeling
- MOLCAD[™] for advanced visualization of molecular surfaces and properties



Influenza virus neuraminidase (1B9V) in complex with an inhibitor (purple capped sticks). The minimized inhibitor has been redocked by Surflex-Dock into the protein (yellow capped sticks) with an rms deviation of 0.645 Angstroms.

Performance

On average 17s per ligand (~3s per rotatable bond).

Hardware and Software Requirements

Surflex-Dock 2.0 is available through the SYBYL interface and as such runs on IRIX[®] (SGI[®]) or Linux[®] (x86). A Windows[®] version of command-line only Surflex-Dock 2.0 is also available.

Acknowledgements

Scientific and Software Partner: Prof. Ajay N. Jain, UCSF Cancer Center, San Francisco, & BioPharmics LLC.

References

1. Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. (2004). Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. *Proteins*, 57, 225-242
2. Welch, W., Ruppert, J. & Jain, A.N. (1996). Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol*, 3, 449-62.
3. United States Patent No. 6,470,305; 20
4. Jain A.N. (2000). Morphological similarity: A 3D molecular similarity method correlated with protein-ligand recognition. *J Comput Aided Mol Des*, 14, 199-213
5. Pham, T. A. and Jain, A.N. (JMC ASAP-Web 2005, to appear in print soon). Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J Med Chem*
6. Jain, A.N. (2003). Surflex: Fully Automatic Flexible Molecular Docking using a Molecular Similarity-Based Search Engine. *J Med Chem* 46, 499-511.
7. Ruppert, J., Welch, W. & Jain, A.N. (1997). Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci* 6, 524-33.
8. Jain, A.N. (1996). Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des* 10, 427-40.

Further Reading

- Cleves, A. E. and Jain, A.N. (2006). Robust Ligand-Based Modeling of the Biological Targets of Known Drugs. *J Med Chem*. 49 2921-2938.
- Jain, A.N. (2004). Virtual Screening in Lead Discovery and Optimization. *Curr Op Drug Discovery & Dev* 7, 396-403.
- Jain, A.N. (2004). Ligand-Based Structural Hypotheses for Virtual Screening. *J Med Chem* 47, 947-961.
- Perkins, E., Sun, D., Nguyen, A., Tulac, S., Francesco, M., Tavana, H., Nguyen, H., Tugendreich, S., Barthmaier, P., Couto, J., Yeh, E., Thode, S., Jarnagin, K., Jain, A.N., Morgans, D. & Melese, T. (2001). Novel inhibitors of poly(ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. *Cancer Res* 61, 4175-83.
- Jain, A.N. (2000). Morphological similarity: A 3D molecular similarity method correlated with protein-ligand recognition. *J Comput Aided Mol Des* 14, 199-213.
- Mount, J., Ruppert, J., Welch, W. & Jain, A.N. (1999). IcePick: A flexible surface-based system for molecular diversity. *J Med Chem* 42, 60-6.
- Ghuloum, A.M., Sage, C.R. & Jain, A.N. (1999). Molecular hashkeys: A novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J Med Chem* 42, 1739-48.
- Jain, A.N., Harris, N.L. & Park, J.Y. (1995). Quantitative binding site model generation: compass applied to multiple chemotypes targeting the 5-HT1A receptor. *J Med Chem* 38, 1295-308.
- Jain, A.N., Dieterich, T.G., Lathrop, R.H., Chapman, D., Critchlow, R.E., Jr., Bauer, B.E., Webster, T.A. & Lozano-Perez, T. (1994). A shape-based machine learning tool for drug design. *J Comput Aided Mol Des* 8, 635-52.



WWW.TRIPOS.COM				CONTACT_US@TRIPOS.COM		
AUSTRALIA +61 (7) 5439 9775	CANADA +1 450 4334500	FRANCE +33 1 69 59 29 49	GERMANY +49 89 45 10 300	JAPAN +81 3 5166 1721	UNITED KINGDOM +44 1 908 650000	UNITED STATES 800-323-2960 +1-314-647-1099
©Tripos, Inc. 2006 All rights reserved. 0075 Printed in USA All trademarks are the property of their respective owners.						