REVIEW OF DOCKING SOFTWARE

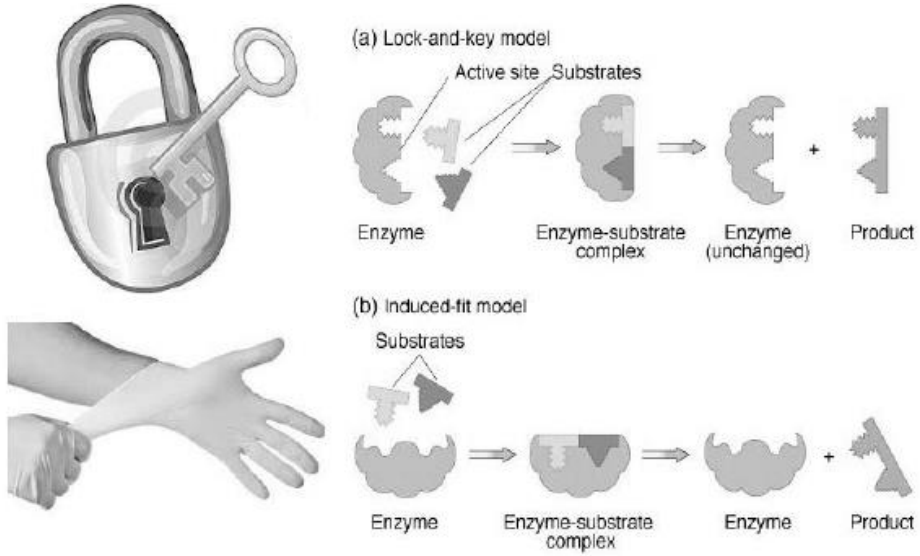
Mateusz Chojnacki

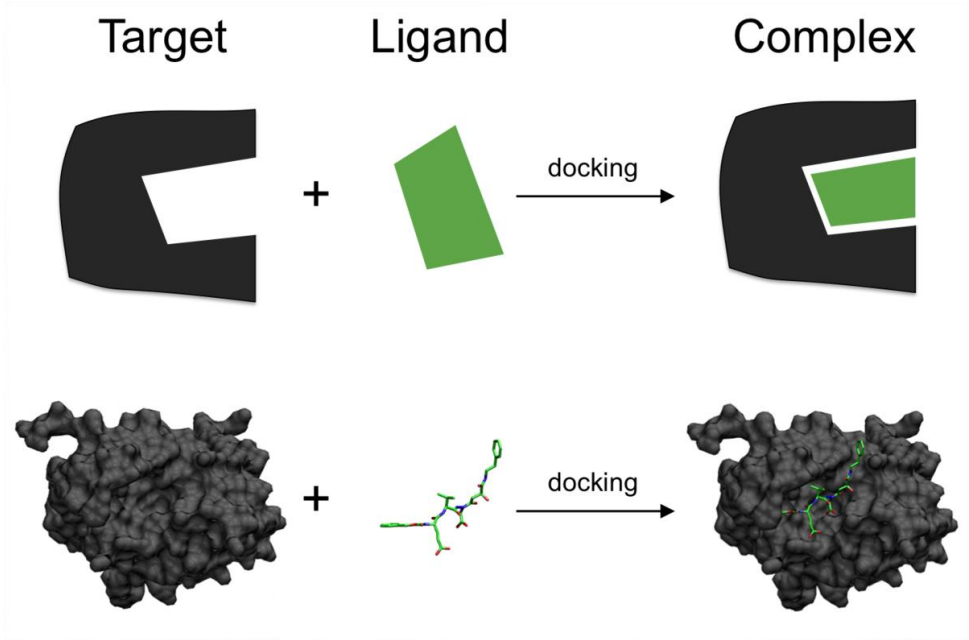
Table of content

Docking problem

Models:

- Lock-and-key
- Hand-in-the-glove
- Induced-fit





https://commons.wikimedia.org/wiki/File:Docking_representation_2.png#/media/File:Docking_representation_2.png

Mukhopadhyay, Mayukh. (2014). A brief survey on bio inspired optimization algorithms for molecular docking. International Journal of Advances in Engineering & Technology. 7. 868-878. 10.7323/ijaet/v7_iss3.

There are many avaialble softwares…

Docking

Software

- Autodock. Free open source EA based docking software. Flexible ligand. Flexible protein side chains. Maintained by the Molecular Graphics Laboratory, The Scripps Research Institute, la Jolla.
- DOCK. Anchor-and-Grow based docking program. Free for academic usage. Flexible ligand. Flexible protein. Maintained by the Soichet group at the UCSF.
- GOLD. GA based docking program. Flexible ligand. Partial flexibility for protein. Product from a collaboration between the university of Sheffield, GlaxoSmithKline plc and CCDC.
- Glide. Exhaustive search based docking program. Exists in extra precision (XP), standard precision (SP) and virtual High Throughput Screening modes. Ligand and protein flexible. Provided by Schrödinger.
- Itzamna. Itzamna is a docking program, identifying active compounds for a given target. You can upload a protein and a docking is performed, either against an in-house database containing more than a million active compounds, or against a user-defined library. Provided by Mind The Byte.
- SCIGRESS. Desktop/server molecular modeling software suite employing linear scaling semiempirical quantum methods for protein optimization and ligand docking. Developed and distributed by Fujitsu, Ltd.
- GlamDock. Docking program based on a Monte-Carlo with minimization (basin hopping) search in a hybrid interaction matching / internal coordinate search space. Part of the Chil^2 suite. Open for general research.
- FlexAID. A small-molecule docking algorithm that accounts for target side-chain flexibility and utilizes a soft scoring function. The pairwise energy parameters were derived from a large dataset of true positive poses and negative decoys from the PDBbind database through an iterative process using Monte Carlo simulations. Precompiled Linux, MacOS and Windows versions are made available by the University of Sherbrooke, Canada.
- GEMDOCK. Generic Evolutionary Method for molecular DOCKing. Program for computing a ligand conformation and orientation relative to the active site of target protein==== Docking - Software ====
- iGEMDOCK. Graphic environment for the docking, virtual screening, and post-screening analysis. Free for non commercial researches. For Windows and Linux.
- HomDock. Progam for similarity-based docking, based on a combination of the ligand based GMA molecular alignment tool and the docking tool GlamDock. Part of the Chil^2 suite. Open for general research.
- ICM. Docking program based on pseudo-Brownian sampling and local minimization. Ligand and protein flexible. Provided by MolSoft.
- FlexX, Flex-Ensemble (FlexE). Incremental build based docking program. Flexible ligand. Protein flexibility through ensemble of protein structure. Provided by BioSolveIT.
- Fleksy. Program for flexible and induced fit docking using receptor ensemble (constructed using backbone-dependent rotamer library) to describe protein flexibility. Provided by the Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen.
- FITTED. (Flexibility Induced Through Targeted Evolutionary Description). Suite of programs to dock flexible ligands into flexible proteins. This software relies on a genetic algorithm to account for flexibility of the two molecules and location of water molecules, and on a novel application of a switching function to retain or displace water molecules and to form potential covalent bonds (covalent docking) with the protein side-chains. Part of the Molecular FORECASTER package and FITTED Suite. Free for an academic site license (excluding cluster).
- FORECASTER. Standalone interface that contains applications to perform docking and more. It includes the FITTED docking program, the sites of metabolism prediction IMPACTS, and the accessory programs to work with the proteins and the ligands. It comes with a java-based graphical interface that integrated all the program into workflows. Provided by Molecular Forecaster Inc.
- VLifeDock. Multiple approaches for protein - ligand docking. Provides three docking approches: Grid based docking, GA docking and VLife's own GRIP docking program. Several scoring functions can be used: PLP score, XCscore and Steric + Electrostatic score. Available for Linux and Windows. Provided by VLife.
- ParaDockS. (Parallel Docking Suite). Free, open source program, for docking small, drug-like molecules to a rigid receptor employing either the knowledge-based potential PMF04 or the empirical energy function p-Score.
- Molegro Virtual Docker. Protein-ligand docking program with support for displaceable waters, Induced-fit-docking, user-defined constraints, molecular alignment, ligands-based screening, and KNIME workflow integration. Distributed by Qiagen.
- DAIM-SEED-FFLD. Free open source fragment-based docking suite. The docking is realized in three steps. DAIM (Decomposition And Identification of Molecules) decomposes the molecules into molecular fragments that are docked using SEED (Program for docking libraries of fragments with solvation energy evaluation). Finally, the molecules are reconstructed "in situ" from the docked fragments using the FFLD program (Program for fragment-based flexible ligand docking). Developed and maintained by the Computational Structural Biology of ETH, Zurich, Switzerland.
- Autodock Vina. MC based docking software. Free for academic usage. Flexible ligand. Flexible protein side chains. Maintained by the Molecular Graphics Laboratory, The Scripps Research Institute, la Jolla.
- VinaMPI. Massively parallel Message Passing Interface (MPI) program based on the multithreaded virtual docking program AutodockVina. Free and open source. Provided by the University of Tennessee.
- FlipDock. GA based docking program using FlexTree data structures to represent a protein-ligand complex. Free for academic usage. Flexible ligand. Flexible protein. Developed by the Department of Molecular Biology at the Scripps Research Institute, la Jolla.
- PharmDock. A protein pharmacophore-based docking program. PharmDock and a PyMOL plugin are made freely available by the Purdue University, West Lafayette, USA.
- FRED. FRED performs a systematic, exhaustive, nonstochastic examination of all possible poses within the protein active site, filters for shape complementarity and pharmacophoric features before selecting and optimizing poses using the Chemgauss4 scoring function. Provided by OpenEye scientific software.
- POSIT. POSIT uses the information from bound ligands to improve pose prediction. Using a combination of approaches, including structure generation, shape alignment and flexible fitting, a ligand of interest is compared to bound ligands and its similarity to such both guides the nature of the applied algorithm and produces an estimate. Both 2D and 3D similarity measures are used in this reliability index. Provided by OpenEye scientific software.
- HYBRID. Docking program similar to FRED, except that it uses the Chemical Gaussian Overlay (CGO) ligand-based scoring function. Provided by OpenEye scientific software.
- idock. Free and open source multithreaded virtual screening tool for flexible ligand docking for computational drug discovery. Developed by the Chinese university of Hong Kong.
- POSIT. Ligand guided pose prediction. POSIT uses bound ligand information to improve pose prediction. Using a combination of several approaches, including structure generation, shape alignment and flexible fitting, it produces a predicted pose whose accuracy depends on similarity measures to known ligand poses. As such, it produces a reliability estimate for each predicted pose. In addition, if provided with a selection of receptors from a crystallographic series, POSIT will automatically determine which receptor is best suited for pose prediction. Provided by OpenEye scientific software.
- Rosetta Ligand. Monte Carlo minimization procedure in which the rigid body position and orientation of the small molecule and the protein side-chain conformations are optimized simultaneously. Free for academic and non-profit users.
- Surflex-Dock. Docking program based on an idealized active site ligand (a protomol), used as a target to generate putative poses of molecules or molecular fragments, which are scored using the Hammerhead scoring function. Distributed by Tripos.
- CDocker. CHARMm based docking program. Random ligand conformations are generated by molecular dynamics and the positions of the ligands are optimized in the binding site using rigid body rotations followed by simulated annealing. Provided by Accelrys.
- LigandFit. CHARMm based docking program. Ligand conformations generated using Monte-Carlo techniques are initially docked into an active site based on shape, followed by further CHARMm minimization. Provided by Accelrys.
- rDock. Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. Free and open source.Developed by the University of Barcelona.
- KIN. Kin is a blind-docking technology. All potential cavities of a given protein are predicted, and a query molecule is docked inside each of them, sorting results by scoring function. Distributed by Mind The Byte.
- Lead Finder. program for molecular docking, virtual screening and quantitative evaluation of ligand binding and biological activity.Distributed by Moltech. For Windows and linux.
- YASARA Structure. Adds support for small molecule docking to YASARA View/Model/Dynamics using Autodock and Fleksy. Provided by YASARA.
- ParaDockS. ParaDockS includes algorithms for protein-ligand docking and is organized that every newly developed scoring function can be immediately implemented. Furthermore, interaction-based classifier, trained on a target-specific knowledge base can be used in a post-docking filter step. An implementation and validation of target-biased scoring methods within the open-source docking framework is implemented. developed and provided free of charge by the University of Halle-Wittenberg, Germany.

https://www.click2drug.org/directory_Docking.php

Small mollecues docking

There are many available software's that can be used to small molecular docking, both with open-source access and commercial, e.g.:

- AutoDock
- AutoDock Vina
- rDock
- Rosetta Ligand
- Schrödinger Glide
- ...

AutoDock

https://ccsb.scripps.edu/

- available on Linux / macOS / Windows
- open source
- under GNU GPL
- first publication: 1990 (AutoDock1)
- current version: AutoDock 4.2.6

- no web server version
- more information:

https://autodock.scripps.edu/

AutoDock

Short description:
A computational docking program based on an empirical free energy force field and rapid Lamarckian genetic algorithm search method.

There are many tools working with AutoDock e.g. results visualisation (Raccoon v1.0), docking of specific cases (Multiple ligand docking, AutoDock covalent docking, ...) and target analasys (AutoLigand, AutoSite).

AutoDock – step by step

Start: protein– PDB file , ligand MOL/MOL2 or SDF file.

1. preparation of coordinate files using AutoDockTools (PDB -> PDBQT, create run.dpf, create macro.gpf)
2. precalculation of atomic affinities using AutoGrid

> autogrid4 -p macro.gpf [-l out_log_macro.glg]

3. docking of ligands using AutoDock,

> autodock4 [-i][-u][-t] -p run.dpf [-l out_log.dlg]

4. analysis of results using AutoDockTools

AutoDockTools has window interface, while AutoGrid and AutoDock are in command line.

AutoDock Vina

https://ccsb.scripps.edu/

AutoDock Vina

- available on Linux / macOS / Windows
- open source
- under Apache License, Version 2.0
- first publication: 2010
- current version: AutoDock Vina 1.2.5

Short description:
A turnkey computational docking program based on a simple scoring function and rapid gradient-optimization conformational search.

- no web server version
- more information:

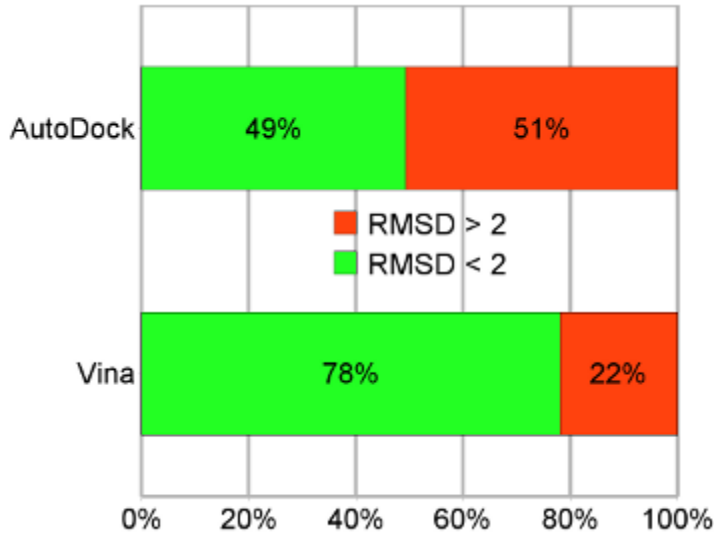
https://vina.scripps.edu/

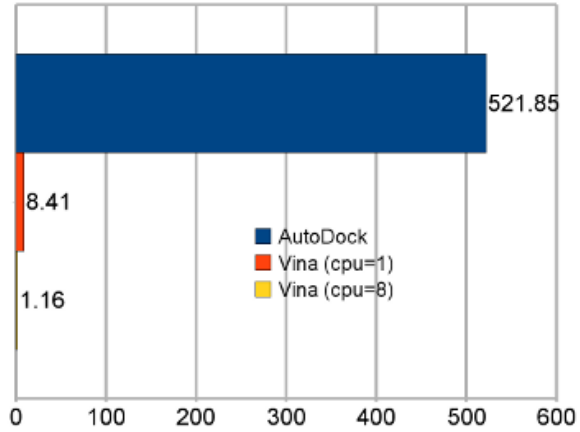
AutoDock Vina vs. AutoDock

- Vina is much more automated than AutoDock and simpler to use.
- Vina support working on CPU.
- Docking steps – the same as in AutoDock, but usually with default Vina parameters; can make grid automaticly.
- Exmplary config.txt file

```
receptor = 1iep_receptorH.pdbqt
ligand = 1iep_ligandH.pdbqt
center_x = 15.19
center_y = 53.903
center_z = 16.917
size_x = 15.0
size_y = 17.25
size_z = 15.0
out = 1iep_ligandH_out.pdbqt
```

- ./vina --config config.txt --exhaustiveness=24





Average time per receptor-ligand pair on the test set. "AutoDock" refers to AutoDock 4, and "Vina" to AutoDock Vina 1.

rDock

rDock

A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids

- available on Linux
- open source (since 2012)
- under GNU-LGPL version 3.0
- first publication: 1998
- current available also in bioconda package

https://www.york.ac.uk/

- no web server version
- more information:

https://rdock.github.io/

rDock step by step

Input: receptor - mol2 file, ligand – sd file

- System Definition (create param file PRMFILE)

- Cavity generation

> rbcavity -was -d -r < PRMFILE >

- Docking

> rbdock -i <INPUT>.sd -o <OUTPUT> -r <PRMFILE> -p dock.prm -n 50

```
RBT_PARAMETER_FILE_V1.00
TITLE gart_DUD

RECEPTOR_FILE gart_rdock.mol2
RECEPTOR_FLEX 3.0

################################################################
### CAVITY DEFINITION: REFERENCE LIGAND METHOD
################################################################
SECTION MAPPER
    SITE_MAPPER RbtLigandSiteMapper
    REF_MOL xtal-lig.sd
    RADIUS 6.0
    SMALL_SPHERE 1.0
    MIN_VOLUME 100
    MAX_CAVITIES 1
    VOL_INCR 0.0
   GRIDSTEP 0.5
END_SECTION

################################
#CAVITY RESTRAINT PENALTY
################################
SECTION CAVITY
    SCORING_FUNCTION RbtCavityGridSF
    WEIGHT 1.0
END_SECTION
```

https://rdock.github.io/docking-in-3-steps/

Rosetta Ligand

https://meilerlab.org/

- available on Linux/macOS
- commercial
- first publication: 2006

- available web server (for academic uses)

https://rosie.graylab.jhu.edu/ligand_docking/submit

- more information:

https://meilerlab.org/rosetta-ligand/

https://www.rosettacommons.org/demos/latest/tutorials/ligand_docking/ligand_docking_tutorial

Rosetta Ligand

Input: receptor – PDB file, ligand – sd file

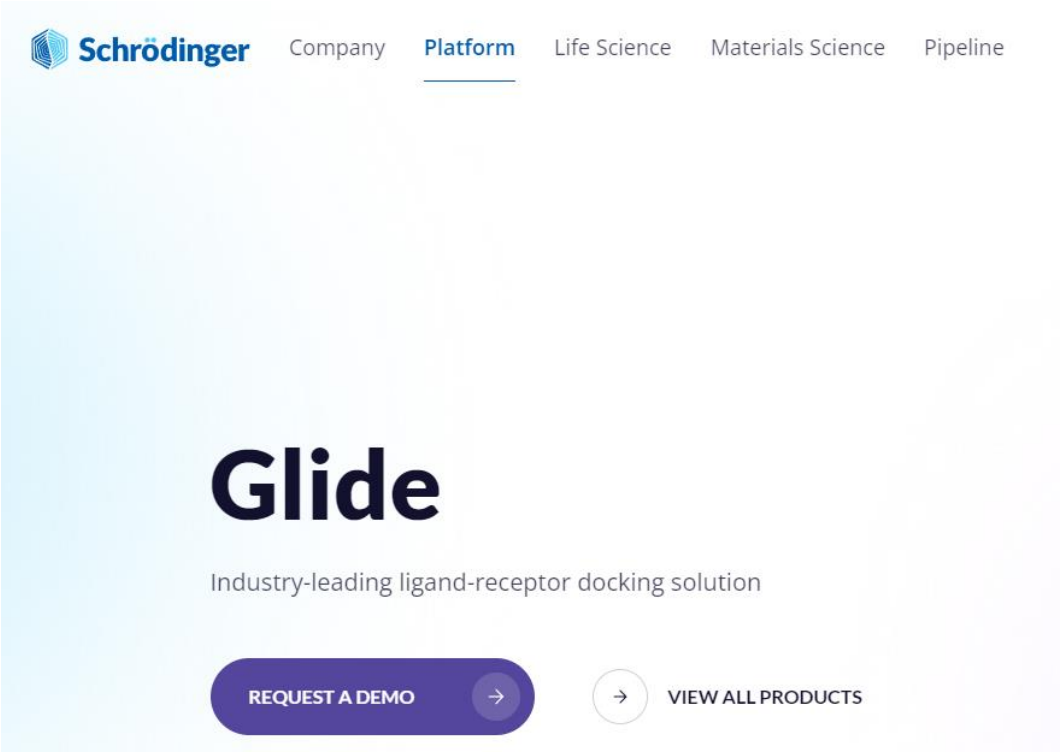
- generate conformers for ligand
- generate parametr file for ligand and its conformers
- prepare RosettaScript for docking
- prepare options file
- run RosettaScript

```
This protocol will simply do low-resolution followed by high-resolution docking.
It will also report the binding energy (ddg) and buried-surface area (sasa) in the score file.
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <ligand_soft_rep weights=ligand_soft_rep>
                        <Reweight scoretype=hack_elec weight=0.42/>
                </ligand_soft_rep>
                <hard_rep weights=ligand>
                        <Reweight scoretype=hack_elec weight=0.42/>
                </hard_rep>
        </SCOREFXNS>
        <LIGAND_AREAS>
                <docking_sidechain chain=X cutoff=6.0 add_nbr_radius=true all_atom_mode=true minimize_ligand=10/>
                <final_sidechain chain=X cutoff=6.0 add_nbr_radius=true all_atom_mode=true/>
                <final_backbone chain=X cutoff=7.0 add_nbr_radius=false all_atom_mode=true Calpha_restraints=0.3/>
        </LIGAND_AREAS>
        <INTERFACE_BUILDERS>
                <side_chain_for_docking ligand_areas=docking_sidechain/>
                <side_chain_for_final ligand_areas=final_sidechain/>
                <backbone ligand_areas=final_backbone extension_window=3/>
        </INTERFACE_BUILDERS>
        <MOVEMAP_BUILDERS>
                <docking sc_interface=side_chain_for_docking minimize_water=true/>
                <final sc_interface=side_chain_for_final bb_interface=backbone minimize_water=true/>
        </MOVEMAP_BUILDERS>
        <MOVERS>
        single movers
                <StartFrom name=start_from chain=X>
                        <Coordinates x=-1.731 y=32.589 z=-5.039/>
                </StartFrom>
                <Translate name=translate chain=X distribution=uniform angstroms=0.01 cycles=50/>
                <Rotate name=rotate chain=X distribution=uniform degrees=360 cycles=1000/>
                <SlideTogether name=slide_together chain=X/>
                <HighResDocker name=high_res_docker cycles=6 repack_every_Nth=3 scorefxn=ligand_soft_rep movemap_builder=docking/>
                <FinalMinimizer name=final scorefxn=hard_rep movemap_builder=final/>
                <InterfaceScoreCalculator name=add_scores chains=X scorefxn=hard_rep native="inputs/7cpa_7cpa_native.pdb"/>
        compound movers
                <ParsedProtocol name=low_res_dock>
                        <Add mover_name=start_from/>
                        <Add mover_name=translate/>
                        <Add mover_name=rotate/>
                        <Add mover_name=slide_together/>
                </ParsedProtocol>
                <ParsedProtocol name=high_res_dock>
                        <Add mover_name=high_res_docker/>
                        <Add mover_name=final/>
                </ParsedProtocol>
        </MOVERS>
        <PROTOCOLS>
                <Add mover_name=low_res_dock/>
          <Add mover_name=high_res_dock/>
                <Add mover_name=add_scores/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

Exemplary RosettaScript for docking

Schrödinger Glide

- available on Linux/macOS/Windows
- commercial
- first publication: 2004



https://newsite.schrodinger.com/platform/products/glide/

- available window (via Maestro) and command line versions

- available web server (with bought licence)
- possible to make University server for Glide

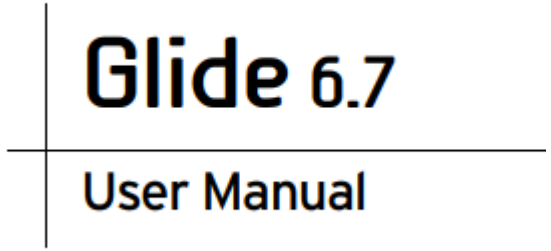
Schrödinger Glide

Input: receptor – PDB file, ligand – sd/PDB/Maestro/mol2 file

- Receptor preparation (e.g. Protein Preparation Wizard)
- Ligand preparation (e.g. Maestro, LigPrep)
- Receptor grid genaration and ligand grid genaration
- Boxes positioning
- Constraints selection (for receptor and ligand)
- Setting docking parameters (e.g. precision, output format,
- Run docking

Insruction step by step:

https://gohom.win/ManualHom/Schrodinger/Schrodinger_2015-2_docs/glide/glide_user_manual.pdf



Peptide docking

Problem of docking peptides with given only sequence is much more complicated problem than docking small molecules, due to multiple ways of peptide's folding according to its environment and co-moledues.

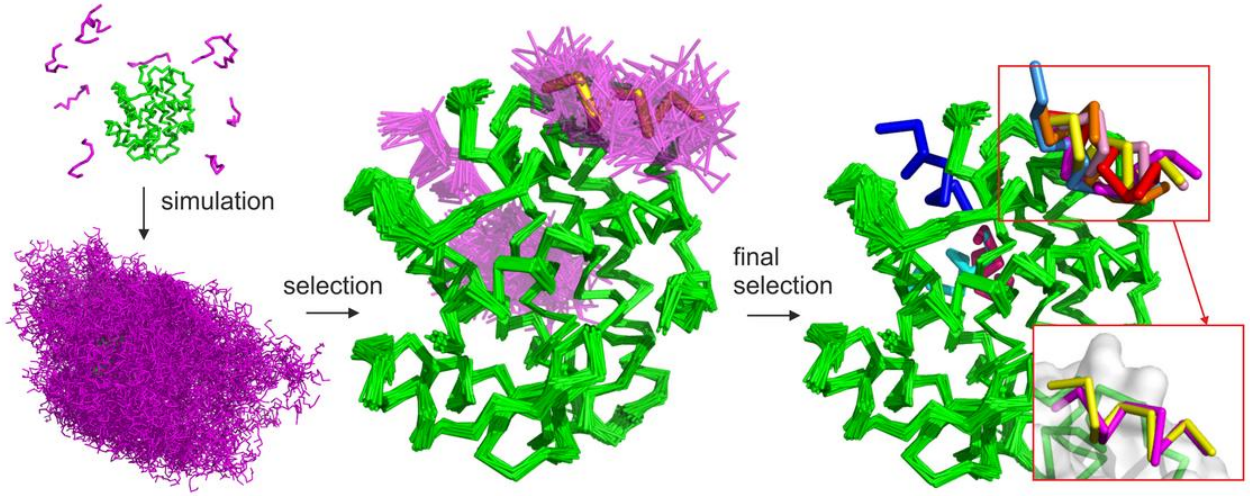
Exemplary softwares:

- CABS-dock
- HPEPDOCK
- HADDOCK2.2
- …

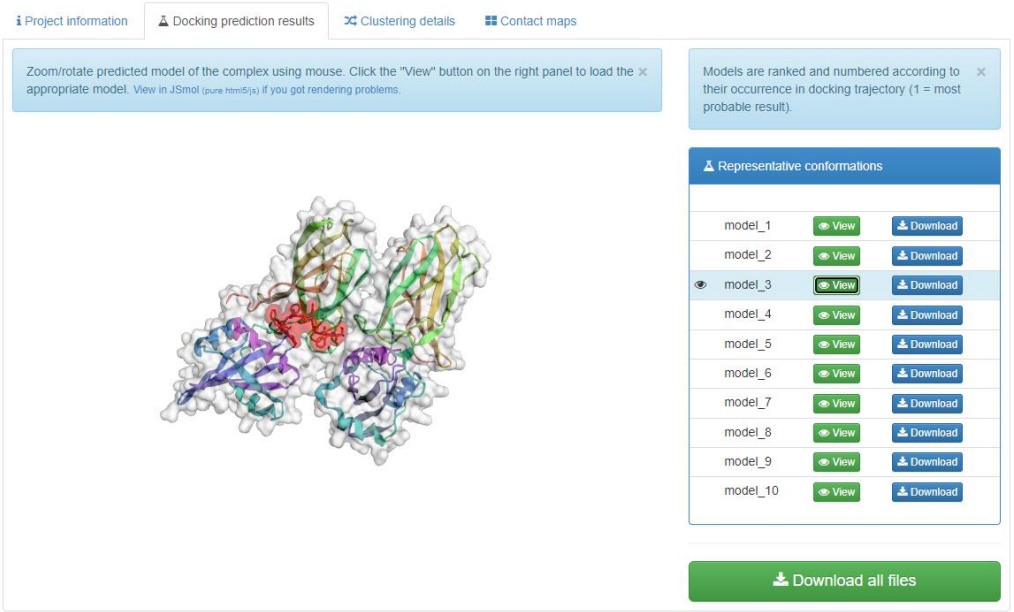
CABS-dock

- available on Linux / macOS / Windows
- open source
- under MIT License
- first publication: 2015

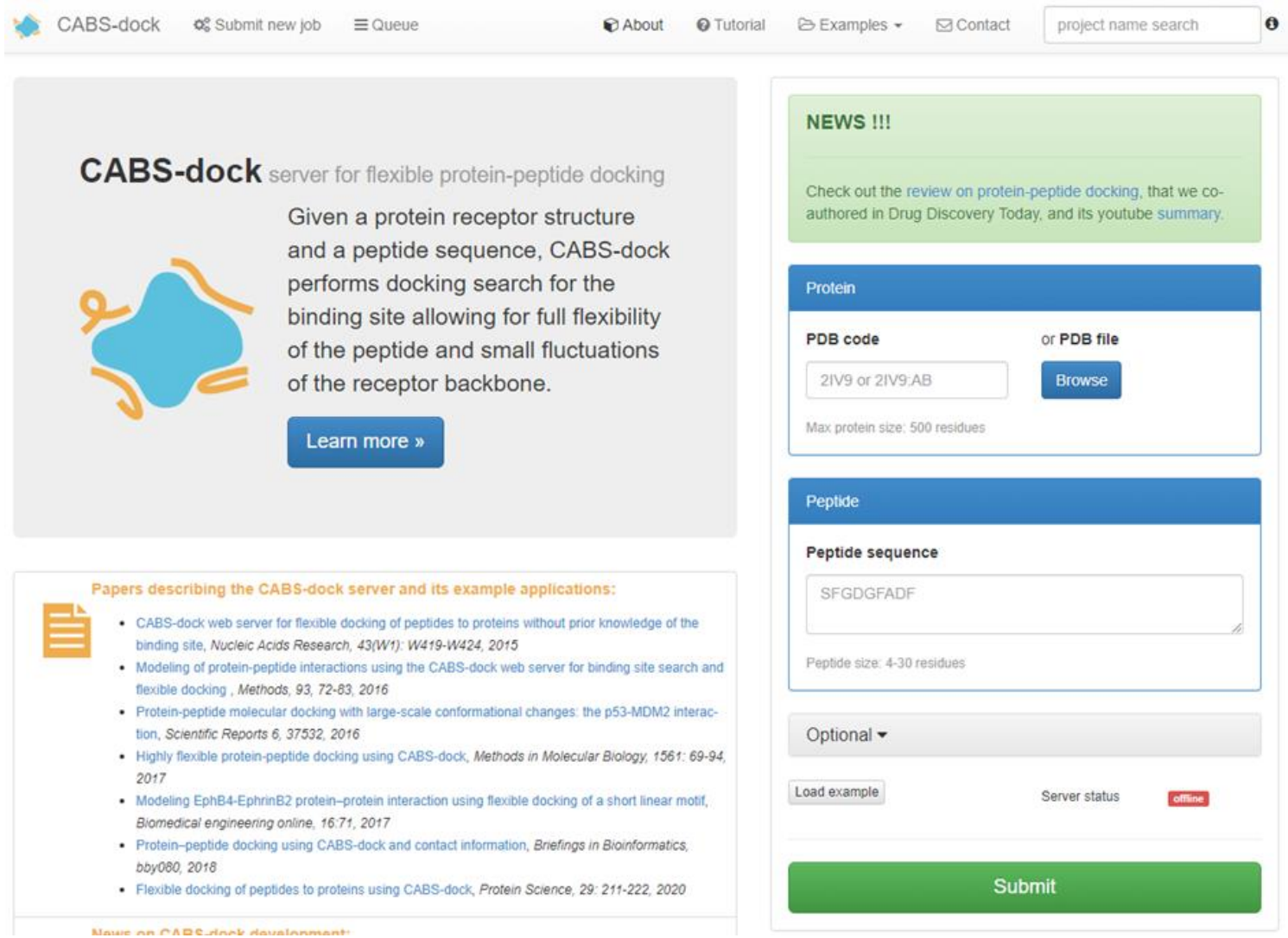
- available on web server:
- https://biocomp.chem.uw.edu.pl/CABSdock/



https://biocomp.chem.uw.edu.pl/CABSdock/learn_more

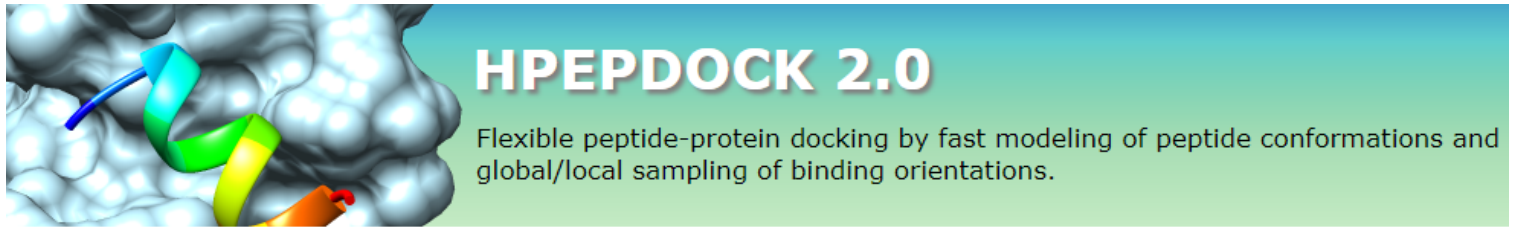


CABS-dock



- available on web server:
- https://biocomp.chem.uw.edu.pl/CABSdock/

HPEPDOCK



- first publication: 2007
- only web server available:
- http://huanglab.phys.hust.edu.cn/hpepdock/
- © Lab of Biophysics and Molecular Modeling, School of Physics Huazhong University of Science and Technology

HPEPDOCK

- Exemplary results:

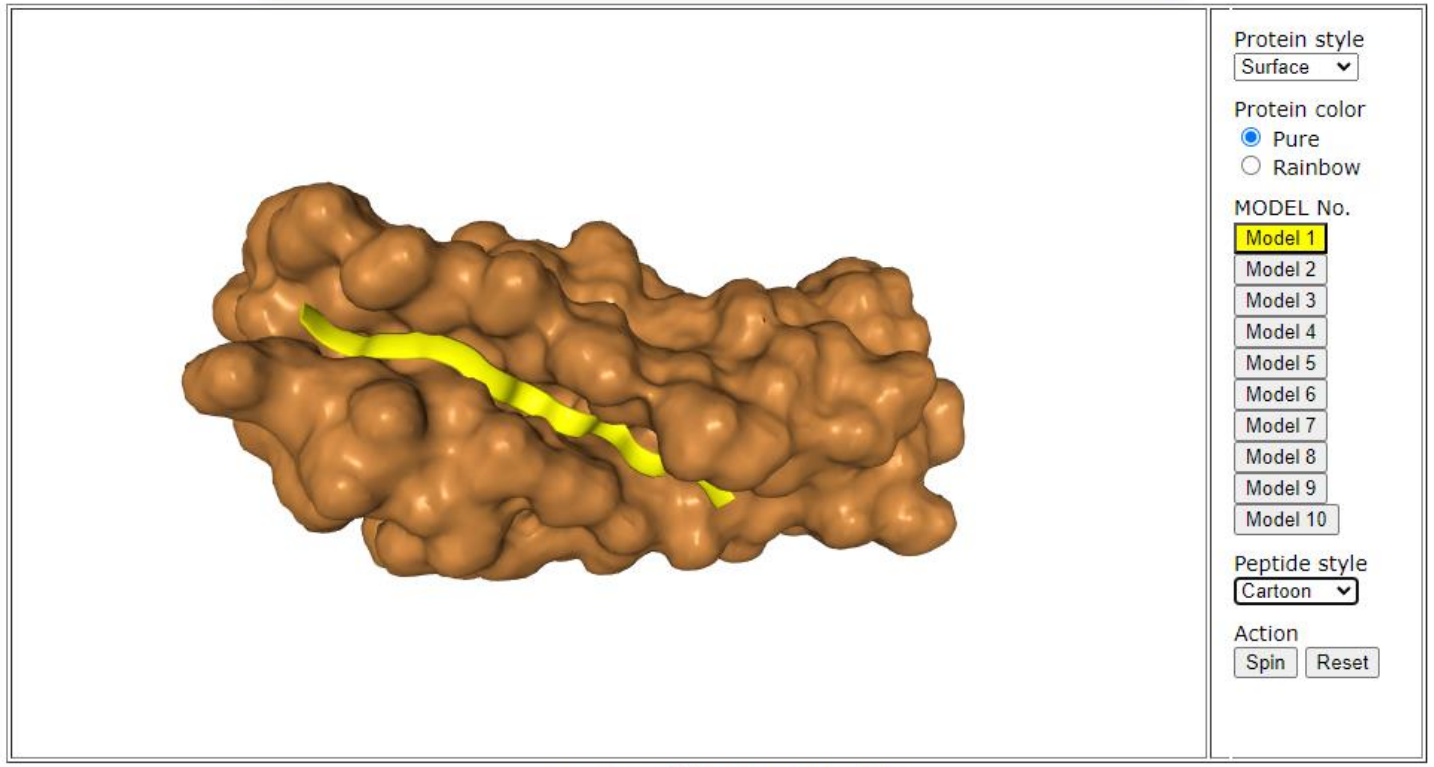
Your HPEPDOCK results for job *example*

Download Files

Receptor PDB file

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20]

Top 10 Predictions Top 100 Predictions All the results in a package



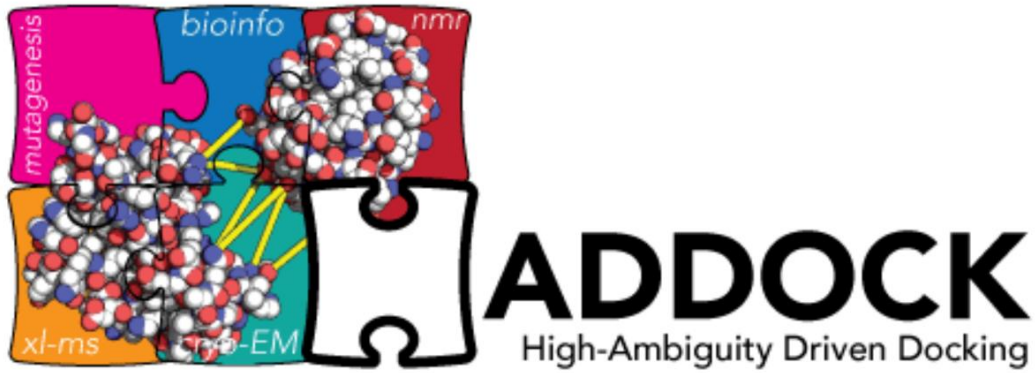
Summary of the top 10 models

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-436.035	-282.055	-281.788	-277.385	-275.499	-265.734	-257.327	-253.375	-252.432	-251.813

(a) Row 1: The ranks of the models.
(b) Row 2: The docking energy scores.

http://huanglab.phys.hust.edu.cn/hpepdock/data/example/

HADDOCK



- available on linux and MacOSX
- free for non-commercial users
- under Apache License, Version 2.0
- first publication: 2003

web server available:

- https://www.bonvinlab.org/software/haddock2.2/haddock-start/

- Works also with DNA/RNA – protein docking

HADDOCK

- can handle docking to 6 separate molecues at the same time
- input PDB/PSF file of peptide is nessecary, even if whole peptide is set as flexible
- to run, you need to prepare files:
  - run.param – base informations about docking
  - run.csn – specific informations about docking

run.csn file is generating by run.param file.

```
# type and location of restrains file (e.g. Ambiguous distance restraints)
AMBIG_TBL=./e2a-hpr_air.tbl

# location of HADDOCK
HADDOCK_DIR=../../

# number of molecues/proteins
N_COMP=2

# location of PDB files
PDB_FILE1=./e2aP_1F3G.pdb
PDB_FILE2=./hpr/hpr_1.pdb

# location of working directory
PROJECT_DIR=./

# protein names
PROT_SEGID_1=A
PROT_SEGID_2=B

RUN_NUMBER=1
```

Exemplary run.param file

Bibliography

- CABS-dock web server for flexible docking of peptides to proteins without prior knowledge of the binding site, Nucleic Acids Research, 43(W1): W419-W424, 2015
- Mateusz Kurcinski, Maciej Pawel Ciemny, Tymoteusz Oleniecki, Aleksander Kuriata, Aleksandra E Badaczewska-Dawid, Andrzej Kolinski, Sebastian Kmiecik, CABS-dock standalone: a toolbox for flexible protein–peptide docking, Bioinformatics, Volume 35, Issue 20, October 2019, Pages 4170–4172, https://doi.org/10.1093/bioinformatics/btz185
- Zhou, P., Jin, B., Li, H., & Huang, S. Y. (2018). HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic acids research*, *46*(W1), W443–W450. https://doi.org/10.1093/nar/gky357
- G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin (2016). "The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes." J. Mol. Biol., 428, 720-725 (2015).
- R. Vargas Honorato, P.I. Koukos, B. Jimenez-Garcia, A. Tsaregorodtsev, M. Verlato, A. Giachetti, A. Rosato and A.M.J.J. Bonvin (2021). "Structural biology in the clouds: The WeNMR-EOSC Ecosystem." Frontiers Mol. Biosci., 8, fmolb.2021.729513 (2021).
- Harris, R., Olson, A., J., Goodsell, D., S. Automated prediction of ligand-binding sites in proteins — 2007 Proteins: Structure, Function and Bioinformatics.
- Computational protein-ligand docking and virtual drug screening with the AutoDock suite.Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ, Nature Protocols, 2016 May, doi:10.1038/nprot.2016.051
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J. Computational Chemistry 2009, 16: 2785-91
- J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling.
- O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461
- Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. 2016 May;11(5):905-19. doi: 10.1038/nprot.2016.051. Epub 2016 Apr 14. PMID: 27077332; PMCID: PMC4868550.
- Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, et al. (2014) rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol 10(4): e1003571. doi:10.1371/journal.pcbi.1003571
- • Fu D., Meiler. J. RosettaLigandEnsemble: A Small-Molecule Ensemble-DrivenDocking Approach. ACS Omega, 3(4): 3655-3664 (2018).
- • DeLuca, S., Khar, K. & Meiler, J. Fully Flexible Docking of Medium Sized LigandLibraries with RosettaLigand. PLoS One 10, e0132508 (2015).
- Lemmon G, Meiler J. Rosetta Ligand docking with flexible XML protocols. Methods Mol Biol. 2012;819:143-55. doi: 10.1007/978-1-61779-465-0_10. PMID: 22183535; PMCID: PMC3749076.
- Glide, version 6.7, Schrödinger, LLC, New York, NY, 2015.

QUESTIONS

THANK YOU FOR ATTENTION