MolDesk Screening ver.1.1

Software package for in silico drug design **www.moldesk.com**

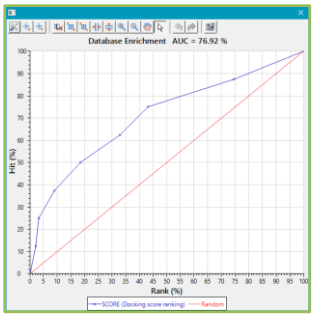
Virtual Screening for 2 million compounds within 0.5 to 2 days on a common 8-core CPU PC
High-speed MD simulation on MPI/GPU (NVIDIA CUDA)

Virtual screening (with parallel processing)

Users can use *myPresto*※1, which includes Multiple Target Screening (MTS), and machine-learning Multiple Target Screening (MTS) as structure-based drug screening, and Docking Score Index (DSI) as ligand-based drug screening.

In these screenings subsets of *Ligandbox*※2 database with 2 million compounds extracted for drug design are scanned at high speed. Users can also perform screening on in-house compound database with several hundred to several million of compounds in 2D SDF file format.

MolDesk sorts the compound list obtained by the drug screenings and provides the docking poses of selected compounds. MolDesk also displays the information of each compound, namely, compound ID, source, source ID, rank, score, chemical formula, 2D chemical structure, molecular mass, number of chiral centers, number of hydrogen bond donors / acceptors, total charge, HOMO, LUMO, and chemical properties such as LogS and LogP.



MolDesk outputs the compound list in CSV / HTML formats that can be read by Excel and used for purchasing the compounds (vendor ID is eliminated and only Namiki/Kishida ID is shown.)

Users can validate the drug screening results by including known active compounds in the screening. Accuracy is confirmed by displaying the database enrichment curve and the area under the curve (AUC) using known active compounds.

MolDesk generates 3D structures of the compounds from the in-house compound library (2D SDF) before performing the drug screening. It generates conformers with chirality and filters out compounds with excess mass weights or unsuitable reactive chemical structures.

Figure 1. Database enrichment curve.

MVO Screening (with parallel processing)

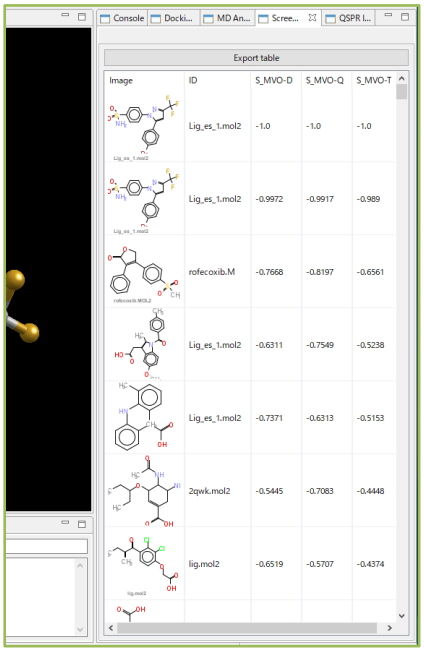


Figure 2. MVO-screening result.

MVO (Maximum Volume Overlap) screening is one of the ligand-based drug screening methods screening compounds with structures similar to the query compound. It estimates volume overlap between the 3D structure of the query compound and that of the selected compound from the database and provides the 3D structure of the overlap.

Three different indexes, S_{MVO-Q}, S_{MVO-D}, and S_{MVO-T}, indicate the overlap between the query compound and the test compound selected from the database. By performing MVO score calculation based on these indexes, you can obtain the ratio of the overlap volume to the query compound's volume, the ratio of the overlap volume to the test compound's volume, and a ratio corresponding to Tanimoto index.
The definitions of each score and the list are shown below:

Table 1. MVO-screening score.

Score Type	Definition	Notes
S_{MVO-D}	**Q∩D/D∩D**	For smaller-size molecules selection.
S_{MVO-Q}	**Q∩D/Q∩Q**	For larger-size molecules selection.
S_{MVO-T}	**-1.0 * Q∩D/Q∪D**	For near-query-molecule-size molecules selection.

Q = Query compound's volume
D = Database compound's volume

※1 *myPresto* is the software of Japan Biological Informatics Consortium developed by the National Institute of Advanced Industrial Science and Technology, JBIC, Osaka University Institute of Protein Research, etc. supported by AMED, Ministry of Economy, Trade and Industry, and NEDO.

※2 *LigandBox* is the database of Japan Biological Informatics Consortium developed as part of myPresto by the National Institute of Advanced Industrial Science and Technology, JBIC, Osaka University Institute of Protein Research, etc. supported by AMED, Ministry of Economy, Trade and Industry, and NEDO.

Docking-score QSAR (with parallel processing)

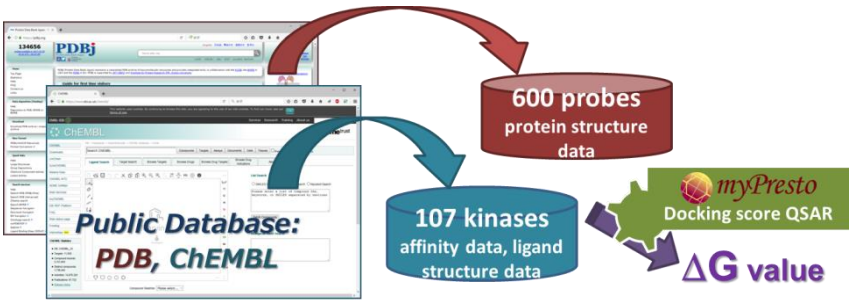


Figure 3. Docking-score QSAR Method.

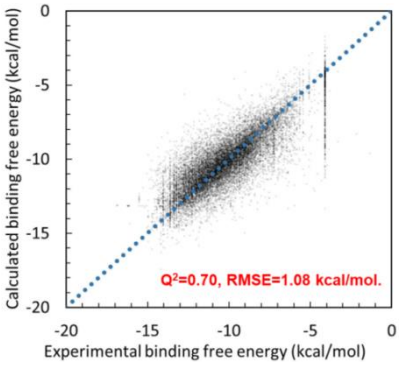


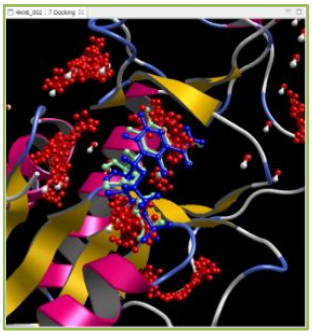
Figure 4. Prediction results of 107kinases.

MolDesk can apply docking-score QSAR in predicting protein-compound binding free energies (ΔG). The docking-score QSAR is a regression method whose descriptors are docking scores of the compound against 600 probe proteins.

MolDesk uses ΔG assay data registered in ChEMBL and protein structures from PDB public databases, and converts the assay data (IC50, %-inhibition, Ki and activity data) to the ΔG values in kcal/mol. The regression model used here is a principal component regression based on M robust estimation and L2 regularization term to reduce over-teaching effect.

As a result, this method shows Q^2=0.70 (RMSE=1.08 kcal/mol) accuracy in predicting ΔG values of 107 kinases on average (Figure 4.)

Pocket search (with parallel processing)

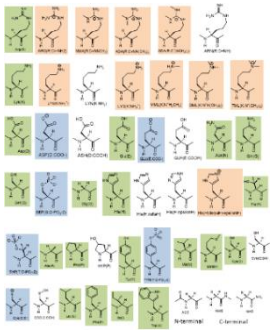


MolDesk can search ligand-binding pockets of a given protein structure with Molecular-docking binding-Site finding (MolSite). This method uses a probe compound library consisting of sets of arbitrarily selected compounds. MolSite performs round-robin docking simulation of the compounds from the compound library against the protein's whole surface, and suggests the protein site that shows the strongest affinity as ligand-binding sites (ligand-binding pocket.)

Figure 5 shows the structure of reconstructed model in a self-docking example (PDBID: 4kn6) with an accuracy of RMSD=1.08A (Blue: experimental data, Green: predicted ligand structure, Red dots: suggested ligand-binding pockets.) Pocket search does not always provide predictions with such high accuracy, but produces much better results compared to existing geometrical approaches. Pocket search runs with thread-parallel processing similarly to virtual screening.

Figure 5. A docking pose in a ligand-binding pocket found by MolSite (PDBID 4kn6.)

Chemical modification of amino-acid residue



Users can replace any kind of amino-acid residues of a protein with 40 kinds of chemically modified residues registered in the program.

For example, when you convert the ARN (arginine in AMBER force field), MolDesk produces a table of candidates for conversion as shown on the right. Users can easily process replacement by selecting respective amino acid residues.

Figure 6. 40 kinds of chemically modified residues.

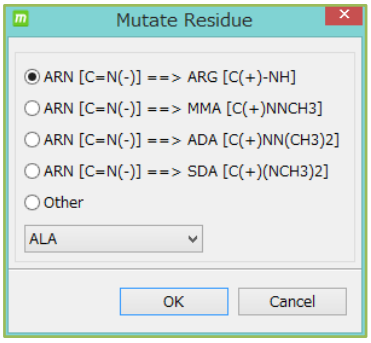
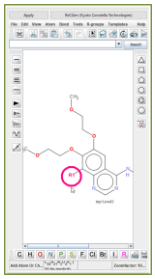


Figure 7. ARN converter window.

JChemPaint (2D compound editing)



As shown in Figure 8, a user can prepare a molecular structure using JChemPaint, and MolDesk converts its 2D molecular structure to 3D molecular structure to be used in user's own drug design calculation.

Figure 8. Editing 2D compounds.

Protein-compound docking simulation

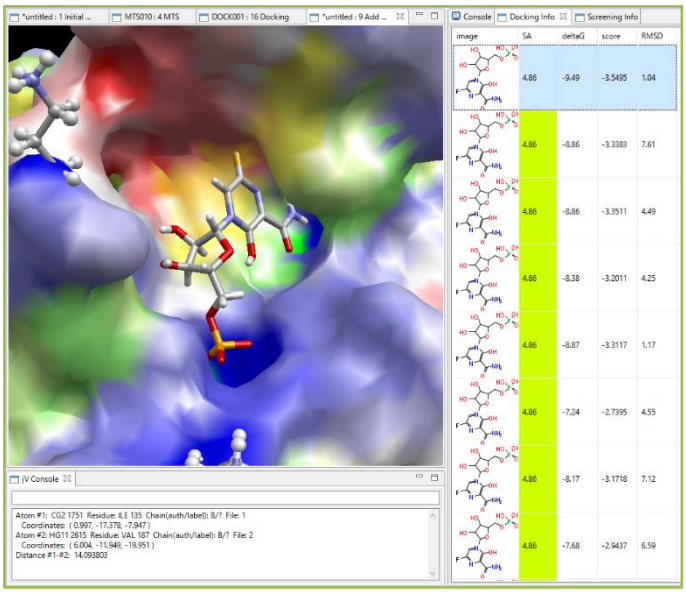


Figure 9. Docking simulation result.

MolDesk shows results of docking simulation in order of docking scores and allows users to visually confirm the docking poses.

MolDesk provides a list with docking scores as well as corresponding ΔG values estimated with a simple regression model and the RMSD value against the initial coordinates of each compound's molecules.

By clicking the score or the arrow keys, the user can view each docking pose. The user can also superimpose multiple docking poses.

Boost docking by Solution NMR data

Solution NMR data (DIRECTION epitope-mapping method) of non-label solution NMR data can improve your protein-compound docking poses.

Y. Mizukoshi, et al., An accurate pharmacophore mapping method by NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **2012**, *51*(6), 1362-1365.

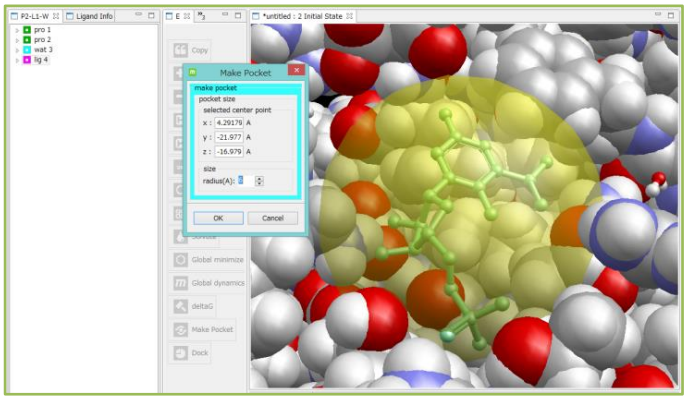


Figure 10. Ligand-binding pocket addressing.

Users can indicate the center of the binding site by either directly entering the coordinates or selecting multiple atoms of the target protein.

MolDesk shows the synthetic accessibility (SA) of all compounds in the list based on their 2D chemical structures. SA value is updated in real time in molecular editing process.

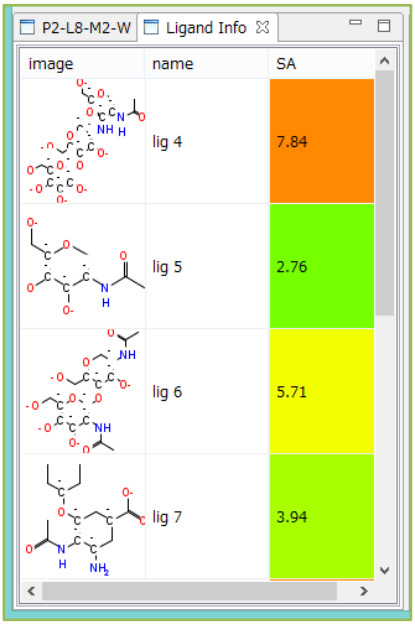
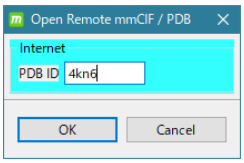


Figure 11. Synthetic accessibility.

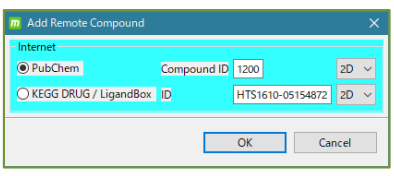
Other functions



Users can remotely download mmCIF and PDB files by entering PDB IDs (4-character codes) in MolDesk and immediately start calculation / simulation.

Figure 12. Protein downloader window.

MolDesk can display precise electrostatic potential surface of the protein provided by eF-site/eF-surf/eF-seek and the surface of the protein's cavity.



Users can remotely download molecular structure files by entering the compound's LigandBox/PubChem ID in MolDesk.

Figure 13. Compound downloader window.

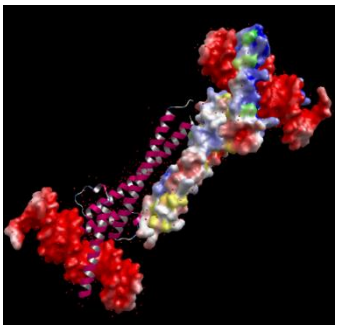


Figure14. Electrostatic potential surface.

MolDesk Basic / MolDesk Screening Function Summary

Function		MolDesk Basic	MolDesk Screening
Input file format	mmCIF・PDB・MOL2・MOL/SDF・SMILES format MOL2・SMILES can be multi-file (including multiple structures)	○	○
Protein editing	H atom addition, elimination of unnecessary parts, end treatment, etc., in preparation for MD simulation	○	○
	40 types of amino acid residues that can be used for epigenetic modification	○	○
Compound editing	Compound editing/Atomic partial charges by AM1-BCC method	○	○
	2D editor (JChemPaint), list for 2D structures and conversion feasibility	○	○
	3D structure generation from SMILES and 2D structure (MOL/SDF format) (output in MOL2 format)	○ Sequential processing	○ Parallel processing
Energy optimization, MD simulation	Energy optimization of protein/compound/complex systems, MD simulation, preparation of systems in water and neutralization by addition of ions, trajectory video, chronological graphs for various energies/temperatures GROMACS can be used as MD calculation program.	○	○
	High speed parallel processing on MPI/GPU (NVIDIA CUDA)		○ Mac disabled
Protein-compound docking simulation	Suggestion for ligand-binding sites, docking simulation, pose improvement by solution NMR experiment (DIRECTION)	○	○
	Manual docking of binding-free energy (energy optimization/prediction of protein-compound binding free energies (ΔG) after users adjust ligand position)	○	○
Pocket search	Simplified version (search through geometrical approach)	○	○
	Advanced version (search by MolSite/parallel processing)		○
Virtual screening	SBDS (target protein required): MTS, machine-learning MTS LBDS (no target protein required): machine-learning DSI (all by parallel processing)		○
	List generation in order of drug screening (2D structures and physical properties), visual confirmation of docking poses, CSV/HTML file output including 2D structure images		○
	Accuracy analysis by database enrichment curve		○
	Screening on imported in-house compound library (2D structures, MOL/SDF format) (3D structure generation/parallel processing)		○
Physical property prediction	Synthetic accessibility	○	○
	Activity prediction by Docking-score QSAR (ChEMBL utilization/parallel processing)		○
Similarity search / substructure search	Similarity search by MVO-screening (parallel processing)		○
	Similarity search/substructure search by Topology Graph Similarity		○

License

[Package Type] Function	Price (Tax excluded)
[MolDesk Screening] All functions (as indicated by the function summary.)	**[Please contact us.] / 1year 1node**
	1 year software subscription.

MolDesk products are supported for Windows 11 / 10 / 8.1 / 8 / 7 / Vista (64bit), Linux (64bit), macOS 10.11 or later.
16/32GB memory is required for 8/48 process parallel calculation.

IMSBIO Co., Ltd.

Contact information:
Sales Division
Owl Tower, 4-21-1, Higashi-Ikebukuro, Toshima-ku, Tokyo 170-0013, Japan
TEL 03-6907-0315 FAX 03-6907-0316
EMAIL: info@imsbio.co.jp URL: http://www.imsbio.co.jp

2022.9.23