



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
(All functions of MolDesk Basic are available.)

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.

Score		Export table							AUC		
Image	rank	ID	Score	Source ID	Ligand Box ID	Formula	Weight	log S	log P	Supplier	ID Number
	1	HTS1508-03579436-02	4.8174	NS-07213367	HTS1508-03579436-02	C21H25N4O	349.458	1.388	4.0763	ENAMINE	Z219349550
	2	HTS1508-01970909-02	4.4553	NS-02926541	HTS1508-01970909-02	C20H22N2O3	336.391	1.4704	2.9269	ENAMINE	Z108660526
	3	HTS1508-02313392-01	4.7091	NS-03806336	HTS1508-02313392-01	C17H20N7	328.444	2.7014	3.2778	ENAMINE	Z195555268
	4	HTS1508-01297838-01	4.3179	NS-01905160	HTS1508-01297838-01	C22H22N3O	344.438	1.2771	3.4332	ENAMINE	Z30501126
	5	HTS1508-03818160-01	4.088	NS-08577078	HTS1508-03818160-01	C17H17N2O2F3	338.329	2.3427	3.215	ENAMINE	Z393480320
	6	HTS1508-04518302-02	4.7931	NS-10566855	HTS1508-04518302-02	C19H28N3O3	344.435	2.3974	2.4917	ENAMINE	Z1538838180
	7	HTS1508-04220538-02	4.7039	NS-10126483	HTS1508-04220538-02	C18H24N3O2	342.423	1.42	2.2163	ENAMINE	Z354468456

Figure 1. MTS Screening Results List

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.)

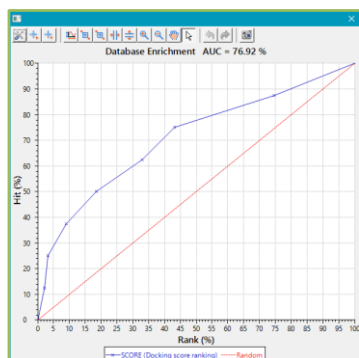


Figure 3. Database enrichment curve.

MolDesk sorts the compound list obtained by the drug screenings and provides the docking poses of selected compounds.

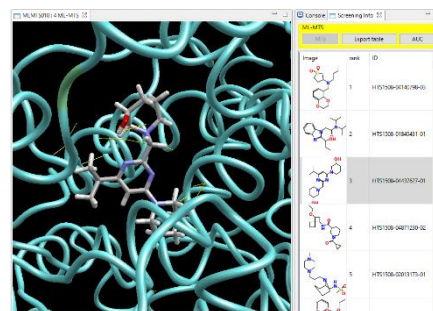


Figure 2. Docking Pose Display

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.

※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.

MVO Screening (with parallel processing)

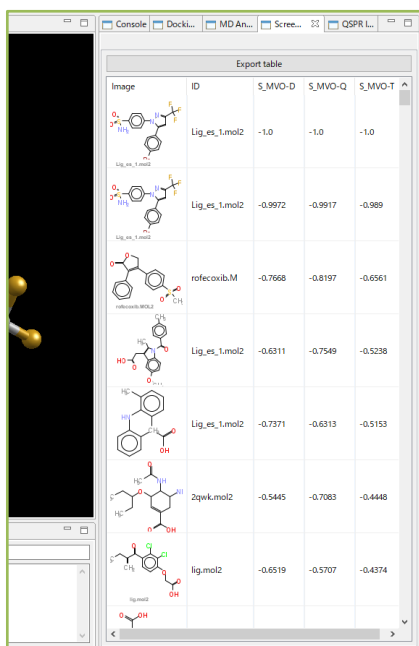


Figure 4. MVO-screening result.

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:

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

Table 1. MVO-screening score.

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

Docking-score QSAR (with parallel processing)

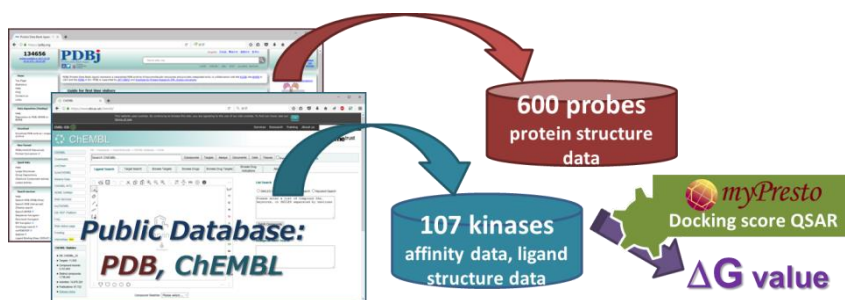


Figure 5. Docking-score QSAR Method.

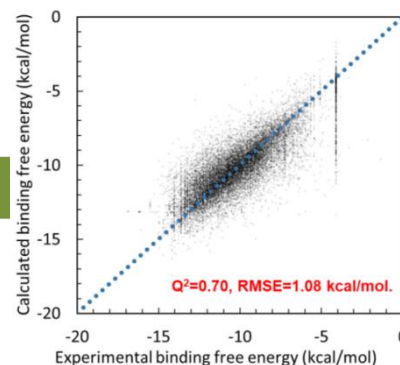


Figure 6. Prediction results of 107kinases.

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, K_i 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 6.)

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)	<input type="radio"/>	<input type="radio"/>
Protein editing	H atom addition, elimination of unnecessary parts, end treatment, etc., in preparation for MD simulation	<input type="radio"/>	<input type="radio"/>
	40 types of amino acid residues that can be used for epigenetic modification	<input type="radio"/>	<input type="radio"/>
Compound editing	Compound editing/Atomic partial charges by AM1-BCC / RESP method	<input type="radio"/>	<input type="radio"/>
	2D editor (JChemPaint), list for 2D structures and conversion feasibility	<input type="radio"/>	<input type="radio"/>
	3D structure generation from SMILES and 2D structure (MOL/SDF format) (output in MOL2 format, parallel processing)	<input type="radio"/>	<input type="radio"/>
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.	<input type="radio"/>	<input type="radio"/>
	High speed parallel processing on MPI/GPU (NVIDIA CUDA)		<input type="radio"/> Mac disabled
Protein-compound docking simulation	Suggestion for ligand-binding sites, docking simulation, pose improvement by solution NMR experiment (DIRECTION)	<input type="radio"/>	<input type="radio"/>
	Manual docking of binding-free energy (energy optimization/prediction of protein-compound binding free energies (ΔG) after users adjust ligand position)	<input type="radio"/>	<input type="radio"/>
Pocket search	Simplified version (search through geometrical approach)	<input type="radio"/>	<input type="radio"/>
	Advanced version (search by MolSite/parallel processing)	<input type="radio"/>	<input type="radio"/>
Virtual screening	SBDS (target protein required): MTS, machine-learning MTS LBDS (no target protein required): machine-learning DSI (all by parallel processing)		<input type="radio"/>
	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		<input type="radio"/>
	Accuracy analysis by database enrichment curve		<input type="radio"/>
	Screening on imported in-house compound library (2D structures, MOL/SDF format) (3D structure generation/parallel processing)		<input type="radio"/>
Physical property prediction	Synthetic accessibility	<input type="radio"/>	<input type="radio"/>
	Activity prediction by Docking-score QSAR (ChEMBL utilization/parallel processing)		<input type="radio"/>
	LogP/LogS/LogD/Papp(Membrane permeability coefficient)/pKa Prediction by Regression Analysis		<input type="radio"/>
Similarity search / substructure search	Similarity search by MVO-screening (parallel processing)		<input type="radio"/>
	Similarity search/substructure search by Topology Graph Similarity		<input type="radio"/>

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.

Operating OS : Windows 11/10(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