systemsDock Operation Manual

Version 2.0
2016 April

systemsDock is being developed by
Okinawa Institute of Science and Technology
http://www.oist.jp/

Integrated Open Systems Unit
http://openbiology.unit.oist.jp/_new/

Acknowledgements

Developers would like to thank the collaborators in The Systems Biology Institute, Tokyo, Japan.

Many thanks to the users who kindly provided us bug reports and feature requests.

Contents

1. Introduction

This manual explains the usage of systemsDock (http://systemsdock.unit.oist.jp).

Top

1.1 What is systemsDock?

systemsDock is a web server for network pharmacology-based prediction and analysis, which applies high-precision docking simulation and molecular pathway map to comprehensively characterize the ligand selectivity and to illustrate how a ligand acts on a complex molecular network. We apply machine learning system to predictively access the ligand-protein binding potentials (for details). For large-scale screening and investigation with ease, a user-friendly GUI interface for systemsDock has been established providing efficient methods for molecule preparation, parameter specification and result inspection. Ligand's binding potentials against individual proteins can be displayed on an uploaded molecular interaction map directly, allowing users to systemically investigate network-dependent effects of a drug or drug candidate.

Top

1.2 Major Features

- It provides a major advance in quality and reliability of assessing protein-ligand interaction, and has already resulted in significantly improved docking simulations (for details).
- Integration with biochemical network to look at drug interactions with potential protein targets involved in a molecular pathway.
- Displaying predicted binding potential on a network map for a comprehensive inspection. A molecular pathway map in SBML format is needed.
- A user friendly GUI for large-scale target proteins screening with ease.
- Finally, it is of course freely accessible to academic research.

Top

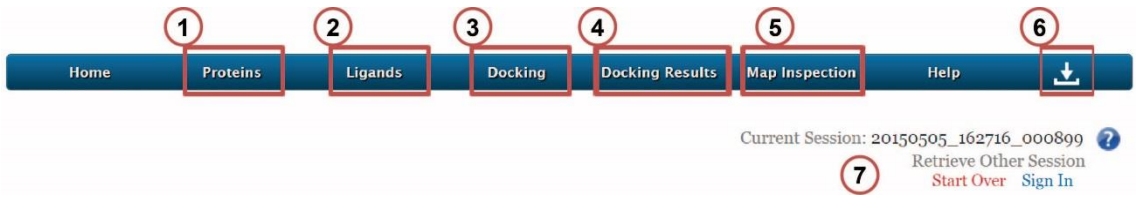
1.3 Target readers

The target readers are researchers who would like to test ligand selectivity and to carry out a pharmacology-based prediction and analysis.

Top

2. Interface operation

systemsDock is designed to perform a screening for a large number of proteins with ease. It is distinguished from other web resources by having a series of efficient **stepwise** methods for molecule preparation, parameter specification and result inspection.



1. Protein and binding site specification.
2. Ligand preparation.
3. Docking simulation.
4. Docking results.
5. Map inspection.
6. File download.
7. Current session ID and Session management.

Top

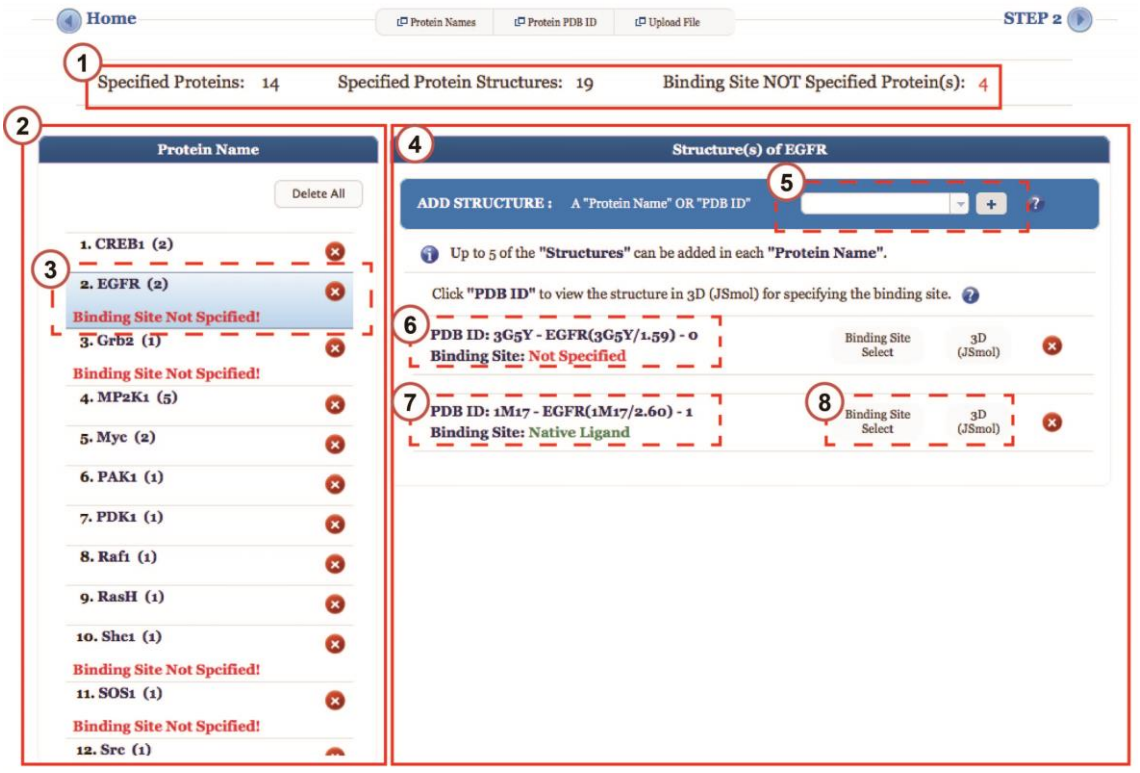
2.1 Protein specification

For network-based screening, systemsDock provides highly efficient flexible options allowing users to test a large number of proteins by specification of 1) protein names or gene symbols, 2) protein PDB IDs or 3) upload a pathway map file in SBML format. The referred protein structures with best determined resolution will be retrieved from an in-house protein identity-to-structure mapping system for docking simulation. Considering the frequent structure modification and data instantaneity, structure files will be dynamically downloaded from RCSB PDB database.

2.1.1 Methods to specify protein identify

1. Input protein names or gene symbols.
2. Directly input PDB IDs.
3. Upload a molecular pathway map file in format of SBML.

2.1.2 Protein structure specification

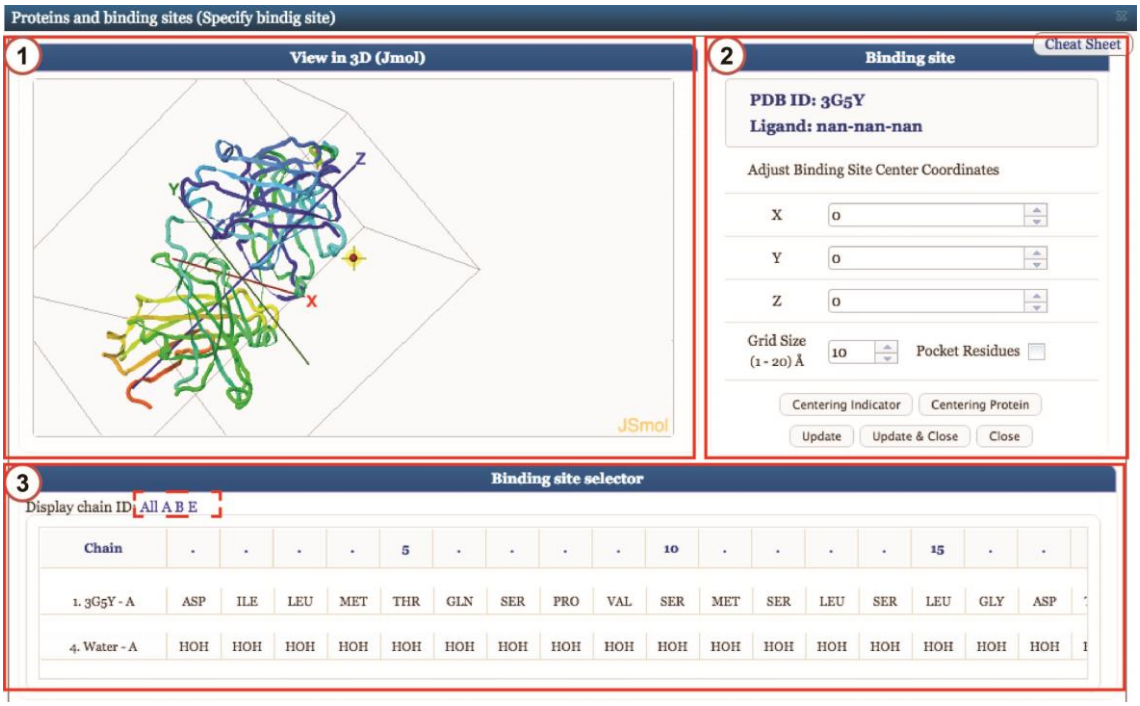


1. Specified proteins are statistically summarized.
2. List of the specified proteins.
3. Indication of binding site status.
4. By clicking on one of the specified protein listed in (3), the structure models of the proteins for docking simulation.

5. Users can add more structures by inputting protein names or PDB IDs.
6. Indicate that the binding site of this structure has not been defined yet. User can click on this entry to define the binding site using an interactive molecular visualizer. See session of **Binding site specification**.
7. Indicate that the binding site of this structure has been defined.
8. Click to proceed the binding specification either by a protein residue table or by an interactive molecular visualizer.

2.1.3 Binding site specification

A binding site for each of proteins will be automatically identified by exploring the position where the biggest native ligand is bound. Alternatively, binding site can be defined through an interactive molecular visualizer by directly clicking on the displayed structure model or residue listed in the sequence table. After clicking on (5) or (6) shown on previous page, the visualizer will pop up:



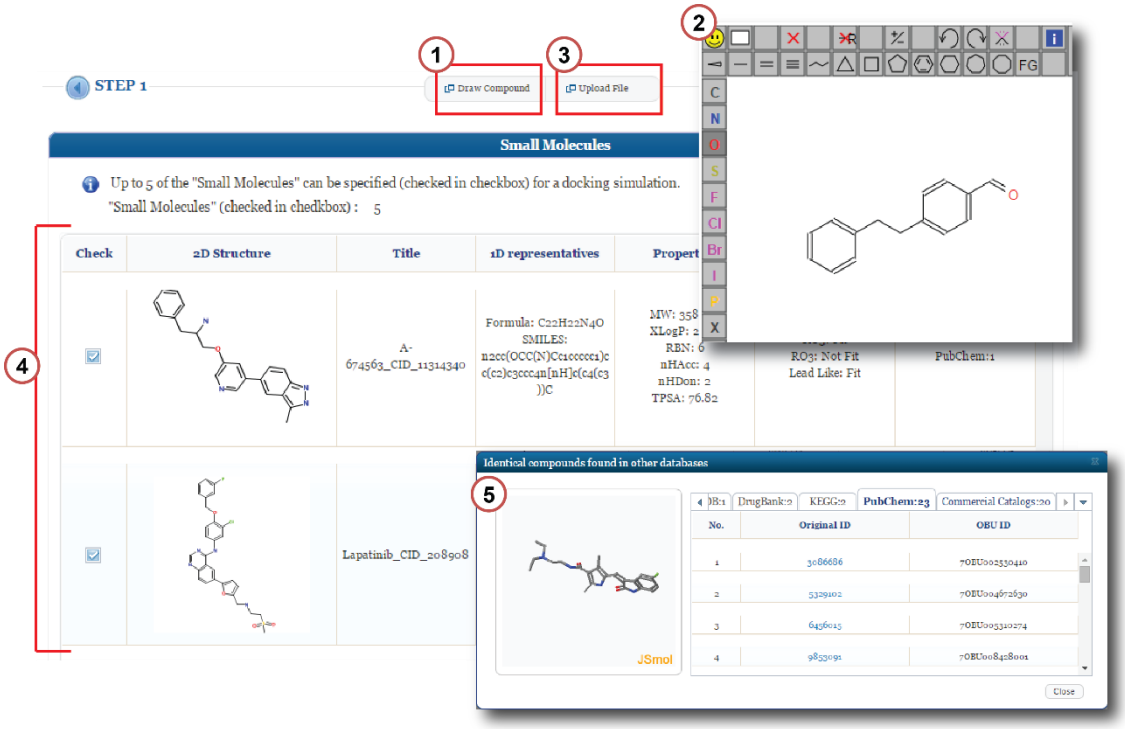
1. Structure model of the specified protein. Directly clicking on the model to define the location of the preferred binding site.
2. Adjust the x-y-z coordinate and the grid size to refine the location of the binding site.
3. List of the protein sequences. Directly clicking on one of the residues or ligands to

define the location of the preferred binding site. By clicking on “Display chain ID” (dashed line square) to list all or specified chains.

Top

2.2 Ligand preparation

systemsDock accepts compounds in commonly used formats, including 2D/3D SDF, Mol2 or SMILES. 2D-3D conversion is done by Ballon. Users can also use a provided web-based molecule editor to compose compounds of interest. Compounds will be displayed in various representatives with a set of calculated molecular properties. There are also convenient links for compounds to external databases including PubChem, DrugBank, BindingDB, TCM Databasae@Taiwan and KEGG as well as over 130 chemical vendors allowing easy access to biological data and identifying commercially available molecules.

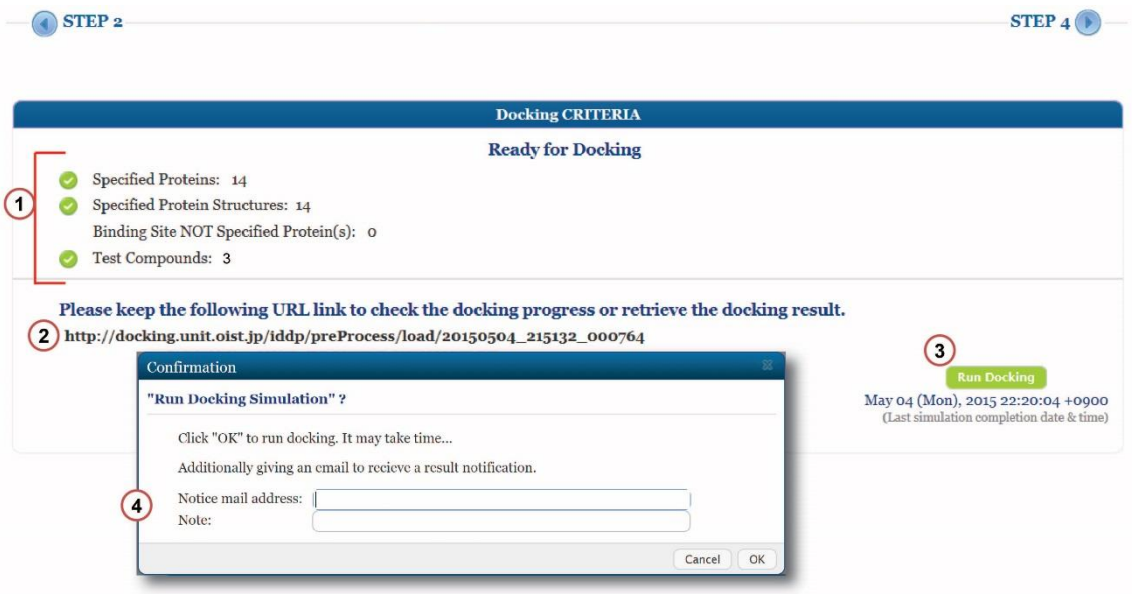


1. Click to pop up a web-based molecule editor.
2. Compose a compound for docking simulation using JSME.
3. Upload a 2D/3D SDF, Mol2 or SMILES file (max. 5 compounds for a screening).
4. Compounds will be displayed in various representatives with a set of calculated molecular properties.
5. Click on an entry of the compound list (4), a window will pop up to show the links for the compound to external databases as well as to virtualize the compound in 3D.

Top

2.3 Docking simulation

systemsDock applies docking simulation (AutoDock VINA) to predict protein-ligand binding potential. To improve the prediction performance, a unique scoring function for docking simulation called docK-IN (combining docking with intelligence) is embedded in systemsDock. docK-IN was initially developed in our previous work to address the critical issue of commonly used docking programs which were seen too inaccurate for a reliable prediction (for details). docK-IN utilizes machine learning algorithm (Random Forest) together with a series of characterized binding interactions and test compound's molecular properties to rescore and rank all of the binding modes generated by the docking tool. The top-score binding mode is selected for each test compound. Upon the molecule specification, the simulation may take long time. Users can retrieve the progress or result simply using a given web link or session ID. An email notification will be sent to user when simulation is done.



1. To confirm the specification for docking simulation.
2. A link is provided for checking the docking progress or retrieving the results later on.
3. While the specification has been done without a problem, the "Run Docking" button will be activated, and clicking on to run docking.
4. Giving an email for notification when simulation is done.

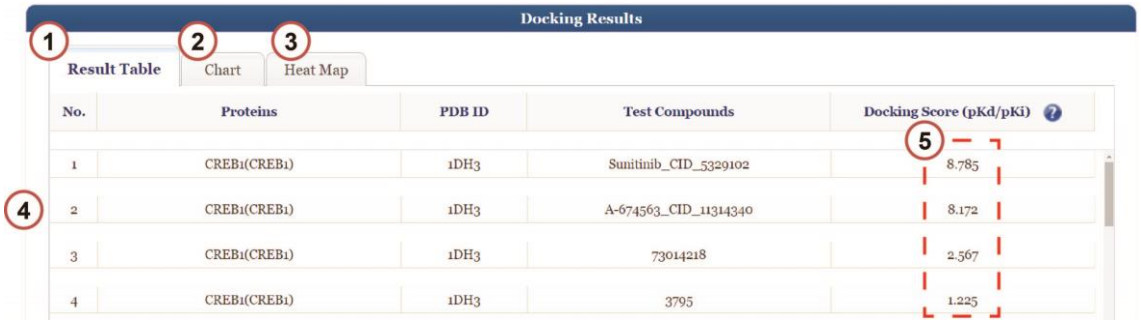
Top

2.4 Docking results

Docking scores, that is the predicted binding affinity, to each of target proteins are displayed in an interactive table and histogram. Unlike other docking methods, the score reported by our docking approach is a negative logarithm of experimental dissociation/inhibition constant value (pKd/pKi) usually ranging from 0 to 10 (i.e. from weak to strong binding), allowing a straightforward indication of binding strength. By clicking on one of the table entries or histogram bars, elaborated molecular binding interactions can be graphically shown in 2D/3D for structure-based investigation. Analyzed by LIGPLOT, amino acid residues involved in the intermolecular interactions will be highlighted for rapid inspection.

Top

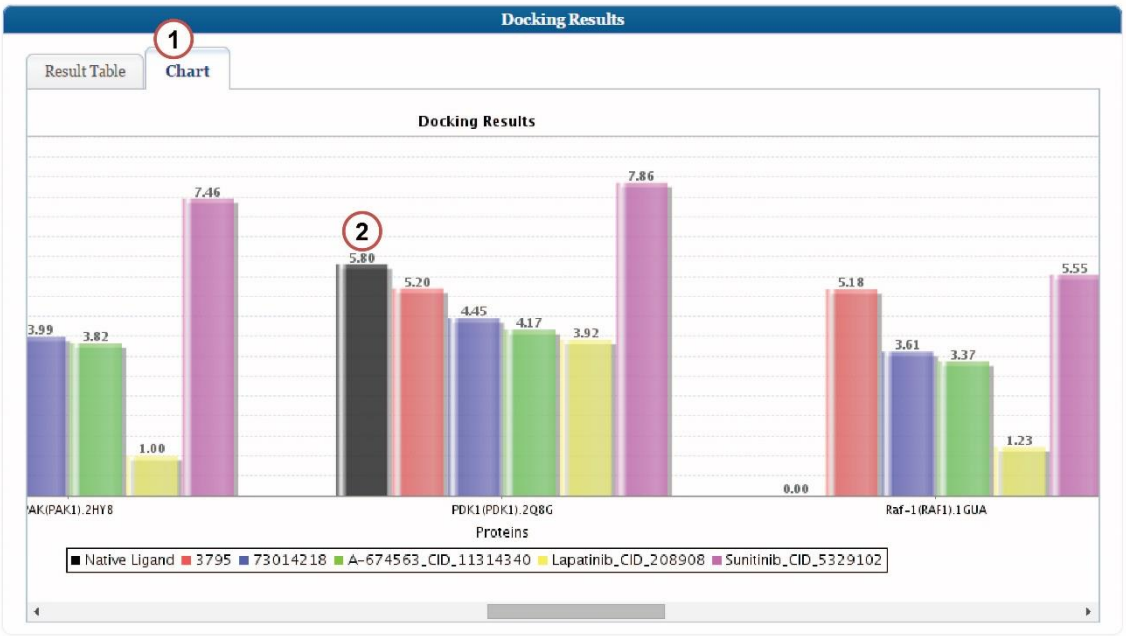
2.4.1 Result table



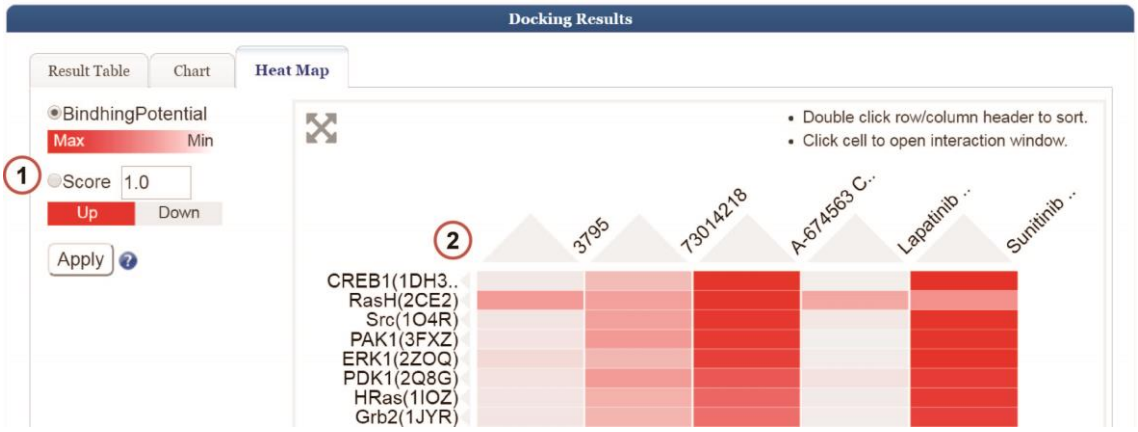
1. Docking results are tabulated.
2. To visualize results in a histogram.
3. To visualize result in a heap map.
4. Docking scores of each pair of protein-ligand binding are listed.
5. By clicking on one of the score entries, elaborated molecular binding interactions can be graphically shown in 2D/3D for structure-based investigation.

Top

2.4.2 Result chart (histogram and heat map)



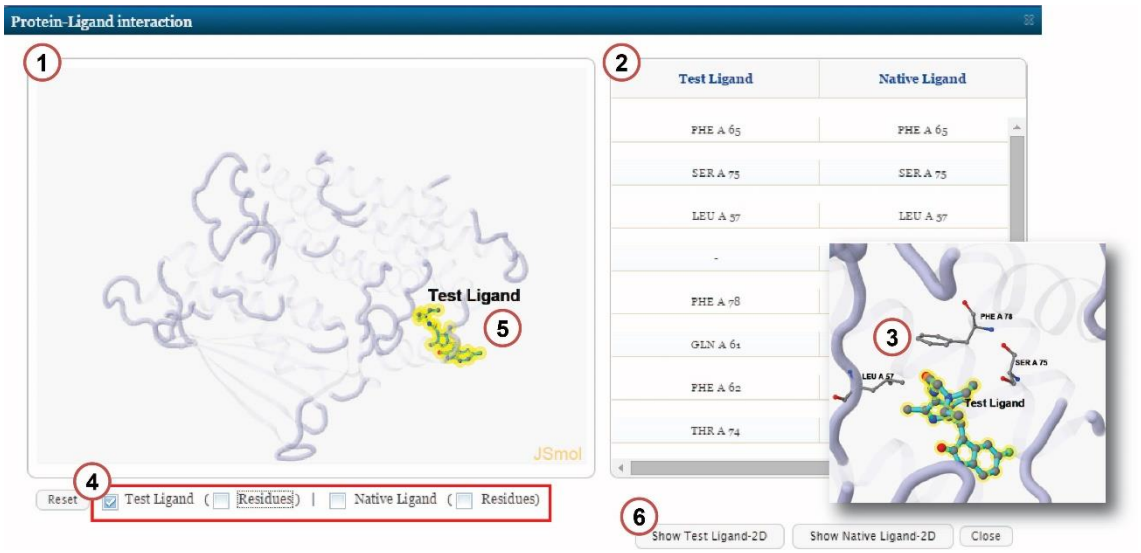
1. Visualize results in a histogram (by JFreeChart). Docking scores of each compound are grouped by proteins.
2. By clicking on one of the bars (other than black one), elaborated molecular binding interactions can be graphically shown in 2D/3D for structure-based investigation.



1. Heat map (by NDV3) with options of 1) to convert the predicted binding affinities into white-red color scales, 2) to set a cutoff as a bioactivity classifier.
2. An interactive heat map with functions sorting, grouping and zooming. Clicking on the cells to elaborate molecular interactions in 2D/3D.

Top

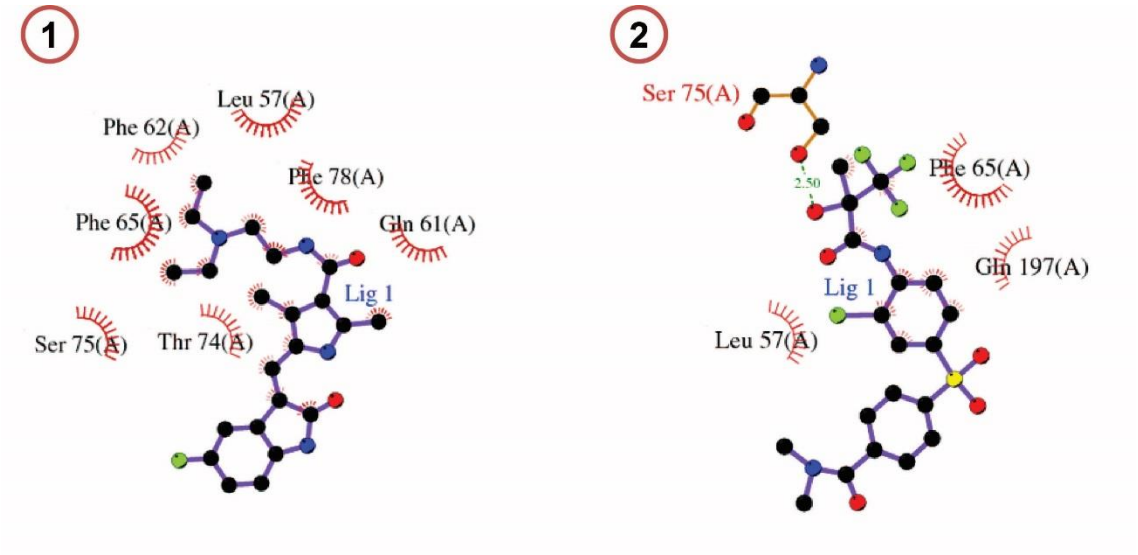
2.4.3 Visualize protein-ligand interaction in an interactive 3D visualizer



1. Visualize protein-ligand interaction using JSmol.
2. Protein residues involved in the binding interaction are listed. As a reference, those interacted with the native ligand, if available, are also listed.
3. Clicking on any of the entries listed in (2) allows users to centre and display the specified residue for closer inspection.
4. Display test/native ligand with their interacted residues for geometrical inspection and comparison with ease.
5. Displayed test ligand.
6. To show the binding interaction in 2D.

Top

2.4.4 Visualize protein-ligand interaction in 2D



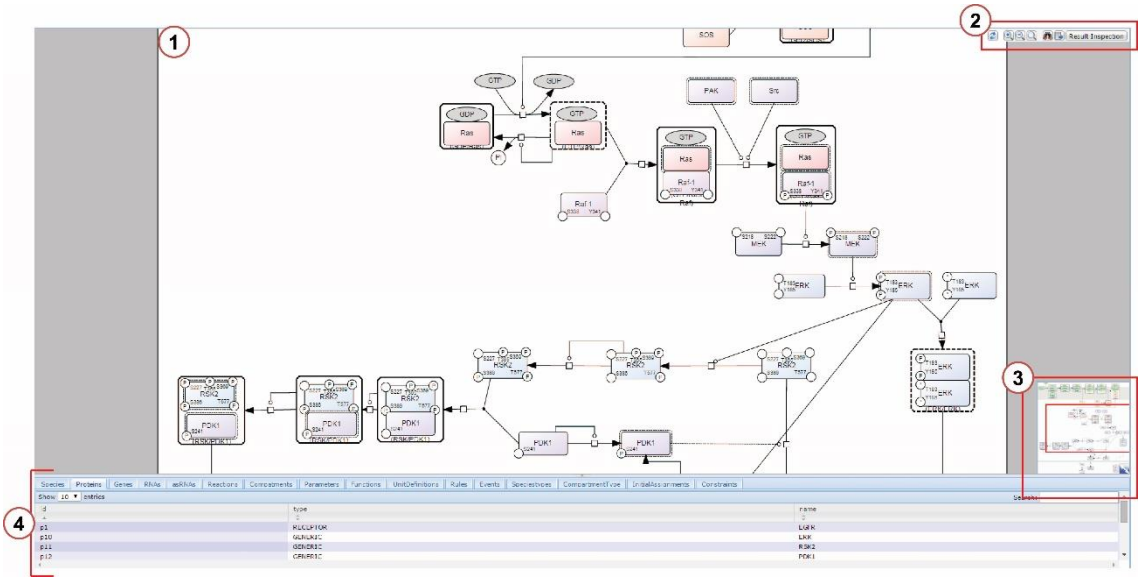
1. Visualize the binding interaction of the test compound in 2D. Interacted residues are identified by LIGPLOT.
2. For a reference, the binding interaction of the native ligand (if available) is also analyzed and displayed.

Top

2.5 Map inspection

Upon the availability of pathway map, docking scores of the test compound against the specified network proteins are optionally converted into a white-to-red color scale (indicating binding strength) or a white/red color classified by a specified cut-off value (indicating active/inactive). The colored results are then projected to the pathway map to directly display predicted binding affinities.

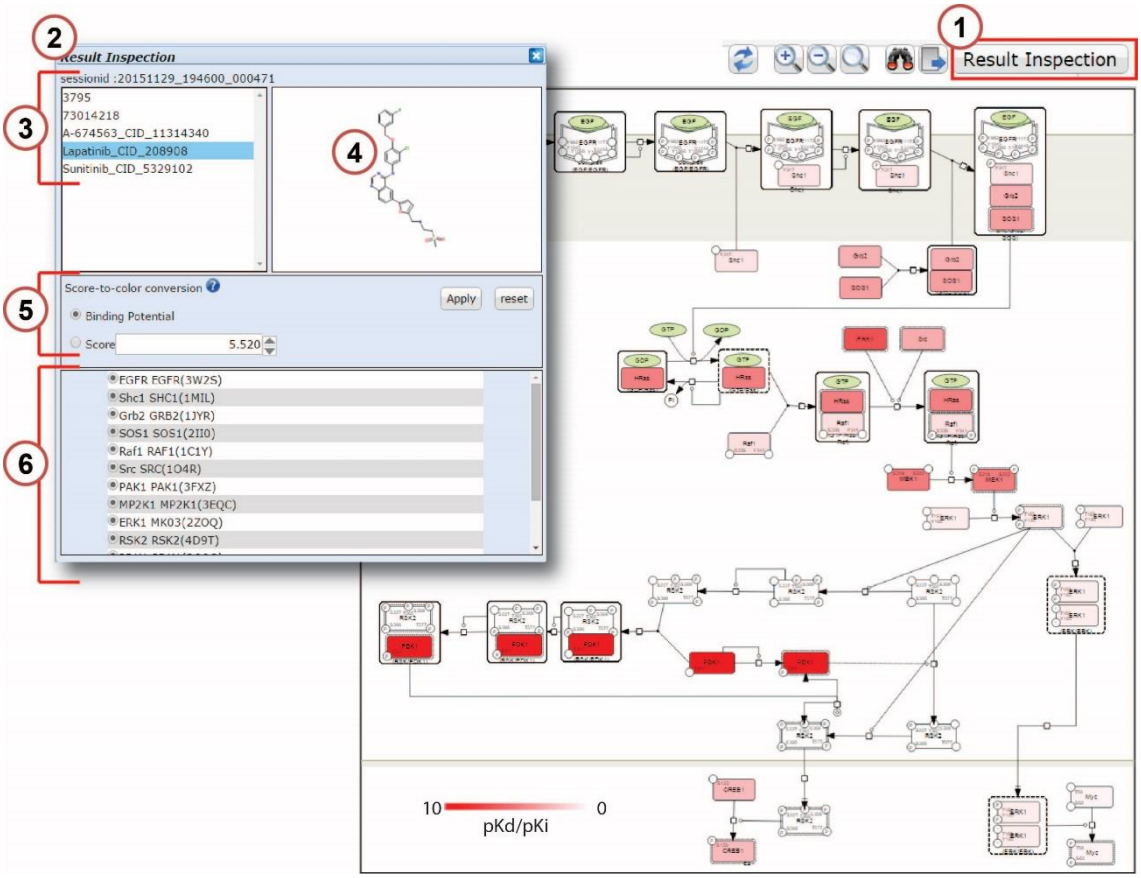
2.5.1 Screen configuration



1. The pathway map provided by user, which displays information on various molecular interactions and reaction networks in format of Systems Biology Graphical Notation (SBGN).
2. Tools for manipulate the map. From left to right, they are functions of 1) refresh the view, 2) zoom in the map, 3) zoom out the map, 4) zoom fit the map to the view, 5) search object (e.g. proteins) by key word, 6) export the map in SVG/PNG format, and 7) convert the docking results to color scales and project on the map for a direct inspection (detailed in session of *Colored map inspection*).
3. Bird's-eye-view function helps navigation for a big pathway map.
4. An interactive object-list panel lists the information of objects shown on the map (e.g. species, proteins, reactions and so on), and simply by clicking on an entry to centre and highlight the object in the view It is most useful for looking for a particular object from a complex map.

Top

2.5.2 Colored map inspection

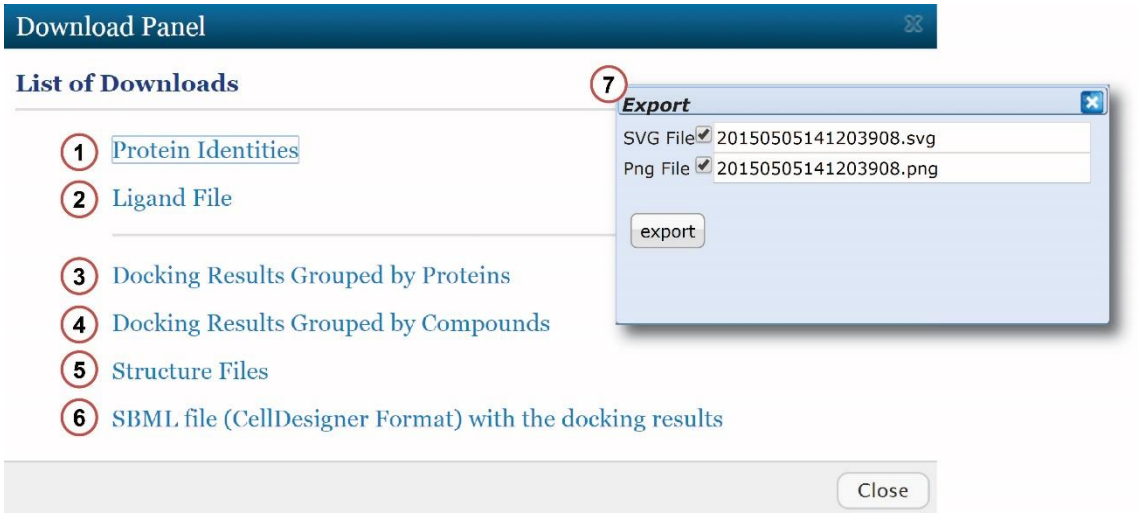


1. Clicking on "Result Inspection" to pop up a window (2) for visualizing selection.
2. A window for selections of molecules (compounds/proteins) and visualizing methods.
3. List of test compounds. Clicking on one of them to display it structure in 2D (4) and its simulation results on the map.
4. 2D representation of the selected compound.
5. Methods of score-to-color conversion. The first option is to convert the predicted binding affinities of the test compound against all of the network proteins into a white-to-red color scale from 0 to 10, as the examples shown on the map. It is useful to make a comparison for binding strength. The second option allows users to set a cutoff value to convert docking score into red or white color, indicating active or inactive. Docking score at 5.52 p*K*d can be a common value to classify compound's binding activity as it is equal to dissociation constant 3 μM.
6. List of proteins. It provides options for cases of that uses have specified multiple structure models of a protein.

Top

2.6 File download

Screening result files from systemsDock are rich in details. Tabulated docking scores (csv file), docked poses (SDF file), processed protein structures (PDB file) and determined binding interactions (png file) as well as the pathway map (SBML file, svg/png file), if available, may be downloaded.



1. A csv file contains information of specified protein identities, including protein names, PDB IDs and binding sites.
2. The structure file of the test compounds in SD format.
3. List of docking scores of test compounds grouped by proteins.
4. List of docking scores of test compounds grouped by compounds.
5. The docking poses of test compounds against each protein in SD format as well as ligand-free protein structure files in PDB format. Those files are ready for inspection by molecule visualizer applications (e.g. PyMol).
6. While a SBML file for protein specification was provided by user, a new SBML file contains docking results of test compounds against each protein can be downloaded. This file can be opened by tools of structured diagram editor (e.g. CellDesigner).
7. At map inspection, the colored pathway map can be exported in formats of svg or png.

Top

2.7 Session management

A unique session ID will be generated for each of the use of systemsDock. Previous molecule preparation or docking results can be retrieved by giving a session ID. Upon the molecule specification, the simulation may take longer time. Users can close the web browser and retrieve the progress or result simply giving the session ID. After a sign-in operation, users can also manage previous sessions.

1. Displaying the current session ID. Users can copy it and keep it for future operation.
2. Clicking it to paste old session ID to retrieve previous works.
3. A sign-in operation for session management. After signed in, the current session ID will be referred for the corresponding account. The user can manage (e.g. retrieve/delete) the previous sessions with ease.

Top

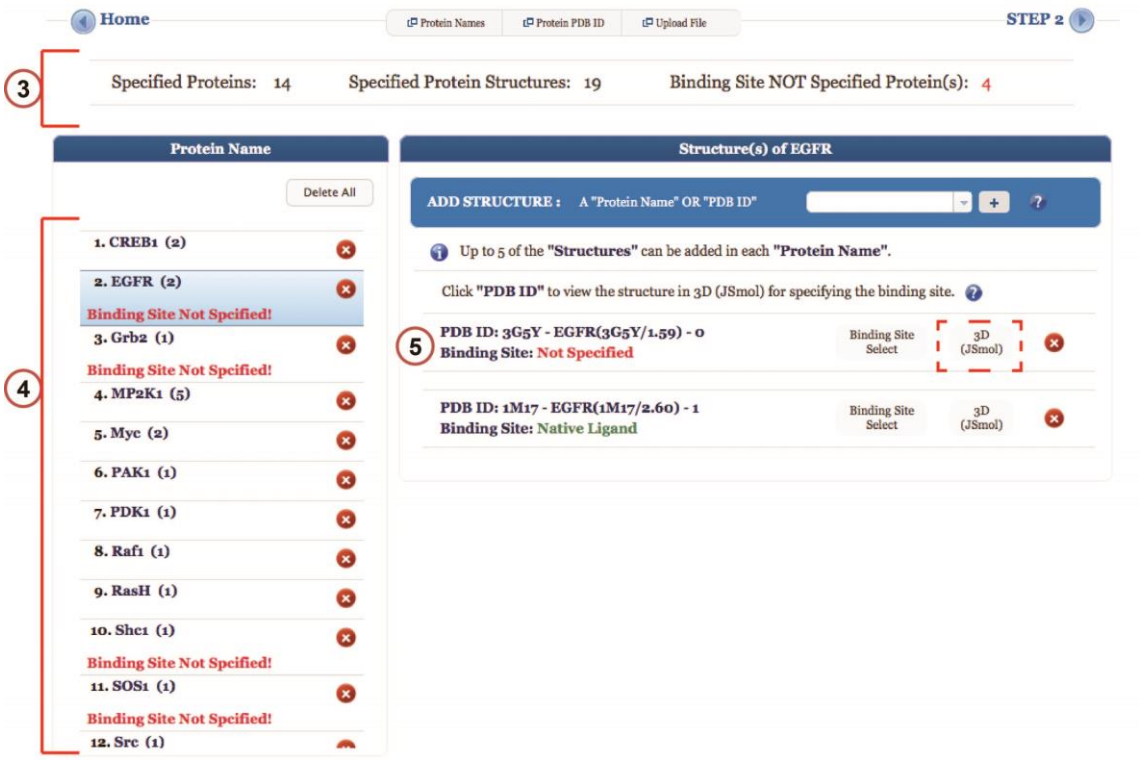
3. Demonstration

The following will demonstrate a screening for three compounds over a simple EGFR pathway. Files of compounds and pathway can be downloaded from here: SDF and SBML.

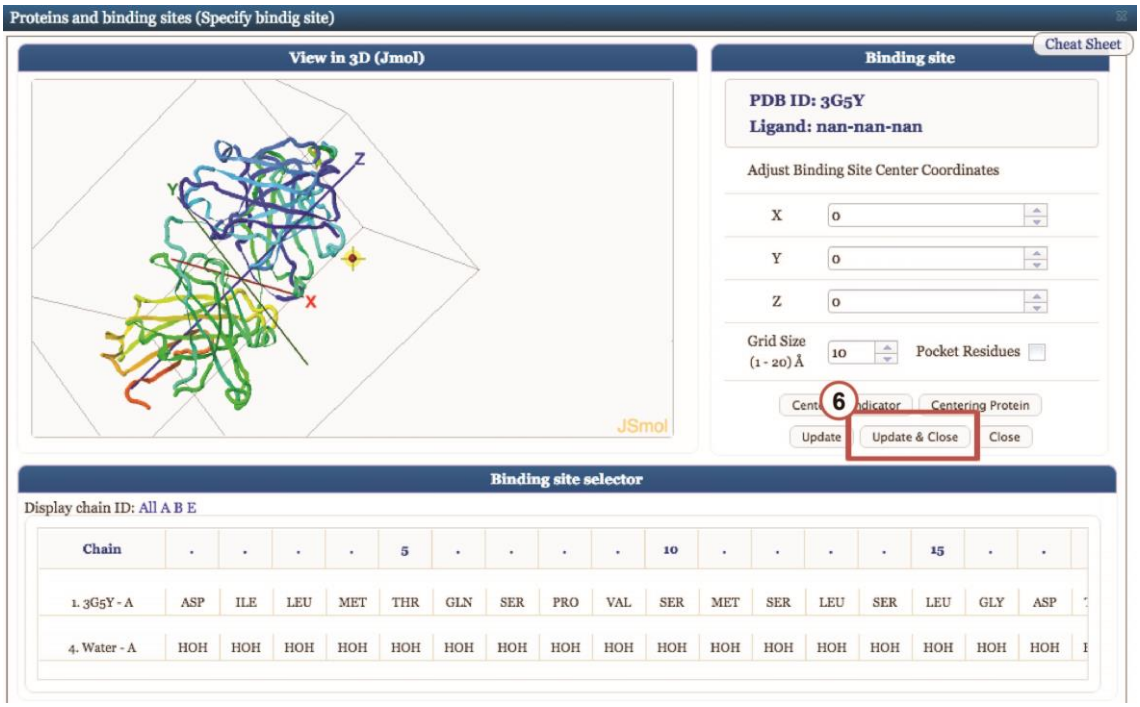
3.1 Protein specification

1. Click the "Protein" shown on the top of home page to enter the Protein Specification page.

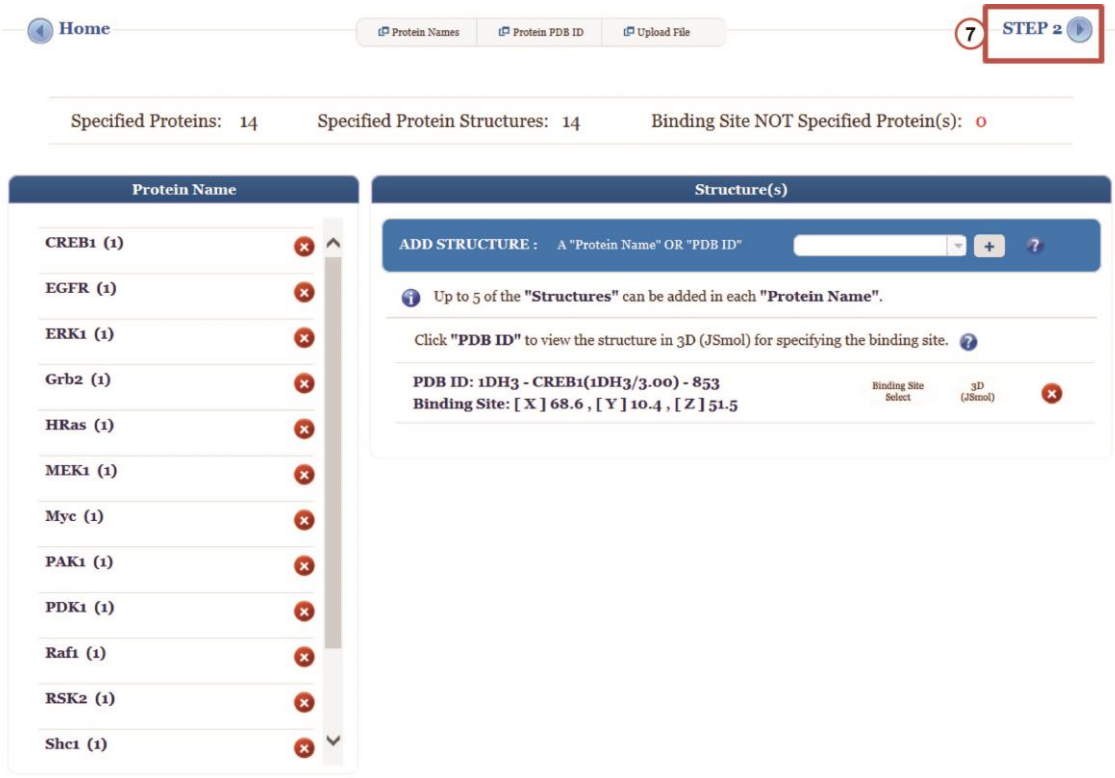
2. Click "Upload File" to upload the EGFR pathway file.



3. There are 14 pathway proteins found in the map. Structures of 12 out of them have been automatically retrieved from our in-house protein identity-to-structure mapping system for docking simulation, but 5 of whose binding sites cannot be identified.
4. The identities of the 14 pathway proteins are listed. Users can click on any of the entries to modify the specification. For example, clicking on the EGFR entry allows to add more EGFR structures or to define the binding site by clicking on 3D (5).
5. To define the binding site as mentioned in session 2.1.3.

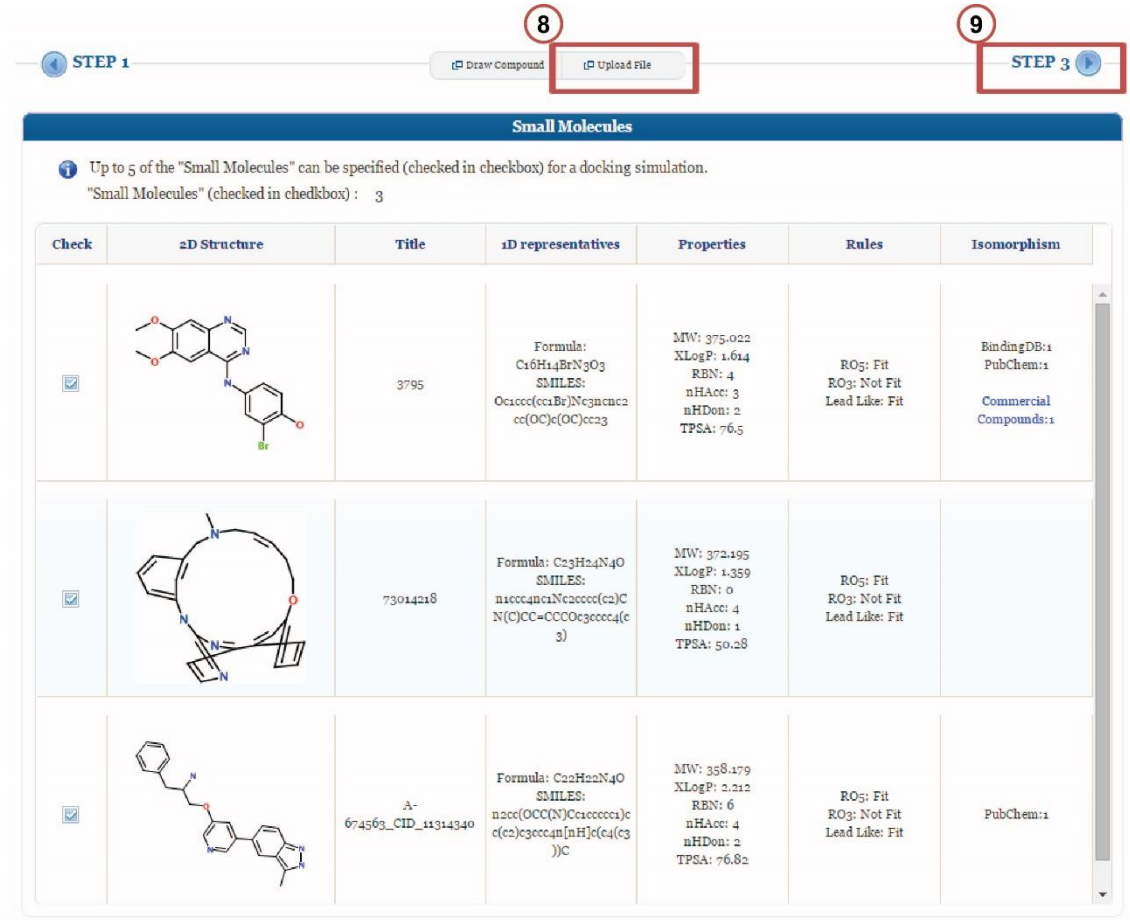


6. Clicking on “Return to Protein List” for next case.



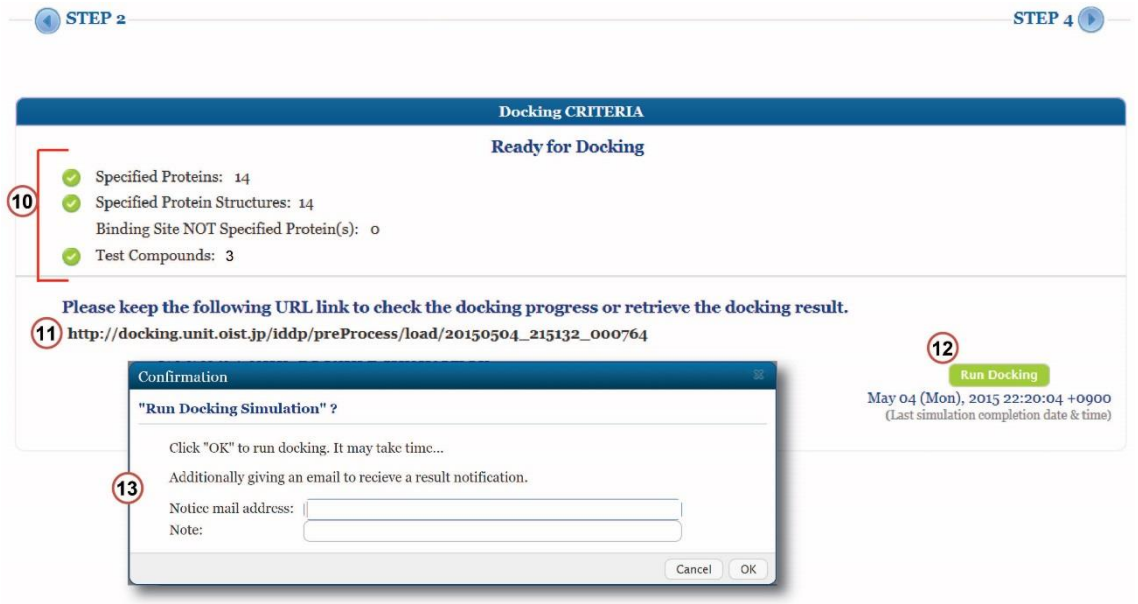
7. After the protein specification has been done, clicking “STEP 2” to process Ligand Preparation

3.2 Ligand preparation



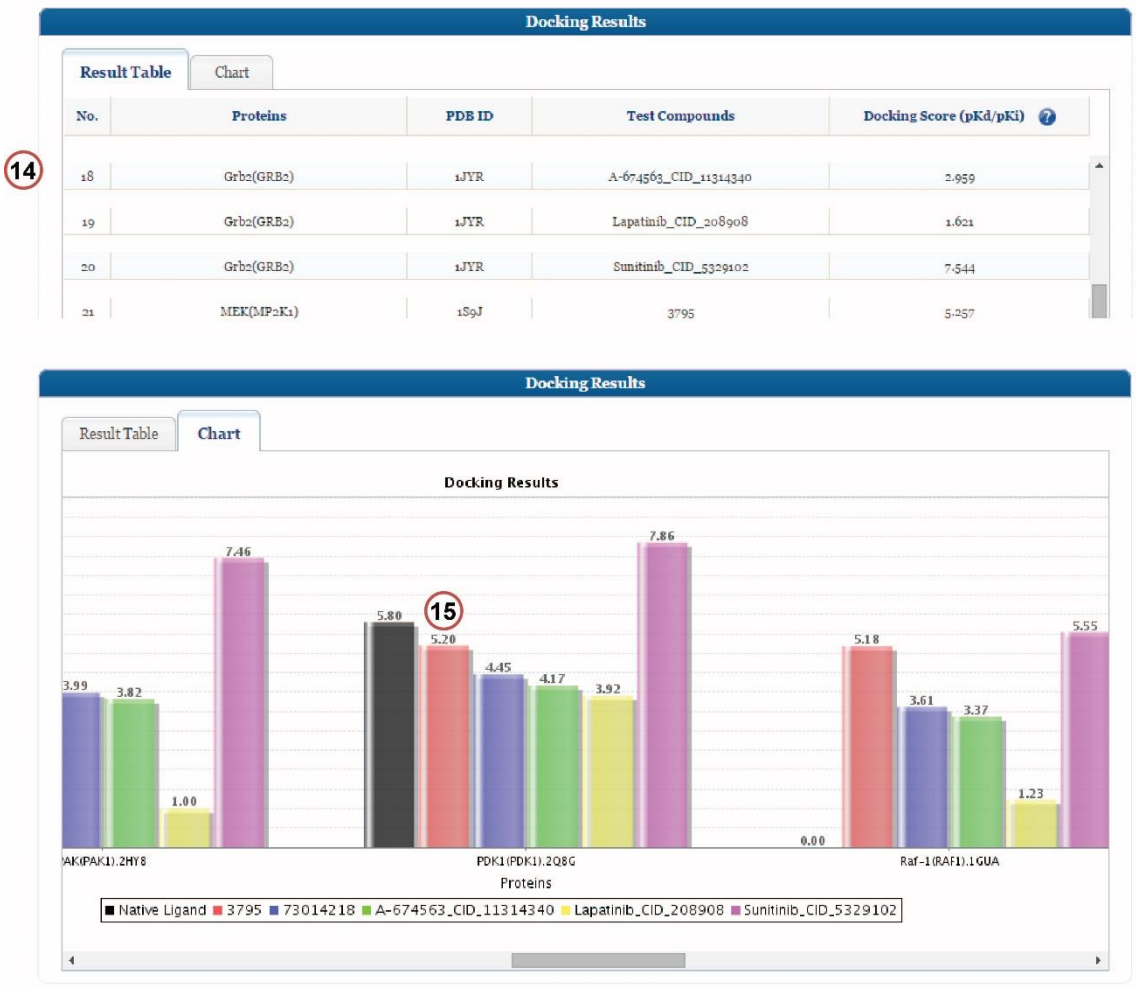
8. Clicking on “Upload File” to upload the test compound file. The compounds will be tabulated together with their chemical properties and others. Users can click on any entry of them to display more details (see session 2.2).
9. Clicking “STEP 3” for docking simulation.

3.3 Docking simulation



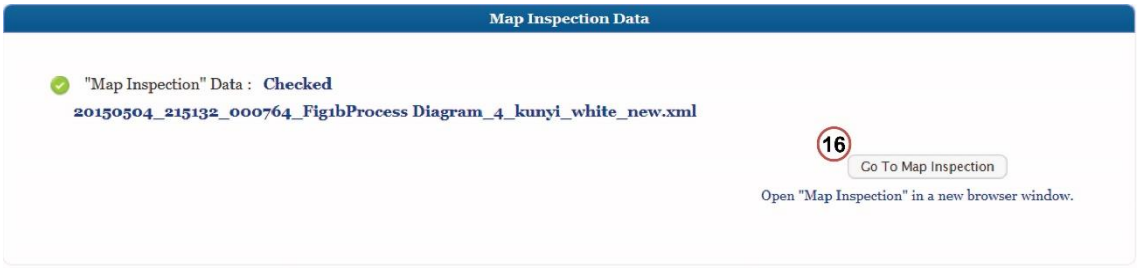
10. To confirm the specification for docking simulation.
11. The simulation might take time. A link is provided for checking the docking progress or retrieving the results.
12. While the specification has been done without a problem, the “Run Docking” button will be activated, and clicking on to run docking.
13. Giving an email for notification when simulation is done

3.4 Docking results

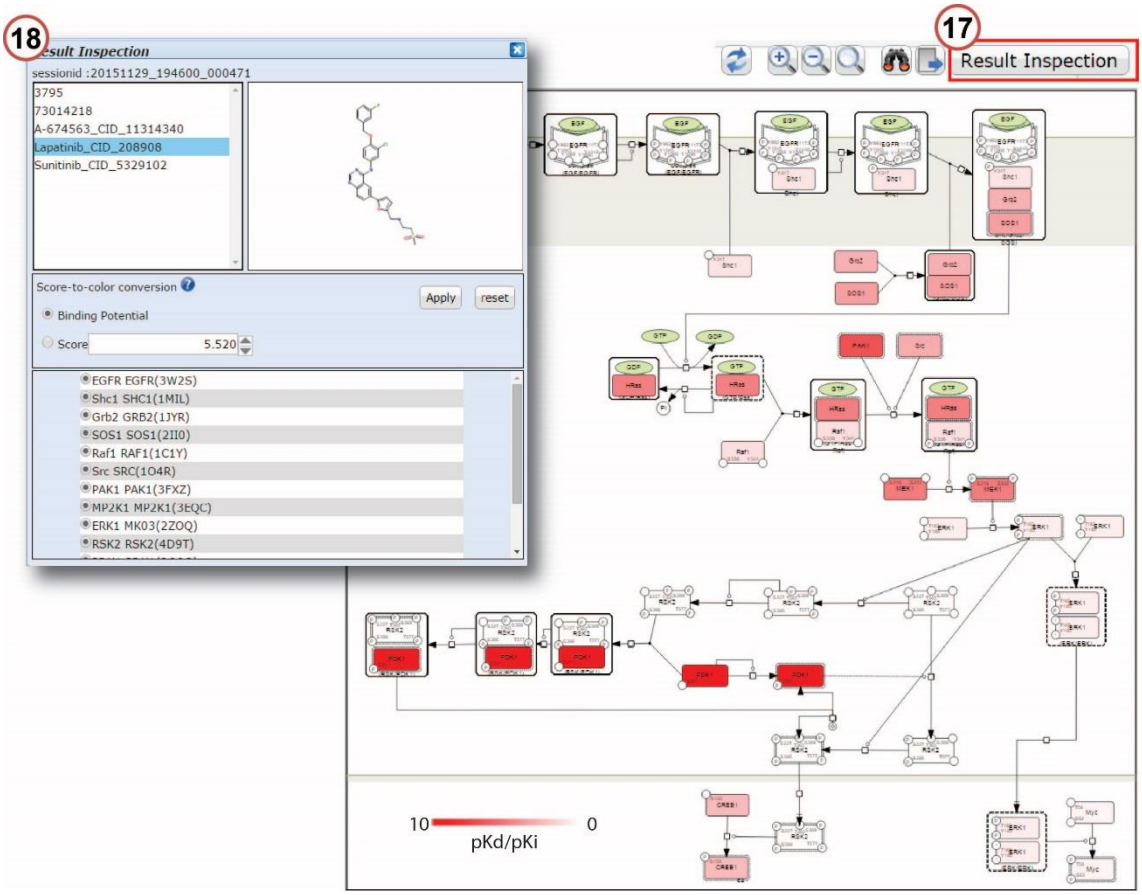


14. Docking scores are listed in an interactive table (see session 2.4.1 for detailed manipulation).
15. An interactive histogram is also available for a comparison (see session 2.4.2 for detailed manipulation).

3.5 Map inspection



16. Upon the availability of pathway map provided by user, click "Go To Map Inspection" for inspecting results over a interactive map.



17. Click on "Result Inspection" to select a test compound and to show its results.
18. Compound specification for result inspection. The map is interactive, allowing users to click on a colored node for more information (see session 2.5.2 Colored map

inspection for details).

Top