

Structure-Based Virtual Screening

Created with: Release 2021-3

Prerequisites: working knowledge of Maestro

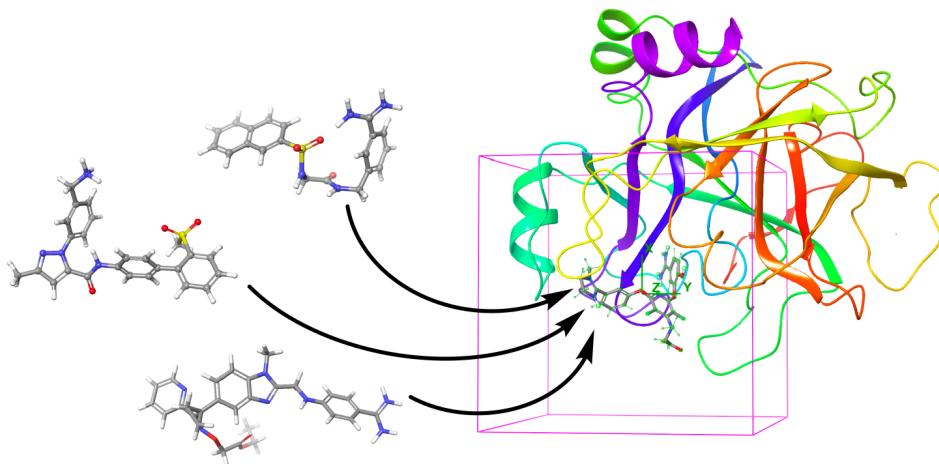
Files Supplied: SBVS Worksheet

Categories: biochemistry, medicinal chemistry

About this Lesson

This lesson will focus on an important early stage of drug discovery in which protein structures and molecular modeling are utilized to identify molecules that can be further developed into drugs.

Using Maestro, students will learn how to perform a virtual screen for potential inhibitors of FXa using the ligand docking application Glide. Students will learn how to generate a protein receptor grid, dock a set of ligands into the receptor grid, and analyze the docking results. Students can then sketch their own inhibitor designs and evaluate their value.



Learning Objectives

- Learn the steps of a molecular docking workflow using Schrödinger's Glide
- Perform a structure-based virtual screen of a small set of ligands
- Design your own inhibitor for FXa and determine its docking score

Standards

- *ACS Guidelines*
 - Biological macromolecules ([Section 5.1](#))
- *ETS Chemistry GRE*
 - Organic Chemistry – Amino acids, Peptides ([3F](#))
- *AAMC MCAT*
 - Structure, function, and reactivity of biologically-relevant molecules ([5D](#))

Assessments

The following types of formative assessments are embedded in this lesson:

- Assessment of student understanding through discussion of warm-up questions and filling in any knowledge gaps about structure-based virtual screen steps
- Visual assessment of student-generated docking scores from their own set of ligands

Warm-Up Questions: To be done on their own or at the beginning of class

Read the article [“Structure-Based Virtual Screening: From Classical to Artificial Intelligence”](#) and answer the following questions.

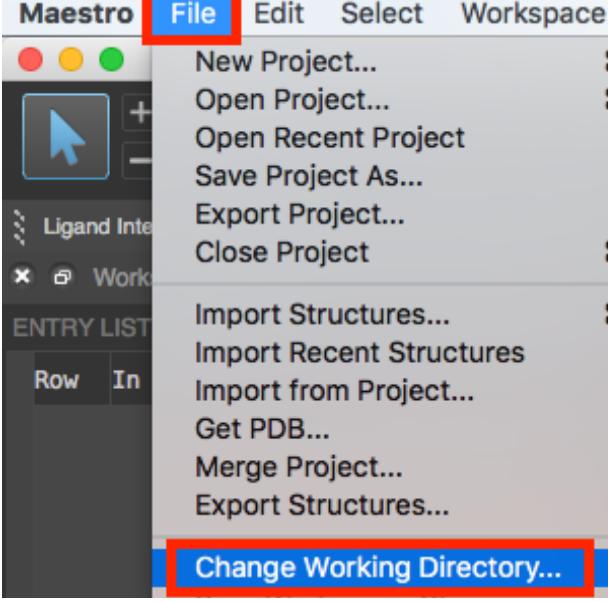
- 1) What is a typical drug development timeline from drug target identification to clinical trials?

- 2) What are some advantages and disadvantages to performing structure-based virtual screens?

Lesson Outline

1. [What you will need for this lesson](#) - p. 3
2. [Introduction to Structure-Based Virtual Screening](#) - p. 5
3. [Generating a Receptor Grid](#) - p. 6
4. [Docking the Cognate Ligand and Screening Compounds](#) - p. 11
5. [Analyzing Results and Binding-Site Characterization](#)- p. 17
6. [Individual Exercises](#) - p. 23
7. [Summary, Additional Resources, and References](#) - p. 24

1. What you will need for this lesson

	<ol style="list-style-type: none">1. Go to the 'Data' folder and open your Class Folder found on the virtual cluster's desktop.2. Right-click on the folder called "Structure_Based_Virtual_Screen" and copy folder to Desktop<ul style="list-style-type: none">• Here, you will find the lesson plan, worksheet, and any additional resources
 <p>Figure 1-1. Open Maestro.</p>	<ol style="list-style-type: none">3. Open Maestro<ol style="list-style-type: none">a. See Starting Maestro if you need help
 <p>Figure 1-2. Change Working Directory option.</p>	<ol style="list-style-type: none">4. Go to File > Change Working Directory5. Find your "Structure_Based_Virtual_Screen" folder that you duplicated to your Desktop, and click Choose

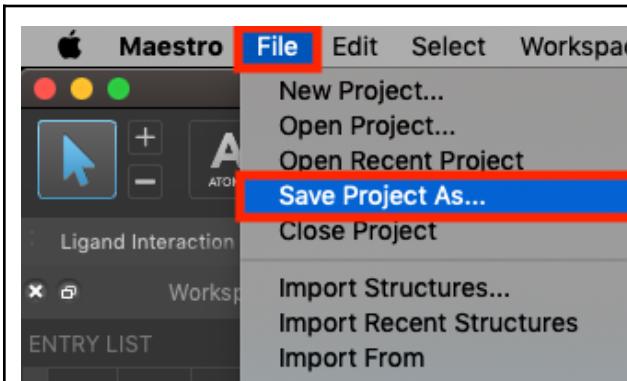


Figure 1-3. Save Project panel.

6. Next, go to **File > Save Project As**
7. Type “**SBVS_tutorial**” and click **Save**
 - a. The project will be titled **SBVS_tutorial.prj**

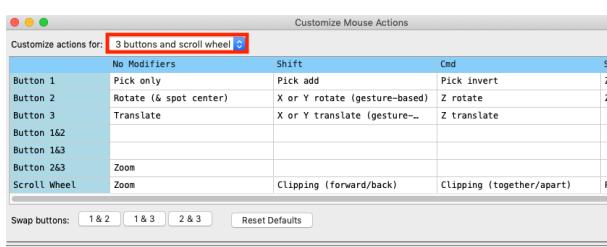


Figure 1-4. Choose the best mouse option for your set up.

8. Finally, check your **Mouse Actions**
 - a. PC : **Edit > Customize Mouse Actions**
 - b. Mac : **Workspace > Customize Mouse Actions**
9. Make sure you have the **best option chosen for your set up**. This lesson was written with a three-button mouse with a scroll wheel, meaning the scroll wheel is a button as well as a wheel. If you do not have a mouse, choose **Trackpad**.

2. Introduction to Structure-Based Virtual Screening

Molecular modeling encompasses a wide array of approaches that impact the early stages of drug discovery: hit ID, hit-to-lead, and lead optimization. Generally, these molecular modeling approaches are broken down into two categories: ligand-based and structure-based approaches. Ligand-based approaches use information from a hit or series of hits to inform on next stages in drug discovery. This information includes molecular fingerprints, shape, charge, and more. Structure-based approaches use information not only from a ligand structure but also from a target structure. A target is another term for a protein or other macromolecule that is being targeted by a drug. Both of these approaches are often used synergistically in all stages of computer-aided drug discovery and in combination with fragment-based approaches as well.

In this lesson, we will be performing a structure-based virtual screen of potential FXa inhibitors. Structure-based approaches allow for the identification of key residues around the ligand and water energetics in the binding site that could modify or enhance the binding of one compound over another. A binding site is a pocket or surface of a target protein where a compound or drug binds to elicit a downstream effect in a disease pathway.

Figure 1 below shows a schematic for the steps involved in a structure-based virtual screen. Structure files obtained from the PDB, vendors, and other sources often lack necessary information for performing modeling-related tasks. Typically, these files are missing hydrogens, partial charges, side chains, and/or whole loop regions. In order to make these structures suitable for modeling tasks, we use the Protein Preparation Workflow to resolve issues. Similarly, ligand files can be sourced from numerous places, such as vendors or databases, often in the form of 1D or 2D structures with unstandardized chemistry. LigPrep can convert ligand files to 3D structures, with the chemistry properly standardized and extrapolated, ready for use in virtual screening.

In this lesson, the protein, cognate ligand, and virtual screening ligands have already been prepared in order to save time. However, these preparation steps are a necessary part of a virtual screen and must be done before docking. Please see the lesson on [Protein-Ligand Interactions](#) for guidance on using the Protein Preparation Workflow and LigPrep.

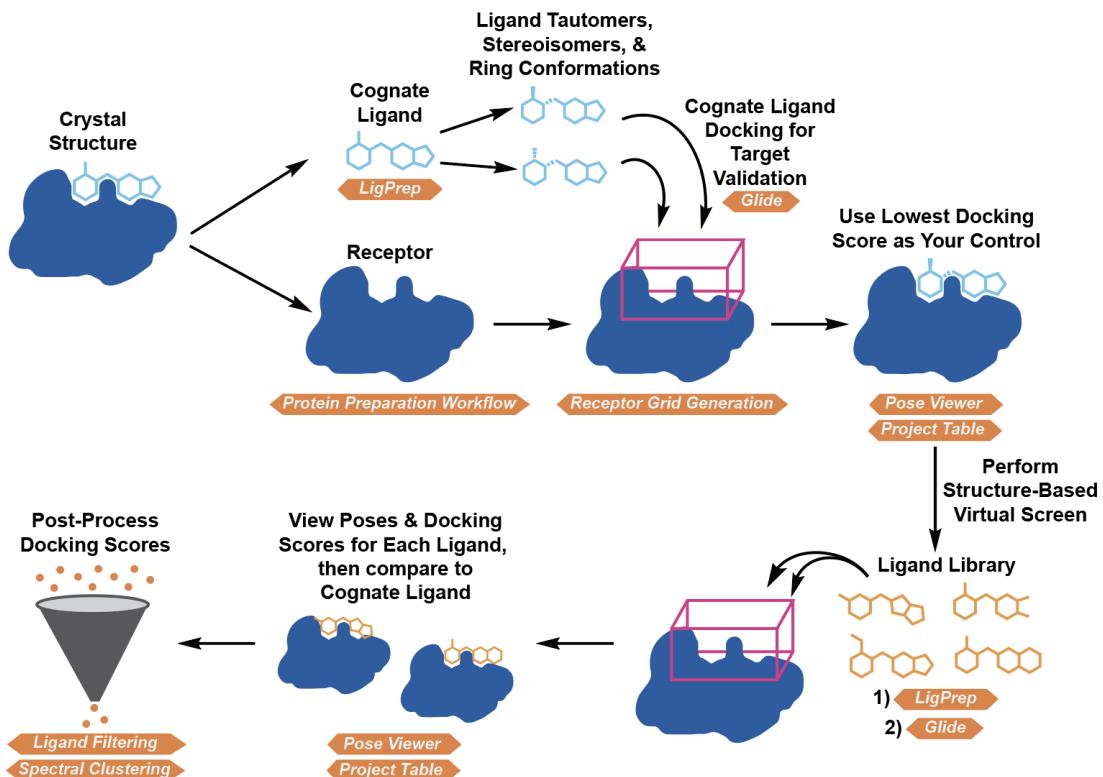


Figure 1. Workflow for a Structure-Based Virtual Screening

3. Generating a Receptor Grid

Grid generation must be performed prior to running a virtual screen with Glide. The shape and properties of the receptor are represented in a grid by fields that become progressively more discriminating during the docking process. To add more information to a receptor grid, different kinds of constraints can be applied during the grid generation stage. For a comprehensive overview of constraint options, see the [grid generation videos](#) on our website or the [Glide User Manual](#) ([Help > Help > User Manuals > Glide User Manual](#)). In this tutorial, we will set a hydrogen bond constraint in our receptor grid.

3.1 Identify the binding site

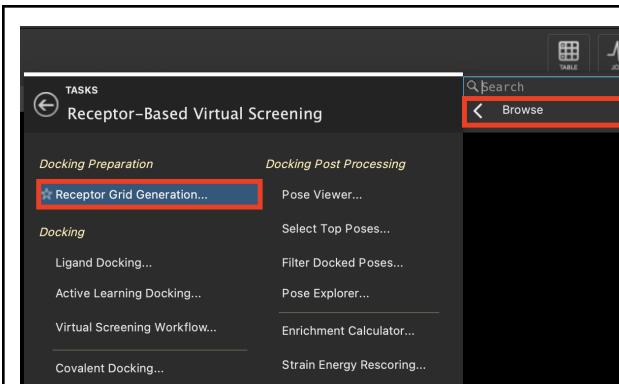


Figure 3-1. Receptor Grid Generation option in Receptor-Based Virtual Screening.

1. Click the **In** circle next to **1fjs_prep_complex** to include it in the Workspace

2. Double-click **Presets**

- **1fjs_prep_complex** is rendered using the Custom Preset

3. Go to **Tasks > Browse > Receptor-Based Virtual Screening > Receptor Grid Generation**

- The Receptor Grid Generation panel opens

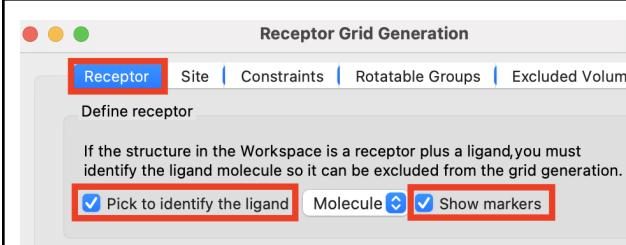


Figure 3-2. The Receptor tab of Receptor Grid Generation.

4. Under Define Receptor, check the boxes for **Pick to Identify the ligand (Molecule)** and **Show Markers**

- A banner in the Workspace will prompt you to click on an atom in the ligand

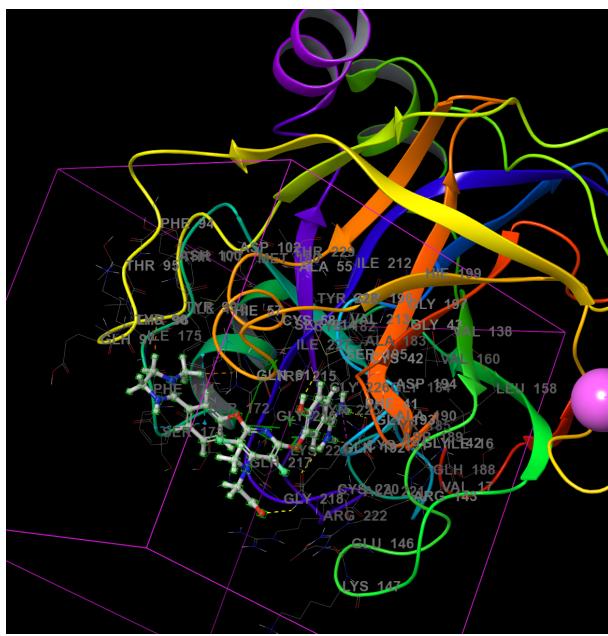


Figure 3-3. The ligand is defined to be excluded from grid generation.

5. Click on the ligand

- The ligand is now highlighted with a purple box around it
 - The ligand will be excluded from the grid generation

Note: The purple bounding box defines the region that the docked molecule(s) can occupy to satisfy the initial stages of docking

3.2 Define the bounding box dimensions

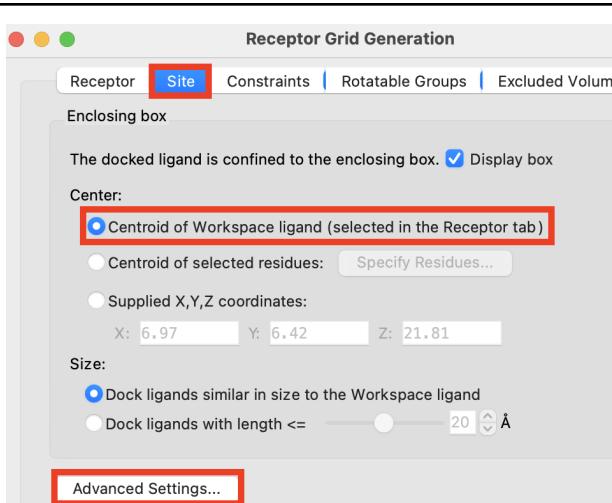


Figure 3-4. The Site tab of Receptor Grid Generation.

1. Click the **Site** tab
 2. Select **Centroid of Workspace ligand (selected in the Receptor tab)**
 3. Click **Advanced Settings**
 - A green inner bounding box appears

Note: The green bounding box defines the region in which the centroid of the docked molecule(s) must occupy to pass the initial stages docking

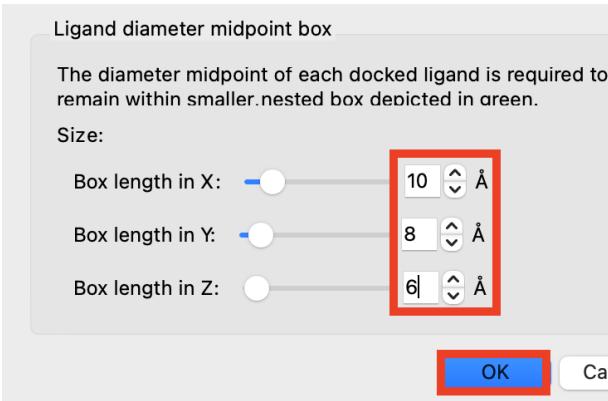
 <p>Ligand diameter midpoint box</p> <p>The diameter midpoint of each docked ligand is required to remain within smaller nested box depicted in green.</p> <p>Size:</p> <p>Box length in X: 10 Å</p> <p>Box length in Y: 8 Å</p> <p>Box length in Z: 6 Å</p> <p>OK</p>	<ol style="list-style-type: none"> 4. Adjust the settings for X, Y, and Z sizes to 10, 8, and 6 Å, respectively. <ul style="list-style-type: none"> ○ The shape of the green box is changed 5. Click OK
--	--

Figure 3-5. Ligand diameter midpoint box panel.

3.3 Set a hydrogen bonding constraint

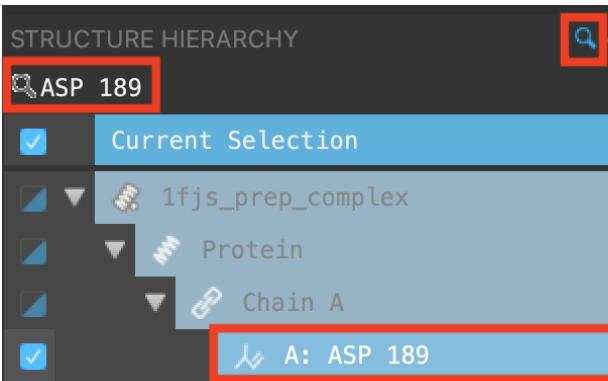
 <p>STRUCTURE HIERARCHY</p> <p>ASP 189</p> <p>Current Selection</p> <p>1fjs_prep_complex</p> <p>Protein</p> <p>Chain A</p> <p>A: ASP 189</p>	<ol style="list-style-type: none"> 1. Type L to zoom to the ligand 2. In the Structure Hierarchy, click the magnifying glass 3. In the search field, type ASP 189 4. Select ASP 189 <p>Note: Please see the Introduction to Structure Preparation and Visualization tutorial for instructions on how to add residue labels and show H-bonds</p>
 <p>Fit: AUTO LIGAND</p> <p>STYLE</p>	<ol style="list-style-type: none"> 5. Under Fit, click Fit view to selected atoms

Figure 3-6. Search in the Structure Hierarchy.

Figure 3-7. Zoom to selected atoms.

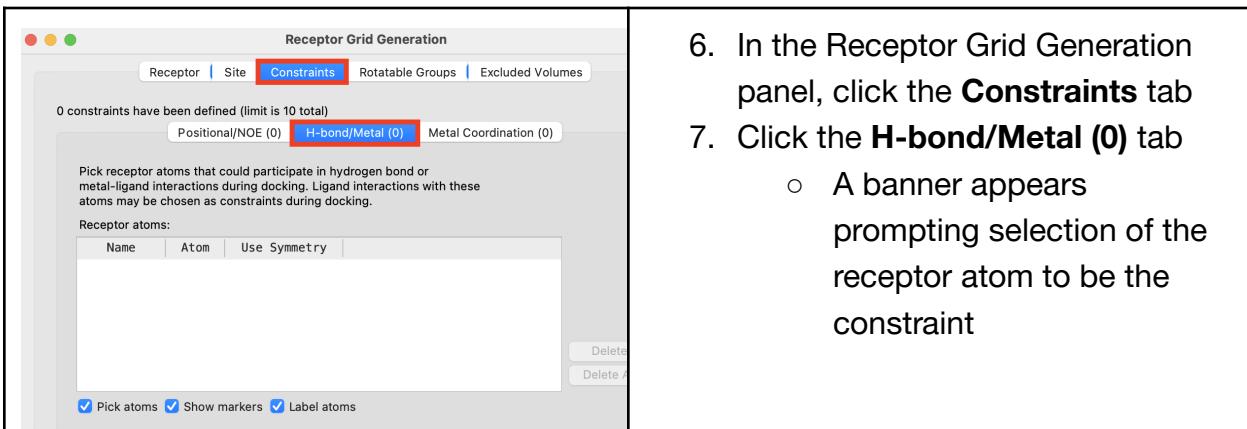


Figure 3-8. The Constraints tab of Receptor Grid Generation.

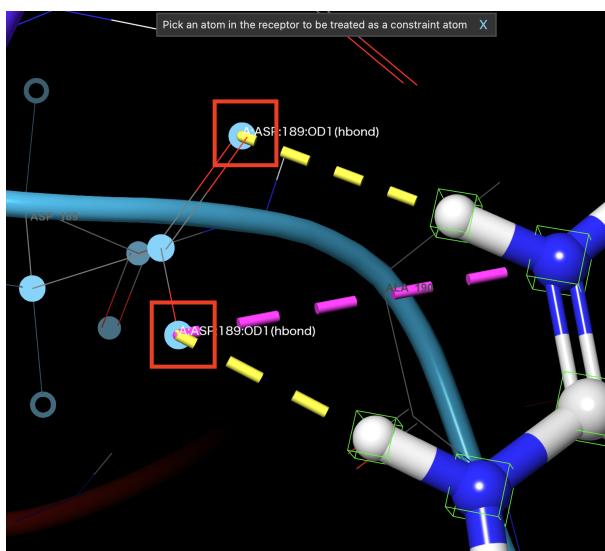


Figure 3-9. Constraint defined on ASP 189.

6. In the Receptor Grid Generation panel, click the **Constraints** tab

7. Click the **H-bond/Metal (0)** tab

- A banner appears prompting selection of the receptor atom to be the constraint

8. Click an **oxygen atom** of the ASP 189 sidechain

- Both oxygens are highlighted
- An H-bond constraint is defined in the Receptor atoms table

Figure 3-10. Run receptor grid generation job.

**9. Change Job name to
glide-grid_1fjs**

10. Click Run

- This job will take about a minute
- A folder named **glide-grid_1fjs** is written to your Working Directory

Question #1:

Why is it necessary to generate a receptor grid? What would happen if you proceeded with docking a ligand without a receptor grid?

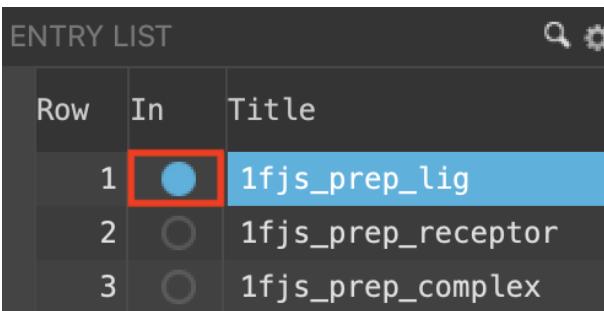
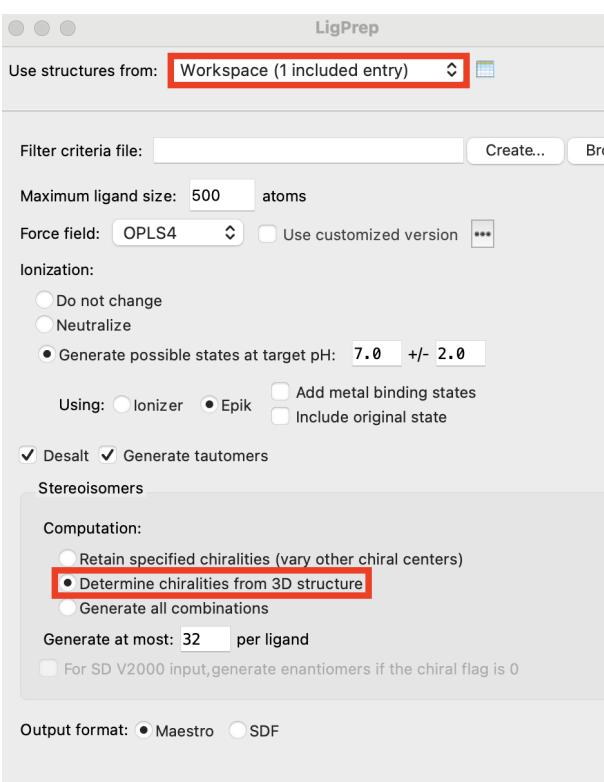
4. Docking the Cognate Ligand and Screening Compounds

The minimum requirements for running a Glide virtual screen are a grid file and a ligand file. It is strongly recommended that the grid file be generated from a protein prepared using the Protein Preparation Workflow and the ligand file be prepared using LigPrep. Additionally, you can choose the scoring function, set ligand- and receptor-based constraints, and define the output. Please see the Glide User Manual for more detail. In this section, we will include the hydrogen bonding constraint that was created in the previous step.

First, we will dock the cognate ligand, which is a helpful way to benchmark a virtual screen of compounds with unknown binding activity against a target. The information

gained from this step can help with evaluating poses and beneficial interactions, which is useful for hit finding. Second, we will dock the screening compounds from a prepared ligand file, `501igs_epik.mae.gz`. Both jobs will use the receptor grid file that was generated in the previous step.

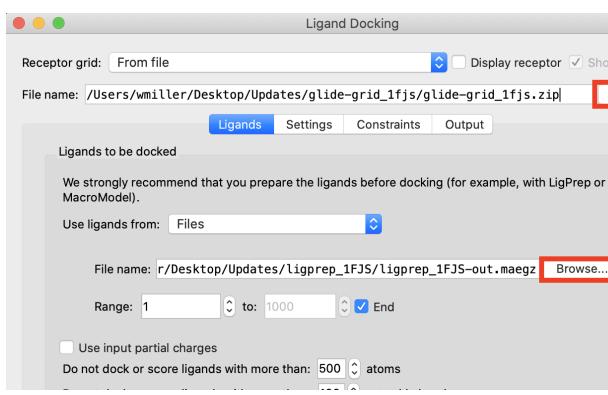
4.1 Prepare the cognate ligand

 <p><i>Figure 4-1. Include 1fjs_prep_lig.</i></p>	<ol style="list-style-type: none"> 1. <u>Include 1fjs_prep_lig in the Workspace</u> 2. Go to Tasks > Browse > LigPrep <ul style="list-style-type: none"> o The LigPrep panel opens 3. For Use structures from, choose Workspace (1 included entry) 4. Under Stereoisomers, select Determine chiralities from 3D structure 5. Change Job name to ligprep_1fjs 6. Click Run <ul style="list-style-type: none"> o A banner appears when the job has been <u>incorporated</u> o A new group is added to the <u>Entry List</u>
 <p><i>Figure 4-2. The LigPrep panel.</i></p>	

Question #2:

Preparing a ligand using LigPrep may produce multiple output structures for each input structure by generating different protonation states, stereochemical outcomes, tautomers, and ring conformations. Why is it important to prepare a ligand before proceeding with docking?

4.2 Dock the cognate ligand



The screenshot shows the 'Ligand Docking' software interface. In the top left, there's a dropdown menu 'Receptor grid: From file'. Below it, a 'File name:' field contains the path '/Users/wmiller/Desktop/Updates/glide-grid_1fjs/glide-grid_1fjs.zip'. A red box highlights this field. Below this are tabs for 'Ligands' (which is selected), 'Settings', 'Constraints', and 'Output'. Under the 'Ligands' tab, there's a note: 'We strongly recommend that you prepare the ligands before docking (for example, with LigPrep or MacroModel)'. A 'Use ligands from:' dropdown is set to 'Files', with a red box highlighting the 'Browse...' button next to it. Below that, a 'File name:' field contains 'r/Desktop/Updates/ligprep_1FJS/ligprep_1FJS-out.maegz', with another red box highlighting the 'Browse...' button. There are range inputs 'Range: 1 to: 1000 End' and a checkbox 'Use input partial charges'. At the bottom, there's a note: 'Do not dock or score ligands with more than: 500 atoms'.

Figure 4-3. The Ligands tab of the Ligand Docking panel.

1. Go to **Tasks > Browse > Receptor-Based Virtual Screening > Ligand Docking**
 - The Ligand Docking panel opens
2. Next to Receptor grid, click **Browse** and choose **glide-grid_1fjs.zip**
3. In the Ligands tab, for Use ligands from, choose **Files**
4. Next to File name, click **Browse** and choose **ligprep_1FJS-out.maegz**

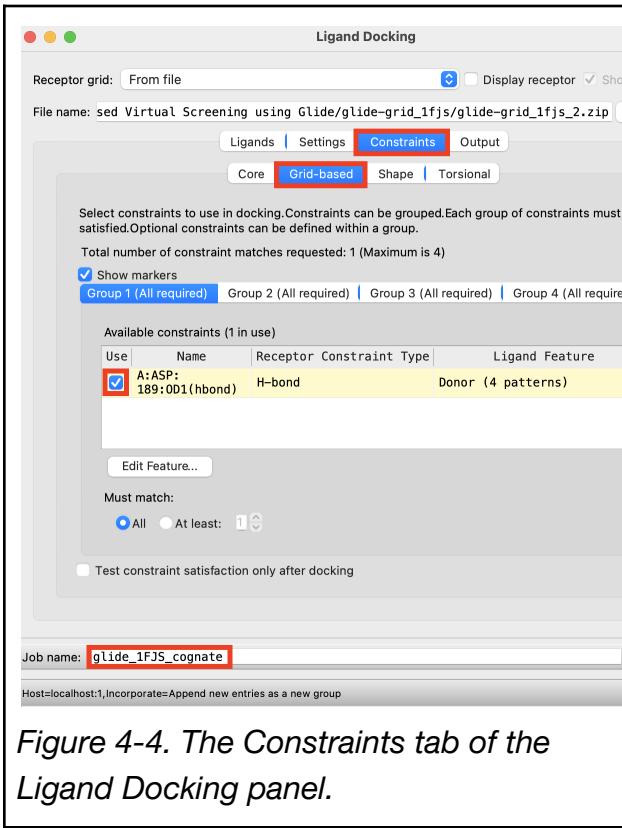


Figure 4-4. The Constraints tab of the Ligand Docking panel.

5. Click the **Constraints** tab
6. Click on the **Grid-based** tab
7. Under Use, **check** the H-bond constraint for ASP 189
8. Change Job name to **glide_1FJS_cognate**
9. Click **Run**
 - This job takes about a minute
 - A banner appears to show that files have been incorporated
 - A new group is added to the Entry List

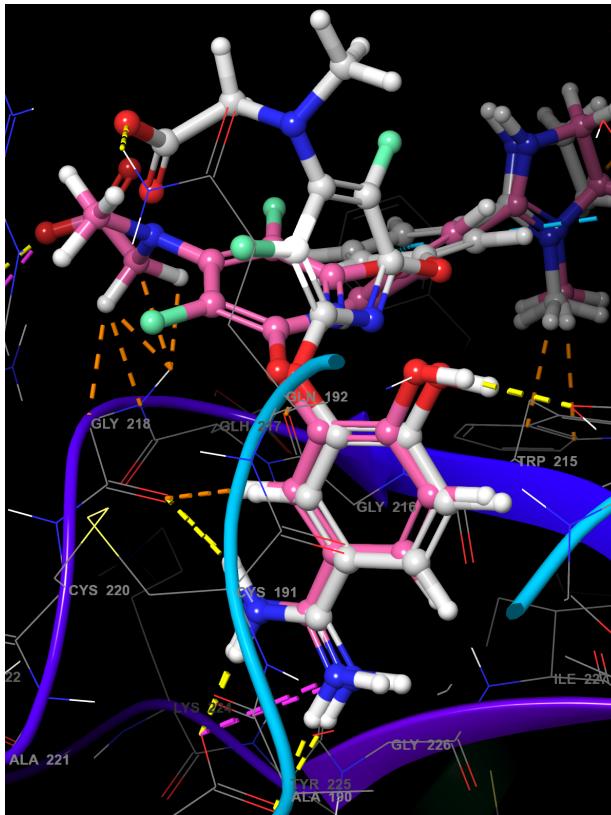


Figure 4-5. Binding pose of the top docked cognate ligand (pink) compared to the crystal structure (gray).

Note: The 1fjs_prep_complex entry is fixed in the Workspace, the top 1fjs_prep_lig entry is included, and the Pose Viewer panel appears

10. Include other ligand results

- H-bonds to ASP 189 are conserved

11. Double-click Presets

12. Double-click the In circle next to 1fjs_prep_complex

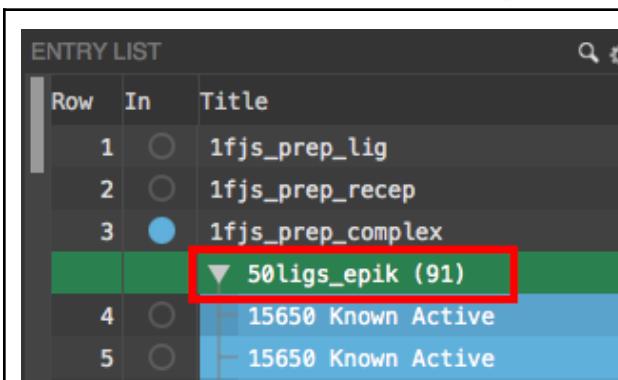
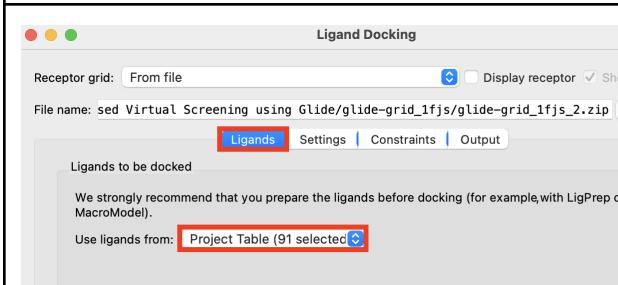
- The entry is no longer fixed in the Workspace

Note: Though only the top ranked result is in strong agreement with the crystallographic pose, all three results accurately capture the pose of the ligand in the binding site (with varying degrees of success in capturing the solvent exposed region)

Question #3:

What important protein-ligand interactions do you see when the cognate ligand is docked? List specific residues and define specific interactions that may play an important role in binding.

4.3 Dock the screening compounds

 <p>Figure 4-6. Select 50ligs_epik in the <u>Entry List</u>.</p>	<ol style="list-style-type: none">1. In the <u>Entry List</u>, select the group 50ligs_epik
 <p>Figure 4-7. Use ligands from selected entries.</p>	<ol style="list-style-type: none">2. In the Ligand Docking panel, click the Ligands tab3. For Use ligands from, choose Project Table (selected entries) <p>Note: Keep glide-grid_1fjs.zip as the receptor grid</p>

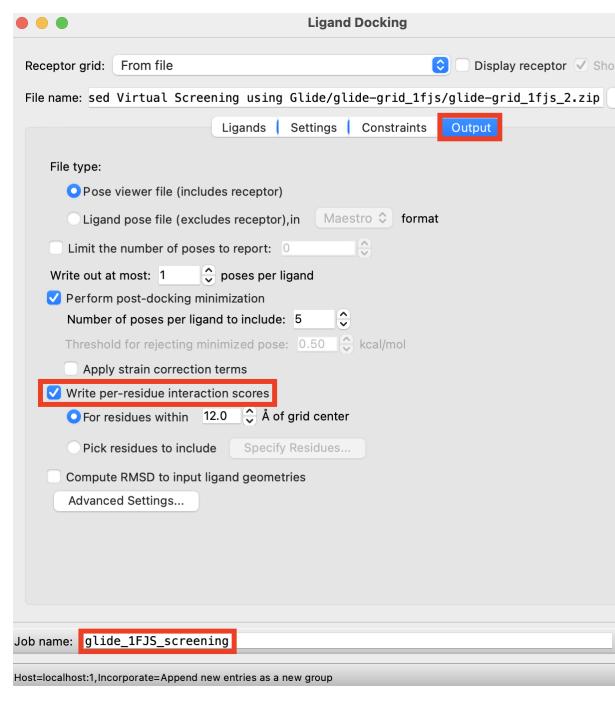


Figure 4-8. The Output tab of the Ligand Docking panel.

4. Click the **Output** tab
5. Check **Write per-residue interaction scores**
6. Change Job name to **glide_1FJS_screening**
7. Click **Run**
 - This job takes a few minutes
 - A banner appears to show that files have been incorporated
 - A new group is added to the Entry List

5. Analyzing Results and Binding-Site Characterization

Multiple Glide docking results can be viewed in the Entry List and be identified by the job name. Docked results will show the receptor in the first row and the docked ligand(s) in the subsequent row(s), where they are ordered by best to worst docking score, or Glide Gscore if Epik state penalties were not applied in LigPrep. The Glide Gscore is broken down by van der Waals electrostatic components and can be seen in the Project Table, using the Property Tree. You can read more about how docking scores/poses are generated [here](#) and [here](#) and what dependencies they have [here](#) and [here](#).

5.1 Visualize the results using Pose Viewer

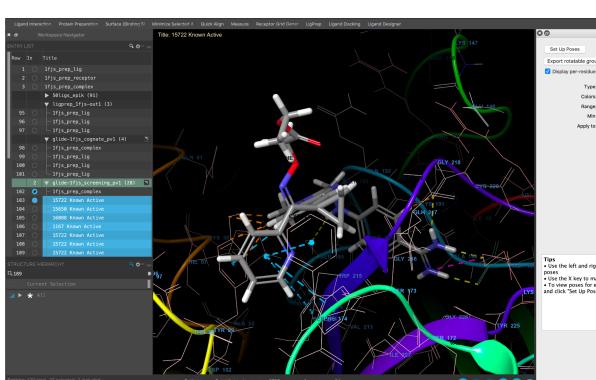


Figure 5-1. Pose Viewer panel.

- Step through the results using the **right** and **left** arrow keys

- Ligand poses are displayed in the **Workspace**
- Residues are colored according to their interaction energies, ranging from green (favorable) to red (unfavorable)

- Close the **Pose Viewer** panel

5.2 Analyze the results

Row	In	Title	docking
1		1fjs_prep_lig	
2		1fjs_prep_receptor	
3		1fjs_prep_complex	
		► 50ligs_epik (92)	
		► ligprep_1FJS-out1 (3)	
		► glide_1FJS-cognate_pvi (4)	
103	2	glide_1FJS_screening_pvi (2_	
104		1fjs_prep_complex	
105		– 15650 Known Active	
106		– 15722 Known Active	
107		– 1167 Known Active	
108		– 16088 Known Active	
109		– 15722 Known Active	

Figure 5-2. Glide Primary properties shown in the Project Table.

- In the **Project Table**, click the **Property Tree** icon

- The Property Tree appears on the right of the **Project Table**

- Click the **All** box twice

- All boxes are deselected

- Click the **Glide** box

- Click **Primary**

- Only the Glide Primary properties are shown

Note: Please see [Knowledge Base Article 1027](#) for more information on the difference between docking score, Glide gscore, and glide emodel score.

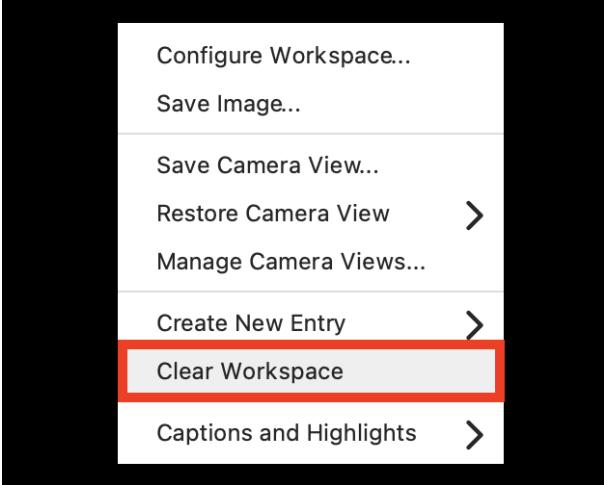
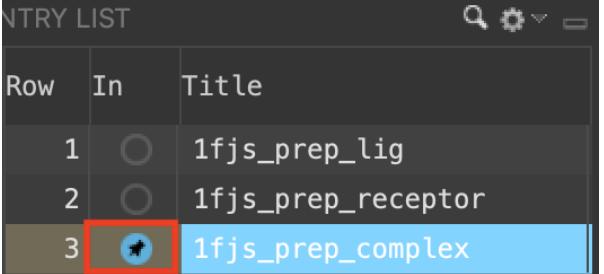
Question #4:

In the table below, select your top 5 ligands with the best docking scores. Remember that the lower the value, the better the docking score. For each ligand, i) write the ligand name and take a screenshot of the pose, ii) list the docking score that can be found in the Project Table, and iii) identify at least 2 types of protein-ligand interactions between that particular ligand and residues within the active site.

Ligand Name & a Screenshot of the Pose	Docking Score (found in Project Table)	Protein-Ligand Interactions

5.3 Identify a binding site with SiteMap

We will analyze the binding site using SiteMap. SiteMap characterizes hydrophilic, hydrophobic, acceptor, and donor regions of a receptor. This is useful for learning more about an active site, predicting a binding site in an apo structure, or identifying possible allosteric sites. SiteMap ranks the potential binding sites with a druggability score, which can be viewed in the [Project Table](#). The output from a Glide virtual screen can be overlaid with SiteMap information to examine how well the docked ligands explore the various regions in the binding cavity. Sites identified by SiteMap can also be used to create receptor grids for virtual screening experiments. This can be useful for exploring sites without a known active compound.

	<ol style="list-style-type: none">1. Right-click an empty area in the Workspace choose Clear Workspace
 <p>Figure 5-4. Fix <i>1fjs_prep_complex</i> in the Workspace.</p>	<ol style="list-style-type: none">2. Double-click the In circle to fix 1fjs_prep_complex (in row 3) in the Entry List3. Go to Tasks > Browse > Structure Analysis > Binding Site Detection<ul style="list-style-type: none">o The SiteMap panel opens

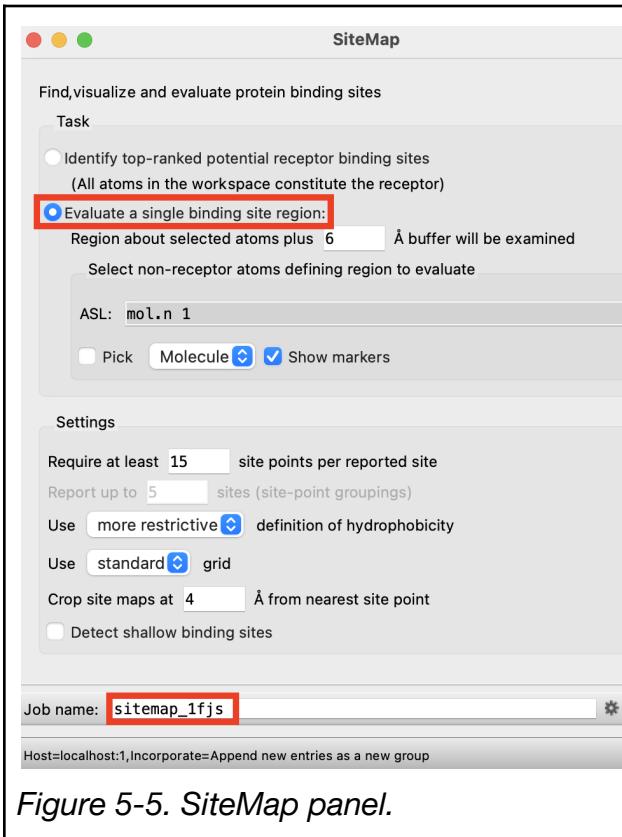


Figure 5-5. SiteMap panel.

4. Under Task, select **Evaluate a Single binding site region**
5. Click on the **ligand** in the Workspace
 - The ligand is highlighted
 - SiteMap removes the ligand from the calculation
6. Change the Job name to **sitemap_1fjs**
7. Click **Run**
 - A banner appears when the job has incorporated
 - A new group is added to the Entry List

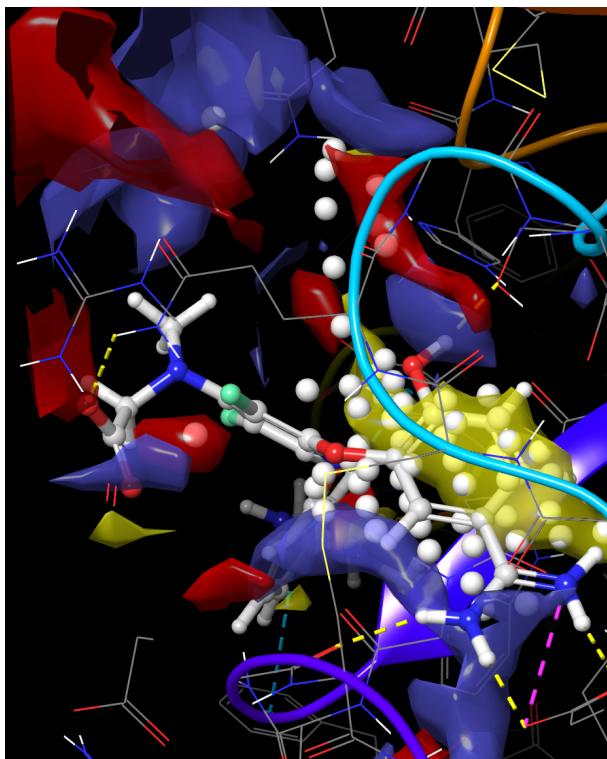


Figure 5-6. SiteMap results in the Workspace.

8. Include the **sitemap_1fjs_site_1**, **sitemap_1fjs_ligand**, and **1fjs_protein**

9. Type **L**

- Various surfaces are shown representing different regions of hydrophilic property; hydrophobic (yellow), acceptor (red), donor (blue)
- The white site-point spheres each represent ~1 Å³

10. In the Entry List, click the **S** next to **sitemap_1fjs_site1** to **toggle** the surfaces associated with the SiteMap

Note: To find all possible binding sites using SiteMap, under Task select **Identify top-ranked potential receptor binding sites**. If you want to detect all pockets, you will need to exclude the cognate ligand from the Workspace.

Question #5:

SiteMap visualization uses a grid of points to identify potential hydrophobic and hydrophilic regions; the hydrophilic regions are further classified into hydrogen-bond donor, hydrogen-bond acceptor, and metal-binding regions, and the surface of the protein is contoured. Take a screenshot of your SiteMap results. Identify which regions of your receptor are hydrophilic and hydrophobic.

6. Individual Exercise

Using the information that you gained from docking the screening ligands, design a new inhibitor that may have a better docking score. Perform LigPrep on your molecule and use Glide to obtain its docking score. Take a screenshot of its pose and paste it below. Then list its docking score. Provide analysis as to why you chose to design this particular inhibitor.

7. Summary, Additional Resources, and References

In this lesson, we completed a workflow for virtual screening using Glide. We generated a receptor grid with a hydrogen bond constraint, which was used in cognate ligand docking as a positive control to set up a virtual screen of test ligands. Then, a series of screening compounds were docked and the results were viewed using Pose Viewer, with known actives being found as the top hits. SiteMap was used to explore the binding site.

For further information, please see:

[Maestro 11 Training Portal](#)

[Introduction to Structure Preparation and Visualization](#)

[Glide User Manual](#)

Glossary of Terms

cognate ligand - a ligand that is bound to its protein target

Entry List - a simplified view of the Project Table that allows you to perform basic operations such as selection and inclusion

included - the entry is represented in the Workspace, the circle in the In column is blue

incorporated - once a job is finished, output files from the working directory are added to the project and shown in the Entry List and Project Table

Project Table - displays the contents of a project and is also an interface for performing operations on selected entries, viewing properties, and organizing structures and data

Scratch Project - a temporary project in which work is not saved, closing a scratch project removes all current work and begins a new scratch project

selected - (1) the atoms are chosen in the Workspace. These atoms are referred to as "the selection" or "the atom selection". Workspace operations are performed on the selected atoms. (2) The entry is chosen in the Entry List (and Project Table) and the row for the entry is highlighted. Project operations are performed on all selected entries

Working Directory - the location that files are saved

Workspace - the 3D display area in the center of the main window, where molecular structures are displayed