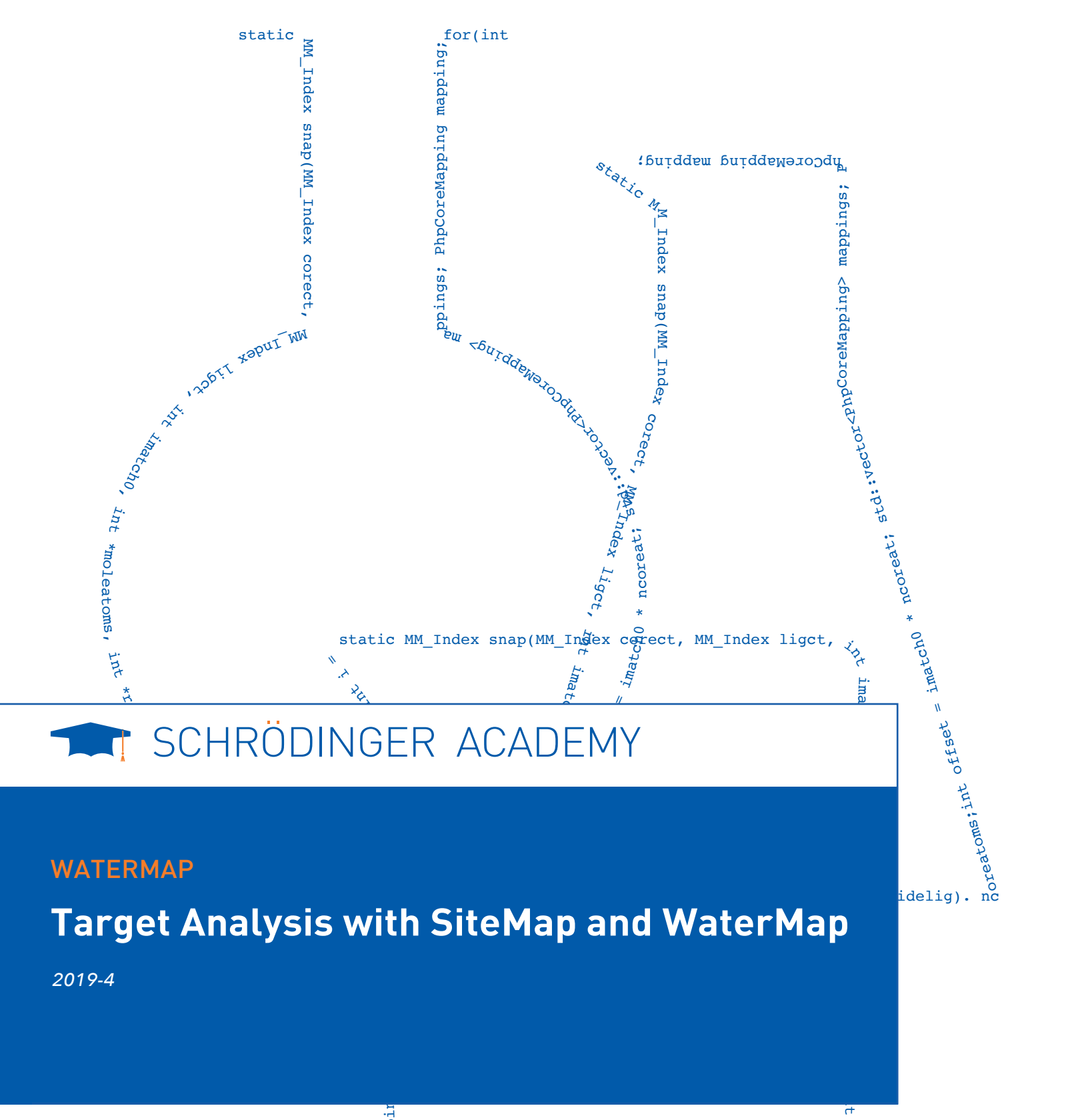


SCHRÖDINGER ACADEMY

WATERMAP

Target Analysis with SiteMap and WaterMap

2019-4

Target Analysis with SiteMap and WaterMap

Created with: Release 2019-4

Prerequisites: Release 2019-4

Categories: Structure-Based Design

Keywords: WaterMap, SiteMap, target analysis

In this tutorial, you will learn how to perform a standard target analysis on a protein-ligand complex using the Protein Reliability Report, SiteMap, and WaterMap.

Words found in the Glossary of Terms are shown like this: Workspace
File names are shown with the extension like this: `1fjs.pdb`
Items that you click or type are shown like this: **File > Import Structures**

This tutorial is written for use of a 3-button mouse with a scroll wheel.

This tutorial consists of the following sections:

1. Creating Projects and Importing Structures

At the start of the session, change the file path to your chosen Working Directory in Maestro to make file navigation easier. Each session in Maestro begins with a default Scratch Project, which is not saved. A Maestro project stores all your data and has a `.prj` extension. A project may contain numerous entries corresponding to imported structures, as well as the output of modeling-related tasks. Once a project is created, the project is automatically saved each time a change is made.

Structures can be imported from the PDB directly, or from your Working Directory using **File > Import Structures**, and are added to the Entry List and Project Table. The Entry List is located to the left of the Workspace. The Project Table can be accessed by **Ctrl+T (Cmd+T)** or **Window > Project Table** if you would like to see an expanded view of your project data.

1. Double-click the **Maestro** icon
    - (No icon? See Starting Maestro)

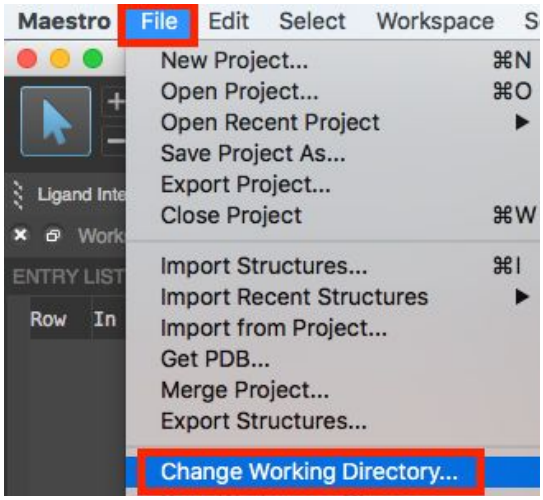


Figure 1-1. Change Working Directory option.

2. Go to **File > Change Working Directory**, find your directory, and click **Choose**

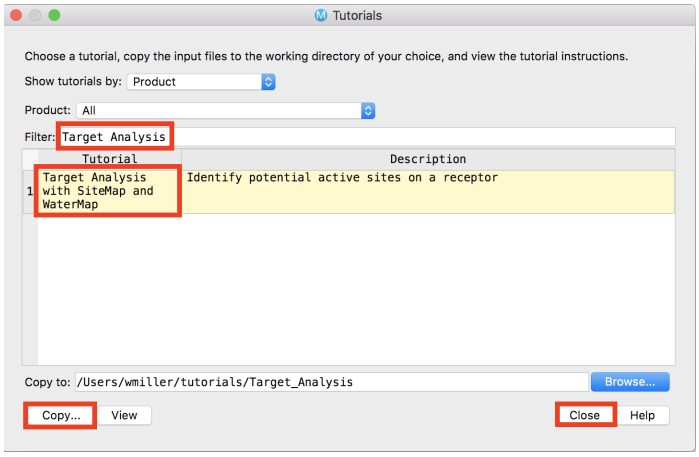


Figure 1-2. Tutorials panel, showing filtered results with "Target Analysis" as a keyword.

3. Go to **Help > Tutorials**
4. Next to Filter, type **Target Analysis**
5. Choose **Target Analysis with SiteMap and WaterMap**
6. Click **Copy**
    - All the tutorial files are copied into your Working Directory
7. Click **Close**

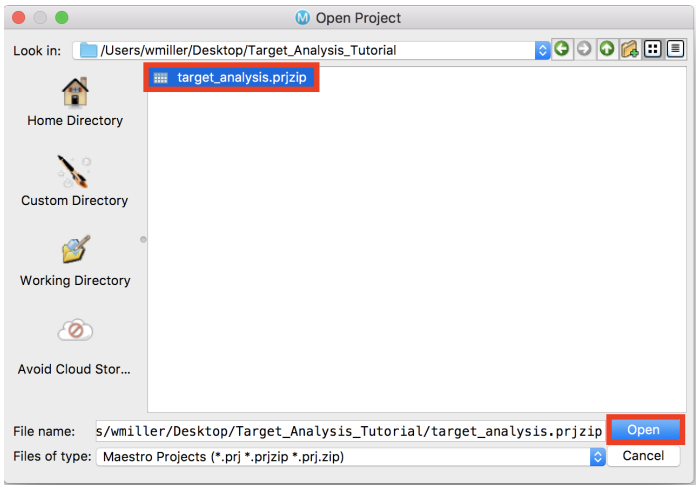


Figure 1-3. Open Project.

8. Go to **File > Open Project** and choose `target_analysis.prjzip`
9. Click **Open**

Note: By default the structure corresponding to the imported file is both included in the Workspace and selected in the Entry List. Please refer to the Glossary of Terms for the difference between included and selected.

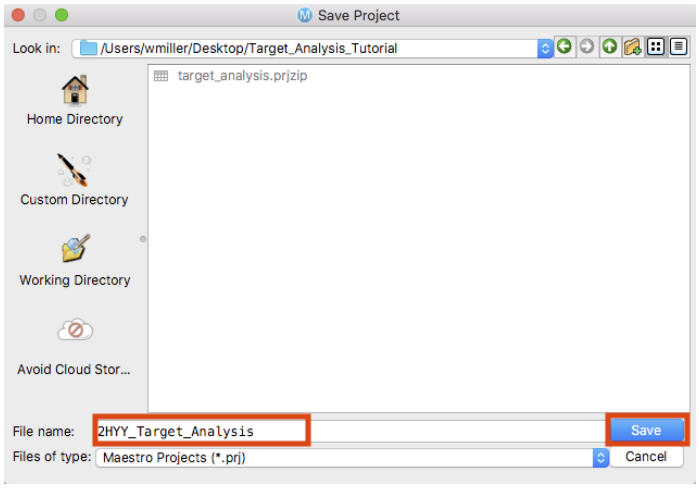


Figure 1-4. Save Project panel.

10. Go to **File > Save Project As**
11. Change the File name to **2HYY_Target_Analysis**, click **Save**
    - The project is now named `2HYY_Target_Analysis.prj`

2. Determining Structure Quality with the Protein Reliability Report

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 Wizard to resolve issues.

Protein reliability is used to determine the amount of protein preparation required prior to use for modeling. While X-ray resolution is generally not a reliable indicator of model quality, values greater than 3Å are often poor models. It is recommended to examine all of the results in the report, and not just focus on the number of green and red bubbles. The factors that contribute the most to structure quality for eventual modeling enablement are the presence of a ligand (and the availability of electron density files), the fit of the ligand and binding site residues to density, missing loops, and missing side chains.

The structure that we will be working with, 2HYY, is ABL Kinase bound with imatinib (Gleevec). Since the structure is a homotetramer, we will only be focusing on one of the chains, Chain A.

2.1 Assess the Quality of the Unprepared Protein

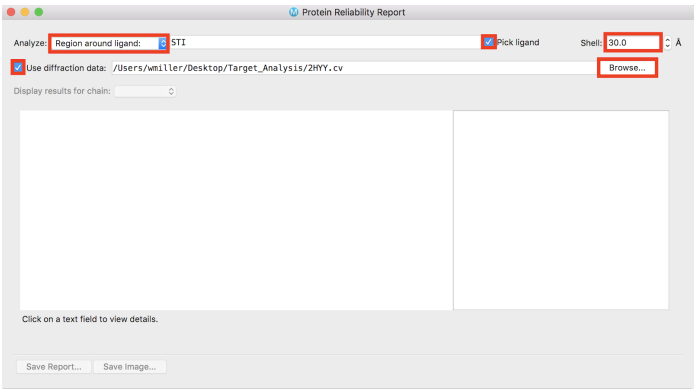


Figure 2-1. Protein Reliability Report panel.

1. Include **2HYY** in the Workspace
2. Go to **Tasks > Browse > Structure Analysis > Protein Reliability Report**
    - The Protein Reliability Report panel opens
3. For Analyze, **Entire structure**
4. Check **Pick Ligand**
5. Select one of the ligands in the Workspace
    - By default, a 10 Å region around each ligand is highlighted
6. Set Shell to **30 Å**
    - This will allow for the quality of the entire protein to be assessed
7. Check **Use diffraction data**
8. Click **Browse**
9. Find `2HYY.cv` and click **Open**

Note: The .cv file was obtained by checking the diffraction data option when importing a protein structure in the Protein Preparation Wizard

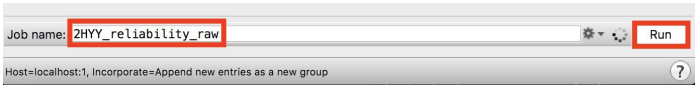


Figure 2-2. Run the Protein Reliability job for unprepared structure.

10. Change Job name to **2HYY_reliability_raw**
11. Click **Run**
    - The job should take ~3 minutes
    - A new entry has been added to the Entry List
    - A banner appears

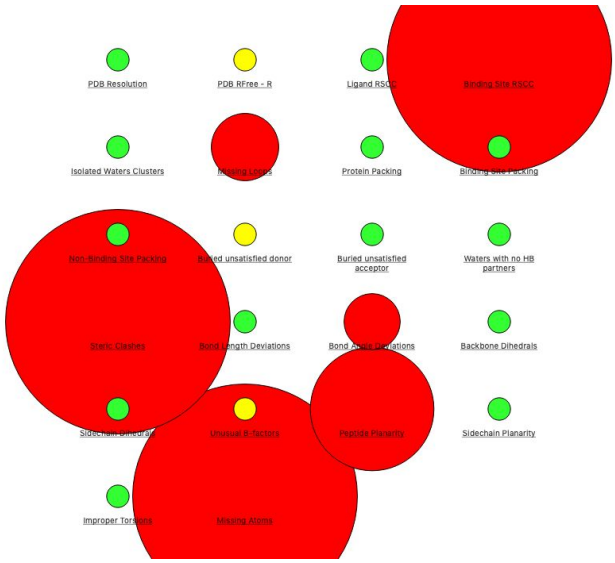


Figure 2-3. Protein Reliability Report of the unprepared protein.

5. On the banner, select **Reliability Report**
   - The Protein Reliability Report panel opens and is populated
6. For Display results for chain, choose **A**

Note: The results can be examined more closely by clicking on the bubbles

Note: There are serious issues with steric clashes, binding site RSCC (real space correlation coefficient) and missing atoms, as is often the case with an unprepared crystal structure.

2.2 Assess the Quality of the Prepared Protein

To save time, the protein has already been prepared. Aside from the standard preparation protocol that can be found in the Introduction to Structure Preparation and Visualization tutorial, “Fill in missing side chains using Prime” was used due to missing side chains in the crystal structure, and all crystallographic waters were kept.

During the preparation stage, it is advisable to retain all crystallographic waters (as with the preparation of all systems for running of Molecular Dynamics simulations). In the WaterMap part of the calculation, you have the option to specify that existing waters should be treated as solvent. This allows you to start with those waters in the right place, while fully solvating the rest of the system. With this setting, both the crystal waters and the other solvent molecules will be treated in the same way and both will be allowed to move freely during the simulation. If, however, there are any doubts about one or more waters, it is advisable to delete them, and allow WaterMap to place them. For apo WaterMap jobs, it is important to note that some of the crystallographic waters are where they are due to interactions with the ligand, which is not present during the simulation. In such a case, starting with a dry complex allows the system setup step in WaterMap to fill the ligand void as needed. With a holo WaterMap, the recommended way to determine whether crystallographic waters interacting with the ligand should be kept is to see if such interactions are supported by electron density.

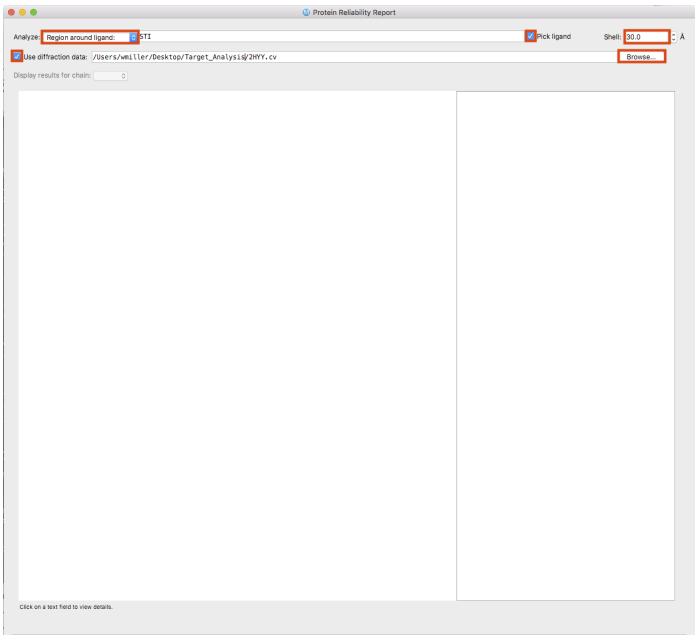


Figure 2-4. Set up Protein Reliability job for prepared structure.

1. Include **2HYY - minimized** in the Workspace
2. Open the Protein Reliability Report panel
3. For Analyze, choose **Region around ligand**
4. Check **Pick Ligand**
5. Select the ligand in the Workspace
6. Set Shell to **30 Å**
    - This will allow for the quality of the entire protein to be assessed
7. Check **Use diffraction data**
8. Click **Browse**
9. Find `2HYY.cv` and click **Open**

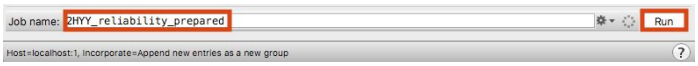


Figure 2-5. Run Protein Reliability job for prepared structure.

10. Change Job name to **2HYY_reliability_prepared**
11. Click **Run**
    - The job should take ~2 minutes
    - A new entry has been added to the Entry List
    - A banner appears

Figure 2-6. Protein Reliability Report of the prepared protein.

12. On the banner, select **Reliability Report**
    - The Protein Reliability Report panel opens and is populated

Note: In the event of a red Binding Site RSCC bubble, it is recommended to examine the size and results closely, and evaluate the residues contributing to the poor score.

Note: The results can be examined more closely by clicking on the bubbles

Note: Protein preparation resolved the issues with steric clashes, missing atoms, and side-chain dihedrals

3. Evaluating Binding Sites with SiteMap

SiteMap can be used to identify potential binding sites on an apo protein, or evaluate a single site of selected residues or a ligand. For more information on SiteMap, please see Overview of SiteMap. When performing target analysis on a ligand-bound structure, it is a best practice to run SiteMap in both modes and look for similar results for the desired site. When analyzing SiteMap results, it is recommended to look for a SiteScore and DScore > 1, a balance > 0.3 and a volume > 225. See How SiteMap Evaluates Sites for more information on the listed properties.

3.1 Evaluate Top-Ranked Binding Sites

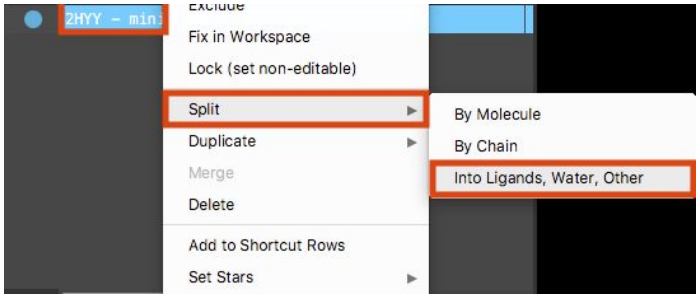
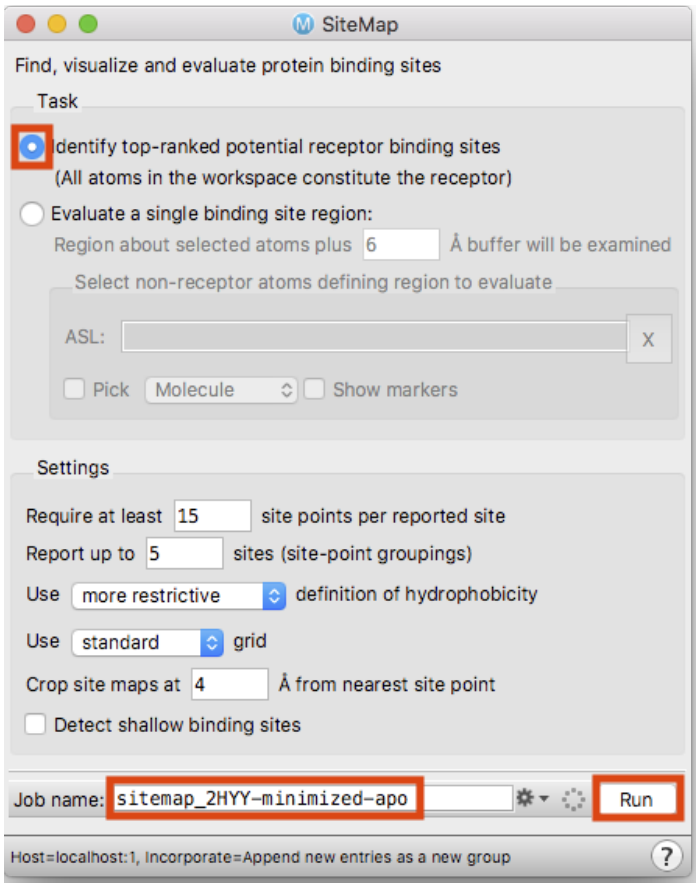


Figure 3-1. Create duplicate entry of 2HYY - minimized.

1. Right-click on **2HYY - minimized** in the Entry List
2. Go to **Split > Into Ligands, Water, Other**
   - A new group is added to the Entry List
3. Double click on 2HYY - minimized_protein to rename it **2HYY - minimized_apo**
4. Include **2HYY - minimized_apo** in the Workspace



5. Go to **Tasks > Browse > Structure Analysis > Binding Site Detection**
   - The SiteMap panel opens
6. For Task, choose **Identify top-ranked potential receptor binding site**
7. Change Job name to **sitemap-2HYY-minimized-apo**
   - The job takes ~1 minute
   - A new group is added to the Entry List

Figure 3-2. Run SiteMap on apo-protein.

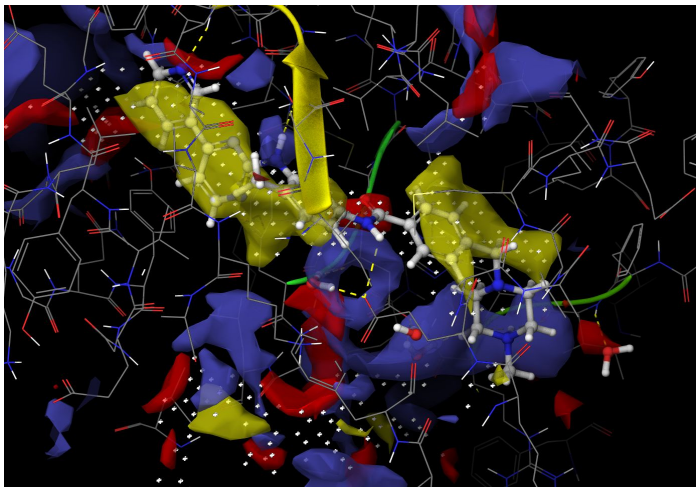


Figure 3-3. Top-ranked SiteMap site overlayed with 2HYY imatinib binding site.

8. Double-click **2HYY - minimized** to fix it in the Workspace
9. Include **sitemap_2HYY-minimized-apo_site_1** in the Workspace
    - The SiteMap surface aligns well well with the imatinib binding site

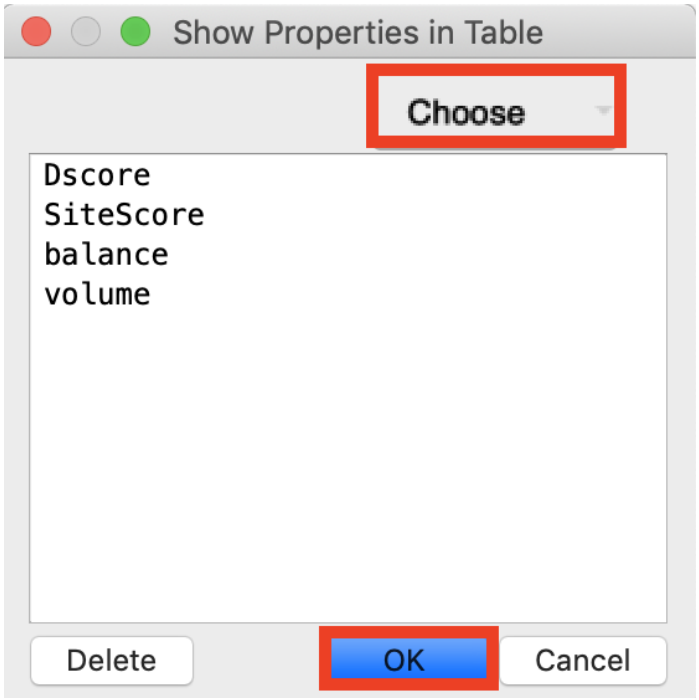


Figure 3-4. Add properties to the Entry List.

10. At the top right corner of the Entry List, click the **Settings button (cog)**
11. Select **Show Property**
    - The Show Property panel opens
12. Click **Choose**
13. Type and select **DScore**
14. Add the **SiteScore**, **balance**, and **volume** properties in the same way
15. Click **OK**

Figure 3-5. Properties added to the Entry List.

Note: The top site, which corresponds with the imatinib binding site, has a DScore, SiteScore, balance, and volume in the desired ranges

3.2 Evaluate a Single Binding Site

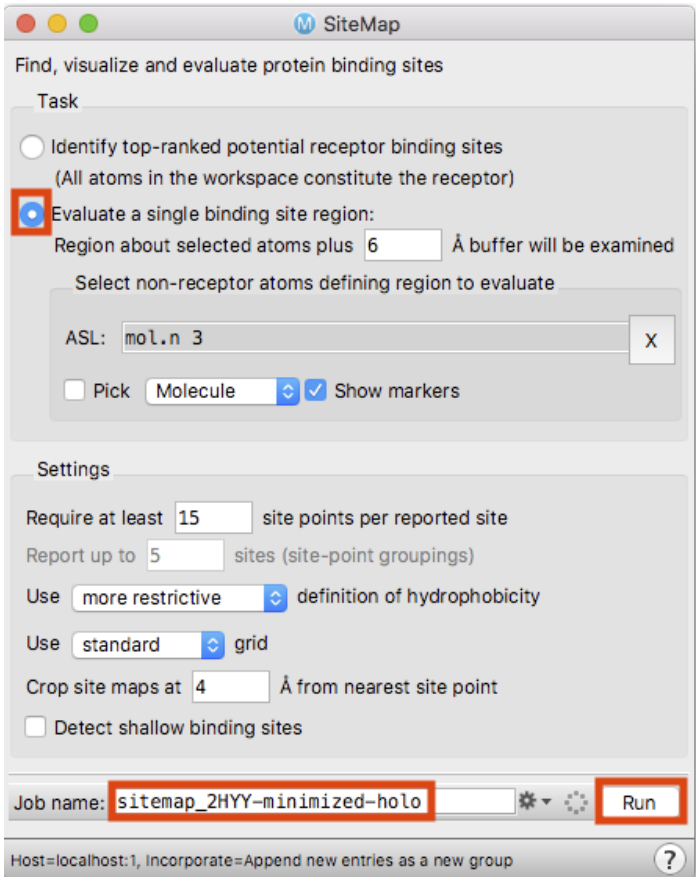


Figure 3-6. Run SiteMap on holo-protein.

1. Include **2HYY - minimized** in the Workspace
2. Open the **SiteMap** panel
3. Choose **Evaluate a single binding site region**
    - A banner appears
4. Select an atom in the **ligand** to define the binding site
5. Change Job name to **sitemap_2HYY-minimized-holo**
6. Click **Run**
    - The job takes ~1 minute
    - A new group is added to the Entry List

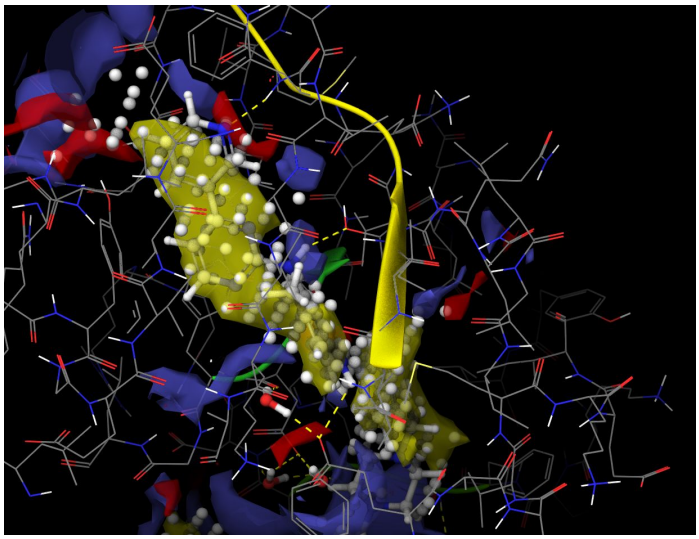


Figure 3-7. SiteMap results overlayed with the ligand and protein.

7. Right-click on the **sitemap_2HYY-minimized-holo_out1** group and choose **Include**
    - The SiteMap results, ligand, and protein are all included in the Workspace

Note: The SiteMap surface aligns well with the imatinib binding pose, and the DScore, SiteScore, balance, and volume values are all in the desired ranges

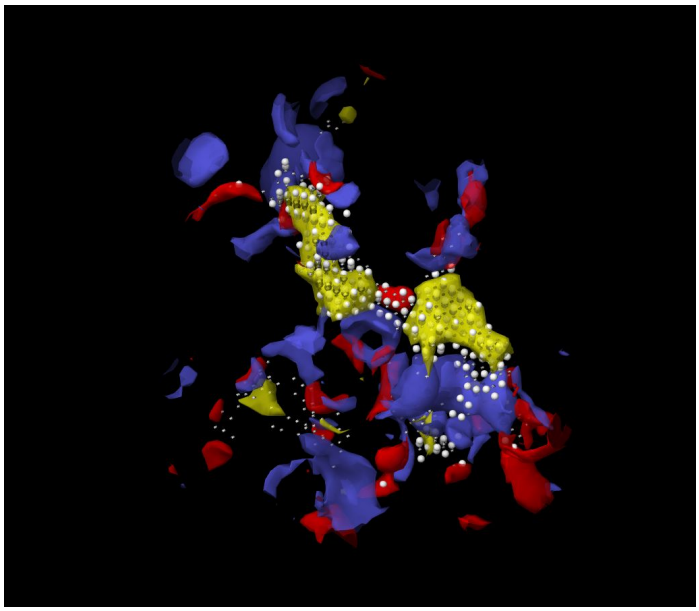


Figure 3-8. Overlay of the top-ranked apo-structure site with the holo-structure site.

8. Right-click and choose **Exclude** to remove the protein and ligand from the Workspace
9. Use Ctrl+click (or Cmd+click) to include **sitemap_2HYY-minimized-apo_site_1**
    - The two maps are now overlayed

Note: The map generated from the apo-structure has a greater volume than the holo-structure map, but both have very strong agreement with the imatinib binding site

4. Assessing Druggability with WaterMap

WaterMap can be used to assess the druggability of a binding site by determining its hydration-site thermodynamics.

The preparation of a system is extremely important prior to performing a WaterMap calculation and can have a profound influence on the quality of results that emerge from it. Any unusual residue charges or incorrectly flipped residues can drastically alter the hydration structure seen in the final water map.

It is a best practice to always run WaterMap on both the apo- and holo-structures. If there are any apo hydration sites that have the same or higher ΔG in the holo map, those sites should be displaced. Additionally, you should consider displacing hydration sites that have high TΔS, but low ΔH.

4.1 Set Up and Run WaterMap Jobs

1. Include **2HYY - minimized** in the Workspace
2. Go to **Tasks > Browse > WaterMap > Perform Calculations**
    - The WaterMap - Perform Calculation panel opens
    - A banner appears

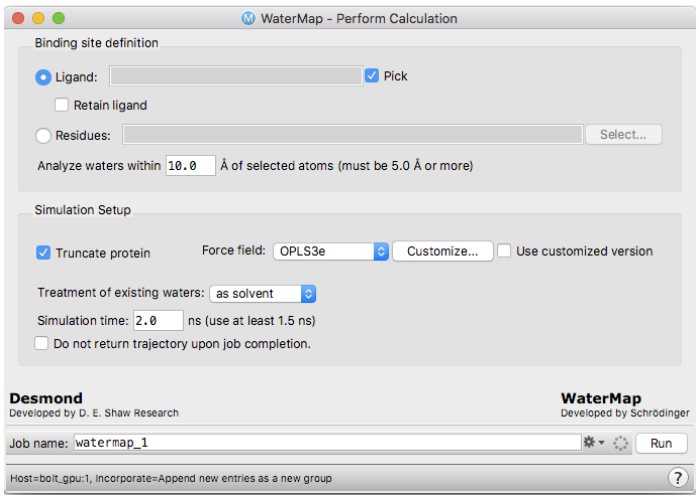


Figure 4-1. WaterMap Perform Calculation panel.

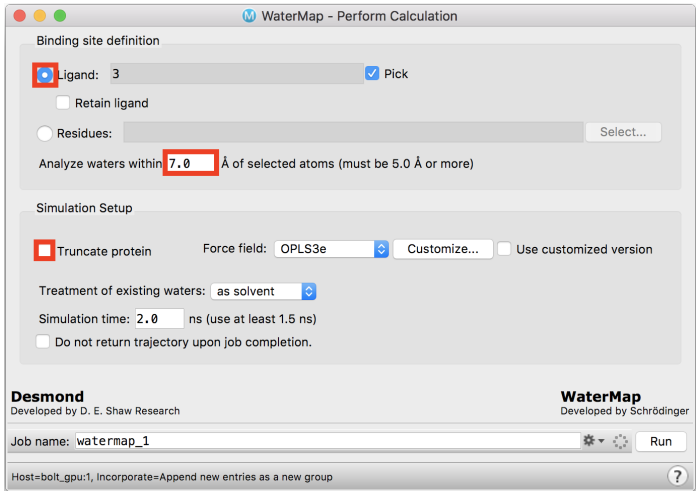


Figure 4-2. Setup WaterMap job.

3. Select an atom in the **Ligand**

Note: Since Retain ligand was not checked, the ligand will be deleted prior to the WaterMap calculation

4. Set Analyze waters to within **7.0 Å**

Note: Unless it is a large binding site, 7.0 Å is the recommended threshold.

5. Uncheck **Truncate protein**
    - Truncation can produce misleading results if it goes through a critical part of the protein.

Note: It is recommended to only consider truncation if working with a very large system and time is a concern

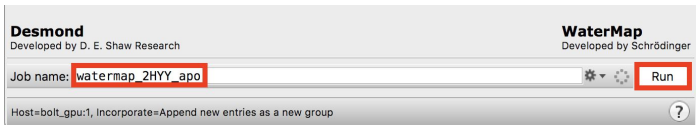


Figure 4-3. Run apo WaterMap job.

6. Change Job name to **watermap_2HYY_apo**
7. Click **Run**
    - This job requires significant GPU resources to run, so we will look at pregenerated results

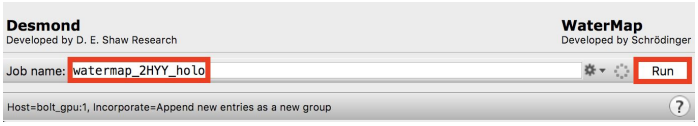


Figure 4-4. Run holo WaterMap job.

8. Check Retain ligand
9. Change Job name to **watermap_2HYY_holo**
10. Click **Run**
    - This job requires significant GPU resources to run, so we will look at pre-generated results

4.2 Analyze apo and holo Hydration Sites

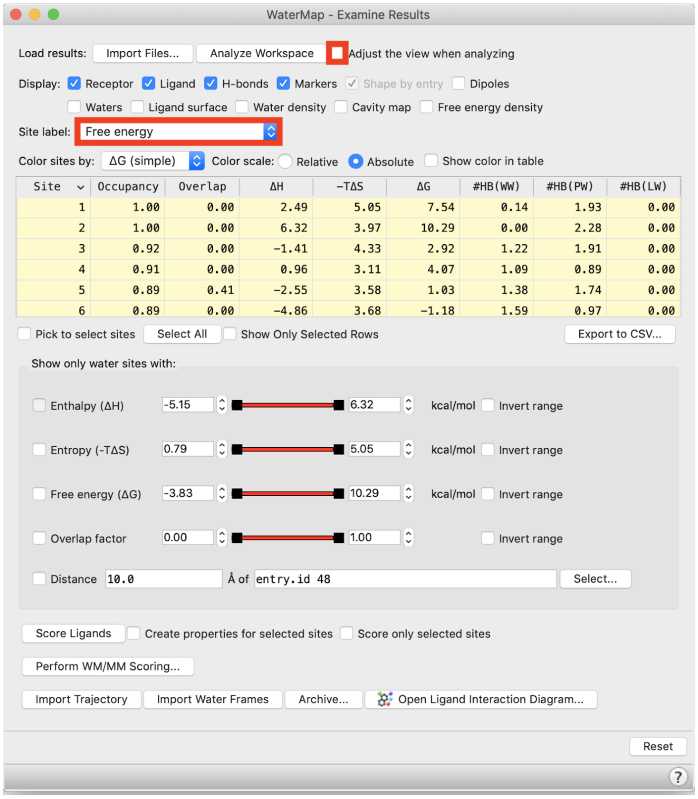
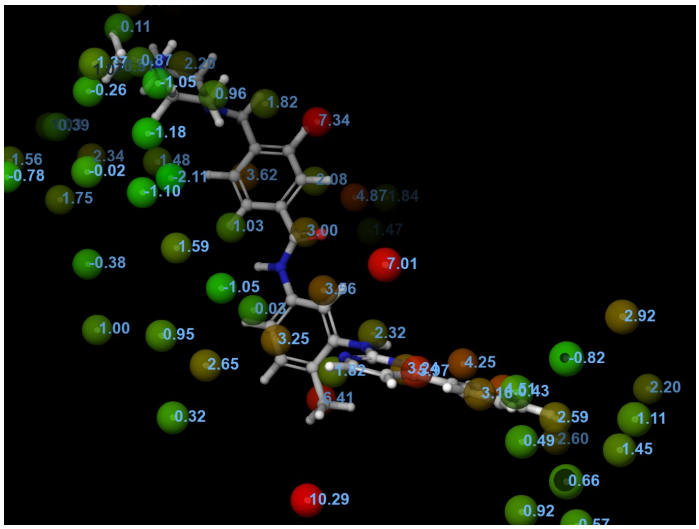
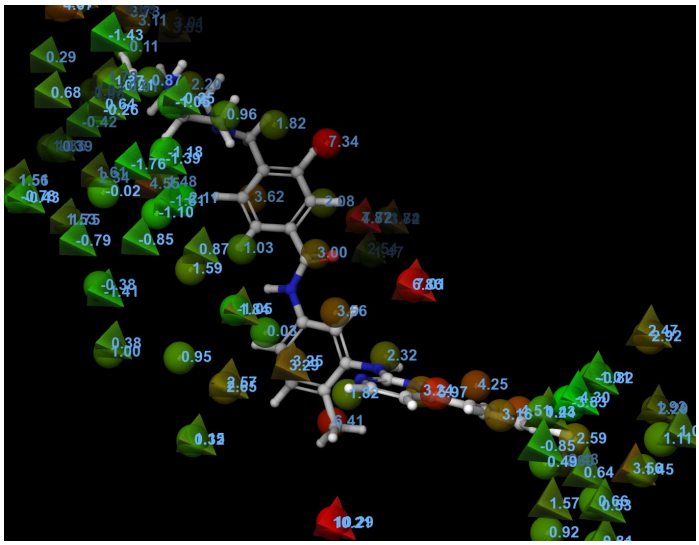


Figure 4-5. WaterMap - Examine Results panel.

1. In the Entry List select and expand **watermap_2HYY_apo_wm**
2. Click the **W** icon next to the WaterMap entry
    - The WaterMap-Examine panel, populated with data, opens and the water map is displayed in the Workspace along with the ligand and receptor
3. For Site label, choose **Free energy**
4. Uncheck **Adjust the view when analyzing**



Note: The apo water map shows a cluster of high energy hydration sites that map out a shape similar to that of Gleevec. The shape of highly unstable hydration sites could give the indication of the shape of potentially potent inhibitors.

Figure 4-6. The apo 2HYY water map overlaid with the protein and ligand.	
 *Figure 4-7. The apo (spherical) and holo (pyramidal) 2HYY water maps overlaid with the protein and ligand*	5. In the Entry List, expand the groups **watermap_2HYY_apo_wm** and **watermap_2HYY_holo_wm** 6. Fix the entries for the **water maps** in both the groups 7. Include all the remaining entries in these groups; **Ctrl+click** on the **In** columns 8. In the WaterMap - Examine Results panel click **Analyze Workspace** 9. Type **L** to zoom to the ligand and inspect the hydration sites in the binding site region ○ The apo water map shows a cluster of high energy hydration sites in the binding site region that are highly unstable. In the holo water map almost all those high energy waters are shown to be displaced by the ligand ○ This implies that a drug-sized molecule (that can be accommodated in this binding site), will receive significant binding from simply occupying the binding site, thus confirming the target's druggability.

4.3 Visualize the Water Map of an Undruggable System (Optional)

In this example we will look at the WaterMap output for PTP-1b (PDB:2QBS), a target considered highly undruggable, to predict the undruggability of the binding pocket. The WaterMap job for the apo system has already been run for convenience.

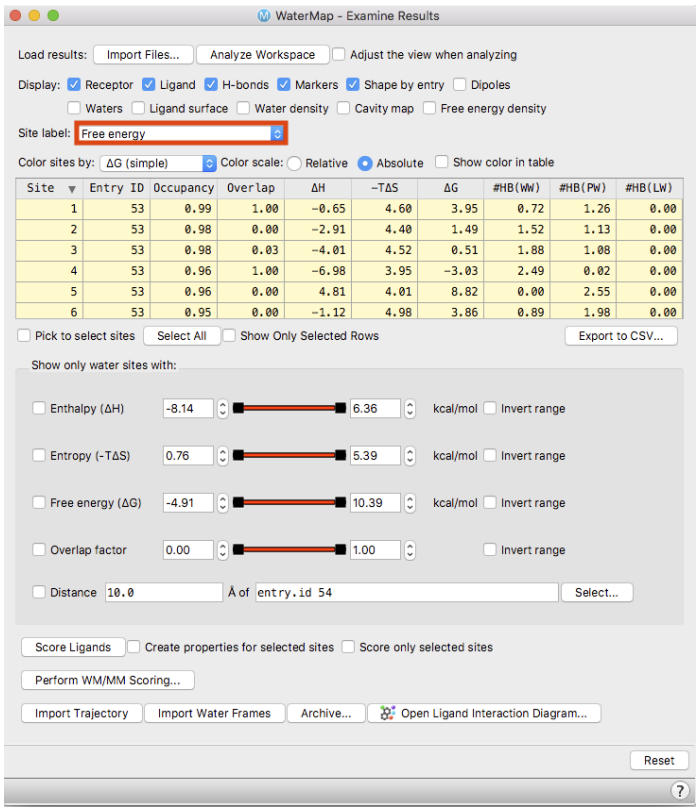


Figure 4-8. WaterMap - Examine Results panel.

1. In the Entry List select and expand **watermap_2BQS_apo_wm**
2. Click the **W** icon next to the WaterMap entry
    - The WaterMap-Examine panel, populated with data, opens and the water map is displayed in the Workspace along with the ligand and receptor
3. For Site label, choose **Free energy**

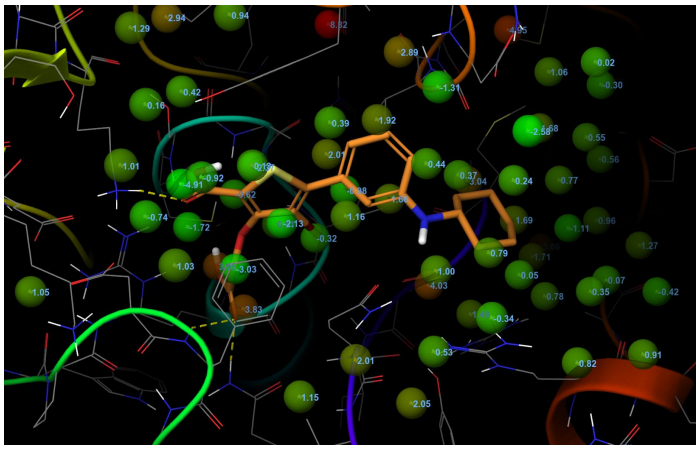


Figure 4-9. The apo 2QBS water map overlaid with the protein and ligand.

4. Type **L** to zoom to the ligand and inspect the hydration sites in the binding site region
    - The core of the binding site contains a large cluster of low energy and stable water molecules. Binding sites with few unstable hydration sites are likely to be either too shallow or too polar to bind a drug-like molecule and this naturally limits the drug-likeness of any ligand
    - The hydrophobic atoms in this region are actually detrimental to binding and only the powerful ionic atoms are capable of replacing some of these water molecules. There are scattered hydrophobic regions away from the main binding site, but reaching these regions requires large and inefficient ligands

5. Conclusion and References

In this tutorial, we used three tools as part of a target analysis workflow.

The protein reliability report serves as a good initial check of protein quality at the start of a structure-based project. It is designed to pictorially highlight a range of issues that will subsequently require more careful consideration during the preparation steps. These range from steric clashes and missing atoms which can be easily resolved, to unsatisfied buried groups and waters lacking hydrogen-bonding partners that should be kept in mind when designing ligands for the site.

Following careful preparation, both SiteMap and WaterMap were used to assess the requirements of the binding site. SiteMap gives you a good sense of the shape, size and composition of your target's primary binding site, as well as the existence of potentially druggable allosteric sites. The generated surfaces represent volumes in the binding site that can be occupied and should be considered during the design process. In this case, both the apo and holo maps had very strong agreement with the imatinib site. WaterMap was also run on both the apo and holo structures where high energy hydration sites in the apo-map were shown to be displaced by the ligand in the holo-map. The shape of the hydration sites in the holo-map was also a good indication of the shape of the potent inhibitor Gleevec.

For further information, please see:

WaterMap User Manual

Introduction to Structure Preparation and Visualization

Identifying Binding Site Requirements and Hit-to-Lead with WaterMap

6. Glossary of Terms

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

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