

# Hands-On Exercises: Docking and Scoring

## Lesson 1: Docking Potential Inhibitors to a Tyrosine Kinase

In docking of small molecules to potential receptors, there are several aspects that should be considered. As discussed in the lecture, we may have either rigid or flexible ligands. The later approach will take more time but will more likely lead to better results. Also, we must bear in mind the distinction between ligand poses and the scoring of the poses. We can have many poses, positions and orientations of the ligand in the binding site, and these are ranked by a scoring function. The methods by which poses are determined can be separated from how the poses are scored.

In this exercise, we will use DS LigandFit (11) to dock ligands to the c-Abl protein kinase catalytic domain examined in an earlier exercise. The native c-Abl is an 1150 residue protein (8). The fusion of the gene for c-Abl with the *bcr* gene results in a constitutive active protein kinase. The result is the manifestation of disease known as chronic myelogenous leukemia (CML) (9). The treatment of CML through protein kinase inhibitors has been an active area of research for several years culminating in the development of STI-571, also known as Gleevac and Imatinib (1). This compound was identified in 1996 by Novartis (2) and has been approved by the U.S. Food and Drug Administration for the treatment of CML.

In this exercise, you will:

- Open the protein you edited in the previous lesson,
- Define the receptor and search for binding sites,
- Edit binding sites,
- Specify the input ligands for docking,
- Perform a rigid fit,
- Perform a flexible fit,
- View the results, and
- Modify the experiment.

### 1. Open the receptor protein

If Discovery Studio is currently open, close all open Windows.

If the program is not open, launch it.

We will open the protein data file taken from the PDB file 1M52 (8) and processed in the earlier exercise.

From the **File** menu, choose **Open**. Navigate to the *DataFiles* folder and select the file *protein.msv*. Click **Open** to load the protein into a new 3D Window. Ensure the **Hierarchy View** is also visible

**Note:** You may have to drag the 3D View to adjust it in order to see the Hierarchy View better.

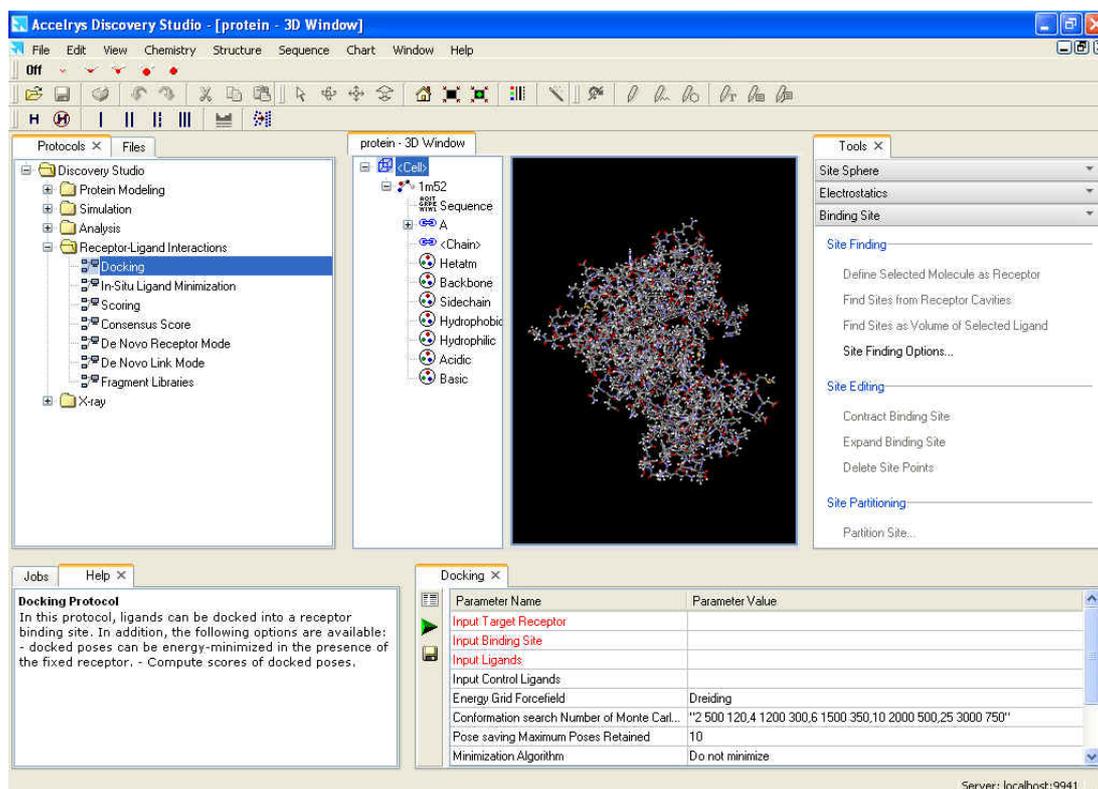
The Abl protein appears in a new 3D View. If necessary, click the **Fit to Screen** button to center the molecule.

## 2. Define the receptor and search for binding sites

Before beginning the docking operation, it is necessary to specify a binding site. DS LigandFit uses a method based on protein shape searching for cavities or holes. Often, the largest identified cavity is or is part of the ligand binding site (6, 7).

**Note:** It is highly recommended that any protein checking and preparation be completed before the target receptor is selected. Ensure that the structure you want to define as the receptor is open in the 3D View. The receptor must have hydrogens added and all atom valencies satisfied so its atoms can be properly typed.

In the **Protocols Explorer**, double click the **Receptor Ligand Interactions** folder to expand it, and then double click **Docking**. The Docking Parameters Explorer appears at the bottom right of the interface (Figure 1).



**Figure 1: The Docking interface layout in Discovery Studio**

The Parameters Explorer is used to set up a particular protocol, in this case a docking calculation. You will need to enter appropriate values for the required parameters which are colored in red. The Parameter Help Explorer will provide guidance on appropriate values for a given parameter when it is selected.

The Tools Explorer in the top right of the interface relates to the current Protocol and allows for quick access to common actions. The Tools Explorer is split up into one or more panels, each a group of logically connected actions. Each Tools Explorer panel is made up of one or more action buttons that automatically enable or disable themselves based on the currently selected view and the current workflow.

In the **Tools Explorer**, ensure the **Binding Site Tool Panel** is displayed. If the Binding Site tool is not displayed, activate its display by selecting **View | Tools Panel | Binding Site** from the menu bar.

In the Hierarchy View click on the *1m52* component to select the whole protein (highlighted in yellow in the 3D View).

This will enable the action buttons in the Binding Site Tool Panel. Click **Define Selected Molecule as Receptor**. An **SBD\_Receptor** component is added to the bottom of the Hierarchy View list.

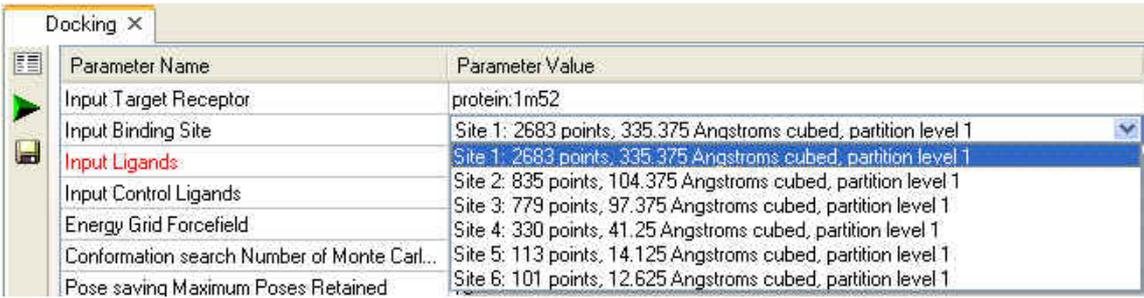
In the **Parameters Explorer**, click on the **Input Target Receptor** parameter. Doing so will automatically enter a value *protein:1m52*.

In the **Site Finding** subsection of the Tool Panel click on **Find Sites from Receptor Cavities**.

6 sites will be listed in the Hierarchy View. Clicking on any of the Site components in the Hierarchy View will highlight the corresponding site in the 3D Window.

Click on **Input Binding Site** parameter in the **Parameters Explorer**.

After a few seconds, a list of cavities appears identified from the cavity site search. Click under the parameter value for the Input Binding Site parameter. The listing shows the size of the binding site by the number of grid points, in cubic Ångstroms, and the partition level. We will examine a site partition problem in a later optional exercise.



Parameter Name	Parameter Value
Input Target Receptor	protein:1m52
Input Binding Site	Site 1: 2683 points, 335.375 Angstroms cubed, partition level 1
Input Ligands	Site 1: 2683 points, 335.375 Angstroms cubed, partition level 1
Input Control Ligands	Site 2: 835 points, 104.375 Angstroms cubed, partition level 1
Energy Grid Forcefield	Site 3: 779 points, 97.375 Angstroms cubed, partition level 1
Conformation search Number of Monte Carl...	Site 4: 330 points, 41.25 Angstroms cubed, partition level 1
Pose saving Maximum Poses Retained	Site 5: 113 points, 14.125 Angstroms cubed, partition level 1
	Site 6: 101 points, 12.625 Angstroms cubed, partition level 1

**Figure 2: Site Listings in the Input Binding Site Parameter**

**Note:** The default site partitioning is now 1, consistent with Cerius2, rather than 0 as DS Modeling 1.2. Note too that 0 is not an accepted value for site partitioning in DS1.5.

Click an empty space in the 3D View to deselect everything. This makes viewing the sites easier. You can also view the sites in the 3D View by selecting each line one-by-one in the list under the Input Binding Site Parameter value. You will see the sites displayed as green grid points. Select the largest site (Site1) as the specified binding site. To view this site without the protein present, double click on the site in the 3D View to select it, and then go to **View | Visibility | Show Only**. Use the **Fit to Screen** button  to get a better view and use the right mouse button to rotate the site.

Notice that the binding site is composed actually of two regions connected by a thin strand, somewhat like a dumbbell.

You can edit binding sites. One approach is to use the Site Editing functionality in the Tools Panel. Take a moment to examine the effects on Site 1 when you click on the **Contract Binding Site** link under Site Editing section. The outermost points defining the binding site are removed if those points are connected to the "outside" of the receptor. Also, the volume in the list of binding sites is updated.

Conversely, the **Expand Binding Site** link expands the site by a layer of site points if those points are not occupied by the receptor and connected to the "outside" of the receptor. **Delete site** link removes the selected site points or the current site.

Defined sites are objects in and of themselves. The tools can be applied to selected site points, or to the whole site if no site points are selected. In fact, site points can be deleted manually, just like molecules and atoms.

Since the original Site 1 is now changed, we should recalculate it. Go to **View | Visibility | Show All** to view the protein in the 3D View. Select the Protein and click **Define Selected Molecule as Receptor**. Then click the **Find Sites from Receptor Cavities** link.

Examine Site 1 again. Rotate the protein until you can see the two lobes of the binding site and can determine which is the larger lobe.

For binding of a single ligand, the site as defined is simply not appropriate. DS LigandFit will try to fit the ligands to the entire defined site, that is, to both lobes simultaneously. We should edit the site.<sup>1</sup>

Select the binding site by double clicking on it and then go to **View | Visibility | Show Only**.

Only the binding site points are visible now.

From the Tool Palette, activate the **Selection** tool. Select the smaller of the two lobes.

When you are satisfied you made the correct selection, go to **Tools | Binding Site | Site Editing | Delete Site Points**.

The larger of the two lobes remains, which is where the ATP binding site lies on c-Abl.

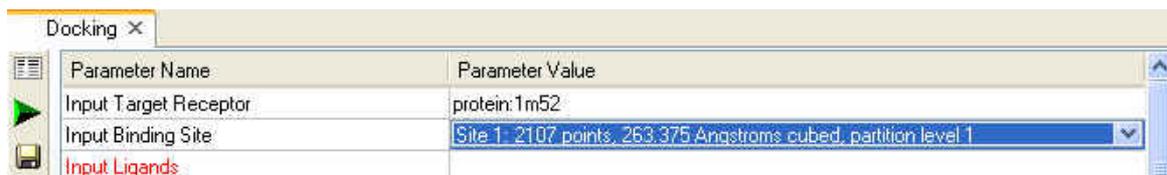
---

<sup>1</sup> In some cases, you may want to use site partitioning and let DS LigandFit search in different parts of the defined binding site. We examine that aspect in a later optional exercise.

The choice and definition of the binding site is a critical step in running DS LigandFit.

To view the protein again, go to **View | Visibility | Show All**

To ensure the edited site is selected for the docking, go to **Input Binding Site Parameter** and scroll down the list to update it. Reselect **Site 1**, and check that the size of site is decreased by the reduction in the number of site points compared to the original site 1 (Figure 3).



**Figure 3: Updated site list for Site 1 after Site Editing.**

Described in this exercise is one approach to defining a binding site, the other approach is to use a bound ligand to help define the site.

### 3. Specify the input ligands

The next step is to specify the ligands to dock to the designated receptor.

When a protein-ligand complex is read from a PDB file, the ligand is considered a part of the protein. You can see this in the Hierarchy View by noticing that the ligand is indicated with a residue icon and is a child of a chain under the protein molecule.

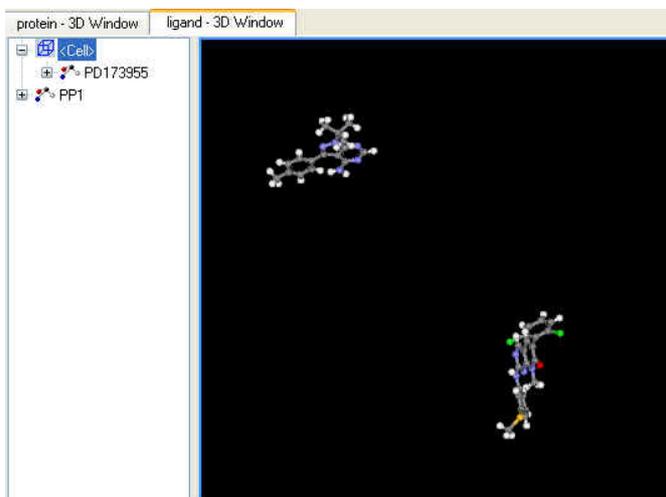
From the **File** menu, select the **Open** command and locate the *ligand.msv* file. Select this file and click **Open**.

A new 3D Window opens and displays the ligand PD173955 that we extracted from the PDB file.

Ensure the **ligand-3D Window** is active.

From the **File** menu, choose **Insert From** and select **File....** In the *DataFiles* subdirectory, select the file *pp1.msv*. Click the **Open** button.

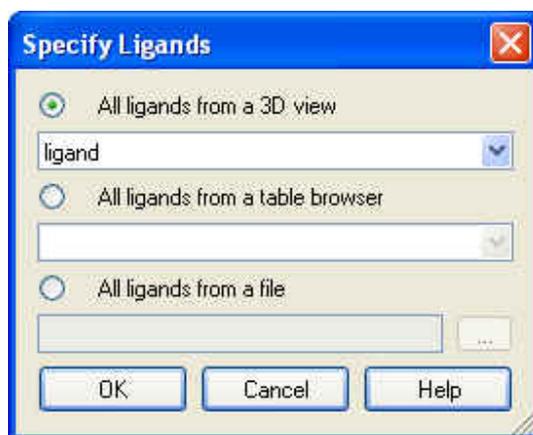
The second ligand is displayed in the same view as the first ligand (Figure 4). If you cannot see both ligands, click in **Fit to Screen** button.



**Figure 4: PD173955 and PP1 ligands displayed in the same 3D View.**

This second ligand, PP1, is a known inhibitor of Src tyrosine kinases and has recently been shown to be active towards Bcr-Abl tyrosine kinase.

In the **Parameter Explorer** click on **Input Ligands** parameter. Press the browser button  that appears to open the **Specify Ligands** dialog. You will see the entry *ligand* appear under **All ligands from a 3D view** (Figure 5).



**Figure 5: The Specify Ligands Dialog.**

This entry contains all molecules in the active 3D Window. If your ligands are in different windows, you will need to load all ligands to be docked into the same 3D Window.

In the Specify Ligands dialog, toggle on **All ligands from a 3D View** option.

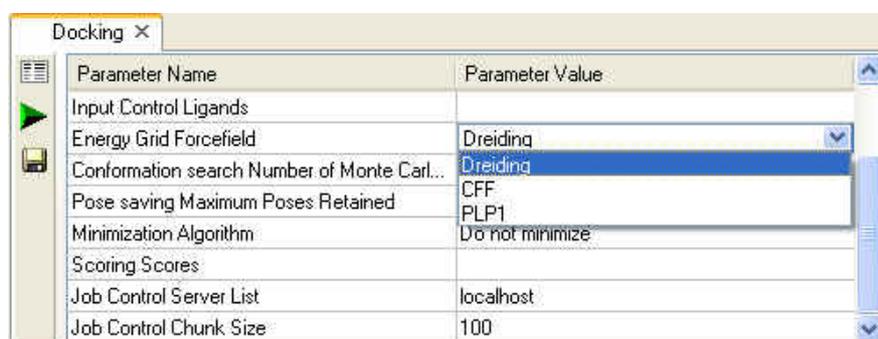
Click OK to close this dialog.

The *ligand* entry will now appear under the Input Ligands parameter setting.

The Input Control Ligand parameter would be used to specify ligands for monitoring the docking experiment. We will examine the use of control ligands in the next exercise.

Click on the **Energy Grid Forcefield** parameter.

Under this Parameter Value, you can choose one of the following three forcefields to be employed for computing ligand-protein interaction energy during docking: Dreiding/Gasteiger, CFF, or PLP1 (Figure 6). We will use the default setting of Dreiding.



**Figure 6: The Energy Grid Forcefield Parameter settings.**

Click on the  button in the Parameters Explorer to view the **Energy Grid Parameters** options.

The Energy Grid Parameters control the grid based docking used in the initial evaluation of the poses. You can change the parameters listed in Table 2 or leave them at their default values.

**Table 2: The Energy Grid Parameters**

<b>Energy Grid Extension from Site</b>	Defines the distance between the site points and the edges of the energy grid. Energy grids are used to compute the interaction energy between the protein and ligand. The grid is usually larger than the site to cover the ligand during docking.  If you have ligands larger than the defined site, you may want to consider enlarging this distance.
<b>Energy Nonbonded Cutoff Distance</b>	Sets the cutoff distance for nonbond interaction calculations.
<b>Energy Grid Outside Binding Site Penalty</b>	If selected, this toggle lowers the ligand contribution to the DockScore to the specified penalty value (kcal/mol per atom) when a ligand atom is outside the defined binding site.
<b>Energy Grid Softened Potential Energy</b>	This toggle employs a softened version of the Van der Waals function for protein-ligand docking. This function behaves like the regular function except that as the distance of separation approaches zero, the softened potential rises to a large value instead of increasing without limit.
<b>Energy Grid Dielectric Constant</b>	Sets the value of the dielectric constant
<b>Energy Grid Distance Dependent Dielectric</b>	Sets a distance dependent dielectric in the electrostatic interaction energy calculation. The interaction energy now has the distance dependence of $1/(Dr^2)$ instead of the usual $1/(Dr)$ . Note that the value of the dielectric constant specified in the field above is still used for D.  This is an approach for mimicking the screening effects of solvent.

We will change one parameter for the Energy Grid.

Set the **Energy Grid Extension from Site** parameter to 5.0 Ångstroms.

Leave all values in the Energy Grid Parameters set to their default values and click the  button to return to the main Parameters Explorer.

The extension of the grid is necessary as PD173955 does extend out of the designated binding site. We are giving the program a bit more leeway in placing the ligand.

#### 4. Perform a rigid fit

In the **Parameter Explorer**, click on **Conformation search Number of Monte Carlo Trial**. Clear the text in the cell and enter a value of 0 to perform a rigid docking.

**Note:** Pressing the Browse button will open the Monte Carlo Trials dialog that allows you to edit the default variable trials per torsion table for the Docking protocol.

The remaining options for docking can be found under the extended Parameter Explorer by pressing the  button and are briefly described in Table 3 below.

**Table 3: Additional Parameter Options for Docking**

<b>Docking RMS Threshold for Ligand/Site Match</b>	When a ligand pose is generated, its shape is compared with the site and all subsites (if a site partition has been performed). If the difference between the shape of the ligand and the site is within the user defined RMS Threshold for Ligand/Site Match, the ligand pose is docked into the matched site/subsite(s).	
<b>Docking Rigid Body SD Iterations</b>	During docking, steepest descent rigid body minimization is applied for the given number of iterations prior to BFGS rigid body minimization.	
<b>Docking Rigid Body BFGS Iterations</b>	During docking, following each steepest descent stage, BFGS rigid body minimization is applied for the given number of iterations.	
<b>Pose Saving DockScore Threshold</b>	When docking is complete, only the retained ligand poses with a DockScore above a user-defined threshold are saved to the database.	
<b>Pose Saving Rigid Body BFGS Iterations</b>	When the docking is complete, all ligand poses in the Save List are further optimized by BFGS rigid body minimization for the given number of iterations.	
<b>Pose Saving RMS Threshold for Diversity</b>	Sets the RMS criteria (in angstroms) for retaining diverse poses. If both DockScore and RMS differences between two ligand poses are within the user defined Score threshold and RMS threshold, they are considered similar and only the one with the better DockScore is retained.	
<b>Pose Saving Score Threshold for Diversity</b>	Sets the score criteria (in kcal/mol) for retaining diverse poses. If both DockScore and RMS differences between two ligand poses are within the user defined Score threshold and RMS threshold, they are considered similar and only the one with the better DockScore is retained.	
<b>Post processing</b>	<b>Pose saving Perform Clustering</b>	If this toggle is on, clustering is performed on retained poses.
	<b>Pose saving Maximum Clusters per molecule</b>	The maximum number of clusters

	<b>Pose saving RMS Threshold For Clusters</b>	The maximum RMS difference allowed within a cluster
--	---	---

We will accept the default options for this run with the exception that this will be a rigid fit.

In the main **Parameters Explorer**, ensure the parameter for **Minimization Algorithm** is set to *Do not minimize*.

Click on the **Scoring Scores** Parameter

This panel allows you to choose various empirical scoring functions to evaluate the ligand poses. This calculation is in addition to the DockScore value mentioned before. If no selections are made at this time, the scores may be calculated from the collected poses later. For this exercise we will select only one scoring function, PLP2 (3, 4).

Under the Scoring Scores parameter value, check the **PLP2** box.

Leave the rest of Parameter Values at their default settings.

Click the **Green Run**  button start the docking experiment.

After the experiment has been started the Jobs View will appear showing the progress of your experiment. This run should take less than two minutes on a Pentium 4 2.0 GHz processor.

## 5. View the results

When the experiment is done, a **Job Completed** dialog appears. Click **OK** in this dialog.

Once the job Status reads Finished, the output files in the specified default folder can be located by clicking on the job name and pressing the  button.

Under the Output folder, information regarding details of the run is shown including access to the log file.

In the Output folder, select and double-click the file *BrowseMolecules.pl*.

This will simultaneously open the results in a Table Browser, as well as a 3D View containing the first docked ligand pose in the binding site.

Right-click any cell of the Table Browser, and select the **Group By...** command. This opens the Group By dialog. In the popup list under **Select columns to group by**, scroll down and select **MOL\_NUMBER**. Click **OK** to close the dialog.

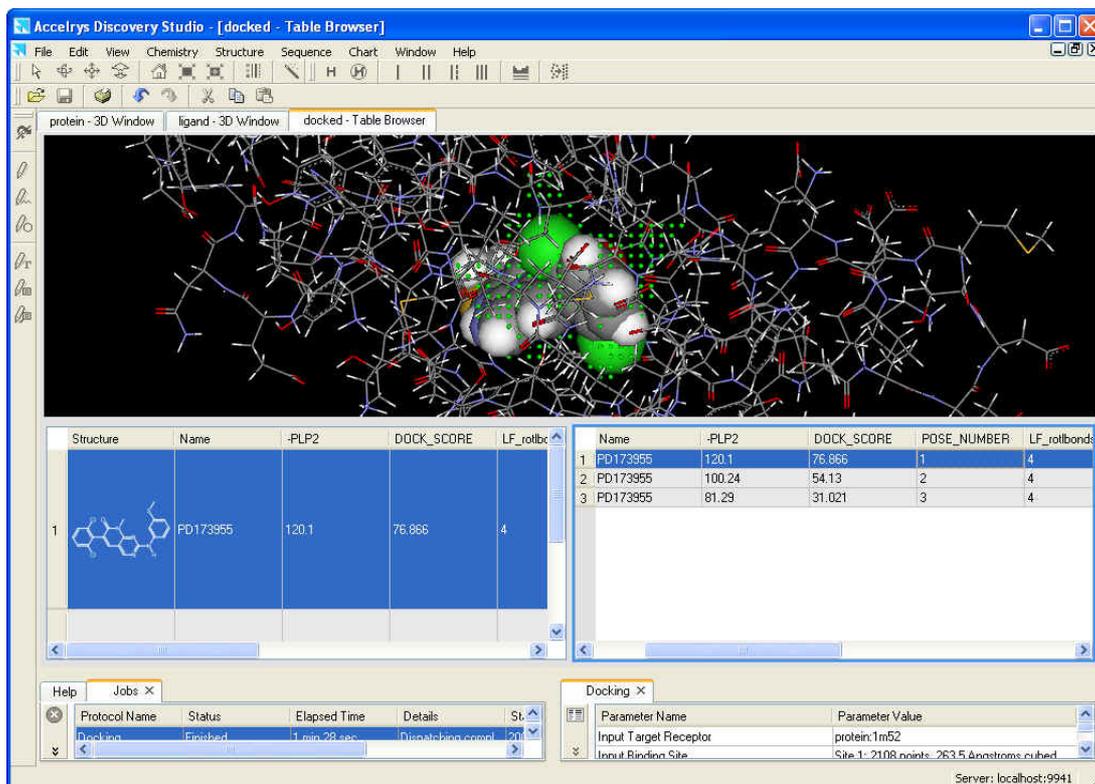
The Table Browser is now split into two views. The left view contains a list of the ligands, and the right view contains a list of all the poses for the first selected ligand (Figure 7).

There will be one row for each ligand. Currently, the table should show two ligands: PD173955 and PP1. The original PD173955 inhibitor should be displayed in the 3D View.

With the mouse in the Table Browser, click on the row number for the PP1 ligand.

Observe that the docked pose changes in the 3D View to show PP1 docked to the binding site.

**Note:** If you are only seeing the binding site then click in the 3D Window and go to **View | Visibility | Show All** to see the protein .



**Figure 7: Interactive display of the docked ligand in the binding site, the Table Browser and the Table of ligand poses.**

To better view the ligand double click on the dots of the binding site and then go to **View | Visibility | Hide**.

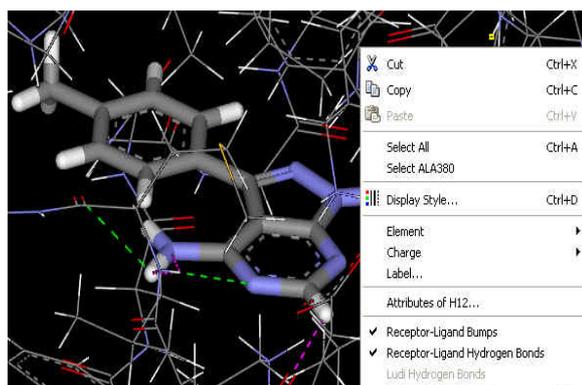
Now let us examine some protein-ligand interactions.

In the 3D View, right-click an empty space and then select **Receptor-Ligand Bumps** from the context menu.

Short contacts between the ligand pose and the protein receptor are displayed in magenta.

Right-click again in an empty space in the 3DView and select **Receptor-Ligand H Bonds** from the context menu.

Green dotted lines appear that indicate hydrogen bond interactions between the ligand and the protein.



**Figure 8: Display of protein-ligand interactions.**

Both the hydrogen bond and bump monitors are visual cues that may be useful for in better understanding the protein-ligand interactions for different ligands and poses.

Expand the Table Browser by using the mouse to drag and resize.

The table has columns for data concerning each ligand. The ligands are sorted by the best DockScore for the best pose. Note that the PLP2 value is also shown as we had specifically requested this score.

Select the row for PD 173955. In the right hand side table you will see this molecule has 3 poses. Select on a different row corresponding to each pose one at a time. As you select different rows, the image in the 3D View is updated to reflect the new ligand pose.

Note that the poses are listed in the decreasing order of their Dock Score.

Click on the row corresponding to the first pose in the right hand side table to view it in the binding site.

This pose should have a DockScore of about 70-80. Note that the pose has monitors for both hydrogen bonds and bumps if they are still visible.

## 6. Evaluate the fit against the original structure

It is of interest to find out the change in position and orientation of the ligand due to rigid docking to the selected site. One way of measuring this change is to compute the RMS difference between this docked pose and the original ligand from the X-ray data. You may perform this calculation using the following steps.

Select the first highest scoring DockScore pose of PD 173955 in the Table Browser so that it is shown in the 3D View.

Double-click this ligand so that it is highlighted and selected in the 3D View.

Ensure only the ligand is displayed by going to **View | Visibility | Show Only**.

Go to **File | Insert From | File...** and select the molecule file *ligand.msv* and then click **Open**. The original inhibitor will be displayed into the current 3D View.

Select the input ligand so it is highlighted in yellow, then go to **Structure | RMSD | Set Reference** menu command.

Now double-click on the current highest scoring pose in the 3D View to select it.

From the **Structure** menu, choose **RMSD** and select **Heavy Atoms** command.

A text window is displayed reporting an RMS difference of approximately 0.30 Å for the highest scoring pose with respect to the reference pose.

Close the text window. Select the second pose for PD 173955 molecule.

Repeat the above operations to choose the reference pose, and then and calculate the RMSD for the second pose.

The RMSD between the reference and pose is only calculated if the following three conditions are fulfilled:

1. The number of atoms match,
2. Each element matches, and
3. The connectivity matches.

Failures to meet these criteria are reported in the RMSD report.

## 7. Perform a flexible search

Thus far we have only performed a rigid body docking of the ligands to the protein. To perform a flexible search we only need to specify a few additional parameters. We will begin by returning to the Parameters Explorer.

In the Parameters Explorer go to the **Conformation search Number of Monte Carlo Trial** parameter. Clear the current value and enter *5000* to perform flexible docking with a fixed number of trials. Ensure this value is registered by selecting another Parameter input cell.

The flexible fit causes the torsion angles of the ligand to be randomly varied during docking. In the case of rigid docking, only the position and the orientation of the ligand are altered to find optimum docking.

Leave the rest of Parameter Values at their default settings.

Click the **Green Run**  **button** start the docking experiment.

After the experiment has been started the Jobs View will appear showing the progress of your experiment. The run will proceed and take approximately five minutes on a Pentium 4 2.0 GHz computer.

## 8. View the results

When the experiment is done, a **Job Completed dialog** appears. Click **OK** in this dialog.

Once the job Status reads Finished, the output files in the specified default folder can be located by clicking on the job name and pressing the  button.

In the Output folder, select and double-click the file *BrowseMolecules.pl* to examine the results.

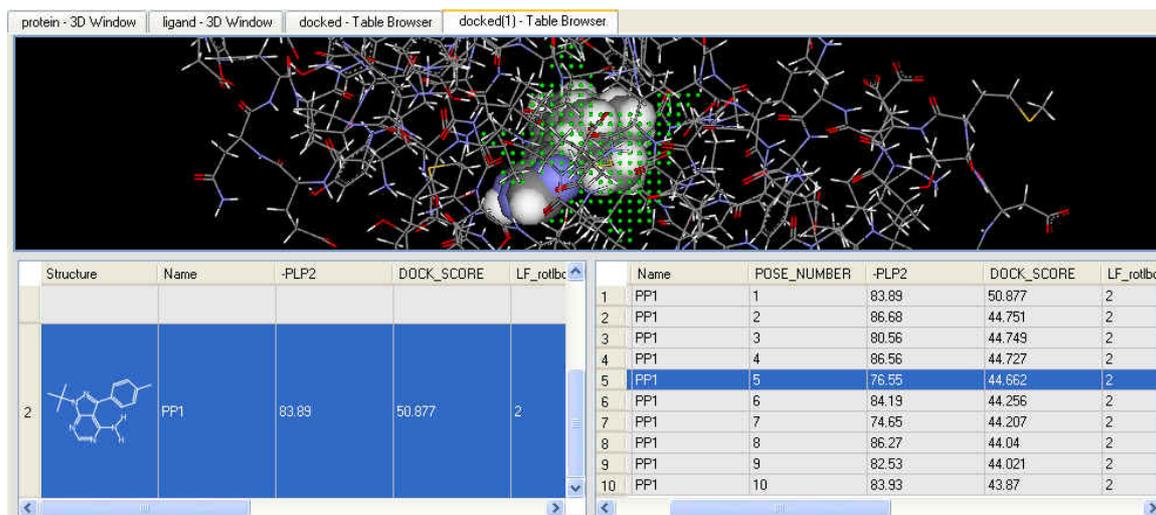


Figure 9: Increased number of poses resulting from flexible docking.

The Table Browser will appear with a 3D View of the binding site with a docked ligand pose.

In the Table Browser, right-click and select the **Group By...** command. Choose **MOL\_NUMBER** from the **Select column to group by** drop down menu in the Group By dialog. Click **OK** to close this dialog.

In the table that appears on the right hand side of the Table Browser; note that both ligands now have 10 entries for the number of poses (Figure 9).

Allowing flexibility of the ligand increases the number of possible poses. The number of poses found depends on the Parameter setting for **Pose saving Maximum Poses Retained** value, and for this experiment this was set to 10.

In structure-based drug design, docking of potential ligands to the receptor is often a crucial step. Whether you dock entire ligands as part of a screening protocol or smaller fragments to use as scaffolds, you want to properly orient the ligand in the receptor. DS LigandFit allows you to rapidly perform docking, generating multiple poses of the ligand. DS LigandScore, then, allows you to evaluate the poses with a range of scoring functions. From the analysis of the docking experiments, you can gain a better understanding of the protein-ligand interactions required for optimal leads.

### **References:**

1. Druker BJ, Lydon NB. 2000. Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* 105: 3
2. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, et al. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* 2: 561
3. Gehlhaar DK, Bouzida D, Rejto PA. 1999. Reduced Dimensionality in Ligand-Protein Structure Prediction: Covalent Inhibitors of Serine Proteases and Design of Site-Directed Combinatorial Libraries. In *Rational Drug Design: Novel Methodology and Practical Applications*, ed. AL Parrill, MR Reddy, pp. 292. Washington, DC: American Chemical Society
4. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, et al. 1995. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 2: 317
5. Hanke JH, Gardner JP, Dow RL, Changelian PS, Brissette WH, et al. 1996. Discovery of a novel, potent, and Src family-selective tyrosine kinase inhibitor. Study of Lck- and FynT-dependent T cell activation. *J Biol Chem* 271: 695
6. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. 1996. Protein clefts in molecular recognition and function. *Prot Sci* 5: 2438
7. Liang J, Edelsbrunner H, Woodward C. 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7: 1884
8. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, et al. 2002. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res* 62: 4236

9. Pendergast AM, Gishizky ML, Havlik MH, Witte ON. 1993. SH1 domain autophosphorylation of P210 BCR/ABL is required for transformation but not growth factor independence. *Mol Cell Biol* 13: 1728
10. Tatton L, Morley GM, Chopra R, Khwaja A. 2003. The Src-selective Kinase Inhibitor PP1 Also Inhibits Kit and Bcr-Abl Tyrosine Kinases. *J Biol Chem* 278: 4847
11. Venkatachalam CM, Jiang X, Oldfield T, Waldman M. 2003. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Mod* 21: 289