

# Discovery Studio 2.0

Tutorials  
Version 1.0

**Note.** Some tutorials require data files that can be downloaded here:  
<http://www.accelrys.com/doc/life/dstudio/20/tutorials/tutorialdata.zip>.

Save the zip file to your computer and then extract the contents to a folder. We recommend that you add this folder to the Files Explorer as a top-level or 'root' folder. To do this, navigate to the folder, right-click the folder, and choose the Add Root Node... command from the context menu. Next, on the dialog, enter a display name for the root folder (e.g., Tutorial Data) and click the OK button. The tutorial data files will then be readily accessible at the top-level of the Files Explorer.

## ZDOCK tutorial

**Purpose:** Learn how to determine and refine protein-protein complex structures using ZDOCK and RDOCK.

**Required functionality and modules:** Discovery Studio Visualizer client, Dock Proteins with ZDOCK, and Refinement with RDOCK.

**Required data files:** 2ptn.pdb and 2sta\_I.pdb.

**Time:** 1 hour 30 minutes.

### Background

ZDOCK is a rigid-body protein-protein docking algorithm based on the Fast Fourier Transform correlation technique that is used to explore the rotational and translational space of a protein-protein system. RDOCK is a CHARMM-based energy minimization procedure for refining and scoring docked poses using energy scoring functions.

### Introduction

In this tutorial, you will perform protein docking with ZDOCK, analyze the docked poses, and refine a set of selected poses with RDOCK. A bovine beta-trypsin structure (PDB ID 2ptn) will be used as the receptor protein, and a trypsin inhibitor CMTI-I (PDB ID 2sta, I chain) as the ligand protein. The X-ray crystal structure of the bovine beta-trypsin and inhibitor CMTI-I protein complex is known (PDB ID 1ppe).

This tutorial covers:

- Setting up a ZDOCK run
- Analyzing ZDOCK results
- Refining docked poses with RDOCK
- Analyzing the binding interface residues using RMSD

### Setting up a ZDOCK run

#### 1. To open the receptor and ligand protein files in the same 3D Window

In the Files Explorer, navigate to and double-click the **2ptn.pdb** data file.

This opens the protein in a new 3D Window.

Next, navigate to the **2sta\_I.pdb** data file and drag and drop it into the same 3D Window.

#### 2. To open the Dock Proteins (ZDOCK) protocol and modify the parameters

In the Protocols Explorer, expand the **Protein Modeling** folder and double-click **Dock Proteins (ZDOCK)**.

This opens the parameters in the Parameters Explorer.

In the Parameters Explorer, click the **Input Receptor Protein** parameter and choose **2ptn:2ptn** from the dropdown list.

Click the **Input Ligand Protein** parameter and choose **2ptn:2sta\_I** from the dropdown list.

Click the **Angular Step Size** parameter and choose **15** from the dropdown list.

**Note.** ZDOCK can run with two Euler angle sampling sizes: 6 degrees or 15 degrees. The 6 degrees search provides more accurate predictions because it samples 54,000 poses. Although the 15 degrees search is not as accurate because it samples only 3600 poses, it finishes faster than the 6 degrees search. For this tutorial, sample size is more important than the speed of the protocol run.

Expand the **Clustering** parameter group.

Click the **RMSD Cutoff** parameter and enter **6.0** for the value.

Click the **Interface Cutoff** parameter and enter **9.0** for the value.

Click the **Maximum Number of Clusters** parameter and enter **60** for the value.

For this small protein inhibitor, a small cluster radius of 6.0 Å for the *RMSD Cutoff* and a smaller *Interface Cutoff* of 9.0 Å produce better clustering results. In this tutorial, a single processor ZDOCK job is performed without specifying blocked residues or filtering binding site residues.

**Note.** ZDOCK can be run with blocking and/or filtering options:

If data suggests that certain residues are not likely to appear in the protein-protein interface, select those residues on the receptor protein, and then set the *Receptor Blocked Residues* parameter to [Create New Group from selection](#). Similarly, select residues to be blocked on the ligand protein for input in the *Ligand Blocked Residues* parameter.

If data suggests that certain residues would appear in the protein-protein interface, select them. Then expand the *Filter Poses* parameter group and set the *Receptor Binding Site Residues* parameter and the *Ligand Binding Site Residues* parameter to [Create New Group from selection](#).

The *Filter Poses function* can also be performed separately after a ZDOCK run, using the Process Poses (ZDOCK) protocol.

### 3. To run the calculation and view the results

In the Protocols toolbar, click the **Run**  button and wait for the job to complete.

The job takes about 15 minutes on a Pentium 4, 2Gb RAM, 2.8GHz machine.

The Job Completed dialog displays when the job is complete.

In the Jobs Explorer, double-click the completed job.

This opens the Report.htm file in an Html Window.

In the *Output Files* section, click the **ZDockResults.msv** link.

## Analyzing ZDOCK results

### 1. To visualize the clustering of poses

In the Hierarchy View, expand **Docked Poses** and **Clusters**.

This displays the groups of clusters and poses. *Cluster\_1* is the largest cluster, followed by progressively smaller clusters.

In the Data Table, click the **ProteinPose** tab.

This displays the following data from the ZDOCK run:

- *ZDock Score* - contains the ZDOCK score (PSC scoring function) of each pose. Pose1 has the highest (best) ZDOCK score.
- *Cluster* - contains the cluster group number for each pose.
- *ClusterSize* - reports the size (number of poses) of a cluster.
- *Density* - represents the number of neighboring poses in the clustering.

For details about the clustering algorithm, refer to the Clustering and analysis of docked protein poses Help topic in Discovery Studio.

In the Tools Explorer, choose **Protein Modeling** from the dropdown list.

This opens the Protein Modeling tool panels.

In the Analyze Docked Proteins tool panel, in the *Clustering* tools group, click **Show Cluster Center** and then click **Show Full Cluster**.

This toggles the display between showing all poses and only the cluster centers.

## 2. To plot the clustering of poses

The following steps illustrate several ways to use plots and the *Analyze Docked Proteins* tools to help view the poses and select a smaller number of ZDOCK poses for further RDOCK refinement.

In the Hierarchy View, press and hold **SHIFT** and click **Cluster\_1** and then **Cluster\_10**.

This selects clusters 1 through 10.

In the Data Table, click the **ProteinPose** tab to make it active. From the menu bar, choose **Chart | 3D Point Plot**.

This opens the Choose Plot Axes dialog.

On the dialog, choose **ZDock Score** for the *X axis*, **Cluster** for the *Y axis*, **Density** for the *Z axis*, and **ZDock Score** for the *Color axis*.

Click the **OK** button.

This opens a 3D Point Plot.

Click the **ZDockResults - 3D Point Plot** tab and drag the window so that you can also view the **ZDockResults - 3D Window**.

In the *View* toolbar, click the **Select** tool and lasso some of the poses in the **ZDockResults - 3D Point Plot** Window with **high ZDock scores**, **high Density values**, and **low Cluster numbers**.

The poses you select in the plot are also selected in the 3D Window.

Click the **ZDockResults - 3D Window** tab to make it active. In the Hierarchy View, in the *Clusters* group, click **Cluster\_2** to select it.

In the Tools Explorer, in the Analyze Docked Proteins tool panel, in the *Display* tools group, click **Show Full Docked Complex**.

This displays *Pose 1* in the 3D Window. This also displays the two names for the Pose 1 protein complex in the Hierarchy View: *Pose1\_2ptn* and *Pose1\_2sta\_1*.

In the Analyze Docked Proteins tool panel, click **Show Next Docked Complex**.

This displays *Pose 3* in the 3D Window. This also displays the two names for the Pose 3 protein complex in the Hierarchy View: *Pose3\_2ptn* and *Pose1\_2sta\_1*.

Click **Show Next Docked Complex** to cycle forward through the poses and click **Show Previous Docked Complex** to cycle back through all the poses that you selected.

Return to display **Pose 1**.

In the Hierarchy View, click **Pose1\_2ptn**.

This selects the entire protein.

In the Analyze Docked Proteins tool panel, click **Select Binding Interface**.

This selects the receptor and ligand binding interface residues.

In the RMSD tools group, click **Define RMSD Reference**.

This creates a group called *Pose1 POSE\_RMSD\_REFERENCE* in the Hierarchy View. This pose will be used for the RMSD calculation.

### 3. To calculate the RMSD

In the Analyze Docked Proteins tool panel, in the *RMSD* tools group, click **Calculate Binding Site RMSD (ZDOCK)**.

This adds a column named *RMSD to Pose1 POSE\_RMSD\_REFERENCE* to the *ProteinPose* tab in the Data Table.

Many top poses in *Cluster\_2* (the second largest cluster) predict the nearest correct binding mode of PDB 1ppe.

**Tip.** In the *ProteinPose* tab of the Data Table, double-click the *Cluster* column to sort by cluster size.

## Refining docked poses with RDOCK

### 1. To set up and run the Refine Docked Proteins (RDOCK) protocol

Close the **ZDockResults - 3D Window** and the **ZDockResults - 3D Point Plot Window**.

Click the **Dock Proteins (ZDOCK) - Html Window** tab to make it active.

In the *Output Files* section, click the **ZDockResults.msv** link to open a new ZDockResults 3D Window.

RDOCK requires that the receptor and ligand proteins are typed, without errors, by the CHARMM Polar H forcefield.

In the Tools Explorer, choose **Simulations** from the dropdown list.

This opens the Simulations tool panels.

In the **Forcefield** tool panel, choose **CHARMM Polar H** from the dropdown list, then click **Apply Forcefield**.

This types the receptor and ligand proteins with the CHARMM Polar H forcefield. The status **ZDockResults typed with CHARMM Polar H** displays in the tool panel.

In the Protocols Explorer, in the **Protein Modeling** folder, double-click the **Refine Docked Proteins (RDOCK)** protocol.

This opens the parameters in the Parameters Explorer.

In the Parameters Explorer, click the **Input Typed Receptor Protein** parameter and choose **ZDockResults:2ptn** from the dropdown list.

Click the **Input Typed Ligand Protein** parameter and choose **ZDockResults:2sta\_I** from the dropdown list.

In this example, a few top poses from *Cluster\_2* will be selected. Looking at the top 10 poses, *Cluster\_2* has four: 1, 3, 4, and 8. In addition, the binding loop is known in the trypsin inhibitor CMTI-I, which is in the N-terminal region with ARG505 as the leading anchor residue. All the poses in *Cluster\_2* show that the N-terminal loop is in the binding interface. Based on this analysis, the top five poses in *Cluster\_2* (1, 3, 4, 8, and 13) will be further refined and rescored with RDOCK.

In the Hierarchy View, in the *Docked Poses* group, press and hold **CTRL** and click poses **1, 3, 4, 8, and 13** to select them.

In the Parameters Explorer, click the **Input Poses** parameter and choose **Create New Group from selection...** from the dropdown list.

The new group is named *Group* by default.

**Note.** Leave the *Dielectric Constant* parameter set to **4.0**. This will be the dielectric value used in the RDOCK CHARMM energy calculation.

In the Protocols toolbar, click the **Run**  button and wait for the job to complete.

The job takes about five minutes on a Pentium 4, 2Gb RAM, 2.8GHz machine.

The Job Completed dialog displays when the job is complete.

## 2. To view the results

In the Jobs Explorer, double-click the job summary.

This opens the Report.htm file in an Html Window and highlights the Output folder of the job in the Files Explorer.

In the Files Explorer, in the Output folder, right-click **Poses.mol2** and choose **Open With | Table Browser**.

Refined poses appear in a Table Browser with the energies calculated by RDOCK listed in the columns.

- $E_{elec1}$  and  $E_{elec2}$  - The electrostatic energy of the protein complex after the first and second CHARMM minimizations.
- $E_{vdw1}$  and  $E_{vdw2}$  - van der Waals nonbond energy of the protein complex after the first and second CHARMM minimizations.
- $E_{sol}$  - The desolvation energy of the protein complex calculated by the ACE method.
- $E_{RDock}$  - The RDOCK score defined as:  $E_{elec2} + \beta * E_{sol}$ .

In the Table Browser, right-click in the **first cell** and choose **Show Structure in 3D Window**.

This displays the refined structure in a Graphics View in the 3D Window.

From the menu bar, choose **View | Hierarchy**.

This displays the Hierarchy View in the 3D Window.

In the Table Browser, click **a row**.

This adds the pose to the Hierarchy View and displays it in the Graphics View.

In the Table Browser, click **each row**.

This adds the poses to the Hierarchy View. When a new row is selected, the other poses are hidden in the Graphics View.

**Tip.** Check the pose checkbox in the Hierarchy View to turn the molecule display on or off in the Graphics View.

### 3. To select the interface residues

In the Hierarchy View, expand the **selected pose** and select the receptor chain **<AminoAcidChain>**.

In the Tools Explorer, choose **Protein Modeling** from the dropdown list.

This opens the Protein Modeling tool panels.

In the Analyze Docked Proteins tool panel, in the *Display* tools group, click **Select Binding Interface**.

This selects the interface residues in the Graphics View.

**Tip.** The *Distance Cutoff* value used to define the binding interface can be set in the Preferences. From the menu bar, choose *Edit | Preferences | Protein Docking*. The default is 5 Å.

## Analyzing the binding interface residues using RMSD

In this section, you will compare docked poses with the crystal structure of the protein complex PDB 1ppe.

### 1. To open PDB 1ppe

From the menu bar, choose **File | Insert From | URL...**

This opens the Insert From URL dialog.

Check that the **PDB ID** radio button is selected, then enter **1ppe** in the textbox.

Click the **Open** button.

This inserts the protein complex in the Graphics View.

In the Hierarchy View, expand **<Cell>**, then expand **1ppe**.

Click **Water** to select it, and then right-click and choose **Cut**.

This removes the water chain from 1ppe.

From the menu bar, choose **Sequence | Show Sequence**.

This opens a Sequence Window.

Before running the RMSD calculation, the poses and 1ppe should be superimposed based on the receptor protein.

### 2. To superimpose the poses and 1ppe

Click the **Poses - Table Browser** tab to make the window active.

In the Hierarchy View, click **chain E** to select it.

This selects the receptor chain of 1ppe.

From the menu bar, choose **Edit | Preferences**.

This opens the Preferences dialog.

Click the **Superimposition** page. Check the **Selected reference residues only** checkbox.

Click the **OK** button.

From the menu bar, choose **Structure | Superimpose | By Sequence Alignment...**

This opens the Sequence Alignment dialog. 1ppe is the reference molecule.

On the dialog, select all of the **poses** in the *Molecules to superimpose* list.

Click the **OK** button.

Click the **Poses - Table Browser** tab to make it active. In the Table Browser, check that the 1ppe receptor chain E is still selected.

In the Tools Explorer, in the Analyze Docked Proteins tool panel, click the **Select Binding Interface** command.

This selects the interface residues.

### 3. To calculate the binding interface residues RMSD values

From the menu bar, choose **Structure | RMSD | By Sequence Alignment...**

This opens the RMSD by Sequence Alignment dialog.

On the dialog, click the **Reference molecule** dropdown list and choose **1ppe**. Click the **Selected residues** radio button, and clear the **Report at residue level** checkbox. Click the **OK** button.

This opens the RMSD Report using the selected residues for poses in an Html Window.

Compare the E\_RDock scores of the docked poses with the RMSD values in the RMSD report. The pose with the lowest RMSD also has the lowest (best) E\_RDock score.

This is the end of the tutorial.

# Contacting Accelrys

## Technical Support

Accelrys is committed to providing you with the best overall product experience. We place the highest priority on our customer interactions and understand that issues may arise that require technical support. Feel free to contact us by phone or email and we will direct your inquiry to a support representative.

### U.S.

Hours: M–F 6:00 AM – 5:30 PM (PST)

- PHONE: 1-800-756-4674 (Toll Free)
- PHONE: 1-858-799-5509

### Europe

Hours: M–F 9:00 AM – 6:00 PM (CET)

- PHONE: +44 1223 228822 (Outside UK)
- PHONE: +44 845 741 3375 (UK)

## Online Support

Accelrys Advantage, our online support system, is available 24/7, offering a Knowledgebase, interactive request tracking, and personal bug tracking.

- WEB SUPPORT: <http://www.accelrys.com/services/support.html>

## Accelrys Locations

### San Diego Office

Accelrys, Inc.  
10188 Telesis Court, Suite 100  
San Diego, CA 92121-4779  
U.S.A.

- PHONE: +1 (858) 799-5000
- FAX: +1 (858) 799-5100
- WEB: [www.accelrys.com](http://www.accelrys.com)

### Cambridge Office

Accelrys, Ltd.  
334 Cambridge Science Park  
Cambridge  
CB4 0WN  
United Kingdom

- PHONE: +44 1223 228500
- FAX: +44 1223 228501

### Tokyo Office

Accelrys KK  
Nishi-shimbashi TS bldg 11F  
Nishi-shimbashi 3-3-1, Minato-ku  
Tokyo, 105-0003  
Japan

- PHONE: 81 3 3578 3861
- FAX: 81 3 3578 3873