

# Getting Started with RDKit and Jupyter Notebooks

Layla Hosseini-Gerami  
Bender Group High-Dimensional Biology IG  
October 2020

# The actual tutorial will be in the Jupyter notebook

However, there are a few steps that need to be done before you can get started, if you have never used a Jupyter notebook before, or in particular RDKit

Please follow these steps before the Jupyter tutorial in case things go wrong, so that we can have a smooth session :)

# How to use the Jupyter Notebook

Download and install Anaconda <https://www.anaconda.com/products/individual>

Two ways to create the environment:

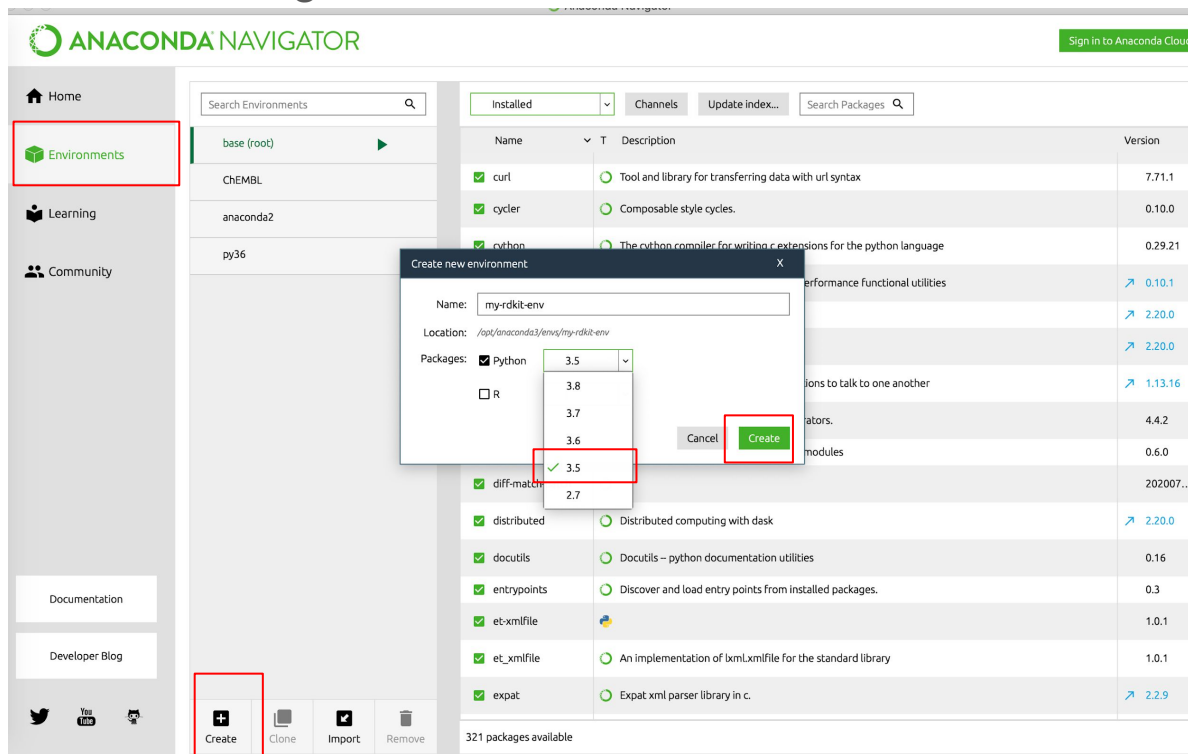
## 1. Command line

```
$ conda create -c conda-forge -n my-rdkit-env rdkit
$ conda activate my-rdkit-env
$ conda install -c conda-forge notebook
$ jupyter notebook
```

# How to use the Jupyter Notebook

**YOU MUST SELECT  
PYTHON 3.5! NOT 3.8!**

## 2. Using the Anaconda navigator



## Add rdkit to list of channels

The screenshot shows the JupyterLab interface with a sidebar on the left displaying a list of environments: "base (root)", "ChEMBL", "anaconda2", "my-rdkit-env", and "py36". The "my-rdkit-env" environment is highlighted with a red box. In the main panel, the "Channels" tab is selected and highlighted with a red box. Below the tab, the text "Manage channels you want Navigator to include." is visible. The "Add..." button is highlighted with a red box. A modal dialog is open, showing a list of channels with "defaults" at the top. At the bottom of the dialog, there are "Cancel" and "Update channels" buttons.

Search Environments

base (root)

ChEMBL

anaconda2

my-rdkit-env

py36

All Channels Update index... RDkit X

Name T Manage channels you want Navigator to include.

Add...

defaults

Cancel Update channels

Make sure you update channels

The screenshot displays the JupyterLab environment management interface. On the left, a sidebar titled "Search Environments" lists several environments: "base (root)", "ChEMBL", "anaconda2", "my-rdkit-env" (highlighted with a green bar and a play button), and "py36". The main panel shows a table of environments with columns for "Name", "Status", and "Channels". The "Channels" column for "my-rdkit-env" is expanded, showing a list of channels: "defaults" and "rdkit". The "rdkit" channel is highlighted with a red box. A dialog box titled "Manage channels you want Navigator to include." is open in the foreground. It contains a list of channels: "defaults" and "rdkit", each with a trash icon to its right. The "rdkit" channel is highlighted with a red box. At the bottom of the dialog, there are two buttons: "Cancel" and "Update channels". The "Update channels" button is highlighted with a red box.

Search Environments

base (root)

ChEMBL

anaconda2

my-rdkit-env

py36

All

Channels

Update index...

RDkit

Name

Manage channels you want Navigator to include.

Add...

defaults

rdkit

Cancel

Update channels


## Now you can search for and install RDKit

Search Environments





base (root)


ChEMBL




anaconda2

my-rdkit-env 



py36

 Create  Clone  Import  Remove

All  Channels Update index...  rdkit 

Name	Description	Version
<input checked="" type="checkbox"/> rdkit		2020.0...
<input type="checkbox"/> rdkit-postgresql		2020.0...
<input type="checkbox"/> rdkit-postgresql95		2017.0...

3 packages available matching "rdkit" 1 package selected

## Install Packages



33 packages will be installed

	Name	Unlink	Link	Channel	
1	rdkit	-	2017.09.3.0	rdkit	^
2	*blas	-	1.0	pkgs/main	
3	*boost	-	1.63.0	rdkit	
4	*bzip2	-	1.0.8	pkgs/main	
5	*cairo	-	1.14.12	pkgs/main	
6	*fontconfig	-	2.13.0	pkgs/main	v

\* indicates the package is a dependency of a selected package

Apply the installed packages :)

Cancel

Apply





Home



Environments



Learning



Community

[Documentation](#)

[Developer Blog](#)

Applications on

my-rdkit-env

Channels

base (root)

ChEMBL

anaconda2

✓ my-rdkit-env

py36

VS C

1.50.1

Streamlined code editor with support for development operations like debugging, task running and version control.

Launch



Glueviz

1.0.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



JupyterLab

2.2.6

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Install



Notebook

6.1.4

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Install



Orange 3

3.26.0

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

Install



Qt Console

4.7.7

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Install



RStudio

1.1.456

A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.

Install



Spyder









4.1.5

Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Install

## Install Jupyter Notebook in this env Simply click “Launch” when it is installed

Applications on my-rdkit-env Channels

 <b>VS Code</b> 1.50.1 Streamlined code editor with support for development operations like debugging, task running and version control. <a href="#">Launch</a>	 <b>Glueviz</b> 1.0.0 Multidimensional data visualization across files. Explore relationships within and among related datasets. <a href="#">Install</a>	 <b>JupyterLab</b> 2.2.6 An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture. <a href="#">Install</a>	 <b>Jupyter Notebook</b> 6.1.4 Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis. <a href="#">Install</a>
 <b>Orange 3</b> 3.26.0 Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox. <a href="#">Install</a>	 <b>Qt Console</b> 4.7.7 PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more. <a href="#">Install</a>	 <b>RStudio</b> 1.1.456 A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks. <a href="#">Install</a>	 <b>Spyder</b> 4.1.5 Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features <a href="#">Install</a>

Now, open the “RDKit\_Tutorial.ipynb” and get started!

If you run into issues, try installing the “conda-forge” channel RDKit by adding “conda-forge” to list of channels

Any other issues please email me [lh605@cam.ac.uk](mailto:lh605@cam.ac.uk)