

Literate programming with CCTBX and PyMOL in Jupyter notebooks

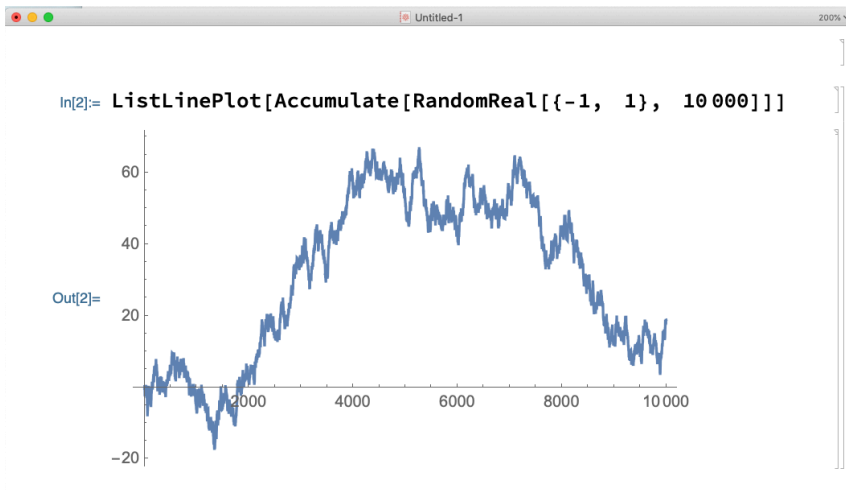
Blaine Mooers, PhD
blaine-mooers@ouhsc.edu
405-271-8300

<http://github.com/MooersLab/ACA2021>

Department of Biochemistry & Molecular Biology
University of Oklahoma Health Sciences Center, Oklahoma City

American Crystallographic Association
Virtual Meeting
August 3, 2021

Cells in Mathematica notebook



Cells in IPython

```
> IPython
Python 3.8.10 | packaged by conda-forge | (default, May 11 2021, 06:39:48)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.23.1 -- An enhanced Interactive Python. Type '?' for help.
```

```
In [1]: from pymol import cmd
```

```
In [2]: from PIL import Image
```

```
In [3]: cmd.fetch("1lw9")
ExecutiveLoad-Detail: Detected mmCIF
Out[3]: '1lw9'
```

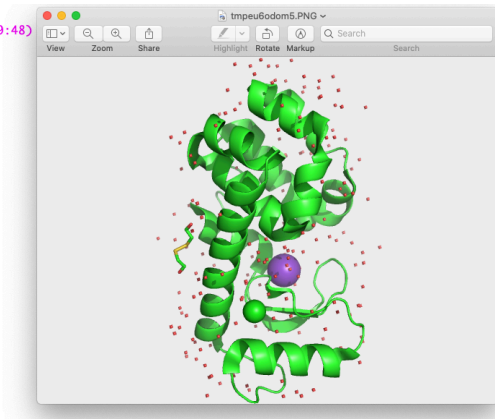
```
In [4]: cmd.orient();cmd.rotate(axis="z", angle="270")
Out[4]: 1
```

```
In [5]: cmd.show("nb_spheres", "resn HOH")
Out[5]: 1
```

```
In [6]: cmd.png("1lw9IPython.png")
Ray: render time: 1.12 sec. = 3217.5 frames/hour (1.12 sec. accum.).
Out[6]: 1
```

```
In [7]: image = Image.open('1lw9IPython.png');image.show()
```

```
In [8]: |
```



Timeline

1988 Mathematica Notebook, Theodore Gray

1990 Python, Guido van Rossum

2001 IPython, Fernando Pérez

2010 IPython moves to ZeroMQ-based architecture. Separate the frontend from the interpreter and supports parallel computing.

2011 IPython notebook

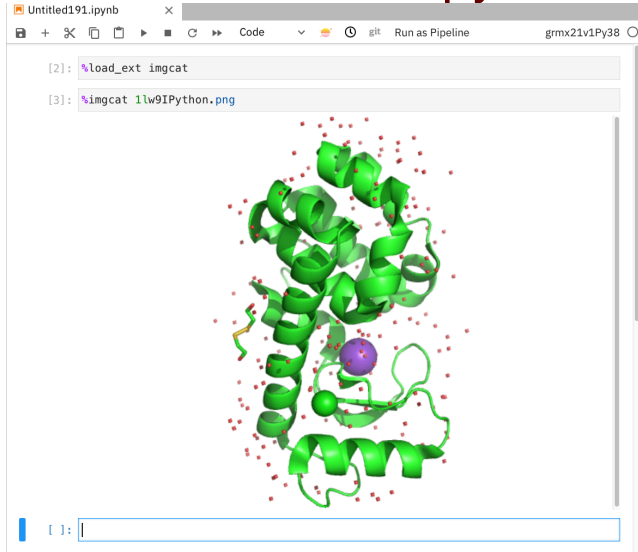
2014 Jupyter Notebook splits off with IPython 4.0.

2016 nteract

2018 JupyterLab

2018 Colab opened to public use

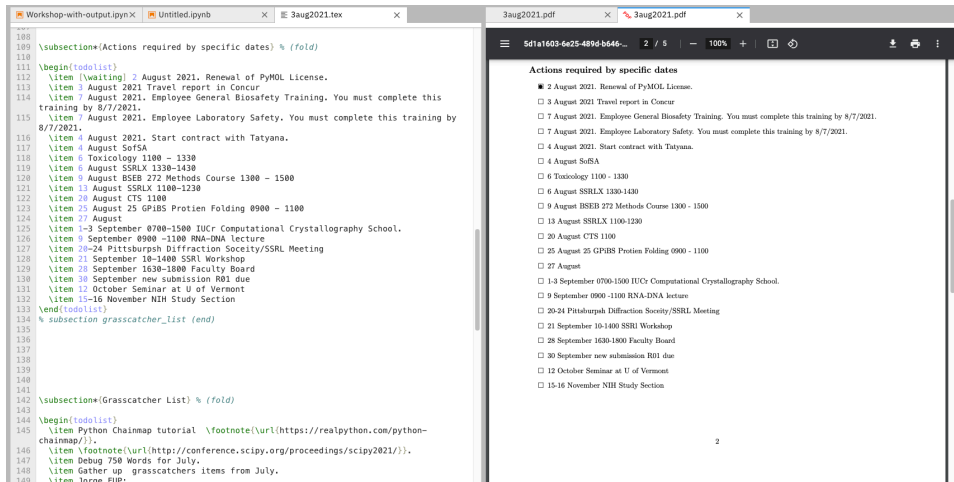
A Jupyter notebook in Jupyter Notebook



A Jupyter notebook in JupyterLab

The screenshot displays the JupyterLab web interface in a browser window. The top navigation bar includes menus for File, Edit, View, Run, Kernel, Git, Snippets, Tabs, Settings, and Help. The left sidebar contains a list of recent files and folders, including "distributed image pr...", "Outlook Web App, light version", "reciprocal space", "Computational Crystallography Tool...", "about:blank", "Settings", "Newline", "File (Houarblah) Faculty Goals 2020.html", "Julia 1.5 Documentation", "Google Calendar - Week of August 23, 2...", "loading - Reload configurations with...", and "How do I reload a file in a buffer? - Emacs...". The main area shows a Jupyter notebook titled "imgcatMagic.ipynb" with two code cells. The first cell contains the command `load_ext imgcat`, and the second cell contains `!imgcat 1lw9IPython.png`. Below the code cells, a 3D visualization of a protein structure is displayed, rendered in green ribbon format, surrounded by a cloud of red dots. The bottom status bar shows the notebook is in "Simple" mode, with 0 lines of code, 2 cells, and a memory usage of 478MB. The kernel is "Fully initial" and the notebook is "Saving co...". The mode is "Co..." and the current line is "Ln 1, ...". The notebook title "imgcatMag..." is also visible in the status bar.

LaTeX with pdf preview in JupyterLab



The screenshot displays the JupyterLab interface with two main panes. The left pane shows a Jupyter Notebook with LaTeX source code, and the right pane shows the rendered PDF document.

Left Pane (Source Code):

```
108
109 \subsection*(Actions required by specific dates) % {fold}
110
111 \begin{todolist}
112   \item [\waiting] 2 August 2021. Renewal of PyMOL License.
113   \item 3 August 2021 Travel report in Concur
114   \item 7 August 2021. Employee General Biosafety Training. You must complete this
training by 8/7/2021.
115   \item 7 August 2021. Employee Laboratory Safety. You must complete this training by
8/7/2021.
116   \item 4 August 2021. Start contract with Tatyana.
117   \item 4 August SofSA
118   \item 6 Toxicology 1100 - 1330
119   \item 6 August SSRLX 1330-1430
120   \item 9 August BSEB 272 Methods Course 1300 - 1500
121   \item 13 August SSRLX 1100-1230
122   \item 20 August CTS 1100
123   \item 25 August 25 GPIBS Protien Folding 0900 - 1100
124   \item 27 August
125   \item 1-3 September 0700-1500 IUCr Computational Crystallography School.
126   \item 9 September 0900 -1100 RNA-DNA lecture
127   \item 20-24 Pittsbupsh Diffraction Society/SSRL Meeting
128   \item 21 September 10-1400 SSRL Workshop
129   \item 28 September 1630-1800 Faculty Board
130   \item 30 September new submission R01 due
131   \item 12 October Seminar at U of Vermont
132   \item 15-16 November NIH Study Section
133 \end{todolist}
134 % subsection grasscatcher_list (end)
135
136
137
138
139
140
141
142 \subsection*(Grasscatcher List) % {fold}
143
144 \begin{todolist}
145   \item Python Chainmap tutorial \footnote{\url{https://realpython.com/python-
chainmap/}}.
146   \item \footnote{\url{http://conference.scipy.org/proceedings/scipy2021/}}.
147   \item Debug 750 Words for July.
148   \item Gather up grasscatchers items from July.
149   \item Jaras FIUP:
```

Right Pane (PDF Preview):

The PDF preview shows the rendered document. The title bar indicates the file path: 5d1a1603-6e25-489d-b646-... 2 / 5. The document content is titled "Actions required by specific dates" and lists various tasks with checkboxes.

Actions required by specific dates

- ☒ 2 August 2021. Renewal of PyMOL License.
- ☐ 3 August 2021 Travel report in Concur
- ☐ 7 August 2021. Employee General Biosafety Training. You must complete this training by 8/7/2021.
- ☐ 7 August 2021. Employee Laboratory Safety. You must complete this training by 8/7/2021.
- ☐ 4 August 2021. Start contract with Tatyana.
- ☐ 4 August SofSA
- ☐ 6 Toxicology 1100 - 1330
- ☐ 6 August SSRLX 1330-1430
- ☐ 9 August BSEB 272 Methods Course 1300 - 1500
- ☐ 13 August SSRLX 1100-1230
- ☐ 20 August CTS 1100
- ☐ 25 August 25 GPIBS Protien Folding 0900 - 1100
- ☐ 27 August
- ☐ 1-3 September 0700-1500 IUCr Computational Crystallography School.
- ☐ 9 September 0900 -1100 RNA-DNA lecture
- ☐ 20-24 Pittsbupsh Diffraction Society/SSRL Meeting
- ☐ 21 September 10-1400 SSRL Workshop
- ☐ 28 September 1630-1800 Faculty Board
- ☐ 30 September new submission R01 due
- ☐ 12 October Seminar at U of Vermont
- ☐ 15-16 November NIH Study Section

Page 2

PNG file in IPython with imgcat magic

pip install imgcat # requires term2

```
%load_ext imgcat
```

```
%imgcat 1lw9IPython.png
```

```
> Ipython
```

```
Python 3.8.10 | packaged by conda-forge | (default, May 11 2021, 06:39:48)
```

```
Type 'copyright', 'credits' or 'license' for more information
```

```
IPython 7.23.1 -- An enhanced Interactive Python. Type '?' for help.
```

```
In [1]: %load_ext imgcat
```

```
In [2]: %imgcat 1lw9IPython.png
```



```
In [3]: |
```


Install PyMOL and cctbx in a same env

```
conda env create -n pc38 python=3.8
conda activate pc38
conda install -c schrodinger pymol-bundle
conda install -c conda-forge cctbx-base
```

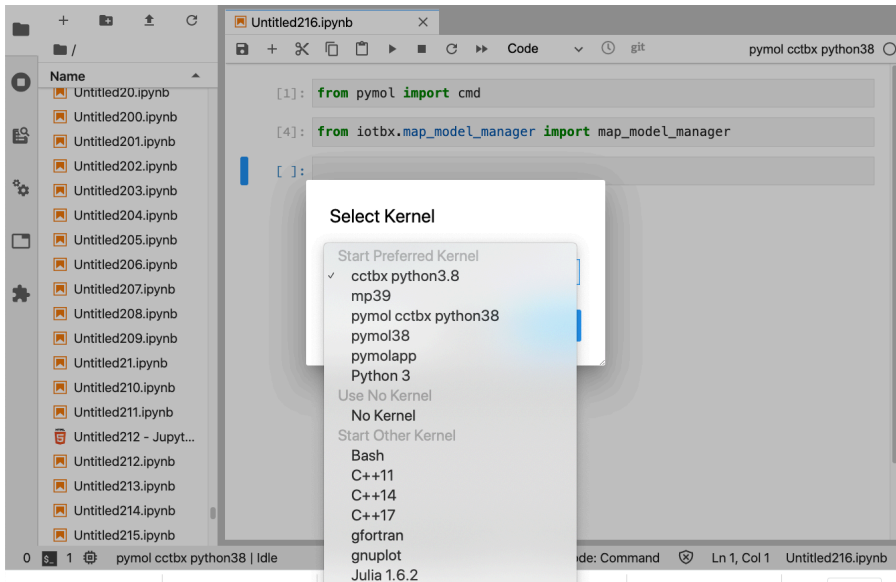
Command to create kernel

```
conda install jupyter
conda install jupyterlab
python -m ipykernel install \
    --user --name pc38 \
    --display-name "pymol cctbx python38"
ls ~/Library/Jupyter/kernels/pc38
logo-64x64.png
logo-32x32.png
kernel.json
```

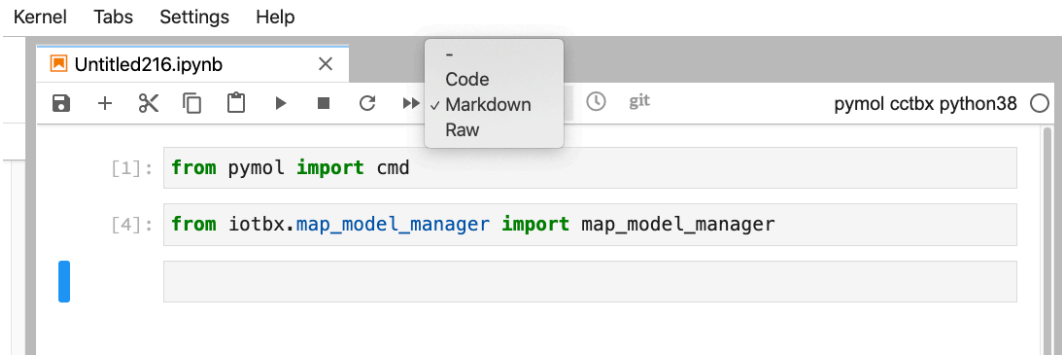
Content of kernel.json

```
{ "argv": [
  "~/opt/anaconda/envs/pc38/bin/python",
  "-m",
  "ipykernel_launcher",
  "-f",
  "{connection_file}"
],
  "display_name": "pymol cctbx python38",
  "language": "python" }
```

Kernel selection



Three kinds of cells in JupyterLab



Left-aligned table

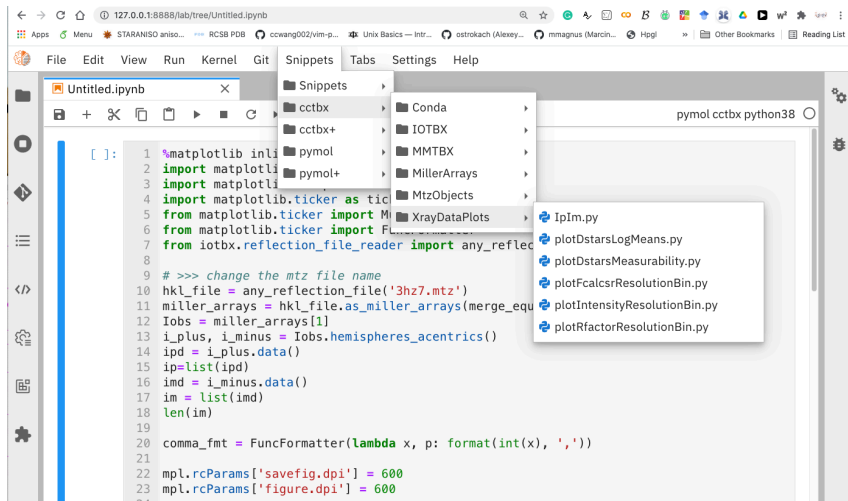
```
In [59]: %%html
<style>
table {float:left}
</style>
```

executed in 5ms, finished 09:27:36 2020-06-17

| Tab trigger | Description |
|----------------|--|
| labelCAs | Label the CA atoms with the Ala333 style format |
| labelMainChain | Label the main chain atoms by resn,resi,atom name. |
| labelWatersHOH | Label waters HOH. |
| labelWatersW | Label waters W. |

| Tab trigger | Description |
|----------------|--|
| labelCAs | Label the CA atoms with the Ala333 style format |
| labelMainChain | Label the main chain atoms by resn,resi,atom name. |
| labelWatersHOH | Label waters HOH. |
| labelWatersW | Label waters W. |

Cascading menu of snippets for cctbx



I(+) vs I(-) plot snippet

<https://github.com/Hekstra-Lab/reciprocalspaceship>

<https://github.com/project-gemmi/gemmi>

3hz7: SirA-like protein from *Desulfitobacterium hafniense*,
Northeast Structural Genomics Consortium, Forouhar et al. 2009.

```
[12]: %matplotlib inline
import matplotlib.pyplot as plt
import matplotlib as mpl
import matplotlib.ticker as ticker
from matplotlib.ticker import MultipleLocator #, FormatStrFormatter
from matplotlib.ticker import FuncFormatter
from iotbx.reflection_file_reader import any_reflection_file

# >>> change the mtz file name
hk1_file = any_reflection_file('3hz7.mtz')
miller_arrays = hk1_file.as_miller_arrays(merge_equivalents=False)
Iobs = miller_arrays[1]
i_plus, i_minus = Iobs.hemispheres_acentrics()
ipd = i_plus.data()
ip = list(ipd)
imd = i_minus.data()
im = list(imd)
len(im)

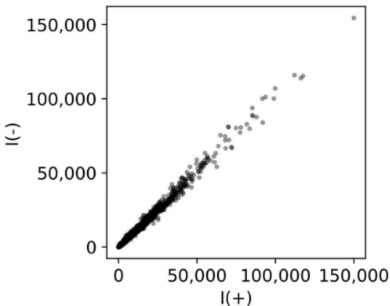
comma_fmt = FuncFormatter(lambda x, p: format(int(x), ','))

mpl.rcParams['savefig.dpi'] = 600
mpl.rcParams['figure.dpi'] = 600

# Set to width of a one column on a two-column page.
# May want to adjust settings for a slide.
fig, ax = plt.subplots(figsize=[3.25, 3.25])
ax.scatter(ip, im, c='k', alpha=0.3, s=5.5)
ax.set_xlabel(r'I(+)', fontsize=12)
ax.set_ylabel(r'I(-)', fontsize=12)
ax.xaxis.set_major_locator(MultipleLocator(50000.))
ax.yaxis.set_major_locator(MultipleLocator(50000.))
ax.get_xaxis().set_major_formatter(comma_fmt)
ax.get_yaxis().set_major_formatter(comma_fmt)

plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
ax.grid(False)

# >>> change name of the figure file
plt.savefig('3hz7IpIm.pdf', bbox_inches='tight')
```



Run pymolshortcuts.py inside Jupyter

<https://github.com/MooersLab/pymolshortcuts>

```
[1]: from pymol import cmd
      from IPython.display import Image

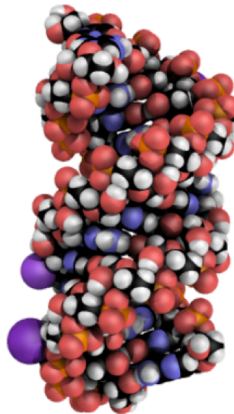
      PATH="/Users/blaine/"
      cmd.do('run /Users/blaine/Scripts/\
PyMOLScripts/pymolshortcuts.py');

      ...
```

The three horizontal dots above are the folded background information that is printed below the cell when the pymolshortcuts.py script is run.

```
[2]: cmd.do("rein;bg_color white;U8;")
      cmd.do("hide nonbonded;A0D;")
      cmd.png("3nd4.png",
              width=300,
              height=450,
              dpi=600)
      Image(filename = PATH + "3nd4.png",
            width=300,
            height=450,
            unconfined=True)
```

[2]:



PyMOL in Jupyter Elyra

File Edit View Run Kernel Git Snippets Tabs Settings Help

</> Code Snippets +

Search...

Eisenberg colorscale 2 F1

H-bond dashes ✓ H-bonds ✓

HOH IPython to display images

Jupyter PE125 PE25 PE33

[python] distance [copy] [run] [edit] [delete]

[python] drawHbonds [copy] [run] [edit] [delete]

[python] hbonddash [copy] [run] [edit] [delete]

```
cmd.do('hide everything, hydrogens;')
cmd.do('hide labels;')
cmd.do('set dash_color, black;')
cmd.do('set dash_gap, 0.4;')
cmd.do('set dash_radius, 0.08;')
```

structuralBiology.ipynb x

Code

```
cmd.do('color green, name CE;')
cmd.do('color brown, resn NAG;')
cmd.do('color red, resi 381;')
cmd.do('remove solvent;')
cmd.do('set specular, 0;')
cmd.do('set ray_trace_gain, 0;')
cmd.do('set ray_trace_mode, 3;')
cmd.do('bg_color white;')
cmd.do('set ray_trace_color, black;')
cmd.do('set depth_cue,0;')
```

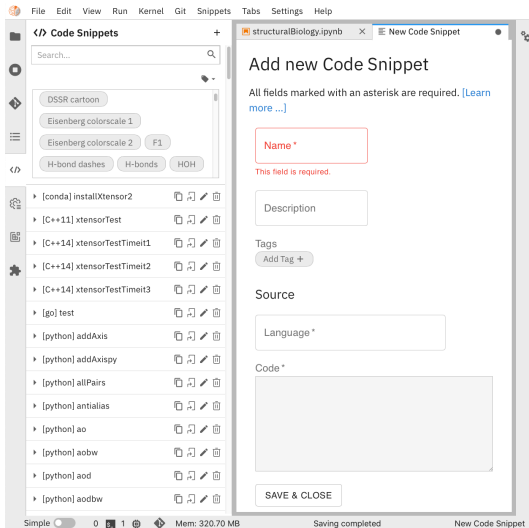
[]:

```
cmd.do('hide everything, hydrogens;')
cmd.do('hide labels;')
cmd.do('# set the color of the dashed')
cmd.do('set dash_color, black;')
cmd.do('set dash_gap, 0.4;')
cmd.do('set dash_radius, 0.08;')
```

[]:

```
cmd.do('hide everything, hydrogens;')
cmd.do('hide labels;')
cmd.do('set dash_color, black;')
cmd.do('set dash_gap, 0.4;')
cmd.do('set dash_radius, 0.08;')
```

Adding a new snippet with Elyra extension



Summary

- Use PyMOL and cctbx with same Python interpreter.
- The kernel.json must have path to correct Python interpreter.
- Jupyter notebooks enable interleaving of code and prose.
- cctbx and pymol snippet libraries on MooersLab GitHub site:

pymol snippets: <https://github.com/MooersLab/jupyterlabpymolpysnips>

cctbx snippets: <https://github.com/MooersLab/jupyterlabccctbxysnips>

References:

- Mooers, B. H. M. (2021) Running CCTBX and PyMOL in the same Jupyter Notebook. Computational Crystallography Newsletter 12, 26-32.
- Mooers, B. H. M. (2021) A PyMOL snippet library for Jupyter to boost researcher productivity. Computing in Science and Engineering. 23(2) 47-53. doi:10.1109/MCSE.2021.3059536.

Acknowledgements

- Mooers Lab: Marina Brown, Francis Acquah
- OUHSC Data Science Workshop

Funding:

- NIH R01CA242845, P20GM103640 (PI Ann West)
- Oklahoma COBRE in Structural Biology (OCSB)
- PHF Team Science Grant with Ian Dunn
- OCAST HR20-002