The Nagai Lab Guide
to making the spliceosome friendly
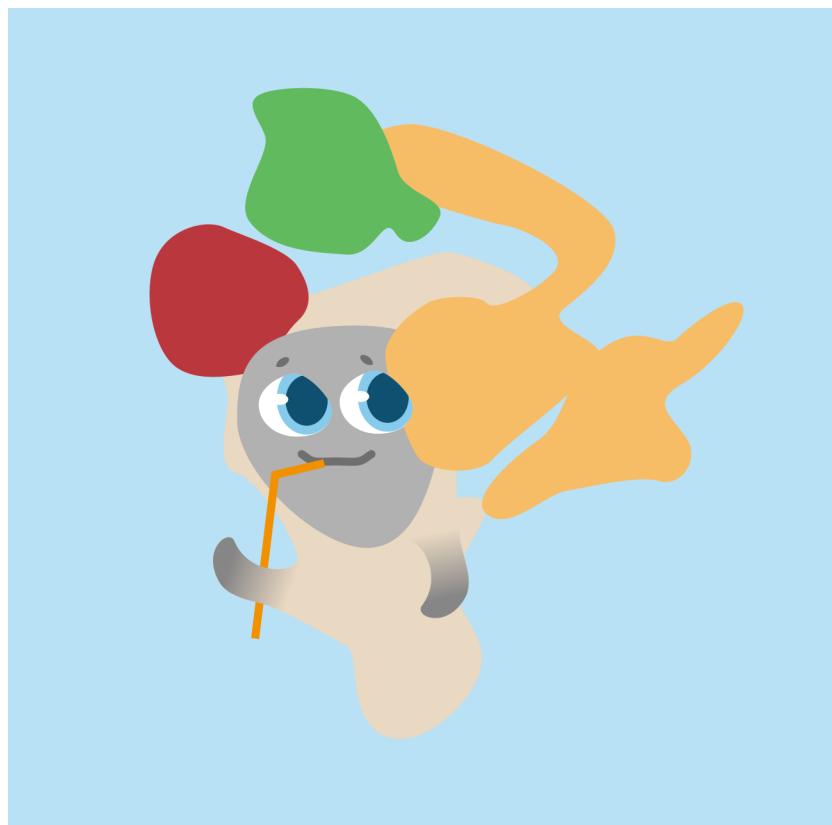
or
How to use PyMOL to explore the structure of the
spliceosome

The spliceosome can look like a rather complicated complex! But by navigating through our spliceosome PyMOL sessions, we hope you will discover for yourself that the spliceosome is really not too complicated to understand. This quick guide will describe how to use PyMOL to explore the spliceosome.

Getting started
You will need PyMOL, which is available for Windows, Mac, and Linux. PyMOL is available either with a license from https://pymol.org/2/ or as an open source version from https://sourceforge.net/projects/pymol/
Note that at the moment, version 2 is only available with the license. All screenshots here will be from PyMOL 2.0 on a Mac, but even if your version looks slightly different all the points will be the same.

We will use the premade PyMOL session for the spliceosomal P complex as an example, which can be downloaded from the Nagai Lab website under Resources (http://www2.mrc-lmb.cam.ac.uk/groups/nagai/resources/). A PyMOL session is a file where the structure has already been subdivided into the individual proteins, and these have been given helpful labels. The structure is also precoloured so that you can easily see the various subcomplexes, and a few selections have been predefined.

Open the P complex session (Psession.pse) by double-clicking in your browser, or using File – Open… in PyMOL.

You should see the whole of P complex. At the top of the PyMOL window you will see a command console. To the right of the main view, you will see a big list of proteins and RNAs. Clicking each name will toggle that molecule's visibility. At the top of the list is the button "all" – this will toggle the visibility of all molecules. At the bottom of the list you will also see some predefined **selections** – you can tell these apart from the molecules because their names are in parentheses.

Command prompt

Viewing window

—All proteins & RNAs

Various RNAs

Various proteins

Some defined selections

—What the mouse does

— What gets selected

Toggle the sequence viewer

Another command prompt
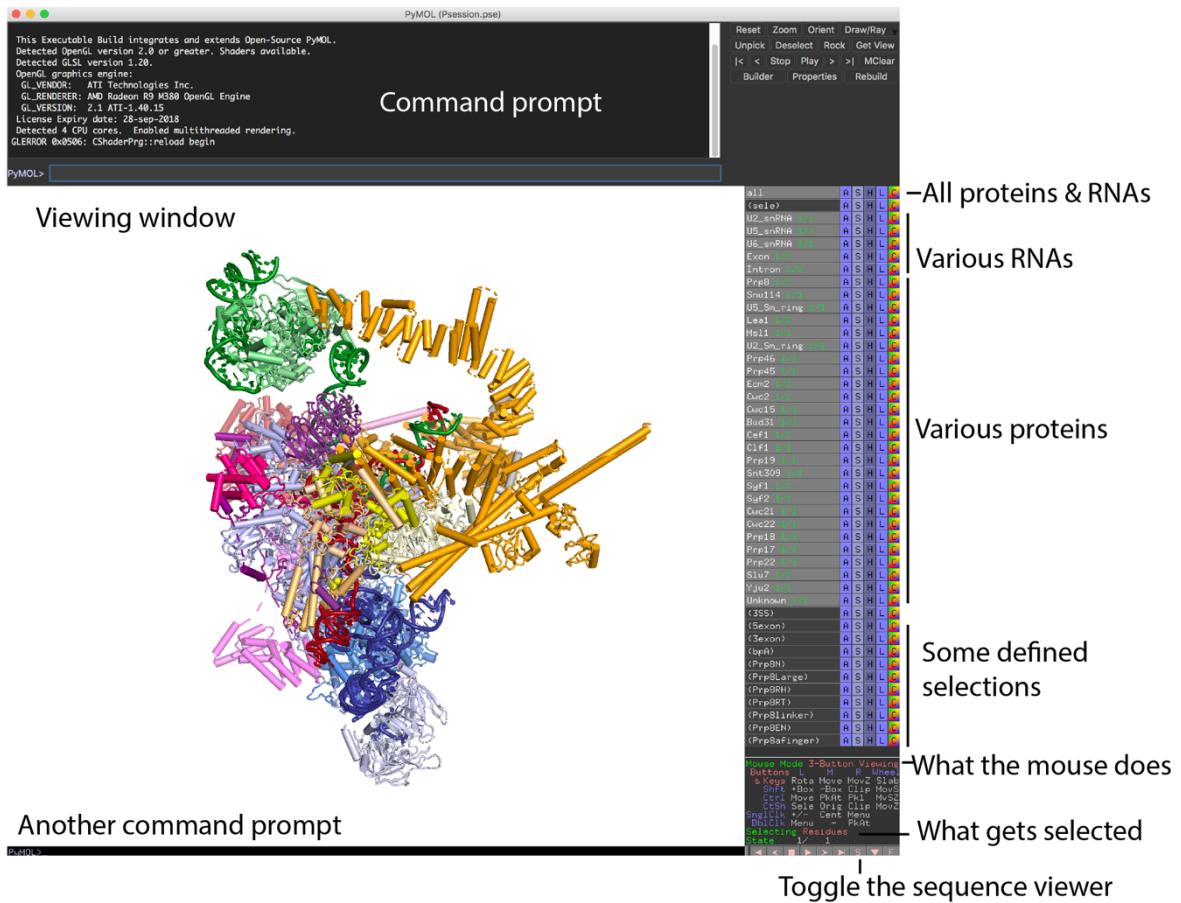
Navigating the molecule
For this step, make sure in the bottom right corner under "Mouse Mode" it says "3-Button Viewing". If it says "3-Button Editing", click on it until it changes.

You will immediately want to explore the structure! Use the mouse to move around the complex. Click and drag with the left mouse button to rotate the view. (Dragging around the edge of the screen will rotate in the plane of the screen, which can be useful sometimes). By default, the view will rotate around the centre of the complex. You can change this pivot point by middle mouse clicking on any point in the spliceosome. The view will then rotate around that point. To zoom, click and drag with the right mouse button, and to pan the view click and drag with the middle mouse button. Take a moment to get used to using the mouse to change the view.

To focus on a protein, middle mouse click on a protein name on the list to the right. For example, if you middle click on "Prp22", it will zoom in on Prp22, and your view will pivot around Prp22. Middle click on "all" to zoom back out and pivot around everything again.

If you scroll with your mouse wheel, you will find a white fog comes towards or away from the screen. This can help show depth. But if you find you can't see anything, keep scrolling until things come into view again. Holding shift and dragging with the right mouse button will adjust the centre of the fog.

Example: focusing on the active site of the spliceosome

The RNA catalytic core of the spliceosome has a fascinating structure that is fun to explore in 3D. But at the moment it's buried under a mass of proteins! Let's first hide everything by clicking on "all" in the right panel. All protein names should then be greyed out, and the whole molecule will disappear. Then to see just the RNAs, in the right panel click on U2_snRNA, U5_snRNA, U6_snRNA, Exon, and Intron. Middle click somewhere in the middle to make sure the view is focused on the centre of the active site. Then zoom in with the right mouse button and rotate the view with the left mouse button until you have a view you like.
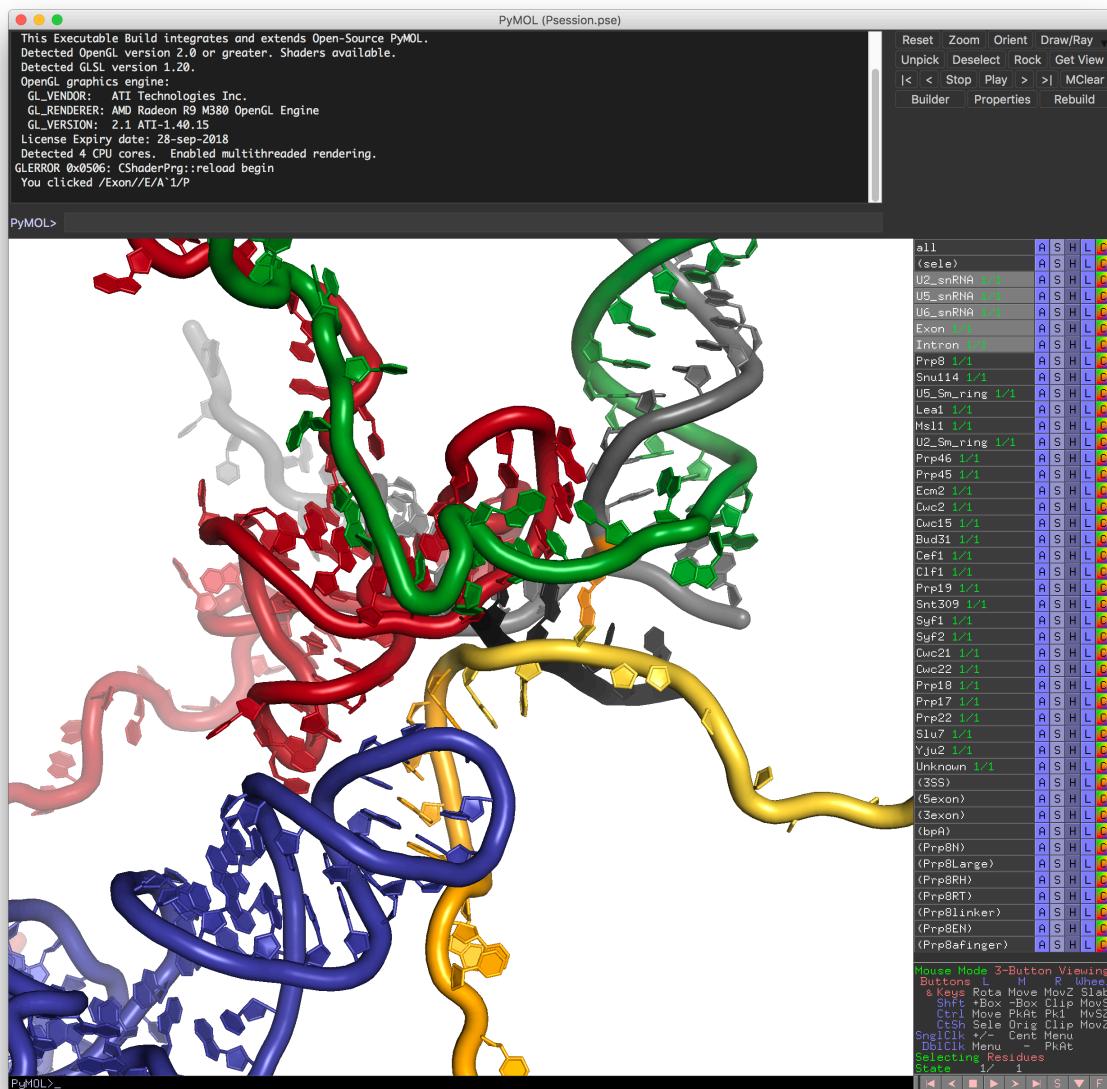
Making selections

For this step, make sure in the bottom right corner under "Selecting" it says "Residues". If it does not (for example it might say "Chains" or "Atoms") click on it until it changes to "Residues".

Now that you're in the active site, you might want to know what all these nucleotides are! In PyMOL we make selections by clicking with the left mouse button. By default this will

select individual residues. If you keep clicking around you will find more and more residues get added to your selection. To clear your selection, just click somewhere in the background. When you make a selection, it appears on the right panel of buttons as (sele). You can also select and deselect your selection by toggling this button.

Click on a nucleotide and look at the console at the top: a prompt will tell you what residue you selected. E.g. If I click on U80 of U6 snRNA, it will tell me

```
You clicked /U6_snRNA//6/U`80/C1'
```

Here, U6_snRNA is the molecule name, 6 is the chain ID in the PDB file, U`80 is the residue name and number, and C1` is the name of the precise atom I happened to click on to select the whole residue.

This method is good for telling you what a particular residue is. But what if you want to find a specific base, like U5 snRNA U96? There are two ways to do this.

The first way: click on the little letter "S" at the very bottom right of the screen (next to some arrows and the letter "F"). This will toggle a sequence view above the viewer, with a grey scroll bar. Slide the bar until you can click on U5 snRNA U96 in the sequence. This will select U96 in the viewer too. If you middle click on U96 in the sequence view, your view will recentre around U96. The problem with this method is if you have all molecules visible, all the sequences will be visible. That's OK at the moment, since we hid most of the proteins, but if they were all visible the sequence view would cover the whole screen!. So it's very difficult to find the residue. (To hide the sequence viewer again, click the letter "S" again)

Click here to turn on the sequence viewer
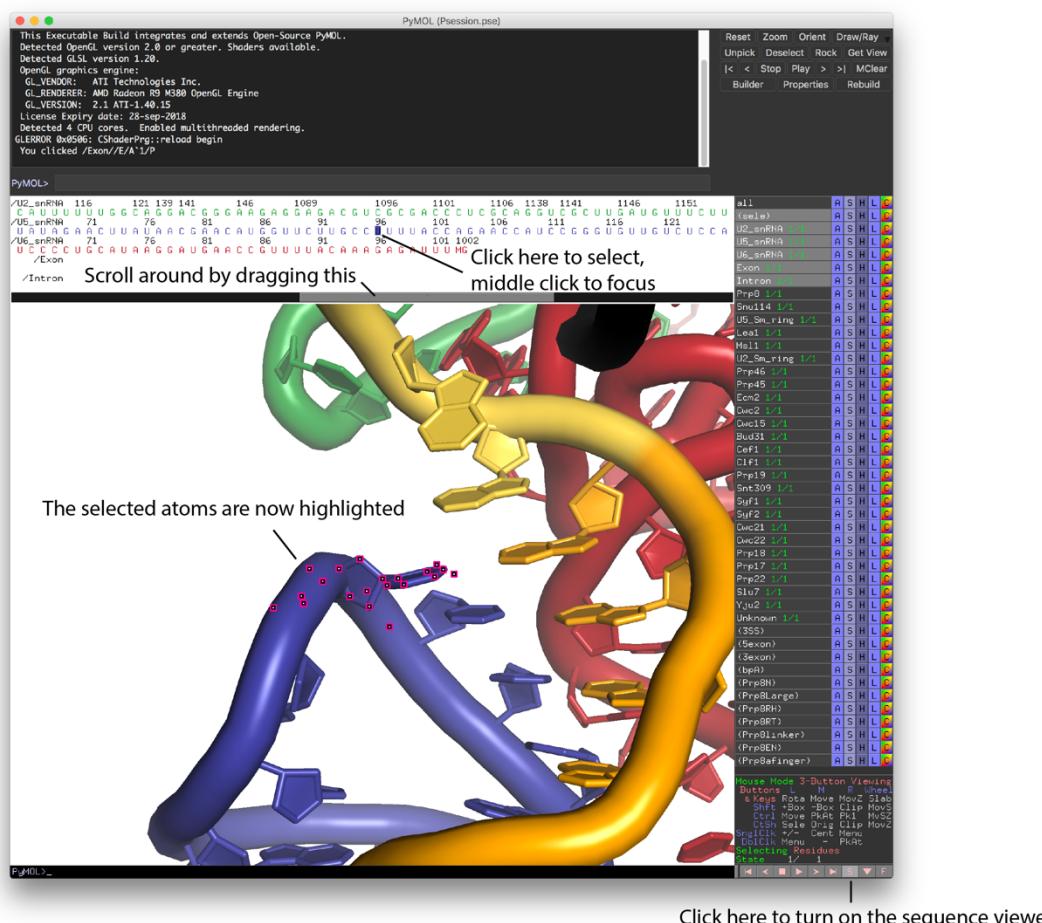
The second way: use the command prompt. Type into the command prompt (after the prompt saying "PyMOL>")

```
select sele, U5_snRNA and resi 96
```

Here "U5_snRNA" is the molecule name, and "resi" stands for "residue ID" (the PyMOL term for residue number). We are giving the selection the default name "(sele)" and it will appear as such on the right panel. So if you click on (sele) you can toggle and untoggle your new selection. This method is good because it is very precise, and can also be used to name selections and select ranges. For example, try this:

```
select U5loop1, U5_snRNA and resi 94-100
```

Now look at the list on the right. At the bottom you should see a new name has appeared "U5loop1". This is because we've defined a new selection called U5loop1, where a **range** of residues has been selected (residues 94 to 100).
Note that if you have your new selection selected, any more residues you select will be added to this predefined selection. So if you want to make a new selection, make sure you deselect everything first!

Changing how the molecules are depicted

So far everything has been in so-called 'cartoon mode.' The data we are visualising here are really a collection of 3D coordinates for various atoms. In cartoon mode, PyMOL is using these coordinates to make a simplified display, e.g. by recognising alpha helices and showing them as cylinders, and by recognising RNA and showing the phosphate backbone as a ribbon. In certain places though we might wish to see a little more detail, so to do this we can adjust the depiction.
For example, let's look at the interactions between the 3' splice site AG dinucleotide of the intron and the branchpoint. These only involve the intron molecule, so click on the other displayed RNAs to hide them. There is a pre-defined selection in this PyMOL session called (bpA). Middle click on this to centre the view around the branchpoint adenosine, and rotate the view to get a better look. We can see that there are non-Watson-Crick base pairs here, but the cartoon view of the bases doesn't show the donor and acceptor groups very well. Click on the four paired residues to select them – or use the sequence viewer to select G1 (the 5'SS +1), A70 (the bpA), A94 (the 3'SS -2) and G95 (the 3'SS -1). Now the selection '(sele)' has four bases.
You will have noticed next to all the selection names and protein names five buttons labelled A, S, H, L, C. These stand for **A**ction, **S**how, **H**ide, **L**abel and **C**olour.
Show and **H**ide are used to show or hide various depictions of the selection. Click the **S** button next to (sele) and click on 'sticks'. This will **S**how the stick representation. Click the **H** button next to (sele) and click on 'cartoon'. This will **H**ide the cartoon representation. The stick representation shows all atoms of the selection, so we can now see the phosphate backbone and the donor/acceptor groups of the bases. To make this even clearer it would be helpful if oxygen, nitrogen, and phosphorous atoms were coloured differently to the carbon atoms. To do this, click on the **C** button next to (sele), hover over 'by element' and

select the first option on the list. This will colour oxygen atoms red, nitrogen atoms blue, and phosphorous atoms orange. (The other options on the list will also change the colour of carbon to various options, e.g. green or blue. We chose the first option because it keeps the overall colour scheme of the selection the same.)


As another example, let's examine how to look at ligand binding. Prp8 has a pocket that binds an inositol hexaphosphate ligand. To find this ligand, hide all molecules and click on Prp8 to show it. Use the sequence viewer to browse to the end of the sequence, where you will see a residue called 'I6P'. (In PDB files, ligands are given short names). Select I6P in the sequence browser, and also middle click on I6P to recentre on it. Again, click on 'S' next to (sele) and use it to show sticks, and then click on 'C' and colour it by element. Adjust the 'view fog' with the mouse scroll wheel to make things clear. To see how the ligand is bound, PyMOL has a really useful tool. Click on 'A' next to (sele) – this will bring up an **A**ctions panel. Hover over 'modify', which will present a list of tools for modifying selections. Hover over 'expand', then click 'by 4A, residues'. This will modify the selection so that it will include all residues within 4 Angstrom of the original selection, i.e. it will select all residues close to the ligand. Now show sticks for this modified selection, and colour by element. Some 'invisible atoms' may have been selected; these belong to Slu7 so also turn on display of Slu7.

When sticks are shown simultaneously with cartoons for proteins, the backbone peptide atoms can look a bit weird as they protrude from the cartoon. To make it look nicer, hide the backbone peptides by typing

```
set cartoon_side_chain_helper, on
```


The Action menu for (sele)
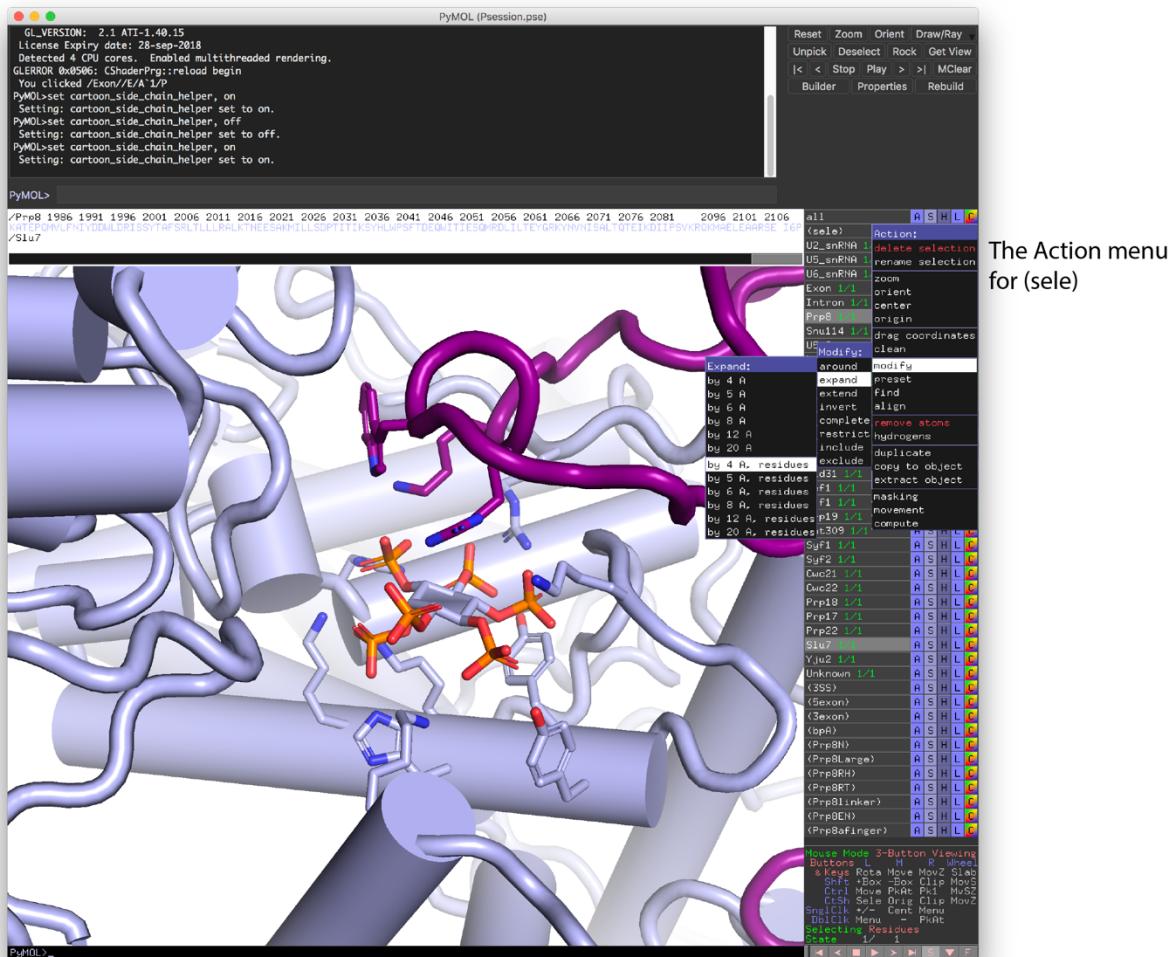
Other display options

PyMOL has many more ways to change how a molecule is depicted, but these are a little bit hidden since the command prompt must be used. Here is a list of some of the more useful commands to play with.
Note that many have the format

```
set some_property, value, selection
```

Where "value" might be a setting like "on" or "off", and selection would be a molecule or part of molecule, or a predefined selection, e.g. "Prp8", "Prp8 and resi 1400-1440", "5exon" etc. If you don't give a selection, the value of the property will be applied to everything.

Command	Note
`set cartoon_cylindrical_helices, off`	Toggles the representation of alpha helices between cylinders and ribbons. We find cylinders are good for large structures because

	they are less complicated, whereas conventional ribbons are good for close ups, particularly when you want to see individual residues.
`set cartoon_tube_radius, 1.2`	Changes the thickness of the RNA backbone when shown as a cartoon – thicker values like 1.2 are good for overall views, thinner values like 0.8 might be helpful for close ups.
`set cartoon_loop_radius, 0.5`	Like tube_radius but for protein loops, it's up to your aesthetic what you like!
`set cartoon_ring_finder, 1`	Changes how RNA bases are drawn in cartoon mode. 0 just shows the backbone, 1 shows the base and ribose, 2 shows the base and a stick for the ribose (like a lollipop)
`set cartoon_ring_transparency, 0.5`	Makes the filled rings in the RNA bases look nice

Making your own PyMOL session

For a large structure like the spliceosome, it is quite difficult to go quickly from a .pdb file downloaded from the PDB to a session with every chain labelled. The PDB file as downloaded will have every residue labelled with its number and a 'chain ID', which will be a letter or number. Some of the chain IDs in our structures are logical, e.g. chain '2' is U2 snRNA, chain '5' is U5 snRNA, chain 'E' is exons, chain 'X' is unknown proteins. Some are necessarily illogical, since there are only so many letters one can use for labelling chains. E.g. chain 'T' is Syf1, and SmB will be chain 'b' if in the U5 Sm ring but chain 'k' if in the U2 Sm ring.

Here are our chain names for the catalytic spliceosomes as an example

A	Prp8	a	Prp18
B	Brr2	c	Slu7
C	Snu114	o	Prp17
D	Yju2	y	Syf2
E	Exon		
F	Cwc25	b	U5 SmB
G	Isy1	d	U5 SmD3
H	Cwc22	e	U5 SmE
I	Intron	f	U5 SmF
J	Prp46	g	U5 SmG
K	Prp45	h	U5 SmD1
L	Bud31	i	
M	Cwc2	j	U5 SmD2
N	Ecm2		
O	Cef1	k	U2 SmB
P	Cwc15	l	U2 SmD1
Q	Prp16	m	U2 SmD2
R	Cwc21	n	U2 SmD3
S	Clf1	p	U2 SmE
T	Syf1	q	U2 SmF

U		r	U2 SmG
V	Prp22		
W	Lea1	s	Snt309
X	unknown	t	Prp19
Y	Msl1	u	Prp19
Z		v	Prp19
		w	Prp19
		x	unknown
		1	U1 snRNA
		2	U2 snRNA
		4	U4 snRNA
		5	U5 snRNA
		6	U6 snRNA

It is because it is so complicated to set up a session that we have provided PyMOL sessions for all of our structures. But you may also want to set up sessions for structures from the other spliceosome structure groups. At the moment, the chain IDs used by different groups are not standardised (although everyone uses chain A for Prp8!) so you will need to look up the PDB database entry itself to find what each chain corresponds to.

For reference, here are the commands used to generate the PyMOL session for P complex. Comments are marked with # and are optional.
After opening the pdb (6EXN.pdb), type (or paste) the following:

```
# set scene #
bg_color white
set ignore_case, off  # So chains a and A will be interpreted as being different

# cartoon rendering #
hide
show cartoon
set cartoon_cylindrical_helices, 1
set cartoon_ring_finder, 3
set cartoon_ring_mode, 3
set cartoon_tube_radius, 1.2
set cartoon_smooth_loops=off
set cartoon_loop_radius, 0.5


# some custom colours, for aesthetic only #
set_color lightgreen, [144, 238, 144]
set_color lemonchiffon, [255,250,205]
set_color cornflowerblue, [100,149,237]


# chain colors #

# snRNA #
select chain 2
extract U2_snRNA, sele
color forest, U2_snRNA

select chain 5
extract U5_snRNA, sele
color deepblue, U5_snRNA

select chain 6
extract U6_snRNA, sele
```

```
color firebrick, U6_snRNA

select chain E
extract Exon, sele
color brightorange, Exon
color yelloworange, Exon and resi 1-30

select chain I
extract Intron, sele
color black, Intron
color grey40, Intron
color black, Intron and resi 91-95

# U5 #

select chain A
extract Prp8, sele
color lightblue, Prp8

select chain C
extract Snu114, sele
color cornflowerblue, Snu114

select chain b or chain d or chain e or chain f or chain g or chain h or chain k
or chain j
extract U5_Sm_ring, sele
color bluewhite, U5_Sm_ring


# U2 #

select chain W
extract Lea1, sele
color lightgreen, Lea1

select chain Y
extract Msl1, sele
color lightgreen, Msl1

select chain k or chain l or chain m or chain n or chain p or chain q or chain r
extract U2_Sm_ring, sele
color palegreen, U2_Sm_ring


# NTC/NTR #

select chain J
extract Prp46, sele
color lemonchiffon, Prp46

select chain K
extract Prp45, sele
color lemonchiffon, Prp45

select chain N
extract Ecm2, sele
color tv_yellow, Ecm2

select chain M
extract Cwc2, sele
color wheat, Cwc2

select chain P
extract Cwc15, sele
color wheat, Cwc15

select chain L
extract Bud31, sele
```

```
color lightorange, Bud31

select chain O
extract Cef1, sele
color brightorange, Cef1

select chain S
extract Clf1, sele
color brightorange, Clf1

select chain t or chain u or chain v or chain w
extract Prp19, sele
color brightorange, Prp19

select chain s
extract Snt309, sele
color brightorange, Snt309

select chain T
extract Syf1, sele
color brightorange, Syf1

select chain y
extract Syf2, sele
color brightorange, Syf2


# Splicing factors #

select chain R
extract Cwc21, sele
color violet, Cwc21

select chain H
extract Cwc22, sele
color violet, Cwc22

select chain a
extract Prp18, sele
color hotpink, Prp18

select chain o
extract Prp17, sele
color violetpurple, Prp17

select chain V
extract Prp22, sele
color firebrick, Prp22

select chain c
extract Slu7, sele
color deeppurple, Slu7

select chain D
extract Yju2, sele
color pink, Yju2

# UNKNOWN #

select chain X
extract Unknown, sele
color grey90, Unknown


# Defining some selections #

select 3SS, Intron and resi 91-95
select 5exon, Exon and resi \-14-\-1
```

```
select 3exon, Exon and resi 1-19
select bpA, Intron and resi 70
color orange, bpA

select Prp8N, Prp8 and resi 1-870
select Prp8Large, Prp8 and resi 871-1827
select Prp8RH, Prp8 and resi 1828-2106
select Prp8RT, Prp8 and resi 871-1375
select Prp8linker, Prp8 and resi 1376-1652
select Prp8EN, Prp8 and resi 1653-1824
select Prp8afinger, Prp8 and resi 1583-1610

# show metals #
sel name ZN
show spheres, sel
color yellow, sel
set sphere_scale, 1.5

# adjust depiction #

set cartoon_gap_cutoff, 30
set cartoon_gap_cutoff, 0, Unknown
viewport 2000,2000

# set front view
set_view (\
     0.674891770,   -0.352280945,    0.648389578,\
    -0.439635426,   -0.897670925,   -0.030115895,\
     0.592642784,   -0.264727056,   -0.760699689,\
    -0.007897735,    0.004491180, -1026.595825195,\
   277.550354004,  267.541595459,  277.006958008,\
   827.363769531, 1226.079711914,  -20.000000000 )
```

Conclusion

Hopefully you are now confident to explore the spliceosome using our PyMOL sessions. We encourage you to have PyMOL open when reading spliceosome structure papers, as being able to interact with a structure in 3D is often a lot clearer and more enlightening than looking at 2D figures. The spliceosome also has many interesting structural features that are not discussed in the structure papers, and these are very fun to explore and discover for yourself.

Please let us know if you have any questions or feedback about this tutorial – we are always happy to help. You can either contact Kiyoshi Nagai directly at kn@mrc-lmb.cam.ac.uk or Max Wilkinson at mwilkin@mrc-lmb.cam.ac.uk

Have fun!