Single molecule data analysis protocol- Audra version

Exactly how you do the data analysis will vary depending ont eh experiment and what you want to get out of it. In particular if you are doing FRET vs just colocalization with DNA, the processes are quite different after the traces\* are made.

\*will explain below what the traces are.

The first thing you want to do for every experiment is determine the protein colocalization with DNA. This is done as follows:

**Step 1: Mapping.**

We need to be able to determine when Red/Green/and Blue spots are co-localized. It turns out that we are working at a quantum level where diffraction of different wavelengths of light through the glass optics of the microscope will make it look like the different colors are in slightly different physical places on the slide, when really they are emanating from the same spot. These differences in diffraction are extremely sensitive to tiny changes in the optics which can be due to temperature in the room, the metal relaxing, expanding, things loosening slightly, or being tightened by Larry, etc. The result of the fact that the diffraction is so sensitive to tiny changes is that we need a way to account for it anew, during each experiment. We do this by creating a “map” between each of the colors by using a DNA with all 3 colors present. Note: in practice you can get away with one mapping per week, but we tend to go once a week or less, so effectively for us, we have to do a mapping new each week. However, if you go on 2 days back to back, you only need one mapping.

The mapping procedure is described in the accompanying video.

**Step2: picking DNA spots**

We want to know when protein is colocalized with DNA, so we need to select the DNA molecules and them map those regions onto the protein fields of view. This process is partially automated but need some manual input to curate the DNA spots (sometimes a 2 DNA molecules will be too close and look like a barbell shape, and the human eye can distinguish this, but the computer cannot).

This process is described in a video.

**Step3: Making a driftlist:**

The slide wiggles around during the experiment, and for the computer to be able to track colocalization, we need to account for that movement. The driftlist is calculated using quantum dot beads attached to the surface of the slide.

Process described in a video.

**Step4: integrating protein intensity at each DNA spot over the course of the experiment.**

We want to know when each protein colocalizes with each DNA molecule. To determine this, we draw a 5x5\* square of pixels around each DNA and the computer will measure the fluorescence intensity over that square at each frame of the video.

Process described in video protocol.

Notes: we typically do this using a 5x5 square of pixels, but Paggard has gotten better results by using a 4x4 square—because there is less background noise in a smaller area. The tradeoff is that if the DNA is wiggling around and is longer than 4 pixels, it could wiggle out of the window. Larry does 3x3, but I think that is probably too small for our 1.3kB DNA. As far as I know, Meg and Simina always used 5x5 squares, but they did not leave documentation, so I can’t be sure.

Note2: for many DNA molecules we have the computer determine the fluorescence intensity using linear interpolation because that is fast (takes ~3-5min).

One could also measure the intensity using a 2d Gaussian— this is a more detailed and accurate way to measure the fluorescence intensity, but it is computationally difficult and is very slow (would take hours for many DNA molecules). However, if we have access to an MIT super computer cluster like they have in the Whitehead, and/or if you wanted to let the process go overnight, larry thinks it is objectively better, it’s just too slow to be practical for them (and for us on our personal computers/ or the current lab computers).

A 2d Gaussian is what we use for the driftlist because that is only 1 bead, instead of hundreds of DNAs, so the computer can handle the task in just a few minutes.

**Step 5a: using the spotpicker program to determine protein-DNA binding events—when a binary output (bound vs not bound) is what you want.**

Use the spotpicker when you want a binary answer of when a protein is bound to the DNA or not bound. This, however, tells you nothing about the intensity of the protein bound, so it is not helpful for FRET, or for counting multiple events on the same DNA. It is very useful for looking at Cdt1, Cdc6, and ORC where we do not expect multiple proteins to colocalize to the DNA at once.

Process described in video protocol.

**Step 5b: determining EFRET:**

For FRET experiments we need to determine the difference in fluorescence intensity between the acceptor and the donor.

Video protocol to come in future.

Written protocol:

* 1. Manually identify regions to analyze: i.e. one red dye, one green dye only. Regions where FRET is possible, but wont be confounded by additional molecules.
     1. Make a table in excel with #Aoi, start of region, end of region

* 1. Create a new data structure with the detrended traces.
  2. Run Paggard EFRET program
  3. Run Paggards mlx Efret plotting program

Step 1: background AOI subtraction

How to do background Aoi subtraction:

* 1. Select your AOIfits file and load into imscroll
  2. Hit the no spots box
  3. Makes sure you are in same frame as made initial aois.
  4. Hit the background button and then draw aoi grid
  5. MUST FIRST do remove close AOIs
     + This must be done first to assign the background AOis to the right DNA spot. For whatever reason, if you do 6 before 5 the program will give an error.
  6. In the dropdown do "retain close aoi" with a radius of ~12
  7. Then int linear interp fit the background aois
  8. Then load you background fits into the plot visualizer window in the lower viewer
  9. Load your actual DNA-aoifits into the upper viewer
  10. Press "plot" on int.aoi for both top and bottom! (must do this, idk why tho)
  11. Then within that drop down go to the bottom selection "make background corrected aoifits"

Step 2: Make EFRET matrix:

NOTE: this program makes some assumptions because it was written specifically for a ring closing fret experiment. For ring closing, if both fluorophores are not bleached, then they necessarily arrive at the same time. So for every event where both colors are present, we chose the frame start as the green frame that was closest. For something where you have FRET between two different molecules and the red could arrive later, then this code may need to be modified.

1. Run AA\_running\_mutlicumulativeInterval\_program.m with your RED cia as "cia" and your green cia as "cia2" to define regions where both fluorophores are present
2. Run AA\_Getting\_FRET\_Matrix\_from\_mcia.m to extract a matrix of events and times
3. Make EFRET matrix
   1. Make sure you have Defined P1=both\_preset\_FRET\_input-- I added this to the previous program so it should be done as long as you have just run that
   2. Load and eval Aoifits\_bk\_subtracted for your FRET and Green channels and define aoifitsGreen and aoifitsRed

[fn fp]=uigetfile

eval(['load ' [fp fn] ' -mat'])

aoifitsGreen=aoifits

[fn fp]=uigetfile

eval(['load ' [fp fn] ' -mat'])

aoifitsRed=aoifits

aoifits.data %to look at it

aoiftis.dataDescription %to se what each column is

dat=draw\_aoifits\_aois\_v1(aoifitsGreen,'y');

datG=dat

dat=draw\_aoifits\_aois\_v1(aoifitsRed,'y');

datR=dat

1. Run: AA\_EFRET\_matrix\_ring\_closing.m
2. If want to look at traces run:

% Cycle through all aoi plots

for indx=1:58 %or however many events you have

figure(15);hold off;plot(datR(:,2,indx),datR(:,8,indx),'r');shg

hold on

figure(15);plot(datR(:,2,indx),datG(:,8,indx),'g');shg

title(num2str(indx))

pause

end

Step 3: run Jeff's program "fret\_hist\_2d\_run17\_EPL\_12\_6\_17\_008\_012.mlx"

1. Load in your EFRET variable if it is not already in the variables window
2. Then click each section or run all on Jeff's program

**Step 6: Plotting your data**

* + 2. Survival curve
       - Use this plot to compare differences in dwell-times, regardless fo number of events. Basically a survival plot shows you the difference in average dwell times for the events that happen, and ignores whether there is a difference in the frequency of events.
       - AA\_survival\_plot.m –compares 2 data sets
    3. Probability density function
       - Use this type of plot to evaluate your dwell times, normalized to # of events and bin widths. This comparison will show you both a difference in dwell times themselves (like if they are shorter on average) and if there is a difference in the likelihood of different events (ie. if all event now switch to just a short form as opposed to being a 2-part distribution of long and short events)
       - AA\_making\_overlapping\_histograms.m –this uses equal bins, but you can change the number of bins
       - AA\_overlapping\_Jeff\_histograms.m –this makes unequal bins that each contain the same % of the data—you can choose how many bins to use
    4. Frequency plot
       - This plot shows both the duration of events, and also the frequency of events at different times—Ie maybe you have the same lengths of events, but there are way fewer events overall in your experimental condition.

Audra\_frequency\_plot.m –compares 2 data sets

Audra\_frequency\_plot\_four\_cias.m –compares 4

Audra\_frequency\_plot\_six\_cias.m –compares 6

These programs will take the cia(s) you have generated from your spotpicker intervals and graph the frequency plot.

* + 1. Time to first binding
       - time\_to\_first\_binding\_fraction\_of\_DNA.m –compares 2 data sets
    2. other operations:
       - extracting Mcm/Cdt1 events where the proteins arrive simutaneously. This should also work for other pairs of proteins, if useful.
         * AA\_silmutaneous\_arrivals\_cias.m (ignore the misspellings)
       - Placeholder
    3. placeholder

**Step7: fitting the data to get rate constants:**

1. First you need to analyze no-DNA spots to ascertain background/nonspecific landings.
2. To fit frequency plots where the data looks like a double exponential (ie on a semi log plot it has 2 slopes):

Audra\_max\_likelihood\_dwelltime\_curve\_fitting.m

Audra\_MLfit\_2exp\_same\_plot.m

Audra\_MLfit\_2exp\_same\_plot\_ORC\_alone\_only.m

Note: I am currently not sure what the difference is between these 3 programs. Will edit once I have figured that out.

Which are derived from codes from Larry called:

Expfalltwo\_mxl.m

Expfalltwo\_mxl\_correction.m

Expfalltwo\_mxl\_with\_background.m

**Step8: photobleaching analysis:**

Not sure how to do this yet. In general, you want to generate a plot of rate vs relative laser intensity, and then find the y intercept. But I am still figuring out hwo to actually do this in a practical sense.