Objective: Develop a pipeline for critical thinking and data exploration

Exercise 1: Synthesize reading on Src and Hsp90 thus far

Any literature dive is disorganized and meandering at first, making it difficult to remember and synthesize everything you've read. I want you to make a summary for everything you've read so far similar to a one page cheat sheet you would make for a test in class. This can be presented in any format (document, slideshow), it must be organized like how you would write a note sheet you would be allowed to use on a test in class (bullet points of info). As a bonus, this summary should be organized for your self-reference, but also organized enough for someone else to review and gain a general understanding (as if you were presenting the background to your project). If you encounter other questions while putting together the summary, this is probably a great point/note to include in your summary! It also likely doesn't require a whole other paper to answer, maybe a Wikipedia summary of the biochemical process will suffice.

An example of things your summary should include (but is by no means a comprehensive list):

- How does Src activation work?
    - What are the differences between v-Src and c-Src activation?
- How does Hsp90 chaperone clients?
- What are the different conformations/states of Src activation?
- What mutations/important positions keep showing up in the literature?
    - How do these positions relate to the different conformations/states of Src?
- Anything else you think adds to your understanding and is important to note in any analyses moving forward

Exercise 2: Document data analysis process

Data analysis is generally an organic process, but rigorous documentation of your thought process can develop your critical thinking skills. You will have identified important positions from your reading summary and I have identified interesting regions via a structural analysis. You can use both to help direct this process.

Example of how VN would go about analyzing a SASA~activity plot

Definition: SASA is defined as the solvent accessible surface area. Residues with low SASA are buried away from the solvent by other regions of the protein and often involved in sidechain packing and stabilizing the fold of the protein. Residues with high SASA are interacting primarily with solvent, and generally less important to maintaining the intrinsic stability of a fold.

General questions: Can SASA coupled with activity data support what we already know about Src regulation? Does it identify new regions on the structure that we didn't previously think was important for Src regulation? Do positions identified through the literature or our structural analysis have a significantly different SASA signature than other positions?

General expectations I expect a SASA~activity plot to look like a triangle where areas of low SASA have a large range of activities for the different mutations observed in the DMS because it is interacting with other parts of the protein and thus mutations can disrupt these interactions. I expect areas of high SASA to have a smaller range of activities because it is not interacting with as many other residues, so there is less potential for mutations that compromise the fold. *I should keep in mind though that folding DOES NOT equal activity though. Loss of folding will lead to a loss of activity, but a loss of activity does not necessarily mean the fold has been lost (the active site could be disrupted, but fold left intact)*

1. First, I'd make the SASA~activity plot to confirm/deny my general expectations
    a. If it doesn't, can I explain why based on my general understanding of biochemistry?
        i. Ex: Maybe it doesn't look like a triangle, but looks like a rectangle. This suggests that both buried and exposed residues are important for activity. I noted earlier that folding != activity. Maybe the high SASA positions interact with ligands and substrates in a specific manner that is compromised when we mutate these positions, and that's why we don't see the triangle I expected.
    b. We won't always be able to come up with an explanation. Maybe the literature will have an answer. If we don't have one right away, we can stick a pin in it and come back later. Try to have some kind of explanation (even a flimsy one) before moving on.
2. Okay, the SASA~activity plot looks like I expected, now I want to see the difference between open/closed SASA plots. I'll make the same plots as before, but on the x-axis I'll have diff(open_SASA - closed_SASA). (Maybe abs(diff(open-closed))? This might be easier to interpret...)
    a. Expectation: I expect positions with a large abs(diff(open-closed)) to have a large range of activity values for the mutants because these positions regulate Src's change between open and closed conformations
    b. If it doesn't look like I expect, maybe we have too many points and the existing trend is to weak to see amidst all the points
    c. If it does look like I expect, great! Anything useful I can get from this plot? Are there outliers from the general trend of the plot? What are these positions? Are they at well described regulatory sites? (activation loop, DFG motif, alphaC helix, etc…)
3. Let's narrow it down to the most extreme cases to see if we can find a trend without noise. Do the abs(diff(open-closed)) of the positions I've identified through my literature search and structural analysis look different from other positions in the protein?
    a. Expectation: I expect these to have the greatest change in SASA because they are key players in Src's regulatory machinery.
    b. If it doesn't look like I expect, did I use the right structures for this analysis? (i.e. did my Src structures include a bound ATP when calculating SASA?)

    c. If it does look like I expect, great! Maybe I can use this to support/explain the process of Src activation in c-Src vs v-Src, as described in the literature
4. Summarize my findings (while referencing the correct plot) in my lab notebook

Exercise 3: Wrestle with VN's hypothesis for Hsp90 inhibited mutants
No one knows the answer, not me or Doug or Dusty. Based on what you've read and what you've seen in the data, do you agree? If not, why? Whatever your answer is, make a case for it.

VN hypothesis: Hsp90 clients are conformationally uncontrolled (while c-Src conformations are strictly controlled by ATP binding and A-loop unfolding). Clients can be inhibited by Hsp90 refolding if mutations disrupt N-lobe folding but leave regulatory interfaces intact (alphaF pocket and SH2-SH3 linker interface). These clients are refolded and monitored by Hsp90 and the regulatory interfaces can then restore this mutant Src to its closed conformation. When Hsp90 is inhibited, these mutants remain in their open/active conformation, leading to this "Hsp90 inhibited" behavior we observe.

When you make your case:
- Summarize my hypothesis and have a clear explanation for it
- If you agree, provide literature and instances in the data that support this hypothesis
- If you don't, provide literature and instances in the data that do not support this hypothesis
    - Provide an alternative explanation that is supported by the literature and data that you used to contradict this hypothesis