I’ve included a few points here but most comments are included in the PDF file (d5\_ejf.pdf).

The manuscript is needs some work to make it clear. There are many details missing or not included in the appropriate sections.

The discussion section is a brief and could be extended more with emphasis on what was learned in this study compared to the literature. This is true for both the computational and the PDZ parts. I can help more on the PDZ meaning.

Several figures need some additional description (see text).

Below are some specific points concerning PDZ structures and the 9 figures.

1. What was the rationale for choosing the 8 PDZ domains and the particular structures (i.e. PDB code)? Several structures were determined with a peptide so the apo state is not actually represented. I think this is true in all the structures. This may not matter for those PDZ domains where it is known that the apo and complex structures do not change much. However, some PDZ domains may have large deviations on the backbone between the apo and complex and these designs may be less reliable (or at least biased). You might compare the PDB structures you used against the “real” apo structure to see different they are. You might be able to argue that the backbone changes are small enough that the changes probably don’t influence the designs much.

It seems that for the initial design (where almost all residues were allowed to change) only apo structures were used. That is, the ligand was removed prior to the simulation? Is this true? This point should be made explicitly in the text. As above, you might be able to argue there is not a large change between the “real apo” and the complex to justify your choice.

Table 1 should include which PDZ domain in the protein the structure corresponds to. For example, the GRIP PDZ domain is PDZ6 (see below).

Below are the PDZ domains and whether or not they have a ligand in the structure. Some of these structures have a ligand fused to their C-term and therefore the structure is actually of a complex. You only see the complex if you look at several copies of the asymmetric unit. Thus, these structures are not of an apo protein and the structure may have some packing issues or backbone structural changes compared to the authentic apo protein. We have seen this in CASK, when you compare 1KWA with our apo structure.

- NHERF PDZ1, 1G9O:DEQL (C-term from other asymmetric unit is interacting with pocket; packing issue?)

- INAD PDZ1, 1IHJ: TEFCA (disulfide in C-term from peptide, known to be artifact)

- GRIP PDZ6, 1N7E:QPASS (C-term bound structure from asymmetric unit, packing issue?)

- Syntenin PDZ2, 1R6J:IMPAF (C-term AF from other asymmetric unit is interacting with pocket; packing issue?)

- DLG2 PDZ2, 2BYG:PTTIY (C-term from other asymmetric unit is interacting with pocket; packing issue?)

- PSD95 PDZ3, 3K82 (internal E from other asymmetric unit is interacting with pocket; less of an issue?)

- CASK PDZ, 1KWA:SYREF (C-term from other asymmetric unit is interacting with pocket; our structure with no ligand shows a difference in the a2 helix orientation)

- Tiam1:SDC1, 4GVD (our structure has a mutation at Q844H, probably not an issue; this is not the apo protein)

2. Figure 1

- A) The yellow text (Cask) is difficult to see. You might refer to the other PDZ domains by protein name rather than PDB code.

- A) Labeling all core positions would be useful as a reference for Figure 3.

- B) What does the gray, dotted line represent? How is this line different than the black, dotted line? What is the criteria for “more distant neighbors”.

- B) Does the length of the line symbolize anything? This sometimes represents how related it is as in an evolutionary tree.

- B) In the figure, proteins are linked by a line that contains two numbers. Presumably, one number is the rms deviation (in Å) and the other is identity scores (%ID). This is not explicitly stated in the legend.

- B) You might use the protein name rather than PDB code.

3. Figure2

- Your sequence alignment is a bit different that the one Young Joo derived. His is a structure-based alignment derived from ~80 carefully chosen PDZ/ligand complexes. Not sure how different the results would be if you used a slightly modified alignment. I suppose it would matter only if the amino acid frequencies differed substantially at many positions.

4. Figure 3 & 4

- Labeling all core positions in Figure 1 would be useful as a reference here. Similarly, another figure with the surface residues labelled might be useful.

5. Figure 5

- You might use the protein name rather than PDB code.

6. Figure 6

- Does core or surface designation mean that only those core (or surface) residues were changed in the simulation or that the similarity score was calculated based only the core (or surface) residues?

7. Figure 7

- It might be useful to divide the figure into three panels (A-C).

- Were all residues but 11 “specificity” residues and G and P were allowed to change? This should be explicitly stated in the Methods. I would helpful to show in the WT and 5 designed sequences which positions were held fixed.

- It might be useful to show the relative stability energy next to each sequence.

- Were any of these sequences determined from the Rosetta design experiments? Were any Rosetta sequences tested beyond the similarity score?

- What is the sequence ID between the 5 models and WT? What is the closest native PDZ domain to each designed sequence? Is Tiam1 the closest relative?

- The structures are shown for the apo proteins only? How do the apo proteins compare to the complexes? It would be useful to let Young Joo and I look at the apo and bound designed structures.

8. Figure 8

- A) The legend indicates that this figure is for Tiam1. However, the sequences shown do not seem to be related to Tiam1 WT. It would be useful to have the WT Tiam1 sequence included for reference. The other PDZ domains could be included in the Supplemental Data.

- I think the symbol (delta) should be on y-axis.

- The color coding in not entirely clear. Shouldn’t the coloring be a gradient like in B? I only see brown, orange, yellow and red. The meaning of red is not described in the legend. What is the meaning if the residue is not colored? If it means it is not changed, than that residue should correspond to Tiam1. It does not seem to.

- B) It would be helpful to show a cartoon of the secondary structure on the 3D image (transparent) for reference. The color scale could be shown here too (related to the point above).

9. Figure 9

- A) The structures (QM and WT) are shown in similar colors making it difficult to tell them apart. I would change the colors to make increase the contrast. Why designate the 4-postion mutant as QM when until now it has been referred to as LKLL. I would stick to one convention.

The 4-positions being designed and their labels are difficult to read. The residues to design could be colored differently to highlight them.

- B) You might include the WT and QM sequence under each logo. For instance, under wildtype backbone you could label it: L911, K912, L915 and L920. For the mutant (QM) you could label it: M911, E912, F915 and V920.

- C) A color scale could be included. What does white represent? You might include 911 in each panel on the bottom and then Apo, SDC1, Caspr4 on top of each plot. The X and Y axes are 912 vs 911, not 912 vs Apo, etc.

- Is the 911/912 plot representative of the correlation between other residues? Why was this particular pair selected for display? This information might be included in the text.