Identify Impactful Mutations Through Rigidity And Energy Minimization Analyses

Loc Truong
Loc.Truong@wwu.edu
Western Washington University

Sean Mcculloch
mcculls5@wwu.edu
Western Washington University

Izzi Piper
piperi@wwu.edu
Western Washington University

Sydney Kaster
kasters@wwu.edu
Western Washington University

Macallan Thorndike
thorndm@wwu.edu
Western Washington University

Kevin Doan
doanm3@wwu.edu
Western Washington University

Matthew Lee
leem53@wwu.edu
Western Washington University

Filip Jagodzinski
jagodzf@wwu.edu
Western Washington University

ABSTRACT

Protein-protein interactions are critical to cellular signaling and trafficking. The PDZ domain is a common module that is involved in such interactions, where binding of the domain to the C terminus of the target facilitates signaling. However, the characterization of targets can be difficult due to transient low-affinity interactions, overlap in targets between different PDZ domains, as well as the fact that a single PDZ domain can recognize multiple targets [6]. Performing wet-lab experiments to better understand the selectivity determinants of PDZ domains through the exhaustive mutation of target residues involved in binding is time consuming and not cost effective. This work investigates the impact of single and double mutations of a target ligand on complex formation using in silico methods, energy minimization, and rigidity analysis.

KEYWORDS

rigidity analysis, energy minimization, PDZ, ligands, mutations

ACM Reference Format:
Loc Truong, Sean Mcculloch, Izzi Piper, Sydney Kaster, Macallan Thorndike, Kevin Doan, Matthew Lee, and Filip Jagodzinski. 2020. Identify Impactful Mutations Through Rigidity And Energy Minimization Analyses. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Cellular function is dependent on proper cellular signaling and trafficking, which are facilitated by protein-protein binding interactions. PDZ domains are involved in such protein-protein binding interactions and are found in high abundance throughout various species. PDZ domains regulate processes such as ion channel signaling, transport, and signal transduction by binding to short amino acid motifs in the target protein, typically at the C terminus [3]. Analysis of the short motifs that the PDZ domains bind to led to PDZ domains being sorted into three classes, though later work demonstrated that PDZ domains bind promiscuously to more targets than previously thought [2]. Thus, much remains unresolved about the scope and identity of the binding motifs of the target protein.

Here we present computational efforts to study how complex formation is disrupted when the target motif is mutated based on the crystal structures of eight ligand-bound PDZ domains. We introduced exhaustive single and double point mutations into the ligand in silico. We then performed energy minimization and rigidity analysis on the mutants and wild-type to determine the impact of each mutation on complex formation based on an energy-minimization metric and a scoring system for rigid cluster size and distribution.

The energy minimization (EM) is done using NAMD, which is one of the two standard molecular dynamics engines (the other being GROMACS). NAMD proceeds by first solvating the simulation box, and then running 500 steps of energy minimization using a leapfrog gradient descent approach (standard).

2 MOTIVATION

Due to their widespread involvement in cell signaling processes, PDZ domains are of medical interest regarding disease pathways. Thus, PDZ domains are being investigated as a therapeutic target, however this requires an intimate understanding of PDZ binding preferences and the ability to predict their endogenous targets, which has not been well established [2]. Additionally, a better understanding of the specificity determinants for the PDZ domain will aid in current studies related to the evolution of the domain over time and its divergence among organisms. The laborious nature of wet lab work and the breadth of proteins possessing PDZ domains makes wet-lab experimental characterization of PDZ binding preferences difficult: in the human proteome alone, the estimated number of PDZ domains ranges from 250 to 400 [1]. The large number of domains to study presents the need for a rapid way to assess PDZ binding. By taking a computational approach, we hope to simplify

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

the study of the binding of PDZ domains to various targets while still garnering meaningful data about the trends regarding binding selectivity and complex formation.

3 RELATED WORKS

There is a lively subset of protein research that aims to rapidly and accurately predict the effects of mutations on protein function, stability, and protein-protein interactions via computational methods. Our project revolves around using energy minimization and rigidity analysis as the computational tools to predict the impact of mutations on such protein-protein interactions. The paper "Low Rank Smoothed Sampling Methods for Identifying Impactful Pairwise Mutations"[5] fits into the broader context of this area of research and served as the basis for the rigidity analysis performed. In this paper, rigidity analysis served as a rapid and effective method for predicting the effects of protein mutations relative to the wild-type. Our project also aims to predict the effects of mutations on protein-protein interactions, and thus Equation 1 presented in this paper served as the basis for our rigidity metric. The paper "Elucidating Which Pairwise Mutations Affect Protein Stability: An Exhaustive Big Data Approach"[4] also fits into the broad context of predicting the effect of mutations on proteins by generating exhaustive mutations of entire proteins. It served as inspiration to use heat maps for data visualization. The heat maps in this paper were used to demonstrate trends and outliers in the rigidity data for a massive data set. The exhaustive single and double point mutations to the ligand also yielded an unwieldy amount of data, and thus we used similar heatmaps to present our data in a more digestible manner. Finally, the paper "PDZ domains and their binding partners: structure, specificity, and modification"[3] served as an important overview of both wet lab and computational approaches that have been carried out thus far regarding the PDZ domain. The paper provided context for the project itself, and thoroughly describes what is known about the biophysical binding preferences of the PDZ domain. This provided the framework for what to expect from our data, and helped reveal that known binding trends are not consistent with our data as it stands. This review will continue to be a useful reference when future data analysis approaches are explored.

4 METHODS

4.1 Ligands and PDB

We have access to eight PDZ (see figure 1) Protein Data Bank (PDB) files, split into three groups. Each group of PDB files contains the same ligand and different group has a slightly perturbed ligand when compare to the other group. The ligands of each groups have are only five-seven residues, allow us to perform single and double points mutation on all the residues of each ligand.

4.2 Rigidity Metric

Our rigidity metric (see equation 1) detailed in [5] where we compute the differences between the wild type (WT) and the mutations(Mut) weighted by each rigid cluster from 1 to the largest ridid cluster (LRC).

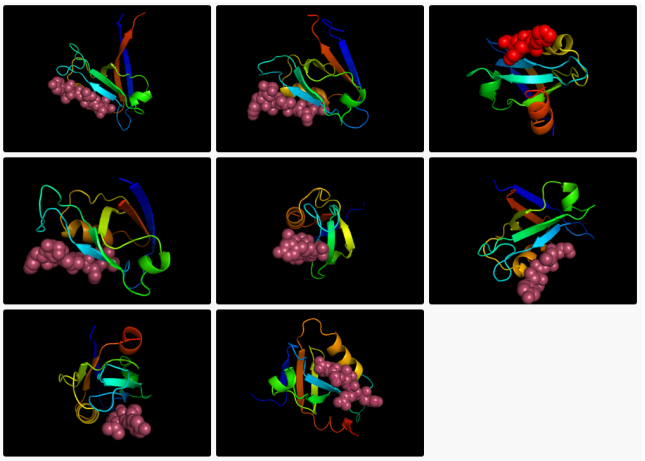

Figure 1: Renderings of the 8 PDZ proteins (segments). Credit: Bodi.

$$RD_{WT \rightarrow mutant} : \sum_{i=1}^{i=LRC} i * (WT_i - Mut_i) \tag{1}$$

To better compare each mutant to each other, we apply the softmax (see equation 2) function and assign each mutant with a value between 0 and 1.

$$softmax(y_i) : \frac{e^{y_i}}{\sum e^{y_i}} \tag{2}$$

4.3 Energy Minimization

For the energy minimization metric we began by calculating a percentage threshold based on the wild type energy minimization process described in equation 3 (Where WT_{10} is the wild type's energy at step 10 and WT_F is the wild type's final minimized energy.). We calculated the threshold to be 95% from the wild type's energy at step 10 to its final minimized energy. We used this range because we found that the potential energy levels (kcal/mol) of the wild type and mutants were highly variable between steps 0-10. By starting at step 10, the threshold is more resistant to the variability found in these early steps. We used the 95% level because we found that differences in the mutants' energy minimization processes were more significant here than at lower percentage thresholds.

$$T : WT_{10} - 0.95 * (WT_{10} - WT_F) \tag{3}$$

We then determined the step at which each wild type and mutant crossed this threshold.

$$WT_T, MT_T \tag{4}$$

To normalize this value, we subtract the wild type's step count from its mutants, giving us a difference in step count. We then take the absolute value of this difference because we are only concerned with the magnitude of the difference. By subtracting the wild type's step count from its mutants, a

$$EM_{WT \rightarrow mutant} : ABS(MT_T - WT_T \tag{5}$$

Table 1: Group 1 PBD, top 10 and bottom 10 single point mutations.

PDB	TOP	MUTATION	RD	NORM RD
2DM8	1	G2837W	89	0.9999999
2DM8	2	G2837M	72	4.14E-08
2DM8	3	G2840C	71	1.52E-08
2DM8	4	G2837F	68	7.58E-10
2DM8	5	G2840W	35	3.53E-24
2DM8	6	G2840R	31	6.47E-26
2DM8	7	G2839M	22	7.98E-30
2DM8	8	G2841Q	17	5.38E-32
2DM8	9	G2843C	16	1.98E-32
2DM8	10	G2843N	14	2.68E-33
2DM8	125	G2841K	-330	1.07E-182
2DM8	126	G2838K	-331	3.95E-183
2DM8	127	G2838R	-331	3.95E-183
2DM8	128	G2838V	-332	1.45E-183
2DM8	129	G2837L	-332	1.45E-183
2DM8	130	G2842L	-332	1.45E-183
2DM8	131	G2837V	-333	5.34E-184
2DM8	132	G2838M	-335	7.23E-185
2DM8	133	G2843R	-341	1.79E-187
2DM8	133	G2841I	-341	1.79E-187

5 RESULTS

This section present results we have on rigidity metric and energy minimization analysis.

5.1 Top outliers by Rigidity Metric

Table 1 presents the top and bottom 10 single point mutations with highest and lowest RD score for 2DM8.Table 2 presents the top and bottom 10 double point mutations with highest and lowest RD score for 2DM8. Notice Tryptophan (W) consistently making the top “disruptive" mutation for both single and double point for 2DM8, however this is not always the case. Table 3 show top and bottom 10 single point mutation for 3RL7, which has the same ligand as 2DM8 but there is not a single Tryptophan as a mutation in the entire table. All of these data can be found on github. for further analysis.

We also noticed that softmax output for top 1 is open overwhelmingly “confident" even thought top 2 is only a few RD points below, perhaps a better function is needed for normalize these kind of output.

Table 2: Group 1 PBD, top 10 and bottom 10 double point mutations.

PDB	TOP	MUTATION	RD	NORM RD
2DM8	1	G2838W.G2843L	133	0.853101
2DM8	2	G2838W.G2840N	130	0.042473
2DM8	3	G2838W.G2842C	130	0.042473
2DM8	4	G2838W.G2841V	130	0.042473
2DM8	5	G2837M.G2838W	128	0.005748
2DM8	6	G2838W.G2842A	128	0.005748
2DM8	7	G2838W.G2840M	128	0.005748
2DM8	8	G2837W.G2839A	127	0.002115
2DM8	9	G2837C.G2838W	124	0.000105
2DM8	10	G2838W.G2839R	121	5.24E-06
2DM8	7573	G2839Q.G2842N	-373	1.51E-220
2DM8	7574	G2839I.G2840Y	-375	2.04E-221
2DM8	7575	G2837P.G2841W	-378	1.02E-222
2DM8	7576	G2841R.G2843R	-396	1.55E-230
2DM8	7577	G2837P.G2839G	-411	4.73E-237
2DM8	7578	G2839P.G2841M	-417	1.17E-239
2DM8	7579	G2839I.G2840T	-427	5.32E-244
2DM8	7580	G2839Q.G2841L	-428	1.96E-244
2DM8	7581	G2838D.G2839Y	-428	1.96E-244
2DM8	7582	G2839I.G2842K	-434	4.85E-247

Table 3: Group 1 PBD, top 10 and bottom 10 single point mutations.

PDB	TOP	MUTATION	RD	NORM RD
3RL7	1	G2840S	301	1
3RL7	2	G2841H	286	3.06E-07
3RL7	3	G2838S	278	1.03E-10
3RL7	4	G2841E	275	5.11E-12
3RL7	5	G2839P	273	6.91E-13
3RL7	6	G2840Y	271	9.36E-14
3RL7	7	G2839E	268	4.66E-15
3RL7	8	G2843K	266	6.31E-16
3RL7	9	G2840H	262	1.15E-17
3RL7	10	G2840P	262	1.15E-17
3RL7	123	G2842K	-687	0
3RL7	124	G2839T	-714	0
3RL7	125	G2840C	-720	0
3RL7	126	G2841N	-725	0
3RL7	127	G2840D	-726	0
3RL7	128	G2837D	-729	0
3RL7	129	G2840T	-736	0
3RL7	130	G2843T	-744	0
3RL7	131	G2842E	-744	0
3RL7	132	G2842I	-746	0
3RL7	133	G2840R	-781	0
3RL7	134	G2843M	-845	0

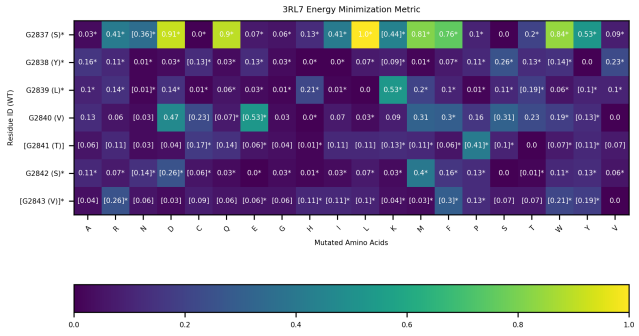

Figure 2: Heatmap of the Energy minimization metric for single point mutations of 3RL7. Note: no patterns were identified in the heatmaps across the 8 pdb files that were investigated.

5.2 Heatmaps

Using matplotlib, heatmaps were generated for both energy minimization and rigidity analysis. An example of one of these heatmaps can be seen in Figure 2. Through the analysis of these heatmaps, the energy minimization metric was determined to be unreliable as as an indication of the impact of a mutation on the proteins ability to function. The heatmaps are generally unifiorm in nature, and no differentiation is reliably made by the metric between mutations of varying size as well as mutations that introduce new hydrogen bonds or hydrophobic interactions.

6 CONCLUSION

Our results are, unfortunately, inconclusive. Computationally, we should see severe impacts on complex formation when the C terminal residue of the ligand is mutated to a large residue such as Tryptophan. This expectation is due to the known preference for V/I/L at the C terminal position for Class 1 PDZ domains, and structural information which shows that V/I/L are involved in tight hydrophobic packing. Since we cannot validate this positive control with our data, there is very little we can say about its meaning. Also, because we have no access to ground-truth, we can not compare to check if the computation output is correct or if we messed anything in between (scripts, pipeline, data processing etc.).

7 FUTURE WORK

The current results of the energy minimization analysis, and to a lesser extent the rigidity analysis, are inconsistent with thoroughly validated experimental data regarding PDZ domain binding preferences. Before further work can be carried out, there are numerous avenues to explore in an attempt to validate this positive control.

One option is to work with outputs of energy minimization other than the total potential energy. The potential energies or other outputs from different step ranges could also be analyzed, such as comparing the relaxation curves of the first ten steps of each mutant. Essentially, further data exploration of the existing data is needed. The velocities of the atoms in the ligand versus the rest of the protein could also be explored. If further data exploration and analysis still results in data inconsistent with known binding preferences of the PDZ domain, a further option to explore is position restrained energy minimization to hold the protein in place while allowing the ligand to move.

REFERENCES

[1] Jeanine F. Amacher, Lionel Brooks, Thomas H. Hampton, and Dean R. Madden. 2020. Specificity in PDZ-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology: X* 4 (2020), 100022. https://doi.org/10.1016/j.yjsbx.2020.100022

[2] Nikolaj Christensen, Jelena Čalyševa, Eduardo Fernandes, Susanne Lüchow, Louise Clemmensen, Linda Haugaard-Kedström, and Kristian Strømgaard. 2019. PDZ Domains as Drug Targets. *Advanced Therapeutics* 2 (04 2019), 1800143. https://doi.org/10.1002/adtp.201800143

[3] Ho-Jin Lee and Jie Zheng. 2010. PDZ domains and their binding partners: Structure, specificity, and modification. *Cell communication and signaling : CCS* 8 (05 2010), 8. https://doi.org/10.1186/1478-811X-8-8

[4] N. Majeske and F. Jagodzinski. 2018. Elucidating Which Pairwise Mutations Affect Protein Stability: An Exhaustive Big Data Approach. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 01. 508–515. https://doi.org/10.1109/COMPSAC.2018.00078

[5] Nicholas Majeske, Filip Jagodzinski, Brian Hutchinson, and Tanzima Islam. 2018. Low Rank Smoothed Sampling Methods for Identifying Impactful Pairwise Mutations. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (Washington, DC, USA) *(BCB '18)*. Association for Computing Machinery, New York, NY, USA, 681–686. https://doi.org/10.1145/3233547.3233714

[6] Jordan Valgardson, Robin Cosbey, Paul Houser, Milo Rupp, Raiden Bronkhorst, Michael Lee, Filip Jagodzinski, and Jeanine Amacher. 2019. MotifAnalyzer-PDZ : A computational program to investigate the evolution of PDZ-binding target specificity. *Protein Science* 28 (10 2019). https://doi.org/10.1002/pro.3741