

Analyzing the Impact of Pleiotrophin's C-Terminal Domain on its Binding Properties to Protein Tyrosine Phosphatase

Bernard Shen

University of Chicago

Introduction

Pleiotrophin (PTN) is a multifaceted protein with crucial roles in various physiological and pathological processes. Belonging to the heparin-binding growth factor family of proteins, it is secreted by the body, and plays numerous roles in many cellular processes, including neural development, angiogenesis, and tissue regeneration [8]. Indeed, the function of PTN and other growth factor proteins is to a great extent dependent on their ability to bind to specific receptors on a cell's surface to influence cellular behavior; and in the case of PTN, the interaction between itself and Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ) is crucial for the signaling process. More specifically, PTN binds to the glycosaminoglycan (GAG) portion of PTPRZ, with a specific affinity to the heparan sulfate (HS) and chondroitin sulfate (CS) components of PTPRZ [5]. The structure of PTP (a necessary approximation for PTPRZ) that was used in this study consists primarily of a dense clump of alpha helices (which I interpreted as PTP-D1 and PTP-D2) with a long turn (a strand that I interpreted as the serine, glycine-rich domain that PTN binds to) that extends outward from the helix clump [1][3].¹ The structure of PTN is fairly simple in comparison; the protein consists of beta sheets, turns, and coils, with structured thrombospondin type-1 repeat (TSR) domains responsible for electrostatic interactions, notably with CS. Ryan et al. identify two of these TSR domains, one associated with the N-terminal and ending at residue 57; the other associated with the C-terminal and beginning at residue 67. Furthermore, they write that the C-terminal domain and hinge (from residue 58 to 66) constitute a robust CS-binding site, and structural elements such as disulfide bonds and basic amino acid clusters contribute significantly to PTN's binding capabilities. Nonetheless, the primary mutation that Ryan et al. identify and discuss involves the truncation of PTN's C-terminal, specifically the removal of residues 115 to 136, which was reported to lower affinities for CSA (one specific type of CS), but not CSE (another type) experimentally [5]. This result motivated my experimental design and choice of second simulation; because PTN is a growth factor, I attempted to model these experimental results using docking and molecular dynamics to better understand why and how the removal of twenty-one residues on the C-terminal may impact PTN's affinities for binding. I should also note that the scientific literature concerning PTN mutations is limited, and Ryan et al.'s report seemed to be the only one discussing how structural mutations to PTN affected its ability to function.

¹ Although fairly little scientific literature exists on the specific subject of PTPRZ structure as it relates to PTN binding, Kuboyama et al. include information about the binding sites of PTPRZ in a diagram; when selecting the protein on AlphaFold, this diagram would form the foundation for my selection of a relevant PTP model (as many models exist, but some did not include the relevant domains, and others appeared to be a different shape than the one modeled in Kuboyama et al.'s paper).

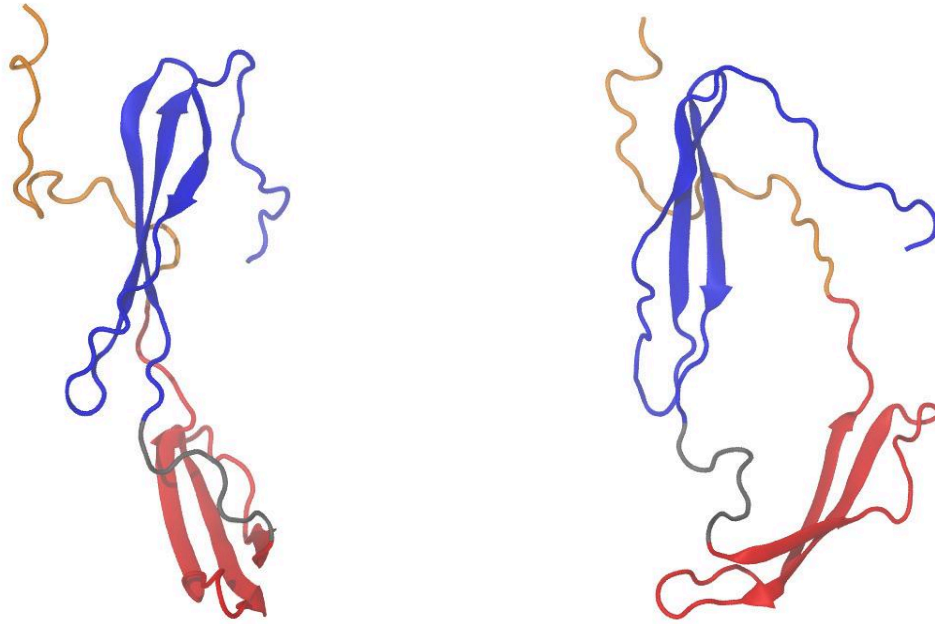


Figure 1. Pleiotrophin [4]

Two (rotated) perspectives of wild-type pleiotrophin. The relevant regions and domains are highlighted in different colors as follows. Blue, N-terminal TSR domain; gray, hinge; red and orange, C-terminal TSR domain; orange, portion of C-terminal that is truncated in the mutation identified by Ryan et. al [5]. Note the relative lack of structure, which contributes to the high degree of flexibility for this protein, and will be particularly relevant in understanding the results of this experiment.

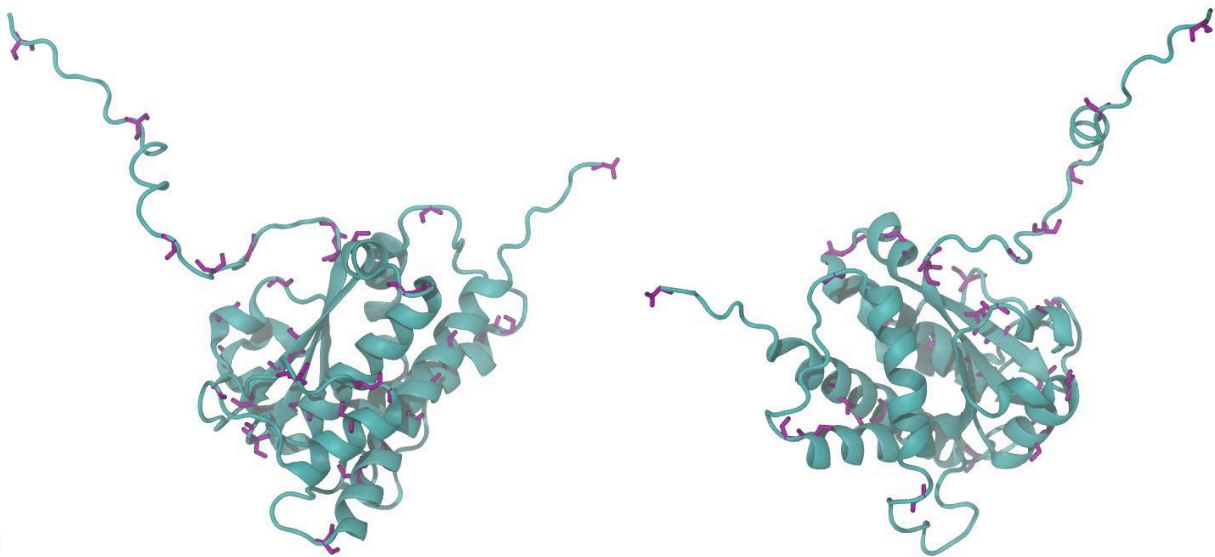


Figure 2. Protein Tyrosine Phosphatase [AlphaFold]

Two (rotated) perspectives of low molecular weight protein tyrosine phosphatase. The relevant region for binding, as highlighted by Kuboyama et al., is the serine, glycine-rich domain [3]. Hence, in these figures, I have highlighted all serine and glycine residues in purple. I interpreted the stem which extends out from the helices as most likely corresponding to this domain, as it most closely resembles the graphic that Kuboyama includes, and appears to have a high concentration of serine and glycine.

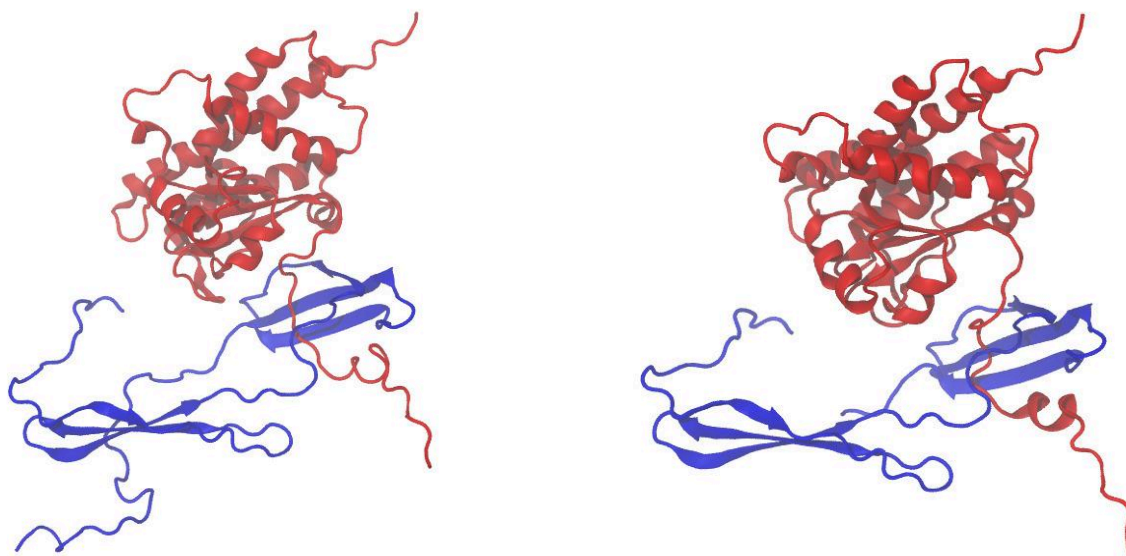


Figure 3. Docked Systems [7]

Left: wild-type PTN (blue) docked to PTP (red); right: mutated PTN (blue) docked to PTP (red). Note that the docking performed by Cluspro 2.0 (protein-protein docking) correctly identified the serine, glycine-rich region on PTP as the binding site for PTN.

Leveraging molecular dynamics simulations for modeling protein-protein interactions, such as that between PTP and PTN, presents unique advantages over experimental approaches, offering enhanced visibility and streamlined data collection; they are able to capture atomic-level details of biomolecular behavior, offering dynamic, three-dimensional insights into processes like conformational changes and ligand binding with femtosecond resolution. MD simulations also allow precise control of parameters, facilitating the study of biomolecules under various conditions. Moreover, this approach presents a compelling tradeoff among different simulation types, balancing notable accuracy at a moderate computational expense. Indeed, MD simulations have recently garnered increased attention because advancements in experimental techniques like X-ray crystallography and cryo-electron microscopy have led to a surge in available biomolecular starting points for MD simulations, and these simulations have become more powerful and accessible due to improvements in computer hardware and more accurate physical models. While the execution of simulations has become relatively straightforward, expertise remains crucial in designing simulations, interpreting results, choosing techniques, and implementing advanced methods [6].

Methods

For my MD simulations, I conducted two distinct setups to explore the interactions of Pleiotrophin (PTN) with its binding partner Protein Tyrosine Phosphatase (PTP). In both setups, I utilized the biological assembly of PTN (PDB ID: 2N6F, obtained by Ryan et al. via NMR spectroscopy [4]) and the AlphaFold structure prediction for low molecular weight PTP [Vajda]. Note that these models are in some regard only approximations for the actual structures (This will be discussed further in the following sections of this report), and I did attempt docking PTN with CSA directly using its biological assembly (PDB ID: 1C4S, obtained by Winter et al. via fiber diffraction). Multiple online docking servers, including ClusPro, SwissDock, and GRAMM, were unable to dock PTN to CSA, which I speculate could be due to the

presence of unfamiliar molecules in the PDB file, notably a Heterogen structure consisting of GC4 and GSU; these ligands/cofactors are not widely recognized by docking algorithms. Hence, I proceeded with AlphaFold's predicted structure for PTP instead of CSA for docking. After downloading the PDB files for these models (PTN and PTP), I created an alternate PDB file for the mutated PTN system, where I removed residues 115 to 136 by manually deleting those entries in a text editor. I then performed protein-protein docking, uploading the respective files to ClusPro (to find where the proteins should be joined in the MD simulation) for both systems. In both systems, ClusPro docked PTN near the expected binding site on PTP. After downloading these new PDB files (wild-type and mutated), which contained both proteins in their docked configurations, I proceeded with the standard steps in a MD simulation.

Reading the PDB files into Visual Molecular Dynamics, I used the AutoPSF generator to add hydrogen atoms to both systems. I then solvated and ionized the systems, padding with ten angstroms of water on each side and neutralizing the total charge with 0.1 mol/L salt concentration. I centered and restrained the protein backbones before submitting them for ten thousand steps of minimization (corresponding to 0.01 nanoseconds) on Midway-3 (using 1 hour and 2 nodes). I should note that all operations performed on the protein on Midway-3 involved the use of a configuration file, which executed all of the actions pertaining to protein simulation, and took the following inputs: coordinates/structure of the system, input/output/restart file name, parameter file name, and cell basis vectors corresponding to the size of the solvated system. After the minimization was complete for each system, I configured them for heat-up and equilibration by exporting the last frame of the minimization trajectory and restraining the backbone. Over the course of two nanoseconds, the heat-up and equilibration step involved bringing the temperature of both systems from 240K to 300K in increments of 5K at a target pressure of 1.0133 bar via Langevin dynamics. Editing the configuration file once again, I submitted the heat-up and equilibration jobs on Midway-3 (using 6 hours and 4 nodes). Following this step, I performed my NPT simulations (constant number of particles, pressure: 1.0133 bar, and temperature: 300K) for both systems. Utilizing two twenty-nanosecond NPT runs for each system, I attained two trajectories, each with 100,000 frames (minor edits had to be made to the configuration file both before and between the NPT runs, primarily corresponding to input and output names). The final task in completing the simulation trajectories was the concatenation of the heat-up and equilibration, NPT1, and NPT2 systems. Using the CatDCD program, I combined the trajectories and reduced the number of frames by a factor of ten. I then unwrapped the trajectories and aligned the concatenated trajectories to the last frame of their corresponding minimizations. From there, I performed my analysis on both systems.

Results and Discussion

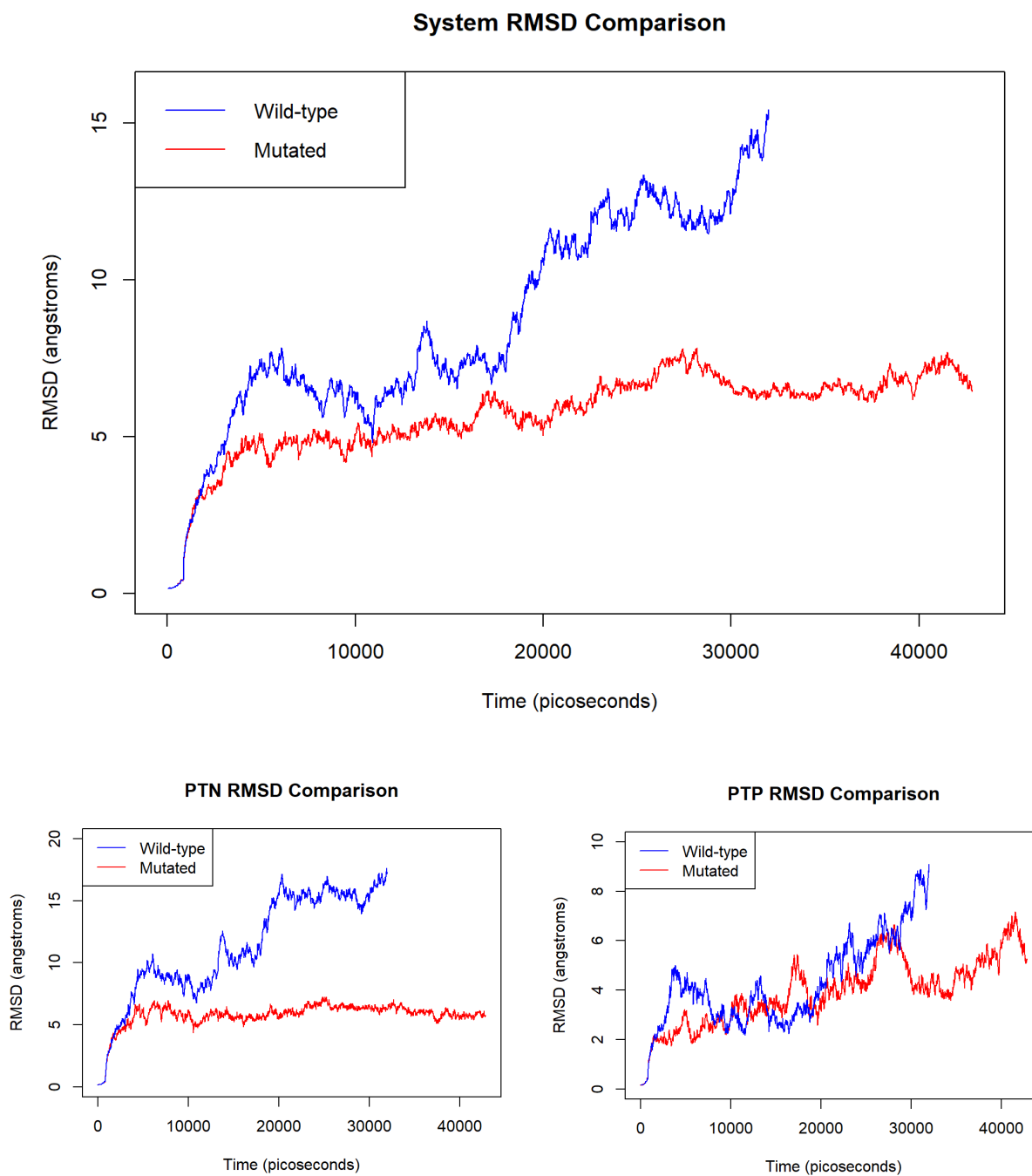


Figure 4. Root Mean Standard Deviation

Top: total system RMSD; bottom: individual RMSD for each protein in the docked system. Note that PTN never reaches a stable equilibrium, and the RMSD for PTN is substantially higher than the RMSD for PTP. Visually analyzing the trajectories, the terminals of the wild-type pleiotrophin seem to flail outward from their original positions after minimizing, explaining the continual increase in the RMSD of the system. This phenomenon is not observed in the mutated system because the C-terminus was truncated. Both systems do exhibit the same initial increase in RMSD, which would seem to be due to the relative flexibility of the PTN.

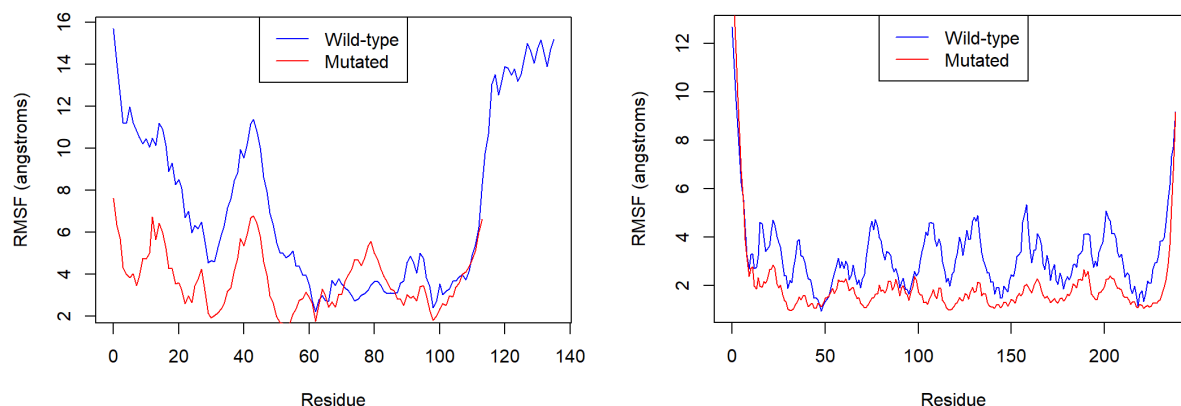


Figure 5. Root Mean Square Fluctuation

Left: PTN RMSF; right: PTP RMSF. The wild-type and mutated systems generally follow the same pattern, although the wild-type generally has much higher RMSF values, particularly near the ends of the first chain. From an analysis of the visual trajectory, these results seem to make sense. The ends of the pleiotrophin protein exhibit an extreme degree of flexibility, and their movement is very noticeable relative to the rest of the system. The system with the truncated terminal, on the other hand, does not exhibit this phenomenon; more interesting, then, are the regions where the RMSF for the wild-type system is lower than that of the mutated system (notably, residues 60 to 85 of PTN and residue 48 and 93 of PTP).

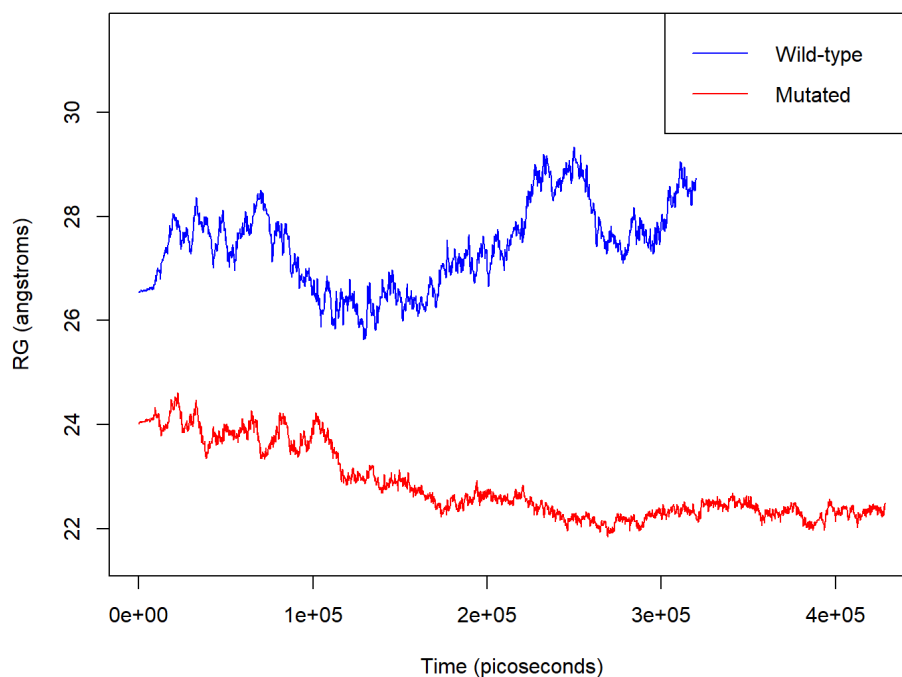


Figure 6. Radius of Gyration

The radius of gyration does not seem too affected by the extension of the ends of the protein over time, seeing as the ends of the protein remain extended outward whilst the radius of gyration decreases. Moreover, this data suggests that the mutated protein becomes more compact with time (though not by much), which is consistent with a visual analysis of the system.

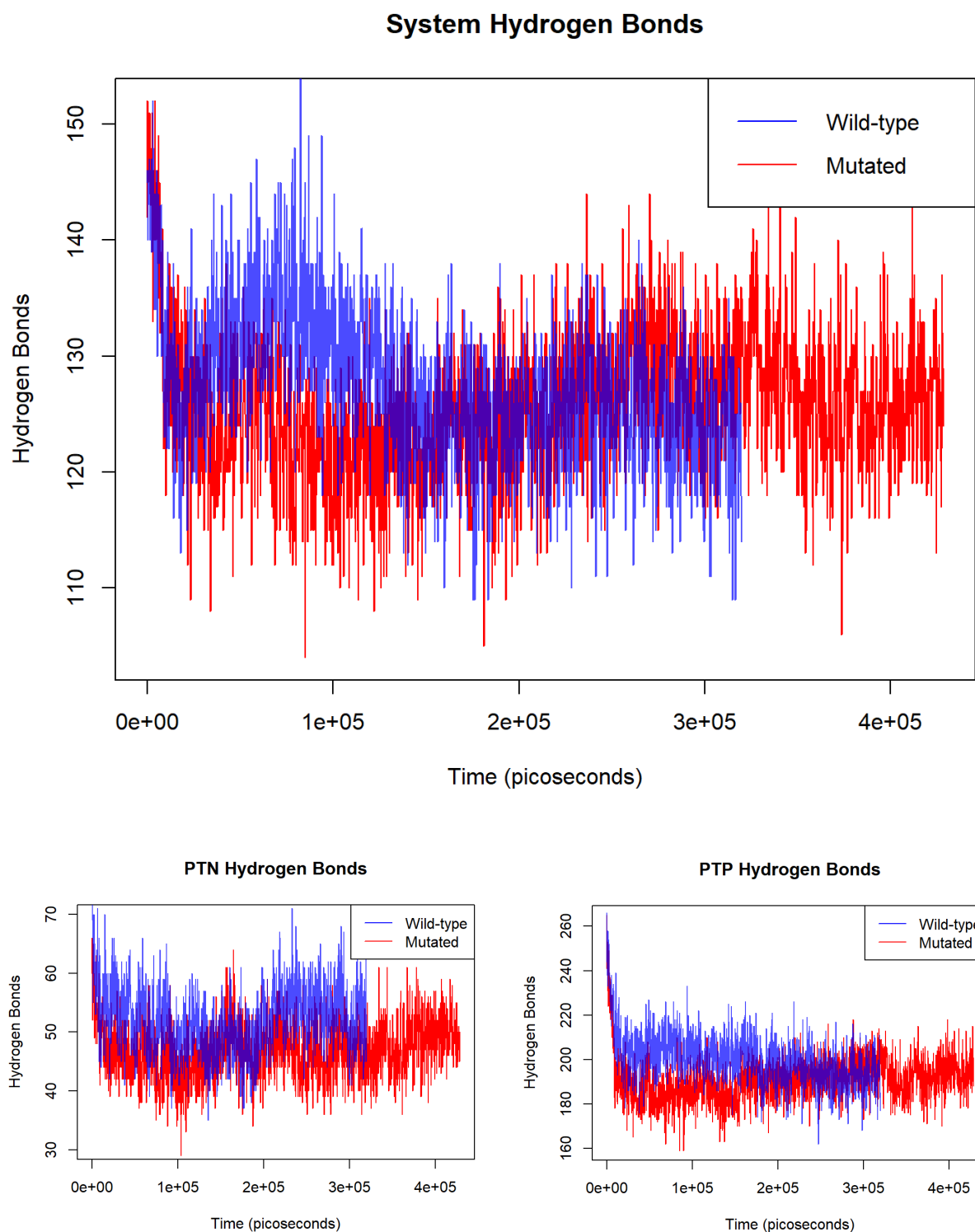


Figure 7. Hydrogen Bonds

Top: total system hydrogen bonds; bottom: hydrogen bonds measured for each protein in the docked system. I used a cutoff distance of 3.5 angstroms and a cutoff angle of 35 degrees for each graph. I do not know why the hydrogen bonds for PTP are higher than the system hydrogen bonds, but after calculating several times I achieved the same result.

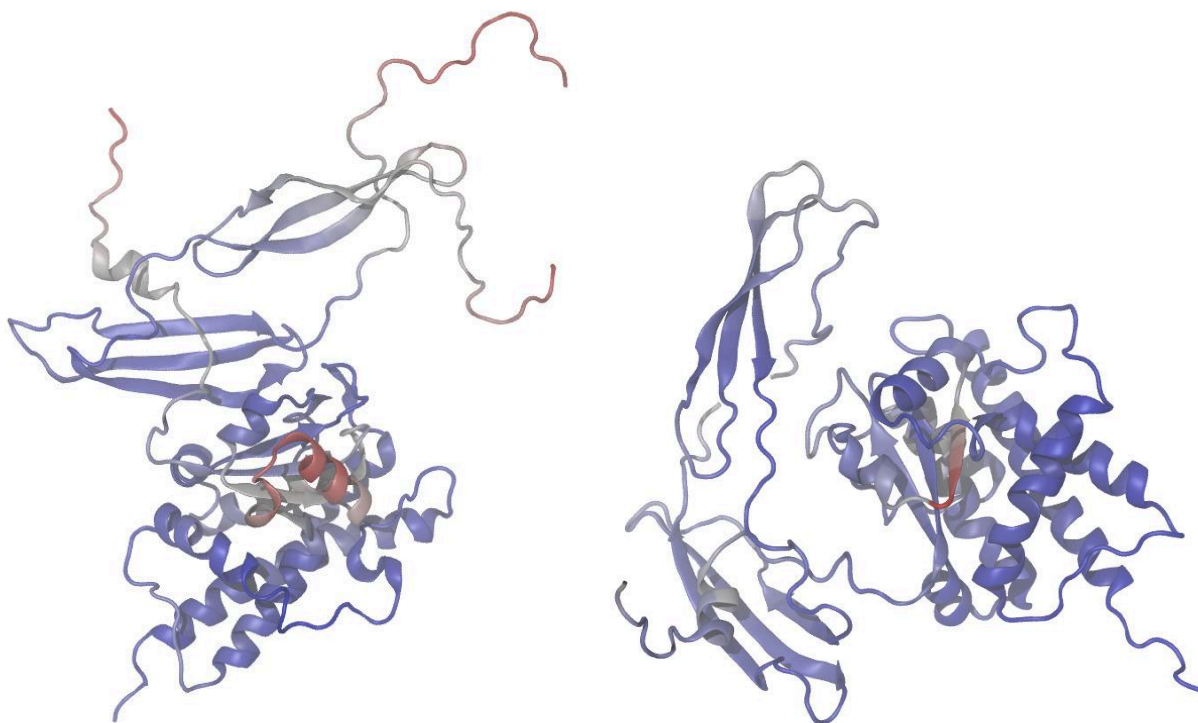


Figure 8. RMSF Coloration

Left: wild-type system RMSF; right: mutated system RMSF (blue is lower, red is higher). Note the coloration of the tails in the wild-type system, which reflects the extent to which those residues moved over the course of the simulation; this result is consistent with a visual analysis of the simulation trajectory. More interesting, then, are the high RMSF values on the interior residues of PTP in either system. For some reason, these residues do not appear so distinctly on the RMSF plots, and do not appear to exhibit more movement than other regions of PTP over the course of the simulation.

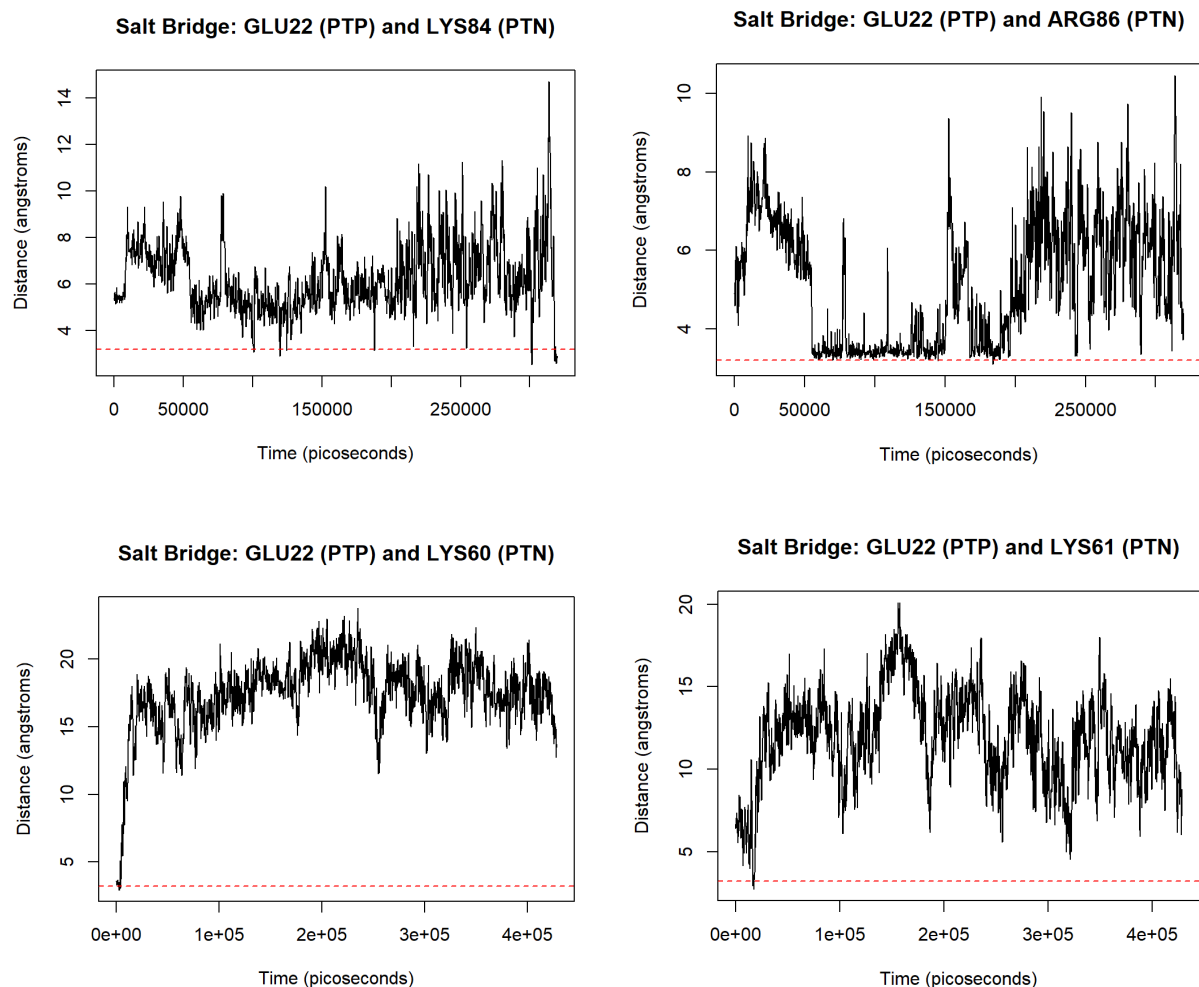


Figure 9. Salt Bridge Analysis

Top: wild-type system; bottom: mutated system. Because my simulation involves the binding of two proteins, I thought it might be interesting to pick a salt bridge that occurs between the two proteins. Despite not being strong (defined by a relatively constant distance) for a majority of the trajectory, these salt bridges represent a key point of contact between the proteins and are more representative of the system than an isolated salt bridge found on one of the proteins. The glutamic acid (22) on PTP forms two notable salt bridges over the course of the trajectory; one with lysine (84) and another with arginine (86). This glutamic acid residue on PTP does form a salt bridge in the alternate system, but with a different set of residues: lysine (60 and 61) on PTN. This difference would seem to be corroborated by research; the truncation of the terminal may have affected the binding site of the residues in the sense that salt bridges were formed at different locations; and in this instance, the salt bridge is much weaker in the mutated protein, which is consistent with the conclusions that researchers reached regarding the reduced affinity (or ability to bond) due to the mutation (though the full extent of this mutation is not clear from this individual salt bridge).

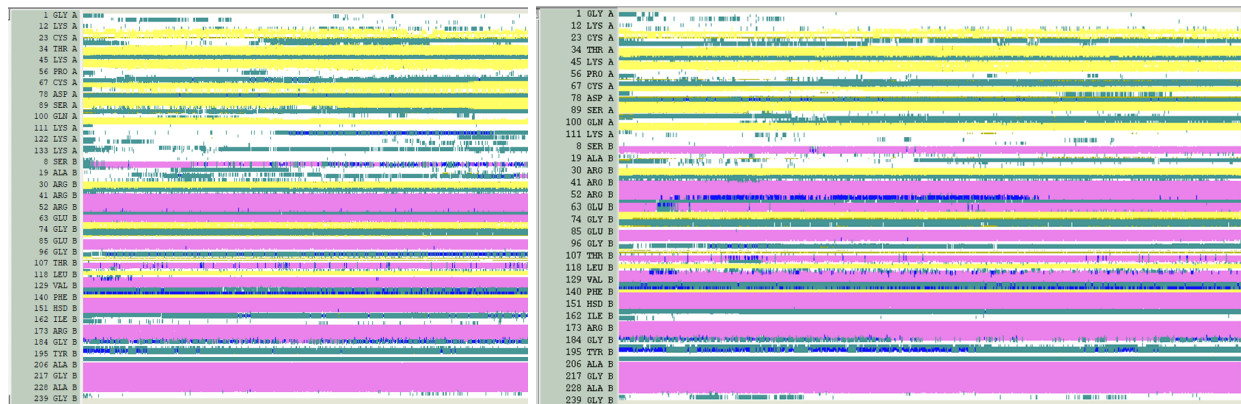


Figure 10. Secondary Structure Analysis

Left: wild-type system secondary structure timeline; right: mutated system secondary structure timeline. As seen above, the secondary structures (notably the observable ones: alpha helices and beta sheets) exhibit fairly minimal change during the simulation trajectory (I am attaching the secondary structure timeline as opposed to images of specific secondary structure changes to demonstrate that this section is not particularly relevant to my system). As the mutation merely cleaved turn residues at the end of the protein, and given that PTN is a flexible protein that lacks much structure to begin with, this result was to be expected.

The dynamical behavior of the wild-type system is primarily characterized by the movement of the tails for both proteins. Whilst the core regions of either protein appear to vibrate and jiggle over the course of the simulation time, the primary binding region remains fairly consistent, without much change in how the proteins are oriented with respect to one another. The tails of both proteins exhibit a lot of scattered and random movement; in particular, over the latter third of the trajectory, the C-terminal tail of PTN appears drawn to the turn structures on PTP at around residue 100, closing to a distance of 2.93 angstroms by the end of the trajectory, a result that might corroborate Ryan et al.'s experimental results as it suggests that the C-terminal tail might be involved in binding. Unfortunately, the extent to which this binding persists is unclear as it only occurs within the last two hundred frames of the concatenated trajectory.

The dynamical behavior of the mutated system is, for the most part, identical; the tails of both proteins exhibit the most motion whilst the cores of the protein and the binding site between the proteins remains mostly in place, with the exception of some minor vibrations over the course of the trajectory. In the mutated system, however, the tails of PTN behave differently; past frame 800, or 160,000 picoseconds, both the N-terminal and truncated C-terminal appear to interact with PTP, and remain in close contact for the remainder of the simulation. Measuring the distance between the residues at the end of the trajectory yields a distance of 1.74 angstroms between Glutamine 4 on the N-terminus of PTN and Arginine 202 on PTP (the closest point of contact). PTP comes to a minimum distance of 10.25 angstroms with the truncated C-terminus, too long of a distance to culminate any meaningful forces. Further examining the simulation, it appears as though the end of the N-terminus (glycine) binds to the end of the truncated C-terminus (lysine), with a distance of only 1.68 angstroms. Hence, it stands to reason that the N-terminus of PTN plays a role in binding to PTP, whilst the truncated C-terminus of PTP merely binds to the N-terminus, potentially acting as a stabilizer, or alternatively, contributing minimally to binding. Indeed, it is worth noting that in the wild-type simulation the

N-terminus does not come remotely close to PTP; whether this is due to the truncation of the C-terminus or inherent randomness in these simulations is unclear.

Note that in the above graphs, the data for my wild-type simulation is truncated due to cross-PBC interactions in the later frames of the trajectory.² Despite padding my docked proteins with ten angstroms of water on each side, a turn on the end of the PTN protein appears to unravel from its initial position over the course of the simulation, eventually coming into close contact with PTPN proteins in neighboring cells. Indeed, after frame 1600 (32000 picoseconds) of the concatenated trajectory, the data is completely unusable (see Figure 11). I hypothesize that this result occurred due to two factors: (1) PTN is a relatively unstructured and flexible protein, consisting mostly of beta sheets and turns, and the terminals that extend off of it are not stabilized by salt bridges or disulfide bonds; (2) the biological assembly of PTN on the RCSB Protein Data Bank represents a highly compacted version of the protein, where the turns on the terminals are neatly tucked next to the beta sheets. Knowing this now, I hypothesize that more padding in the solvation stage (twenty angstroms of water instead of ten) would have prevented these cross-PBC interactions; due to limited time, I was unable to perform the simulation again and fix the issue.

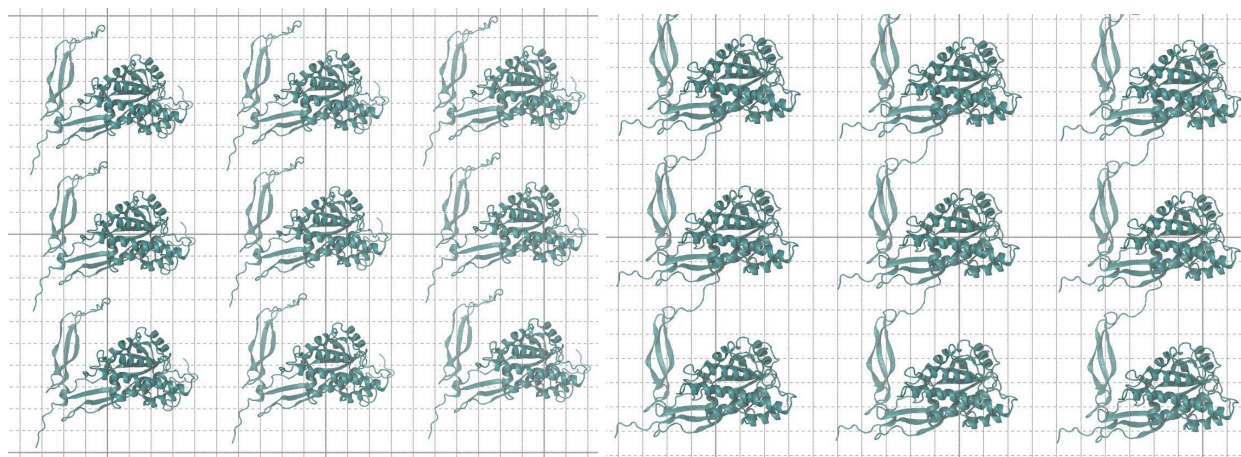


Figure 11. Cross-PBC Interactions

Left: last acceptable frame of the simulation (approximately ten angstroms of distance between all parts of the proteins and before cross-PBC contact); right: cross-PBC interactions observed after frame 1600 (approximately thirty-two nanoseconds). The data after frame 1600 for the wild-type system was discarded.

Examining the differences (or lack thereof) in secondary structure, hydrogen bonds, RMSF, and RMSD between the two simulations, it is not entirely clear whether the experimental result attained by Ryan et al. can be verified; indeed, comparing the RMSD graphs, it is clear that the truncation of the C-terminus has a noticeable impact on the broader movements of PTN, and does not seem to affect the movement of PTP. The RMSD for the wild-type system is notably higher than that of the mutated system, which is explained by the flexibility of the C-terminus and its contribution to RMSD. While a lower RMSD in the wild-type system may have served as evidence in support of Ryan et al.'s conclusion

² I spoke to professor Haddadian about this. Since the cross-boundary interactions did not occur until fairly late in the simulation, he said it would be acceptable to simply discard those data points, as opposed to rerunning the entire simulation.

(tighter binding), the fact that the wild-type system attains a higher RMSD is not necessarily an outright departure from Ryan et al.'s result. The radius of gyration results essentially indicate the same thing: the wild-type system is more flexible, adopting a wider mass footprint over more of the trajectory. However, just because the C-terminal exhibits a lot of movement does not necessarily imply that the binding of the proteins is worse. The RMSF outcome is nonetheless more contradictory to the experimental conclusion; across both the PTN and the PTP proteins, the wild-type system exhibits a higher RMSF than the mutated system, except between residues 70 and 85 in PTN and residues 49 and 96 in PTP. These results indicate that the wild-type system exhibits more movement on average, which contradicts an assumption for stronger binding and hence a more tightly constrained system. Strangely, the plot of the RMSF does not appear to agree with the colored protein (the tails of PTP should be more red than its center), and this plot does not provide any additional insights to verifying the experimentation.

Analyzing the hydrogen bond graphs, it appears that the number of hydrogen bonds in either system as a whole did not vary by much, although the wild-type trajectory appears to follow a slightly different trajectory to the mutated trajectory; the gain and loss is hard to explain from a visual analysis. The hydrogen bond graphs corresponding to each system provide more insights, namely showing that the wild-type system generally has more hydrogen bonds, which one would expect given that it has twenty-one more residues. Whether this, alone, is enough to explain the higher number of hydrogen bonds is unclear; perhaps more hydrogen bonds were formed between PTN and PTP during some phases of the wild-type trajectory, indicating a stronger connection between the two proteins.

To identify the important residues in my system, I examined the visual trajectory to find which specific residues constituted binding sites over the course of the simulation. In the wild-type system, these amino acids occurred around index 62, 76, and 100 on PTN and index 12, 65, and 99 on PTP. In the mutated system, these amino acids occurred around index 5, 60, and 113 on PTN and index 18, 83, and 203 on PTP. Using the binding site description from Kuboyama et al. as a foundation for my investigation allowed me to further focus my analysis more specifically: for the wild-type system, I chose to examine residues 60 to 65 and 98 to 103 on PTN, as well as residues 10 to 15 on PTP; for the mutated system, I chose to examine residues 58 to 65 and 98 to 103 on PTN, as well as residues 14 to 22 on PTP. It is important to note that, while Ryan et. al note the importance of the C-terminus in PTN binding to PTPRZ, using these residues as a basis for comparison would not be effective because there is nothing to compare them against in the mutant system.

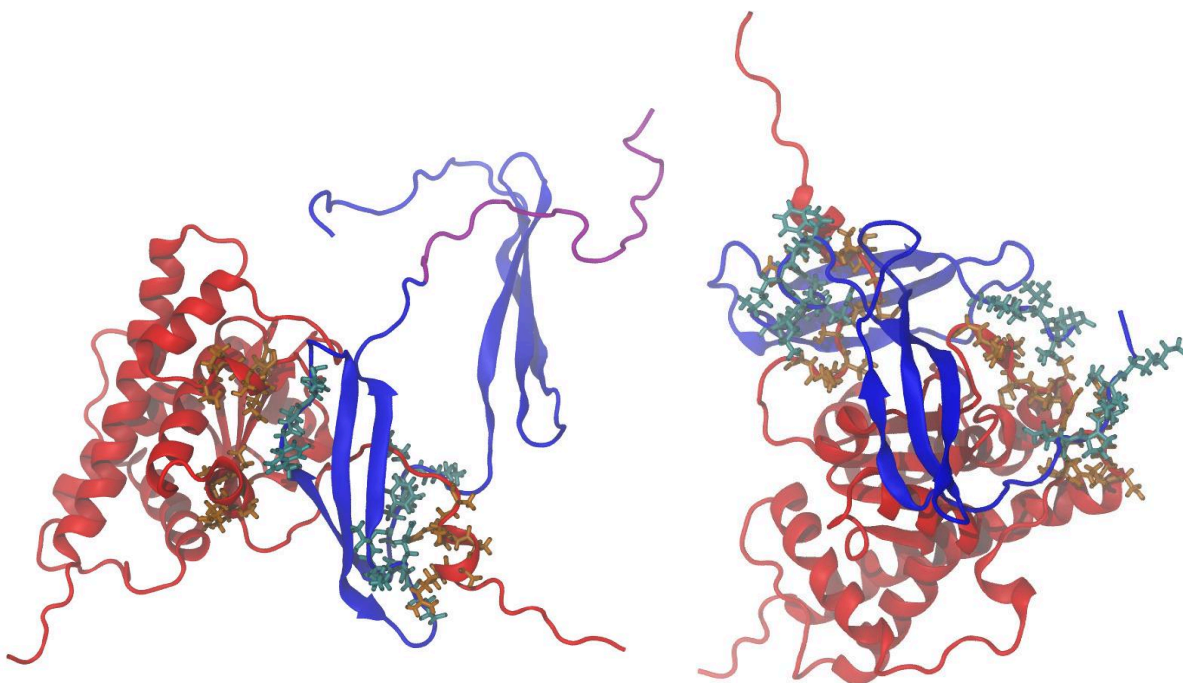


Figure 12. Residues of Interest

Left: regions where the two proteins appear to make contact in the wild-type system; right: regions where the two proteins appear to make contact in the mutated system (PTN is shown in blue, PTP is shown in red, contact residues on PTN are shown in cyan, contact residues on PTP are shown in orange, the C-terminal region of mutation is highlighted in purple in the wild-type system). These regions are important because they determine how well the two proteins can attach to one another.

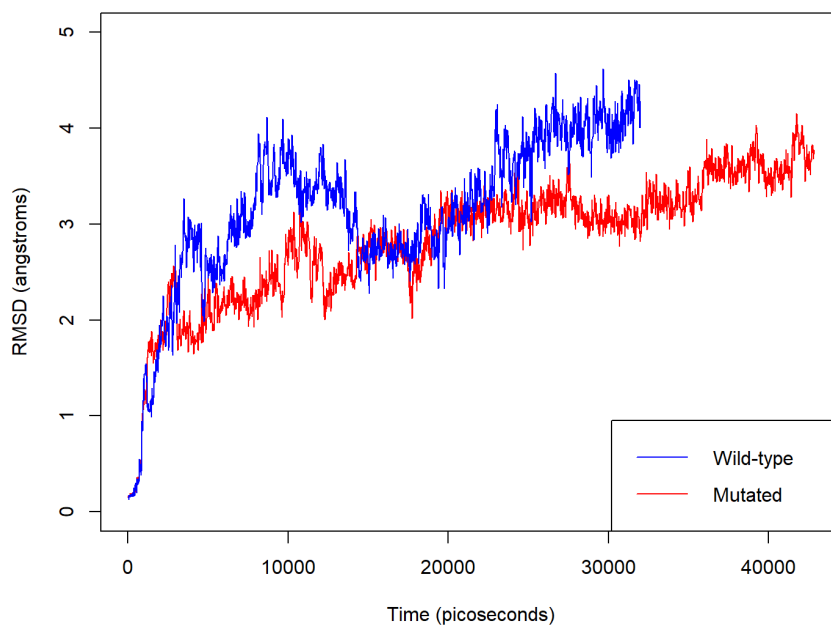


Figure 14. Binding Region RMSD

Here I compare the RMSD of the binding site residues in the wild-type system (residues 60 to 65 and 98 to 103 on PTN; residues 10 to 15 on PTP) to the mutated system (residues 58 to 65 and 98 to 103 on PTN; residues 14 to 22 on PTP).

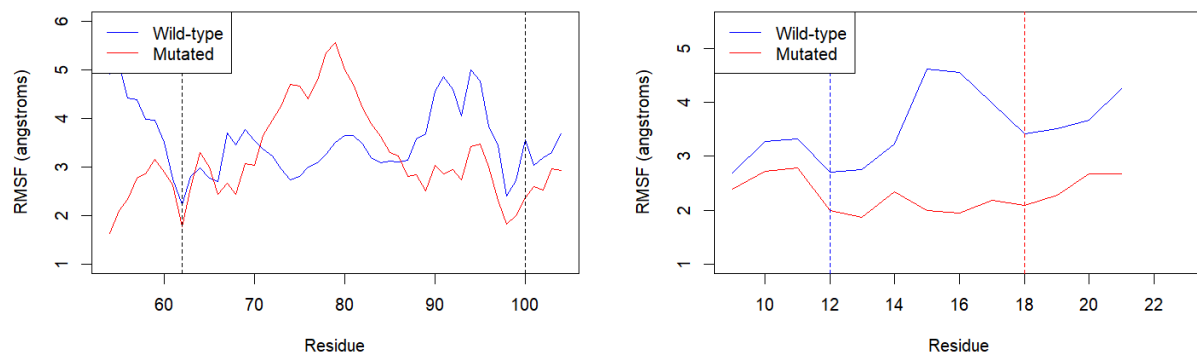


Figure 15. Binding Region RMSF

Left: PTN binding site RMSF; right: PTP binding site RMSF. The black dotted lines correspond to the approximate region where the residues are closest. On PTP the binding locations are slightly different between the wild-type and mutated system, so I indicate each region of proximity with a dotted line drawn in its corresponding color.

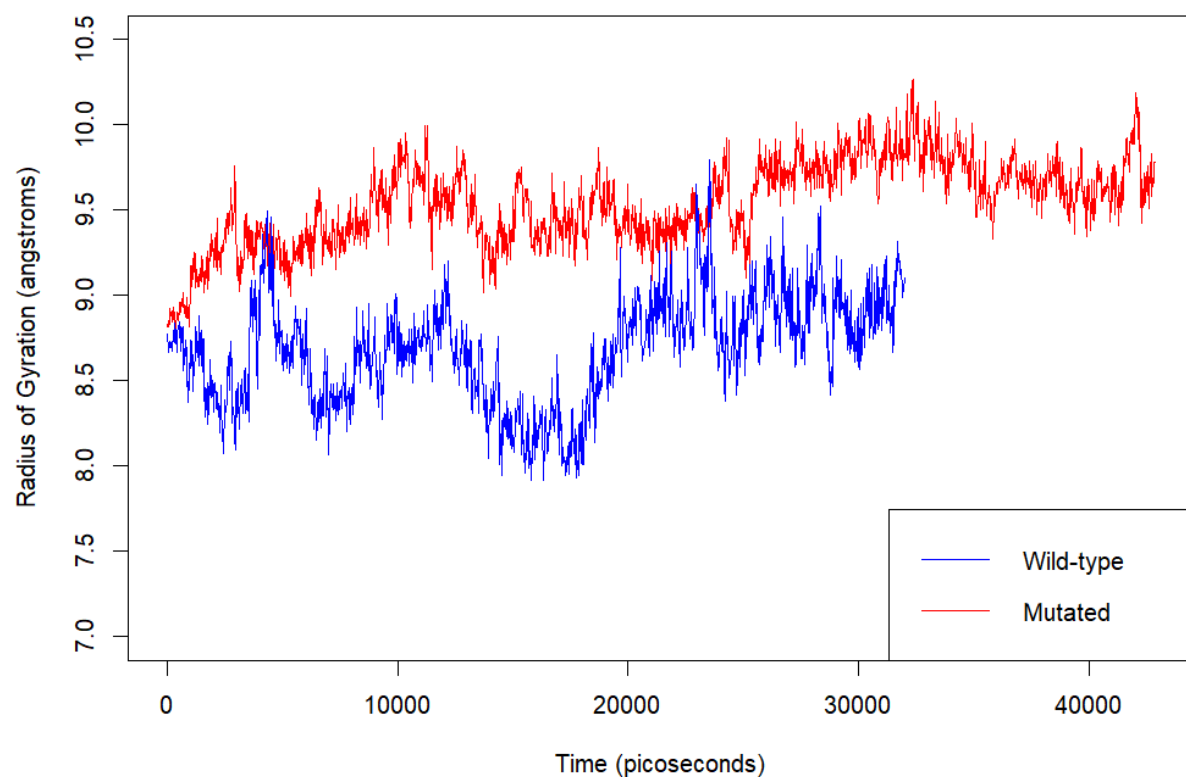


Figure 16. Binding Region Radius of Gyration

The residues near the binding sites of the wild-type system have a lower radius of gyration for the majority of the trajectory, indicating a tighter mass and a better connection between the two proteins (using the same set of amino acids as in the RMSD graph). These radii may not be comparable, however, because the residues used to compute them (in either case) are not the same, which may throw off the calculation.

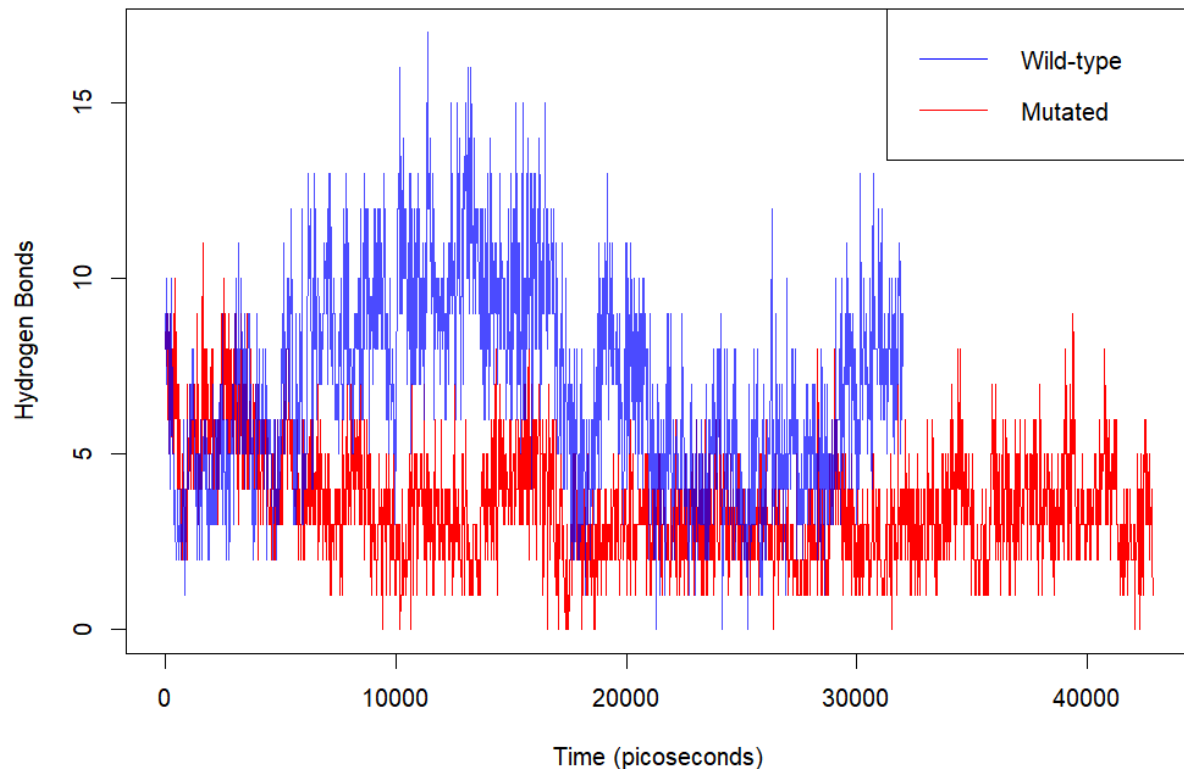


Figure 17. Binding Region Hydrogen Bonds

I equated the number of residues across systems to make this a more fair comparison, though the specific residues I chose for each system varied slightly to best correspond to their respective binding sites.

The MD simulation appears to have revealed some degree of nuance to either system; indeed, while the binding sites and general shape of the systems appear almost identical in the docked PDB, MD simulation reveals a layer of added detail when examining this mutation. For instance, the use of these methods provided insight into the high degree of flexibility that PTN exhibits, particularly in its wild-type state. However, having collected data regarding the movement, density, and secondary structures of these simulations, it is still difficult to pinpoint the role that the C-terminus plays in PTN-PTP interaction. In some respects, the mutated system appears to achieve a better connection: the RMSF is lower across nearly all residues. On the other hand, the wild-type system is preferred in terms of hydrogen bonds and salt bridge formation, which could constitute key mechanisms for protein-protein binding (though this is just speculation, and the salt bridge metrics do not serve as reliable benchmarks for making a judgment on this matter; further, the exact mechanisms of binding between PTN and PTP are unknown). The role of the C-terminus in facilitating these interactions, nonetheless, is uncertain. Hence, while the modification of the protein does produce different results in the simulation, these results mostly appear to be symptoms of random movements in the long and flexible strand of residues constituting the C-terminus. Furthermore, the relevance of this outcome to the experimental results obtained by Ryan et al. appears to be fairly minimal, and the main effect that the mutated residue achieves is less flexibility and random movement.

An additional analysis I chose to perform for this system is a contact map. Using this tool allows for a visualization of the residue-residue interactions and distances throughout the simulation, which would seem beneficial for understanding which residues are in close contact during the binding process.

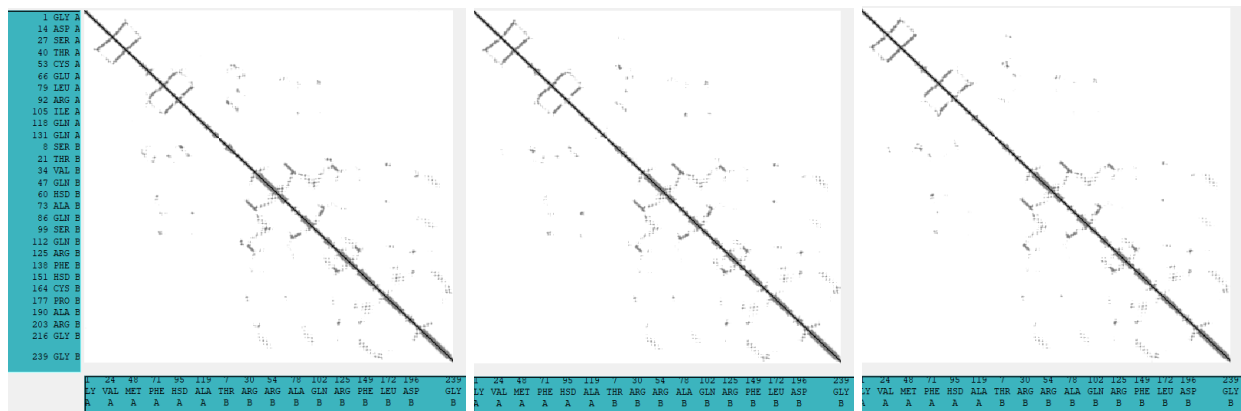


Figure 18. Wild-type System Contact Maps

Left: frame zero (0 picoseconds); middle: frame 800 (16,000 picoseconds); right: frame 1600 (32,000 picoseconds).

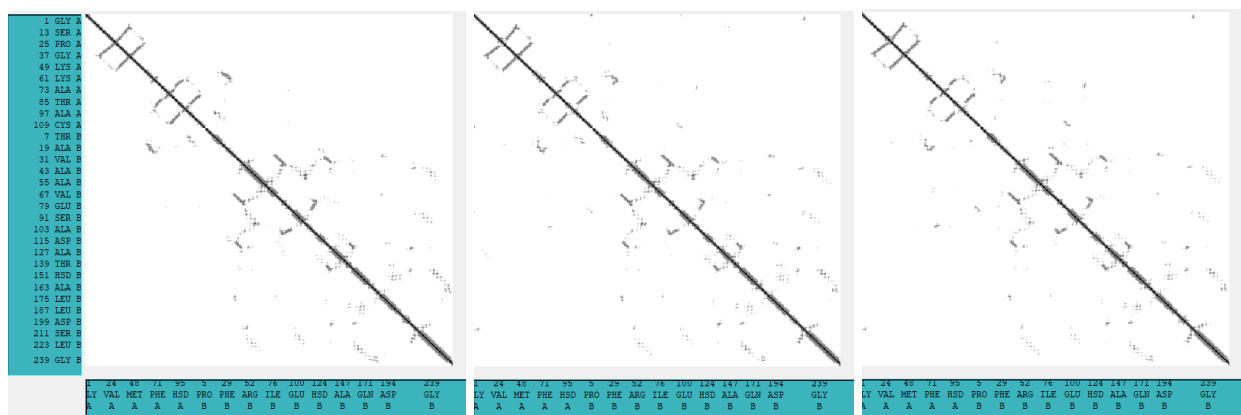


Figure 19. Mutated System Contact Maps

Left: frame zero (0 picoseconds); middle: frame 1050 (21,000 picoseconds); right: frame 2142 (42,860 picoseconds).

From a visual examination of the contact maps (figure 19), it appears that neither trajectory causes an extreme alteration in the configuration of the protein (in terms of proximity between amino acids). For the most part, secondary structures are conserved, which would explain the persistent patterns through the timeline. Looking at the lower left and upper right of each graph provides insights into the contact between these proteins; in the mutated system, between the first and second graph there are two notable instances of proximity that develop on the upper right side of the graph, one corresponding to the N-terminus coming close to PTP in the trajectory, and the other connecting residue 60 of PTN to 120 of PTP. The wild-type system almost appears to exhibit the opposite effect, with instances of proximity between PTP and PTN fading over time (note that, while the third frame may add

a few additional points of proximity, the general pattern stays the same, and the degree of proximity for the amino acids decreases across most points).

Another additional analysis I performed was VolMap, which I used to visualize the electrostatic potentials around the binding sites of PTN and PTP. Indeed, having a physical map of the electrostatic charges in either system could provide insight into how or why these proteins interacted the way that they did, as charges can either repel or attract each other, producing forces in the system.

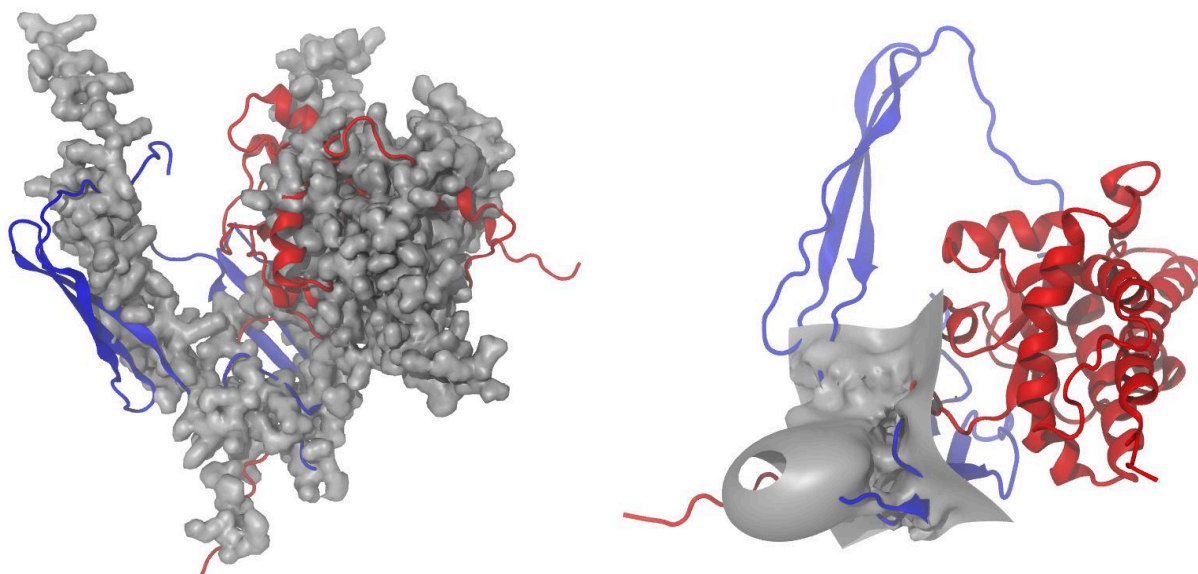


Figure 20. Volumetric Maps

Left: volumetric map for the wild-type system; right: volumetric map for the mutated system.

Strangely, performing identical operations for constructing the volumetric map yielded completely different results in either system, and I am not entirely sure of how to interpret this output. Despite specifying the binding site residues, the volumetric map of the wild-type system seems to encompass the entire protein. Moreover, the mutated system takes on a rather bizarre shape that is much flatter and smoother than its wild-type counterpart. Due to the relative similarities between the binding site interactions of PTN and PTP, it is difficult to assume that these vastly different depictions are accurate representations of electrostatic forces in my simulations.

Having initially selected my two simulations in an attempt to validate the experimental result achieved by Ryan et al., the primary conclusions that can be drawn from my analysis are not entirely contradictory to their findings, but also insufficient to corroborate them. The trajectories of the simulation fail to address the intricacies of the questions I was trying to answer about the binding properties of PTN. I attribute this shortcoming to the lack of a precise model for PTPRZ domains and my insufficient knowledge regarding the system I was trying to model (though, to be fair, scientific literature on this binding interaction is very limited and most of the research on this topic is rather challenging to comprehend). Indeed, the usefulness of MD simulation is heavily dependent on how it is set up. I mentioned in the methods section of this report that I attempted protein docking to CSA, which is a

ligand/cofactor that is directly mentioned by Ryan et al., and would have potentially provided a more accurate and meaningful result to the one that I attained. I dedicated a lot of time to searching for the most accurate model (PTPRZ) to dock with, but many of the PTPRZ models I stumbled across on the RCSB protein data bank included only one domain of PTPRZ (and none of these domains were the serine, glycine-rich domain that I was searching for, per Kuboyama et al.'s diagram on PTN-PTPRZ binding) [3]. In the end I settled on a model that looked close enough (low molecular weight PTP), but in hindsight, I do not think this model bore enough resemblance to the protein used in the experimentation to replicate what was performed by Ryan et al. (after doing more research on the PTP model I attained from AlphaFold, I discovered that the model I chose actually originates from a plant bacteria, not human cells). Indeed, Ryan et al. are very detailed in their conclusion, noting that the mutation in PTN decreases affinities for CSA and not CSE; such a specific outcome further speaks to how necessary it is to find the right model [5].

Hollingsworth and Dror mention several, more broader pitfalls of MD simulations. They write that force fields used in MD simulations are inherently approximate, so they may not capture all aspects of molecular interactions, leading to potential inaccuracies in simulation results. Moreover, covalent bonds are generally not formed or broken during MD simulations (quantum molecular mechanics simulations are required for this), limiting dynamical changes. Finally, as I experienced first-hand, accurate simulations rely on the availability of precise experimental protein structures as initial conditions. When designing a simulation, researchers are limited by available experimental structures that define what is possible when exploring specific biological phenomena [6].

Conclusions

Analyzing PTN and its binding partner PTPRZ through molecular dynamics simulations has offered some degree of insight into their interactions. My simulations aimed to further my understanding of how truncating PTN's C-terminus (residues 115 to 136) affects the protein's binding affinity with PTP (an approximation for PTPRZ). In my simulations, the dynamical behavior of both wild-type and mutated systems revealed motions in protein tails, with PTN's C-terminal potentially interacting with PTP in the wild-type system. However, the significance and persistence of this binding remains uncertain, as my trajectory encountered cross-PBC interactions that distorted data after 32,000 picoseconds into the simulation. Moreover, analysis of structural parameters indicated differences between systems, but interpreting these differences to either validate or challenge the experimental findings of Ryan et al. proved challenging. In comparing my wild-type system to mutated one, the former exhibited increased flexibility while the latter displayed lower overall movement. Hydrogen bond analysis further hinted at potential differences in binding mechanisms, favoring Ryan et al.'s conclusions. Contact maps provided insights into residue-residue interactions, showcasing distinctions between the wild-type and mutated systems, and I used VolMap to visualize electrostatic potentials (although I found it difficult to interpret the output of this plug-in).

Nevertheless, challenges in comprehending PTN-PTP interactions and the absence of a precise PTPRZ model complicated the analysis. Despite these limitations, performing MD simulations provided nuanced perspectives on the effects of C-terminal truncation in PTN. Acknowledging the constraints of

MD simulation, such as the need for precise model selection and accurate experimental structures, whilst also recognizing the ongoing advancements in the field and continuous updates to online databases, is crucial to gaining meaningful biological insights. While the MD simulations I performed on PTN and PTP offered some interesting findings, further validation and refinement of simulation parameters are essential for advancing our understanding of PTN-PTP interactions.

Realistically, if I were to study this protein further, I might attempt binding with a structure that is documented in scientific literature and also available online; I want to see what binding with this protein really looks like when I get the model selection right. Alternatively, I might pursue a different path entirely and investigate the metabolic implications of PTN in peripheral organs, which is a phenomenon documented by research and seemingly more applicable to day-to-day life. Simulations under various conditions could explore PTN's effects on preadipocyte (precursor cells to fat cells) proliferation, mammary gland development, and liver structural integrity. By investigating how PTN contributes to glucose and lipid homeostasis, and insulin sensitivity, I could uncover the multifaceted roles of PTN in maintaining metabolic equilibrium [2]. While modeling at such a large scale can pose challenges for molecular dynamics simulations, the use of coarse graining and other techniques may aid in simplifying these models.

References

- [1] AlphaFold. 2022. Low molecular weight protein-tyrosine-phosphatase ptp. *AlphaFold structure prediction*: <https://alphafold.ebi.ac.uk/entry/A0A251YLR6>
- [2] Ballesteros-Pla, C., Sánchez-Alonso, M. G., Pizarro-Delgado, J., Zuccaro, A., Sevillano, J., & Ramos-Álvarez, M. P.. 2023. Pleiotrophin and metabolic disorders: insights into its role in metabolism. *Frontiers in endocrinology*, 14, 1225150. <https://doi.org/10.3389/fendo.2023.1225150>
- [3] Kuboyama K, Fujikawa A, Suzuki R, Tanga N, Noda M. 2016. Role of Chondroitin Sulfate (CS) Modification in the Regulation of Protein-tyrosine Phosphatase Receptor Type Z (PTPRZ) Activity: PLEIOTROPHIN-PTPRZ-A SIGNALING IS INVOLVED IN OLIGODENDROCYTE DIFFERENTIATION. *J Biol Chem*. 291(35): 18117-28. doi: 10.1074/jbc.M116.742536. Epub 2016 Jul 21. PMID: 27445335; PMCID: PMC5000061.
- [4] Ryan, E., Shen, D., & Wang, X. 2015. 2N6F. *RCSB Protein Data Bank*: <https://www.rcsb.org/structure/2N6F>
- [5] Ryan, E., Shen, D., & Wang, X. (2016). Structural studies reveal an important role for the pleiotrophin C-terminus in mediating interactions with chondroitin sulfate. *The FEBS Journal*, 283 (8), 1488-1503. <https://doi.org/10.1111/febs.13686>
- [6] Hollingsworth, S. A. Dror, R. O. 2018. Molecular Dynamics Simulation for All. *Neuron*. 99: 1129-1139.
- [7] Vajda Lab and ABC Group. Cluspro 2.0 protein-protein docking. *Boston University and Stony Brook University*: <https://cluspro.org/login.php>
- [8] Wang X. Pleiotrophin: 2020. Activity and mechanism. *Adv Clin Chem*. 98:51-89. doi: 10.1016/bs.acc.2020.02.003. Epub 2020 Mar 12. PMID: 32564788; PMCID: PMC7672882.