# R19Q Mutation in the 2L9N SBDS Protein

Ucer, Nusret Efe

March 15, 2025

1 Introduction

Shwachman–Diamond Syndrome is an autosomal recessive disease that affects many parts of the body, including bone marrow, pancreas, bones, immune and the central nervous system [1], with an increased risk of progression to myelodysplastic syndrome [2]. In most cases, SDS is associated with mutations in the Shwachman–Bodian–Diamond Syndrome gene, located on chromosome 7q11 and encoding for a protein, named SBDS, structurally organized into three highly conserved domains visible in Figure 1: the N-terminal FYSH domain, the central helical domain, and the C-terminal domain [3]. SBDS proteins are highly flexible, with the N-terminal FYSH domain and the C-terminal ferredoxin-like domain capable of undergoing substantial rotational adjustments with respect to the central domain [4]. SBDS plays an important role in the assembly of mature ribosomes—last step of ribosome assembly—and ribosome biogenesis, interacting with another protein called Elongation Factor Like-1 (EFL1) and triggering the GTP-dependent release of eIF6 (eukaryotic initiation factor 6) from 60S pre-ribosomes in the cytoplasm. The release of eIF6 activates ribosomes for translation competence—the ability to properly assemble and function in protein synthesis—and allows for the 80S ribosome assembly. In addition to ribosome assembly, SBDS also helps recycle eIF6 back to the nucleus, where it helps prepare new ribosomes. Given its crucial role in mature ribosome assembly and ribosome biogenesis, the SBDS protein aptly belongs to the ribosome maturation factor family. The protein is said to be primarily expressed in both the nucleus and the cytoplasm. SBDS localization is cell-cycle dependent, with nucleolar localization during G1 and G2 (first and second growth phases of the cell) and diffuse nuclear localization during S (Synthesis) phase [5]. As ribosome biogenesis and assembly mainly occurs during G1 and G2 phases, it is only natural that the SBDS protein be localized in the nucleus during those phases. Carrying an important role in ribosome biogenesis, SBDS interacts primarily with ribosomal RNA (rRNA), specifically with the 60S ribosomal subunit. By assisting in the maturation and assembly of ribosomal subunits, SBDS ensures they are correctly assembled and functional. While SBDS mainly interacts with rRNA in this process, it may also engage with other ribosomal proteins. Other than its interaction with EFL1 protein and rRNA, SBDS does not interact with any ligands or conventional co-factors.
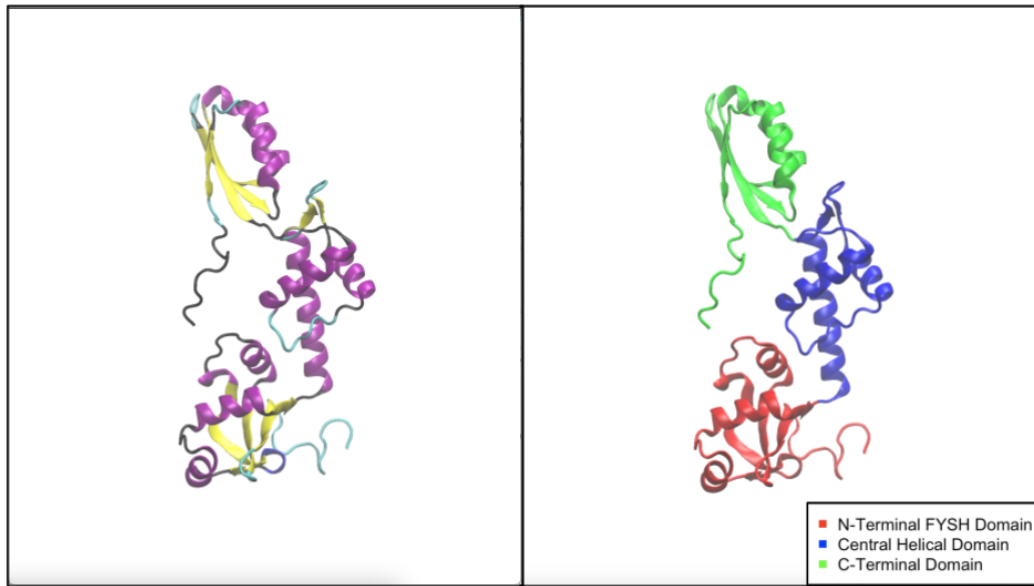
Figure 1. Wild Type SBDS Protein (2L9N) Colored by Secondary Structure (left) and Domain (right).

To sufficiently interact with EFL1 and rRNA, and the C-terminal domain shares structural homology with known RNA-binding domains, more specifically the RNA Recognition Motif (RRM). The role of the C-terminal in RNA binding is further supported by its extension in several plant SBDS orthologs, which seems to play a role in RNA interaction as they contain putative RNA-binding U1-type zinc fingers [6]. On the other hand, there are studies which argue that it is not the ferredoxin-like C-terminal domain but the novel N-terminal FYSH domain that is involved in RNA binding. Findings of Oliveira et. al. show that most strongly perturbed residues in SBDS are centered on the surface of the β-sheet formed by the two SHHP motifs on the N-terminal domain, suggesting that RNA makes its major contact with this region, more specifically, with possible binding sites Arg22, Lys25, and Arg26; whereas no residue in the C-terminal domain shows significant chemical shift changes in the presence of RNA, indicating that the C-terminal domain shows an extremely low affinity for RNA [6]. Other studies on RRM's role in protein-protein interaction also support that the RRM motifs do not necessitate that the C-terminus domain play a role in RNA interaction. In addition, it is also possible that the C-terminal domain still play a supporting role in RNA interactions even if not as major as the N-terminus FYSH domain. The last domain of SBSD, the central domain, has not been reported to be exclusively active in such processes, however, it is worth mentioning that as it contains a common winged helix-turn-helix motif, it is reported to be highly flexible and that its flexibility plays an important role to the functioning of the N-terminal FYSH domain.
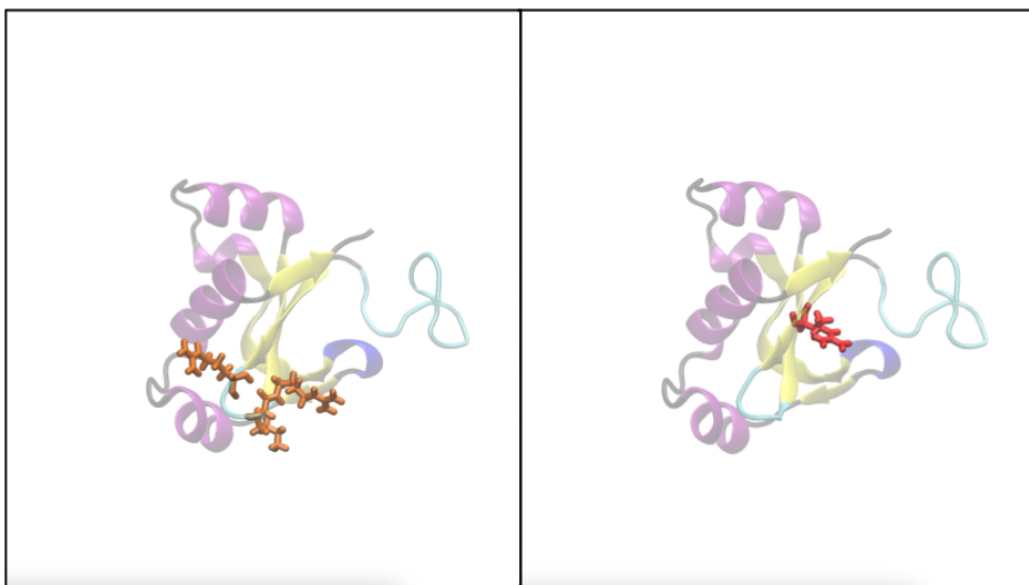
Figure 2. Side-to-Side Depiction of RNA Binding Sites on the N-Terminal FYSH Domain (left) and the Site of R19Q Mutation on the Wild Type N-Terminal FYSH Domain (right).

In light of the role of N-terminal FYSH domain in RNA binding and how most of the disease-related mutations in human SBDS have been reported to localize to the N-terminal domain, in close proximity to the RNA binding region, Oliveira et. al. suggests that mutations R19Q, K33E, and K67E occur in charged residues of the domain, possibly perturbing the RNA interaction by changing the basic character of the surface patch [6]. Spinetti et. al. further expands on the effects of R19Q which was previously shown by Gijsbers et. al. to have decreased the affinity of the second binding event by an order of magnitude, by conducting a 450ns molecular dynamics simulation and concurring with their hypothesis that SBDS function is governed by an allosteric mechanism involving domains I and III [3]. On the other hand, though much previously, Shammas et. al. had predicted that the third group of mutations (among which are R19Q, K33E, and K67E) affects the surface electrostatic potential or locally alters surface epitopes and should not disrupt the overall fold of the protein [5]. Therefore, I decided to take it upon myself to verify the findings of Spinetti et. al. at a much smaller extent, confirming that R19Q does indeed affect the fold of SBSD. In doing so, this paper predicts how a non-binding site mutation such as R19Q which was previously predicted to only affect the surface electrostatic potential impacts the overall fold of the protein and inherently, alters its function.

Molecular Dynamics Simulations have allowed us to put hypotheses such as that of Shammas et. al. to test, confirming or contradicting such claims. As this paper demonstrates, given a structure, what would have been predictions twenty years ago can now be simulated. With recent developments, an undergraduate such as myself, with the access to right hardware such as Midway3, the right force fields such as CHARMM, and the right software such as NAMD, can thereby verify the findings or test further studies of a post doctoral fellow such as Shammas. Highly specialized and advanced research of twenty years ago, requiring qualified and expert researchers, can now, to some extent, be conducted by students with minimal skills to prepare a protein in VMD. Its larger scale applications as done by Spinetti et. al., going up to much larger time scales can not only confirm hypotheses but provide immense insight into what normally would have been extremely tedious and expensive experiments or impossible to visualize effects. With MD, provided femtosecond by femtosecond

visualization at an atomic level resolution, in addition to the ability to regulate environmental conditions with certainty and impose conditions that cannot be tested in reality, allosteric mechanisms as done by Spinetti et. al. or endless other aspects of a biological system can be tested and explored. MD is thereby now widely used, as it was done in this paper, to explore the functions of proteins or to determine the structure of a protein.
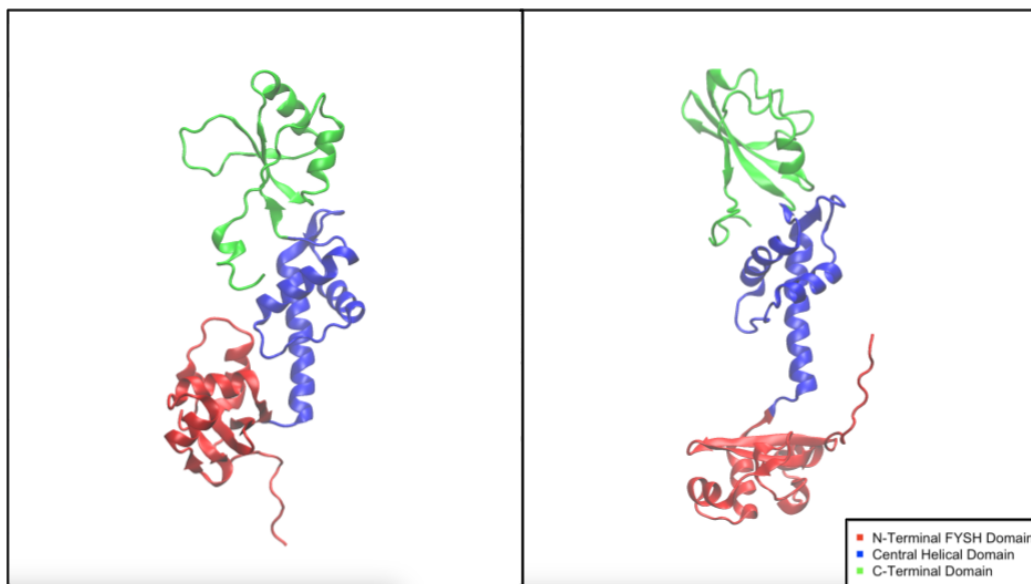
Figure 3. Comparison of Wild-Type SBDS Protein (2L9N) Colored by Domain (left) and Alternative Type SBDS Protein with R19Q Mutation Colored by Domain (right) after 40ns of MD Simulation.

2 Methods

To simulate Wild-Type and Alternate-Type SBDS Protein, first the Protein Databank File of SBDS protein was attained, PDB code 2L9N. As this is an NMR based file, it was noted that there were possible 20 frames, and in light that all frames were confirmed to be essentially the same, the first frame was picked for further simulation. Extracted and saved the coordinates of the first frame, the protein was then prepared for simulation. First, VMD's "Automatic PSF (Protein Structure File) builder" was used to build hydrogen atoms that might not have shown on the NMR. Then, using VMD's "Solvent Box" extension, 2L9N was put into a solvent box of water with 10 angstrom of padding on each side, rotating water molecules to minimize volume. Then, to neutralize the net charge of the solvent box and to simulate natural conditions, VMD's "Add Ions" extension was used to neutralize any net charge and add a 0.10 mol/L concentration of NaCl. Then the protein was centered using the Tkl console, determining the vector distance from the center and adjusting the pdb file accordingly to center 2L9N. Lastly, restraints were added to the protein backbone using the "Beta" column of the pdb file. The preparation of 2L9N was thereby concluded.

Then a minimization simulation was run on 2 nodes on the Midway3 supercomputer using CHARMM 36 parameters. Proceeding minimization, a heatup and equilibration job was run on, once again on 2 nodes on the Midway3 supercomputer and using CHARMM 36 parameters. Initialized with the previously set constraints and at at a temperature of 240

Kelvin, the temperature was increased by 5 K every 10 picoseconds up until a temperature of 300 K was reached, and it was lastly left to stabilize for 0.1 nanoseconds. Following the heatup, the restraints were gradually removed: first, they were scaled to 0.75 times its size, left to stabilize for 0.10 nanoseconds, then scaled to 0.5 times its size, left to stabilize for 0.10 nanoseconds, then scaled to 0.25 times its size, left to stabilize for 0.12 nanoseconds, then scaled to 0.125 times its size, left to stabilize for 0.15 nanoseconds, then scaled to 0.05 times its size, left to stabilize for 0.15 nanoseconds, and lastly removed, left to stabilize for 2 nanoseconds. The simulations were determined to run under NpT—a constant particle number N, a pressure p fluctuating around an equilibrium value p and a temperature T fluctuating around an equilibrium value T—conditions. For the parameters of NpT, the Langevin piston target pressure was set to be 1.01325 bar, the piston oscillation period to be 100 fs, and the piston decay period to be 50 fs. The overall Langevin Damping coefficient was set to be 1/ps. The minimization, heatup, and equilibrium simulations for 2L9N were thereby concluded. Lastly, two batches of 20 nanosecond NpT simulations, NPT1 and NPT2, were ran on the latest equilibrated 2L9N with a target pressure of 1.01325 bar and a target temperature of 300K. In total, this made over 40ns of simulation with a timestep of 1fs throughout the simulation, and at least the last 40ns were accepted to be at equilibrium.

To obtain the Alternate Type SBDS protein 2L9N-R19Q, the processes were started as it was done with the wild type but following the use of VMD's Automatic PSF builder", VMD's built-in "Mutate Residue" extension was used before proceeding further to mutate Arginine (R) of residue id 19 to Glutamine (Q). Thereby 2L9N-R19Q was obtained. After confirming that the 19th residue had been successfully mutated from Arginine to Glutamine, R19Q, all processes and simulations were repeated as it was done for the Wild Type.

3   Results and Discussion

3.1   Broad Analysis

After more than 40ns of NpT MD simulation, the two variants of SBDS protein have differed on some fronts and stayed similar on others; however, it is safe to say that the R19Q mutation has had a structural effect on the Wild-type 2L9N contrary to Shammas et. al. had predicted and in agreement with what Spinetti et. al. had observed in their own simulation. As previously predicted by Oliveira et. al., SBSD loses its functionality with regard to RNA interaction and binding as the N-terminal FYSH domain separates from the central helical domain and extends outwards, localizing and changing its shape as a result of the mutation, visible in the comparison in Figure 3. The main cause of this change seems to rise from the
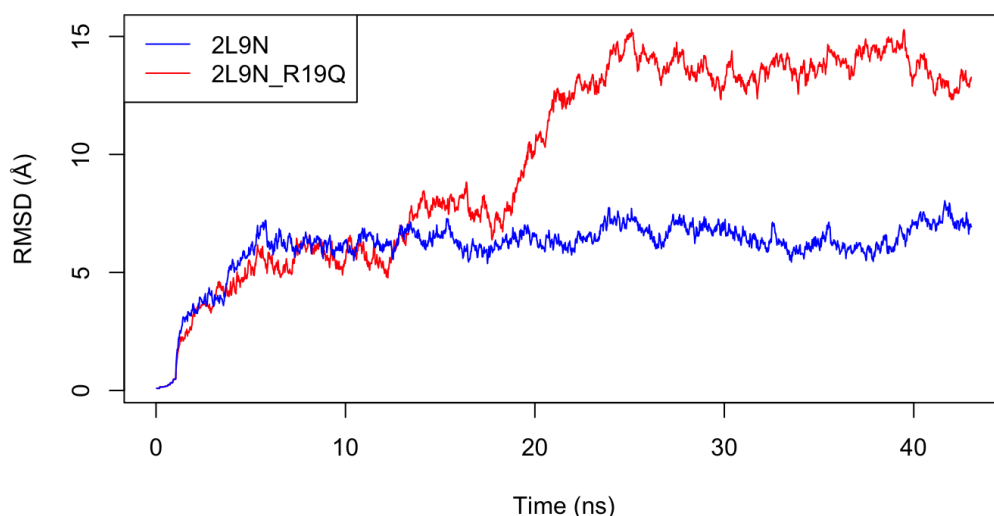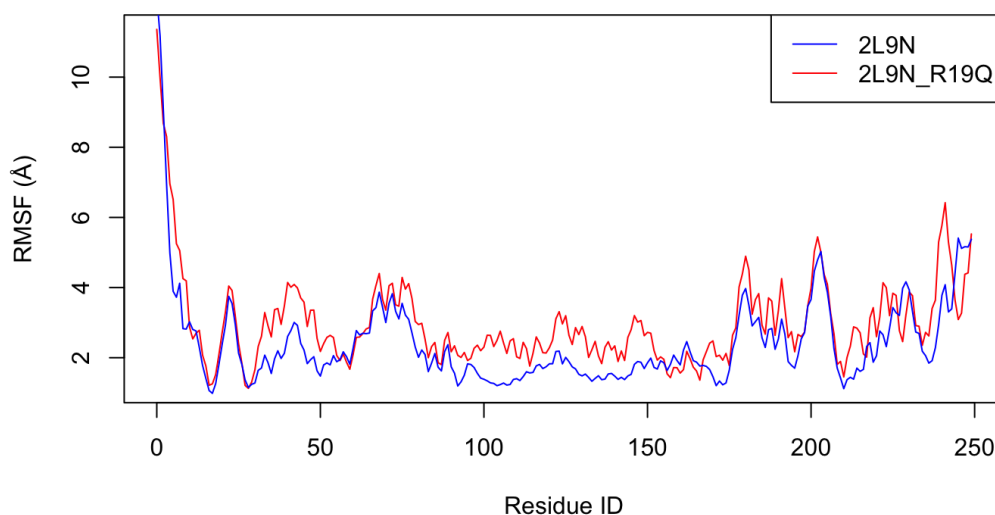
Figure 4. Comparison of Root Mean Squared Distance Values of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

For the approximately first half of the simulation, the RMSD values are quite similar as visible on the graph. Up until around 13ns, it can even be argued that the alternate type has a lower average RMSD value. However, starting after around 13ns and taking off around 17ns, the RMSD value of the alternate type 2L9N (2L9N-R19Q) increases drastically whereas the RMSD value of the wild-type 2L9N remains relatively stable. The reason to this sudden fluctuation and increase in RMSD values of the alternate type 2L9N is because of the dissociation of the N-terminal FYSH domain of the 2L9N from the central helical domain and its extension outwards. In the wild type and the start of the alternate type simulation, the N-terminal FYSH domain is curved towards the helical central domain and is seemingly attached there. The mutation seems to change that interaction and stability, allowing for the N-terminal FYSH domain to move freely and extend outwards from the central helical domain. That is the main reason why we observe, from approximately 17ns and onwards, an increase in RMSD. Once the motion ends, and the extended N-terminal FYSH domain stabilizes, and we can see the RMSD equilibrate, yet still in a more volatile manner as the N-terminal FYSH domain is not as constrained as it was in the wild-type, therefore fluctuating more violently. The first portion of the RMSD values are simply low as that is the part in which the protein is constrained and cannot move freely, therefore should be disregarded in the behavioural analysis and comparison of the alternate and wild-type.
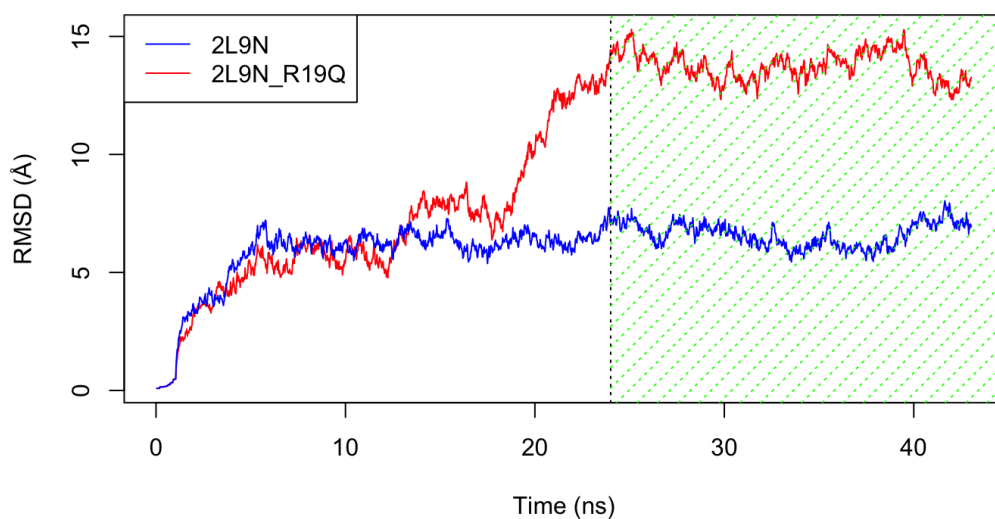
Figure 5. Designation of the Stable Region of the Comparison of Root Mean Squared Distance Values of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q) in Green.

Figure 6. Comparison of Root Mean Squared Fluctuation Values of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q) within the Stable RMSD Region (24ns onwards).

The RMSF values within the stable part of the RMSD designated in Figure 5 do not differ too much from one another; however, there are still important regions to note. The first important region to note is the region in between around resid 30 to 60. This region has a relatively higher average of RMSF. That is a portion of the N-terminal FYSH domain that was bent

and connected towards the central helical domain. This is the part that separates from the rest and extends outwards from the protein in the alternate type. Therefore, as it is a less restrained and thereby more unstable structure in the alternate type, it is logical that it has a higher RMSF value in that region. Another important region to note is the region in between around resid 100 to 150. This region has, once again, a relatively higher average of RMSF. That is the middle domain of the protein, to which the N-terminal FYSH was previously bent towards and connected to. The assumption is that the folded structure with the N-terminal FYSH and the central helical domain made both domains more stable. In the absence of such a bend, the regions are more fragile and volatile, therefore have higher RMSF values even in stable regions of the RMSD.
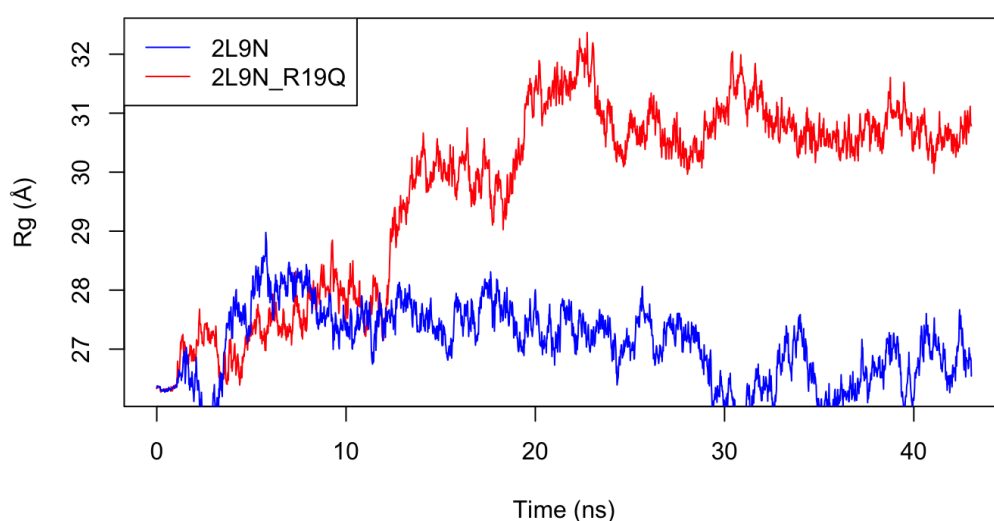

Figure 7. Comparison of Radius of Gyration Values of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

Radius of gyration effectively visualizes the effects of the R19Q mutation in the alternate type 2L9N. As visible, past approximately 12ns, the radius of gyration shoots up in the alternate type whereas it gradually decreases in the wild type. Because the alternate type misses the fold that holds the N-terminal FYSH bent towards the central helical domain, by definition of Rg, it is only natural that it shoots up as the N-terminal FYSH dissociatiates and extends outwards. Because in the alternate type an entire domain of the protein moves away from the COM, Rg is noticeably higher in that type. When both types are stable and in equilibrium, the difference in their shapes and how the orientation in which the N-terminal FYSH is held is noticeable in the plot; therefore, the Rg plot successfully confirms the findings of RMSF, and vice versa.
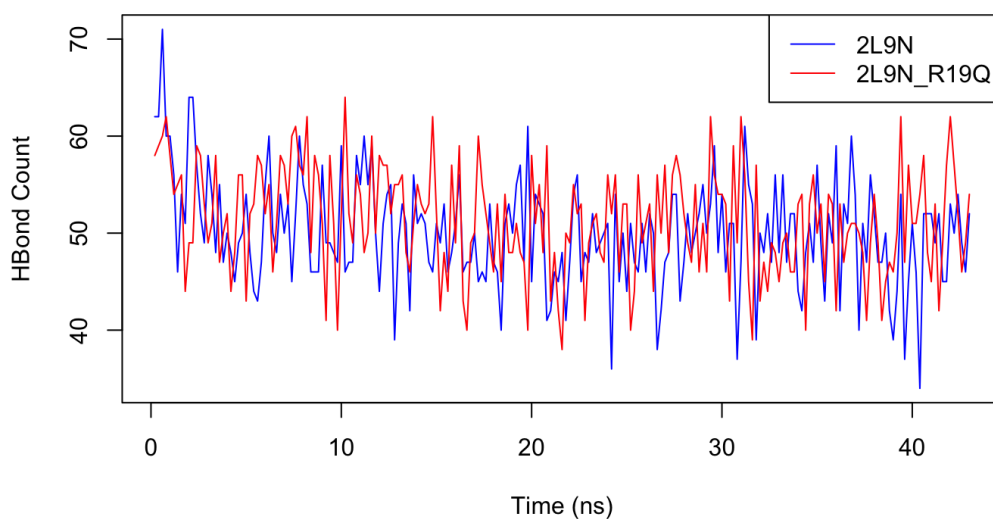
Figure 8. Comparison of the Number of hydrogen bonds wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).
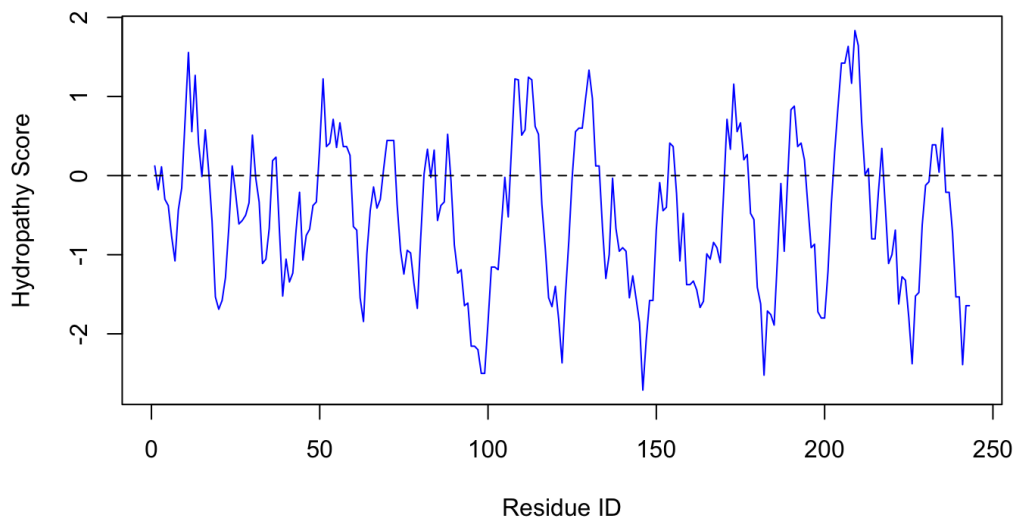

Figure 9. Kyte&Doolittle Hydropathy Score plot of wild type 2L9N with window size 9 and linear weight variation.

The hydrogen count of wild type and alternate type 2L9N don't seem to differ, they both seem to follow an average of 55 H-bonds. Contrary to RMSD and Rg plots, the swing of the N-terminal FYSH in the alternate type is not as visible in the H-bond count plot; spikes in the 10-20ns range do not have permanent effects. Consequently, the main problem with the deformation of the alternate type can be figured not to be the loss of hydrogen bonds but the

loss of salt bridges. Moreover, due to the abundance of hydrophilic residues visible in Figure 9, it is not improbable that different H-bonds form as 2L9N changes conformation.
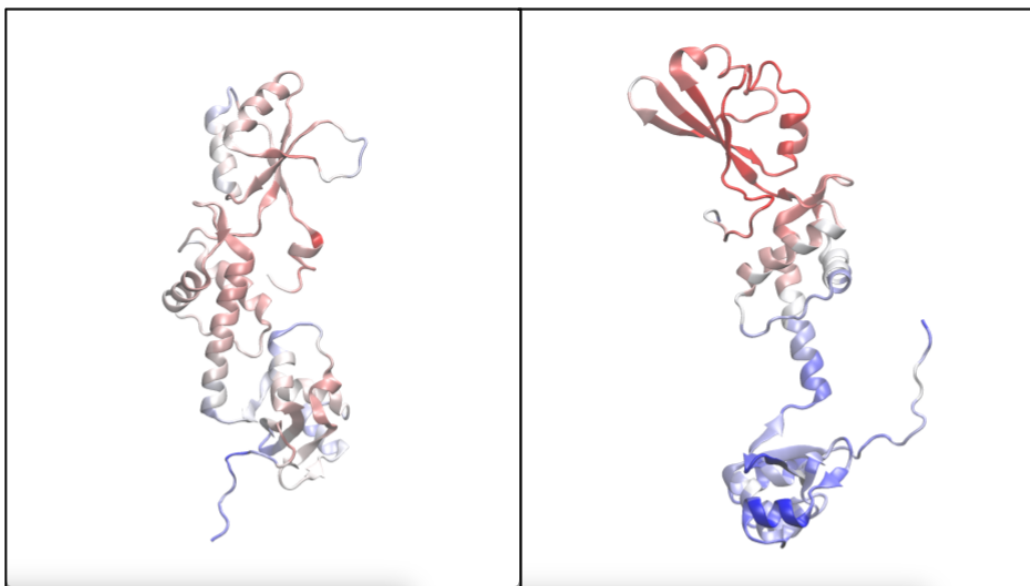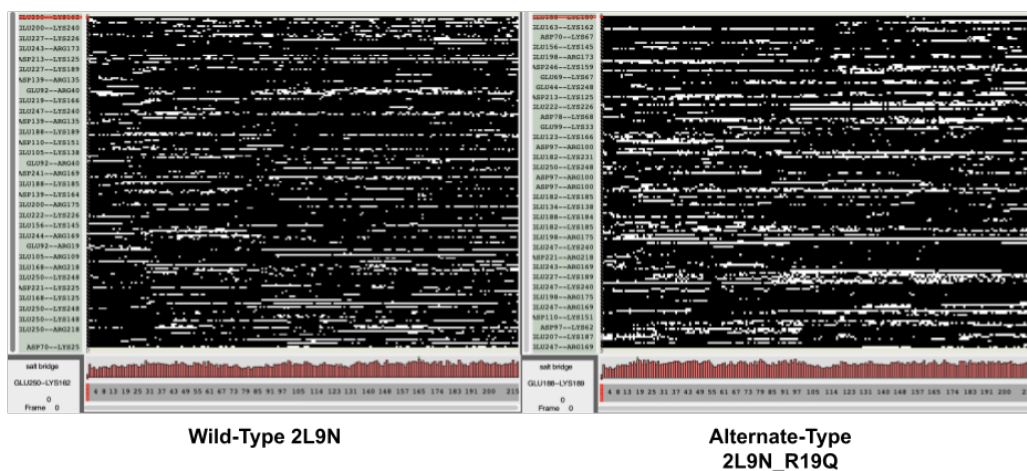

Figure 10. Comparison of wild type 2L9N (left) and alternate type 2L9N (right) by RMSF coloring during the dissociation of the N-Terminal FYSH Domain
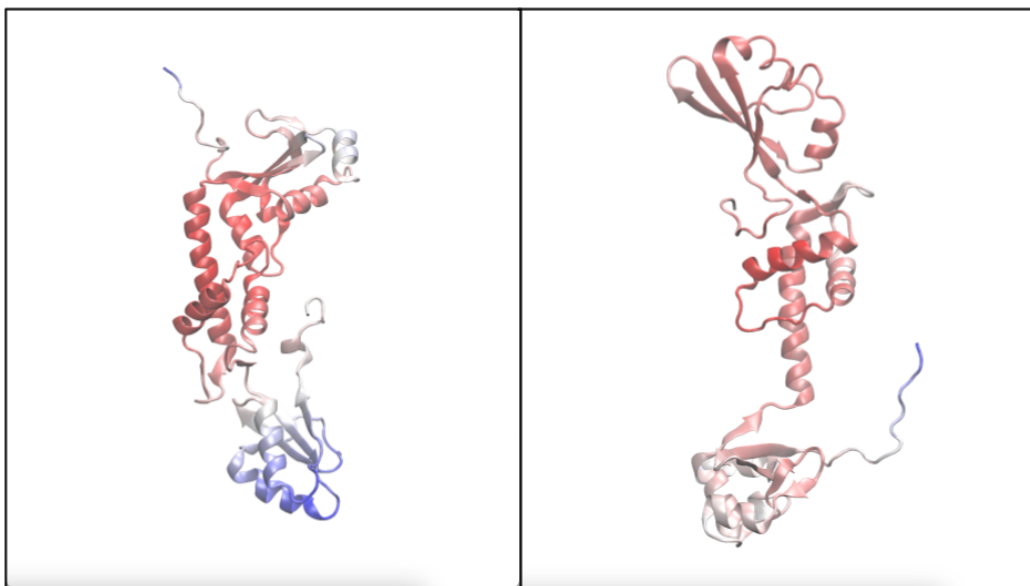

Figure 11. Comparison of wild type 2L9N (left) and alternate type 2L9N (right) by RMSF coloring after the dissociation and equilibration of the N-Terminal FYSH Domain

The main dissociation of the N-Terminal FYSH Domain from the central helical domain, evident in Figure 4 and Figure 7, approximately happens in between 10ns and 20ns. Within the scope of this effect of R19Q mutation, Figure 10 shows a colored comparison of wild type and alternate type 2L9N where mobility is associated with blue and stability, red. As expected from its swing, the N-Terminal FYSH Domain is shown to be considerably more blue, i.e. mobile in the alternate type. Figure 11, concerning the RMSF values after the

dissociation of the N-terminal FYSH domain and its stabilization, provides insight into what the alternate type 2L9N loses proceeding its movement in Figure 10. Clearly noticeable, the N-terminal FYSH domain is much more mobile in the wild type 2L9N. It was previously mentioned by Oliveira et. al. that the flexibility and mobility of the 2L9N was integral to its functionality. In the light that N-Terminal FYSH Domain is much more mobile and flexible in the wild type, it can be inferred that by losing its dynamic with the central helical domain, the alternate type most probably loses its function. Providing even further insight, Figure 11 also shows that it is not only the N-terminal domain but also the C-terminal domain that becomes more rigid because it also extends outwards in response to the movement of the N-terminal. This narrative further supports the loss of functionality in the alternate type due to loss of flexibility, mobility, and dynamic with the central helical domain.


Figure 12. Comparison of salt bridge timelines of wild type 2L9N (left) and alternate type 2L9N (right)
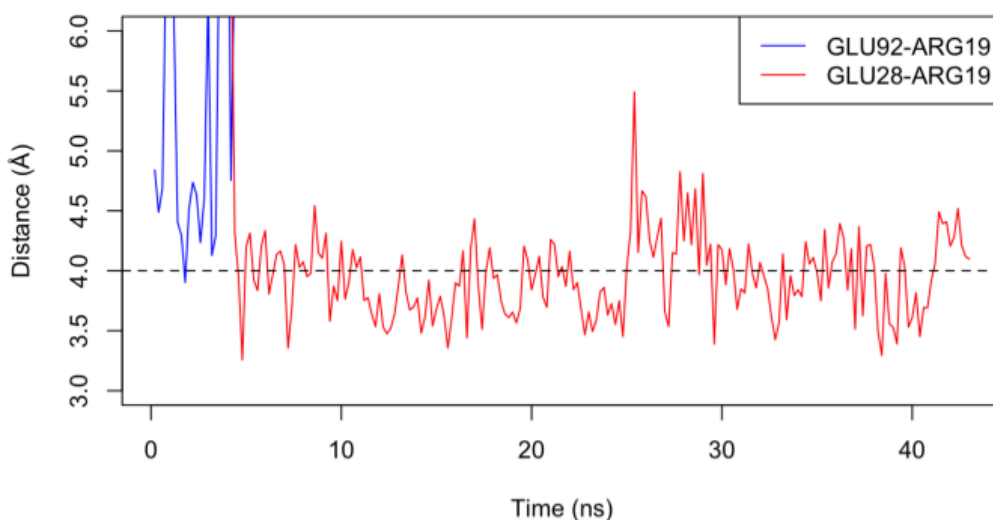
Figure 13. Distance plots of GLU28-ARG19 and GLU92-ARG19 salt bridges in the wild type 2L9N with an approximate strong salt bridge limit.

11

As aforementioned, the region spanning approximately from resid 30 to 60 is of integral interest because that's the region where the N-terminal FYSH domain connects to the central helical domain. In this context, residue 19 and the GLU92-ARG19 and GLU28-ARG19 salt bridges, directly associated with where the mutation occurs, is of integral importance to the comparison of the wild and alternate types. However, further analysis shows that GLU92-ARG19 is not nearly a strong salt bridge, or after the constraints are released, not a salt bridge at all. GLU28-ARG19, however, is relatively strong (especially if the upper limit is taken as 4.0 angstroms as given with the black dashed line) and consistent as visible in Figure 13, ensuring that the N-terminal domain stays connected on a second end to the central domain and prevents the N-terminal domain from extending outwards. Because R19Q mutates the Arginine to Glutamine, the GLU28-ARG19 salt bridge no longer exists in the alternate type, no longer fixing the beta sheet that holds the alpha helix with polar attraction in place, helping keep the N-terminal domain curved towards the central domain. The importance of the GLU28-ARG19 salt bridge is evident in that it exists throughout. In the alternate type, in the salt bridge's abscence the beta sheet deforms, causing the polar attractions to deterioate and the N-terminus to dissociate.
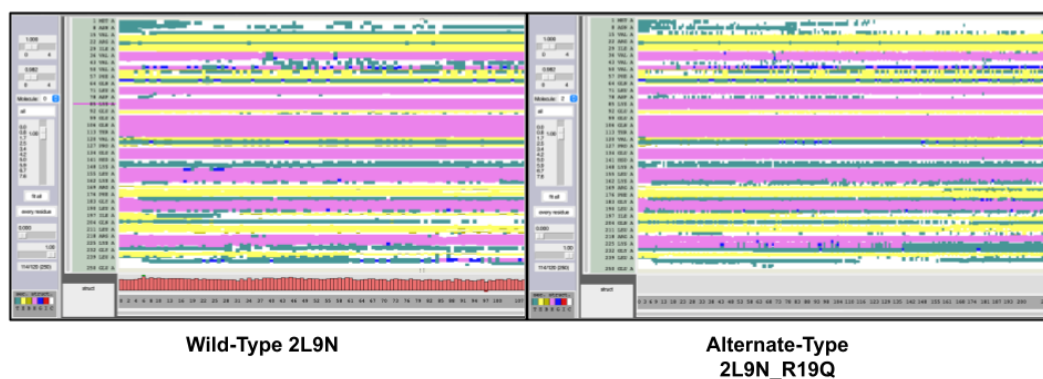

Figure 14. Comparison of secondary structure timelines of wild type 2L9N (left) and alternate type 2L9N (right)

There are no large secondary structure changes in neither the wild type or the alternate type. As visible, all colors indicating secondary structures remain very consistent throughout the simulation in the wild type whereas beta sheets especially around residue 50-60 are quite unstable in the alternate type relative to the wild type. There are only certain moments in which the turns are compromised in the wild type but they are usually regained. The part where the mutation occurs, the beta-sheet that ensures that the connecting alpha-helix is stable enough to connect to the main body by polarity does not suffer from changes in the wild type because of the GLU28-ARG19 salt bridge whereas the beta sheet arond resid 50-60 is comparatively unstable and keeps fluctuating in the alternate type.

Figure 15. Highlight of important residue ranges on the wild type 2L9N

The important regions of the protein mostly concern the beta-sheet in which the R19Q mutation occurs as highlighted in yellow (residue id 16 to 32) in Figure 15, and possibly the point at which the N-terminal domain (residue id 40 to 43) and the central helical domain (residue id 147 to 150) seemingly connect. Further assessment will mostly focus on the beta-sheet in which the R19Q mutation occurs as changes within that region cause other regions of interest to lose functionality.
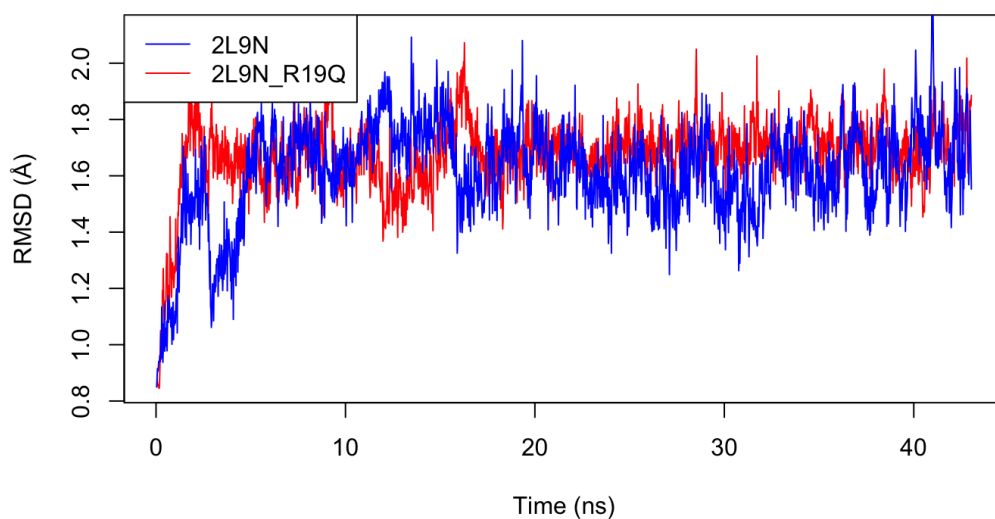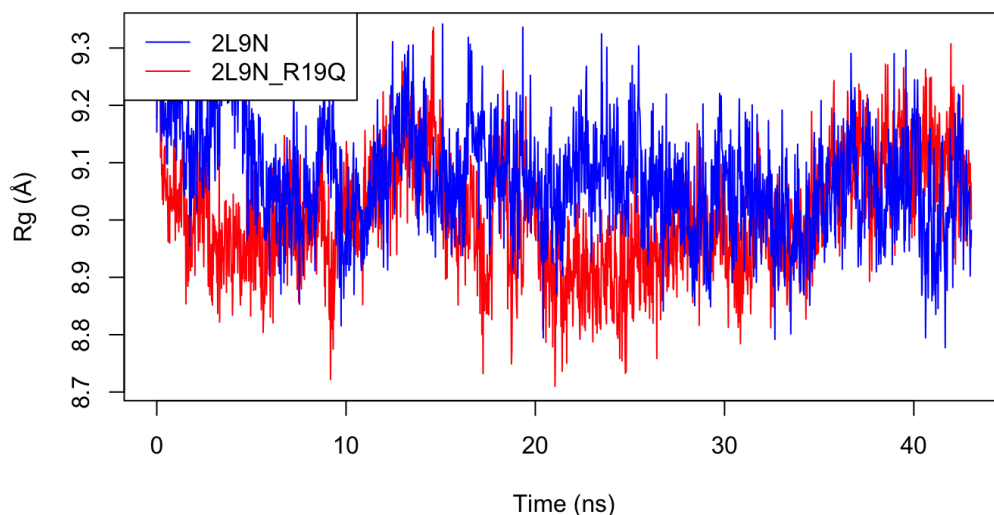
Figure 16. Comparison of Root Mean Squared Distance Values of residues 16 to 32 of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

The wild type 2L9N has an average RMSD of 1.614243 whereas the alternate type has an average RMSD of 1.674047, extracted from Figure 16. Though this difference may not be too indicative, it increases after the alternate type stabilizes, up around 20ns where the alternate type's average increases to an RMSD value of 1.705197 whereas the wild type's average only increases to 1.625101, opening up the difference to a more substantial magnitude. As one would expect for it to be revealed later in the H-bond count, this 0.08 angstrom difference should have mattered for beta sheet stability.
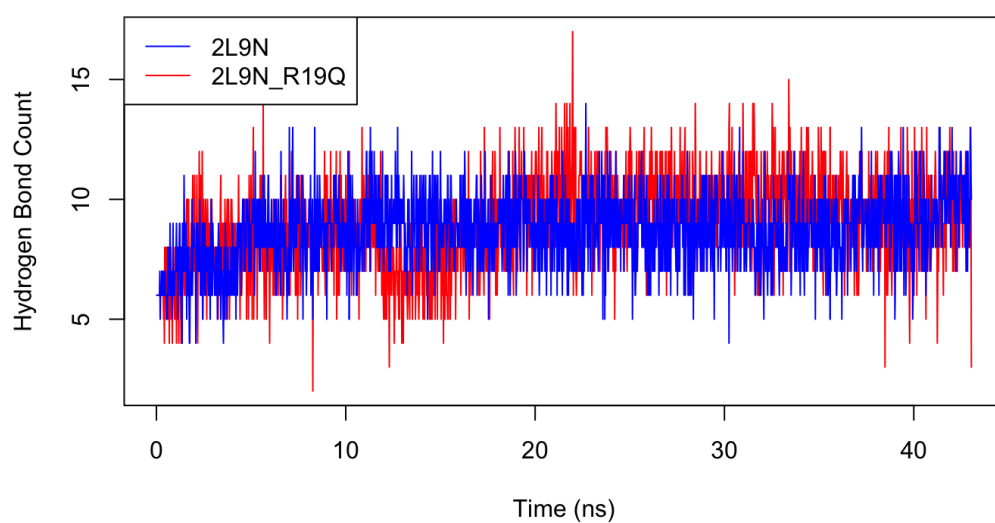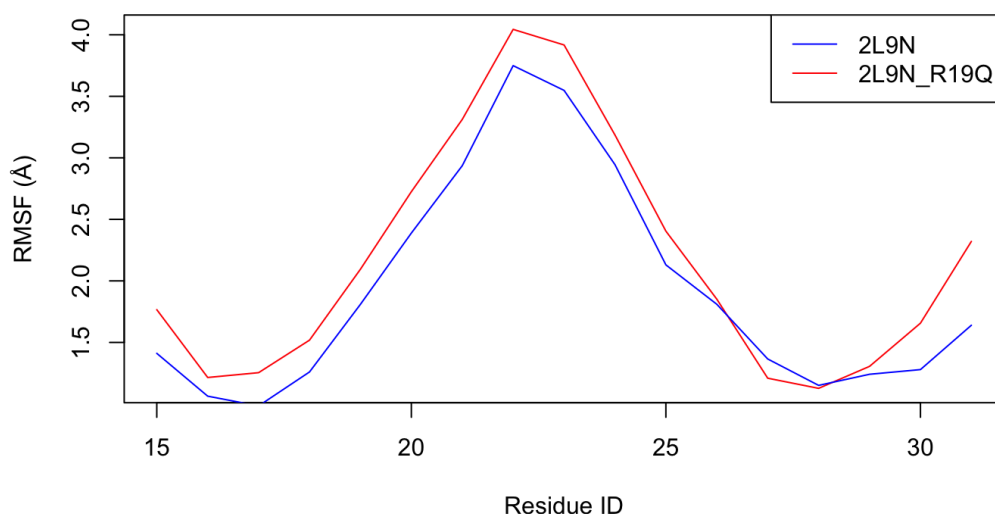

Figure 17. Comparison of Hydrogen Bond Counts of residues 16 to 32 of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

14

Interestingly, there are no differences in the hydrogen bond counts within the beta sheet of interest. Certainly, there are visual differences in that the beta sheet deforms as the simulation progresses in the alternate type, however, that is not necessarily reflected in the region of interest where the GLU28-ARG19 salt bridge is no longer associated with. In fact, whereas the wild type has a grand average of 8.690056 hydrogen bonds and an average of 8.892454 after it stabilizes, the alternate type has a grand average of 8.976766 hydrogen bonds and an average of 9.766696 after it stabilizes. This can possibly be interpreted as an increase in bonding in response to the loss of a salt bridge, especially taking into consideration that Glutamine is polar; however, to fully observe the mutations affect on the count of H-bonds would require more careful analysis than averages and sums as they are most probably dependent on one another.

Figure 18. Comparison of Radius of Gyration of residues 16 to 32 of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

In agreement with the simulation, the Rg value of the wild type 2L9N is on average higher. Though this may be deceiving at first, in some way implying the wild type is less stable, reflecting the trend by which the N-terminal domain gets more compact, the mutation may have an effect on the volume of the domain, restricting the effectiveness of RNA binding sites located within the area of interest. Further analysis will be conducted.

Figure 19. Comparison of Root Mean Square Fluctuation of residues 16 to 32 of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q) within the Stable RMSD Region (24ns onwards).

Unfortunately, the RMSF close-up comparison does not provide more insight than an average of 0.25 angstrom difference in fluctuation which does not give way to much interpretation. It should be noted, however, that including Figure 19, most plots indicate that the mutation has destabilized the protein.

3.2 Evaluation

MD simulation and analysis has provided extensive and valuable insight into the effects of the R19Q mutation, especially in terms of details of the magnitude of the effect and its implications on different domains in the protein. With regard to its broader implications of structural change, it would have been near impossible to predict that a change in a beta sheet which is not in relation to the part where the two domains interact could have had such a large effect. That is most probably why Shammas et. al. did not predict such an effect as well. On the static PDB, the R19Q mutation looks as if it will have no effect. However, with prior insight that binding sites occur in residues 22, 25, and 26, and that the N-terminal domain is the most important domain in the protein, suspicions would have arisen easier, information Shammas et. al. once again did not have. Furthermore, MD analysis and data gave important insights into not only how the mutation affects the mutated site but the entire protein as well. The C-terminal domain is heavily affected by the mutation as the mutation has grave implications on the overall inter-domain dynamics and flexibility as it was shown in Figure 11 and that could never have been figured from the PDB alone. Without such knowledge, half of the included figures would not have made sense, once again proving how critical it was to see the implications of the mutation simulated and its implicit dynamics observed.

Now it can be safely said that the R19Q indeed has effects on the overall fold of the 2L9N protein and affects the flexibility and domain interaction of the protein in a grand scale. Furthermore, it can also be said that the mutation causes the most structural change

within the N-terminal domain, an important part of the protein in terms of functionality as it contains the RNA binding sites that are of upmost importance to the protein's function. Possible changes in its volume and shape, raised by the close analysis of the region introduce the need for further analysis in that regard, after which a better conclusion will be reached in terms of binding site functionality.
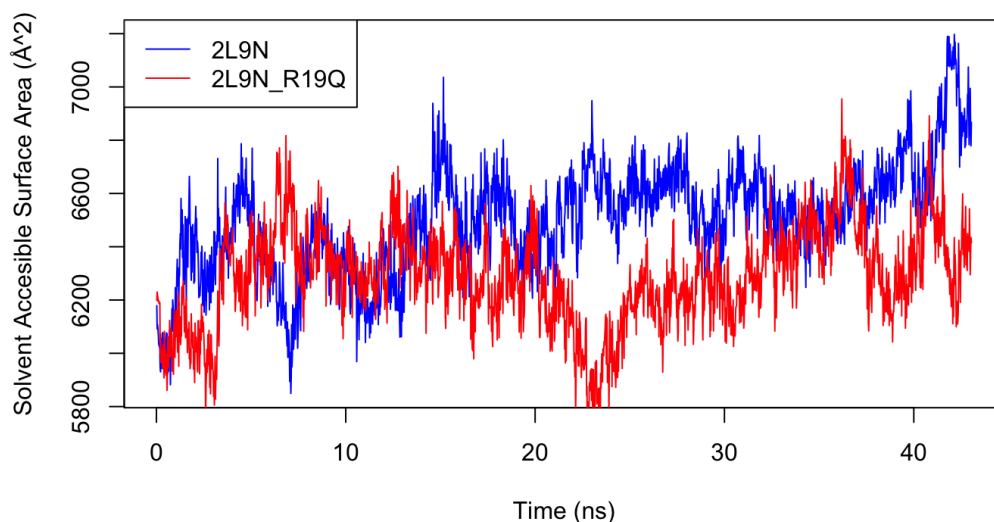
3.3   Further Analysis

Wild-Type 2L9N   Alternate-Type 2L9N_R19Q
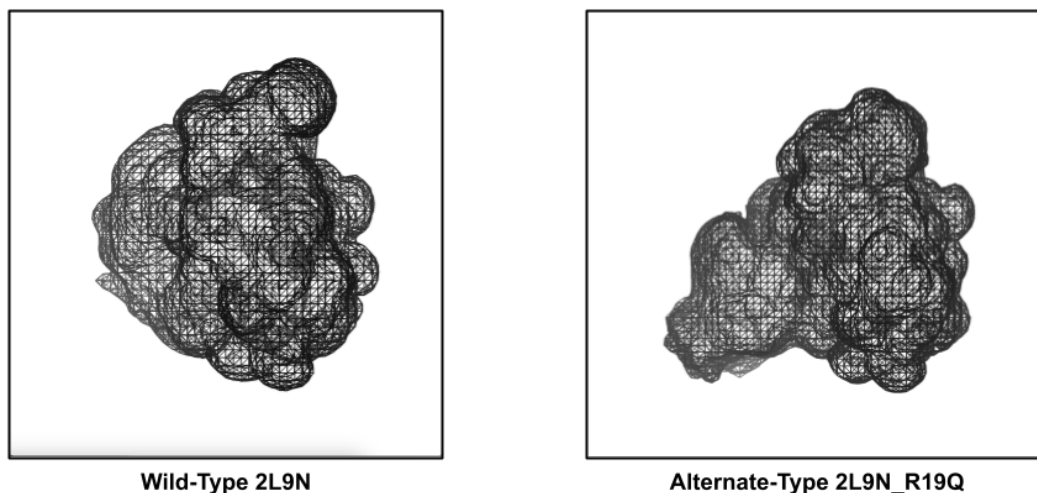
Figure 20. Comparison of Volume Mappings of the N-terminal FYSH domains (residue id 1-95) of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

Wild-Type 2L9N   Alternate-Type 2L9N_R19Q
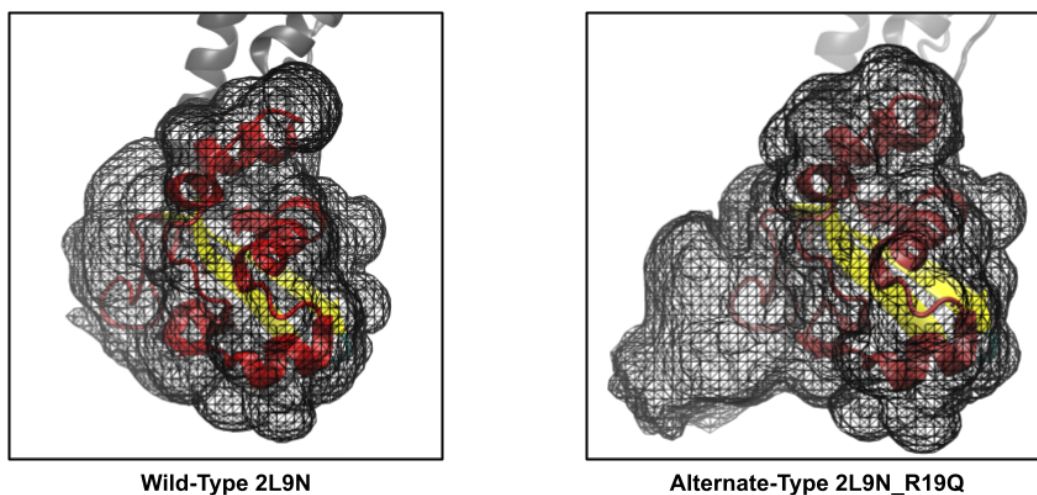
Figure 21. Comparison of Volume Mappings of the N-terminal FYSH domains (residue id 1-95) of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q) with visible colored domains.

As aforementioned, R19Q subjects the N-terminal domain to structural change, and as found in Figures 16, 17, and 18, indicatively with regards to its volume. A higher average RMSF, more hydrogen bonds, and a lower average Rg point to a decrease in volume because if the alternate type is more different than its original state compared to the wild type (indicated by the average RMSD) combined with a lower average Rg, then it should have shrunk in volume

17

and accordingly changed in shape. This analysis is conducted to visualize how the average volume throughout the simulation changes and is shaped, indicating loss of function for the binding sites in accordance. With such a visualization, obtained using VMD's "VolMap Tool", seeing that the normally spherical domain has been deformed and shrunk we can infer that the RNA binding sites located within the domain will become less active as its access will be more restricted.

Figure 22. Comparison of Solvent Accesible Surface Area of the N-terminal FYSH domains (residue id 1-95) of wild type 2L9N (labeled 2L9N) and alternate type 2L9N (labeled 2L9N-R19Q).

Now that it's been shown in the previous further analysis that the volume has shrunk and the shape has deformed, its effects on accessibility of binding sites should be observed. A higher Solvent Accessible Surface Area (SASA) generally indicates greater flexibility and potential for interactions, while a lower SASA suggests a more compact conformation. By using the Solvent Accessible Surface Area script, it is observed in Figure 22 that the N-terminal domain of wild type 2L9N has higher SASA compared to the shrunk alternate type, better able to interact and bind to RNA. Not only that, but as it was aforementioned that flexibility of especially the N-terminal was of integral importance to the functionality of the protein and its bindings sites, a lower SASA in the alternative type is all the more indicative of a loss of function. Furthermore, destabilizations in the hydrophobic core (as SASA provides surface area) can also be observed in the alternate type, fluctuating more aggressively compared to the wild type; likewise indicating that the conformational changes may reduce the effectiveness of the active site.

3.4 Final Discussion

Beginning the comparison and analyses of the wild and alternate type 2L9N, my expectations were that the N-terminal domain would be affected heavily by the mutation and that that would destabilize the protein. The simulation has to a great extent confirmed this: the protein was destabilized, its functionality was reduced, and N-terminal domain was susceptible

to these altercations. On the other hand, the mutation affected other domains to a much greater extent than I had expected and the importance of the flexibility of the protein was overlooked in my expectations. The domain-domain interaction was extremely important as it was emphasized in the sources and was probably responsible for a large extent of the loss of functionality. By showing this, MD was extremely insightful and it helped to confirm a lot of analysis done by myself and others in their studies. It was also assuring to see that my simulation was supported by other studies, providing similar results and conclusions. However, where MD fell short was the lack of complex functional analysis, for which I'd either have to conduct much more complex further analyses or different simulations. For example, an experiment like that of Oliveira et. al. in which they observe domain function using NMR in an RNA saturated solution would be extremely insightful, however, that would require a whole new MD simulation be conducted. MD thereby falls short in that it has its limitations in replicating reality and its results cannot be used as flexible as a real prepared protein. To use MD in such a way, one would either have to prepare multiple systems piling up an immense computational cost or prepare extremely intricate simulations in which unconfirmed effects could be hard to model. Furthermore, Spinetti et. al.'s work show that the alternate type, though it seems to stabilize around 20ns, destabilizes once more around 50ns and even more in the future. That is why they have run 450ns of MD on the mutation. Because of the time problem with MD, I was not able to recognize that I had hit such a local energy minima, and if not for Spinetti et. al., it would have been very probable that this conformation be designated the favorable conformation for the alternate type. Due to time and size limitations, it is also a shortcoming of MD that we can't necessarily know we've hit certain limits without immense cost.

4   Conclusion

In this paper, the wild and alternate type of the SBDS protein was extensively discussed, analyzed, and compared and contrasted. The 2L9N base model was used to simulate both systems, with the alternate type's 19th residue mutated from Arginine to Glutamine. Its implications were loss of flexibility in the overall protein, an overall destabilization of the protein domains, and a predicted loss of functionality in the RNA binding site in the N-terminal FYSH domain, possibly due to a decrease in volume and solvent accessible surface area indicated by further analyses. The alternate type is significantly less stable. The overall findings are verified by other sources whereas its detailed explanations can use an further discussion. The heavily inspected movement of the N-terminal domain imitates the findings of Spinetti et. al. and values such as RMSD and RMSF fit their plots.

For further studies, I would suggest a similar experiment like that of Oliveira et. al. be done with the wild and alternate type: submerging the equilibrated systems in rRNA saturated solutions and analyzing domain based and binding site activation. This would shed further light into the role of the C-terminal domain for the SBDS protein as well as that has been yet to be completely resolved. Not only that, but more importantly this will provide insight onto the role of the shape and volume of the N-terminal domain in RNA binding, possibly confirming predictions made in the Further Analysis section of Discussion. However, it should be noted that the equilibrated alternate type here is, as indicated by Spinetti et. al., a local minima and that before proceeding with any such experiments the truly equilibrated alternate type of 2L9N should be attained. If provided a truly equilibrated wild and alternate type, further studies could saturate the solution with rRNA and graph the interactions of the

designated binding sites or overall any other sites with the rRNA. If ran for a significant while, this study will provide what for what parts of the protein does rRNA possess most affinity and what parts of the protein possess most affinity for rRNA. Thereby we can confirm for once and for all the functions of domains of SBDS in rRNA binding and observe the mutations affects on rRNA binding affinity, possibly providing insight onto why binding site functionality may be impaired.

5 Citations

[1] Dror, Y., Donadieu, J., Koglmeier, J., Dodge, J., Toiviainen-Salo, S., Makitie, O., Kerr, E., Zeidler, C., Shimamura, A., Shah, N., Cipolli, M., Kuijpers, T., Durie, P., Rommens, J., Siderius, L., Liu, J. M. (2011). Draft consensus guidelines for diagnosis and treatment of Shwachman-Diamond syndrome. Annals of the New York Academy of Sciences, 1242, 40–55. https://doi.org/10.1111/j.1749-6632.2011.06349.x

[2] Finch, A. J., Hilcenko, C., Basse, N., Drynan, L. F., Goyenechea, B., Menne, T. F., González Fernández, A., Simpson, P., D'Santos, C. S., Arends, M. J., Donadieu, J., Bellanné-Chantelot, C., Costanzo, M., Boone, C., McKenzie, A. N., Freund, S. M., Warren, A. J. (2011). Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. Genes development, 25(9), 917–929. https://doi.org/10.1101/gad.623011

[3] Spinetti, E., Delre, P., Saviano, M., Siliqi, D., Lattanzi, G., Mangiatordi, G. F. (2022). A Comparative Molecular Dynamics Study of Selected Point Mutations in the Shwachman-Bodian-Diamond Syndrome Protein SBDS. International journal of molecular sciences, 23(14), 7938. https://doi.org/10.3390/ijms23147938

[4] Savchenko, A., Krogan, N., Cort, J. R., Evdokimova, E., Lew, J. M., Yee, A. A., Sánchez-Pulido, L., Andrade, M. A., Bochkarev, A., Watson, J. D., Kennedy, M. A., Greenblatt, J., Hughes, T., Arrowsmith, C. H., Rommens, J. M., Edwards, A. M. (2005). The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. The Journal of biological chemistry, 280(19), 19213–19220. https://doi.org/10.1074/jbc.M414421200

[5] Shammas, C., Menne, T. F., Hilcenko, C., Michell, S. R., Goyenechea, B., Boocock, G. R., Durie, P. R., Rommens, J. M., Warren, A. J. (2005). Structural and mutational analysis of the SBDS protein family. Insight into the leukemia-associated Shwachman-Diamond Syndrome. The Journal of biological chemistry, 280(19), 19221–19229. https://doi.org/10.1074/jbc.M414656200

[6] de Oliveira, Juliana Ferreira, Maurício L.Sforça, Tharin M.A.Blumenschein, Mauricio B.Goldfeder, Beatriz G.Guimarães, Carla Columbano Oliveira, Nilson I.T.Zanchin,and Ana-Carolina Zeri."Structure,Dynamics,and RNA Interaction Analysis of the Human SBDS Protein." Journal of Molecular Biology396,no.5(2010):1053–1069.https://doi.org/10.1016/j.jmb.2009.12.039.