

*Computational analysis of human prion protein's allosteric potential reveals potential pharmacological chaperone binding site for V189I and V203I PrP<sup>Sc</sup>*

Reagan McNeill Womack

Dr. Joshua Ring, advisor

Department of Chemistry, Lenoir-Rhyne University

Computational analysis of human prion protein's allosteric potential reveals potential pharmacological chaperone binding site for V189I and V203I PrP<sup>Sc</sup>

## Abstract

---

Chaperone molecules facilitate folding pathways to assist in proteins folding to their native conformation. For proteins that do not fold spontaneously, chaperonins may be required for folding to occur. This idea can be applied to already misfolded proteins by attaching a protein to a misfolded protein that destabilizes the misfolded conformation and assists in folding the protein back into its native conformation. Prion diseases are rare neurodegenerative diseases characterized by tissue deposition of a misfolded isoform of the cellular prion protein. The purpose of this study was to identify a potential site for pharmacological chaperone molecules to bind to on human prion protein that may aid in stabilizing the native conformation of the protein or protect a normally folded protein from being attacked by the infectious isoform in the presence of V189I or V203I mutations. Two structures of normal cellular prion protein (PrP<sup>C</sup>) were obtained from the Protein Databank (PDB) for analysis (PDB ID: 1QM2 and 1I4M). Eris, a protein stability prediction server, was used to perform an exhaustive analysis of missense mutations on prion protein. The results from this analysis were compared to the list of infectious missense mutations to the PRNP (Prion Protein) gene obtained from the Human Gene Mutation Database (HGMD). The ProteinLens webserver was used to compute Markov transient times for each residue with the source site set as the site of a missense mutation. These results were then used to test potential regulatory sites by exploring the effect of small probe binding and simulated mutation with the AlloSigMA server. Data were extracted from each server and analyzed using statistical methods in R and Python to locate any abnormalities and inconsistencies in the data. The results from these computational analysis tools suggest residues 149, 166, 167, 195, and 197 of human prion protein are all potential targets for pharmacological chaperone molecule binding.

# INTRODUCTION

---

Prion diseases are a class of neurodegenerative diseases caused by aggregation of a misfolded prion protein [1]. When a mutation in the gene encoding prion protein (PRNP) occurs, sometimes the prion protein does not fold into its native conformation ( $\text{PrP}^C$ ). When the protein misfolds due to one of many infectious mutations, it forms a stiff  $\beta$ -sheet rich conformation that aggregates and leads to neurodegeneration and a quickly progressing and fatal disease [2]. While not proven, the current working theory suggests that a misfolded prion protein ( $\text{PrP}^{Sc}$ ) could attach to normal cellular  $\text{PrP}^C$ , causing it to misfold and add to the growing chain of  $\beta$ -sheets, forming a structure called an amyloid fibril [3].

The focus of this research was to use known models of allosteric [4] and predicted effects of pharmacological chaperones [5] to predict potential regulatory sites that may prevent misfolding of human prion protein. Various computational methods were used to identify one potential binding site that could prevent  $\text{PrP}^C$  from misfolding in the presence of one of two mutations, V189I or V203I.

This project focuses on applying existing computational methods developed for identifying allosteric sites on allosteric proteins to the non-allosteric Human Prion Protein to identify a potential drug target site. While Prion Diseases are rare, applying methods such as those outlined in this report may prove beneficial in identifying drug target sites for other proteins such as synuclein and amyloid-beta peptide [6]. These two proteins, which are the site of disease for Parkinson's Disease and Alzheimer's Disease respectively, have shown prion-like transmissibility in various studies [7].

# BACKGROUND

---

Computational research allows for more people to collaborate at once without wasting resources or worrying about ethical concerns. While computational research has many applications, model-informed drug development (MIDD) was the focus of this research project. MIDD has the potential to streamline the entire drug development process and, when applied successfully, result in a much clearer and more accurate starting point for in-lab research [8].

## Prion Diseases

Prion Diseases, formerly known as transmissible spongiform encephalopathies (TSEs), are rare neurodegenerative diseases characterized by tissue deposition of a misfolded isoform of the cellular prion protein [9]. The agents of prion diseases are “prions,” or proteinaceous

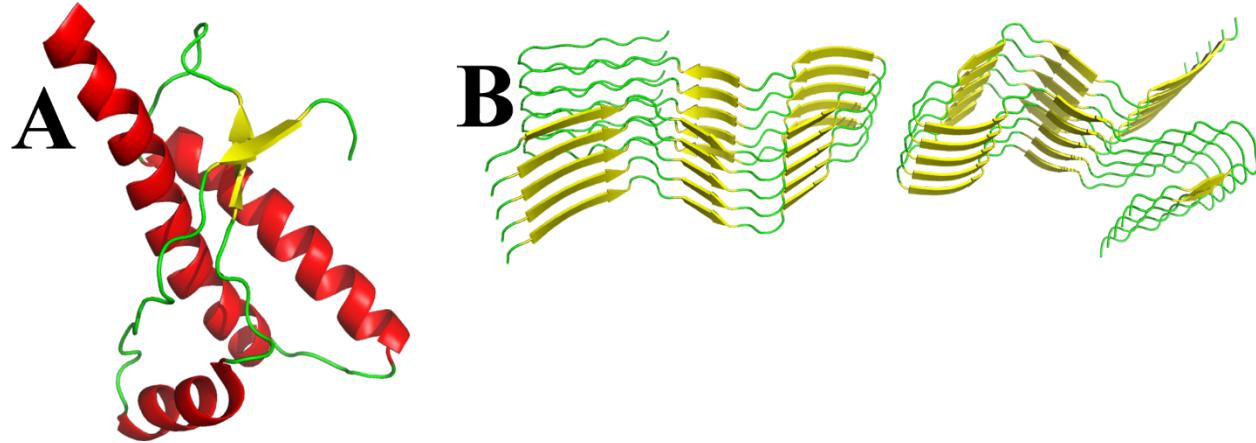
infectious particles that are devoid of nucleic acids. Prions are abnormal isoforms of cellular prion protein ( $\text{PrP}^{\text{C}}$ ), a cell surface protein present expressed in the nervous system. These abnormal isoforms differ from the normal protein only in their secondary structure (with an exception for inherited cases), meaning the amino acid sequence coding for the protein remains the same in the abnormal isoform as that for  $\text{PrP}^{\text{C}}$  [9].

Human prion diseases are especially rare prion diseases, affecting approximately 1-2 persons per million every year, and are often characterized by rapidly progressing and fatal central nervous symptoms [6]. Most human prion diseases are sporadic, making up about 85% of the cases. Approximately 10-15% of human prion disease cases can be classified as inherited and the small percentage that remains can be classified as transmitted [6].

The first described transmitted form of human prion disease was Kuru, which spread among the Fore people of Papua New Guinea in the 1950s and was linked to cannibalistic rituals performed by the people [10].

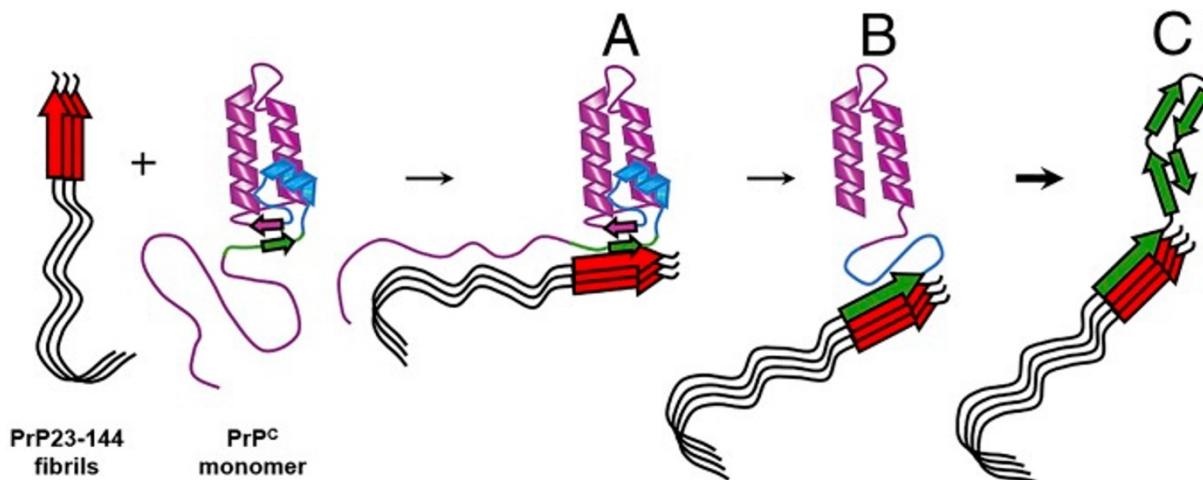
### Prion protein isoforms form infectious amyloid fibrils

The secondary structure of the abnormal isoform of prion protein ( $\text{PrP}^{\text{Sc}}$ ) is characterized by a stiff  $\beta$ -sheet rich conformation (see *Figure 1B*) that aggregates and leads to neurodegeneration and a quickly progressing and fatal disease [2].



**Figure 1.** Comparison of  $\text{PrP}^{\text{C}}$  and  $\text{PrP}^{\text{Sc}}$  structures. (A) shows the structure of 1QM2 [11], a representation of  $\text{PrP}^{\text{C}}$ ; (B) shows the structure of 6LNI [12], a representation of the  $\beta$ -sheet rich conformation of  $\text{PrP}^{\text{Sc}}$ .

This  $\beta$ -sheet rich conformation is also referred to as an amyloid fibril, which have been shown to be infectious [3]. The discovery of these amyloid fibrils led to the current working theory of prion pathogenesis, where an infectious prion attaches to a normal  $\text{PrP}^{\text{C}}$  and adds it to a growing structure that provides explanation for how the neurodegeneration seen in these diseases occurs [3].



**Figure 2. Current working theory for prion pathogenesis [3].**

Prion diseases aren't the only amyloid diseases (collectively referred to as amyloidoses); Alzheimer's disease and Parkinson's disease are both amyloidoses [13]. Methods used to develop treatments for prion diseases may prove beneficial for treating these diseases and any other amyloidoses.

Amyloid fibrils are stabilized by aromatic and hydrophobic side chains, meaning binding to these side chains before or after a misfold occurs may prove beneficial in breaking up the amyloid fibril. Drugs that can correct CTFR misfolding events (causing Cystic Fibrosis) have been developed; these drugs are often called "pharmacological chaperones" [13].

## Prion protein folds under kinetic control

Competing protein folds can take place under kinetic control or thermodynamic control. In many cases, proteins fold under thermodynamic control, meaning the resulting protein fold is the most stable conformation [14]. For prion protein, however, evidence shows that it is under kinetic control as the native conformation of the protein is not the most stable conformation [15]. This means that the native conformation of prion protein is simply the quickest conformation for the protein to fold to [16].

The actual function of normal cellular prion protein is not well understood and only a surface level understanding is available for the function of PrP<sup>C</sup> in the peripheral nervous system. PrP<sup>C</sup> is highly expressed in both the peripheral nervous system and the central nervous system, which explains why the aggregation of mutant variants leads to neurodegenerative diseases. While the function of PrP<sup>C</sup> is still being investigated, it is proposed to have a physiological role in sleep, memory formation, calcium homeostasis, and many other cellular regulation systems [17].

## Creutzfeldt-Jakob Disease

Creutzfeldt-Jakob Disease (CJD) is the most common human prion disease worldwide, particularly sporadic CJD (sCJD) [18]. CJD can be classified as sporadic, transmitted, or inherited, much like how human prion diseases are classified [18].

Variant CJD (vCJD) was first described in 1996 following dietary exposure to Bovine Spongiform Encephalopathy (more commonly known as “mad cow disease”) [19]. vCJD can be classified by transmission, as the disease only occurs from exposure to another prion disease.

## Allosteric binding sites regulate protein activity

A protein’s activity is regulated primarily by its active site. When the protein’s unique ligand binds to the active site, a shape change is induced so that the protein can be used to do work. Separate from the active site, some proteins (called allosteric proteins) also have supplemental binding sites called allosteric binding sites. Allosteric binding sites affect a protein’s overall activity along with the active site. When a ligand binds to an allosteric binding site, like with the active site, a shape change occurs that increases or decreases the protein’s activity. Allosteric binding sites can also affect the thermodynamics of a cellular reaction when the protein in question is involved.

## Missense mutations can alter allosteric response

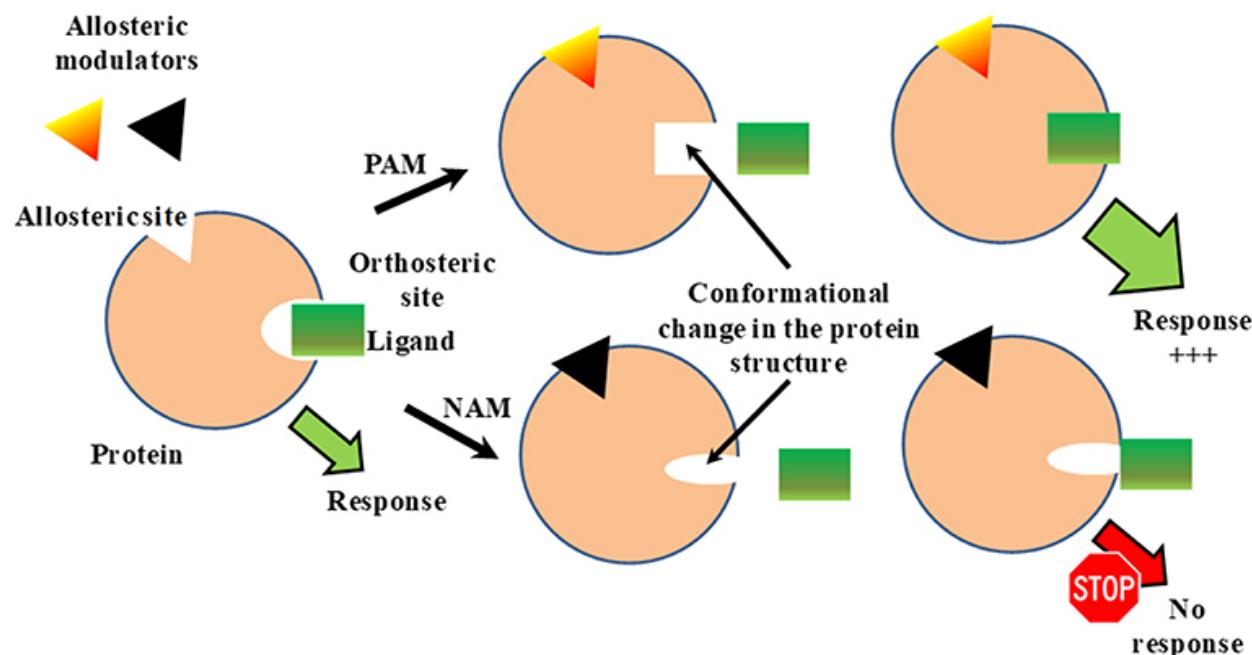
One possible cause of inherited disorders is missense mutations. Missense mutations, defined as a single base pair substitution, cause a different amino acid to be coded into the polypeptide. This is often referred to as an amino acid substitution. This mutation can cause improper folding of the polypeptide into its final conformation in the protein, which can have disastrous effects. Proper protein folding allows the protein to carry out its function and bind with other molecules in the most thermodynamically favorable way.

Missense mutations may or may not affect the conformation of an allosteric binding site. Since proteins are made up of multiple polypeptide chains, a missense mutation may only affect one domain or subunit of the protein. If an allosteric binding site is on a separate subunit, then it likely won’t be affected by the missense mutation.

A missense mutation can also alter the effect of a ligand binding to an allosteric binding site. One example of this is outlined in Hu et. al [20]. In this article, wild-type and mutant receptors were compared in the protein activity following the binding of a compound related to Arey et. al.’s Compound 1 to the Human  $\text{Ca}^{2+}$  Receptor [21]. It was found that the related compound acted “as a negative allosteric modulator on the wild-type receptor but as a positive modulator” on the mutant receptor [20]. In another article studying some of the same conditions and body systems, it was found that allosteric modulation instead rectified signaling abnormalities caused by mutations on the associated GPCRs [22].

## Many drug molecules are allosteric modulators

Allosteric modulators are molecules that bind to a receptor to alter the receptor's response to a stimulus. One study has shown that GPCR modulation has the possibility to overcome abnormalities such as those caused by missense mutations indirectly [22]. In other words, binding to an allosteric site may rectify the effects of a mutation that occurs on a separate location on the molecule.



*Figure 3. Overview of positive and negative allosteric modulators [23].*

Allosteric modulators are also reversible, meaning that they can be added or removed from the target protein to regulate protein activity and subsequent substrate binding [4]. This is especially important in the case of PrP. Since the goal is to find a way to prevent prion aggregation, a reversible modulator allows for other regulatory molecules to bind to the same molecule [4]. Once an allosteric modulator binds to a protein and a shape change occurs, several different options are available to regulate the protein. Additional modification of the protein can occur through covalent modification. A simple modifying group can be added to the protein, aided by the conformation change induced by the allosteric modulator, that could help the protein fold back into the native conformation. The same approach can be used to modify the mutant amino acid into one that doesn't have such a disastrous effect [4].

## Covalent modifications control protein location and interaction

Protein phosphorylation is one of many forms of covalent modification to a protein. Protein phosphorylation involves attaching a phosphate group to one or more side chains on the protein. Since each phosphate group carries two negative charges, the addition of a phosphate group to part of the protein can induce a conformational change. One example is a positively

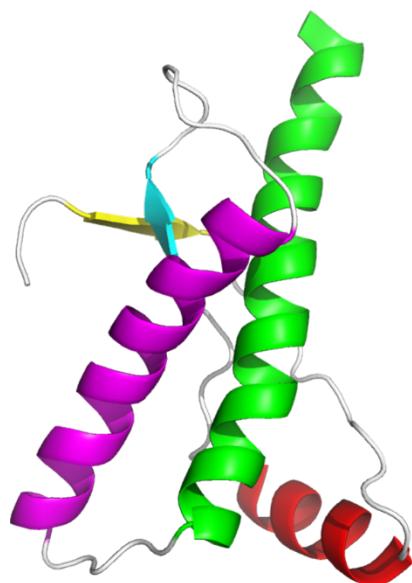
charged amino acid on another area of the protein being attracted to the negatively charged phosphate group, causing the protein to fold on itself to bring these two groups together [24]. Phosphate groups aren't the only groups that can be added to a protein in covalent modification. Attaching ubiquitin, for example, targets a protein for degradation. Through using a reversible covalent modifier, we can control a protein's activity or stability by adding or removing a covalent modifier group [24].

## Pharmacological chaperone proteins can assist in protein folding

Chaperone proteins facilitate correct folding pathways to assist in proteins folding to their native conformation. For proteins that do not fold spontaneously, chaperonins may be required for folding to occur [13]. This idea can be applied to already misfolded proteins by attaching a chaperone protein to a misfolded protein that destabilizes the misfolded conformation and assists in folding the protein back into its native conformation. Other chaperone proteins "protect" proteins and their native conformation [13]. In cases where a prion disease is detected early, a chaperone protein may prove useful in protecting the cellular prion protein from the misfolded protein.

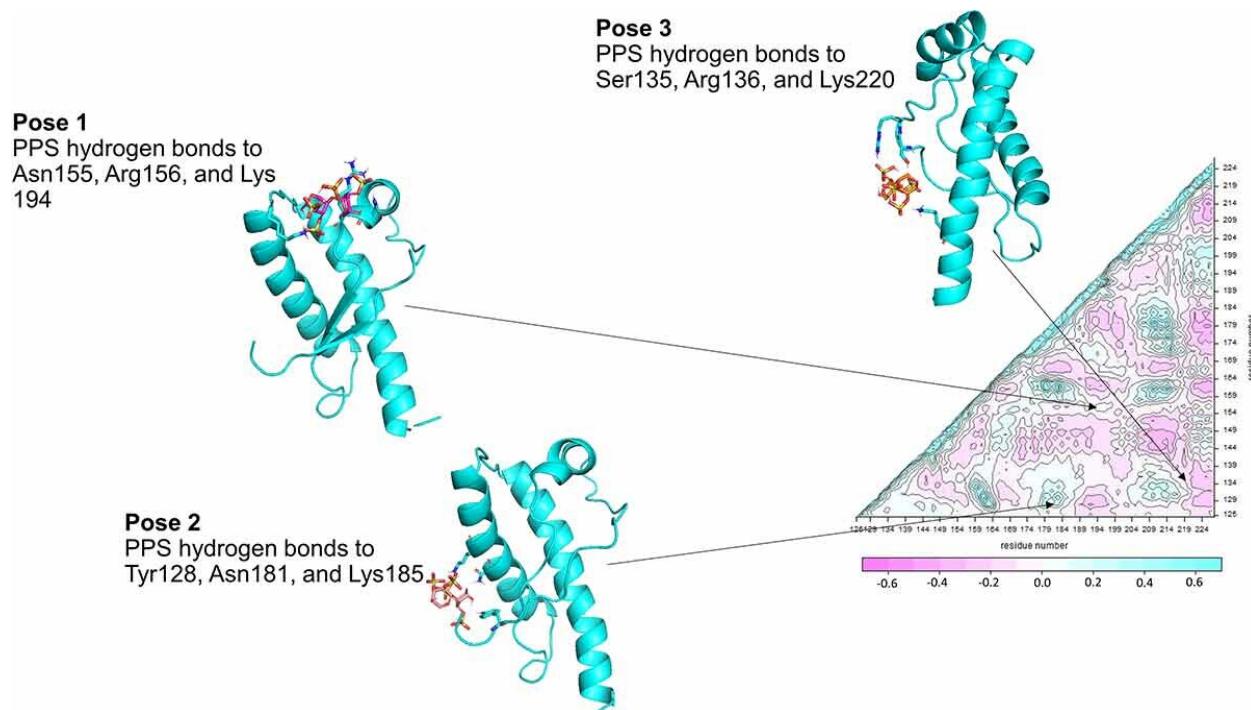
## The structure of PrP<sup>C</sup> provides insight into regulation

PDB files for 1QM2 and 1I4M were obtained from the Protein Databank and explored in PyMOL. The structures from these files were compared to that of hamster prion protein explored by Soto et. al. who identified two distinct subdomains of Prion protein [25]. *Figure 4* shows the structure of 1QM2 color coded by secondary structure.



*Figure 4. Secondary structure classifications in PrP<sup>C</sup>. β-sheet 1 (β1, Y128-G131) is shown in yellow, α-helix 1 (α1, D144-M154) in red, β-sheet 2 (β2, V161-R164) in cyan, α-helix 2 (α2, N173-K194) in magenta, and α-helix 3 (α3, E200-R228) in green.*

Soto et. al. found that the non-conserved loop connecting  $\alpha 2$  and  $\alpha 3$  has the largest conformational flexibility in the C-terminus of the protein [25]. This led to their proposition that residues preceding this  $\alpha 2$ - $\alpha 3$  loop are the site at which misfolding of the protein begins [25]. With this information, Soto et. al. generated three binding poses for the pentosan polysulfate (PPS) molecule on hamster prion protein in which it acted as a pharmacological chaperone [25]. These three binding poses are shown in *Figure 5*. These poses were later referenced when determining the validity of some results.



*Figure 5. Binding poses of PPS on hamster PrP<sup>C</sup> and dynamic cross correlation matrix averaged over the NMR ensemble of hamster PrP<sup>C</sup> [25].*

## Using principles of allostery on non-allosteric proteins

Prion proteins are not allosteric proteins and there is no targetable active site that we are aware of on Human Prion Protein. Allosteric models may still prove useful in applying to non-allosteric proteins, however. The metabolic enzyme DHFR has previously been shown to have a network of residues that link the active site to numerous distance surface sites [26]. This may also be the case with Human Prion Protein. If the location at which a missense mutation is modeled on cellular prion protein is designated as the “active site,” we could potentially identify surface sites that affect the structure and stability of this active site, or the site of mutation in the context of this research.

## Allosteric computational models

Several technologies are now available that can predict the location of allosteric binding sites and results of amino acid substitutions. One possible research approach uses such technologies to map out target binding sites and possibly even reverse the effects of an amino acid substitution. This can prove useful in the field of model-informed drug development [8]. If an allosteric site can be identified and modeled, synthetic drugs can be modeled to fit the target site prior to chemically synthesizing the compound in lab. Computing the effects of amino acid substitutions (as opposed to observing them in lab) reduces the amount of time spent on refining synthesized drug molecules. If it is known what effect certain amino acid substitutions have on the binding affinity of a protein and its ligand, negative effects from some missense mutations may be avoided altogether. An allosteric map can prove to be very useful in model-informed drug development as it allows for more specific targeting of the issue at hand, rather than simply treating symptoms. Further development in complex machine learning algorithms is still needed to have a more detailed understanding of allosteric landscapes [27].

# METHODS

---

All methods described in this paper were done using a computer connected to the internet. The methods described can be done on any operating system and a combination of a Windows desktop and Mac OS laptop were utilized throughout this project. Files were saved to a flash drive for ease of access across all devices before being uploaded to a public Git repository that can be accessed at the link included at the end of this report.

## Accessing PDB files through the Protein Databank

The Protein Databank is a public website that stores all PDB files produced through research. A PDB file includes coordinates of every atom in a protein structure, allowing the structure to be visualized and manipulated by any program with the capability of reading a PDB file. Each structure uploaded to the Protein Databank has a unique PDB-ID. The two proteins used in this project were 1I4M and 1QM2. 1QM2 is an isolated monomer of PrP<sup>C</sup> and had better validation scores on the Protein Databank, and thus was used for most of this research.

## Identifying infectious point mutations to PrP<sup>C</sup>

The Human Gene Mutation Database (HGMD) provides a comprehensive list of known mutations to PRNP (gene that encodes Human Prion Protein) with published material. A summary of all missense and nonsense mutations to PRNP that are classified as pathogenic by

the HGMD were listed in an Excel workbook and used to classify point mutations tested as known pathogenic or unknown.

## Exhaustive analysis of effect of point mutation on protein stability

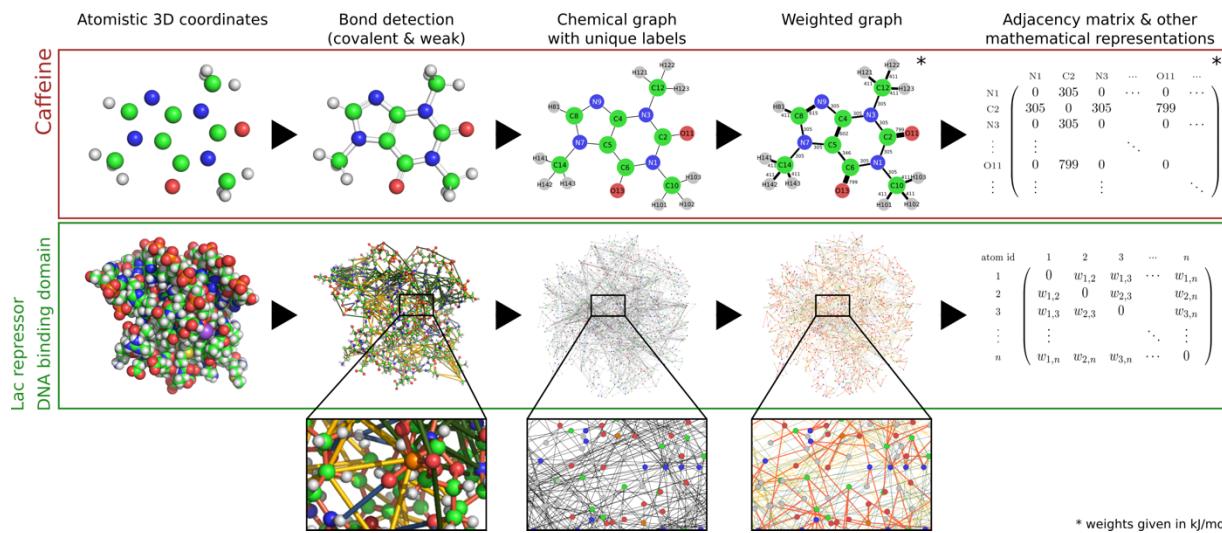
The Eris webserver utilizes the Medusa modeling suite [28] to calculate protein stability changes caused by mutations specified by the user. The result is a  $\Delta\Delta G$  value that indicates whether the mutation induced stabilizes or destabilizes the conformation encoded by the PDB file. Eris was used to analyze the stability change in the protein model from every possible single amino acid substitution. Substitutions involving cysteine were excluded as Eris cannot account for contribution of disulfide bonds introduced [29].

Output using different parameters were tested and no strong correlation between  $\Delta\Delta G$  value and parameters used was found when a correlation analysis was done in R. Final  $\Delta\Delta G$  values used were from applying backbone relaxation and allowing Eris to calculate stability with a flexible backbone method [30].

Using the list of known infectious mutations obtained from the HGMD, two mutations of interest were identified through this exhaustive analysis and were used in the following methods.

## Markov transient analysis as a method of identifying significant sites of regulation

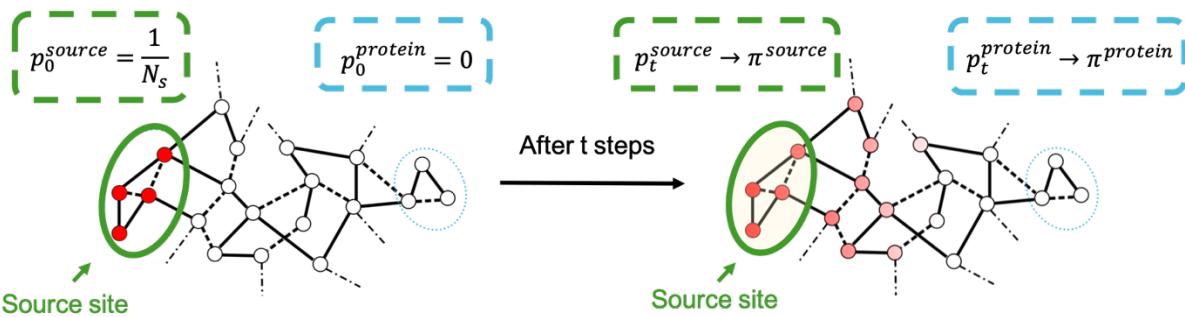
The ProteinLens webserver constructs an atomistic graph of a given biomolecular structure that can be used to explore long-range communication between specific sites [31]. *Figure 4* gives a visual diagram of how this atomistic graph is constructed.



*Figure 6. Schematic overview of graph construction pipeline [32].*

From this atomistic graph, ProteinLens provides two methods of analyzing long-range communication within a protein. This project utilized calculated Markov transient times to identify potential sites of regulation for the mutations of interest. Markov transient analysis is a method that allows one to scan communication across all scales within a biomolecule. In other words, Markov transient analysis allows all residues, regardless of their distance from the source site, to be scored equally. From this analysis, specific pathways that weave throughout the protein can be identified. This methodology is explained in more detail in Amor et. al. [33].

Characteristic transient time  $t_{1/2}$  is a way of measuring how connected two sites are to one another. ProteinLens calculates  $t_{1/2}$  and allows the user to explore various levels of significance and identify sites of interest. In short,  $t_{1/2}$  gives a measure of how many time steps it takes for a signal to travel from one site on the protein to another, regardless of how far away the two sites are from one another. Markov transient analysis takes into consideration the strength of bonds between atoms and robustness of different amino acids, providing a unique measure of connectivity that can be used to identify allosteric and regulatory sites on a protein.



**Figure 7. Modeling the spread of probability across a graph [32].**

When submitting a new job to ProteinLens, one of the mutation sites of interest was designated as the source site and all excess molecules were stripped from the PDB file to analyze only the protein structure. This resulted in two separate jobs, one for each mutation site of interest. The top 3 residues scored through Markov transient analysis for each mutation site of interest were noted and tested using AlloSigMA.

## Testing regulatory site binding effect on stability of mutated protein

With the AlloSigMA webserver, the residues identified through Markov transient analysis can be validated and different simulations can be run to compute the theoretical effect binding on these sites has on the initiation site of misfold on PrP<sup>Sc</sup> and the mutation sites of interest. AlloSigMA calculates the simulated allosteric response to various conditions set by the user (such as binding to a certain site or mutating a certain residue). This response is output as an allosteric free energy difference  $\Delta g_i$  [34]. Residues that show a strong conformational change in response to the condition set will have a positive  $\Delta g$  while residues that show a constraining response will have a negative  $\Delta g$  [34].

## Verifying residues from Markov transient analysis with an allosteric signaling map

An allosteric signaling map was built using the PDB file for 1QM2. From this, individual mutations that represent the mutations of interest were simulated. Mutations that add bulk to the structure were designated as “UP-mutations” and mutations that remove bulk from the structure were designated as “DOWN-mutations.” If the residues noted from the Markov transient analysis are in fact strongly connected to these mutation sites, then a strong change in allosteric free energy (positive or negative) should be observed. The free energy change at residue 129 and theorized initiation site of misfold were also noted [25].

## Simulating binding of a small molecule to proposed regulatory sites

To test the potential regulatory sites identified from Markov transient analysis and verified with the allosteric signaling map, AlloSigMA was used to simulate the binding of a small 3-residue probe to the residues of interest and adjacent residues. The change in free energy at the mutation residues of interest and theorized initiation site of misfold were noted. Residues that showed a constraining response when probed were concluded to be potential regulatory sites for pharmacological chaperone binding as binding to these residues would introduce more rigidity to the normal form of PrP.

## Visualizing protein and proposed regulatory sites

PyMOL run in an Anaconda environment was utilized to produce figures of PrP that are shown throughout this report. Any protein structures that are not from outside sources were produced using PyMOL. All project files and related script to reproduce these figures are available in the public Git repository linked at the end of this report.

## Organization and data analysis

R code was utilized to identify any outliers or unique trends in the data. R was also used to run correlation analysis and verify the completeness of the data collected. All R scripts were produced using RStudio run in an Anaconda environment.

Microsoft Excel was used to organize data extracted from web servers and run quick analysis before exporting files as .csv and analyzing them in R.

Python code was utilized to produce small programs to track hours worked on this project, make quick notes during data collection, and extract data from files that cannot be read by RStudio or Microsoft Excel. All Python scripts were produced using Spyder run in an Anaconda environment.

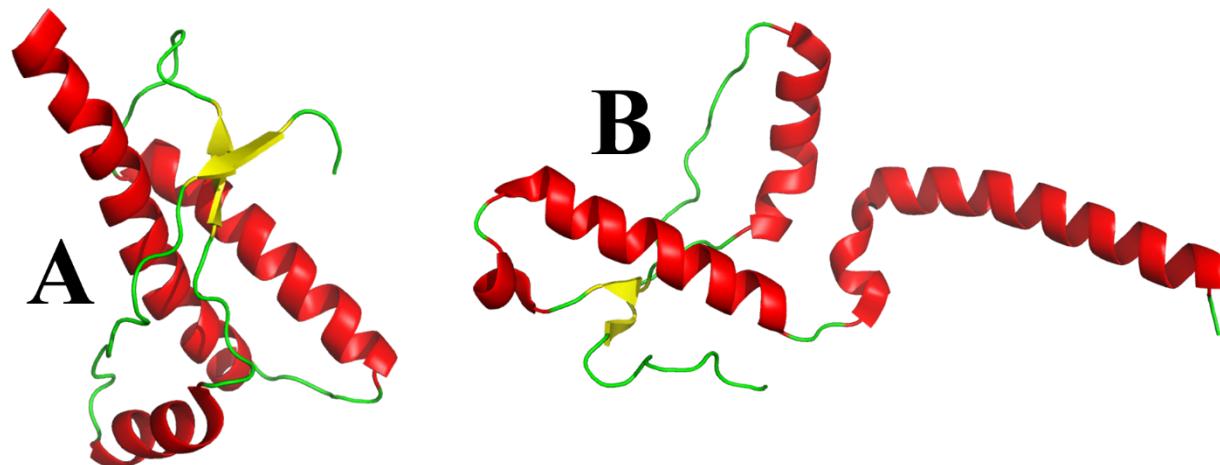
All files produced for data analysis purposes are available in the public GitHub repository linked at the end of this report.

# DATA AND RESULTS

---

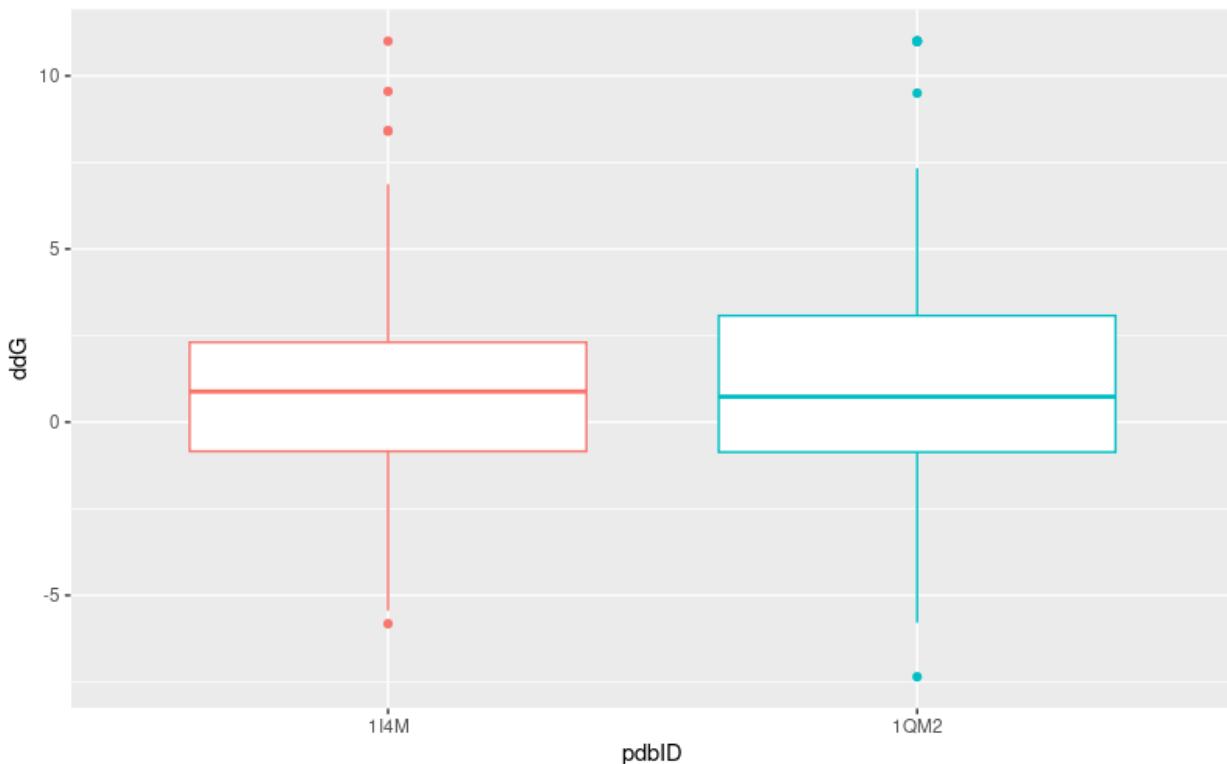
## V189I and V203I identified as mutations of interest

The PDB file for 1QM2 contained an isolated monomer of PrP<sup>C</sup> while 1I4M contains two PrP<sup>C</sup> monomers joined by  $\alpha$ 3 on one monomer wrapping around  $\alpha$ 3 on the other. The difference between these two files is shown in *Figure 8*.



**Figure 8. Comparison of two PrP<sup>C</sup> structures.** (A) Structure of 1QM2, derived from solution NMR [11]; (B) Structure of 1I4M derived from x-ray diffraction [35]; 1I4M displays a unique fold that was found to make up a human prion protein dimer, the length of each tertiary structure is different from that of a single human prion protein monomer shown in (A).

Since 1QM2 is an isolated monomer of PrP<sup>C</sup> and provides the proposed “ideal” fold for PrP<sup>C</sup>, it is the more ideal structure to perform analysis on. A Multiple Sequence Analysis using NCBI BLAST [36] was run on each structure, and it was also found that 1QM2 provides a structure that is both more complete and more similar to other prion protein structures found. 1I4M was previously exhaustively analyzed with Eris by Redler et. al. [29], so to support the usage of 1QM2 throughout the majority of this research, both structures were exhaustively analyzed with Eris and the distribution of  $\Delta\Delta G$  for each was compared with R. Only  $\Delta\Delta G$  values for the known infectious missense and nonsense mutations from the HGMD (see *Supplementary Table 1*) were computed for this initial analysis. *Figure 9* shows the distribution of  $\Delta\Delta G$  values for each structure when one mutation is induced. Once again, any mutations involving cysteine residues were omitted due to Eris’s restrictions.



**Figure 9.  $\Delta\Delta G$  distributions for infectious mutations applied to each of 2 PDB files (1QM2 and 1I4M).**

Using a single-factor ANOVA, it was found that  $\Delta\Delta G$  was not significantly affected by the structure used. This validates the usage of 1QM2 in computational analysis. All data collection done from this point forward was done with 1QM2 unless noted otherwise.

There are 2 main parameters to be set in Eris: backbone pre-relaxation and backbone flexibility. The methods for which these two parameters are set are outlined in Yin et. al. along with experimental report on the effects of these two parameters [30]. It was found that keeping the protein's backbone fixed versus allowing it to move and rotate freely had no significant effect on the  $\Delta\Delta G$  distribution for the proteins explored, but a flexible backbone method has the potential to resolve atomic clashes. Yin et. al. also introduced a pre-relaxation step in the Eris protocol that optimizes the protein's backbone structure to have the least strain throughout the whole protein [30]. This significantly improved the side-chain packing of the protein core, resulting in more accurate  $\Delta\Delta G$  evaluations by Eris.

For each infectious mutation outlined in *Supplementary Table 1*, 4 jobs were submitted using 1QM2 as the structure and varying in backbone flexibility and relaxation parameters. Each set of parameters was given a specific trial code for analysis purposes. Trial codes and their associated parameters are outlined in *Table 1* and the output values are shown in *Supplementary Table 2*.

**Table 1.** Table of parameters outlining settings used for each Eris suite trial code in Supplementary Table 2.

Trial Code	Parameters
Trial 1	Fixed backbone, no backbone pre-relaxation
Trial 2	Fixed backbone, pre-relaxation on
Trial 3	Flexible backbone, no backbone pre-relaxation
Trial 4	Flexible backbone, pre-relaxation on

Since backbone pre-relaxation and allowing the backbone to be flexible both increase the time it takes for a job to run, analyzing the  $\Delta\Delta G$  distributions for each trial parameter provides insight into whether it is beneficial to lengthen the run time for each job. Using a single-factor ANOVA, it was found that  $\Delta\Delta G$  was not significantly affected by the parameters set when submitting the job. This meant that the exhaustive analysis of 1QM2 could be performed with a fixed backbone and no backbone pre-relaxation, shortening the run time for each job significantly.

All R-script files and .csv files housing data obtained from Eris suite can be accessed through the public Git repository linked at the end of this report.

From the exhaustive analysis of change in protein stability due to point mutations, it was found that V189I and V203I both had relatively high positive  $\Delta\Delta G$  values and a relatively small standard deviation amongst all 4 trials (*Supplementary Table 2*). V189I and V203I both cause the same type of sporadic Creutzfeldt-Jakob Disease (sCJD), the MM/MV 1 variant [37] [38]. One important characteristic of this variant is that in addition to the presence of one mutation (such as V189I or V203I), it can be heterozygous for the polymorphic M129V mutation [18]. M129V was also observed throughout this project, especially in conjunction with other mutations.

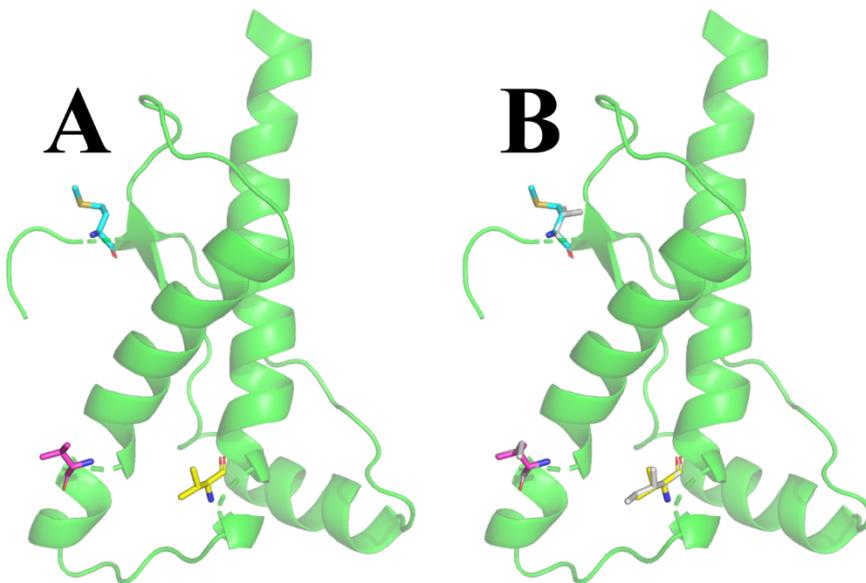
In contrast, a mutation related to V203I, V203G, has been identified, but the individual was homozygous for valine at codon 129 [39]. This mutation had a much higher average  $\Delta\Delta G$  than any of the other mutations previously mentioned (*Supplementary Table 2*), but with the lack of a definitive diagnosis for the individual found with the mutation, the effects from this mutation were merely observed and further analysis was not performed. By referencing current literature on the classification of CJD, it was determined that the individual discovered with the V203G mutation may have had a rarer VV2 sCJD variant, although this could not be supported by an autopsy to show the structure of PrP<sup>Sc</sup> in the brain of the individual [2] [18].

The stability of the protein when one of the two mutations was present alongside the M129V mutation was predicted using Eris and compared with the results previously obtained. *Table 2* shows an increase in  $\Delta\Delta G$  when the M129V mutation was also present, alluding to this mutation further destabilizing the protein with an existing mutation. Residues of interest are visualized in *Figure 10*.

**Table 2.**  $\Delta\Delta G$  values for two mutations of interest with and without M129V polymorphism using Trial 4 parameters (Supplementary Table 3).

Mutation	$\Delta\Delta G$ Values	
	1QM2 Trial 4	Trial 4 with M129V
V189I	2.89	4.87
V203I	1.58	3.06

Table 2 shows that with both one of the mutations of interest and the M129V polymorphism present, the protein is predicted by Eris to be further destabilized, indicated by a more positive  $\Delta\Delta G$  when M129V is present.

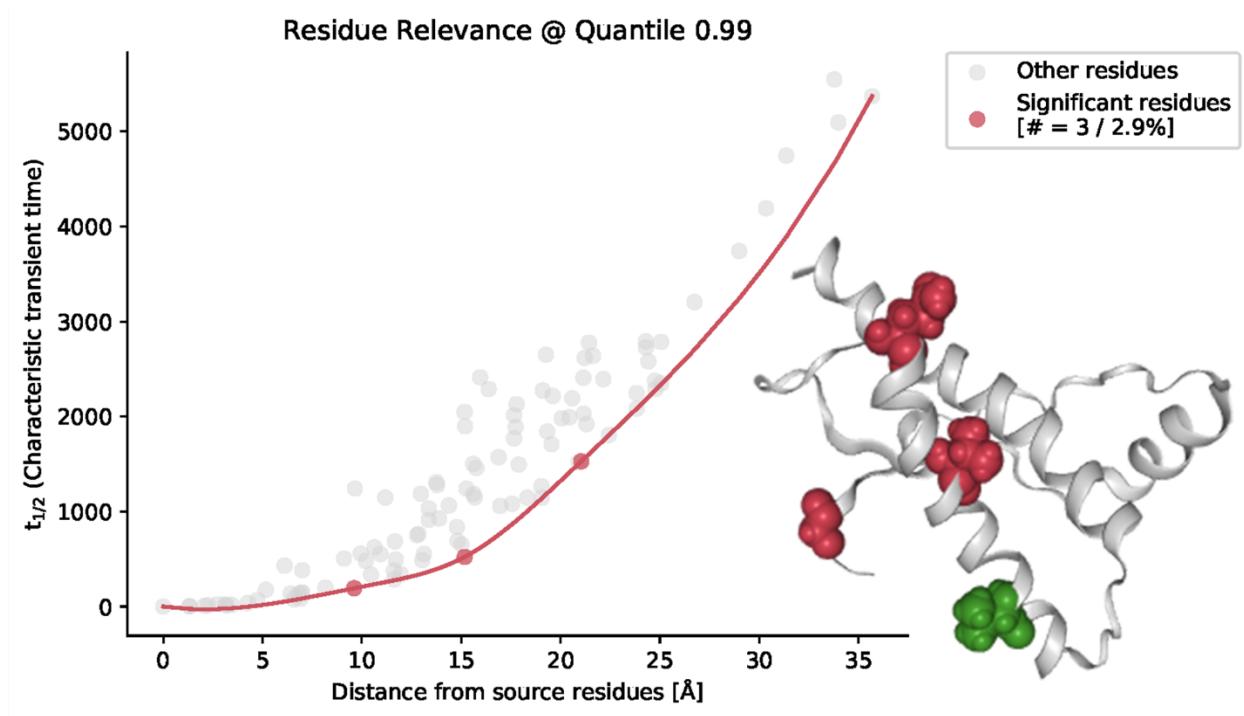


**Figure 10.** Structure of 1QM2 with residues of interest shown in sticks structure. (A) and (B) show M129 in cyan, V189 in magenta, and V203 in yellow. (B) overlays the mutated residue (valine for residue 129, isoleucine for residues 189 and 203) in white on top of the native residue representation.

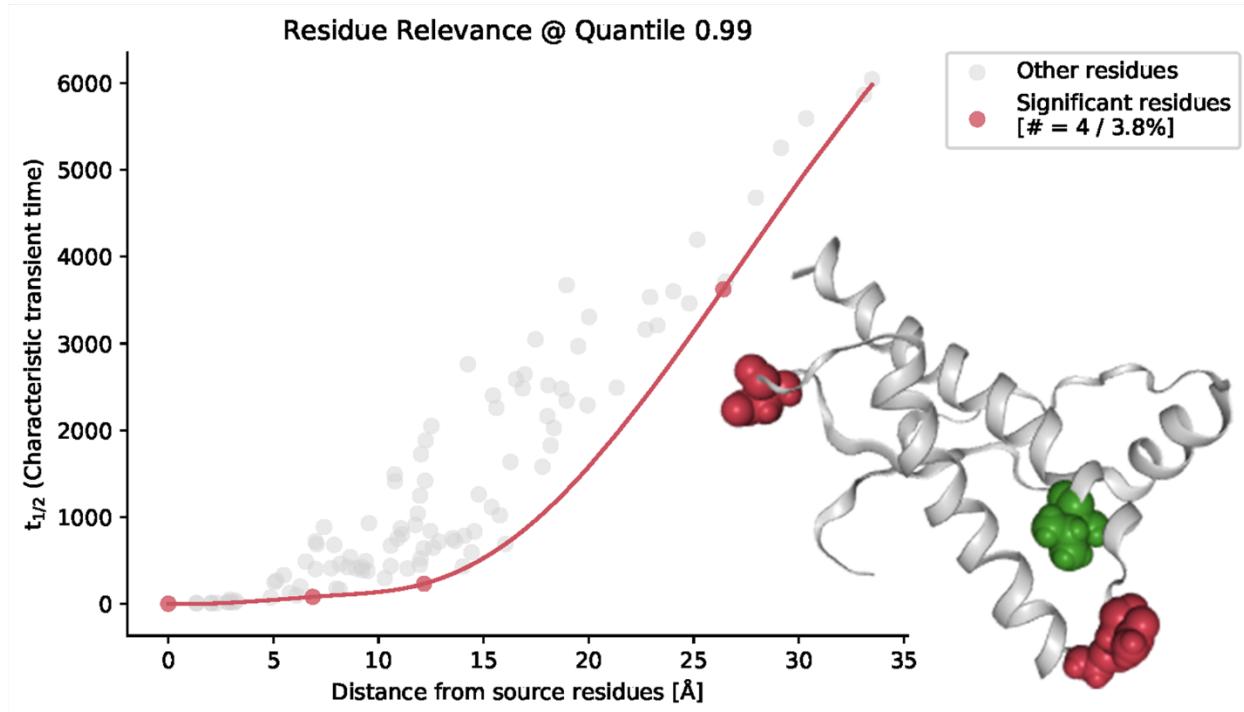
Using the secondary structure classifications from Figure 4, it is known that residue 129 is located on  $\beta$ 1, residue 189 is located on  $\alpha$ 2 before the non-conserved loop and residue 203 is located on  $\alpha$ 3 just after the non-conserved loop. Knowing these residues are close to the proposed site of misfold initiation [25], it would make sense that we would observe destabilization of the native structure, as indicated by a positive  $\Delta\Delta G$  value.

### Mutations of interest may be regulated by residues in the same region of PrP<sup>C</sup>

Two ProteinLens sessions were created with either residue 189 or 203 designated as the source residue. Markov transient plots (Figures 11 & 12) were generated and used to identify residues with strong predicted connectivity to the source residue.



**Figure 11.** Markov transient plot used to identify residues with strong connectivity to residue 189. The graph shown visualizes the  $t_{1/2}$  for every residue on 1QM2 in relation to residue 189. The red fit line follows the 99<sup>th</sup> quantile of residues. The protein structure to the right shows the source residue (V189) in green and the three most relevant residues according to the Markov transient plot in red. These three relevant residues are the three residues that fall on the 99<sup>th</sup> quantile fit line.



**Figure 12.** Markov transient plot used to identify residues with strong connectivity to residue 203. The graph shown visualizes the  $t_{1/2}$  for every residue on 1QM2 in relation to residue 203. The red fit line follows the 99th quantile of

residues. The protein structure to the right shows the source residue (V203) in green and the three most relevant residues according to the Markov transient plot in red. These three relevant residues are the three residues that fall on the 99th quantile fit line.

Residues with low characteristic transient time ( $t_{1/2}$ ) relative to its distance to the source site can be considered strongly connected to the source site. A low  $t_{1/2}$  means fewer time steps for a signal to travel from the source site to the residue being analyzed. At the 99% quantile, three residues are strongly connected to the source site for each site of interest (189 or 203). For residue 189, Valine180, Glycine127, and Glutamine172 were identified as potential regulatory sites. Asparagine197, Glycine195, and Aspartic Acid167 were identified as potential regulatory sites for residue 203. The distance from the source site and  $t_{1/2}$  for each of these sites are shown in *Table 3*.

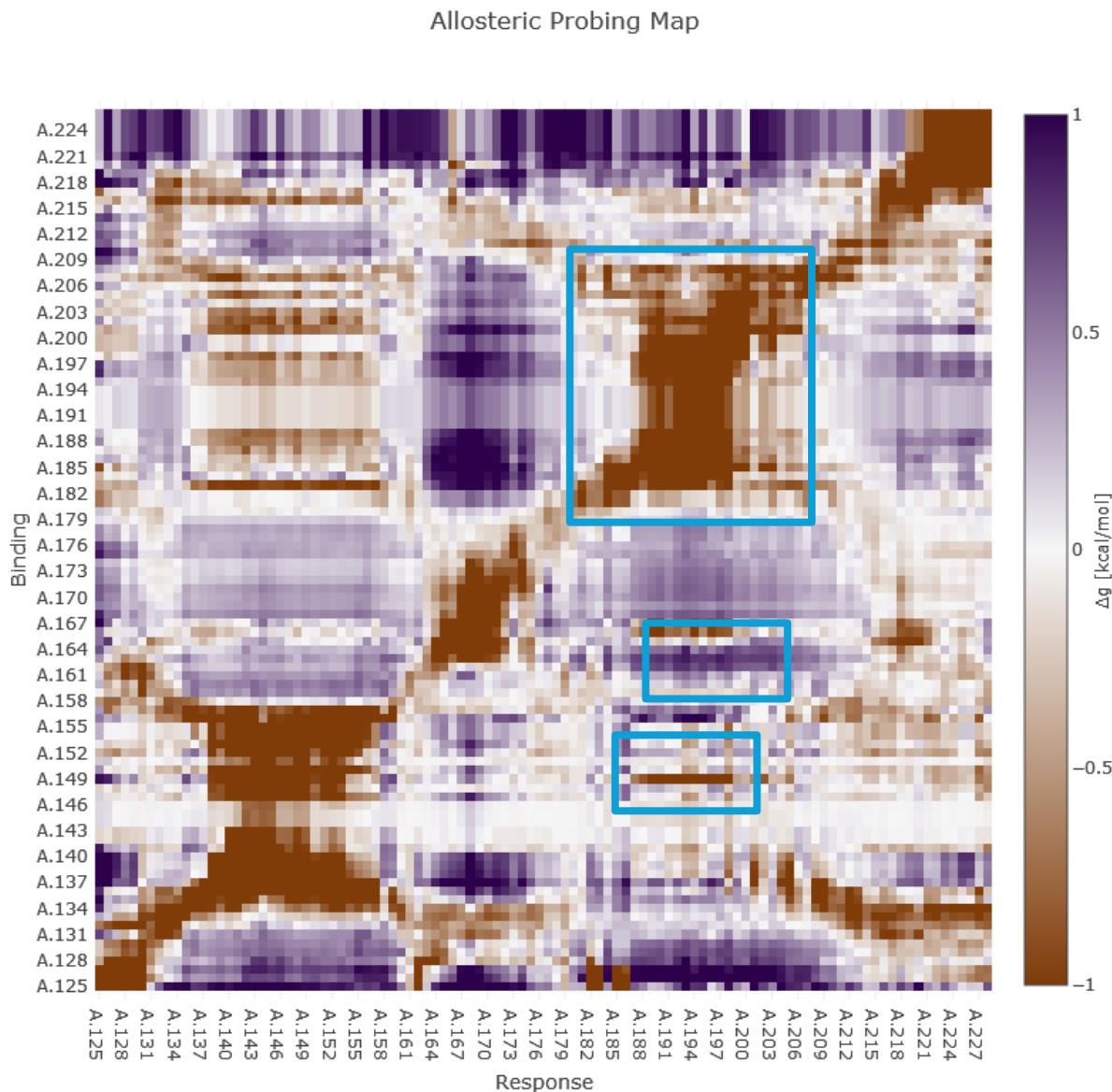
**Table 3.** Potential regulatory sites for mutation sites of interest on PrPC identified with ProteinLens.

Source (Mutation) Site	Potential Regulatory Site	Characteristic Transient Time ( $t_{1/2}$ )	Distance from Source Site (Å)
V189	V180	193.75	9.61
	G127	523.29	15.18
	Q172	1529.12	21.05
V203	N197	79.86	6.87
	G195	230.86	12.17
	D167	3623.75	26.41

G195 and N197 are part of the  $\alpha$ 2- $\alpha$ 3 loop, so introducing rigidity to these residues by binding or inducing a mutation that adds bulk to the protein may prove beneficial in stabilizing the native conformation with V203I mutation present. Previous molecular dynamics simulations have suggested the G127V mutation prevents stable  $\beta$ -sheets and dimers from forming, preventing disease propagation [40]. This mutation adds bulk to the structure, so binding to this residue may have a similar effect as the G127V mutation. With this in mind, simulated binding analysis was performed on these residues and the polymorphic residue 129 to observe their predicted effect on the stability of the mutation sites of interest and the  $\alpha$ 2- $\alpha$ 3 loop.

## Simulated probe binding reveals potential regulatory sites

AlloSigMA can predict the effects binding a small 3-residue probe to one site on a protein has on the stability of the rest of the protein. *Figure 13* shows the effect binding a small 3-residue probe on one residue has on a responding residue. A negative  $\Delta g$  (allosteric free energy change) indicates a constraining response on the responding residue, which may indicate stabilizing of the native structure in that region.



**Figure 13.** Allosteric probing map of 1QM2 calculated using AlloSigMA. The boxed areas show regions of residues that restrict conformational changes from occurring on the responding residues when a small 3-residue probe is bound to a binding residue.

From Figure 13, some interesting responses were noted. As expected, binding to a residue in or around the region where misfolding is predicted to begin (residues 195-199) shows a constraining response on the residues in this region. Interestingly, binding to residues 166 or 167 also had a constraining response on the non-conserved  $\alpha$ 2- $\alpha$ 3 loop. Residue 149 also seemed to have a constraining response in this region when bound. Residue 149 was not previously noted as being strongly connected to the mutation sites of interest, but it may have strong connectivity to the site of misfold initiation. Table 4 gives a summary of the allosteric free

energy change on the mutation sites of interest and the misfold initiation site resulting from binding to a single residue of interest.

**Table 4. Allosteric free energy change on sites of interest resulting from 3-residue probe binding on potential regulatory sites.**

<b>Responding Residue</b>	<b>Allosteric free energy change (<math>\Delta g</math>) from probing at site:</b>								
	<b>127</b>	<b>129</b>	<b>149</b>	<b>166</b>	<b>167</b>	<b>172</b>	<b>180</b>	<b>195</b>	<b>197</b>
<b>189</b>	1.159	0.664	-1.292	-1.024	-1.045	0.509	-0.039	-1.337	-1.855
<b>195</b>	1.172	0.724	-1.626	-0.589	-0.881	0.550	0.249	-2.485	-3.216
<b>196</b>	1.281	0.648	-1.504	-0.974	-0.961	0.531	0.015	-1.781	-2.463
<b>197</b>	1.364	0.744	-1.762	-0.868	-0.910	0.613	-0.162	-2.395	-2.835
<b>198</b>	0.921	0.722	-1.628	-0.827	-0.864	0.530	0.167	-2.056	-2.527
<b>199</b>	1.188	0.593	-0.454	-0.307	-0.043	0.410	-0.042	-0.275	-0.887
<b>203</b>	1.118	0.691	-0.308	-0.372	-0.105	0.431	-0.234	-0.560	-0.355

Residues 149, 166, 167, 195, and 197 all had a constraining effect on residues of interest. The residues of interest tested were the two mutation sites of interest (189 and 203) and the residues that make up the  $\alpha$ 2- $\alpha$ 3 loop. Residue 172 was identified through Markov transient analysis as being a potential regulatory site for residue 189. A positive value for  $\Delta g$  indicates enhanced conformational flexibility on the responding residue when a small probe is bound to the probing site. *Table 4* shows that binding to residue 172 results in a positive allosteric free energy response on residue 189 and all other residues tested, but not as strong of a positive response as binding to residue 127. Residue 127 was also identified as a potential regulatory site of residue 189. The positive response seen in *Table 4* for these residues indicates that binding to these residues may induce a conformational change rather than restricting the native conformation. If the results from ProteinLens are correct in classifying residue 172 and 127 as being highly connected to residue 189, then one could predict binding to these residues would lead to a drastic conformational change at and around residue 189. This contradicts my hypothesis that binding to residue 127 stabilizes the native conformation. Knowing prion protein folds under kinetic control, it may be that the introduction of the G127V mutation simply makes it more difficult for PrP to fold into the  $\beta$ -sheet rich prion conformation, rather than stabilizing the native conformation if PrP were to fold under thermodynamic control.

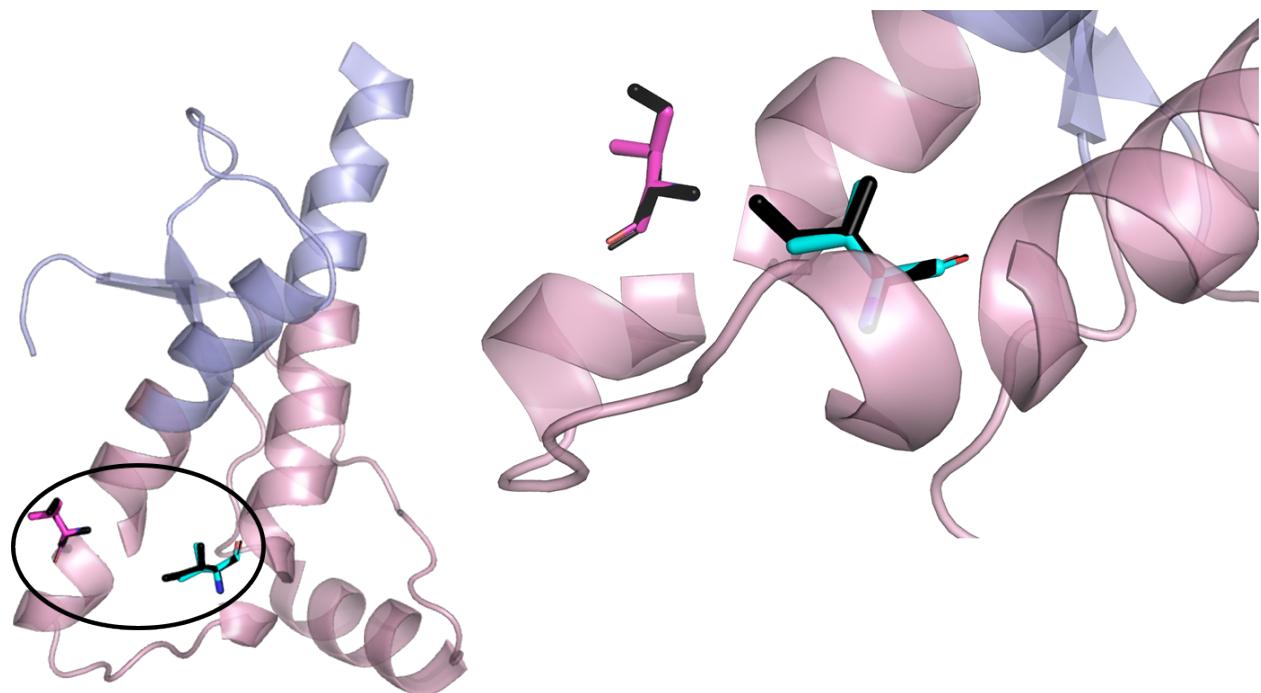
## DISCUSSION

If the effects of mutations and subsequent small-molecule binding can be predicted and simulated using digital software, in-lab experimentation can be reduced and more specific. This project aimed at providing proof of concept that allosteric models can be applied to non-allosteric proteins for the purpose of exploring potential regulatory sites for drug molecules.

The results from the above computational analysis methods have mirrored that of what we already know about prion protein, with few exceptions. Since these analysis methods are largely based on allosteric proteins that assume thermodynamic control, it may be that prion protein's kinetic folding is not accurately simulated with these methods. These methods may still have some validity, but in-lab testing, and further molecular dynamics simulations need to be performed for any conclusion to be made on how valid these methods are for use on non-allosteric proteins that fold under kinetic control.

## Proposed binding sites for pharmacological chaperones

One study concluded that prion protein can be divided into two distinct subdomains, separated by a disulfide bridge from C179 on  $\alpha 2$  to C214 on  $\alpha 3$  [25]. These two subdomains are visualized in *Figure 14* and the two mutations of interest are noted.



**Figure 14. Mutations of interest on 1QM2 showing 2 distinct subdomains. Residue V189 in magenta, V203 in cyan; black sticks structure shows conformational change to isoleucine from mutation.**

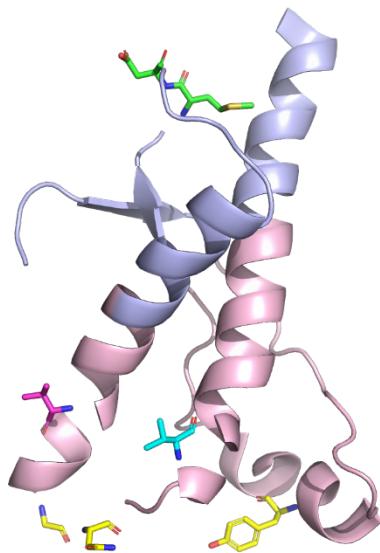
Both mutations of interest are located on the lower subdomain that includes the  $\alpha 2$ - $\alpha 3$  loop. Many infectious mutations listed in *Supplementary Table 1* are also part of this subdomain, providing insight into the region of prion protein that is the most susceptible to misfolding.

The same study that identified these two subdomains also found 3 PPS binding poses (*Figure 5*) that connect the two subdomains as a method of stabilizing the native conformation [25]. Binding to a site on each subdomain was able to assist the protein in folding to the native conformation, but all three binding poses were located on the C-terminus of PrP. The question

still remains as to how the N-terminus of PrP affects binding site availability on the C-terminus [25].

Using these binding poses as a guide, I predict binding to one of residues 166 or 167 and binding to one of residues 149, 195, or 197 would have the greatest stabilizing effect on the native conformation of human prion protein. Future work is needed to test these binding sites and synthesize an appropriate chaperone protein that can bind to these sites effectively.

The proposed sites for pharmacological chaperone binding are shown in *Figure 15*. Site A, located on the upper subdomain seen in *Figure 14*, is shown in green and includes residues 166 and 167. Site B, located on the lower subdomain consisting of the  $\alpha$ 2- $\alpha$ 3 loop, is shown in yellow and includes residues 149, 195, and 197. The two mutation sites of interest, V189 and V203, are shown in magenta and cyan respectively.



*Figure 15. Proposed binding sites for pharmacological chaperones on IQM2.*

## Moving Forward

The methods outlined in this report are not exclusive to human prion protein. Methods such as these can be applied to any protein and may prove useful in identifying potential regulatory sites on proteins that cause other diseases. Alzheimer's disease and Parkinson's both have similar methods of pathogenesis to prion diseases [41]. If the methods outlined in this report are also applied to the proteins responsible for the progression of these diseases, regulatory sites could be identified that weren't previously known, allowing for new treatments to be developed.

## Replicating results with other mammalian prion protein structures

This report outlines the usage of two human prion protein structures in computational analysis. To verify the results seen, not only should in-lab experimentation be conducted, but computational analysis of other mammalian prion protein structures should be performed. Hamster prion protein is one of utmost importance because of its application in drug testing. It may also prove beneficial to analyze bovine, ovine, and cervid prion protein structures because of the prion disease transmission risk between these species and humans [42] [43] [44].

## Identifying drug molecule characteristics through docking simulations

Docking simulations allow researchers to explore how small molecules may bind to a target protein. If the binding sites outlined in *Figure 15* are used as sites for docking analysis, libraries of small molecules can be tested against the surface structure of human prion protein. The best docking poses can be explored to help identify what atoms within the binding sites are the most crucial for strong binding to the protein to occur. This information can serve as a guide when synthesizing potential drug molecules. These potential drug molecules can then be tested in lab to see if it acts as a pharmacological chaperone or “protects” PrP<sup>C</sup> from prions.

## Simulating and analyzing prion-resistant mutations

Current methods of treatment for prion diseases target the symptoms of the disease and not the root cause, largely because of the lack of research and conclusive answers as to what purpose prion protein serves and how prion pathogenesis works [17]. One promising discovery, however, is of a completely prion disease-resistant genotype, V127M129 [40]. Molecular dynamics simulations suggested that the G127V mutation prevents stable β-sheets from forming [40].

Replicating these molecular dynamic simulations and conducting further analysis may uncover why this mutation prevents stable β-sheets from forming in the first place. We could then attempt to replicate this same effect in other ways – such as inducing other mutations or binding to certain residues on PrP<sup>C</sup>.

# ACKNOWLEDGEMENTS

I would like to thank my research and academic advisor, Dr. Joshua Ring, for his continued guidance and support throughout this project. I would also like to thank the personal investigators and peer mentors of the summer 2023 MIV-REU program for their expertise in the field of computational sciences and insight into the current techniques used for docking analysis. Finally, I would like to thank the Appalachian College Association and their Ledford Scholars Program for funding this first stage of my research.

# RESOURCES

---

## RESOURCES FOR COMPUTATIONAL METHODS

**Protein Databank (PDB):** The PDB can be accessed through the following link: <https://www.rcsb.org>. PDB files can be accessed using their unique PDB-ID and downloading the desired files from the structure page.

**Human Gene Mutation Database (HGMD):** The HGMD can be accessed through the following link: <https://www.hgmd.cf.ac.uk/ac/index.php>. Once logged in, a user can search for a specific gene (such as PRNP) through the search bar, which will open the associated webpage for that gene. From this webpage, the user can access detailed information about documented mutations on that gene.

### Documented Webservers

Eris Suite: <https://dokhlab.med.psu.edu/eris/login.php>

ProteinLens: <https://www.proteinlens.io/webserver/index>

AlloSigMA: <http://allosigma.bii.a-star.edu.sg/home/>

### Data Analysis and Visualization

**Anaconda Navigator:** <https://www.anaconda.com/anaconda-navigator>

RStudio and Spyder were installed in the main Anaconda environment used throughout this project. To install and use these programs independently, use the links below.

**RStudio:** <https://posit.co/download/rstudio-desktop/>

**Spyder:** <https://www.spyder-ide.org>

PyMOL can be installed as an independent program at the link below. For this project, PyMOL was installed using the ‘schrodinger’ Anaconda channel onto a new environment separate from the analysis environment.

**PyMOL:** <https://pymol.org>

More information on installing and using programs built through Anaconda Navigator can be found at the following link: <https://docs.anaconda.com>.

Microsoft Excel was used extensively to view, organize, and process data prior to analysis. Microsoft Excel can be accessed and installed along with the full Microsoft Office suite at the following link: <https://www.microsoft.com/en-us/microsoft-365>

## GitHub

Data and script files used throughout this research can be found at the following link:  
<https://github.com/oREAGanoWombat/computational-analysis-of-human-prion-protein>

## REFERENCES AND SUGGESTED READINGS

- [1] S. Sakaguchi, "Prion pathogenesis revealed in a series of the special issues "Prions and Prion Diseases"," *International Journal of Molecular Sciences*, vol. 23, p. 6490, 2022.
- [2] S. Baiardi, M. Rossi, A. Mammana, B. S. Appleby, M. A. Barria, I. Cali, P. Gambetti, E. Gelpi, A. Giese, B. Ghetti, J. Herms, A. Ladogana, J. Mikol, S. Pal, D. L. Ritchie, V. Ruf, O. Windl and S. Capellari, "Phenotypic diversity of genetic Creutzfeldt-Jakob disease: a histo-molecular based classification," *Acta Neuropathologica*, vol. 142, pp. 707-728, 2021.
- [3] J.-K. Choi, I. Cali, K. Surewicz, Q. Kong, P. Gambetti and W. K. Surewicz, "Amyloid fibrils from the N-terminal prion protein fragment are infectious," *PNAS*, vol. 113, no. 48, pp. 13851-13856, 29 November 2016.
- [4] D. L. Nelson, M. M. Cox and A. A. Hoskins, "Regulatory Enzymes," in *Lehninger Principles of Biochemistry*, 8 ed., A. Thorne, C. Korsvik and K. Hughes, Eds., New York, NY: Macmillan Learning, 2021, pp. 213-223.
- [5] D. Grasso, S. Galderisi, A. Santucci and A. Bernini, "Pharmacological chaperones and protein conformational diseases: Approaches of computational structural biology," *International Journal of Molecular Sciences*, vol. 42, no. 6, p. 5819, 2023.
- [6] R. Sánchez-Valle, "Prion diseases," *European Neuropsychopharmacology*, vol. 55, pp. 1-3, 2022.
- [7] E. Lauwers, G. Lalli, S. Brandner, J. Collinge, V. Compernolle, C. Duyckaerts, G. Edgren, S. Haïk, J. Hardy, A. Helmy, A. J. Ivinson, Z. Jaunmuktane, M. Jucker, R. Knight, R. Lemmens, L.-C. Lin, S. Love, S. Mead, V. H. Perry and B. De Strooper, "Potential human transmission of amyloid  $\beta$  pathology: surveillance and risks," *Neurology*, vol. 19, no. 10, pp. 872-878, 2020.
- [8] A. Mitra and Y. Wang, "Applications of model informed drug development (MIDD) in drug development lifecycle and regulatory review," *Pharmaceutical Research*, vol. 39, no. 8, pp. 1663-1667, 2022.
- [9] P. Parchi, L. de Boni, D. Saverioni, M. L. Cohen, I. Ferrer, P. Gambetti, E. Gelpi, G. Giaccone, J.-J. Hauw, R. Höftberger, J. W. Ironside, C. Jansen, G. G. Kovacs, A. Rozemuller, D. Seilhean, F. Taglianini, A. Giese and H. A. Kretzschmar, "Concensus classification of human prion disease histotypes allows reliable identification of molecular

- subtypes: an inter-rater study among surveillance centres in Europe and USA," *Acta Neuropathol.*, vol. 124, pp. 517-529, 2012.
- [10] J. Collinge, J. Whitfield, E. McKintosh and J. Beck, "Kuru in the 21st century - an acquired human prion disease with very long incubation periods," *The Lancet*, vol. 367, no. 9528, pp. 2068-2074, 2006.
- [11] R. Zahn, A. Liu, T. Luhrs and R. Riek, "NMR solution structure of the human prion protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 1, pp. 145-150, 1999.
- [12] L. Q. Wang, K. Zhao, H. Y. Yuan, Q. Wang, Z. Guan, J. Tao, X. N. Li, Y. Sun, C. W. Yi, J. Chen, D. Li, D. Zhang, P. Yin, C. Liu and Y. Liang, "Cryo-EM structure of an amyloid fibril formed by full-length human prion protein," *Nat Struct Mol Bio*, vol. 27, pp. 598-602, 2020.
- [13] D. L. Nelson, M. M. Cox and A. A. Hoskins, "Protein Denaturation and Folding," in *Lehninger Principles of Biochemistry*, 8 ed., A. Thorne, C. Korsvik and K. Hughes, Eds., New York, NY: Macmillan Learning, 2021, pp. 128-136.
- [14] J. M. Karty, "Kinetic Control, Thermodynamic Control, and Mad Cow Disease," in *Organic Chemistry Principles and Mechanisms*, New York, W. W. Norton & Company, Inc., 2014, p. 598.
- [15] I. V. Baskakov, G. Legname, S. B. Prusiner and F. E. Cohen, "Folding of prion protein to its native  $\alpha$ -helical conformation is under kinetic control," *Journal of Biological Chemistry*, vol. 276, no. 23, pp. 19687-19690, 2001.
- [16] J. M. Karty, "Kinetic Control, Thermodynamic Control, and Reversibility," in *Organic Chemistry Principles and Mechanisms*, New York, W. W. Norton & Company, Inc., 2021, pp. 508-511.
- [17] M. Wulf, A. Senatore and A. Aguzzi, "The biological function of the cellular prion protein: An update," *BMC Biology*, vol. 15, no. 1, p. 34, 2017.
- [18] Y. Iwasaki, "Creutzfeldt-Jakob disease," *Neuropathology*, vol. 27, pp. 174-188, 2017.
- [19] J. Collinge and M. Rossor, "A new variant of prion disease," *The Lancet*, vol. 347, no. 9006, pp. 916-917, 1996.
- [20] J. Hu, J. Jian, S. Constanzi, C. Thomas, W. Yang, J. H. M. Feyen, K. A. Jacobson and A. M. Spiegel, "A missense mutation in the seven-transmembrane domain of the human

Ca<sup>2+</sup> receptor converts a negative allosteric modulator into a positive allosteric modulator," *The Journal of Biological Chemistry*, vol. 281, no. 30, pp. 21558-21565, 2006.

- [21] B. J. Arey, R. Seethala, Z. Ma, A. Fura, J. Morin, J. Swartz, V. Vyas, W. Yang, J. K. Dickson and J. H. M. Feyen, "A novel calcium-sensing receptor antagonist transiently stimulates parathyroid hormone secretion in vivo," *Endocrinology*, vol. 146, no. 4, pp. 2015-2022, 2005.
- [22] V. N. Babinsky, F. M. Hannan, C. M. Gorvin, S. A. Howles, M. A. Nesbit, N. Rust, A. C. Hanyaloglu, J. Hu, A. M. Spiegel and R. V. Thakker, "Allosteric modulation of the calcium-sensing receptor rectifies signaling abnormalities associated with g-protein  $\alpha$ -11 mutations causing hypercalcemic and hypocalcemic disorders," *Signal Transduction*, vol. 291, no. 20, pp. 10876-10885, 2016.
- [23] E. Vavers, L. Zvejniece, T. Maurice and M. Dambrova, "Allosteric modulators of sigma-1 receptor: A review," *Frontiers in Pharmacology*, vol. 10, p. 223, 2019.
- [24] B. Alberts, K. Hopkin, A. Johnson, D. Morgan, M. Raff, K. Roberts and P. Walter, "How Proteins are Controlled," in *Essential Cell Biology*, New York, W. W. Norton & Company, 2019, pp. 149-158.
- [25] P. Soto, G. M. Gloeb, K. A. Tsuchida, A. A. Charles, N. M. Greenwood and H. Hendrickson, "Insight into the conserved structural dynamics of the C-terminus of mammal PrPC identifies structural core and possible structural role of pharmacological chaperones," *Prion*, vol. 17, no. 1, pp. 55-66, 2023.
- [26] K. A. Reynolds, R. N. McLaughlin and R. Ranganathan, "Hotspots for allosteric regulation on protein surfaces," *Cell*, vol. 147, no. 7, pp. 1564-1575, 2011.
- [27] G. M. Verkhiver, "Making the invisible visible: Toward structural characterization of allosteric states, interaction networks, and allosteric regulatory mechanisms in protein kinases," *Current Opinion in Structural Biology*, vol. 71, pp. 71-78, 2021.
- [28] F. Ding, N. V. Dokholyan and E. Shakhnovich, "Emergence of protein fold families through rational design," *PLoS Computational Biology*, vol. 2, no. 7, 2006.
- [29] R. L. Redler, J. Das, J. R. Diaz and N. V. Dokholyan, "Protein destabilization as a common factor in diverse inherited disorders," *Journal of Molecular Evolution*, vol. 82, no. 1, pp. 11-16, 2016.

- [30] S. Yin, F. Ding and N. V. Dokholyan, "Modeling backbone flexibility improves protein stability estimation," *Structure*, vol. 15, no. 12, pp. 1567-1576, 2007.
- [31] "Welcome to ProteinLens," Yaliraki Lab, Imperial College London, 2019. [Online]. Available: <https://www.proteinlens.io/webserver/index>.
- [32] "Background," 2019. [Online]. Available: <https://www.proteinlens.io/webserver/background>.
- [33] B. Amor, S. N. Yaliraki, R. Woscholski and M. Barahona, "Uncovering allosteric pathways in caspase-1 using Markov transient analysis and multiscale community detection," *Molecular Biosystems*, vol. 10, p. 2247, 2014.
- [34] "AlloSigMA Tutorial," Bioinformatics Institute, Agency for Science, Technology, and Research, [Online]. Available: <http://allosigma.bii.a-star.edu.sg/tutorial/#>.
- [35] V. C. Yee, K. J. Knaus, M. Morillas, W. Swietnicki, M. Malone and W. K. Surewicz, "Crystal structure of the human prion protein reveals a mechanism for oligomerization," *Nature Structural & Molecular Biology*, vol. 8, no. 9, pp. 770-774, 2001.
- [36] "Standard Protein BLAST," [Online]. Available: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- [37] G. Di Fede, M. Catania, C. Atzori, F. Moda, C. Pasquali, A. Indaco, M. Grisoli, M. Zuffi, M. C. Guaita, R. Testi, S. Taraglio, M. Sessa, G. Gusmaroli, M. Spinelli, G. Salzano, G. Legname, R. Tarletti, L. Godi, M. Pocchiari and et al., "Clinical and neuropathological phenotype associated with the novel V189I mutation in the prion protein gene," *Acta Neuropathologica Communications*, vol. 7, no. 1, pp. 1-11, 2019.
- [38] I. Gandoglia, L. Strada, A. Poleggi, A. Castaldi, M. Del Sette and E. Di Maria, "Penetrance of the V203I variant of the PRNP gene: report of a patient with stroke-like onset of Creutzfeldt-Jakob disease and review of published cases," *Prion*, vol. 16, no. 1, pp. 19-22, 2022.
- [39] G. Schelzke, K. Stoeck, S. Eigenbrod, E. Grasbon-Frodl, L. M. Raddatz, C. Ponto, H. A. Kretzschmar and I. Zerr, "Report about four novel mutations in the prion protein gene," *Dementia and Geriatric Cognitive Disorders*, vol. 35, pp. 229-237, 2013.
- [40] Z. Zheng, M. Zhang, Y. Wang, R. Ma, C. Guo, L. Feng, J. Wu, H. Yao and D. Lin, "Structural basis for the complete resistance of the human prion protein mutant G127V to prion disease," *Scientific Reports*, vol. 8, no. 1, p. 13211, 2018.

- [41] R. Morales, C. Duran-Aniotz, J. Castilla, L. D. Estrada and C. Soto, "De novo induction of amyloid- $\beta$  deposition in vivo," *Molecular Psychiatry*, vol. 17, no. 12, pp. 1347-1353, 2011.
- [42] P. Campbell, "Bovine spongiform encephalopathy (BSE) - mad cow disease," *Biopolimery i Kletka*, vol. 15, no. 2, pp. 93-102, 1999.
- [43] M. T. Osterholm, C. J. Anderson, M. D. Zabel, J. M. Scheftel, K. A. Moore and B. S. Appleby, "Chronic wasting disease in cervids: Implications for prion transmission to humans and other animal species," *mBio*, vol. 10, no. 4, 2019.
- [44] N. Hunter, "Scrapie," *Molecular Biotechnology*, vol. 9, no. 3, pp. 225-234, 1998.
- [45] K. Peoc'h, P. Manivet, P. Beaudry, F. Attane, G. Besson, D. Hannequin, N. Delasnerie-Lauprêtre and J.-L. Laplanche, "Identification of three novel mutations (E196K, V203I, E211Q) in the prion protein gene (PRNP) in inherited prion diseases with Creutzfeldt-Jakob Disease phenotype," *Mutation in Brief*, vol. 323, 2000.
- [46] B.-H. Jeong, Y.-C. Jeon, Y.-J. Lee, H.-J. Cho, S.-J. Park, D.-I. Chung, J. Kim, S. H. Kim, H.-T. Kim, E.-K. Choi, K.-C. Choi, R.-I. Carp and Y.-S. Kim, "Creutzfeldt-Jakob disease with the V203I mutation and M129V polymorphism of the prion protein gene (PRNP) and a 17 kDa prion protein fragment," *Neuropathology and Applied Neurobiology*, vol. 36, pp. 558-563, 2010.
- [47] R. Diaz-Espinoza and C. Soto, "High-resolution structure of infectious prion protein: the final frontier," *Nat Struct Mol Biol*, vol. 19, no. 4, pp. 370-377, 2014.
- [48] L. C. Walker, "Prion-like mechanisms in Alzheimer disease," *Handb Clin Neurol*, vol. 153, pp. 303-319, 2018.
- [49] J. Ma, J. Gao, J. Wang and A. Xie, "Prion-like mechanisms in Parkinson's disease," *Front Neurosci*, vol. 13, p. 552, 2019.
- [50] J. D. F. Wadsworth, S. Joiner, J. M. Linehan, M. Desbruslais, K. Fox, S. Cooper, S. Croiner, E. A. Asante, S. Mead, S. Brandner, A. F. Hill and J. Collinge, "Kuru prions and sporadic Creutzfeldt-Jakob disease prions have equivalent transmission properties in transgenic and wild-type mice," *Proceedings of the National Academy of Sciences*, vol. 105, no. 10, pp. 3885-3890, 2008.

## SUPPLEMENTARY TABLES

*Supplementary Table 1. List of missense/nonsense mutations on PRNP.*

Codon Change	Amino Acid Change	Codon Number	Phenotype
AGT-GGT	Ser-Gly	17	Alzheimer disease, late-onset
CCG-CTG	Pro-Leu	39	Frontotemporal lobar degeneration
CAG-CCG	Gln-Pro	52	*Creutzfeldt-Jakob syndrome
GGT-AGT	Gly-Ser	54	*Wilson disease, neurological, modifier of
CCT-TCT	Pro-Ser	84	Gerstmann-Sträussler syndrome
AGT-AAT	Ser-Asn	97	Dementia
CCG-CTG	Pro-Leu	102	Gerstmann-Sträussler syndrome
CCA-ACA	Pro-Thr	105	Creutzfeldt-Jakob syndrome
CCA-CTA	Pro-Leu	105	Gerstmann-Sträussler syndrome
CCA-TCA	Pro-Ser	105	Prion disease
GGT-GTT	Gly-Val	114	Prion disease
GCA-GTA	Ala-Val	117	Gerstmann-Sträussler syndrome
GGC-GTC	Gly-Val	127	**Prion disease, resistance to
<b>ATG-GTG</b>	<b>Met-Val</b>	<b>129</b>	<b>**Gerstmann-Sträussler syndrome</b>
GGA-GTA	Gly-Val	131	Gerstmann-Sträussler syndrome
AGT-ATT	Ser-Ile	132	Prion disease
GCC-GTC	Ala-Val	133	Spongiform encephalopathy, familial
ATC-ATG	Ile-Met	138	Dementia
GGC-AGC	Gly-Ser	142	Creutzfeldt-Jakob syndrome
TAT-TAG	Tyr-STOP	145	Gerstmann-Sträussler syndrome
CGT-CAT	Arg-His	148	Creutzfeldt-Jakob syndrome
CGT-TGT	Arg-Cys	156	Dementia
CAA-TAA	Gln-STOP	160	Dementia
TAC-TAG	Tyr-STOP	162	Diarrhoea, autonomic failure & neuropathy
TAC-TAG	Tyr-STOP	163	Prion disease
GAT-AAT	Asp-Asn	167	Prion disease
AAC-AGC	Asn-Ser	171	*Schizoaffective disorder
AAC-AAG	Asn-Lys	173	*Creutzfeldt-Jakob syndrome
GTG-GGG	Val-Gly	176	*Gerstmann-Sträussler syndrome
GAC-AAC	Asp-Asn	178	Creutzfeldt-Jakob syndrome
GTG-ATC	Val-Ile	180	Creutzfeldt-Jakob syndrome
ACA-GCA	Thr-Ala	183	Spongiform encephalopathy, familial
CAC-CGC	His-Arg	187	Encephalopathy, familial
ACG-AAG	Thr-Lys	188	Dementia

ACG-AGG	Thr-Arg	188	Prion disease
ACG-GCG	Thr-Ala	188	Creutzfeldt-Jakob syndrome
<b>GTC-ATC</b>	<b><u>Val-Ile</u></b>	<b><u>189</u></b>	<b><u>Creutzfeldt-Jakob syndrome</u></b>
ACC-ATC	Thr-Ile	193	Creutzfeldt-Jakob syndrome
AAG-GAG	Lys-Glu	194	Creutzfeldt-Jakob syndrome
GAG-AAG	Glu-Lys	196	Creutzfeldt-Jakob syndrome
GAG-GCG	Glu-Ala	196	Creutzfeldt-Jakob syndrome
TTC-GTC	Phe-Val	198	Dementia, neurodegenerative
TTC-TCC	Phe-Ser	198	Gerstmann-Straeussler syndrome
GAG-AAG	Glu-Lys	200	Creutzfeldt-Jakob syndrome
GAG-GGG	Glu-Gly	200	Creutzfeldt-Jakob syndrome
GAC-AAC	Asp-Asn	202	Gerstmann-Straeussler syndrome
GAC-GGC	Asp-Gly	202	Dementia
<b>GTT-ATT</b>	<b><u>Val-Ile</u></b>	<b><u>203</u></b>	<b><u>Creutzfeldt-Jakob syndrome</u></b>
<b>GTT-GGT</b>	<b><u>Val-Gly</u></b>	<b><u>203</u></b>	<b><u>*Creutzfeldt-Jakob syndrome</u></b>
CGC-CAC	Arg-His	208	Creutzfeldt-Jakob syndrome
CGC-TGC	Arg-Cys	208	Dementia, neurodegenerative
GTT-ATT	Val-Ile	210	Creutzfeldt-Jakob syndrome
GAG-CAG	Glu-Gln	211	Creutzfeldt-Jakob syndrome
GAG-GAC	Glu-Asp	211	Gerstmann-Straeussler syndrome
CAG-CAC	Gln-His	212	*Creutzfeldt-Jakob syndrome
CAG-CCG	Gln-Pro	212	Gerstmann-Straeussler syndrome
ATC-GTC	Ile-Val	215	Creutzfeldt-Jakob syndrome
CAG-CGG	Gln-Arg	217	Gerstmann-Straeussler syndrome
TAC-AAC	Tyr-Asn	218	Gerstmann-Straeussler syndrome
GAG-AAG	Glu-Lys	219	**Creutzfeldt-Jakob syndrome, resistance to
GCC-GTC	Ala-Val	224	***Creutzfeldt-Jakob syndrome
TAC-TAA	Ala-STOP	226	Prion disease
CAG-TAG	Gln-STOP	227	Prion disease
ATG-ACG	Met-Thr	232	Gerstmann-Straeussler syndrome
ATG-AGG	Met-Arg	232	Creutzfeldt-Jakob syndrome
CCA-TCA	Pro-Ser	238	Prion disease
* not proven			
** in association with			
*** increased risk			
<i>Supplementary Table 1</i> gives a list of all mutations listed as missense or nonsense by the Human Gene Mutation Database (HGMD). Target mutations from this research are <b><u>underlined and bold</u></b> .			

*Supplementary Table 2.*  $\Delta\Delta G$  values for all mutations listed in Supplementary Table 1, sorted from highest to lowest average  $\Delta\Delta G$ ; mutations of interest are in bold and underlined.

mutID	1QM2 Trial 1	1QM2 Trial 2	1QM2 Trial 3	1QM2 Trial 4	Average 1I4M	Average $\Delta\Delta G$	Standard Deviation
V176G	2.11	5.86	3.35	6.61	9.34	6.91	3.05
Q212P	11.00	11.00	11.00	11.00	1.22	6.11	5.29
N171S	7.19	4.56	2.08	6.03	4.77	4.87	1.68
<b><u>V203G</u></b>	<b><u>4.31</u></b>	<b><u>3.88</u></b>	<b><u>1.87</u></b>	<b><u>3.73</u></b>	<b><u>5.29</u></b>	<b><u>4.37</u></b>	<b><u>1.60</u></b>
E200G	0.95	6.31	11.00	5.52	1.29	3.62	3.80
H187R	9.50	7.33	0.99	6.31	1.15	3.59	3.79
F198S	4.73	4.55	3.37	5.10	2.21	3.32	1.46
<b><u>V189I</u></b>	<b><u>2.59</u></b>	<b><u>3.03</u></b>	<b><u>3.66</u></b>	<b><u>2.89</u></b>	<b><u>3.24</u></b>	<b><u>3.14</u></b>	<b><u>1.21</u></b>
D202G	4.89	4.73	3.88	3.65	1.47	2.88	1.69
G142S	1.54	0.50	2.37	0.19	4.22	2.69	1.95
<b><u>V203I</u></b>	<b><u>7.21</u></b>	<b><u>3.52</u></b>	<b><u>-4.25</u></b>	<b><u>1.58</u></b>	<b><u>3.11</u></b>	<b><u>2.56</u></b>	<b><u>3.34</u></b>
<b><u>M129V</u></b>	<b><u>5.75</u></b>	<b><u>3.46</u></b>	<b><u>3.13</u></b>	<b><u>3.78</u></b>	<b><u>1.05</u></b>	<b><u>2.54</u></b>	<b><u>2.08</u></b>
N173K	1.61	2.11	3.42	2.92	2.50	2.51	1.33
I215V	3.06	5.21	3.78	4.90	-0.35	1.95	2.85
E200K	2.49	1.97	2.64	1.92	1.31	1.78	0.93
R208H	1.97	3.93	0.51	-3.46	2.63	1.68	2.34
T188K	0.61	1.92	-2.37	2.07	2.01	1.28	1.68
R148H	2.28	3.82	-1.26	3.12	0.53	1.26	1.67
K194E	0.55	1.05	-0.37	0.04	2.13	1.22	1.15
E211Q	-0.32	2.23	1.74	1.23	0.95	1.09	1.12
T188R	0.05	0.75	1.38	1.51	1.04	0.98	0.67
E219K	-0.55	1.36	0.25	1.45	1.14	0.88	0.69
V180I	-0.14	-0.14	-0.05	-0.05	1.64	0.77	1.15
E196K	-0.79	-0.19	-0.19	0.63	1.14	0.50	0.78
T188A	1.42	0.50	-1.91	0.72	0.80	0.49	1.24
D178N	0.95	1.28	0.36	-0.12	0.13	0.38	2.11
E211D	-1.83	-1.26	0.79	-0.11	1.32	0.36	1.39
Q212H	3.18	-1.86	-3.03	-1.23	1.43	0.35	2.21
A133V	0.06	1.58	-0.08	1.97	-0.51	0.19	1.67
D202N	1.98	-0.12	0.04	0.49	-0.73	-0.06	0.97
T183A	-3.22	-1.22	-0.34	0.26	0.88	-0.12	1.53
V210I	-0.14	-0.14	-0.05	-0.05	-0.75	-0.42	0.64
I138M	-1.65	-0.98	-3.09	-1.87	0.14	-0.88	1.54
D167N	-0.99	-0.99	-0.84	-0.84	-0.92	-0.92	0.08
G131V	-2.18	-2.50	-2.28	0.63	-0.57	-1.08	1.25
E196A	-0.97	-1.04	-1.51	-0.53	-1.22	-1.12	0.84
F198V	-1.15	-1.15	-0.43	-0.43	-1.57	-1.18	1.11

A224V	-1.56	-1.56	-1.17	-1.17	-1.37	-1.37	0.21
Y218N	3.63	2.06	1.33	2.31	-5.46	-1.56	4.22
S132I	-1.05	-1.68	-0.94	0.83	-2.58	-1.64	1.26
Q217R	-2.21	-2.16	-1.89	-1.06	-2.19	-2.01	1.23
T193I	-5.79	-1.71	-1.78	-0.80	-1.74	-2.13	2.24
G127V	-0.61	-3.13	-7.35	-2.62	-1.61	-2.52	2.19

The data in *Supplementary Table 2* was analyzed using R to determine if trial parameters had a significant effect on  $\Delta\Delta G$  values. Trial code was found to be statistically not significant on the resulting  $\Delta\Delta G$  value.

*Supplementary Table 3. Table of parameters outlining setting used for each Eris suite trial code Supplementary Table 2.*

<b>Trial Code</b>	<b>Parameters</b>
Trial 1	Fixed backbone, no backbone pre-relaxation
Trial 2	Fixed backbone, pre-relaxation on
Trial 3	Flexible backbone, no backbone pre-relaxation
Trial 4	Flexible backbone, pre-relaxation on