Finding the Structure of CPEB

Alison Saunders and Professor Daniel Cox
UC Davis Physics REU
(Dated: August 18, 2009)

The work of Eric Kandel suggests that prions may be involved in the formation of memory; prions contain misfolded and normal protein forms to code for "present" or "not present," similarly to binary. The neurons of aplysia contain the protein CPEB. CPEB exhibits prion-like properties when expressed in yeast cells and CPEB's action is almost impossible to reverse. I hence look at CPEB as a potential molecular basis for memory formation. I first thread the known protein sequence around a β-helical structure; threading is preformed by hand, and by a program written to minimize the energy cost of building the structure. I then analyze the stability of the thread using the molecular dynamics program AMBER9. I also analyzed a protein of only glutamine (PolyQ) in a β-helical structure, because the sequence of CPEB is comprised of many glutamine. I found PolyQ to be stable in a left-handed β-helical structure with eighteen residues per turn. A candidate structure for CPEB was located with the same β-helical structure.

INTRODUCTION

Relevant Biological Concepts

Proteins form through what is known as "The Central Dogma of Molecular Biology." A DNA sequence leads to RNA transcription, which leads to the expression of a protein through 20 naturally occurring amino acids. Amino acids are commonly referred to as "residues," and consist of a carboxyl functional group (COOH), an amine functional group (NH_2), and a side chain, known as the R group. Figure 1 shows a typical amino acid structure. The R group is what defines each of the 20 naturally-occurring amino acids; many are defined by various traits, including polar, hydrophilic, hydrophobic, and charged. Amino acids link together, and the sequence of amino acids in a protein defines the protein's primary structure. A protein's secondary structure arises from the form the strands take. Examples of an α-helix and a β-sheet can be seen in Figure 2 and in Figure 3. Both α-helices and β-sheets are typical secondary protein structures. The tertiary structure of a protein describes the overall configuration of various secondary structures contained within the protein.

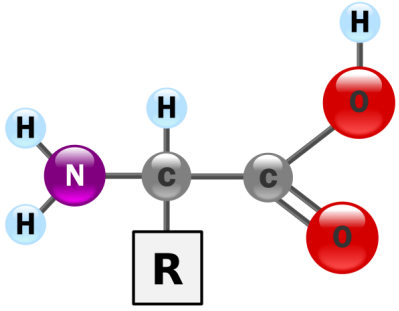

FIG. 1: The structure of an amino acid with an amine functional group (NH_2), a carboxyl functional group (COOH), and the R group. Image taken from Wikipedia.

FIG. 2: A cartoon of an α-helix obtained from Wikipedia.

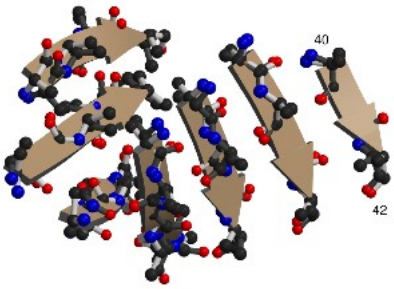

FIG. 3: A cartoon of a β-sheet obtained from Wikipedia.

Prions form an interesting group of proteins due to their properties. A prion is a misfolded protein that causes other proteins with the same primary structure to similarly misfold. Prions also introduce amyloid fibrils, which allow prions to polymerize into long strands. The ability to polymerize and the difficulty in reversing the polymerization makes prions dangerous to organisms. Several diseases, including Kuru and Creutzfeldt-Jakob Disease, are caused by prions. Humans contain a natural prion, labeled as PrP^C, which contains three α-helices.

The misfolded form, PrP^{Sc}, may take the form of a beta helix, which appears in Figure 4.

FIG. 4: A cartoon of a left-handed β-helix as seen from above.

The left-handed β-helix (LBH) appears naturally in bacterial enzymes, but is an uncommon protein structure. Three sides of six resides each make up one turn of the LBH, and the six residues are labeled as L1 through L6 (see Figure 5.) The position of the residue on a turn

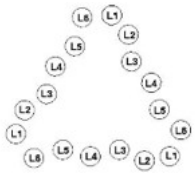

FIG. 5: The labeling of LBH residues.

determines what residues are likely to be located there by potential energy analysis, as further discussed in the technical background section. Some data from electron microscopy supports the idea of the misfolded human prion as taking an LBH structure. The LBH is also a candidate for amyloid fibril structures, as the β structure allows for the fast formation of strong hydrogen bonds between proteins.

Motivation

Eric Kandel predicted that the molecular basis for memory may rely on proteins like prions. He suggested that the prion mechanism for memory functions similarly to binary; misfolded signifies "present" (1), while normally folded signifies "not present" (0). The protein CPEB shows prion-like properties when experimentally

FIG. 6: The sequence of CPEB. Figure taken from Si et al.

expressed in yeast cells; misfolded proteins are passed on to daughter yeast cells. [1] CPEB arises from neurons of aplysia, a sea slug variety. Because the action of CPEB polymerization is hard to reverse and because CPEB is found in neurons, CPEB offers a candidate as a protein involved in long term memory formation. The glutamine (Q) rich sequence of CPEB can be seen in Figure 6. Glutamine is of potential interest to the amyloid community because long strands of glutamine in proteins cause Huntington's disease in humans. [2] Research shows that a protein of glutamine, called PolyQ, acts very stably under molecular dynamics simulations when put into a typical LBH structure. [3] The stability of LBH PolyQ found by Stork et al. was done on a molecular dynamics simulation program called CHARMM. I seek to reproduce the results using AMBER9.

Because the β-helix offers a candidate structure for amyloid fibrils and Stork et al. found PolyQ to be very stable around an LBH structure, I run molecular dynamics simulations to find candidates structures for CPEB. All structures are formed around an LBH template with eighteen residues per turn.

TECHNICAL BACKGROUND

As mentioned, I use the molecular dynamics software AMBER9 (Assisted Model Building with Energy Refinement) to test for stable models of the CPEB protein.

To begin, I thread the sequence of amino acids around the LBH structure. "Threading" signifies the placement of residues around the LBH turns in such a way that minimizes the energy cost of building and sustaining such a protein structure. I employ a program developed by a graduate student Youval Dar that scores various threads around an LBH structure. The threading program seeks to create the highest score possible by adding and subtracting points for the following:

Plus or Minus	Element
+/-	charged residues out/in
+/-	hydrophobic residues in/out
+/-	polar residues out/in
-	deviations from ideal loop volume
-	cut corner
-	loop
+	side chain hydrogen binding

The greatest cost in energy of LBH formation is when charged residues point inwards. Other penalties arise from the situations mentioned in the table. An ideal volume for the LBH is found by theoretical modeling, and deviations outside of that create less stable LBH structures. Helix corners are cut by L6 or L1 deletion to create greater energy gains for other residues around the LBH

structure. Loops are series of residues that appear between L6 and L1, or L1 and L2, to help make the other amino acids more stable, similarly to cutting corners.

The threading program finds the highest scored threads for a specific primary structure. A series of programs titled Jackal is then used to change residue names, cut corners, and create loops around a standard LBH template. I then place the protein into AMBER9 to run the simulation. I first add ions of Na+ and Cl- to neutralize the molecule. Another important aspect of protein simulation is to run a simulation in water, as opposed to in vacuum; water mimics the protein's true environment. With AMBER9, explicit solvent can be used by adding water molecules to the system. Implicit solvent can also be used through a Born solution model. Once I hydrate the protein, either implicitly or explicitly, I heat the water and protein to 300 K. The final molecular dynamics simulation runs for 10 ns, with output intervals of 20 ps. The AMBER9 output files contain the coordinates for all atoms in the system. AMBER9 obtains the coordinates by analyzing the potential energy functions of each atoms.

There are two categories of potential energies: bonded and non-bonded. The bonded energies depend on bond length (l), bond angle (defined by three atoms, θ), and bond torsion (defined by four atoms, ϕ). The potential energy functions are represented as follows:

$$V = \sum_{bonds} \frac{1}{2} k_b \left(l - l_0\right)^2 + \sum_{angles} k_a \left(\theta - \theta_0\right)^2$$

$$+ \sum_{torsions} \frac{1}{2} V_n \left[1 + \cos(n\phi - \gamma)\right]$$

where V is the potential energy, k_a and k_b act as effective spring constants, and γ is a phase shift in oscillation.

The non bonded energies depend on coulombic interactions and Van der Waals forces. The potential energies functions are represented as follows:

$$V = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left(\epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_1 q_2}{4\pi\epsilon_0 r_{ij}} \right)$$

where ϵ is the well-depth, σ is the equilibrium distance, and q is the charge on the atom.

AMBER preforms the analysis on the potential energy functions for each atom. It then finds the position of the atoms through the potential energy functions. The force on each atom is the negative sum of the gradient of the potential energies:

$$\vec{F_i} = -\sum_{j=i} \frac{\partial U_{ij}}{\partial \vec{r_i}},$$

the velocity of the atoms can be found from the force:

$$\vec{v_i} = \int \frac{\vec{F_i}}{m_i} dt$$

and then the position can be found from the velocity:

$$\vec{r_i} = \int \vec{v_i} dt \,.$$

AMBER also accounts for temperature within its calculations by way of the Langevin equation. A simple version of the Langevin equation follows:

$$m\dot{v}(t) + \mu v(t) = f_s(t) + f_r(t)$$

where m is the mass of the particle, μ is the coefficient of friction due to the surrounding liquid, v is the velocity of the particle, f_s is the systematic force being applied to the particle, and f_r is the random force acting on the particle due to random thermal motion. The Langevin equation assures that the root mean squared velocities of all atoms abide by the equipartition principle, allowing AMBER to account for temperature in its force field calculations.

Data analysis from AMBER appears in the root mean squared deviation (RMSD) of the molecule. The RMSD sums over the motion of all atoms, taking out side-to-side motion and rotational motion of the whole molecule. RMSD squares the velocities, then divides the summation by the number of atoms in the system. The RMSD then provides a mean measure of much each atom in the molecule changed its position. A stable protein typically exhibits an RMSD plot with a horizontal line of best fit found at about 1-2 Å.

RESULTS AND DISCUSSION

PolyQ

To reproduce the results of Stork et al., I ran molecular dynamics simulations with two-turn, three-turn, and six-turn PolyQ with an 18 residue per turn LBH structure.

Two turns of glutamine produced in a β-helical form provided an unstable protein. The RMSD can be seen in Figure 7, and shows that the protein never reached a stable equilibrium point. The instability of two-turn LBH PolyQ is inline with the findings of Si et al.

Three turns of glutamine in a β-helical structure did produce stable results. The RMSD in Figure 8 shows the protein settles into a solid structure at about 1000 picoseconds without much fluctuation after that point.

Figure 9 shows the three-turn PolyQ after minimization and then after the molecular dynamics simulation. While PolyQ's RMSD suggests stability, PolyQ's structure no longer resembles the ideal LBH after the final simulation. The results of an altered structure may suggest that PolyQ is attempting to obtain a different shape. Replica exchange models may be run to determine if there is a favorable for three-turned PolyQ.

I also simulated a six-turn PolyQ LBH. The RMSD, shown in Figure 10, implies the structure never found

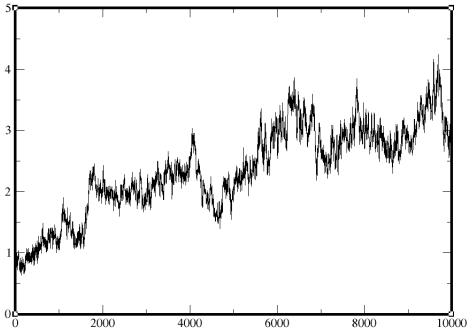

FIG. 7: The RMSD plot of two-turn PolyQ. The vertical axis is in Angstroms and the horizontal axis measures time in picoseconds.

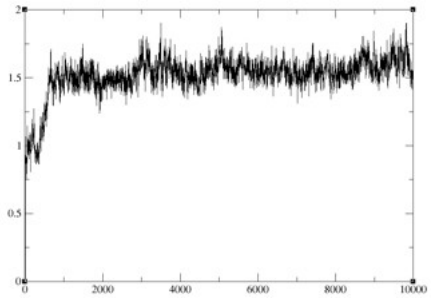

FIG. 8: The RMSD graph of three stranded PolyQ. The vertical axis is in Angstroms and the horizontal axis is in picoseconds.

a completely stable equilibrium point. However, the RMSD does not go above 1.5 Å, suggesting that six-turn PolyQ found reasonable stability in LBH form. Figure 11 shows the change in structure during the molecular dynamics simulation. The two LBH look very similar in appearance, implying that the LBH structure may provide stability to this protein.

PolyQ with six turns then appears more stable in the LBH structure than does PolyQ with three turns, by way of looking at the proteins' appearances before and after molecular dynamics simulations. The finding of greater stability with more turns supports the finding of Ogawa et al.; there is a certain critical number of residues required for the β-helical structure to be stable. More research could be done to find the exact number glutamine residues required to make a stable LBH structure. These results could also be compared to the number of residues involved in the onset of Huntington's disease, to provide insight into the disease's action.

CPEB

I then ran simulations of the gulatmine-rich CPEB in an LBH form, given the stability found with PolyQ.

The first molecular dynamics simulations I performed

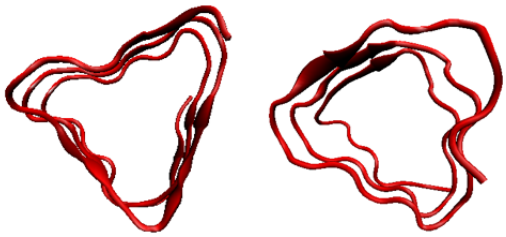

FIG. 9: Three-turn PolyQ after minimization (left) and after molecular dynamics simulation (right).

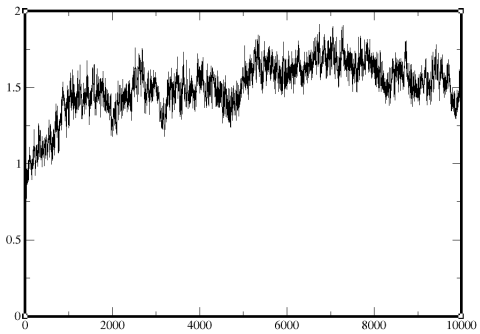

FIG. 10: The RMSD for six-turn PolyQ. The vertical axis is in Angstroms and the horizontal axis measures time in picoseconds.

were using threads provided by the threading program. The first thread appeared as follows:

$$MQAM|-AVASQ|SPQTVD|QAISVK|$$

$$(TDYEDN)QQEHIP|SNFEIF|RRINAL|$$

$$(LD)NSLEAN|-NVSCS|QS$$

where the letters are the 1-letter codes for amino acids, the vertical lines represent divisions between sides, the hyphens represent corner cuts, and residues in parenthesis represent loops. The RMSD, seen in Figure 12, shows the thread to have created a completely unstable protein.

The unsuccessful threading of CPEB led to analysis of the threading program. The program takes into account the volume of the residue side chains as a factor for analyzing potential threads. However, the program was using the masses of the amino acids to calculate the volumes, assuming constant density. The assumption of constant density proved incorrect, and the true volume of glutamine was then used. The threads produced after the volume correction were different and the scores for forming potential LBH's were increased. The runs done with newer threads are discussed below.

Another method of threading an amino acid sequence onto a protein is to thread by hand. This involves making a spread sheet and using the guidelines the threading

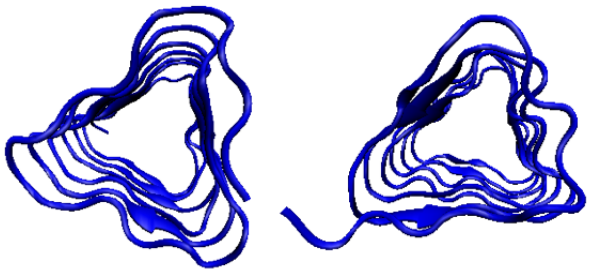

FIG. 11: Six-turn PolyQ after minimization (left) and after molecular dynamics simulation (right).

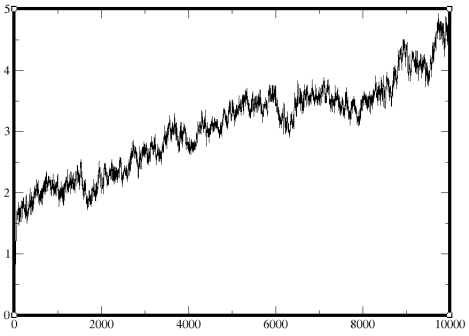

FIG. 12: The RMSD for the first simulation of CPEB. The vertical axis is in Angstroms and the horizontal axis measures time in picoseconds.

program uses to find a potential thread. Initial results from this method were more successful than the initial results from the threading program. The thread used was:

$$MQAMA|VASQSP|QTVDQA|ISVKTD|$$

$$(YEDNQQEHIP)SNFEIF|RRINAL$$

$$|LDNSLE|ANNVSC| - SQS$$

The RMSD, seen in Figure 13, seems reasonable, but still does not indicate a desired structure. Other hand-threaded results were far less desirable.

More threads arrived from the threading program once the volume element of glutamine was corrected. Initial threads from the corrected program were unsuccessful. The first thread run contained many large loops, and is shown as follows:

$$MQAMAV|ASQSPQ|TVDQAI|$$

$$(SVKTYDYEDNQQEHIP)SNFEIF|$$

$$RRINAL|LDNSLE|(ANNVSCSQSQSQQQQ)$$

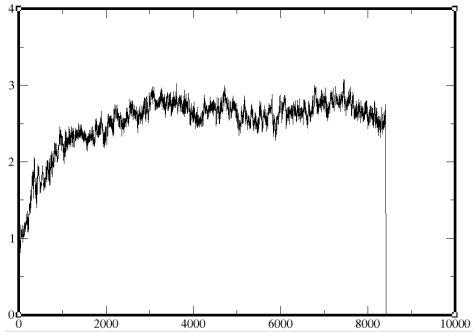

FIG. 13: The RMSD plot of hand threaded CPEB. The vertical axis is in Angstroms and the horizontal axis measures time in picoseconds.

$$QQTQQQ|QQQQQQ|QQQQQH$$

Figure 14 shows the RMSD of the thread with larger loops from the corrected program, and suggests an unstable protein structure.

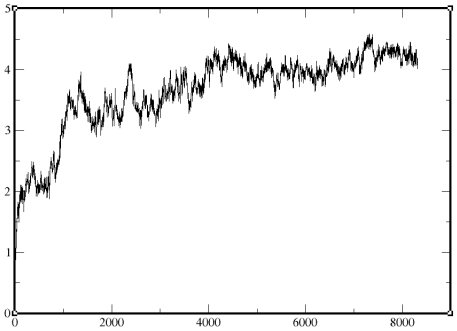

FIG. 14: The RMSD plot of CPEB with bigger loops, and ran with the improved threading program. The vertical axis measures Angstroms and the horizontal axis measures time in picoseconds.

Noting that the loop size was large, I eliminated the loops from the RMSD calculation. Figure 15 shows the RMSD with the loops eliminated. Unfortunately, the structure still did not show stability.

Another thread provided by the program had a maximum loop length of six residues. The thread appeared as follows:

$$MQAMAV|(ASQSP)QTVDQA|ISVKTD|$$

$$YEDNQQ|EEIFRR|INALLD|NSLEAN|$$

$$N(VSCSQ)SQSQQ|QQQQTQ$$

Again the RMSD plot did not suggest stability. Figure 16 shows the RMSD of the above thread.

I decided to again mask out loops from the RMSD calculations. Figure 17 shows the RMSD with the loops masked.

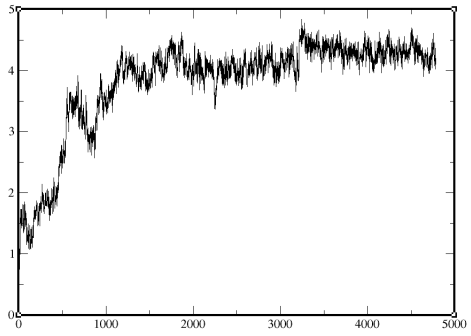

FIG. 15: The RMSD plot of CPEB (larger loops from the improved threading program) with loops masked out of RMSD calculations. The vertical axis measures Angstroms, and the horizontal axis measures frames of 20 picoseconds.

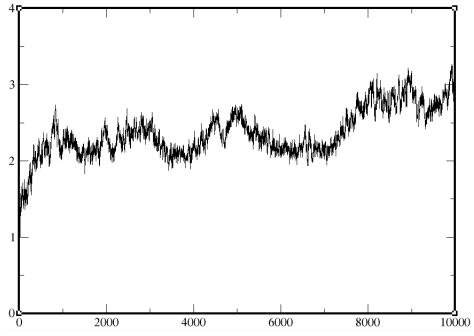

FIG. 16: The RMSD plot of CPEB ran with the improved threading program. The vertical axis is in Angstroms and the horizontal axis measures time in picoseconds.

This RMSD of the protein with all small loops masked never exceeds a value of 2 Å, but the value fluctuates, implying the LBH structure with this thread does not stay stable under molecular dynamics simulations. Noting that large loops occur in the beginning of all threads, the threading program was rerun with only 118 amino acids, removing the initial looped section. The thread then appeared as follows:

$$EIF|RRINAL|LDNSLE|-ANNVS|CSQSQS|$$

$$QQQQQQ|TQQQQQ|QQQQQQ|QQQHLQ|$$

$$-QVQQQ|RLLKQQ|-QQQAQ|-RQQIQ|$$

$$QQLLQQ|QQQKQQ|LQQQQQ|QEQLQQ|$$

$$QQLQLQ|QQLQQQ|-LQHIQ|KEPSSH$$

The first three turns produced an RMSD that did not suggest stability; the plot never created a horizontal pattern, and the RMSD values fluctuated up to 6 Å. Figure 18 shows the RMSD plot.

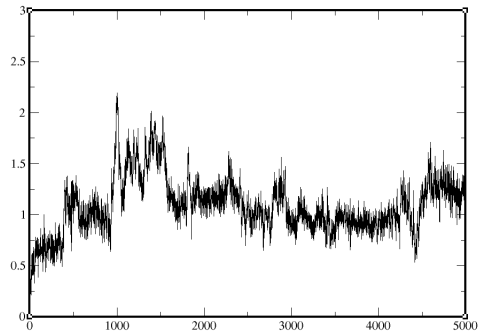

FIG. 17: The RMSD plot of CPEB ran with the improved threading program. In this case, the loops are masked out of the RMSD calculations. The vertical axis measures Angstroms and and horizontal axis is in picoseconds.

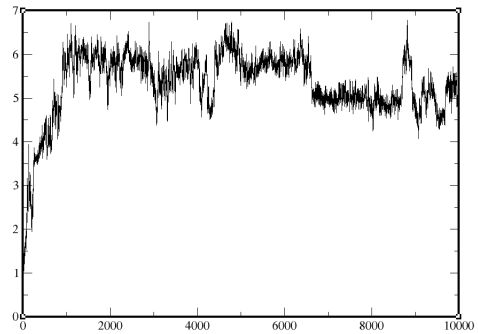

FIG. 18: The RMSD plot of the first three turns of CPEB, threaded without the initial 42 residues. The vertical axis is measured in Angstroms and the horixontal axis is measured in picoseconds.

A visual examination of the protein after simulation provided some insight into why the RMSD suggested instability. Figure 19 shows the structure after simulation, as seen from above.

FIG. 19: The first three turns of the final thread after simulation, as seen from above.

Clearly, the first and the last sides did not stay affixed to the main structure; they moved about throughout the simulation. The thread I ran did begin and end within the middle of the protein strand, implying that the ends

of the thread might indeed not stay stable. I decided to analyze the RMSD without consideration of the first and last six residues. Figure 20 shows the RMSD with the ends masked, and suggests a very stable structure.

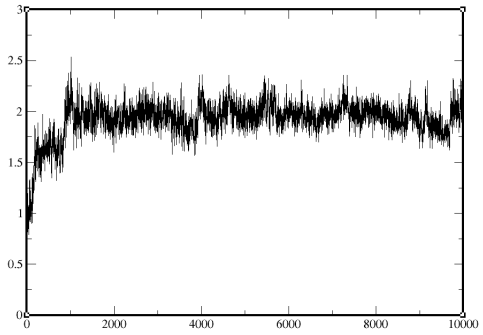

FIG. 20: The RMSD plot of the first three turns of CPEB with the first and last six residues masked. The vertical axis is measured in Angstroms and the horixontal axis is measured in picoseconds.

The RMSD reaches a value of 2 Å and stays constant thereafter. The stability of this structure leads to more analysis to be performed. The thread should be re-run using explicit solvent, as this structure was analyzed using implicit solvent. Likewise, the full thread should be run under molecular dynamics simulations. The potentially stable thread should then be completed, by including the first region of residues that were removed to aid in the creation of the stable thread by the threading program. This thread offers a potential candidate for CPEB, as the RMSD does indeed suggest stability.

CONCLUSION

Overall, a candidate thread for the structure of CPEB was located. Further work remains to be done with the candidate structure. The three strands found to be stable should be run with explicit solvent, and the full thread should also be run with explicit solvent. Once a stable structure is located, it remains to be experimentally tested. There also should be more work done with PolyQ to determine the minimum amount of residues in the β-helical structure. The minimum number of residues required for stability should also be tested experimentally. Data from theoretical and experimental work should be looked at in accordance with the number of glutamine present with the onset of Huntington's Disease.

The end goal of finding the structure of CPEB is to provide insight into amyloid structure and function. If a stable structure is found, other proteins exhibiting similar prion-like properties should be examined in a β-helical structure. Knowledge of amyloid structure can then lead to information on amyloid aggregation kinetics. With knowledge of structure, function, and kinetics, the molecular basis for memory may be located in various species, offering insight into the perpetually perplexing function of the human mind.

[1] K. Si, S. Lindquist, and E. R. Kandel, Cell **115**, 879 (2003).
[2] H. Ogawa *et al.*, Computational Biology and Chemistry **32**, 102 (2008).
[3] M. Stork, A. Giese, H. A. Kretzschmar, and P. Tavan, Biophysical Journal **88**, 2442 (2005).