Copyright

by

James Edward Jennings

2019

THE EFFECTS OF QUANTUM TUNNELING ON THE CONFORMATIONAL ENTROPY OF THE CAS9 PROTEIN

by

James Edward Jennings, BS

THESIS

Presented to the Faculty of

The University of Houston-Clear Lake

In Partial Fulfillment

Of the Requirements

For the Degree

MASTER OF SCIENCE

in Physics

THE UNIVERSITY OF HOUSTON-CLEAR LAKE

DECEMBER 2019

THE EFFECTS OF QUANTUM TUNNELING ON THE CONFORMATIONAL ENTROPY OF THE CAS9 PROTEIN

by

James Edward Jennings

APPROVED BY

Van Eric Mayes, PhD, Chair

Samina Masood, PhD, Committee Member

Paul Withey, PhD, Committee Member

APPROVED/RECEIVED BY THE COLLEGE OF SCIENCE AND ENGINEERING:

David Garrison, Ph.D, Interim Associate Dean

Miguel Gonzalez, Ph.D, Dean

Acknowledgements

I would like to think my father Jim Jennings and my brother William Jennings for helping me out with the python code.

ABSTRACT

THE EFFECTS OF QUANTUM TUNNELING ON THE CONFORMATIONAL ENTROPY OF THE CAS9 PROTEIN

James Edward Jennings University of Houston-Clear Lake, 2019

Thesis Chair: Van Eric Mayes

The Cas9 protein is a protein that is used in bacteria to edit out viral DNA. Cas9 is very precise and is used in laboratories all over the world for the purposes of editing gene sequences. In order for Cas9 to function properly it needs to fold into a specific shape. These shapes are referred to as conformations and the entropy corresponding to the number of conformations that a polymer chain can fold into is referred to as the conformational entropy. Until recently, the current model used for calculating the conformational entropy failed to incorporate the effects due to hydrogen bonds. Now a research group from Chicago has constructed a model that incorporates these effects. It has been demonstrated that the effects due to quantum interactions without tunneling effects influence the conformational entropy by an order of magnitude of 0.7% for backbone interactions and 10% for sidechain interactions respectively. The effects due to tunneling further influence the backbone entropy by an order of magnitude of 0.08%.

v

TABLE OF CONTENTS

List of Tablesv	/11
List of Figuresvi	 111
Chapter Pag	ge
CHAPTER I: THE CLASSICAL CONFORMATIONAL ENTROPY	.1
Introduction Background The Bead Spring Model The Lattice Model	.1 .2 .5 [1
CHAPTER II: QUANTUM MECHANICAL PERTURBATION2	21
Quantum Mechanical Interactions2Backbone-Backbone Interactions2Sidechain-Sidechain Interactions3Solvent Interactions3Energy Eigenstates4	21 28 34 38 42
CHAPTER III: COMPARISON OF THE CLASSICAL AND SEMI-CLASSICAL ENTROPIES	45
Conformational Entropy Correction	45 46 51 57
CHAPTER IV:	53
Conclusion	53
REFERENCES	54
APPENDIX A: AMINO ACID TABLES	56

LIST OF TABLES

Table	Page
Table 1: Entropies for the FoldX model, the Morse Potential, and the quantum tunn model.	eling 62
Table 2: Total energies for the FoldX model, the Morse Potential, and the quantum tunneling model.	62
Table A1: Hydrogen bond probabilities	66
Table A2: Hydrogen acceptors	67
Table A3: Sidechain vector combinations	68

LIST OF FIGURES

Figure	Page
Figure 1: Bead Spring Model	2
Figure 2: Lattice Model	3
Figure 3: The structure of an Alpha helix	29
Figure 4: The structure of a Beta strand	29

CHAPTER I: THE CLASSICAL CONFORMATIONAL ENTROPY

Introduction

One very important concept in the field of Polymer Physics is the concept of hydrogen bonds. In the field of Polymer Physics, Hydrogen bonds are what maintain a protein's conformation. Proteins are made up of 20 types of amino acids. All amino acids contain two components. These components are the side chain and the backbone respectively. The backbone of every amino acid has the same structure. The side chain of every amino acid however is different. Some side chains can form hydrogen bonds while other side chains are incapable of forming hydrogen bonds. Hydrogen bonds come in four different varieties. These varieties are N—H bonds, O—H bonds, S—H bonds and π —H bonds respectively. For the purposes of this research only the N-H and O-H bonds will be taken into account. The interactions due to the S—H bonds and π —H bonds will be set to zero. There are two different types of hydrogen bonds that form between the components of the amino acids. They are hydrogen bonds between side chains and hydrogen bonds between backbones. In addition to these bonds, hydrogen bonds can form between side chains and solvents. For the purposes of this research the system under study will be constrained so that the only solvent-sidechain interactions are those involving water.

Background

The conformational entropy of a polymer chain is the entropy due to the number of geometries that the polymer chain can assume. These geometries are referred to as conformations. As the number of possible geometries increases so does the conformational entropy. There are two different models for computing the conformational entropy. These models are referred to as the bead spring model and the lattice model. Both models treat the polymer as an ideal chain where it is assumed that there are no interactions between the amino acids.

In the bead spring model, the polymer is broken into a set of sub chains which are treated as harmonic springs. Each sub chain is a series of identical sequences of amino acids called repeat units. The ends of these sub chains are referred to as the beads within the model. A diagram pertaining to this model is given below.



Figure 1: Bead Spring Model

The number of conformations that a polymer chain can assume is dependent on the distance between the ends of the polymer chain. This is referred to as the end to end distance. As the end to end distance is constrained to a larger value, the number of ways in which the polymer chain can fold up become increasingly limited. Hence, the number of available conformations decreases as a result as well as the entropy. Since higher values for the conformational entropy correspond to lower end to end distances, lower values for the end to end distance are more probable than high values. The probability that the end to end distance will be some specified value *R* is given by the formula P(R). In a nutshell, the bead spring model is an approximation which treats the polymer chain as a sequence of harmonic springs. These harmonic springs represent the sub chains. The sub chains in turn are a series of repeat units, where the repeat units are sequences of amino acids. Each harmonic spring is held together by beads, where each bead is an amino acid at the end of each sub-chain. The distance between the two ends of the polymer is constrained to a specific value *R*. The segments of the polymer chain are free to move about randomly since it is treated as an ideal chain.

In the lattice model the polymer chain is constrained so that each of the amino acids occupy a site on a 3-dimensional lattice. Since the polymer is treated as an ideal chain, the amino acids do not interact with each other in the model. As a result, the probabilities for each bond vector to take on a certain value are all equal for every amino acid. The amino acids in the polymer chain are each held together by harmonic springs. So long as the number of amino acids in this model is sufficiently large the lattice model is equivalent to the bead spring model. A diagram pertaining to the lattice model is given below for the two-dimensional case.

	0	0	•	0	0	0	0	0	0
	0	0	•	0	0	0	0	0	0
	0	0		•	-	0	0	0	0
	0	0	0	0	•	0	0	0	0
\mathbf{v}	0	0	0	0		•	•	-	0
	•	-•	0	0	0	0	0	•	0
	0	•	0	0	•	•	•	-	0
	0	•	-•-	•	-•	0	0	0	0
	0	0	0	0	0	0	0	0	0
—	l.	-	_						
F	l∿∏				X	→			
						-			

Figure 2: Lattice Model

Within this model the end to end distance is constrained to some specified constant *R*, just like in the bead spring model. Within the lattice model there is an equation corresponding to the probability that the polymer chain will take on a certain specified distance. This is the same equation as that corresponding to the bead spring model. Both equations are referred to as P(R) where P(R) is the probability for the end to end distance to take on some specified value *R*. As stated previously the end to end distance is the distance between both ends of the polymer chain.

To sum it all up, the lattice model is a model in which the amino acids are each fixed onto points on a three-dimensional lattice. These amino acids are linked together by peptide bonds which are treated as harmonic springs in the model. Just like in the bead spring model, the distance between the ends of the polymer chain are constrained to a specific distance R. Within the lattice model the amino acids are free to move onto different lattice sites, while the end to end distance is fixed.

In summary, there are two models for approximating the classical conformational entropy of a protein. These models are referred to as the bead spring model and the lattice model. The bead spring model treats the polymer as a series of sub chains held together by the amino acids at their ends. The sub chains as well as the amino acids mentioned are referred treated as harmonic springs and beads respectively.

The lattice model on the other hand treats the polymer as a chain of amino acids held together by peptide bonds. The peptide bonds are treated as harmonic springs within this model. Within this model the amino acids are constrained to occupy sites on a threedimensional lattice. Both models will be used to calculate the entropy of the Cas9 protein.

The Bead Spring Model

The Classical Conformational Entropy is given by the following equation.

$$S_0(R) = ln\left[\frac{Z_0}{\mathcal{V}} \times P(R)\right]$$

Where Z_0 is the number of conformations that a polymer can take on regardless of constraints. P(R) is the probability that the distance between both ends of the polymer will take on a distance R. This is referred to as the end to end vector. The number of conformations that a polymer can take on when constraints are taken into account is dependent on the end to end vector. Large distances correspond to less possible conformations and small distances correspond to more possible conformations. The formula for P(R) is given below.

$$P(R) = \int dr_0 \dots \int dr_N \delta(r_N - r_0 - R) P(\{r_i\}), \qquad P(\{r_i\}) = \frac{1}{Z_0} e^{-\beta \mathcal{H}_0}$$

Where $P(\{r_i\})$ is the probability of finding the conformation of an ideal chain. The Dirac delta $\delta(r_N - r_0 - R)$ requires that the polymer be constrained such that the only geometries that are considered when computing the conformational entropy $S_0(R)$ are the geometries where the distance between the vectors r_N and r_0 is equal to R. In order to accomplish this task, one must abandon the notion of certainty. Instead of assuming that the total number of conformations Z_0 is fixed, there is a probability P(R)that there will be a certain number of conformations Z_0 . This is due solely to the fact that the set of conformations Z_0 corresponding to the case where $r_N - r_0 = R$ is not the only possible outcome. There are plenty of other conformations that correspond to cases where $r_N - r_0 \neq R$. In other words the case where $r_N - r_0 = R$ is not the only probability. One must give up certainty when calculating the entropy $S_0(R)$. The probability that the total number of conformations will be Z_0 is given by the formula $\Omega(R) = Z_0 \times P(R)$. The formula for Z_0 is given below along with the Hamiltonian $\mathcal{H}_0(r_0, \dots, r_N)$.

$$Z_{0} = \int dr_{0} \dots \int dr_{N} e^{-\beta \mathcal{H}_{0}(r_{0},\dots,r_{N})}, \mathcal{H}_{0} = \frac{3k_{B}T}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}$$

Where Z_0 is the total number of conformations regardless of the constraint $r_N - r_0 = R$. Likewise, $\mathcal{H}_0(r_0, ..., r_N)$ is the Hamiltonian of the system. This Hamiltonian is equivalent to the potential energy of the peptide bonds. In the equation $\mathcal{H}_0(r_0, ..., r_N)$ above N is the total number of sub chains and r_i and r_{i+1} are the end to end vectors corresponding to the *ith* sub chain. T is the temperature of the system. k_B is the Boltzmann constant and b is the average length of the sub-chains. Next the terms Z_0 and P(R) will be evaluated. With the use of a gaussian integral given to the left, we obtain the following calculations for Z_0 .

$$\int dr e^{-a|r-b|^2} = \left(\frac{\pi}{a}\right)^{3/2}, \qquad Z_0 = \mathcal{V}\left(\frac{2\pi b^2}{3}\right)^{3N/2}$$

Likewise, the calculations for P(R) are given below. Since the integral for P(R) is computationally intensive the following equations for the Dirac Deltas $\delta(r)$, $\delta(r_N - r_0 - R)$ will be employed to streamline these calculations. These equations are given below.

$$\delta(r) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r-a)}, \qquad \delta(r_N - r_0 - R) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_N - r_0 - R)}$$

In the equations above the term $\delta(r)$ corresponds to the integral form of the Dirac delta in three dimensions. Likewise, the integral for $\delta(r_N - r_0 - R)$ is also given above in the top right. This equation will appear frequently for the calculations for the integral term P(R). Another integral that will be used repeatedly will be the Fourier transform of a gaussian, which is given by the following equation below.

$$\int dr e^{-a|r-c|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a} - iq \cdot c}$$

As stated previously the formulas for P(R) and $P(\{r_i\})$ are given by the following equations below. In the bottom equation the Dirac Delta function $\delta(r_N - r_0 - R)$ is converted into an integral, as stated previously in the paragraphs above.

$$P(R) = \int dr_0 \dots \int dr_N \delta(r_N - r_0 - R) P(\{r_i\}), \qquad P(\{r_i\}) = \frac{1}{Z_0} e^{-\beta \mathcal{H}_0}$$
$$P(R) = \frac{1}{(2\pi)^3} \int dr_0 \dots \int dr_N \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_N - r_0 - R)} \frac{1}{Z_0} e^{-\beta \mathcal{H}_0}$$

Below are the calculations for the integral P(R). While these are mostly simplifications, they are necessary in order to avoid miscalculations. This will be explained later when the formula for the Fourier transform is displayed. These calculations are continued in the following pages.

Below are the calculations for the P(R) integral. In order to streamline calculations even further the Fourier transform of a gaussian will be employed in order to evaluate P(R).

$$P(R) = \frac{1}{(2\pi)^{3}Z_{0}} \int dr_{0} \dots \int dr_{N} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_{N} - r_{0} - R)} e^{-\beta \mathcal{H}_{0}}$$

$$P(R) = \frac{1}{(2\pi)^{3}Z_{0}} \int dr_{0} \dots \int dr_{N} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_{N} - r_{0} - R)} e^{-\beta \frac{3k_{B}T}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}}$$

$$P(R) = \frac{1}{(2\pi)^{3}Z_{0}} \int dr_{0} \dots \int dr_{N} \int_{-\infty}^{+\infty} dq e^{-iq \cdot (-r_{N} + r_{0} + R)} e^{-\beta \frac{3k_{B}T}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}}$$

$$P(R) = \frac{1}{(2\pi)^{3}Z_{0}} \int dr_{0} \dots \int dr_{N} \int_{-\infty}^{+\infty} dq e^{-\beta \frac{3k_{B}T}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}} e^{-iq \cdot (-r_{N} + r_{0} + R)}$$

$$P(R) = \frac{1}{(2\pi)^{3}Z_{0}} \int dr_{0} \dots \int dr_{N} \int_{-\infty}^{+\infty} dq e^{-\beta \frac{3k_{B}T}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}} e^{-iq \cdot (-r_{N} + r_{0} + R)}$$

Again, the Fourier transform of a gaussian is given by the following equation for convenience.

$$\int dr e^{-a|r-c|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a} - iq \cdot c}$$

As stated in the previous pages this equation will be deployed repeatedly in order to obtain accurate results for the integral P(R). Since $e^{-iq \cdot r}$ is given to the left of the gaussian term $e^{-a|r-c|^2}$ in the integral $\int dr e^{-a|r-c|^2} e^{-iq \cdot r}$, the term $e^{iq \cdot (r_N - r_0 - R)}$ in the integral P(R) was shifted to the left in order to prevent confusion and avoid miscalculations. The calculations for P(R) are continued below and will be continued in the following pages.

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_1 \dots \int dr_N \int_{-\infty}^{+\infty} dq \left(\frac{2\pi b^2}{3}\right)^{3/2} e^{-\frac{|q|^2 2b^2}{4 \cdot 3} - iq \cdot r_1} e^{iq \cdot r_N} e^{-iq \cdot R}$$

$$\times e^{-\beta \frac{3k_B T}{2b^2} \sum_{i=1}^{N-1} |r_i - r_{i+1}|^2}$$

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_1 \dots \int dr_N \left[e^{-\beta \frac{3k_B T}{2b^2} \sum_{i=1}^{N-1} |r_i - r_{i+1}|^2} e^{-iq \cdot r_1} \right]$$

$$\times \left(\frac{2\pi b^2}{3}\right)^{3/2} e^{-\frac{|q|^2 2b^2}{4 \cdot 3}} e^{iq \cdot (r_N - R)}$$

As was stated in the previous pages the formula for the Fourier transform of a gaussian is given below along with the integral for P(R).

$$\int dr e^{-a|r-c|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a} - iq \cdot c}$$

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_2 \dots \int dr_N \left[\left(\frac{2\pi b^2}{3}\right)^{3/2} e^{-\frac{|q|^2 2b^2}{4\cdot 3} - iq \cdot r_2} \right] \times \left[e^{-\beta \frac{3k_B T}{2b^2} \sum_{i=2}^{N-1} |r_2 - r_{i+1}|^2} \right] \times \left[\left(\frac{2\pi b^2}{3}\right)^{3/2} e^{-\frac{|q|^2 2b^2}{4\cdot 3}} e^{iq \cdot (r_N - R)} \right]$$

After multiple integrations are performed for P(R), the integral is distilled to the following equation given below.

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_{N-1} \dots \int dr_N \left[\left(\frac{2\pi b^2}{3} \right)^{3(N-1)/2} e^{-(N-1)\frac{|q|^2 2b^2}{4\cdot 3}} e^{-iq \cdot r_{N-1}} e^{iq \cdot (r_N - R)} \right] \\ \times e^{-\beta \frac{3k_B T}{2b^2} |r_{N-1} - r_{N+1}|^2}$$

The equations for the Dirac delta term as well as the Fourier transform are displayed below for convenience. The first term below corresponds to the Fourier transform displayed previously. Beneath this equation are the integrals for the Dirac delta equations $\delta(r_N - r_0 - R)$ and $\delta(r_{N-1} - r_0 - R)$ respectively.

$$\int dr e^{-a|r-c|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a} - iq \cdot c}$$
$$\delta(r_N - r_0 - R) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_N - r_0 - R)}$$
$$\delta(r_{N-1} - r_0 - R) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot (r_{N-1} - r_0 - R)}$$

With the use of the equations given above for the Fourier transform and the Dirac Delta terms $\delta(r_N - r_0 - R)$ and $\delta(r_{N-1} - r_0 - R)$ the integral P(R) is reduced to the following equation below. The calculations for P(R) will be continued on the following pages along with the computation for the classical conformational entropy $S_0(R)$.

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_N \left[\left(\frac{2\pi b^2}{3} \right)^{3(N-1)/2} e^{-(N-1)\frac{|q|^2 2b^2}{4\cdot 3}} e^{-\frac{|q|^2 2b^2}{4\cdot 3}} e^{-iq\cdot R} e^{iqr_N} \right]$$

In order to compute the classical conformational entropy $S_0(R)$ it is necessary to make use of the equations for both Z_0 and P(R). These equations, along with $S_0(R)$ are displayed beneath the calculations for P(R).

$$P(R) = \frac{1}{(2\pi)^3 Z_0} \int dr_N \left[\left(\frac{2\pi b^2}{3} \right)^{3N/2} e^{-N \frac{|q|^2 2b^2}{4 \cdot 3}} e^{-iq \cdot R} e^{iqr_N} \right]$$
$$P(R) = \frac{1}{(2\pi)^3} \left(\frac{6\pi}{Nb^2} \right)^{3/2} e^{-\frac{|R|^2 6}{4Nb^2}}, \qquad P(R) = \frac{1}{(2\pi)^{6/2}} \left(\frac{6}{Nb^2} \right)^{3/2} e^{-\frac{3}{2Nb^2}|R|^2}$$

As stated in the previous pages P(R) is the probability for the end to end vector of the polymer chain to have a value of R. Likewise, Z_0 is the total number of conformations for the polymer chain when the constraint P(R) is not taken into account.

Below are the calculations for Z_0 , P(R) and $S_0(R)$.

$$Z_0 = \mathcal{V}\left(\frac{2\pi b^2}{3}\right)^{3N/2}, \qquad P(R) = \left(\frac{3}{2\pi N b^2}\right)^{3/2} e^{-\frac{3}{2Nb^2}|R|^2}, \qquad S_0(R) = \ln\left[\frac{Z_0}{\mathcal{V}} \times P(R)\right]$$

This in turn gives us the following calculations for the conformational entropy. The term C(N) is a constant that is given in terms of the number of amino acids. *b* is the average length of the sub chain. The number of sub chains is given by the term *N*. As stated previously *R* is the end to end vector of the polymer chain. k_B is Boltzmann's constant and $S_0(R)$ is the classical conformational entropy.

$$S_0(R) = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \right], \qquad S_0(R) = ln \left[\left(\frac{2\pi b^2}{3} \right)^{3N/2} \left(\frac{3}{2\pi N b^2} \right)^{3/2} e^{-\frac{3}{2Nb^2} |R|^2} \right]$$
$$S_0 = -\frac{3k_B}{2Nb^2} |R|^2 + C(N)$$

Hence the length of the end to end vector *R* decreases the conformational entropy as expected. While the bead spring model appears relatively simple it has its limits. First of all, it only applies when the number of sub chains is large. More importantly however it cannot be used to incorporate the effects due to Hydrogen bonds. This is due to the fact that the entropy is described in terms of the vectors corresponding to the amino acid segments instead of the vectors corresponding to the peptide bonds, otherwise known as the bond vectors. The lattice model however does describe the conformational entropy in terms of the bond vectors. This is the model that will be discussed in the following pages.

The Lattice Model

Since the Perturbation term corresponding to the Hydrogen bonds is in terms of the bond vectors the bead spring model, which treats the sub chains as single units, is insufficient for taking into account the quantum mechanical effects due to hydrogen bonds. Among these effects are the breaking of hydrogen bonds due to quantum tunneling. In order to incorporate quantum mechanical effects, a different model will be used. This model is referred to as the lattice model and it is equivalent to the bead spring model when the number of amino acids is large. Below is the formula for P(R) where P(R) is the probability for the end to end vector to have a value of R.

$$P(R) = \int du_0 \dots \int du_{N-1} \,\delta\left(R - \sum_{i=0}^{N-1} u_i\right) P_N(\{u_i\})$$

Likewise, $P_N(\{u_i\})$ is the probability of finding a certain chain conformation. u_i is the bond vector for the *ith* peptide bond, where the peptide bonds are the bonds between each amino acid in the polymer chain. The equation P(R) is equal to the product of the probabilities for each of the N bond vectors b to take on a certain value. These bond vectors correspond to the peptide bonds between the amino acids. Hence the formula for $P_N(\{u_i\})$ is given below.

$$P_N(\{u_i\}) = \prod_{j=0}^{N-1} P_1(u_j)$$

In the equation above $P_1(u_j)$ is the likelihood for the bond vector of the *j*th amino acid to take on a specific value. Since it is assumed that the peptide units do not interact the probabilities associated with each value of the *j*th bond vector are all equal, where the peptide units are the amino acids. When the number of amino acids is sufficiently large the probabilities for the directions of the bond vectors are all equal. Since there are six total directions that the bond vector *a* can take on the value of $P_1(u_j)$ is equal to $\frac{1}{6}$. The set corresponding to these bond vectors is given by $b = \{\pm b_x, \pm b_y, \pm b_z\}$. Within the set $b = \{\pm b_x, \pm b_y, \pm b_z\}$, b_x, b_y, b_z and $-b_x, -b_y, -b_z$ are the positive and negative values of the bond vector. The bond vectors are the directions for the peptide bonds between each peptide unit. As mentioned earlier each peptide unit is an amino acid in the Cas9 polymer chain.

Since the calculations for P(R) are computationally intensive, the Dirac Delta term within P(R) for the lattice model must be rewritten as an integral. Since the Dirac Delta equation within P(R) is given in three dimensions, the integral formula for the Dirac Delta must also be written in three dimensions. The integral formulation for the Dirac Delta function is given below. The formula for P(R) is also given below for convenience.

$$\delta(r) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot r}, \qquad P(R) = \int du_0 \dots \int du_{N-1} \,\delta\left(R - \sum_{i=0}^{N-1} u_i\right) P_N(\{u_i\})$$

In the equation for P(R) above, the term $P_N(\{u_i\})$ is the product of the probabilities for each bond vector *b* to take on a specific value out of the set of the bond vectors $b = \{\pm b_x, \pm b_y, \pm b_z\}$. The equation for $P_N(\{u_i\})$ is given below where $P_N(\{u_i\})$ is the joint probability distribution for the *N* bonds.

$$P_N(\{u_i\}) = P_1(u_0)P_1(u_1) \dots P_1(u_{N-1})$$

Likewise the terms $P_1(u_0)$, $P_1(u_1) \dots P_1(u_{N-1})$ are the probability distributions for a single bond corresponding to the bond vectors u_0, u_1 and u_{N-1} .

Since the polymer in the lattice model is treated as an ideal chain, the probabilities for each bond vector to take on a specific value are all equal to each other, where there are a total of six bond vectors. This in turn leads to the following equation for $P_1(u)$ where $P_1(u)$ is the probability distribution for any of the bond vectors.

$$P_{1}(u) = \begin{cases} \frac{1}{6} & (u \in nearest neighbor vectors) \\ 0 & (otherwise) \end{cases}$$

Since the cas9 polymer is going to be treated as an ideal chain, the terms

 $P_1(u_0), P_1(u_1) \dots P_1(u_{N-1})$ are all equal to $P_1(u)$, whereas stated previously $P_1(u)$ is the probability distribution for any of the bond vectors. This in turn leads to the following formula for $P_N(\{u_i\})$ below.

$$P_N(\{u_i\}) = \left(P_1(u)\right)^N$$

This in turn makes it possible to rewrite the formula for P(R) as the following equation.

$$P(R) = \frac{1}{(2\pi)^3} \int dq \, e^{iqR} \prod_{j=0}^{N-1} \int du_j e^{-iqu_j} P_1(u_j)$$

The formulas for $P_1(u_j)$ and $P_N(\{u_i\})$ are also given below for convenience along with $\delta(r)$.

$$P(R) = \int du_0 \dots \int du_{N-1} \,\delta\left(R - \sum_{i=0}^{N-1} u_i\right) P_N(\{u_i\}), \qquad P_1(u_j) = P_1(u)$$

$$P(R) = \frac{1}{(2\pi)^3} \int dq \, e^{iqR} \prod_{j=0}^{N-1} \int du_j e^{-iqu_j} P_1(u_j)$$

$$\delta(r) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} dq e^{iq \cdot r}$$

The equation for P(R) above is really just the inverse Fourier transforms of many Fourier transforms. The equations corresponding to these Fourier transforms are given below. $\tilde{P}_1(q)$ is the Fourier transform for a single probability distribution $P_1(u)$. Likewise, $[\tilde{P}_1(q)]^N$ is the product of N such transforms.

$$\tilde{P}_1(q) = \int du P_1(u) e^{-iq \cdot u}$$
, $[\tilde{P}_1(q)]^N = \prod_{j=0}^{N-1} \int du_j e^{-iqu_j} P_1(u_j)$

This leads to the following equation for P(R) given below.

$$P(R) = \frac{1}{(2\pi)^3} \int dq \, e^{iqR} [\tilde{P}_1(q)]^N$$

The formulas for $[\tilde{P}_1(q)]^N$ and $\tilde{P}_1(q)$ are both given below for convenience. The computations for P(R) will be continued in the following pages.

$$\tilde{P}_1(q) = \int du P_1(u) e^{-iq \cdot u}$$
, $[\tilde{P}_1(q)]^N = \prod_{j=0}^{N-1} \int du_j e^{-iqu_j} P_1(u_j)$

As was stated in the previous page the formula for P(R) is given in the following form.

$$P(R) = \frac{1}{(2\pi)^3} \int dq \, e^{iqR} \prod_{j=0}^{N-1} \int du_j e^{-iqu_j} P_1(u_j)$$

Next the equation for $P_1(u)$ will be computed along with the formulas for $\tilde{P}_1(q)$. As stated previously the formula for $P_1(u)$ is given by the following equation.

$$P_{1}(u) = \begin{cases} \frac{1}{z} & (u \in nearest neighbor vectors) \\ 0 & (otherwise) \end{cases}$$

This in turn can be rewritten as the following series where a sums through the set of the nearest neighboring vectors b.

$$\tilde{P}_1(u) = \frac{1}{6} \sum_a \delta(u-a), \qquad b = \left\{ \pm b_x, \pm b_y, \pm b_z \right\}$$

In order to streamline the calculations for $[\tilde{P}_1(q)]^N$, approximation techniques will need to be employed. The approximations that follow in these pages hold so long as $N \gg 1$.

In order to make use of such approximations, the following series for e^x will be deployed for the purposes of calculating P(R). The calculations for $\tilde{P}_1(q)$ are also given beneath the series for e^x below and will be continued in the following pages.

$$e^{x} = \sum_{n=0}^{\infty} \frac{x^{n}}{n!} = 1 + x + \frac{x^{2}}{2!} + \frac{x^{3}}{3!} + \cdots$$
$$\tilde{P}_{1}(q) = \int du \left[\frac{1}{6} \sum_{a} \delta(u-a) \right] e^{-iq \cdot u}$$

The equations for $\tilde{P}_1(q)$ that were stated in the bottom of the previous page are restated below for convenience.

$$\tilde{P}_1(q) = \int du P_1(u) e^{-iq \cdot u}, \qquad \tilde{P}_1(q) = \int du \left[\frac{1}{6} \sum_a \delta(u-a)\right] e^{-iq \cdot u}$$

These integrals in turn lead to the following equation for the Fourier transform $\tilde{P}_1(q)$. As stated previously the function $\tilde{P}_1(q)$ is the Fourier transform of the probability distribution $P_1(u)$ where $P_1(u)$ is the probability distribution for a single bond vector a.

$$\tilde{P}_1(q) = \frac{1}{6} \sum_a e^{-iq \cdot a}$$

Since it is assumed that $N \gg 1$, $[\tilde{P}_1(q)]^N$ can be reduced to the result $e^{-\frac{Nb_0^2|q|^2}{6}}$ with the use of the following approximation scheme given below.

$$\begin{split} [\tilde{P}_1(q)]^N &= \left[\frac{1}{6} \sum_a 1 - iq \cdot a - \frac{1}{2}(q \cdot a)^2 + \cdots\right]^N \\ &\sim \left[\frac{1}{6} \left(6 - \frac{1}{2} \left\{2q_x^2 b_0^2 + 2q_y^2 b_0^2 + 2q_z^2 b_0^2\right\}\right)\right]^N \sim \left(1 - \frac{b_0^2}{6} \left(q_x^2 b_0^2 + q_y^2 b_0^2 + q_z^2 b_0^2\right)\right) \\ &\sim \left[1 - \frac{b_0^2 |q|^2}{6}\right]^N \sim e^{-\frac{N b_0^2 |q|^2}{6}} \end{split}$$

Hence $[\tilde{P}_1(q)]^N$ is given by the following equation for $N \gg 1$.

$$[\tilde{P}_1(q)]^N = e^{-\frac{Nb_0^2|q|^2}{6}}$$

Just like in the previous case for the lattice model the Fourier Transform of a Gaussian will be employed to streamline calculations. The formulas corresponding to this transform are given below. The calculations for P(R) will be carried out in the next page. As stated previously P(R) is the probability for the end to end vector to have a value R.

$$\int dr e^{-a|r-c|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a} - iq \cdot c}, \qquad \int dr e^{-a|r|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a}}, \qquad \int dr e^{-a|r|^2} e^{iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a}}, \qquad \int dr e^{-a|r|^2} e^{iq \cdot r} = \left(\frac{\pi}{a}\right)^{3/2} e^{-\frac{|q|^2}{4a}},$$

The calculations for the term P(R) are given below.

$$P(R) = \frac{1}{(2\pi)^3} \int dq e^{iq \cdot R} e^{-\frac{Nb_0^2 |q|^2}{6}} = \frac{1}{(2\pi)^3} \left(\frac{6\pi}{Nb_0^2}\right)^{3/2} e^{-\frac{6|q|^2}{4Nb_0^2}}$$
$$P(R) = \frac{1}{(2\pi)^3} \int dq e^{iq \cdot R} e^{-\frac{Nb_0^2 |q|^2}{6}} = \frac{1}{(2\pi)^{6/2}} \left(\frac{6\pi}{Nb_0^2}\right)^{3/2} e^{-\frac{6|q|^2}{4Nb_0^2}}$$
$$P(R) = \left(\frac{3}{2\pi Nb_0^2}\right)^{\frac{3}{2}} e^{-\frac{3|R|^2}{2Nb_0^2}}$$

While both the bead spring model and the lattice model have their uses, they are not without limits. The partition function Z_0 in the bead spring model is not written in terms of the position vectors of the amino acids. Since the hydrogen bonds are interactions between these amino acids any model that incorporates hydrogen bonds must be in terms of these position vectors. While The lattice model is written in terms of these position vectors it fails to account for the energies due to the peptide bonds. Since the quantum mechanical interaction is a sum of the energies of the hydrogen bonds, the lattice model is not sufficient either. Another model must be obtained for incorporating the quantum mechanical perturbation. In order to construct such a model, the strengths of both the bead spring model as well as the lattice model must be combined into a single model. This model will be written in terms of the position vectors of the amino acids; however, it will also account for the potential energy of the peptide bonds. In order to achieve such a model, the partition function Z_0 will be rewritten in terms of the position vectors of the amino acids $r'_0, ..., r'_N$ as well as the potential energy for the peptide bonds \mathcal{H}_0' . Below are the equations for position vectors of the amino acids u_i , i = 0, ..., N - 1as well as the partition function Z_0 . The Hamiltonian \mathcal{H}_0 is also given below.

$$Z_{0} = \int dr_{0} \dots \int dr_{N} e^{-\beta \mathcal{H}_{0}(r_{0},\dots,r_{N})}, \qquad \mathcal{H}_{0} = -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2}$$
$$u_{i} = r_{i+1}' - r_{i}'$$

In order to convert from one model to the other a coordinate transformation must be performed on the integral Z_0 . This will be done in the following page. The coordinate transformation from the coordinates $r_0 \dots r_N$ to the coordinates $u_0, \dots u_{N-1}$ is given below.

$$dr_0 \dots dr_N = \left| \frac{\partial(r_0, \dots, r_N)}{\partial(r_0, u_0, \dots, u_{N-1})} \right| dr_0 du_0 \dots du_{N-1} = dr_0 du_0 \dots du_{N-1}$$

In order to convert the bead spring model to the lattice model the Hamiltonian \mathcal{H}_0 must be rewritten in terms of the bond vectors u_0, \ldots, u_{N-1} . This is due to the fact that the Hamiltonian \mathcal{H}_0 is in terms of the position vectors for the ends of the sub chains. The calculations below correspond to the conversions of \mathcal{H}_0 from the bead spring model to the lattice model where the indices *i'* and *N'* are dummy indices. As stated previously *R* is the end to end vector. These calculations are continued on the next page.

$$\begin{aligned} \mathcal{H}_{0} &= -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i} - r_{i+1}|^{2} = -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i+1} - r_{i}|^{2} = -\frac{3}{2b^{2}} |r_{N} - r_{0}|^{2} \\ &- \frac{3}{2b^{2}} |r_{N} - r_{0}|^{2} = -\frac{3}{2b^{2}} |r_{N}' - r_{0}'|^{2}, \qquad -\frac{3}{2b^{2}} |r_{N}' - r_{0}'|^{2} = -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i+1}' - r_{i}'|^{2} \\ \mathcal{H}_{0} &= -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i+1}' - r_{i}'|^{2}, \qquad \mathcal{H}_{0} = -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |u_{i}|^{2} \\ \mathcal{H}_{0} &= -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |-u_{i}|^{2}, \qquad \mathcal{H}_{0} = -\frac{3}{2b^{2}} \sum_{i=0}^{N-1} |r_{i}' - r_{i+1}'|^{2} \end{aligned}$$

Next the partition function Z_0 is rewritten in terms of the position vectors of the amino acids $r'_0, ..., r'_N$. The following Jacobian is given below along with the following equation for u_i .

$$u_i = r'_{i+1} - r'_i, \qquad \left| \frac{\partial(r_0, u_0, \dots, u_{N-1})}{\partial(r'_0, r'_1, \dots, r'_N)} \right| = 1$$

This in turn leads to the following equations for Z_0 and \mathcal{H}_0 where Z_0 and \mathcal{H}_0 are rewritten in terms of the position vectors for the amino acids r'_0, r'_1, \dots, r_N .

$$Z_0 = \int dr_0 \dots \int dr_N e^{-\beta \mathcal{H}_0(r_0,\dots,r_N)}, \qquad \mathcal{H}_0 = -\frac{3}{2b^2} \sum_{i=0}^{N-1} |r_i' - r_{i+1}'|^2$$

Hence when the number of sub chains is large the two models have the same formalism. The main difference is that one model sums over sub chains while the other sums over the position vectors of the amino acids. Since the number of amino acids is very large for each protein it is necessary to break it down into sub structures in order to understand it, this is especially the case for Cas9 which contains 1368 amino acids. These substructures are referred to as the primary structure, secondary structure and primary structure respectively. These structures will be discussed in the following pages below.

Protein Structure

The primary structure is the sequence of the amino acids in the polymer chain. The secondary structures are referred to as beta strands and alpha helices respectively. The tertiary structure is the structure of an entire conformation.

Every amino acid can be broken up into two sections. These sections are the backbone and the sidechain. Another name for the sidechain is the R group. The backbone of every amino acid is identical. The sidechains, however, are different for every amino acid.

There are three different cases in which hydrogen bonds can form. In the first case, hydrogen bonds form between the backbones of two different amino acids. This is possible for each and every pair of amino acids. In the second case hydrogen bonds form between the sidechains of two different amino acids. In the third case side chains form hydrogen bonds with a solvent. In each of these cases certain requirements need to be met in order for a hydrogen bond to form.

In order for two backbones to form a hydrogen bond there are two conditions that must be met. In the first condition the vectors pertaining to the side chains of each amino acid must be parallel to each other. In the second condition, the sequence of amino acids containing the two amino acids that are adjacent to each other must be in a strand state. This will be explained in detail later.

While the hydrogen bonds pertaining to backbones can occur between each amino acid, this is not the case for sidechains. Only certain combinations of sidechains are capable of forming hydrogen bonds. Even when two such sidechains are in contact it is not guaranteed that a hydrogen bond will form. Depending on the form of the sidechain, a hydrogen bond may or may not form. If the sidechain only comes in one form, then the probability is equal to one. The same cases hold if the form of the sidechain is irrelevant. If the form of the side chain of an amino acid determines whether or not a hydrogen bond will form, then the probability that a hydrogen bond will form is less than one. So long as the number of amino acids is sufficiently large these probability terms hold.

19

Just like for the sidechain interactions there are only certain amino acids that can interact with solvents. In addition, the only component of an amino acid that can interact with the solvent is the side chain. The sidechains must also be pointing towards an empty lattice site within the lattice model. In order to streamline calculations, the system will be constrained so that the only solvent interacting with the side chains will be water.

In a nutshell there are three different forms of hydrogen bonds. These are referred to as backbone interactions, side chain interactions and solvent interactions. The backbone interactions occur between the backbones of the amino acids. The side chain interactions occur between sidechains of the amino acids. Likewise, the solvent interactions occur between the side chains of the amino acids and a corresponding solvent. In this case study the solvent will be water. These terms together make up a correction that will be referred to as the quantum mechanical perturbation and it accounts for the effects due to the hydrogen bonds. Among these effects are the effects due to quantum tunneling. These corrections will be discussed in the following pages. Then the quantum mechanical correction computed.

CHAPTER II:

QUANTUM MECHANICAL PERTURBATION

Quantum Mechanical Interactions

The term that accounts for the quantum mechanical interactions is given by G_{QM} , where G_{QM} is the Gibbs free energy corresponding to the hydrogen bonds. This includes both the potential energy of the peptide bonds as well as the total energy pertaining to the hydrogen bonds. There are three different forms of hydrogen bonds that will be discussed in more detail later on. These are the hydrogen bonds due to the backbone interactions, sidechain interactions and solvent interactions. In addition to these interactions are steric interactions which will be explained in detail later. The equation for the Gibbs free energy G_{QM} is given below.

$$G_{QM} = b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + g \cdot \Delta G_{kon}$$
$$+ h \cdot T \Delta S_{mc} + k \cdot T \Delta S_{sc} + a \cdot \Delta G_{vdw} + f \cdot \Delta G_{el} + l \cdot \Delta G_{clash}$$

In the equation above $\Delta G_{steric} = a \cdot \Delta G_{vdw} + f \cdot \Delta G_{el} + l \cdot \Delta G_{clash}$ is the contribution to the free energy due to steric effects. $a \cdot \Delta G_{vdw}$ is the Gibbs free energy due to the Van der Waals forces, $f \cdot \Delta G_{el}$ is the Gibbs free energy due to the electrostatic interactions and $l \cdot \Delta G_{clash}$ is the Gibbs free energy due to steric clashes. The equation G_{QM} is rewritten below in terms of the substitution $\Delta G_{steric} = a \cdot \Delta G_{vdw} + f \cdot \Delta G_{el} + l \cdot \Delta G_{clash}$ for convenience.

$$G_{QM} = b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + g \cdot \Delta G_{kon}$$
$$+ h \cdot T \Delta S_{mc} + k \cdot T \Delta S_{sc} + \Delta G_{steric}$$

The terms $h \cdot T\Delta S_{mc}$ and $k \cdot T\Delta S_{sc}$ are the contributions to the free energy due to the hydrogen bonds between the backbones of the amino acids as well as the sidechains of the amino acids.

The substitutions $T\Delta S_{bb} = h \cdot T\Delta S_{mc}$, and $T\Delta S_{sc} = k \cdot T\Delta S_{sc}$ are included below in the term corresponding to the free energy G_{QM} for convenience.

$$G_{QM} = \Delta G_{steric} + b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + g \cdot \Delta G_{kon} + T \Delta S_{bb} + T \Delta S_{sc}$$

In the equation above $T\Delta S_{bb}$ is the entropic penalty due to backbone interactions. $T\Delta S_{sc}$ likewise, is the entropic penalty due to the sidechain interactions, where the constant k has been absorbed into the equation for S_{sc} .

The terms $b \cdot \Delta G_{solvH}$ and $c \cdot \Delta G_{solvP}$ together account for the interactions between the amino acid side chains as well as the solvent. This in turn gives us the following equation below, where $\Delta G_{solvent} = b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP}$. In order to constrain the model even further it is assumed that the only amino acids that interact with water are the polar amino acids. Therefore $b \cdot \Delta G_{solvH} = 0$. The modifications for G_{QM} are given below.

$$G_{QM} = \Delta G_{steric} + \Delta G_{solvent} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + g \cdot \Delta G_{kon} + T \Delta S_{bb} + T \Delta S_{sc}$$

Since the term $g \cdot \Delta G_{kon}$ is only applicable for interactions between different polypeptide chains it is set to 0. The term G_{QM} can then be rewritten in the following form below.

$$G_{QM} = \Delta G_{steric} + \Delta G_{solvent} + d \cdot \Delta G_{wb}$$
$$+ e \cdot \Delta G_{hbond} + T \Delta S_{bb} + T \Delta S_{sc}$$

The model will now be further constrained so as to assume that the water molecules form up to one bond each with the cas9 protein. The term that accounts for multiple hydrogen bonds (see the paper The FoldX web server: an online force field) with water is given by $d \cdot \Delta G_{wb}$. Since the model is constrained so as to assume that the water molecules form up to one bond each with the cas9 protein the term $d \cdot \Delta G_{wb}$ is set to zero such that $d \cdot \Delta G_{wb} = 0$. This leads to the following equation below.

$$G_{QM} = \Delta G_{steric} + \Delta G_{solvent}$$
$$+ e \cdot \Delta G_{hbond} + T \Delta S_{bb} + T \Delta S_{sc}$$

The model will be further constrained so as to assume that the amino acids each only form one hydrogen bond with each other. The term that accounts for such interaction is given by ΔG_{hbond} . ΔG_{hbond} is therefore set to zero. Therefore $\Delta G_{hbond} = 0$. This leads to the following equation below for ΔG_{QM} . As stated previously ΔG_{QM} is the difference in the free energy due to the quantum mechanical interactions.

$$\Delta G_{QM} = \Delta G_{steric} + \Delta G_{solvent} + h \cdot T \Delta S_{bb} + k \cdot T \Delta S_{sc}$$

In the equation for ΔG_{QM} above, ΔG_{steric} is the free energy due to steric hinderance, $\Delta G_{solvent}$ is the free energy due to the solvent interactions and $T\Delta S_{bb}$ and $T\Delta S_{sc}$ are the free energy penalties due to the backbone-backbone and sidechainsidechain interactions respectively. In the following pages the quantum mechanical perturbation ΔS_{QM} will be obtained from ΔG_{QM} .

This leads to the following equations below for ΔG_{QM} and ΔS_{QM} .

$$\begin{split} \Delta G_{QM} &= \Delta G_{steric} + \Delta G_{solvent} + T\Delta S_{bb} + T\Delta S_{sc}, \qquad \Delta G_{QM} = -T\Delta S_{QM} \\ \Delta S_{QM} &= -\frac{1}{T} \Delta G_{steric} - \frac{1}{T} \Delta G_{solvent} - \Delta S_{bb} - \Delta S_{sc}, \qquad -\frac{1}{T} \Delta G_{solvent} = \Delta S_{solvent} \\ \Delta S_{QM} &= \Delta S_{steric} + \Delta S_{solvent} - \Delta S_{bb} - \Delta S_{sc}, \qquad -\frac{1}{T} \Delta G_{steric} = \Delta H_{steric} - \Delta S_{steric} \\ \text{The calculations for the total conformational entropy } S_{total} are given below. \\ S_{total} &= S_0 + S_{QM}, \qquad S_{total} = S_0 - S_{bb} - S_{sc} + S_{water} + S_{steric}, \\ S_{amino} &= S_{bb} + S_{sc}, \qquad S_{total} = S_0 - S_{amino} + S_{water} + S_{steric}, \\ G_{amino} &= H_{bb} - TS_{sc}, \qquad G_{amino} = -TS_{amino} \\ -\frac{G_{amino}}{T} &= -\frac{H_{bb}}{T} + S_{sc}, \qquad G_{amino} = -k_B T ln(Q_{amino}), \qquad ln(Q_{amino}) = -\beta G_{amino} \\ Q_{amino} &= e^{-\beta G_{amino}}, \qquad Q_{amino} = e^{-\beta H_{bb} + \beta T S_{sc}}, \qquad TS_{amino} = -G_{amino} \\ S_{total}(R) &= k_B ln \left[\frac{Z_0}{\mathcal{V}} \times P(R)\right] - k_B ln(Q_{amino}) + S_{water} + S_{steric}, \\ S_{total}(R) &= k_B ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \frac{1}{Q_{amino}}\right] + S_{water} + S_{steric} \end{split}$$

As stated previously the formula for the total conformational entropy S_{total} is given by the equation below.

$$S_{total} = S_0 - S_{bb} - S_{sc} + S_{water} + S_{Steric}$$

In the term S_{total} above, S_{bb} is the entropy corresponding to the backbone penalty and S_{sc} is the entropy corresponding to the sidechain penalty. The sum $S_{amino} = S_{bb} + S_{sc}$ is a function of the partition function Q_{amino} . The entropy S_{amino} is derived from the free energy equation G_{amino} , and G_{amino} is the free energy corresponding to the backbone and sidechain interactions. Likewise, H_{bb} is the potential energy corresponding to the backbone conformation. As we will see later, when quantum mechanical effects are taken into account a kinetic energy term KE is added to the potential energy H_{bb} such that $H_{bb} \rightarrow KE + H_{bb}$. In the following pages the equations for S_{sc} , S_{bb} , $S_{solvent}$ and S_{steric} will be calculated in order to derive the formula for the total entropy $S_{total} =$ $S_{protein} + S_{water}$.

Below are the calculations for the side chain conformational entropy S_{sc} .

$$S_{sc} = -k_B \sum_{i} p_i lnp_i, \qquad p_i = \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}}, \qquad i = 0, \dots, 1367$$

$$S_{sc} = -k_B \sum_i \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} ln\left(\frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}}\right)$$

$$S_{sc} = -k_B \sum_i \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} ln\left(\frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}}\right)$$

$$S_{sc} = -k_B \sum_i \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} ln\left(e^{-\beta E_{sc,i}}\right) + k_B \sum_i \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} ln\left(\sum_i e^{-\beta E_{sc,i}}\right)$$

$$S_{sc} = -\frac{1}{T} \sum_i \frac{E_{sc,i} e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} + k_B \sum_i \frac{e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} ln\left(\sum_i e^{-\beta E_{sc,i}}\right)$$

$$S_{sc} = -\frac{1}{T} \sum_i \frac{E_{sc,i} e^{-\beta E_{sc,i}}}{\sum_i e^{-\beta E_{sc,i}}} + k_B ln\left(\sum_i e^{-\beta E_{sc,i}}\right)$$

In the calculations above the term $E_{sc,i}$ is the hydrogen bonding energy corresponding to the *i*th amino acid.

When the $E_{sc,i}$'s are all very small this equation reduces to the following approximation below.

$$S_{sc} \cong -\frac{1}{T} \frac{\sum_{i} E_{sc,i} e^{-\beta E_{sc,i}}}{\sum_{i} e^{-\beta E_{sc,i}}} = -\frac{1}{T} \overline{E}_{sc}$$

Assuming that $E_{sc,i}$ is small the entropy S_{sc} reduces to the following equation below. The formula for the total energy due to the sidechain interactions \overline{E}_{sc} is also given below.

$$S_{sc} = -\frac{1}{T} \sum_{i} E_{sc,i} , \qquad \overline{E}_{sc} = \frac{1}{T} \sum_{i} E_{sc,i}$$

This in turn leads to the following equation for S_{sc} below.

$$S_{sc} = -\frac{1}{T}\bar{E}_{sc}$$

The free energy $G_{amino} = H_{bb} - TS_{sc}$ where $G_{sc} = -TS_{sc}$ can then be rewritten in the following form below.

$$G_{amino} = H_{bb} + \bar{E}_{sc}$$

In the equation for G_{amino} above, S_{sc} and $p_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$ both sum over all of the 1367 amino acids of the cas9 protein, S_{water} . The calculations pertaining to the S_{water} term will be given in the following pages.

As stated previously the solvent interaction term S_{water} is constrainted so as to assume that the only solvent that interacts with the amino acids is water. In the calculations below, $\Delta G_{a \rightarrow b}$ is the change in the Gibbs free energy due to the water interactions. Likewise, Q_b and Q_a are the partition functions corresponding to the terms G_b and G_a respectively. The Hamiltonians for the partition functions Q_b and Q_a are given by both H_b and H_a . The calculations for the entropy contribution for the solvent interactions S_{water} are given below. These calculations will be continued in the following pages.

$$\Delta G_{a \to b} = G_b - G_a, \qquad \Delta G_{a \to b} = -k_B T ln\left(\frac{Q_b}{Q_a}\right), \qquad \Delta G_{a \to b} = -k_B T ln\left(e^{-(H_b - H_a)/k_B T}\right)$$
$$\Delta G_{a \to b} = -k_B T ln\left(e^{-\frac{H_b}{k_B T}}e^{\frac{H_a}{k_B T}}\right) = -k_B T ln\left(e^{-\frac{H_b}{k_B T}}/e^{-\frac{H_a}{k_B T}}\right)$$

The calculations corresponding to the free energy are continued below.

$$\Delta G_{a \to b} = -k_B T \left[ln \left(e^{-\frac{H_b}{k_B T}} \right) + ln \left(1/e^{-\frac{H_a}{k_B T}} \right) \right]$$
$$\Delta G_{a \to b} = -k_B T \left[ln \left(e^{-\frac{H_b}{k_B T}} \right) - ln \left(e^{-\frac{H_a}{k_B T}} \right) \right]$$
$$G_b = -k_B T ln \left(e^{-\frac{H_b}{k_B T}} \right), \qquad G_a = -k_B T ln \left(e^{-\frac{H_a}{k_B T}} \right)$$

Where $\Delta G_{a \to b}$ is the change in the binding energy between a given molecule and amino acid. This is the energy required to break a bond with a water molecule. The formulas pertaining to the entropy due to the hydrogen bonds with water are given below.

$$S_{water} = -\frac{G_{water}}{T}$$
, $G_{water} = H_{water}$, $S_{water} = -\frac{H_{water}}{T}$

In the equations above, H_{water} is the Hamiltonian corresponding to the hydrogen bonds between the sidechains and water.

The partition function Q_{water} is given by the following equation below, where Q_{water} is the partition function corresponding to the energies between the sidechains and water.

$$Q_{water} = e^{-\frac{H_{water}}{k_B T}}$$

This in turn gives leads to the following calculations for the total conformational entropy S_{total} below, which will be continued in the following page.

$$S_{total} = S_0(R) + k_B ln(Q_{amino}) + k_B ln(Q_{water}) + S_{steric}$$

$$S_{total} = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \right] - k_B ln(Q_{amino}) + k_B ln(Q_{water}) + S_{steric}$$

$$S_{total} = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \right] - k_B ln \left(\frac{Q_{water}}{Q_{amino}} \right) + S_{steric}$$

$$S_{total} = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) e^{\beta E_{bb} + \beta E_{sc} - \beta E_{water}} \right] + S_{steric}$$

$$S_{total} = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) e^{\beta E_{bb} + \beta E_{sc} - \beta E_{water}} \right] + S_{steric}$$

For the purposes of this paper the value of S_{steric} will be set to some unknown constant C(N). This in turn leads to the following equation for S_{total} .

$$S_{total} = ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \frac{e^{-\beta E_{water}}}{e^{E_{bb} - E_{sc}}} \right] + C(N).$$

Where E_{water} is the quantum mechanical energy due to the interactions with water. E_{bb} is the correction due the interactions between the backbones and E_{sc} is the correction due to the interactions between the sidechains. Only certain sidechains are allowed to interact with each other and are different for every amino acid. The backbones however are the same for every amino acid and the backbone interactions can take place between any amino acid pair. Likewise, for the water interactions only some of the amino acids are capable of interacting with water. In order to streamline calculations, the system will be constrained so that the only solvent that will be interacting with the amino acids will be water. The first term that will be discussed will be the energy due to the backbone interactions E_{bb} . Then the energy corresponding to the side chain interactions E_{sc} will be evaluated. Afterwards the energy corresponding to the interactions between the amino acids with water E_{water} will be computed. Then the entropies for the cases corresponding to the quantum mechanical corrections with and without tunneling will be computed and the extent effects will be determined on the total conformational entropy of the Cas9 protein. In the following pages the energy due to the backbone interactions E_{bb} will be computed and its overall influence on the conformational entropy of Cas9 determined.

Backbone-Backbone Interactions

The Hamiltonian corresponding to the interactions between the amino acids is given below.

$$\mathcal{H}_{amino} = E_{bb} + E_{so}$$

Where E_{bb} is the energy corresponding to the hydrogen bonds between the backbones and E_{sc} is the energy corresponding to the hydrogen bonds between the sidechains. Each term is a linear combination of tensor elements. These tensor elements determine whether or not a hydrogen bond will form based upon the factors that where discussed earlier. The following series was obtained from the research paper "A Simple Model that Captures Protein Folding, Aggregation and Amyloid Formation".

$$E_{bb} = \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} \epsilon_{hb} H_{ij} C_{ij}$$

In the equation above, $H_{ij} = 1$ if the *i*th and *j*th amino acids have formed a hydrogen bond, otherwise $H_{ij} = 0$. Likewise, the term $C_{ij} = 1$ if, the *i*th and *j*th amino acids are in contact, and the *i*th and *j*th amino acids are not adjacent to each other on the polymer chain. If on the other hand the *i*th and *j*th amino acids are not in contact, then $C_{ij} = 0$. Both of these matrix elements are given below.

$$H_{ij} = \begin{cases} 1 \text{ if } s_i, s_j = \text{strand and } \hat{d}_i = \hat{d}_j \\ 0 \text{ otherwise} \end{cases}$$
$$C_{ij} = \begin{cases} 1 \text{ if } |\vec{p}_i - \vec{p}_j| = 1 \text{ and } |i - j| > 1 \\ 0 \text{ otherwise} \end{cases}$$

Within the definitions for the matrix elements H_{ij} and C_{ij} above, the terms \hat{d}_i and \hat{d}_j are the vectors for the ith and *j*th side chain. Likewise, \vec{p}_i and \vec{p}_j are the position vectors for the *i*th and *j*th amino acids. The terms s_i and s_j are elements within the set $A_{sec,s} = \{strand, coil\}$ such that $s_i, s_j \in \{strand, coil\}$, where $A_{sec,s}$ is the set of secondary structures. The elements *strand* and *coil* are the beta strands and alpha helices respectively. Likewise, the terms s_i and s_j are the states containing the *i*th and *j*th amino acid respectively. Each state corresponds to a secondary structure. These secondary structures are the beta strand and the alpha helix respectively.


Figure 3: The structure of an Alpha helix

Likewise, a diagram for the structure of a beta strand is given below.



Figure 4: The structure of a Beta strand

While the model for the energy E_{bb} is simple it is limited. It is only useful when one can take into account the directions in which the side chains are pointing. This is computationally intensive. In order to streamline calculations an approximation will be employed that allows for the directions \hat{d}_i , \hat{d}_j of the bond vectors to be ignored. So long as the number of amino acids is sufficiently large this approximation holds, as the directions of the bond vectors which are assumed to be random cancel each other out.

The Energy corresponding to the hydrogen bonds between the backbones is rewritten below as the series E_{bb} . The terms B_{ij} , X^{ij} and Y_{ij} are also displayed below for convenience and will be discussed in detail in the following paragraph.

$$\begin{split} E_{bb} &= \frac{1}{2} \sum_{i=1}^{N-2} \sum_{j=1}^{N-2} A^{ij} B_{ij} X^{ij} Y_{ij} \overline{E}^{0x} \\ B_{ij} &= \frac{1}{2} \delta_{r_{i+1} - r_{j+1}, u}, \qquad A^{ij} = \delta_{r_{i+2} - r_{j+0}, u} \\ X^{ij} &= \delta_{r_i - r_j, u}, \qquad Y_{ij} = \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ 1 \ otherwise \end{cases} \end{split}$$

The term $A^{ij}B_{ij}$ accounts for whether or not the sub chain containing the *i*th and *j*th amino acid is in a strand state. Together with the fraction $\frac{1}{2}$ these terms replace the matrix element H_{ij} given in the previous page. Likewise the terms Y_{ij} and $\delta_{r_i-r_j,u}$, both take the place of the matrix element C_{ij} . In order for the sub chain containing the *i*th and *j*th amino acid to be in a strand state there must be at least two or three consecutive pairs of amino acids that are in contact. Below are diagrams corresponding to the term $A^{ij}B_{ij}X^{ij}$.

$$\begin{aligned} r_{i+0} & \longrightarrow r_{i+1} \longrightarrow r_{i+2} \cdots \searrow, \qquad r_{j+0} \longrightarrow r_{j+1} \longrightarrow r_{i+2} \cdots \searrow \\ r_{j+2} & \leftarrow r_{j+1} \leftarrow r_{j+0} \cdots \swarrow, \qquad r_{i+2} \longrightarrow r_{i+1} \longrightarrow r_{i+0} \cdots \checkmark \end{aligned}$$

In both of these diagrams, the position vectors r_{i+0} and r_{j+2} are both adjacent to each other. The position vectors r_{i+1} and r_{j+1} are adjacent to each other. The same case holds for the position vectors r_{i+2} and r_{i+0} which are also adjacent to each other. This is the structure of a beta strand.

The equation for the term B_{ij} is stated below for convenience.

$$B_{ij} = \frac{1}{2} \delta_{r_{i+1} - r_{j+1}, u}$$

Within the term B_{ij} above the fraction accounts for the likelihood that the bond vectors will be parallel to each other and oriented in the same direction. If the sub chain is in a strand state then $A^{ij}B_{ij}X^{ij}$ is equal to one, otherwise it is equal to zero.

Within the term $A^{ij}B_{ij}X^{ij}$, $A^{ij} = 1$ if $r_{i+2} - r_{j+0} = u$. Likewise if $r_{i+2} - r_{j+0} \neq u$ then $A^{ij} = 0$. This leads to the following relation below.

$$A^{ij} = \begin{cases} 1 \ if \ r_{i+2} - r_{j+0} = u \\ 0 \ if \ r_{i+2} - r_{j+0} \neq u \end{cases}$$

Just like the matrix element A^{ij} , the term $B_{ij} = \frac{1}{2}$ if $r_{i+1} - r_{j+1} = u$. Likewise if $r_{i+1} - r_{j+1} \neq u$ then $B_{ij} = 0$. This leads to the following relation below for B_{ij} .

$$B_{ij} = \begin{cases} 1 \ if \ r_{i+1} - r_{j+1} = u \\ 0 \ if \ r_{i+2} - r_{j+0} \neq u \end{cases}$$

Within the term $A^{ij}B_{ij}X^{ij}$ the matrix element $X^{ij} = 1$ if $r_i - r_j = u$. Likewise if $r_i - r_j \neq u$ then $X^{ij} = 0$. This leads to the following relation below.

$$X^{ij} = \begin{cases} 1 \text{ if } r_i - r_j = u \\ 0 \text{ if } r_i - r_j \neq u \end{cases}$$

In order to account for double counting a term $\frac{1}{2}$ was introduced in the equation for E_{bb} . The equation E_{bb} is restated below for convenience.

$$E_{bb} = \frac{1}{2} \sum_{i=1}^{N-2} \sum_{j=1}^{N-2} A^{ij} B_{ij} X^{ij} Y_{ij} \overline{E}^{0x}$$

As stated previously the matrix element Y_{ij} is given by the following relation below.

$$Y_{ij} = \begin{cases} 0 \ if \ i = j+1 \ or \ j = i+1 \\ 1 \ otherwise \end{cases}$$

The term $Y_{ij} = 0$ when the *i*th and *j*th amino acids are adjacent to each other on the polymer chain, otherwise the term $Y_{ij} = 1$. The term Y_{ij} was introduced in the summation E_{bb} because amino acids that are adjacent to each other on the polymer chain do not interact with each other. Hence even if $A^{ij}B_{ij}X^{ij}$ is equal to $\frac{1}{2}$ a hydrogen bond may not

be present between the *i*th and *j*th amino acids. This is precisely the case for $Y_{ij} = 0$. This corresponds to the case where the *i*th and *j*th amino acids are adjacent on the polymer chain.

If all of these conditions mentioned above are met, then the likelihood of a hydrogen bond forming is equal to 1/2. For a large number of amino acids, the energy for a single hydrogen bond can be set to the average of all of the eigenstates. These eigenstates correspond to the different possible energies of the O—H bond between the ith and *j*th backbone. The average energy for these eigenstates is given by the equation for \overline{E}^{Ox} below.

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}$$

Where $O^{ij} = 1$ if the hydrogen bond pertaining to the *i*th and *j*th amino acid contains an Oxygen atom and equals zero otherwise. Likewise, $N^{ij} = 1$ if the hydrogen bond pertaining to the *i*th and *j*th amino acid contains a Nitrogen atom and equals zero otherwise.

In summary there are two different types of interactions between the amino acids. These are referred to as backbone interactions and sidechain interactions. In the case of backbone interactions, the sidechains corresponding to the pair of amino acids in a hydrogen bond must be pointing in the same direction. The term that accounts for this is given by the fraction 1/2. Also, the sequences of amino acids containing the hydrogen bond must be in the form of a beta strand. The term that accounts for this is given by $A^{ij}B_{ij}$. Likewise the term that accounts for whether or not the *i*th and *j*th amino acids are adjacent to each other is given by X^{ij} . In order to ensure that the *i*th an *j*th amino acids are not adjacent to each other in the polymer chain the term Y_{ij} is included in the summation E_{bb} , where E_{bb} is the equation for the sum of the backbone interactions. A similar term holds for the side chain interactions and is given by the following equation below.

$$E_{sc} = \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} C_{ij} K_{ij} M_{a_i a_j}$$

This equation will be discussed in the following pages and the energy E_{sc} will be evaluated.

Sidechain-Sidechain Interactions

The Energy corresponding to the hydrogen bonds between the sidechains is given by the following equation.

$$E_{sc} = \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} C_{ij} K_{ij} M_{a_i a_j}$$

The term C_{ij} determines for whether or not the *i*th and *j*th amino acid are in contact. The matrix K_{ij} determines whether or not the *i*th and *j*th amino acid are facing each other or parallel to each other. The matrix elements C_{ij} and K_{ij} are both given below.

$$C_{ij} = \begin{cases} 1 \text{ if } \left| \vec{p}_i - \vec{p}_j \right| = 1 \text{ and } |i - j| > 1 \\ 0 \text{ otherwise} \end{cases}$$

$$K_{ij} = \begin{cases} 1 \text{ if } \hat{d}_i = -\hat{d}_j \text{ and } |(\vec{p}_i + \hat{d}_i) - (\vec{p}_j - \hat{d}_j)| = 1\\ 1 \text{ if } \hat{d}_i = \hat{d}_j \text{ and } |(\vec{p}_i + \hat{d}_i) - (\vec{p}_j - \hat{d}_j)| = 1\\ 0 \text{ otherwise} \end{cases}$$

In the equations above, term K_{ij} accounts for whether or not the sidechains are pointing in the same direction or are parallel to each other. If the *i*th and *j*th side chains are parallel to each other then $\hat{d}_i = \hat{d}_j$ and $|(\vec{p}_i + \hat{d}_i) - (\vec{p}_j - \hat{d}_j)| = 1$ and the term $K_{ij} = 1$. Likewise if the *i*th and *j*th side chains are pointing towards each other then $\hat{d}_i = \hat{d}_j$ and $|(\vec{p}_i + \hat{d}_i) - (\vec{p}_j - \hat{d}_j)| = 1$ and $K_{ij} = 1$.

Just like the term C_{ij} within the summation for E_{bb} the matrix element C_{ij} within E_{sc} accounts for whether or not the *i*th and *j*th amino acids are adjacent to each other on the amino acid chain. If this is the case, then $C_{ij} = 0$. Otherwise $|\vec{p}_i - \vec{p}_j| = 1$ and |i - j| > 1 thereby setting $C_{ij} = 1$.

Lastly, the matrix element $M_{a_i a_j}$ corresponds to the binding energies between the *i*th and *j*th amino acid.

Since this model accounts for the directions of the bond vectors for each of the *i*th and *j*th amino acids, this model is insufficient for the purposes of this study which ignore such effects. In order to create a less computationally intensive model these effects must be ignored. This approximation holds so long as the number of amino acids is large. The equation corresponding E_{sc} corresponding to this approximation where the directions of the amino acids are ignored is given below.

$$E_{sc} = \frac{1}{2} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} \mathcal{H}^{ij}$$

Where the terms R_{ij} , Y_{ij} and \mathcal{H}^{ij} are given below. M_{ij} is the probability that a hydrogen bond will form between the *i*th and *j*th amino acid. The matrices \mathcal{H}^{ij} and M^{ij} were both calculated in the python interpreter. The matrix elements corresponding to the summation E_{sc} are given below. These matrices will be discussed in detail later.

$$\begin{aligned} R_{ij} &= \frac{3}{16} \delta_{r_i - r_j, u}, \qquad Y_{ij} = \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ 1 \ otherwise \end{cases} \\ \mathcal{H}^{ij} &= O^{ij} \overline{E}^{ox} + N^{ij} \overline{E}^{Ni} \end{aligned}$$

In the equation for the term \mathcal{H}^{ij} above, O^{ij} is equal to one if an O—H bond can form between the *i*th and *j*th amino acid and is equal to zero otherwise. Likewise, the term N^{ij} is equal to one if an N—H bond can form between the *i*th and *j*th amino acid and is equal to zero otherwise. The fraction $\frac{1}{2}$ accounts for double counting. The equations for \overline{E}^{Ox} and \overline{E}^{Ni} are both given below.

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}, \qquad \bar{E}^{Ni} = \frac{\sum_{n=0}^{1} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{1} e^{-\beta E_n^{Ni}}}$$

In the relations above the term \overline{E}^{Ox} is the average value for the O—H bonds and the term \overline{E}^{Ni} is the average value for the N—H bonds. So long as the number of hydrogen bonds is sufficiently high these approximations hold. There are three eigenvalues corresponding to the O—H bond and two eigenvalues corresponding to the N—H bond. The term \overline{E}^{Ox} sums through the energy eigenstates for the O—H bond, meanwhile the term \overline{E}^{Ni} sums through the energy eigenstates for the N—H bond. These eigenstates will be discussed later in the section "Energy Eigenstates".

The term R_{ij} is equal to $\frac{3}{16}$ when the amino acids are one-unit vector apart and is equal to 0 otherwise. This is due to the fact that there are three different combinations for the directions of the adjacent side-chain vectors to arrange themselves in order for the hydrogen bond to form. These correspond to the two cases where the side chains of the ith and *j*th amino acids are parallel along with the case where the side chains are pointing towards each other. There are 16 total different combinations of arrangements for these side chain vectors assuming they are adjacent to each other, where the side chain vectors are the vectors corresponding to the side chains discussed previously. Likewise, M^{ij} is the likelihood that a hydrogen bond will form between the *i*th and *j*th amino acid. This is due to the fact that just because two sidechains are in contact does not mean a hydrogen bond will form between them. This is due to the following reasons. First of all, in order for a hydrogen bond to form between two sidechains the hydrogen atom in one of the sidechains must be in contact with either the nitrogen atom or the oxygen atom in the other sidechain. In some cases, only the hydrogen atom in the *i*th sidechain can form a hydrogen bond with the nitrogen or oxygen atom within the *j*th side chain. There are other cases in which only the nitrogen atom or oxygen atom in the *j*th chain can form a hydrogen bond with the hydrogen atom within the *i*th side chain. In order to account for these facts, probabilities are assigned to each of the elements within the M^{ij} matrix element given in the summation E_{sc} . Also, there are cases where the amino acids need to come in specific forms in order to form hydrogen bonds at all. The matrix element M^{ij} accounts for this fact as well. So long as the number of amino acids is large this approximation holds. Likewise, the term $\frac{3}{16}$ accounts for the probability that the side chains will either be facing each other or be parallel to each other. For more information consult the tables in the appendix.

In summary the hydrogen bonds between the side chains must meet certain requirements in order to form. First of all, the amino acids must either be pointing towards each other or be parallel to each other. Second of all, the amino acids in question must be capable of forming hydrogen bonds. Third of all, if necessary, the amino acid(s) that are in contact must be in the right form, the matrix that accounts for this is given by M_{ij} . The matrix M^{ij} accounts for whether or not the *i*th and *j*th amino acid can form a hydrogen bond. If they can then the matrix is equal to the probability that that hydrogen bond will form for the reasons discussed earlier. The matrix R_{ij} is equal to $\frac{3}{16}$ when the amino acids are in contact and is equal to 0 otherwise. The fraction $\frac{3}{16}$ is the probability that the side chains corresponding to the *i*th and *j*th amino acid will either be parallel or facing each other. Likewise, the matrix \mathcal{H}^{ij} is equal to the hydrogen bonding energy between the *i*th and *j*th amino acids. Lastly the series E_{sc} is the total energy corresponding to the hydrogen bonds between the sidechains. I order to account for all of the interactions due to hydrogen bonds, the interactions between the amino acids with water must be computed along with the backbone interactions and side chain interactions. Just like with the side chain interactions there are only certain side chains that can interact with water. This term will be computed in the following pages.

Solvent Interactions

The $E_{solvent}$ term accounts for the interactions of the side chains of the amino acids with water. These amino acids are referred to as hydrophilic amino acids. There are three different types of hydrophilic amino acids each corresponding to three different types of R groups. These are positively charged side chains, negatively charged sidechains and side chains with alcohol groups. These groups of amino acids correspond to the sets E_+ , E_- and E_β and are displayed below.

$$E_{\beta} = \{E_{Ser}, E_{Thr}\}, \qquad E_{+} = \{E_{Lys}, E_{Arg}, E_{His}\}, \qquad E_{-} = \{E_{Asp}, E_{Glu}\}$$
$$E_{PU} = \{E_{Asn}, E_{Gln}\}$$

Where E_+ and E_- are sets containing the energies for the positively and negatively charged R groups. Likewise, the term E_{β} is the set containing the amino acids with beta-hydroxyl groups. Lastly the term E_{PU} is the set containing the energies of the amino acids with polar uncharged groups. Since Serine and Threonine contain Hydrogen, Carbon and Oxygen atoms in their sidechains the only hydrogen bond that can form between them and water is the O—H bond. Hence the energy E_{β} is given by the following equation.

$$E_{\beta} = \frac{\sum_{n=0}^{2} E_{n}^{Ox} e^{-\beta E_{n}^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_{n}^{Ox}}}$$

Likewise, Lysine, Arginine and Histidine are positively charged. Therefore, the Hydrogen atoms and only the hydrogen atoms on these sidechains can interact with water. Hence the energy corresponding to the set E_+ is given by the following equation.

$$E_{+} = \frac{\sum_{n=0}^{2} E_{n}^{Ox} e^{-\beta E_{n}^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_{n}^{Ox}}}$$

Since Aspartate and Glutamate are negatively charged, only the Oxygen atoms are going to interact with water. Hence the energy for E_{Asp} are E_{Glu} corresponds to the O—H bond and only the O—H bond.

$$E_{-} = \frac{\sum_{n=0}^{2} E_{n}^{Ox} e^{-\beta E_{n}^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_{n}^{Ox}}}$$

The calculations for these terms will be continued in the following pages along with the energy $E_{solvent}$ corresponding to the solvent interactions

Asparagine and Glutamine both contain a N—H dipole and an O—H dipole. Both of these dipoles can interact with water. Because of this, the energy for E_{PU} must be set to the likelihood that an N—H bond will form plus the likelihood that an O—H bond will form. This is the average energy corresponding to the amino acids E_{Asn} and E_{Gln} respectively.

The equation corresponding to this energy is given below

$$E_{PU} = \frac{\sum_{n=0}^{1} E_{l}^{PU} e^{-\beta E_{l}^{PU}}}{\sum_{n=0}^{1} e^{-\beta E_{l}^{PU}}}$$

In the term for E_{PU} above the energies E_0^{PU} and E_1^{PU} are given by the following equations below.

$$\bar{E}_{0}^{PU} = \frac{\sum_{n=0}^{2} E_{n}^{Ox} e^{-\beta E_{n}^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_{n}^{Ox}}}, \qquad \bar{E}_{1}^{PU} = \frac{\sum_{n=0}^{2} E_{n}^{Ni} e^{-\beta E_{n}^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_{n}^{Ni}}}$$

These relations are the same as those for the energies E^{0x} and E^{Ni} . These equations are restated below for convenience.

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}, \qquad \bar{E}^{Ni} = \frac{\sum_{n=0}^{2} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ni}}}$$

The equation pertaining to the energy $E_{solvent}$ is given below. The calculations for this term will be completed in the following page.

$$E_{solvent} = \frac{1}{2} \sum_{solv}^{N_{solv}} \sum_{i}^{N} M_{a_i, solv} \cdot K_{i, solv}$$

In the summation $E_{solvent}$ above, the term $M_{a_i,solv}$ is the energy between the solvent and the *i*th amino acid a_i . Likewise, the term $K_{i,solv}$ indicates whether or not an interaction between the solvent and the *i*th amino acid a_i occurs.

The equation for the matrix element $K_{i,solv}$ is given below.

$$K_{i,solv} = \begin{cases} 1 \ if \ \left(\vec{p}_i + \hat{d}_i\right) = \vec{p}_{solv} \\ 0 \ otherwise \end{cases}$$

In the matrix element $K_{i,solv}$ above the term \vec{p}_i is the position vector for the *i*th amino acid. Likewise, \hat{d}_i is the vector for the sidechain of the *i*th amino acid.

$$E_{solvent} = \sum_{i=0}^{N-1} E_i - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} W^{ij} E_i$$

In the equation above, E_i is the energy for the hydrogen bonds with water. W^{ij} accounts for how many amino acids are touching the *i*th amino acid. These terms along with the usual Y_{ij} term are given below.

$$\begin{split} W^{ij} &= \frac{1}{4} \delta_{r_i - r_j, u}, \qquad E_i = \begin{cases} \overline{E}^{Ox} \ if \ E_i \in E_\beta \cup E_+ \cup E_- \\ \overline{E}^{Hb} \ if \ E_i \in E_{PU} \\ 0 \ otherwise \end{cases} \\ Y_{ij} &= \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ E \ otherwise \end{cases}, \end{split}$$

In the set of equations above the term $W^{ij} = \frac{1}{4}$ if the *i*th and *j*th amino acids are in contact otherwise $W^{ij} = 0$. This accounts for how many empty lattice sites are adjacent to the *i*th amino acid. If there are 4 amino acids adjacent to the *i*th amino acid, then the likelihood that a hydrogen bond will form is zero. If there are 3 adjacent amino acids, then the likelihood that a hydrogen bond will form is equal to 1/4. Likewise, if there are 2 adjacent amino acids then the likelihood that a hydrogen bond will form is equal to 2/4. If there are 3 adjacent amino acids, then the likelihood that a hydrogen bond will form is equal to 3/4. Lastly if there are 4 adjacent amino acids, then the likelihood that a hydrogen bond will form is equal to 4/4. In summary there are four different sets of amino acids that can form hydrogen bonds. These sets are the positively charged amino acids E_+ , negatively charged amino acids E_- , polar uncharged amino acids E_{PU} as well as the set of amino acids containing beta hydroxyl groups E_β . In order for a hydrogen bond to form between the *i*th amino acid and water, there must be an empty lattice site adjacent to *i*th amino acid. This is accounted for by the term W^{ij} . The term W^{ij} accounts for the probability that the sidechain will be facing an empty lattice site. So long as the number of amino acids is large this approximation holds. Lastly the term $E_{solvent}$ is the total energy for the hydrogen bonds between the amino acid side chain and water, where water is the solvent of choice for this system.

You may have noticed that there were terms \overline{E}^{Ox} and \overline{E}^{Ni} that were given in E_{bb}, E_{sc} and $E_{solvnet}$ when computing the quantum tunneling correction. The terms \overline{E}^{Ox} and \overline{E}^{Ni} are displayed below for convenience.

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}, \qquad \bar{E}^{Ni} = \frac{\sum_{n=0}^{2} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ni}}}$$

These terms correspond to the average energies of the eigenstates. In order to make sense of the terms \overline{E}^{Ox} and \overline{E}^{Ni} it is necessary to compute the energies of these eigenstates. These eigenstates correspond to the different possible energies that the hydrogen bonds can take on. These energy eigenstates are given by integrals and these integrals are written in terms of a potential referred to as the Morse potential. The Morse potential accounts for the breaking of the hydrogen bonds. However, it does not account for the effects due to quantum tunneling. Another set of equations must be employed for these purposes. This will be discussed in the following pages. Afterwards the quantum mechanical perturbation will be added to the term given by $S_0(R)$. This term was discussed in the previous sections of this text and it corresponds to the classical conformational entropy due to the peptide bonds. These corrections will be discussed in the following section "Energy Eigenstates".

Energy Eigenstates

In order to evaluate the energy corresponding to the hydrogen bonds it is necessary to evaluate the eigenstates pertaining to them. These are referred to as the energy eigenstates. The equation for the nth eigenstate of a hydrogen bond is given by the equation E_n below.

$$E_n = \frac{\int_{ymin}^{ymax} \Psi_n^* \widehat{H}_M \Psi_n dy}{\int_{ymin}^{ymax} \Psi_n^* \Psi_n dy}$$

Within the term E_n above, Ψ_n and Ψ_n^* are the wavefunction and conjugate wavefunction and \hat{H}_M is the Hamiltonian for the hydrogen bond. The Hamiltonian \hat{H}_M for the hydrogen bonds is given below. This is the Hamiltonian corresponding to the Morse potential which accounts for the breaking of the hydrogen bonds.

$$\hat{H}_{M} = -\frac{\beta^{2}\hbar^{2}}{2\mu}\frac{\partial^{2}}{\partial y^{2}} + D_{e}\left(e^{-2(y-y_{eq})} - 2e^{-(y-y_{eq})}\right)$$

In the equation for \hat{H}_M above, D_e is the disassociation energy, \hbar is the planck constant, $y_{eq} = \beta x_{eq}$ where x_{eq} is the internuclear equilibrium distance and μ is the reduced mass of the system. The internuclear equilibrium distance x_{eq} is the distance between the nuclei of the two atoms engaged in a hydrogen bond. The term β is an arbitrary constant related to the width of the potential well. Below is the wavefunction $\Psi(y)_n^{(1)}$ for the nth state.

$$\Psi(y)_{n}^{(1)} \cong \frac{1}{\sqrt{V_{n+1}(y) - \epsilon_{0}^{(n+1)}}} e^{\int \sqrt{V_{n+1}(y) - \epsilon_{0}^{(n+1)}} dy}$$

In the term for $\Psi(y)_n^{(1)}$ above, $V_{n+1}(y)$ and $\epsilon_0^{(n+1)}$ are the potential and total energies for the nth excited state. The equations for these formulas are given below.

$$V_{n+1}(y) = \Lambda^2 \left(e^{-2(y - y_{eq})} - 2e^{-(y - y_{eq})} \right) + 2n\Lambda e^{-(y - y_{eq})}$$
$$\epsilon_0^{(n+1)} = -\left(\Lambda - \frac{2n+1}{2}\right)^2$$

In the following pages the terms $P_{total}(\{r_i\})$ and $P_{total}(R)$ will finally be computed and their corresponding values determined.

. Below is the equation $P_{total}(\{r_i\})$ for the probability of finding a certain conformation. This term takes into account the quantum mechanical corrections E_{bb} , E_{sc} and $E_{solvent}$.

$$P_{total}(\{r_i\}) = \frac{1}{Z_0 e^{-\beta E_{bb} - \beta E_{sc}}} e^{-\beta \mathcal{H}_0(r_0, \dots, r_N) - \beta E_{solvent}}$$

Likewise the equation for $P_{total}(R)$ is given by the following integral below.

$$P_{total}(R) = \int dr_0 \dots \int dr_N \delta(r_N - r_0 - R) \frac{1}{Z_0} e^{-\beta \mathcal{H}_0}$$

Just like in the previous case for $P_{total}(R)$ in order to compute $P_{total}(R)$ it is necessary to deploy the equations for the Dirac delta and the Fourier transform. The familiar equations for the Dirac delta as well as the Fourier transform for a Gaussian are displayed below for convenience.

$$\int_{-\infty}^{+\infty} dq e^{iq \cdot x} = 2\pi\delta(x), \qquad \int dr e^{-a|r-b|^2} e^{-iq \cdot r} = \left(\frac{\pi}{a}\right)^{\frac{3}{2}} e^{-\left(\frac{|q|^2}{4a} + iq \cdot b\right)}$$

Whereupon the terms Z_0 and $P_{total}(R)$ are combined to produce the following calculations for the total conformational entropy $S_{total}(R)$. As stated previously C(N) is an arbitrary constant in terms of N.

$$S_{total}(R) = k_B ln \left[\frac{Z_0}{\mathcal{V}} \times P_{total}(R) \right]$$

$$S_{total}(R) = k_B ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \frac{e^{-\beta \mathcal{H}_0(r_0, \dots, r_N) - \beta E_{solvent}}}{e^{-\beta E_{bb} - \beta E_{sc}}} \right]$$

$$S_{total}(R) = k_B ln \left[\frac{Z_0}{\mathcal{V}} \times P(R) \right] - \beta k_B E_{solvent} + \beta k_B E_{bb} + \beta k_B E_{sc} + C(N)$$

$$S_{total}(R) = (0.00057922750 \ kgs^{-2}K^{-1}) |R|^2 - \frac{E_{solvent}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T} + C(N)$$

In a summary the hydrogen bonds do in fact influence the conformational entropy of the Cas9 protein. In the following pages the effects due to quantum tunneling will be computed along with the effects due to the hydrogen bonds. The effects due to quantum tunneling are a further approximation on the effects due to the hydrogen bonds. Both interactions are quantum mechanical in nature and the tunneling correction is a correction to the effects due to the hydrogen bonds. Both of these interactions will be calculated in the following pages for a specific case in which the Cas9 protein is in the form of a giant beta strand.

CHAPTER III:

COMPARISON OF THE CLASSICAL AND SEMI-CLASSICAL ENTROPIES

Conformational Entropy Correction

The equations for the total conformational entropy $S_{total}(R)$ are given below. The constants for *b* and k_B are also displayed.

$$S_{total}(R) = S_0(R) - \frac{E_{solvent}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T} + S_{steric} + C(N)$$

$$S_{total}(R) = k_B ln \left[\frac{Z_0}{V} \times P(R) \right] - \frac{E_{solvent}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T} + S_{steric} + C(N)$$

$$b = 1.32 \cdot 10^{-10} m, \qquad k_B = 1.38064852 \cdot 10^{-23} J \cdot K^{-1}$$

This leads to the following formula below.

$$\begin{split} S_{total}(R) &= (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 \\ &- \frac{E_{solvent}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T} + S_{steric} + C(N) \\ S_{total}(R) &= 1.0092377196 \cdot 10^{-23} \\ &- \frac{E_{solvent}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T} + S_{steric} + C(N) \end{split}$$

The terms E_{bb} , E_{sc} and $E_{solvent}$ are the energies of the hydrogen bonds due to backbone interactions, sidechain interactions and solvent interactions. E_{Steric} is the steric energy penalty. For the purposes of this paper water will be the only solvent that interacts with the Cas9 protein. The steric energy accounts for those cases where the sidechains of adjacent amino acids on the sidechain clash. For the purposes of this paper the steric term S_{steric} will be absorbed into the constant C(N). The following pages contain the corrections due to both the hydrogen bonds as well as the tunneling corrections. These terms are given by the equations $\mathcal{H}_{qm_{-}M}$ and $\mathcal{H}_{qm_{-}T}$. Where $\mathcal{H}_{qm_{-}M}$ is the correction due to the Morse potential and $\mathcal{H}_{qm_{-}T}$ is the correction due to quantum tunneling. These terms will be computed in the following pages.

Tunneling Corrections

Below are the equations for the energy corresponding to the backbone hydrogen bonds.

$$\begin{split} E_{bb} &= \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} A^{ij} B_{ij} X^{ij} Y_{ij} \, \bar{E}^{0x} \\ B_{ij} &= \frac{1}{2} \, \delta_{r_{i+1} - r_{j-1}, u}, \qquad A^{ij} = \delta_{r_{i-1} - r_{j+1}, u} \\ X^{ij} &= \delta_{r_i - r_j, u}, \qquad Y_{ij} = \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ E \ otherwise \end{cases}, \end{split}$$

Where the equation for \overline{E}^{Ox} is given below. The average energies of the hydrogen bonds for the O—H and N—H pairs are given at the end of this study.

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{1} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}$$

In the equation for \overline{E}^{Ox} above E_n^{Ox} is the equation for the nth energy eigenstate for the O—H bond. The calculations for the term E_{bb}/T are given below where $\frac{E_{bb}}{T} = S_{bb}$ is the entropic penalty for the backbone interactions. These values are given below along with the value for \overline{E}^{Ox} .

$$\frac{E_{bb}}{T} = \frac{1}{2T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \delta_{r_i-r_j,u} Y_{ij} \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}$$
$$\bar{E}^{Ox} = -6.66101 \times 10^{-19} J$$

This leads to the following summations for $\frac{E_{bb}}{T}$ below.

$$\begin{aligned} \frac{E_{bb}}{T} &= (-6.66101 \times 10^{-19} J) \frac{1}{2T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \, \delta_{r_i-r_{j},u} Y_{ij} \\ \frac{E_{bb}}{T} &= -1.66525 \times 10^{-19} J \frac{1}{T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \, \delta_{r_i-r_{j},u} Y_{ij} \, O^{ij} \\ \frac{E_{bb}}{T} &= -5.58810 \times 10^{-22} J \, K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \, \delta_{r_i-r_{j},u} Y_{ij} \, O^{ij} \end{aligned}$$

The following pages contain the calculations for the side chain contribution E_{sc} .

Below is the summation for the energy corresponding to the sidechain hydrogen bonds.

$$E_{sc} = \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} \mathcal{H}^{ij}$$
$$R_{ij} = \frac{3}{16} \delta_{r_i - r_j, u}, \qquad S_{ij} = \frac{3}{16} \delta_{r_i - r_j, u}, \qquad Y_{ij} = \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ 1 \ otherwise \end{cases}$$

Whereupon substituting the equation for \mathcal{H}^{ij} one obtains the following equation for $\frac{E_{sc}}{T}$ where $\frac{E_{sc}}{T} = S_{sc}$ is the entropic penalty for the side chain interactions.

$$\frac{E_{sc}}{T} = \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij} \overline{E}_n^{Ox} + \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij} \overline{E}_n^{Ni}$$

This in turn leads to the following summations below. The values for \overline{E}_n^{Ox} and \overline{E}_n^{Ni} are also given beneath the summations for $\frac{E_{sc}}{T}$

$$\begin{split} \frac{E_{sc}}{T} &= \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij} \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}} \\ &+ \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij} \frac{\sum_{n=0}^{2} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ni}}} \\ &= \frac{E_{sc}}{T} = \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij} \overline{E}_n^{Ox} \\ &+ \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij} \overline{E}_n^{Ni} \\ &= -6.66101 \times 10^{-19} J, \qquad \overline{E}^{Ni} = -5.21342 \times 10^{-19} J \end{split}$$

The calculations for $\frac{E_{sc}}{T}$ will be continued in the pages below.

Given the values for \overline{E}^{Ox} and \overline{E}^{Ni} in the previous page the equation for $\frac{E_{sc}}{T}$ is equal to the following summation below.

$$\frac{E_{sc}}{T} = -6.66101 \times 10^{-19} J \frac{1}{2} \frac{3}{16T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij}$$
$$+ -5.21342 \times 10^{-19} J \frac{1}{2} \frac{3}{16T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij}$$

The values for \overline{E}^{Ox} and \overline{E}^{Ni} are restated below for convenience.

$$\bar{E}^{0x} = -6.66101 \times 10^{-19} J, \qquad \bar{E}^{Ni} = -5.21342 \times 10^{-19} J$$

With the computation of basic arithmetic, the above equation $\frac{E_{sc}}{T}$ above is distilled to the following summation below.

$$\frac{E_{sc}}{T} = -2.09554 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij}$$
$$-1.64013 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij}$$

The calculations for the energy E_{water} will be carried out in the following pages. The calculations corresponding to the E_{water} term are given below along with the matrices E_i , Y_{ij} and W^{ij} .

$$\frac{E_{water}}{T} = \sum_{i=0}^{N-1} E_i - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} W^{ij} E_i$$

= j + 1 or j = i + 1 $W^{ij} = \frac{1}{2} \sum_{i=0}^{N-1} E_i = \overline{E} O_i$

$$Y_{ij} = \begin{cases} 0 \ if \ i = j+1 \ or \ j = i+1 \\ E \ otherwise \end{cases}, \qquad W^{ij} = \frac{1}{4} \delta_{r_i - r_j, u}, \qquad E_i = \overline{E}^{Ox} O_i + \overline{E}^{Hb} H_i$$

The calculations for $E_{solvent}$ will be continued on the

This leads to the following expansion for the summation $E_{solvent}$ given above. The values for \overline{E}^{Ox} and \overline{E}^{Hb} are also given below.

$$\frac{E_{water}}{T} = \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Ox} O_i + \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Hb} H_i - \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Ox} Y_{ij} W^{ij} O_i - \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Hb} Y_{ij} W^{ij} H_i$$
$$\bar{E}^{Ox} = -6.66101 \times 10^{-19} J, \qquad \bar{E}^{Hb} = -5.21342 \times 10^{-19} J$$

Given the values for \overline{E}^{Ox} and \overline{E}^{Ni} above the equation for $\frac{E_{sc}}{T}$ is equal to the following summation below.

$$\frac{E_{solvent}}{T} = -6.66101 \times 10^{-19} J \frac{1}{T} \sum_{i=0}^{N-1} O_i - 5.21342 \times 10^{-19} J \frac{1}{T} \sum_{i=0}^{N-1} H_i$$
$$-6.66101 \times 10^{-19} J \frac{1}{4} \frac{1}{T} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i - r_j, u} O_i - 5.21342 \times 10^{-19} J \frac{1}{4} \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Hb} Y_{ij} \delta_{r_i - r_j, u} H_i$$

This in turn leads to the following quantum mechanical perturbation term for $\frac{E_{solvent}}{T}$ below, whereas stated previously T = 298 K.

$$\frac{E_{solvent}}{T} = -2.2354 \times 10^{-21} \frac{J}{K} \sum_{i=0}^{N-1} O_i - 1.74947 \times 10^{-21} \frac{J}{K} \sum_{i=0}^{N-1} H_i$$

-5.58810 × 10⁻²² $\frac{J}{K} \sum_{i=0}^{N-1} Y_{ij} W^{ij} O_i - 4.37367 \times 10^{-22} \frac{J}{K} \sum_{i=0}^{N-1} \bar{E}^{Hb} Y_{ij} W^{ij} H_i$

These calculations will be continued on the next page.

Summing up all of the terms E_{bb} , E_{sc} and $E_{solvent}$ such that $\mathcal{H}_{QM} = E_{bb} + E_{sc} + E_{solvent}$ leads to the following summation for \mathcal{H}_{qm_T} below.

$$\begin{aligned} \mathcal{H}_{qm_{-}T} &= -5.58810 \times 10^{-22} J \, K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \, \delta_{r_{i}-r_{j},u} Y_{ij} \, O^{ij} \\ &- 2.09554 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i}-r_{j},u} M^{ij} Y_{ij} O^{ij} \\ &- 1.64013 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i}-r_{j},u} M^{ij} Y_{ij} N^{ij} \\ &(-2.2354 \times 10^{-21} J) \sum_{i=0}^{N-1} O_{i} - (1.74947 \times 10^{-21} J) \sum_{i=0}^{N-1} H_{i} \\ &- (5.58810 \times 10^{-22} J) \sum_{i=0}^{N-1} Y_{ij} W^{ij} O_{i} - (4.37367 \times 10^{-22} J) \sum_{i=0}^{N-1} \overline{E}^{Hb} Y_{ij} W^{ij} H_{i} \end{aligned}$$

Where the equation above \mathcal{H}_{qm_T} corresponds to the tunneling correction for the energies of the hydrogen bonds within the Cas9 protein. As stated previously the term corresponding the Morse potential is given by \mathcal{H}_{qm_M} and its energies are lower than those corresponding to the correction \mathcal{H}_{qm_T} . The perturbation \mathcal{H}_{qm_M} corresponding to the Morse potential will be discussed in the next pages.

Corrections due to the Morse Potential

Below are the calculations for all of the different possible energies corresponding to the E_{bb} term for the quantum corrections that ignore tunneling. Just like in the case for tunneling it is assumed that T = 298 K

The equation corresponding to the energy of the backbone hydrogen bonds is given below. The equation for \overline{E}^{Ox} is also given below where \overline{E}^{Ox} is the average energy for the O—H bonds. This equation holds so long as the number of amino acids is large.

$$E_{bb} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij} \bar{E}^{0x}$$

$$B_{ij} = \frac{1}{2} \delta_{r_{i+1}-r_{j-1},u}, \qquad A^{ij} = \delta_{r_{i-1}-r_{j+1},u}$$

$$X^{ij} = \delta_{r_i-r_j,u}, \quad Y_{ij} = \begin{cases} 0 \ if \ i = j+1 \ or \ j = i+1 \\ E \ otherwise \end{cases}$$

$$\bar{E}^{0x} = \frac{\sum_{n=0}^{1} E_n^{0x} e^{-\beta E_n^{0x}}}{\sum_{n=0}^{2} e^{-\beta E_n^{0x}}}$$

In the set of equations above, \overline{E}^{ox} is the average energy eigenvalue corresponding to the nth energy state. Its value is given at the end of this study.

Below are the calculations for the Hydrogen bonding energies where the equation for \overline{E}^{Ox} has been substituted into $\frac{E_{bb}}{T}$, where as stated previously $\frac{E_{bb}}{T}$ is the entropic penalty for the backbone interactions. The value for the constant \overline{E}^{Ox} is also given below.

$$\frac{E_{bb}}{T} = \frac{1}{2T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \,\delta_{r_i-r_j,u} Y_{ij} \,O^{ij} \frac{\sum_{n=0}^2 E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^2 e^{-\beta E_n^{Ox}}}$$
$$\bar{E}^{Ox} = -6.66651 \times 10^{-19} J$$

This leads to the following summation for $\frac{E_{bb}}{T}$.

$$\frac{E_{bb}}{T} = -6.66651 \times 10^{-19} J \frac{1}{2T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \,\delta_{r_i-r_j,u} Y_{ij}$$

These calculations will be continued in the following pages below.

After the computation of basic arithmetic, the following summation is obtained for $\frac{E_{bb}}{T}$.

$$\frac{E_{bb}}{T} = -5.59271 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \,\delta_{r_i-r_j,u} Y_{ij}$$

Below are the calculations for all of the different possible energies corresponding to the E_{sc} term for the case where the effects due to tunneling are ignored. The equation for \mathcal{H}^{ij} is also displayed below.

$$E_{sc} = \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} \mathcal{H}^{ij}$$

$$R_{ij} = \frac{3}{16} \delta_{r_i - r_j, u}, \qquad S_{ij} = \frac{3}{16} \delta_{r_i - r_j, u}, \qquad Y_{ij} = \begin{cases} 0 \ if \ i = j + 1 \ or \ j = i + 1 \\ 1 \ otherwise \end{cases}$$

$$\mathcal{H}^{ij} = O^{ij} \overline{E}_n^{ox} + N^{ij} \overline{E}_n^{Ni}$$

Next the term E_{sc} is expanded by substituting the value for the matrix element \mathcal{H}^{ij} into $\frac{E_{sc}}{T}$ where $\frac{E_{sc}}{T}$ is the entropic penalty for the side chain interactions.

$$\frac{E_{sc}}{T} = \frac{3}{16} \frac{1}{2T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij} \overline{E}^{Ox} + \frac{3}{16} \frac{1}{2T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij} \overline{E}^{Ni}$$

In the following pages the term $\frac{E_{sc}}{T}$ will be expanded even further to include the equations for \bar{E}_n^{Ox} and \bar{E}_n^{Ni} . These equations are given below, along with the values for \bar{E}_n^{Ox} and \bar{E}_n^{Ni} .

$$\bar{E}^{Ox} = \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}}, \qquad \bar{E}^{Ni} = \frac{\sum_{n=0}^{2} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ni}}}$$
$$\bar{E}^{Ox} = -6.66651 \times 10^{-19} J, \qquad \bar{E}^{Ni} = -5.21342 \times 10^{-19} J$$

The calculations for $\frac{E_{sc}}{T}$ will be continued in the following pages.

Below is the expansions for the term $\frac{E_{sc}}{T}$. The values for the average energies \bar{E}^{Ox} and \bar{E}^{Ni} are also inserted into the bottom equation for $\frac{E_{sc}}{T}$, where the values for \bar{E}^{Ox} and \bar{E}^{Ni} are $\bar{E}^{Ox} = -6.66651 \times 10^{-19} J$ and $\bar{E}^{Ni} = -6.66651 \times 10^{-19} J$ respectively.

$$\begin{split} \frac{E_{sc}}{T} &= \frac{3}{16T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij} \frac{\sum_{n=0}^{2} E_n^{Ox} e^{-\beta E_n^{Ox}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ox}}} \\ &+ \frac{3}{16T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij} \frac{\sum_{n=0}^{2} E_n^{Ni} e^{-\beta E_n^{Ni}}}{\sum_{n=0}^{2} e^{-\beta E_n^{Ni}}} \\ \frac{E_{sc}}{T} &= -6.66651 \times 10^{-19} J \frac{3}{16T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij} \\ &- 5.21342 \times 10^{-19} J \frac{3}{16T} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij} \end{split}$$

With the computation of basic arithmetic, the term $\frac{E_{sc}}{T}$ is rewritten as the following equation below. As stated previously it is assumed that T = 298 K.

$$\frac{E_{sc}}{T} = -2.09727 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} O^{ij}$$
$$-1.64013 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i - r_j, u} M^{ij} Y_{ij} N^{ij}$$

Below are the calculations corresponding to the $E_{solvent}$ term. The formula for $E_{solvent}$ is given by the equation below.

$$E_{water} = \sum_{i=0}^{N-1} E_i - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} W^{ij} E_i$$
$$Y_{ij} = \begin{cases} 0 \ if \ i = j+1 \ or \ j = i+1 \\ E \ otherwise \end{cases}, \qquad W^{ij} = \frac{1}{4} \delta_{r_i - r_j, u}, \qquad E_i = \bar{E}^{Ox} O_i + \bar{E}^{Hb} H_i$$

The calculations for $E_{solvent}$ will be continued in the following pages.

Below are the calculations for $\frac{E_{water}}{T}$ below, where $E_i = \overline{E}^{Ox}O_i + \overline{E}^{Hb}H_i$ has been substituted into $\frac{E_{water}}{T}$. The values for \overline{E}^{Ox} and \overline{E}^{Ni} are also given below.

$$\begin{split} \frac{E_{water}}{T} &= \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Ox} O_i + \frac{1}{T} \sum_{i=0}^{N-1} \bar{E}^{Hb} H_i + \frac{1}{T} \frac{1}{4} \sum_{i=0}^{N-1} \bar{E}^{Ox} Y_{ij} W^{ij} O_i + \frac{1}{T} \frac{1}{4} \sum_{i=0}^{N-1} \bar{E}^{Hb} Y_{ij} W^{ij} H_i \\ Y_{ij} &= \begin{cases} 0 \ if \ i = j+1 \ or \ j = i+1 \\ E \ otherwise \end{cases}, W^{ij} = \frac{1}{4} \delta_{r_i - r_j, u}, \qquad E_i = \bar{E}^{Ox} O_i + \bar{E}^{Hb} H_i \\ W^{ij} &= \frac{1}{4} \delta_{r_i - r_j, u}, \qquad O_i = \begin{cases} 1 \ if \ E_i \in E_\beta \cup E_+ \cup E_- \\ 0 \ otherwise \end{cases} \\ \bar{E}^{Ox} &= -6.66651 \times 10^{-19} J, \qquad \bar{E}^{Ni} = -5.21342 \times 10^{-19} J \end{split}$$

This in turn leads to the following equations below where the values for $\overline{E}^{Ox} = -6.66651 \times 10^{-19} J$ and $\overline{E}^{Ni} = -5.21342 \times 10^{-19} J$ have both been substituted into the formula for $\frac{E_{water}}{T}$ where $\frac{E_{water}}{T}$ is the entropic penalty for the interactions with water.

$$\frac{E_{water}}{T} = -6.66651 \times 10^{-19} J \frac{1}{T} \sum_{i=0}^{N-1} O_i - 5.21342 \times 10^{-19} J \frac{1}{T} \sum_{i=0}^{N-1} H_i$$
$$-6.66651 \times 10^{-19} J \frac{1}{4T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} \delta_{r_i - r_j, u} O_i$$
$$- 5.21342 \times 10^{-19} J \frac{1}{4T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} \delta_{r_i - r_j, u} H_i$$

With the computation of basic arithmetic, the above equation is reduced to the following summation below.

$$\frac{E_{water}}{T} = -2.23708 \times 10^{-21} J \cdot K^{-1} \sum_{i=0}^{N-1} O_i - 1.74947 \times 10^{-21} J \cdot K^{-1} \sum_{i=0}^{N-1} H_i$$

+5.59271 × 10⁻²² J · K⁻¹ $\sum_{i=0}^{N-1} Y_{ij} \delta_{r_i - r_j, u} O_i + 4.37367 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i - r_j, u} H_i$

This in turn leads to the following equation for the quantum mechanical correction below.

$$\begin{aligned} \frac{\mathcal{H}_{qm}}{T} &= -5.59271 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \, \delta_{r_i-r_{j},u} Y_{ij} \\ &- 2.09727 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i-r_{j},u} M^{ij} Y_{ij} O^{ij} \\ &- 1.64013 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i-r_{j},u} M^{ij} Y_{ij} N^{ij} \\ &- 2.23708 \times 10^{-21} J \cdot K^{-1} \sum_{i=0}^{N-1} O_i - 1.74947 \times 10^{-21} J \cdot K^{-1} \sum_{i=0}^{N-1} H_i \\ &+ 5.59271 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i-r_j,u} O_i + 4.37367 \times 10^{-22} J \cdot K^{-1} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i-r_j,u} H_i \end{aligned}$$

It is important to note that the formula for the conformational entropy provides entropy values for specific protein geometries. Depending upon the conformation that the protein takes on, certain hydrogen bonds will form while other hydrogen bonds will fail to form. The hydrogen bonds between the amino acids correspond to specific values for the Kronecker Delta matrix elements in the E_{water} , E_{bb} and E_{sc} terms respectively. In order to allow the data in this model to be compared to experimental results the model will be constrained so as to assume that the conformation that the protein assumes is that of a beta strand. This means that backbone, sidechain and solvent interactions will all be at play. Since the shape of the protein is that of a beta strand, all the backbone interactions are present between every adjacent amino acid. The same holds for both the sidechain interactions as well as the solvent interactions. The entropy formulas corresponding to these geometries are given below.

$$S_{QM}(R) = (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 - \frac{E_{water}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T}$$
$$S_T(R) = (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 - \frac{E_{water}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T}$$

The equations $S_{QM}(R)$ and $S_T(R)$ are restated below for convenience.

$$S_{QM}(R) = (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 - \frac{E_{water}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T}$$
$$S_T(R) = (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 - \frac{E_{water}}{T} + \frac{E_{bb}}{T} + \frac{E_{sc}}{T}$$

In the set of equations above $S_{QM}(R)$ is the entropy corresponding to the Morse potential and $S_T(R)$ is the entropy corresponding to the tunneling correction. Likewise, $\mathcal{H}_{H2O,T}$ is the perturbation corresponding to tunneling and $\mathcal{H}_{H2O,QM}$ is the perturbation corresponding to the Morse potential. In order to determine the impact of both the quantum mechanical effects as well as the tunneling corrections, it is necessary to compare the two results. As stated previously these results are given by both \mathcal{H}_{qm_T} as well as \mathcal{H}_{qm_M} .

Comparison of the two quantum mechanical corrections

Below are the expansions corresponding to the $S_{QM}(R)$ term and the $S_T(R)$ term respectively.

$$\begin{split} S_{QM}(R) &= (0.00057922750 \ kgs^{-2}K^{-1})|R|^2 \\ &- 5.59271 \times 10^{-22}J \cdot K^{-1} \frac{1}{T} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_{i+1}-r_{j-1},u} \delta_{r_{i-1}-r_{j+1},u} \ \delta_{r_i-r_j,u} Y_{ij} \\ &- 2.09727 \times 10^{-22} \ J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i-r_j,u} M^{ij} Y_{ij} O^{ij} \\ &- 1.64013 \times 10^{-22} \ J \cdot K^{-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \delta_{r_i-r_j,u} M^{ij} Y_{ij} N^{ij} \\ &- 2.23708 \times 10^{-21} \ J \cdot K^{-1} \sum_{i=0}^{N-1} O_i - 1.74947 \times 10^{-21} \ J \cdot K^{-1} \sum_{i=0}^{N-1} H_i \\ &+ 5.59271 \times 10^{-22} \ J \cdot K^{-1} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i-r_j,u} O_i + 4.37367 \times 10^{-22} \ J \cdot K^{-1} \sum_{i=0}^{N-1} Y_{ij} \delta_{r_i-r_j,u} H_i \end{split}$$

With the use of the Python interpreter the calculations of the entropies for ΔS_{bb} , ΔS_{sc} and ΔS_{solv} for the quantum mechanical and tunneling corrections are given below.

$$S_{bb_M} = -7.65083 \cdot 10^{-19} J/K$$
$$S_{sc_M} = -1.67133 \cdot 10^{-19} J/K$$
$$S_{water_M} = -1.17391 \cdot 10^{-18} J/K$$

$$S_{bb_T} = -7.64451 \cdot 10^{-19} J/K$$
$$S_{sc_T} = 1.67133 \cdot 10^{-19} J/K$$
$$S_{water_T} = -1.17767 \cdot 10^{-18} J/K$$

Likewise, the calculations for the entropies from the experimental data are given below.

$$\begin{aligned} G_{bb} &= -48.30 \frac{kcal}{mol}, \qquad G_{sc} = -18.40 \frac{kcal}{mol}, \qquad G_{solv-polar} = 73.62 \frac{kcal}{mol} \\ G_{amino} &= G_{bb} + G_{sc}, \qquad G_{water} = H_{water}, \qquad G_{amino} = H_{bb} + \bar{E}_{sc} \\ G_{sc} &= -TS_{sc}, \qquad S_{sc} = -\frac{1}{T} \bar{E}_{sc}, \qquad \bar{E}_{sc} = -TS_{sc}, \qquad G_{sc} = \bar{E}_{sc} \\ G_{amino} &= G_{bb} + \bar{E}_{sc}, \qquad G_{bb} + G_{sc} = H_{bb} + \bar{E}_{sc}, \qquad G_{bb} + G_{sc} = H_{bb} + G_{sc} \end{aligned}$$

This leads us to the following equations for the free energy.

$$G_{bb} = H_{bb}, \qquad G_{sc} = \overline{E}_{sc}, \qquad G_{water} = H_{water}$$

The Hamiltonians corresponding to the experimental data are therefore given by the following equations below.

$$H_{bb} = -48.30 \frac{kcal}{mol}, \qquad \bar{E}_{sc} = -18.40 \frac{kcal}{mol}, \qquad H_{water} = 73.62 \frac{kcal}{mol}$$

This leads to the following values for the Hamiltonians for 1 hydrogen bond.

$$\begin{aligned} H_{bb} &= -48.30 \frac{kcal}{mol} \cdot \frac{mol}{6.02214076 \cdot 10^{23} \text{ particles}} \cdot 1 \text{ particle} = -8.020 \cdot 10^{-23} \text{ kcal} \\ H_{bb} &= -8.020 \cdot 10^{-23} \text{ kcal} \cdot \frac{4184 \text{ J}}{1 \text{ kcal}} = -3.356 \cdot 10^{-19} \text{ J} \end{aligned}$$

$$\bar{E}_{sc} = -18.40 \frac{kcal}{mol} \cdot \frac{mol}{6.02214076 \cdot 10^{23} \text{ particles}} \cdot 1 \text{ particle} = -3.055 \cdot 10^{-23} \text{ kcal}$$
$$\bar{E}_{sc} = -3.055 \cdot 10^{-23} \text{ kcal} \cdot \frac{4184 \text{ J}}{1 \text{ kcal}} = -1.278 \cdot 10^{-19} \text{ J}$$

$$73.62 \cdot \frac{kcal}{mol} \cdot \frac{mol}{6.02214076 \cdot 10^{23} \text{ particles}} \cdot 1 \text{ particle} = 1.223 \cdot 10^{-22} \text{ J kcal}$$
$$H_{water} = 1.223 \cdot 10^{-22} \text{ J kcal} \cdot \frac{4184 \text{ J}}{1 \text{ kcal}} = 5.117032 \cdot 10^{-19} \text{ J}$$

In order to make sure that the calculations for S_{bb_M} , S_{sc_M} , S_{water_M} , S_{bb_T} , S_{sc_T} and S_{water_T} have any validity they must be compared to results from the FoldX model. In order to accomplish this the values for H_{bb} , \overline{E}_{sc} and H_{water} must be inserted into the equations for E_{bb} , E_{sc} and $E_{solvent}$ respectively. Remember that H_{bb} , \overline{E}_{sc} and H_{water} are the constants for the energies pertaining to the backbone, side chain and solvent interactions respectively. The values for these constants are given below along with the summations E_{bb} , E_{sc} and $E_{solvent}$.

$$E_{bb} = \frac{1}{2} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij} \overline{E}^{Ox} , E_{solvent} = \sum_{i=0}^{N-1} E_i - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Y_{ij} W^{ij} E_i$$
$$E_{sc} = \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij} \overline{E}^{Ox} + \frac{1}{2} \frac{3}{16} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij} \overline{E}^{Ni}$$

$$H_{bbF} = -3.356 \cdot 10^{-19} J$$
$$H_{scF} = -1.278 \cdot 10^{-19} J$$
$$H_{waterF} = 5.117 \cdot 10^{-19}$$

This in turn leads to the following equations for $E_{bb_{-}E}$ and $H_{sc_{-}E}$ below.

$$E_{bbF} = -3.356 \cdot 10^{-19} J \frac{1}{2T} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij}$$

$$E_{scF} = -1.278 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij}$$

$$-1.278 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij}$$

Remember that since the experimental data was taken for a large volume of hydrogen bonds, the effects due to the directions of the side chains, which are incorporated into E_{bb} and E_{sc} by means of the fractions $\frac{1}{2}$ and $\frac{3}{16}$ respectively are included within the values for $H_{bb_{a}E}$ and $H_{sc_{a}E}$.

In other words, the terms $H_{bb_{-}E}$ and $H_{sc_{-}E}$ are roughly equal to the following values below for $N \gg 1$ hydrogen bonds, as is the case when using reduced units.

$$H_{bbE} \cong \frac{1}{2} \bar{E}^{Ox}, \qquad H_{scE} \cong \frac{3}{16} \bar{E}^{Ox}, \qquad H_{bbE} \cong \frac{3}{16} \bar{E}^{Ni}$$

For convenience a comparison between the equations for E_{bb_E} , E_{bb_M} and E_{bb_T} , as well as E_{sc_E} , E_{bb_M} and E_{sc_T} is given below.

$$E_{bbF} = -3.356 \cdot 10^{-19} J \frac{1}{2T} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij}$$
$$E_{bbM} = -3.333 \cdot 10^{-19} J \frac{1}{2T} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij}$$
$$E_{bbT} = -3.333 \cdot 10^{-19} J \frac{1}{2T} \sum_{i=1}^{N} \sum_{j=1}^{N} A^{ij} B_{ij} X^{ij} Y_{ij}$$

$$\begin{split} E_{scF} &= -1.278 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij} \\ &- 1.278 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij} \\ E_{scM} &= -1.250 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij} \\ &- 10.00 \cdot 10^{-19} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij} \\ E_{scT} &= -1.250 \cdot 10^{-19} J \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} O^{ij} \\ &- 10.00 \cdot 10^{-19} \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{ij} M^{ij} Y_{ij} N^{ij} \end{split}$$

Likewise, the average energies corresponding to a single hydrogen bond are given below. Keep in mind that these estimations hold so long as the number of amino acids is hydrogen bonds is sufficiently large.

$$H_{bbT} = -6.66101 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{1}{2} = -2.079 \ eV$$
$$H_{scT} = -6.66101 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{3}{16} = -0.780 \ eV$$
$$H_{waterT} = -6.66101 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{3}{4} = -3.123 \ eV$$

$$H_{bbM} = -6.66651 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{1}{2} = -2.080 \ eV$$
$$H_{scM} = -6.66651 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{3}{16} = -0.780 \ eV$$
$$H_{waterM} = -6.66651 \cdot 10^{-19} J \cdot \frac{6.242 \cdot 10^{18} \ eV}{1 \ J} \cdot \frac{3}{4} = -3.123 \ eV$$

For convenience the equations above are distilled to the following formulas below. The values from the FoldX model are also displayed below.

$$\begin{split} H_{bbF} &= -2.095 \; eV, \qquad H_{bbM} = -2.080 \; eV, \qquad H_{bbT} = -2.079 \; eV \\ H_{sc_F} &= -0.798 \; eV, \qquad H_{scM} = -0.780 \; eV, \qquad H_{scT} = -0.780 \; eV \end{split}$$

$$H_{water_F} = 3.192 \ eV, \qquad H_{waterM} = -3.121 \ eV, \qquad H_{waterT} = -3.121 \ eV$$

The following values for the entropy where obtained from the python interpreter corresponding to the value for the energy obtained from the FoldX model.

$$S_{bbF} = -4.8078554 \frac{eV}{K}, \qquad S_{scF} = -1.165934 \frac{eV}{K}, \qquad S_{waterF} = -5.85547 \frac{eV}{K}$$

For convenience the values for the entropies S_{bbM} , S_{bbT} , S_{scM} , S_{scT} , S_{waterM} and S_{waterT} are all given below.

$$S_{bb_{-}M} = -4.77527 \frac{eV}{K}, \qquad S_{sc_{M}} = -1.0432 \frac{eV}{K}, \qquad S_{water_{-}M} = -7.356 \frac{eV}{K}$$
$$S_{bb_{-}T} = -4.77133 \frac{eV}{K}, \qquad S_{sc_{T}} = -1.0432 \frac{eV}{K}, \qquad S_{water_{-}T} = -7.3504 \frac{eV}{K}$$

The tables corresponding to the results are given below.

Table 1: Entropies for the FoldX model, the Morse Potential, and the quantum tunneling model.

	FoldX Model	Morse Potential	Tunneling
Backbone Entropy	$-7.70303 \cdot 10^{-19} J/K$	$-7.65083 \cdot 10^{-19} J/K$	$-7.64451 \cdot 10^{-19} J/K$
Sidechain Entropy	$-9.38828 \cdot 10^{-19} J/K$	$-1.67133 \cdot 10^{-19} J/K$	$-1.67133 \cdot 10^{-19} J/K$
Solvent Entropy	$-9.38828 \cdot 10^{-19} J/K$	$-1.17391 \cdot 10^{-18} J/K$	$-1.17767 \cdot 10^{-18} J/K$

Table 2: Total energies for the FoldX model, the Morse Potential, and the quantum tunneling model.

	FoldX Model	Morse Potential	Tunneling
Backbone Energy	$-3.356 \cdot 10^{-19} J$	$-3.333 \cdot 10^{-19} J$	$-3.331 \cdot 10^{-19} J$
Sidechain Energy	$-1.278 \cdot 10^{-19} J$	$-1.250 \cdot 10^{-19} J$	$-1.250 \cdot 10^{-19} J$
Solvent Energy	$-5.117 \cdot 10^{-19} J$	$-5.000 \cdot 10^{-19} J$	$-5.000 \cdot 10^{-19} J$

CHAPTER IV:

Conclusion

Below are the percent changes corresponding to the entropic corrections for the backbone interactions, side chain interactions and solvent corrections respectively.

$$\frac{(-4.8078554 + 4.77527)}{-4.8078554} \times 100\% = -0.677753\%$$
$$\frac{(-1.165934 + 1.0432)}{-1.16435} \times 100\% = 10.5409885\%$$
$$\frac{(-5.85547 + 7.356)}{-5.85547} \times 100\% = -25.626123949\%$$

When the effects due to quantum tunneling are taken into account one obtains an increase in the conformational entropy. This is to be expected since quantum tunneling has the potential to break hydrogen bonds. This has already been shown to be the case with water molecules. Not only does quantum tunneling have an influence on the breaking of hydrogen bonds but the more hydrogen bonds form, the greater the influence of quantum tunneling effects. It must be stated however that steric interactions were not taken into account. When the effects due to steric interactions are accounted for it may very well be the case that tunneling ceases to have an influence. What we do know however is that in certain cases the effects due to quantum tunneling have more of an influence than the effects due to the potential energy between the peptide bonds. It is thus still inconclusive whether or not quantum tunneling overall has any significant impact. Further advances in the research could take into account the effects of quantum tunneling on the Van der Waals forces as well as the electrostatic forces if any are present. This would be a way of establishing whether or not quantum tunneling has an overall impact on the total conformational entropy. The results however still hold that quantum mechanical effects still have an overall positive impact on the total conformational entropy. Again, however this is only the case when the Van der Waals interactions are set to zero. The benefits of this research are that the algorithms for calculating the conformational entropy may be further optimized by taking into account the effects due to quantum tunneling.

REFERENCES

- 1. Kawakatsu, T. (1965) Statistical Physics of Polymers. Sendai, Japan: Springer
- C. S. dos Santos, E. Drigo Filho, R. M. Ricoota. Int. J. Quantum Chem, 12, 115, (2015)
- 3. Sanne Abeln, Michele Vendruscolo, Christopher M. Dobson, Daan Frenkel,
- 4. R. K. Pathria, Paul D. Beale, Statistical Mechanics. Burlington, USA: Elsevier
- 5. Nouredine Zettili, Quantum Mechanics. Jacksonville, USA: Wiley
- 6. Arthur M. Lesk, *Introduction to protein architecture*. New York, USA: Oxford University Press
- 7. Kyungsun Kim, Richard A. Friesner, J. Am. Chem. Soc, 119, (1997)
- 8. Griffiths, David. Introduction to Quantum Mechanics, Portland, Prentice Hall, Inc., 1995
- J. Richardson, C. Perez, S. Lobsiger, A. Reid, B. Temelso, G. Shields, Z. Kisiel, D. Wales, B. Pate, S. Althorpe, "Concerted hydrogen-bond breaking by quantum tunneling in the water hexamer prism." *Science*. 18, March 2019. 1310-1313.
- Michael Baxa, Esmael Haddadian, John Jumper, Karl Freed, Tobin Sosnick, "Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations." 13, Oct 2014.
- 11. Garcia LG, Pereira Araujo, "Entropy Reduction Effect imposed by hydrogen bond formation on protein folding cooperativity, evidence from a hydrophobic minimalist model." *Physical Review*. 1 Nov 2005.
- 12. Onur Pusuluk, Gokhan Torun, Cemsinan Deliduman, "Quantum Entanglement in Hydrogen Bonds and its usage as a resource in Molecular Recognition." *Biological Physics.* 2 Jul 2018.
- 13. Moran Laurence, Horton Robert, Scrimgeour Gray. Principles of Biochemistry. W. Ross MacDonald School Resource Services
- 14. F. Mallamace, C Corsaro, D. Mallamace, S. Vasi, C. Vasi, P. Baglioni, S. Buldyrev, S. Chen, E. Stanley, "Energy Landscape in protein Folding and Unfolding." *Proceedings of the National Academy of Sciences of the United States of America*. 22 Dec 2015. 3159-3163.
- 15. J. Caro, K. Harpole, V. Kasinath, J. Lim, J. Granja, K. Valentine, K. Sharp, J. Wand, "Entropy in Molecular Recognition by Proteins." *Proceedings of the National Academy of Sciences of the United States of America*. 6563-6568.
- 16. Judson Thomas, Stephan, Abstract Algebra Theory and Application.
- 17. Jinfeng Zhang, Jun S Liu "On Side Chain Conformational Entropy of Proteins." *PloS Computational Biology*. Dec 2006
- 18. Thirmulali, D. Reddy, G. Straub, J. E. "Role of water in Protein Aggregation and Amyloid Polymorphism." Acc Chem Res. 45, 1, 83-92
- Petukhov, Cregut, D., Soares, C.M. and Serrano, L. (1999). "Local Water bridges and Protein Conformational stability." *Protein Sci 8. 1999, 8(10),* 1982-9.
- Huggins, D. J. "Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations" *Biophysics J.* 2015. 108(4), 928-936
- Buß, O., Rudat, J., Ochsenreither, K.: "FoldX as Protein EngineeringTool: Better than Random Based Approaches?" *Comput. Struct. Biotechnol. J.* 16, 25-33 (2018).
- 22. Ainavarapu SR et al., "Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds". *Biophysics J.92:225-233.* (2007).
- 23. J.M. Jumper, N.F. Faruk, K.F. Freed, T.R. Sosnick, "Accurate Calculation of side chain packing and free energy with applications to protein molecular dynamics" *PloS Comput Biol* 14(12): e1006342 (2018)
- 24. H. Reza et al., "QM/MM Calculations provide insight into the mechanism of bioluminescence triggering in ctenophore photoproteins". PLOS ONE, 12(8), E0182317 (2017)

APPENDIX A:

AMINO ACID TABLES

Tabl	e A	1: E	lydr	oger	ı bo	nd	pr	obal	bilit	ies.
				<u> </u>						

	А	-	L	Μ	V	F	W	Υ	Ν	С	Q	S	Т	D	Е	R	Н	К	G	Ρ
А	0	0	0	1	0	1	0	0	1	1	1	0	1	1	1	0	0	1	1	0
Ι	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Μ	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	1	1	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	1	0	0	0	0	2	0	0	0.6	0	0.6	1	1	0.5	0.5	0	0	1	1	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Υ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ν	1	0	0	1	0	0.6	0	0	2	1	1	1	1	1	1	0	0	1	0.8	0
С	1	0	0	1	0	0	0	0	1	2	1	1	1	1	1	0	0	1	1	0
Q	1	0	0	1	0	0.6	0	0	1	1	2	1	1	1	1	0	0	1	0.8	0
S	0	0	0	1	0	1	0	0	1	1	1	2	1	1	1	0	0	1	1	0
Т	1	0	0	1	0	1	0	0	1	1	1	1	2	1	1	0	0	1	1	0
D	1	0	0	1	0	0.5	0	0	1	1	1	1	1	2	1	0	0	1	1	0
Е	1	0	0	1	0	0.5	0	0	1	1	1	1	1	1	2	0	0	1	1	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Н	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
К	1	0	0	1	0	1	0	0	1	1	1	1	1	1	1	0	0	2	1	0
G	1	0	0	1	0	1	0	0	0.8	1	0.8	1	1	1	1	0	0	1	0	0
Ρ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A1 contains the probability that a Hydrogen bond will form between the sidechains of two amino acids. The top row and the left column contain all 20 amino acids. The elements in each cell correspond to the probability that a hydrogen bond will form between the two amino acids when their sidechains are in contact.

				\sim																
	А	Ι	L	Μ	V	F	W	Υ	Ν	С	Q	S	Т	D	Е	R	Н	Κ	G	Ρ
А	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ι	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Μ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	Ρ	0	0	0	0	Р	0	0	Ρ	0	Ρ	Ρ	Ρ	Ρ	Ρ	0	0	Ρ	Ρ	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Υ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ν	Ν	0	0	0	0	Ν	0	0	Ν	Ν	Ν	Ν	Ν	Ν	Ν	0	0	Ν	Ν	0
С	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	Ν	0	0	0	0	Ν	0	0	Ν	Ν	Ν	Ν	Ν	Ν	Ν	0	0	Ν	Ν	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Т	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Е	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Н	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
К	Ν	0	0	0	0	n/a	0	0	Ν	Ν	Ν	Ν	Ν	Ν	Ν	0	0	Ν	Ν	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ρ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A2: Hydrogen acceptors.

Table A2 contains the hydrogen acceptors for each amino acid pair. Both the top row and the left column contain all 20 amino acids. Each cell corresponds to the hydrogen bond acceptors where N, O and P stand for the nitrogen bond, hydrogen bond and (Pi bond?) respectively.

	\uparrow	\rightarrow	\downarrow	\leftarrow
\uparrow	\checkmark	Х	Х	Х
\rightarrow	Х	Х	Х	\checkmark
\downarrow	Х	Х	\checkmark	Х
	Х	Х	X	X

Table A3: Sidechain vector combinations.

Table A3 lists the sidechain vector combinations that permit a hydrogen bond. The up and down arrows correspond to side chains pointing up and down respectively. The x's correspond to side chains pointing into the page. The *'s correspond to the side chains pointing out of the page. In order to remove the conformations with the sidechains pointing in the same direction for the consecutive side chains, a steric penalty is applied. This corresponds to the cases where the sidechains are pointing in the same direction.