6-2011

Effects of Arginine on the Kinetics of Bovine Insulin Aggregation Studied by Dynamic Light Scattering

Michael M. Varughese
Union College - Schenectady, NY

EFFECTS OF ARGININE ON THE KINETICS OF BOVINE INSULIN AGGREGATION STUDIED BY DYNAMIC LIGHT SCATTERING

By

Michael M. Varughese

****

Submitted in partial fulfillment
of the requirements for
Honors in the Department of Biological Sciences
and done in conjunction with the Department of Physics and Astronomy

UNION COLLEGE
June, 2011

ABSTRACT

VARUGHESE, MICHAEL Effects of Arginine on the Kinetics of Bovine Insulin Aggregation Studied by Dynamic Light Scattering. Departments of Biological Sciences and Physics and Astronomy. June, 2011.

ADVISORS: Professor Jay Newman and Professor Brian Cohen

In the fields of protein science and medicine, understanding the kinetics of protein aggregation are significant in the research and treatment of certain amyloid diseases such as Alzheimer's. Previous studies have suggested that arginine can increase the solubility of certain proteins, suppress protein aggregation, and assist in the refolding of aggregated proteins; however, the molecular mechanisms by which arginine can influence protein aggregation are still unclear. Bovine insulin was employed as a model system for further understanding the effects of arginine on protein aggregation.

Using Dynamic Light Scattering (DLS), we studied the concentration-dependent and temperature-dependent suppression of aggregation in insulin by means of arginine. Arginine concentrations from 10mM to 500mM were shown to have produced a concentration-dependent increase in the lag time of the aggregation, which is the period preceding protein aggregation. DLS measurements of insulin in the presence of arginine from 60°C to 85°C showed a significant increase in the aggregation delay for samples with arginine compared to control samples without arginine. Arginine samples were shown to have delayed aggregation by up to a factor of 7.5. From Arrhenius analysis, we also found that the activation energy of 1mM insulin was 17 ± 5 kcal/mol while the energy of the insulin samples with 500mM arginine was higher (26 ± 3 kcal/mol). These energy values are in accordance with the energy associated with β-sheet formation, which is about 0.5 kcal/mol/residue (or ~25 kcal/mol for monomeric insulin). The 9 kcal/mol difference may quantify the barrier effect of arginine on insulin aggregation.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my thesis advisor, Professor Jay Newman, for his help in completing this thesis from beginning to end. I had the privilege of first working with Professor Newman as his research assistant during the summer after my freshman year in 2008. He taught me about the theory of DLS, how to use the DLS apparatus, and how to collect and analyze DLS data. I was able to work with him again in the summer of 2010 in order to pursue an honors in biology through this biophysical research. It was a pleasure to work with an advisor that is highly knowledgeable about this line of research, was constructively critical during the writing process, and was extremely flexible with his availability to help me through this whole process. In addition, I would like to acknowledge and thank my biology reader, Professor Brian Cohen, for overseeing my research in accordance with the requirements for an honors thesis in the Department of Biological Sciences. Lastly, I would like to thank both the Department of Biological Sciences for their approval of this biophysical research and the Department of Physics and Astronomy for offering a supportive environment for my research.

Table of Contents

1. Introduction

1.1. Insulin

Insulin, a key component in glucose metabolism, is a small macromolecule composed of 51 amino acids. It consists of two chains, an A chain, composed of an α-helix and a B chain, composed of an α-helix and a β-sheet. These two chains are connected via disulfide bonds at the ends of the A chain and the C-terminal of the B chain (Hua & Weiss, 2004, p. 21449).

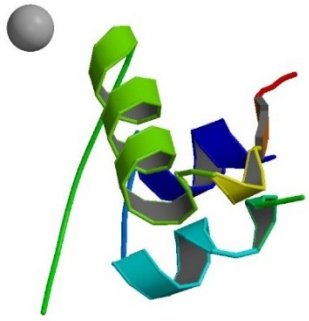

Figure 1. Structure of human insulin containing the A chain (shown in shades of blue) and the B chain (shown in green) with disulfide bonds in between (Timofeev, Baidus, Kislitsyn, & Kuranova).

Insulin is synthesized as Zn^{2+}-coordinated hexamers in human β-cells. It has been proposed that this occurs in order to protect insulin from fibrillation during *in vivo* storage. However, when in use, insulin functions as a Zn^{2+}-free monomer (Hua & Weiss, 2004). In addition, insulin structure is relatively conserved throughout different species. In this research, bovine insulin, which differs from human insulin in 3 amino acids, was used. The hexameric form of bovine insulin is shown below in Figure 2.

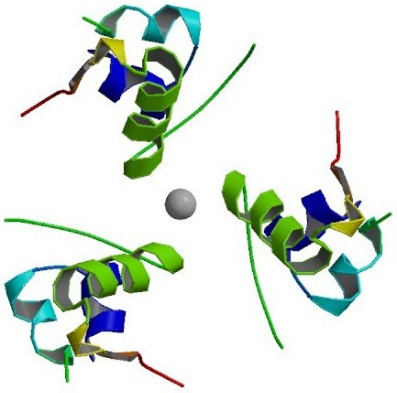

Figure 2. Hexameric form of bovine insulin (Jaimohan, Naresh, & Mandal).

Along with the hormone glucagon, insulin controls the glucose concentration in blood. It is produced in the islets of Langerhans located in the pancreas. It attaches itself to receptors on the surface of cells and increases glucose uptake by those cells, namely liver and muscle cells. When the function of insulin or insulin-producing cells is impaired, diabetes can develop. Diabetes comes in two types: type I diabetes, which occurs when one's immune system destroys the cells in the pancreas where insulin is produced; and type II diabetes, which occurs when insulin receptors lose sensitivity. Type I diabetes is treated with insulin injections or insulin transfusions. The problem that emerges is that insulin, like many other proteins, will undergo aggregation under certain conditions. This aggregation starts with the normal monomer state and advances into the dimer and later fibril stages of aggregation. In type II diabetes, insulin aggregates may form in the pancreas. Insulin aggregation renders the insulin useless for the conversion of glucose in the body into useful energy.

The aggregation of insulin has been observed via several scientific techniques and has a generally accepted pathway. Figure 3 below illustrates the pathway of insulin from its monomeric form to advanced fibril bundles.

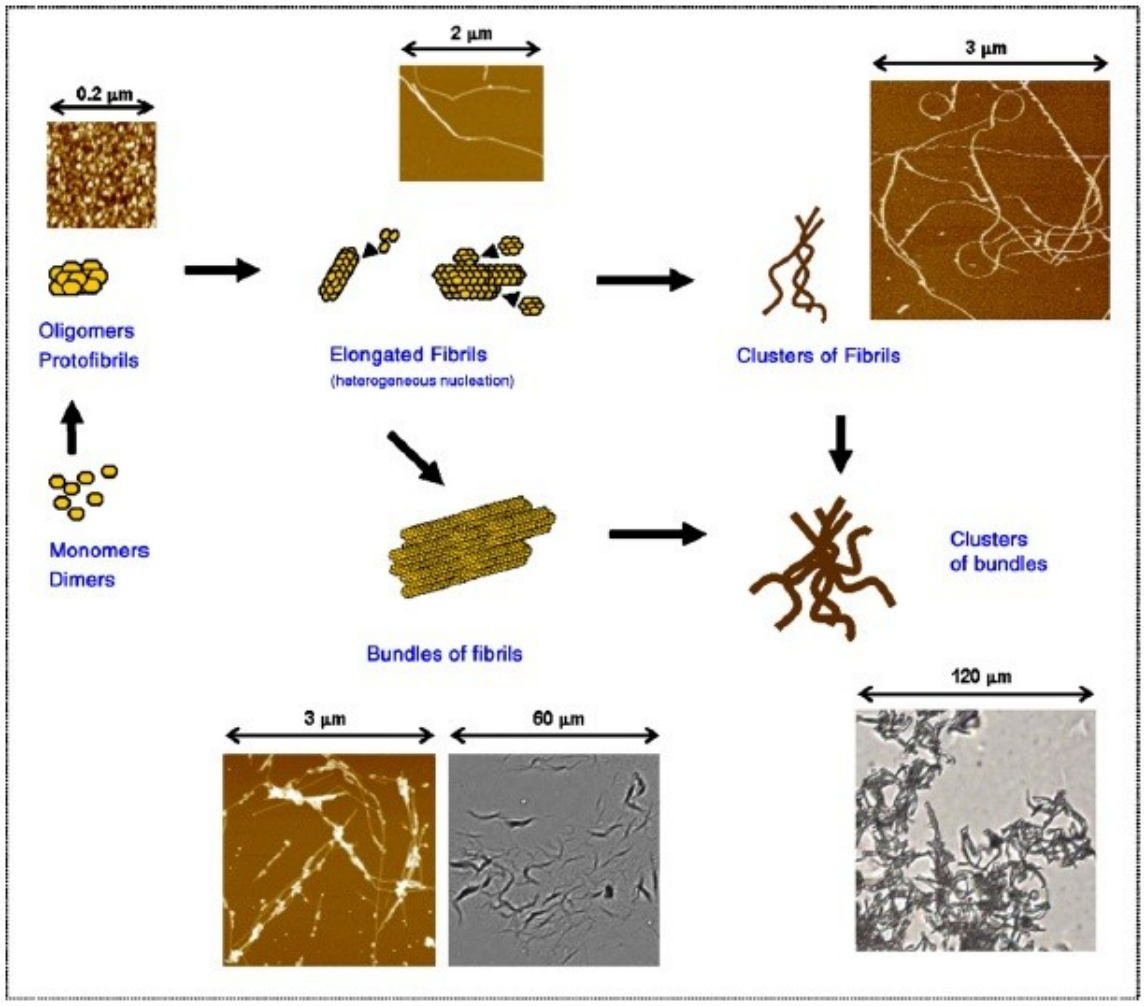

Figure 3. The aggregation pathway of insulin (Mauro, et al., 2007).

Insulin is in a category of proteins that can form amyloids that have been implicated in amyloidosis. Amyloidosis is characterized by the abnormal self-assembly and deposition of proteins into insoluble aggregates possessing β-sheet structures (Bouchard, Zurdo, Nettleton, Dobson, & Robinson, 2000, p. 1960). The aggregation of insulin can be used to learn about the aggregation seen in other similar proteins. It is thought that a population of unfolded or partially folded intermediates is the link between the various amyloid-associated disorders (Merlini & Bellotti, 2003). These amyloid diseases include Alzheimer's disease, the spongiform

encephalopathies, and type II diabetes (Bouchard, Zurdo, Nettleton, Dobson, & Robinson, 2000, p. 1960); (Thompson, 2003).

In the aggregation of insulin, the first step is nucleation, which is the formation of oligomeric centers known as nuclei from prefibrillar aggregates (Mauro, et al., 2007, p. 266); (Smith, Sharp, & Roberts, 2008); (Ahmad, Uversky, Hong, & Fink, 2005). The time it takes to form stable nuclei is referred to as the lag time for the aggregation (Nielsen, et al., 2001, p. 6036). From the nuclei, fibrils form through elongation and create the characteristic β-sheet structures. These fibrils develop into higher level structures such as bundles and clusters as seen in Figure 3 or spherulites (Krebs, MacPhee, Miller, Dunlop, Dobson, & Donald, 2004). Finally, a gel-like matrix is formed by the fibrils and other structures.

1.2. L-Arginine

L-Arginine, or (S)-2-Amino-5-guanidinopentanoic acid as designated by the International Union of Pure and Applied Chemistry (IUPAC), is an amino acid commonly found in the human body. The structural formula of L-Arginine is shown in Figure 4.

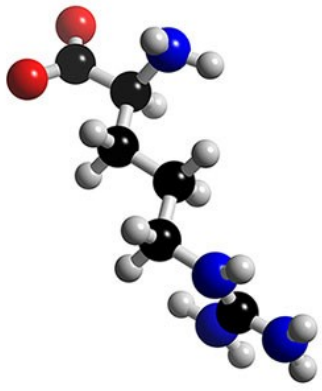

Figure 4. The chemical structure of L-Arginine.

L-Arginine has been shown to affect protein refolding, protein-protein interactions, and protein aggregation (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1301). With regard to protein refolding, L-Arginine appears to interrupt disulfide bond formation which causes partly folded polypeptide chains to remain soluble for polypeptides with free thiol groups *in vitro*. However, L-Arginine was also shown to suppress aggregation and allow refolding and disulfide formation in certain concentrations of periplasmically secreted proteins *in vivo* (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, pp. 1302-1303). L-Arginine has also been shown to assist in eluting antibodies from Protein-A or –G affinity chromatography. Affinity chromatography is used to purify antibodies by attracting the antibodies to the column's proteins and allowing the other proteins to wash away. A problem arises in trying to remove the antibodies from the column wall. Under low pH, it is possible to remove the antibodies, but at that pH the antibodies are also destabilized. By using a buffer containing Arginine, the antibodies can be eluted from the column wall without compromising their stability (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1303). Lastly and most importantly for this research, arginine has been reported to suppress the aggregation of unfolded proteins such as lysozyme and bovine serum albumin (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1304). Little is still known about the true mechanism of arginine's inhibition of aggregation, but research has shown that arginine, in certain amounts and at high temperatures, inhibits both heat-induced and dilution-induced aggregation in certain proteins (Shiraki, Kudou, Fujiwara, Imanaka, & Takagi, 2002, p. 595). One research group suggests that arginine may have an affinity for side chains that are responsible for protein aggregation and may also interact with aromatic side chains that are buried in native proteins that are responsible for aggregation (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1305). Other research has suggested that arginine can behave in both a stabilizing and destabilizing fashion; it can contract

the unfolded states of certain proteins in order to stabilize them, but can also effectively bind to intermediate protein structures of other proteins in order to destabilize higher order protein structures. (Ghosh, Sharma, & Chattopadhyay, 2009, p. 1142).

2. Theory of Dynamic Light Scattering

Dynamic light scattering (DLS) is a technique that is utilized to ascertain information about the shape and size of molecules or macromolecules that are in solution. In order to achieve this, light is directed toward the sample in solution, and the light that is scattered from the sample is analyzed at various angles. More specifically, the scattered light will fluctuate in intensity due to the Brownian motion of the molecules, the random movement of the scatterers. By using what is known as an autocorrelation function, where these intensity fluctuations are correlated with intensity fluctuations at slightly later times, the translational diffusion coefficient can be obtained. A larger diffusion coefficient means that the particle has more rapid Brownian motion, typically implying that it is smaller in size. The translational diffusion coefficient can be used to obtain a hydrodynamic radius, which is the radius of a sphere which behaves hydrodynamically like the irregularly shaped particle. Additionally, a Perrin or form factor, F, can be calculated. This factor provides information about the native shape of the scatterer being studied (Dahneke, 1983, p. 5).

DLS requires the incoming light beam to be monochromatic, intensity-stable, coherent (in-phase), and plane-polarized. The most practical light source used in such experiments is a laser. Now we will delve into the more mathematical premises of dynamic light scattering. The initial light wave has a wavevector, $\vec{k}_0$, while the scattered light wave is at some angle, θ, from

the incoming wave and can be represented by a wavevector, $\vec{k}_s$, where (Cummins, 1974, p. 286):

$$\vec{k} = \frac{2\pi}{\lambda}, \quad \text{[Eq. 1]}$$

where λ is the wavelength of the corresponding light wave. Note that since the frequencies of the light waves are nearly the same, $\vec{k}_0$ and $\vec{k}_s$ are approximately the same magnitude. The electric field strength of the scattered wave is represented as E and is determined to be:

$$\boldsymbol{E} = \sum_j \boldsymbol{A_0} e^{i(\phi_j - \omega t)}, \quad \text{[Eq. 2]}$$

where A_0 is the amplitude of the electric field, ω is the angular frequency of the incoming light wave, and ϕ_j is the phase of the field scattered from the j^{th} scatterer. This assumes that each molecule, j, in the solution is independently interacting with the incoming light wave; thus the summation of each molecule's interaction with the light wave results in the total electric field strength detected, and the amplitudes of the electric fields are equal for each molecule.

The difference between $\vec{k}_0$ and $\vec{k}_s$ is called the scattering vector, which is represented as $\vec{q}$, as seen in Figures 5 and 6.

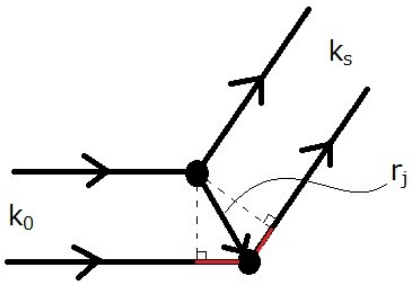

Figure 5. Geometry of a light scattering experiment. The red line segments denote the path difference between the two incident light rays.

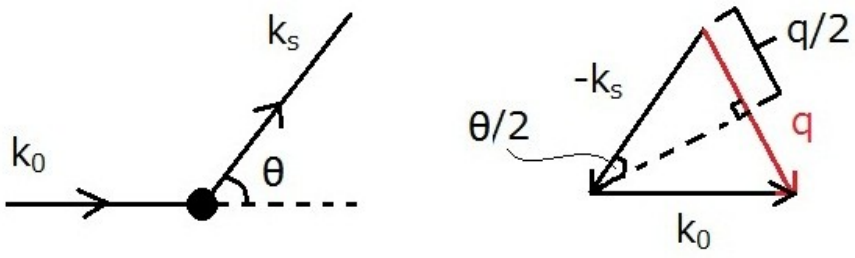

Figure 6. Trigonometric Geometry.

If an origin is defined, such that ϕ=0, then the ϕ_j from the j^{th} molecule is the dot product of the scattering vector, $\bar{q}$, and the distance, $\bar{r_j}$, from the origin to the center of mass of the molecule. As seen in Figure 5, the path difference between the incident rays is highlighted in red. Since the dot product is related to the scalar projection of a vector onto another vector, the path difference along $\hat{k}_0$ can be described through the dot product of $\bar{r_j}$ and $\hat{k}_0$, while the path difference along $\hat{k}_s$ can be described through the dot product of $\bar{r_j}$ and $-\hat{k}_s$. The total path difference is the sum of the path differences along both the incident and the scattering wavevectors. The path difference is directly related to the phase difference; the wavevectors are defined in terms of the wavelength of the incoming light and the dot product of the wavevectors and r_j describe the path difference of the scattered light waves. Since $\bar{k}_0 - \bar{k}_s = \vec{q}$, , ϕ_j is the dot product of $\bar{q}$ and $\bar{r_j}$. Therefore, Equation 2 can be rewritten as:

$$\boldsymbol{E = \sum_j A_0 e^{i([\vec{q}\cdot\vec{r}_j]-\omega t)}}, \quad \text{[Eq. 3]}$$

Using trigonometric relationships as seen in Figure 6, the scattering vector can be solved for through simple algebraic rearrangement:

$$\boldsymbol{\frac{q}{2} = k_0 sin\left(\frac{\theta}{2}\right)}, \quad \text{[Eq. 4]}$$

The scattering vector is determined to be

$$q = \left(\frac{4\pi n_0}{\lambda_0}\right) sin\left(\frac{\theta}{2}\right), \quad \textbf{[Eq. 5]}$$

where λ0 is the wavelength of the incoming light in a vacuum and n0 is the index of refraction of the medium (the solution) (Cummins, 1974, p. 288).

Ultimately, it is the intensity of the scattered light wave that will yield information about the sample. In order to obtain the intensity of the light, the light is sent to a phototube, where photons will strike a metal photocathode and scatter electrons which in turn will scatter more electrons within the photomultiplier tube until a measurable current is able to be detected. This current is proportional to the intensity of light. Furthermore, this intensity is related to the electric field strength from Equation 1 as follows:

$$\boldsymbol{I(t)} =< \boldsymbol{A_0^2 \sum_j \sum_{j'} e^{i\vec{q}\cdot\left(\vec{r}_j - \vec{r}_{j'}\right)}} >, \quad \text{[Eq. 6]}$$

Since the intensity of the light measured by the phototube is constantly fluctuating, due to the Brownian motion of the scatterers, the correlation between the intensity at some time, t, and the intensity at some later time, t+τ, will reveal information about the motion of the molecules in the medium. The comparison of the correlation of the intensity with itself is called autocorrelation, and the autocorrelation function for intensity is given by

$$\boldsymbol{G^{(2)}(\tau)} =< \boldsymbol{I(t)I(t+\tau)} >. \quad \text{[Eq. 7]}$$

An autocorrelation function also exists for the electric field strength and is given by:

$$\boldsymbol{G^{(1)}(\tau)} =< \boldsymbol{E^*(t)E(t+\tau)} >. \quad \text{[Eq. 8]}$$

These two autocorrelation functions are related through the equation:

$$\boldsymbol{g^{(2)}(\tau) = \left|g^{(1)}(\tau)\right|^2 + 1}, \quad \text{[Eq. 9]}$$

where the small g functions are normalized:

$$g^{(1)}(\tau) = \frac{G^{(1)}(\tau)}{G^{(1)}(0)} \quad \text{[Eq. 10]}$$

$$g^{(2)}(\tau) = \frac{G^{(2)}(\tau)}{[G^{(1)}(0)]^2}. \quad \text{[Eq. 11]}$$

Substituting Eq. 2 into Eq. 8 yields,

$$G^{(1)}(\tau) = < \sum_{j=1}^{N} A_0 e^{-iq \cdot r_j(0)} \sum_{j=1}^{N} A_0 e^{iq \cdot r_{j'}(\tau)} > e^{-i\omega_0 \tau}, \quad \text{[Eq. 12]}$$

where N is the total number of particles. Assuming that we are working with dilute solutions, the positions of the different molecules in the solution are not correlated so cross terms are eliminated and a factor of N is introduced to remove the summation:

$$G^{(1)}(\tau) = N|A_0|^2 < e^{iq \cdot (r(\tau) - r(0))} > e^{-i\omega_0 \tau}. \quad \text{[Eq. 13]}$$

Due to random diffusion, the following is true (Cummins, 1974, p. 296):

$$< e^{iq \cdot (r(\tau) - r(0))} > = e^{-D_T q^2 \tau}, \quad \text{[Eq. 14]}$$

where D_T is the translational diffusion coefficient. Thus Equation 13 now becomes:

$$G^{(1)}(\tau) = N |A_0|^2 e^{-D_T q^2 \tau} e^{-i\omega_0 \tau}. \quad \text{[Eq. 15]}$$

Combining Eq. 9, 10, and 15, we see that:

$$g^{(1)}(\tau) = \frac{G^{(1)}(\tau)}{G^{(1)}(0)} = e^{-D_T q^2 \tau} e^{-i\omega_0 \tau} \quad \text{[Eq. 16]}$$

so that,

$$g^{(2)}(\tau) = e^{-2D_T q^2 \tau} + 1 = e^{-2\Gamma \tau} + 1. \quad \text{[Eq. 17]}$$

Also note that $D_T q^2$ is often referred to as Γ, the decay constant. As we can see, the normalized intensity autocorrelation function has a range from 2 to 1. When τ=0, $g^{(1)}(\tau)=1$ and $g^{(2)}(\tau)=2$; also, when τ is large (approaches infinity), $g^{(1)}(\tau \rightarrow \infty)= 0$ and $g^{(2)}(\tau \rightarrow \infty)=1$. Mathematically, there are two possible means of obtaining the translational diffusion coefficient: either by keeping τ fixed

(Eq. 5); or by keeping θ fixed (q fixed) and varying only τ, which is the more commonly used method. In order to determine D_T using the first method, the following equation, which is the modified form of Eq. 17, is used:

$$\frac{\ln\left(g^{(2)}(\tau)-1\right)}{\tau} = -2D_T q^2. \quad \text{[Eq. 18]}$$

Notice that this equation shows that there is a linear relationship between $\ln[g^{(2)}(\tau)-1]$ and q^2. D_T is determined by analyzing the slope of this plot, since D_T is the negative slope. Similarly, there is a linear relationship between $\ln[g^{(2)}(\tau)-1]$ and τ as shown below:

$$\frac{\ln\left(g^{(2)}(\tau)-1\right)}{q^2} = -2D_T\tau. \quad \text{[Eq. 19]}$$

Again D_T is determined by analyzing the slope of this plot, since D_T is the negative slope.

Related to this method of obtaining the translational diffusion coefficient is the intensity autocorrelation time, τ_c:

$$\tau_c = \frac{1}{2D_T q^2} = \frac{1}{2\Gamma}. \quad \text{[Eq. 20]}$$

According to Einstein, the translational diffusion coefficient in one dimension can be written as

$$D_T = \frac{\langle r^2\rangle}{2t}, \quad \text{[Eq. 21]}$$

where $\langle r^2\rangle^{1/2}$ is the root mean square distance travelled by a particle in time t due to its Brownian motion. As seen in Equation 5, q has units of inverse distance and has a dependence on the scattering angle, medium, and wavelength. Combining equations 20 and 21 shows that the intensity correlation time, τ_c, is the average time it takes a particle to diffuse a distance of 1/q. Thus large particles will take a longer τ_c to diffuse a distance of 1/q than smaller particles.

By knowing the translational diffusion coefficient, the hydrodynamic radius of the molecule can be calculated through the Stokes-Einstein equation:

$$D_T = \frac{kT}{6\pi\eta r_h}, \quad \text{[Eq. 22]}$$

where k is the Boltzmann constant, T is the absolute temperature of the solution, η is the viscosity of the solvent, and r_h is the radius of the molecule. It should also be noted that this assumes the molecule is spherical in shape (Cummins, 1974, p. 296). The translational diffusion coefficient can also be written as:

$$D_T = \frac{kT}{f}, \quad \text{[Eq. 23]}$$

and,

$$f_{sph} = 6\pi\eta r_h, \quad \text{[Eq. 24]}$$

where f is the frictional coefficient of the scatterer, f_{sph} is the frictional coefficient of a spherical molecule of radius, r_h, in a solution with a fluid viscosity of η. The Perrin, or shape factor, F, is useful in determining information about the molecule's shape. The shape factor is the following:

$$F = \frac{f}{f_{theoretical}} = \frac{R_H}{R_H^{Theoretical}}, \quad \text{[Eq. 25]}$$

where,

$$R_H^{Theoretical} = \left[\frac{3M(V_S+h)}{4\pi N_A}\right]^{\frac{1}{3}}, \quad \text{[Eq. 26]}$$

where $f_{theoretical}$ is the theoretical frictional coefficient of a sphere whose radius is calculated from its molecular mass, R_H is the hydrodynamic radius, $R_{H,theoretical}$ is the solution to Equation 26, M is the molecular mass, V_S is the partial specific volume of the molecule, h is the hydration of the molecule, and N_A is Avogadro's number. F compares the frictional behavior of a molecule with that of an anhydrous sphere of the same molecular mass. The closer F is to 1, the more spherical

the molecule is. The farther F deviates from 1, the more non-spherical the scatterer is. This is how the Perrin factor helps gain insight into the shape of the molecule.

Alterations are made to these basic equations in order to analyze molecules that are non-spherical or molecules that are subjected to conditions which do not satisfy the several assumptions made during the use of the above equations and relationships.

Our analysis thus far has dealt with a monodisperse solution, a solution with uniformly sized scatterers; however, it is much more likely that a biological solution is polydisperse, having varied sized scatterers. For a polydisperse solution, the measured diffusion coefficient is the average of the diffusion coefficient of each scatterer's size. Equation 16, which is a single exponential, becomes a superposition of exponentials. This is written as

$$\left|g^{(1)}(\tau)\right| = \int_0^\infty F(\Gamma)e^{-\Gamma\tau}d\Gamma, \quad \text{[Eq. 27]}$$

where Γ is the decay constant mentioned previously in Equation 17, and F(Γ) is the distribution of decay rates (Cummins, 1974, p. 303). The exponential $e^{-\Gamma\tau}$ is expanded about the average day rate, <Γ>, and the natural logarithm is taken to obtain

$$\ln\left(g^{(1)}(\tau)\right) = C_0 - C_1\tau + C_2\tau^2 - C_3\tau^3 + \cdots, \quad \text{[Eq. 28]}$$

where,

$$C_n = (-1)^n \int (\Gamma - <\Gamma>)^n F(\Gamma)d\Gamma, \quad \text{[Eq. 29]}$$

where, C_n are the amplitudes, known as the cumulants of $g^{(1)}(\tau)$. A polydispersity parameter can be determined. The following is a second degree cumulant polydispersity parameter:

$$p = \frac{C_2}{\Gamma^2}, \quad \text{[Eq. 30]}$$

where $\bar{\Gamma}$ is the average decay rate. By use of the Laplace transform and the CONTIN program, the populations of the different sized scatterers are determined. This is used to determine how many different populations exist. This analysis also determines the order of the exponential fit that should be used for the semi-log plot of g versus time. This semi-log plot is fit to increasing order polynomials similar to those seen in Equation 28.

Returning to DLS theory under the assumption that there is a monodisperse solution, but when the scatterers are not small or spherical, a form factor, P(θ), is used to determine the scatterer's shape. When scatterers are not small or spherical, the intensity of the scattered light is less than expected due to interference, and is reduced by this angle-dependent form factor. The form factor can be graphed in terms of u,

$$u = \frac{qL}{2}, \quad \text{[Eq. 31]}$$

where L is the length of the particle. The following graph shows the form factors of different shaped particles (Cummins, 1974, pp. 291-292):

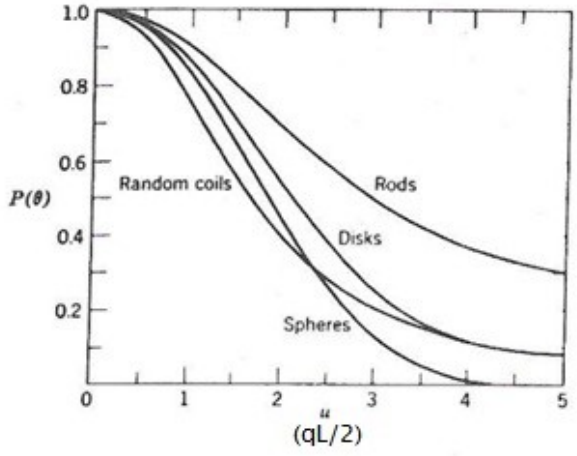

Figure 7. Form Factor P(θ)

Thus fitting data to the form factors can be used to determine the scatterer's shape.

These additions to the basic theory of dynamic light scattering are useful in the analysis of more complex macromolecules such as proteins. Using these additions, it is possible to predict and determine shape and size.

3. Materials and Methods

3.1. Sample Preparation

The standard insulin samples used in the experiments of this thesis were samples of 1mM bovine insulin in an aqueous solution of 20% (v/v) acetic acid and 0.1M NaCl. Optical glass cuvettes were labeled and weighed on a balance. Then in a centrifuge tube, powdered bovine insulin was dissolved in the aqueous solution of acetic acid and NaCl. For every cuvette, 5.8 mg of powdered bovine insulin (purchased from Sigma-Aldrich # I5500, CAS 11070-73-8 and stored at -20^{o}C) was dissolved in 1 mL of solvent at 4°C; samples were prepped in groups ranging from 3 to 13 samples at a time. The centrifuge tube was spun at 5000 rpm in a JA-25.50 rotor of an Avanti Model J-25 centrifuge at 4^{o}C in order to allow any large particulates to sink to the bottom. Working in a cold room at 4^{o}C, the insulin solution was then carefully extracted near the meniscus via a Pasteur pipette into a glass syringe equipped with a 13 mm diameter 0.22 micron pore sized filter (Millex-GV Low Protein Binding Durapore (PVDF) Filter Unit- 0.22 µm, Cat #SLGVX13NL). The solution was then filtered into individual glass cuvettes (about 0.8 mL to 1 mL of solution per cuvette). The cuvettes were then weighed again and the difference in mass was used to determine the exact mass and volume of the insulin sample filtered into it. The samples were stored at approximately 2°C. When in use, samples were allowed to equilibrate to room temperature and then either placed straight into the DLS apparatus or first augmented with varying doses of L-Arginine as next described (purchased from Sigma-Aldrich #A5006, CAS 74-79-3).

Some experimental samples were further prepared by the addition of a measured volume of a 4M L-Arginine solution (aqueous solution of L-Arginine in 20% (v/v) acetic acid and 0.1M NaCl). By using the mass of the insulin solution sample, the volume of the solution was calculated. Using this volume, the necessary volume of 4M L-Arginine solution was calculated in order to obtain concentrations of L-Arginine from 10mM to 500mM, and the exact volume of 4M stock solution was added to the cuvette via a micropipette.

$$Let:$$

$$C_i = the\ concentration\ of\ stock\ solution\ of\ L-Arginine$$
$$C_f = the\ desired, final\ concentration\ of\ L-Arginine$$
$$V_i = the\ volume\ of\ initial\ solution\ of\ 1mM\ Insulin$$
$$V_a = the\ volume\ of\ L-Arginine\ stock\ solution\ added\ to\ the\ insulin\ solution$$

$$C_i V_a = C_f(V_i + V_a)$$
$$V_a = \frac{C_f}{(C_i - C_f)} V_i$$

By determining V_a, the necessary volume of L-Arginine stock solution can be added to the 1mM insulin sample. In order to produce a sample of 1mM insulin with 500mM of L-Arginine from a 4M stock solution, the desired volume of stock solution is 14.3% of the volume of initial insulin solution. However at low temperature (below roughly 40°C), 4M L-Arginine in 20% (v/v) acetic acid and 0.1M NaCl is not very soluble and will not be stable in solution thereby eventually crystallizing out of solution. Due to this instability, a modification was made to the sample preparation procedure.

In order to use a lower concentration stock solution of L-Arginine without causing a significant dilution effect to the 1mM insulin solution, the concentration of the starting solution of insulin was increased in to 1.5mM, and the volume of initial insulin solution was reduced by approximately 40%. Additionally, the stock concentration of L-Arginine was reduced to 1.5M. By

doing so, a larger volume of stock solution can be added to the sample in order for the sample to contain both 1mM of insulin and 0.5M of L-Arginine and have a total volume of roughly 1mL. These changes ultimately bypass any problems with solubility or crystallization of the stock solution.

3.2. Sample Analysis

Data acquisition involved first measuring the intensity of the laser beam and entering that and other information about the sample such as the sample temperature into the DLS software. The count rate data and other parameters such as diffusion coefficient and polydispersity were collected and calculated in real-time by analyzing the autocorrelation function while the sample was running in the DLS apparatus. Combining the autocorrelation function with known theoretical relationships provides useful information such as the diffusion coefficient as seen in Equation 18. Samples were fully prepped immediately before being placed in the DLS apparatus; control samples were brought to room temperature before being placed in the DLS water bath; arginine samples were made immediately before being placed in the DLS water bath (arginine from a stock solution was added to the cuvette via micropipette). The DLS software records scattered photon intensities for set time intervals—for these experiments, intervals of 2 minutes were used. At the end of the interval, various parameters are fully calculated and a CONTIN analysis is completed. For each sample, the DLS software was automated to run for hours (using progressive 2 minute runs) until a significant increase in count rate and effective diameter is observed. These runs are saved and combined together to form a report that shows the changes in various parameters over time from the first time interval to the last. The CONTIN analysis, after the completion of the time interval, can be graphed using different variables such as intensity, number, and volume versus diameter or diffusion

coefficient. These graphs show the relative populations of all the scattering species. From the data collected, different information about the sample such as the diffusion coefficient, the size of the aggregates, and the average count rate were calculated. The lag time, that is the time it took for the solution to begin aggregation into large-diameter particles, was calculated by comparing when the count rate and baseline difference began to show rapid increases.

In order to standardize the method of obtaining the lag time, the count rate plots of all runs were edited and fit to the following expression:

$$\boldsymbol{Count\ Rate} = e^{a(t-t_0)}, \quad \text{[Eq. 32]}$$

where count rate is the dependent variable, time (t) is the independent variable, and t_0 is the time value representative of the lag time for the aggregation. The variable *a* represents some experimental factor. By fitting the count rate plots to this equation, the lag time, t_0, can be calculated for each condition of aggregation and be compared to other conditions less arbitrarily.

4. Results

4.1. Preliminary Studies

Before analyzing insulin, initial studies were done using polystyrene latex spheres (PLS) of known diameter. These runs were used to check that the DLS system was properly aligned and that the data collected by the DLS software was being analyzed correctly. A solution of 0.089 micron (89 nm) and 0.53 micron (530 nm) PLS, mixed together in known amounts, was placed into the DLS apparatus. Using the CONTIN analysis provided through the Brookhaven DLS Software, the diameters of the solutes in the sample can be ascertained. Figure 8 shows the CONTIN histogram that displays the calculated diameters of the solutes.

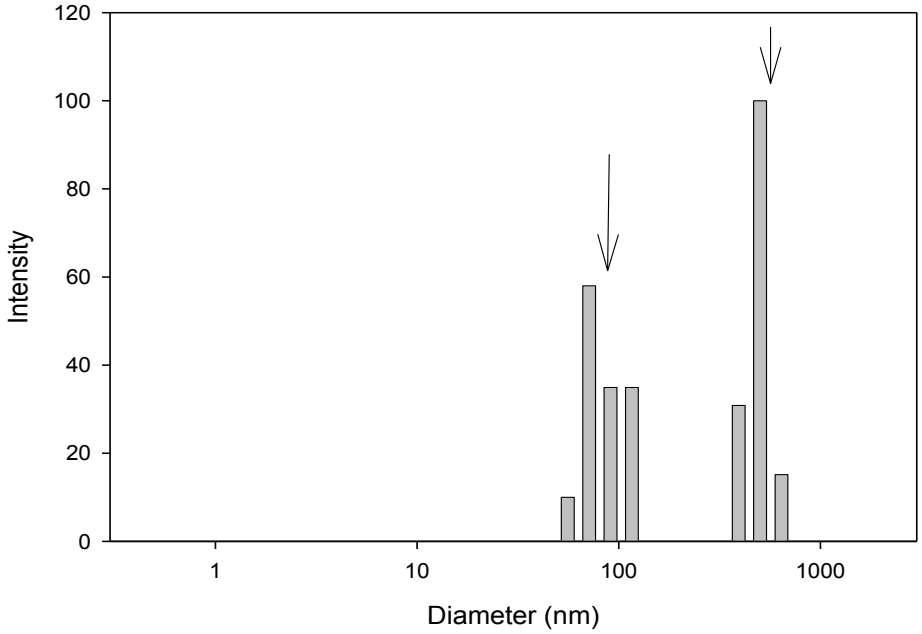

Figure 8. Mix of polystyrene latex spheres. CONTIN histogram illustrating the calculated diameters of the PLS in the sample. The arrows indicate the true diameters of the PLS used, 89 nm and 530 nm, respectively.

The CONTIN analysis shows two groups of scatterers, one averaging between 80 and 90 nm and another averaging between 500 and 600 nm. Assuming a baseline count of over 10^6, CONTIN analyses can be used to accurately determine the mean diameter of scatterers in a sample.

In addition to CONTIN, the DLS software also graphs an autocorrelation function as data is being collected. The autocorrelation is a function of τ, or the delay time. In the case of a sample with uniformly sized scatterers, a single exponential fit is used to extract useful information from the raw data as seen in Figure 9.

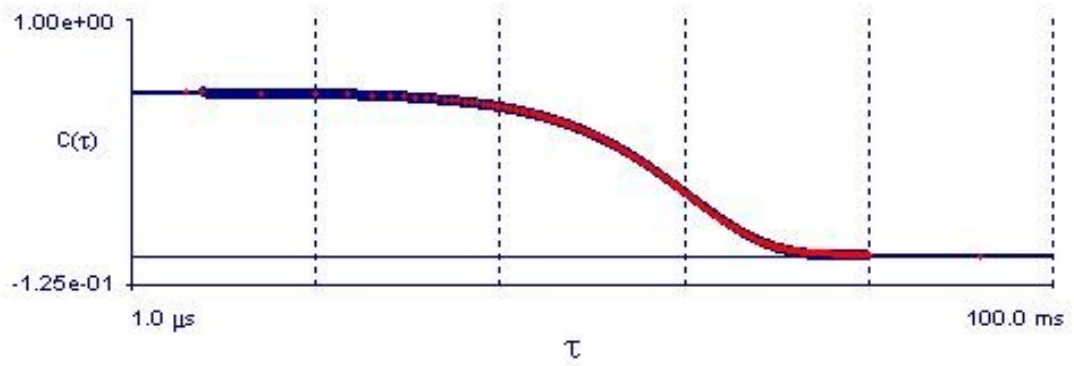

Figure 9. The log-log plot of the autocorrelation function of a solution of 0.53 micron polystyrene latex spheres.

The autocorrelation function can also be plotted as the log of the autocorrelation function versus the delay time, τ, as seen in Figure 10. The autocorrelation function fits a linear regression when plotted as a log-linear plot. This shows the relationship that was described earlier in Equation 18. The slope of this linear fit is the decay constant of the autocorrelation function for this sample.

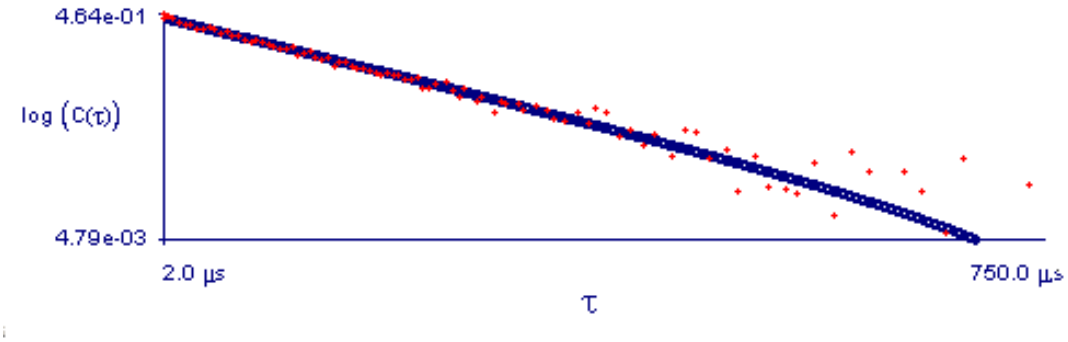

Figure 10. The log-linear plot of the autocorrelation function of a solution of 0.089 micron polystyrene latex spheres.

The diffusion coefficient that was extracted from the decay constant of the autocorrelation function seen in Figure 10 of the 0.089 micron PLS was $(5.5 \pm 0.04)\ x10^{-8}\ cm^2sec^{-1}$. This diffusion coefficient yields an effective diameter (twice the hydrodynamic radius) of 82.5 nm. The

nominal diameter (given on the sample) of the PLS is 0.089 microns or 89 nm. This is only a difference of about 7.9%.

The polydispersity of the 89 nm PLS sample is another factor that is analyzed by the Brookhaven DLS software. For this sample, the polydispersity parameter was determined to be 0.059. As mentioned previously, the polydispersity parameter indicates the diversity of species of scatterers in a sample. Since there is only one species of scatterers in the sample (the 89 nm PLS), the polydispersity parameter was expected to be close to zero and was in fact low.

Another factor to observe in the 89 nm PLS sample is the range of delay times. The autocorrelation function has 2 orders of magnitude from 15 microseconds to 40 milliseconds. Since the scatterer species is relatively small, it is expected that the correlation between the intensity information of the scatterers would decay quickly since smaller scatterers will move more quickly (undergo more Brownian motion) in a short period of time compared to a larger scatterer. This is seen when comparing Figure 9 with Figure 10. Figure 9 is the plot of a 530 nm PLS sample, which has 5 orders of magnitude in the delay time, and Figure 10 is the plot of an 89 nm PLS sample, which has only 2 orders of magnitude. Since the 530 nm PLS are larger than the 89 nm PLS, the 530 nm PLS will undergo Brownian motion slower than the 89 nm PLS, therefore the correlation of the intensity information for the 530 nm PLS will decay slower (correlation times differ by a factor of 6 between the 530 nm and 89 nm PLS samples, which is related to the ratio of the PLS sizes) compared to the 89 nm PLS samples.

Since the solutions are monodisperse, a single exponential fit is sufficient; however, if the solution is polydisperse, then a multi-exponential fit is used. This is seen below in Figure 11 with a sample of 1mM insulin in its standard solvent.

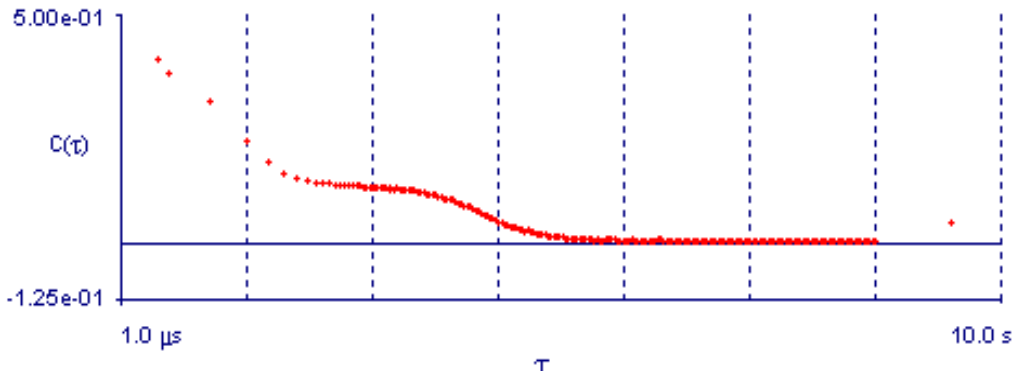

Figure 11. The log-log plot of the autocorrelation function of a solution of 1mM insulin.

The polydispersity of this insulin sample is due to the nature of insulin. It can be in many forms: monomers, dimers, oligomers, fibrils, and clusters of fibrils. Since there are many forms, each with its own characteristic size, the sample will most likely possess a mix of these forms. This mixture is what causes the polydispersity. Since these forms have different sizes, they undergo Brownian motion differently, which causes a multi-exponential regression to the autocorrelation function.

The polydispersity of the insulin sample with an autocorrelation function seen in Figure 11 would be expected to be higher than that of both the 530 nm and 89 nm PLS samples, which were monodisperse. The polydispersity of the insulin sample is over 1 as opposed to the polydispersities of the PLS samples which are much lower than 1.

In terms of the delay time range, the insulin sample has 7 orders of magnitude from 1 microsecond to 10 seconds. This shows how insulin is a much larger scatterer than the PLS – in particular the aggregate form - . Since the insulin aggregate is much larger (on the order of 10^3 nm) than the PLS, the delay time decays at an even slower rate than both the 530 nm and 89 nm PLS samples.

4.2. Insulin (Control) Samples

After the initial studies using PLS, 1mM insulin samples were run as control samples at different temperatures. As seen below in Figure 12, the initial aggregation of insulin follows an exponential growth. In order to characterize the lag time, the plots for each sample were fit to Equation 32 after having their count rates adjusted to a baseline of 0. With the exception of the higher temperature runs, which seem to congregate near one another, the control runs seem to follow a trend.

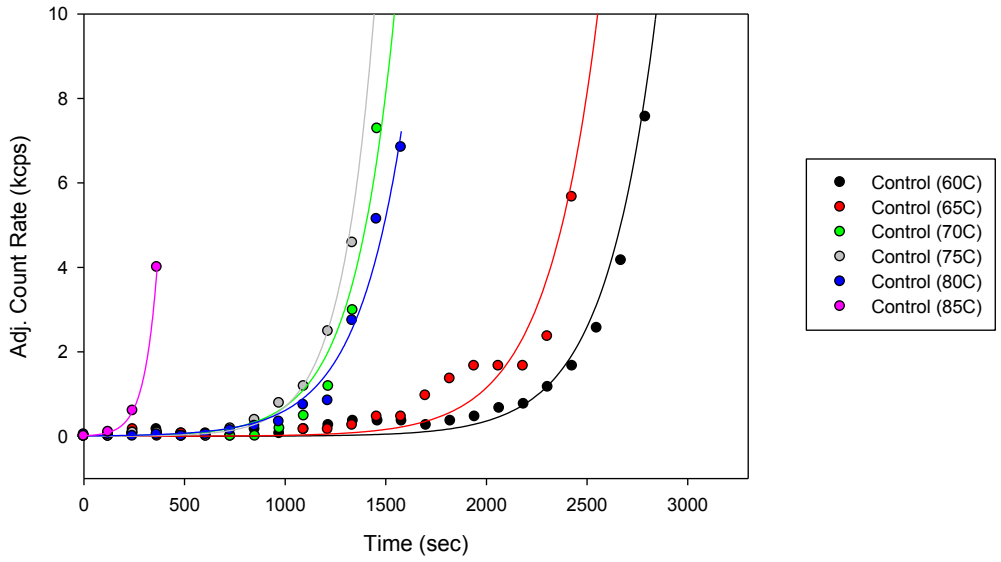

Figure 12. Count rate regressions for insulin control samples. The regression plot for the count rates of the control samples (1 mM insulin) at different temperatures, which were fit to an exponential growth equation (Eq. 32).

It is important to note that this analysis focuses on the initial growth rates of the insulin samples. As seen in Figure 13 below, the growth rate of the insulin aggregates changes over time and no longer follows an exponential growth (in this example, aggregates must have settled out of solution over the course of the aggregation).

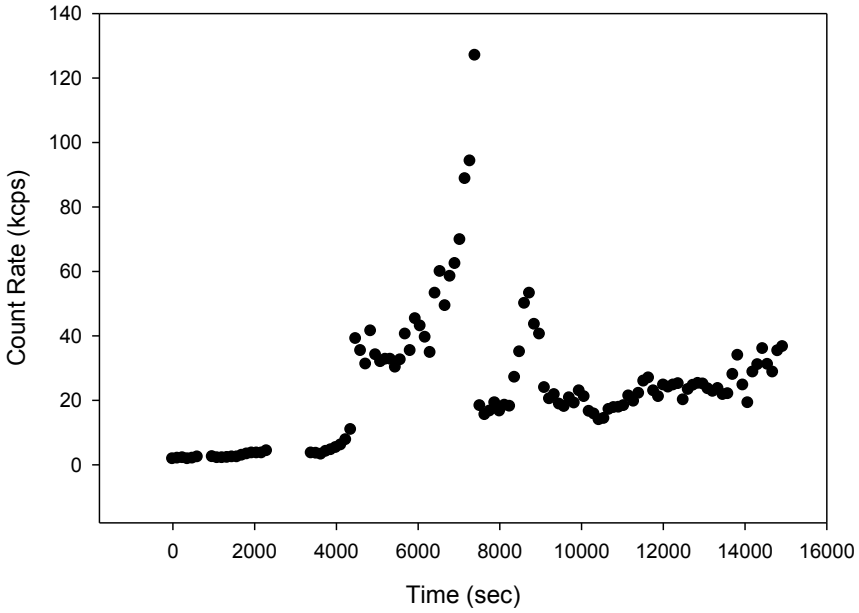

Figure 13. The count rate plot for an insulin control sample at 65°C. This plot shows the changes in growth rate of the insulin aggregates over a span of 4 hours—the initial growth occurred in the first 1.5 hours (~5400 sec).

After extracting the lag time from the fitted equation, the lag times were plotted, as seen in

Figure 14.

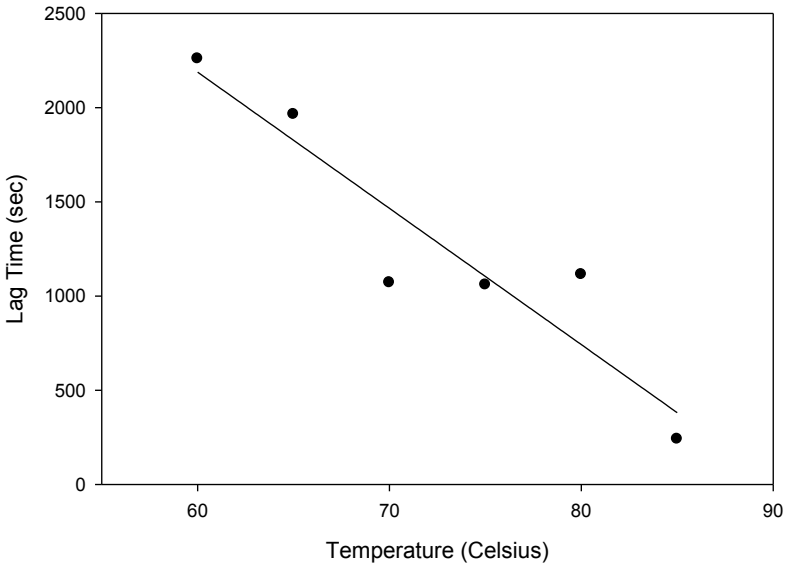

Figure 14. Lag time vs. temperature of insulin control samples. The plot of the lag times for the control samples over a range from 60°C to 85°C.

The trend suggested by the count rate graph is further supported by the lag time data, which shows that as temperature increases, the lag time decreases at a relatively proportional rate. This temperature dependence of insulin aggregation has been documented before in a previous Union College Honors Thesis regarding insulin and DLS (Mahmood, 2008).

In addition to the lag time, the exponential fits for the insulin controls (Eq. 32) also yield an experimental factor, α. This α value is related to the rate of intensity increase and it seems to have a temperature dependence (Figure 15). This factor can be compared to the α values of the arginine runs (below) to see if there are any sort of relationships between them.

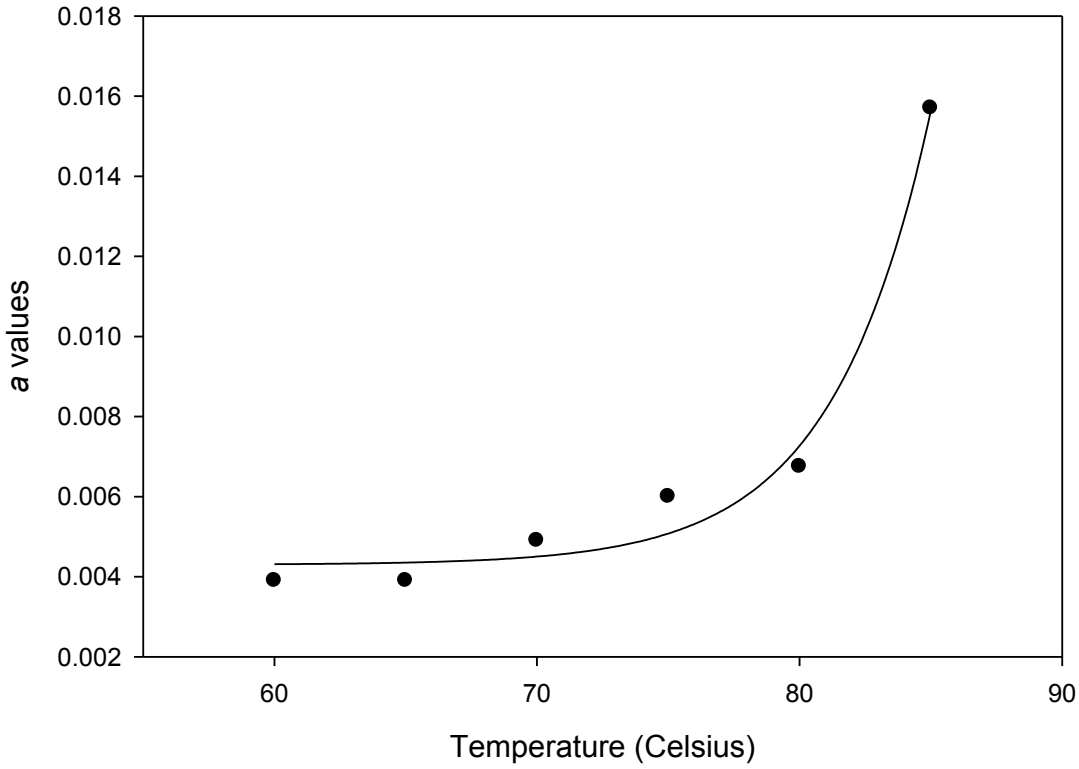

Figure 15. Experimental factor a vs. temperature of insulin control samples. An exponential fit best describes the temperature dependence of this factor.

In addition to looking at photon count rate data, the Brookhaven DLS software also calculates other parameters such as the diffusion coefficient, as described by Equation 23, in real-time. The diffusion coefficient can be used to generate other metrics such as the hydrodynamic radius, shown through Equation 24, and doubled to yield the effective hydrodynamic diameter. The effective diameter data for the same insulin control samples shown in Figure 12 are shown below in Figure 16.

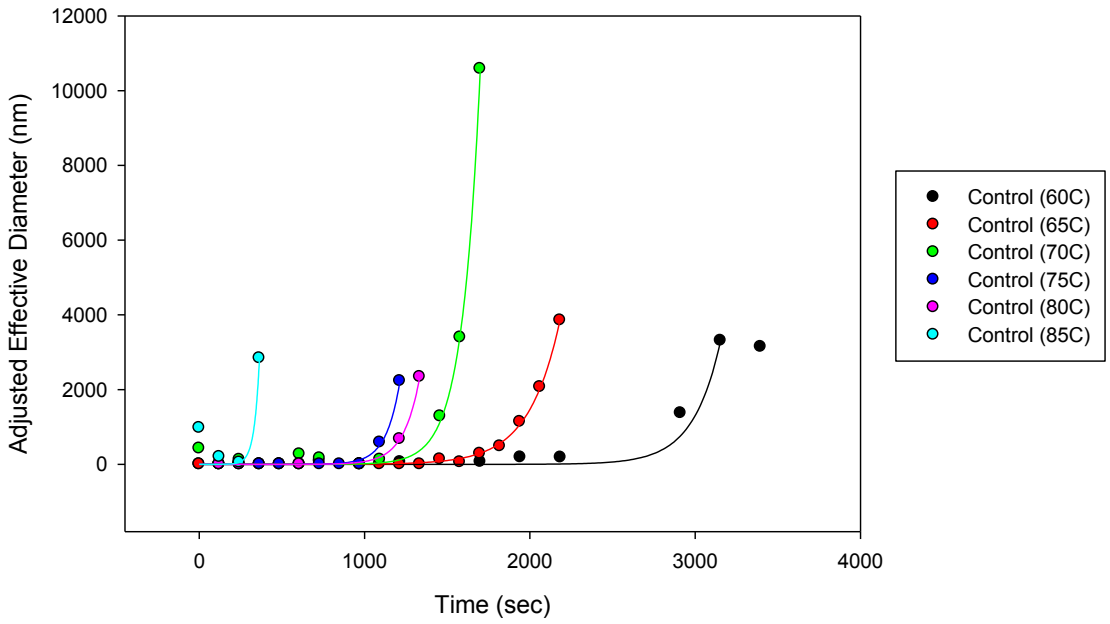

Figure 16. Adjusted effective diameter plots for insulin control samples. The time dependence plots for the effective diameter of the control samples (1mM insulin) at different temperatures, which were adjusted and fit to an exponential growth equation (Eq. 32).

The relationships and trends seen in the count rate plots are also seen in the effective diameter plots. As seen in Figure 16, the plots start at a baseline adjusted effective diameter (which was normalized to start at 0) and increased to larger-sized aggregates; the existence of this baseline shows how there are no large-sized aggregates early on before the lag time of the aggregation. The effective diameter plots were fit to the same exponential growth equation as the count rate plots with the replacement of the count rate parameter with the effective diameter parameter. Since the effective diameters are calculated values, these data plots are much noisier than count rate plots when trying to find trends and fit to smooth curves. The effective diameter plots require higher degrees of editing by checking suspected outlier data points against background noise to better filter the data plot. Nevertheless, the effective diameter plots offer an independent source to verify the lag times obtained through the count

rate plots. As seen in Figure 17, the lag times obtained from the effective diameter are compared to the lag times obtained from the count rates. The trend seen in the count rate lag times was seen also in the effective diameter lag times—this is supported by the similarities in the slopes of each regression, which differed by 19%. The lag times are fairly similar with an average difference of 485 sec. between the two plots. This comparison shows how the lag times for insulin samples can be obtained through two independent methods of analysis.

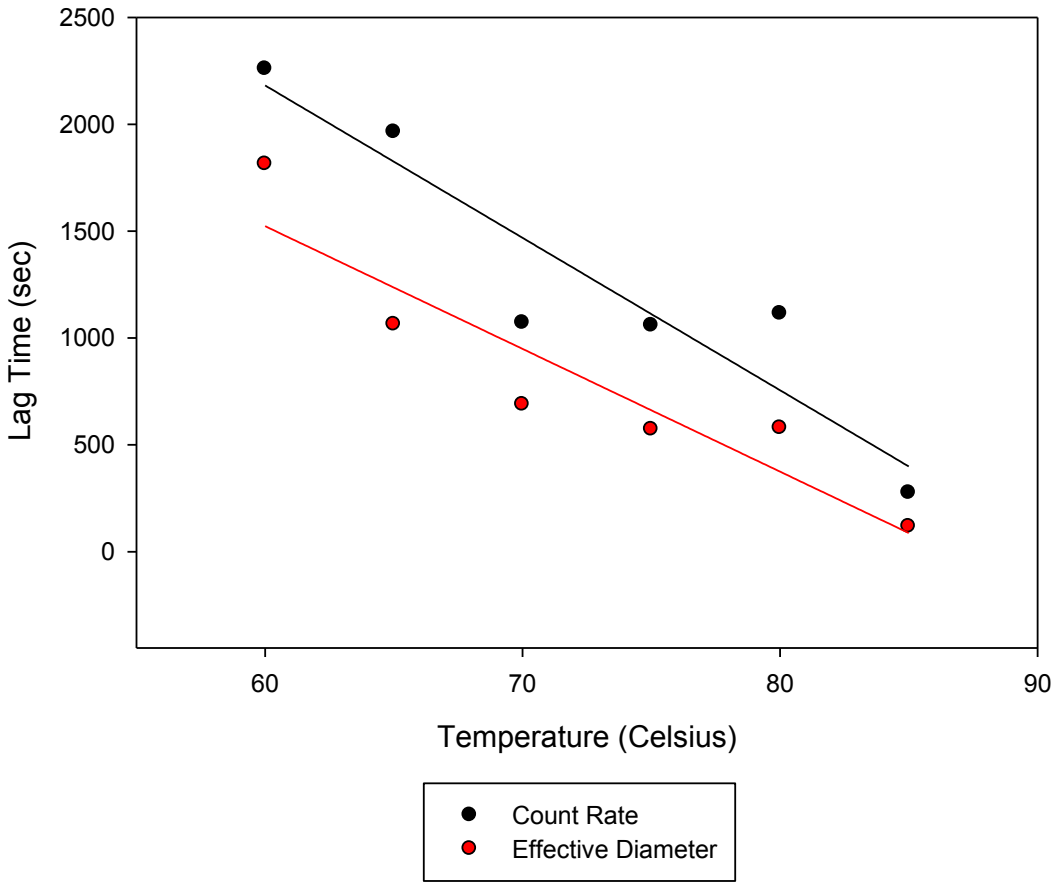

Figure 17. Comparison of the lag times calculated via count rate plots and effective diameter plots.

Since the effective diameter plot was adjusted in a similar manner to the count rate plot, the exponential fit of the effective diameter data (described by Eq. 32) allows for the extraction

of the experimental factor, *a*. Figure 18 shows the comparison of the *a* factors from both the count rate and effective diameter plots. The count rate data fits to a 3-parameter exponential equation (including a y-intercept), while the effective diameter data fits to a 2-parameter exponential equation (without a y-intercept). Despite the differences in the fits of the two plots, *a* seems to hold its temperature dependence; as temperature increases, *a* increases as well.

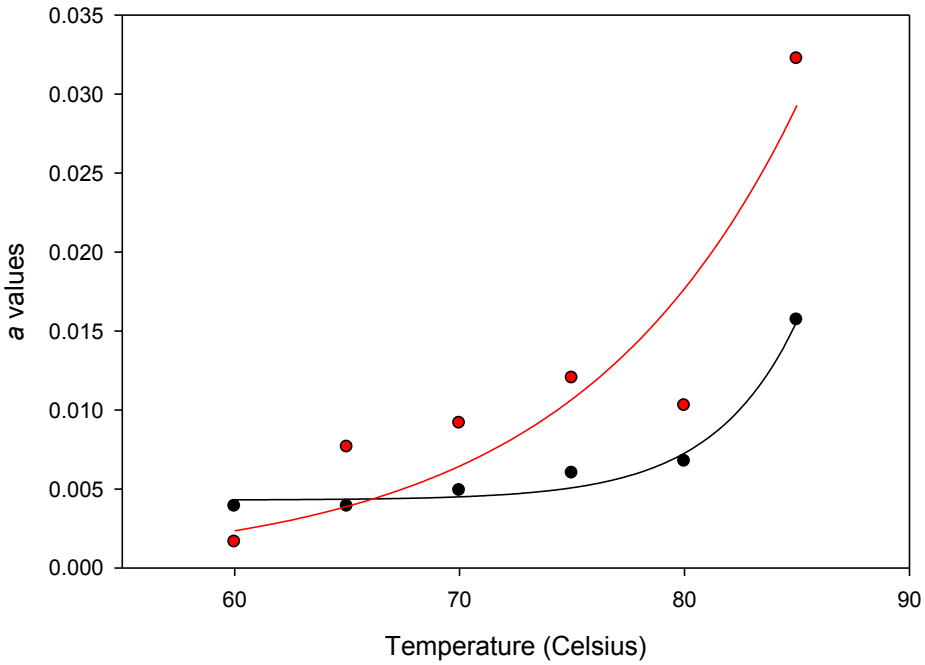

Figure 18. Comparison of the experimental a factor extracted from the exponential fits of the count rate data (shown in black) and the effective diameter data (shown in red).

Table 1 below shows a summary of the values extracted from the exponential fits of the insulin control plots.

Table 1. Summary table of the lag times and *α* values from the insulin control samples over various temperatures.

Temperature (Celsius)	Number of samples	Count Rate Lag Time (sec)	Effective Diameter Lag Time (sec)	Count Rate *α* values	Effective Diameter *α* values
60	2	2300 ± 46	1800 ± 52	(4.3 ± 0.5) x 10^{-3}	(2.0 ± 0.5) x 10^{-3}
65	1	2000	1100	3.9 x 10^{-3}	7.7 x 10^{-3}
70	4	1400 ± 200	900 ± 230	(4.8 ± 0.5) x 10^{-3}	(6.9 ± 3.2) x 10^{-3}
75	2	1100 ± 88	600 ± 95	(6.5 ± 0.7) x 10^{-3}	(1.3 ± 0.1) x 10^{-2}
80	2	1100 ± 42	550 ± 43	(6.6 ± 0.3) x 10^{-3}	(1.01 ± 0.03) x 10^{-2}
85	2	280 ± 25	98 ± 27	(1.32 ± 0.04) x 10^{-2}	(2.96 ± 0.04) x 10^{-2}

4.3. L-Arginine Concentration Dependence

Based on prior research, L-Arginine has been shown to influence the protein aggregation process (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1301). In this work, we have studied the effects of L-Arginine concentrations over a range from 10mM to 500mM. As seen in Figure 19, the count rate curves follow a specific trend; as the concentration of arginine increases, the lag time of the insulin aggregation increased as well. This increase in lag time, presumably caused by the presence of arginine, suggests that arginine does indeed influence the aggregation of insulin by increasing the amount of time between different stages in the aggregation pathway of insulin.

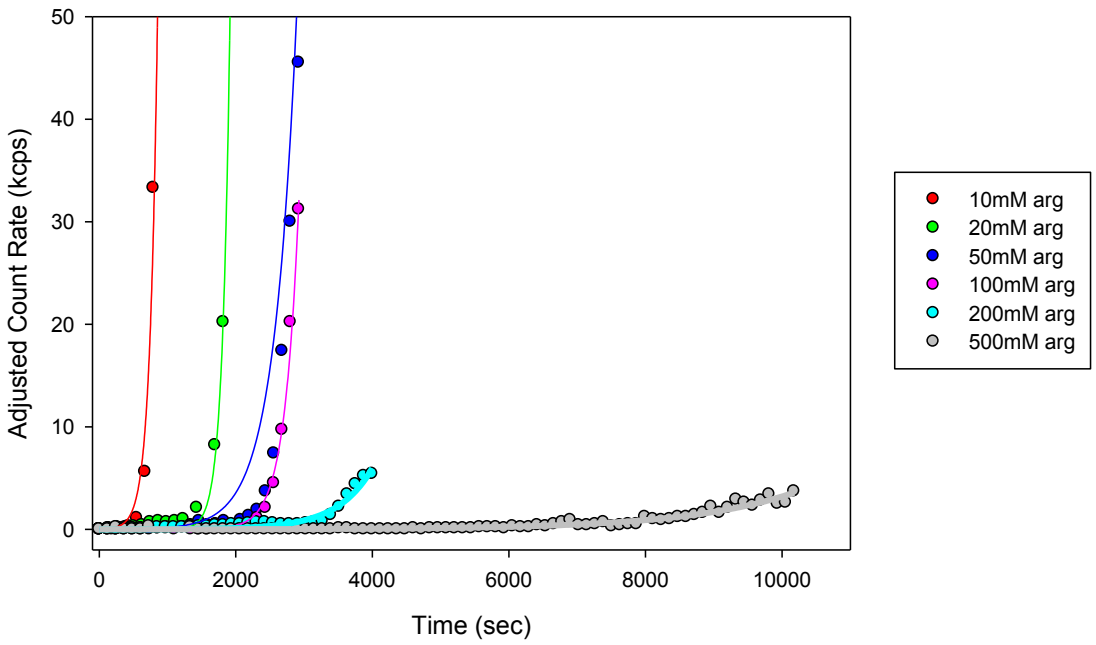

Figure 19. Concentration-dependent count rate data and fits to 18. The different data are for 1mM insulin samples with different concentrations of L-Arginine all measured at 70oC. The count rates have been slightly adjusted to all start at zero and then fit to 18 shown as the smooth lines. The fits at the two highest arginine concentrations are shown with thicker lines for clarity.

It is important to note that at high arginine concentrations, the count rate did not go up as dramatically as seen with the samples with lower arginine concentrations. One reason for not observing a high count rate increase in the higher arginine concentration samples is because insulin aggregates sedimented to the bottom of the cuvette instead of forming a stable gel-like matrix within the solution.

As seen in Figure 3, the aggregation of insulin starts from smaller monomers and dimers to oligomers and finally to long fibrils and fibril systems. In arginine samples with concentrations of 100mM or higher, insulin aggregation was different at the later stage of aggregation from that seen in the insulin controls and low arginine concentration samples, both of which form a gel-like matrix of fibrils at the later stages of insulin aggregation. The aggregation seen in these

high arginine concentration samples did not lead to a gel-like matrix at the end of the aggregation pathway of the insulin. Instead, the aggregates precipitated to the bottom of the cuvette or collected at the bottom of the meniscus, due to surface tension. These aggregation characteristics may suggest that the aggregation of the insulin in these high arginine concentration environments may include the formation of long fibrils or bundles that form very quickly which leads to sedimentation of these aggregates instead of further interactions between these fibrils into the normal gel-like matrix.

In order to verify the lag time data obtained from the count rate plots, plots of the change in effective diameter were used (Figure 20). The effective diameter plots were used to extract lag times for the aggregation of each sample; these lag times are independent of the lag times obtained via the count rate plots.

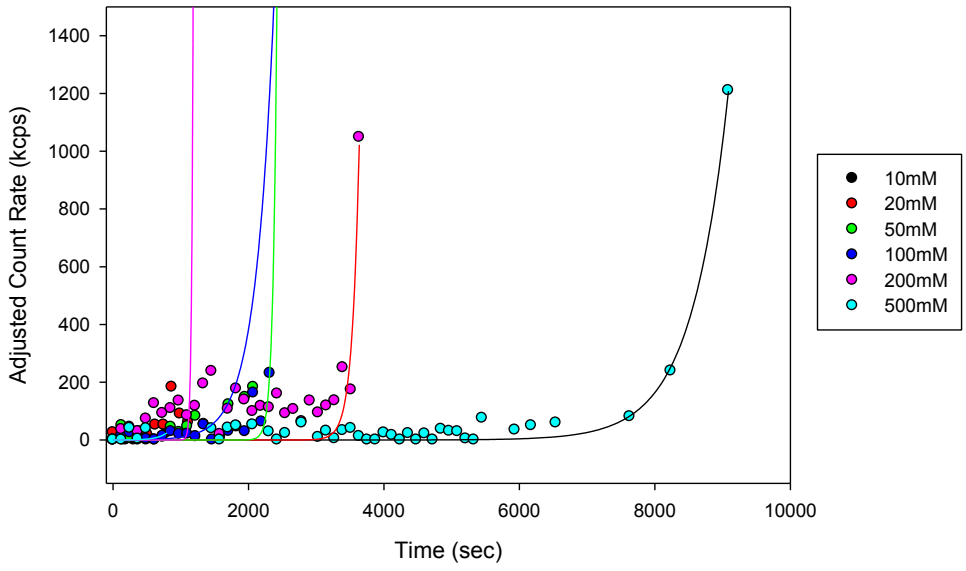

Figure 20. The effective diameter plots for 1mM insulin samples with different concentrations of arginine.

Once again, by extracting the lag time from the exponential fit of the count rates and effective diameters, the trends in L-Arginine concentration and lag time can be determined.

Figure 21 plots the lag times against the concentrations of arginine for both sources (count rate plots and effective diameter plots).

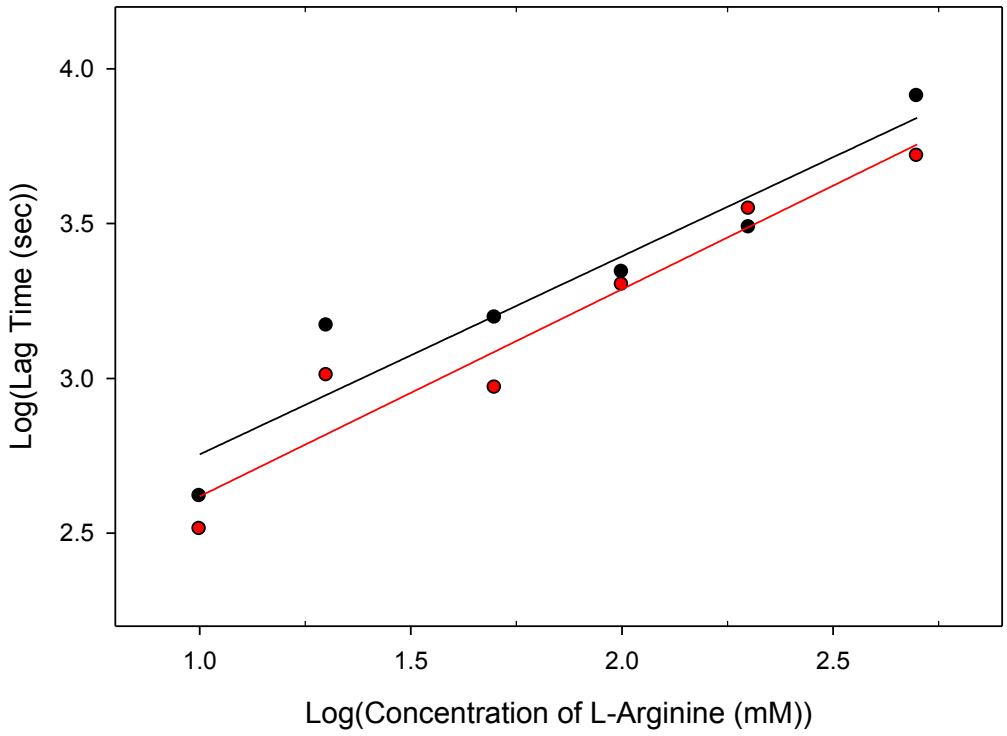

Figure 21. Lag time of different concentrations of L-Arginine samples for 1mM insulin at 70°C.The black data represent the count rate data while the red data represent the effective diameter data.

The plot shows a positive relationship between the log of arginine concentration and the log of the lag times; as the concentration of L-Arginine increases, the lag time increases as well. The linear regressions for the count rate plot and the effective diameter plot are

$$Log\ (Count\ Rate\ Lag\ Times) = 2.1 + (0.64) * Log(Arg\ Conc.)\text{,} \quad \text{[Eq. 33]}$$

and

$$Log\,(Effective\ Diameter\ Lag\ Times) = 1.9 + (0.67) * Log(Arg\ Conc.)$$ [Eq. 34]

The slopes of these fits differ by about 4.6% from each other. The two fits seem to be parallel with the effective diameter fit being offset by small factor (about 0.2 based on the fits). The similar slopes are expected since the two plots are analyzing the same samples. The shift in the effective diameter fit is most likely due to the noise of the effective diameter and the higher sensitivity of the effective diameter data in comparison to the count rate data to changes in the intensity of scattered photons analyzed by the DLS apparatus.

The extrapolation of the linear trend to an arginine concentration of 0 results in a lag time of approximately 670 sec., which is a 60% difference from the previously determined lag time of 1100 sec for a 1mM insulin control sample at 70°C. This difference may indicate that a linear extrapolation from the lower limit of this data set is a poor representation of the true trend at low arginine concentrations. This result verifies the previous studies that have also determined that increased arginine concentrations delay protein aggregation (Tsumoto, Umetsu, Kumagai, Ejima, Philo, & Arakawa, 2004, p. 1301).

Table 2 summarizes the calculated values for the arginine concentration experiments.

Table 2. Summary of lag times for each of the different concentrations of arginine at 70°C.

Arginine Concentration (mM)	Number of samples	Count Rate Lag Time (sec)	Effective Diameter Lag Time (sec)
0	4	1400 ± 200	900 ± 230
10	1	420	330
20	2	1600 ± 110	1100 ± 170
50	2	1300 ± 420	1000 ± 130
100	3	2400 ± 140	2100 ± 88
200	2	3000 ± 160	3000 ± 710
500	2	8900 ± 1000	5700 ± 610

4.4. Temperature Dependence

For the insulin control samples, the lag time of the insulin aggregation decreased as the temperature increased (Figure 14). The same trend is seen with the insulin samples that contained arginine, which were all 500mM in concentration, over a range of 60°C to 85°C. The lower temperature limit of 60°C was established because the aggregation of insulin at temperature lower than 60°C takes a long time to occur (on the order of many hours to days or more). This long lag time was observed previously (Mahmood, 2008). The upper temperature limit of 85°C was established because the external heating apparatus used to heat the water bath of the DLS apparatus was limited to 85°C.

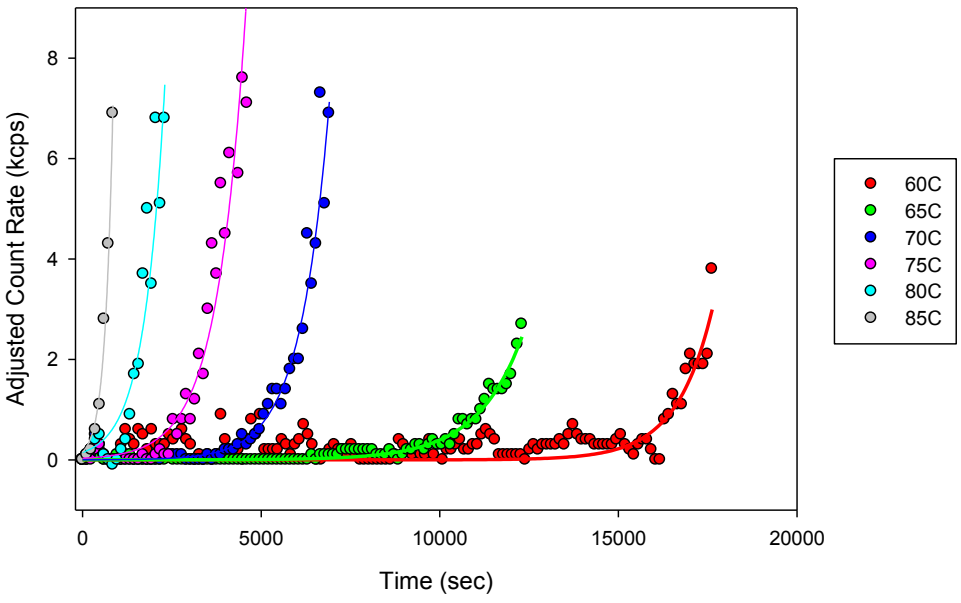

Figure 22. Temperature-dependent count rate data. The count rates for insulin samples with 500mM arginine at different temperatures ranging from 60°C to 85°C.

The lag time plot in Figure 22 seems to suggest an inverse relationship between temperature and lag time. The red regression curve represents the trend of the 500mM arginine samples,

while the black line represents a linear fit to the insulin control sample data. The arginine data fit an exponential equation reasonably well with

$$Lag\ time\ (sec) = (1.5\text{x}\ 10^{7}{}_{\mathbf{sec}}) * e^{-(0.11)*(T(°C))}, \quad \text{[Eq. 35]}$$

The linear regression of the control samples is represented by the linear equation:

$$Lag\ time = (6500\ \text{sec.}) - 72 * (\text{Temp}(°\text{C})), \quad \text{[Eq. 36]}$$

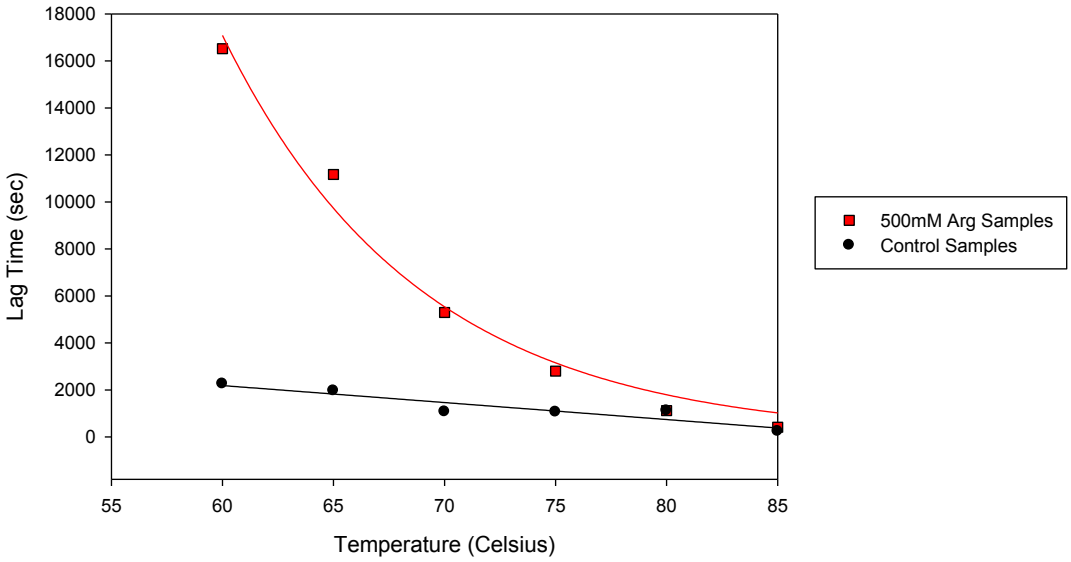

Figure 23. Lag time of 500mM L-Arginine samples and control samples at different temperatures.

Since the lag times for the arginine samples follow an exponential decay and the lag times for the control samples follow a linear trend, the ratio of lag times of arginine samples compared to control samples follows an exponential decay as seen in Figure 24. The ratio is described by the equation below:

$$Ratio\ of\ lag\ times = (-3.0) + (99) * e^{(-0.04)*(Temp\,(°C))}, \quad \text{[Eq. 37]}$$

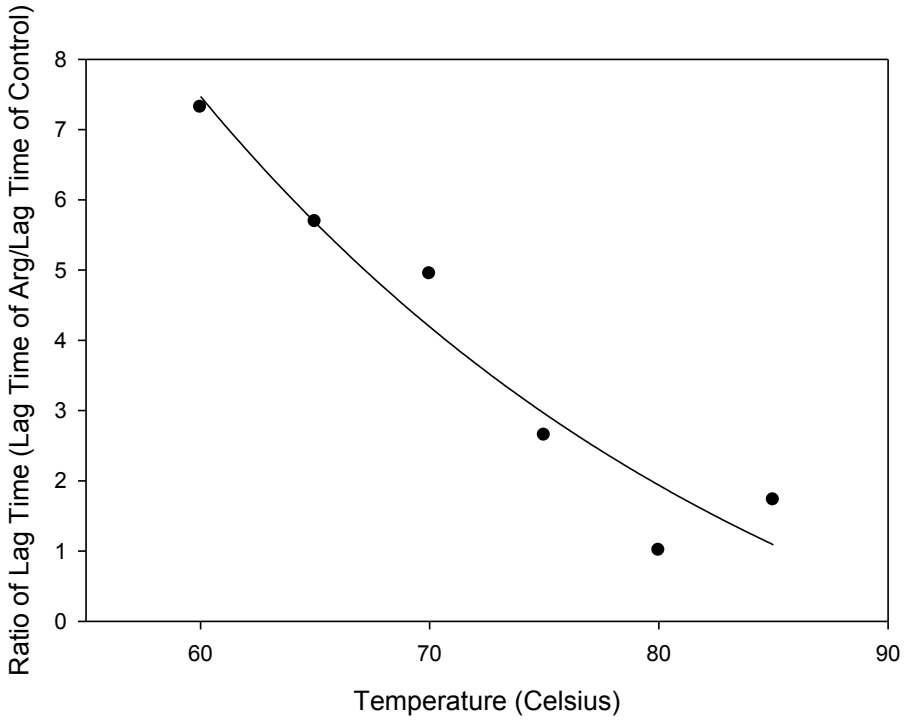

Figure 24. Ratio of lag times between arginine samples and control samples.

Figure 24 shows the effectiveness of the L-arginine in inhibiting aggregation as a function of temperature. Using both Figure 23 & Figure **24**, it may be possible to extrapolate the data to other temperatures outside the experimental range. In particular, extrapolation to 37°C, predicts a tremendous inhibition in the presence of L-arginine. Running actual DLS experiments at this temperature, however, is problematic because the aggregation of a 1mM insulin sample would normally take a day or more to fully aggregate. If arginine does in fact inhibit tremendously at this temperature, then it would most likely take several days in order to observe any aggregation. In that length of time, it is difficult to maintain the stability of the sample temperature due to issues in running the water heating unit for a prolonged period of time. Also, the power of the argon-ion laser could experience significant fluctuations over a long period of time. These issues would undoubtedly lead to reproducibility problems over several days time.

As with the other sample types, the temperature-dependent arginine plot can be independently verified by looking at the effective diameter plot of the same sample.

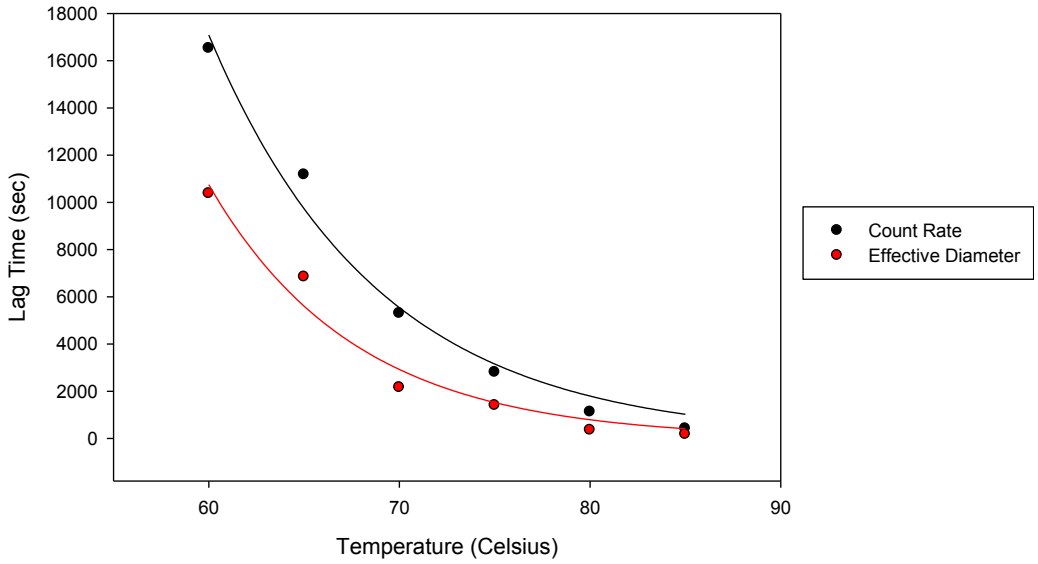

Figure 25. Comparison of the lag times for the count rate plot and the effective diameter plot for the 500mM arginine runs over temperatures ranging from 60°C to 85°C.

As seen in the other comparison plots, the effective diameter verifies the trend found in the count rate plot. It is interesting to note that for all the comparisons of count rate and effective diameter, the effective diameter data seem to always indicate a shorter lag time than the comparable count rate data. This may be due to a characteristic of DLS; the DLS software seems to be able to detect changes in hydrodynamic radius faster (or more sensitively) than the increase in the count rate. If the effective diameter is observe to increase before the count rate, then it would account for the observation of the effective diameter plots being below the count rate plots.

Table 3 below provides a summary of the lag times determined for 500mM arginine samples over various temperatures.

Table 3. Summary of lag times for 500mM arginine samples using the modified procedure.

Temperature (Celsius)	Number of samples	Count Rate Lag Time (sec)	Effective Diameter Lag Time (sec)
60	1	17000	10000
65	1	11000	6800
70	2	5200 ± 150	2300 ± 160
75	2	2800 ± 47	1400 ± 37
80	2	1600 ± 610	611 ± 360
85	1	420	170

4.5. Arrhenius Plots

By using the exponential fit (Eq. 32) to extract a standardized lag time for each sample (both arginine and control) at different temperatures, an Arrhenius analysis was done. An Arrhenius analysis is used to evaluate the relationship between a rate (usually the reaction rate of a process) and temperature. An Arrhenius relationship follows the general form

$$Reaction\ rate = Ae^{-E_a/RT}, \quad \text{[Eq. 38]}$$

where A is a frequency factor, E_a is the activation energy of the reaction, R is the gas constant, and T is the temperature in Kelvin. An Arrhenius plot of both the control samples and the arginine samples are illustrated below in Figure 26.

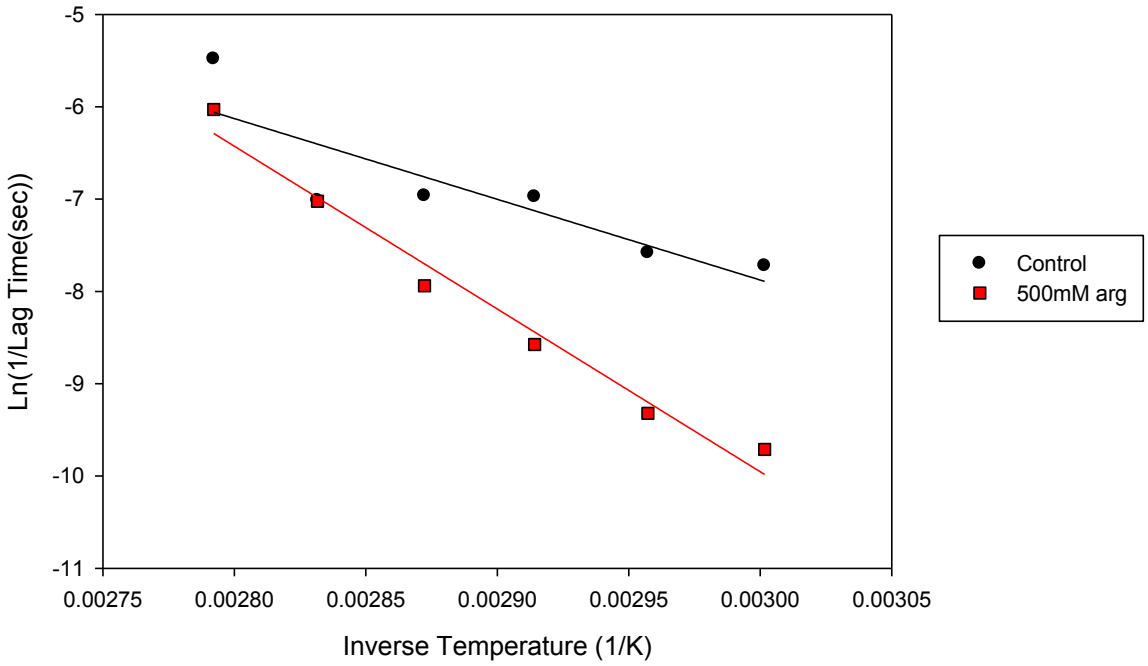

Figure 26. Arrhenius plots of insulin samples with and without L-Arginine.

By plotting the natural log of the inverse of the lag time against the inverse of temperature, the linear fit of the data will yield information about the activation energy for each sample type as seen below.

$$Ln\left(\frac{1}{Lag\ time}\right) = Ln(A) + \left(\frac{-E_a}{R}\right) * \left(\frac{1}{Temperature\ (K)}\right), \quad \text{[Eq. 39]}$$

The control samples fit to the following equation:

$$Ln\left(\frac{1}{Lag\ time\ of\ control\ samples}\right) = (18) + (-8800) * \left(\frac{1}{Temperature}\right), \quad \text{[Eq. 40]}$$

The arginine samples fit to this equation:

$$Ln\left(\frac{1}{Lag\ time\ of\ arginine\ samples}\right) = (43) + (-18000) * \left(\frac{1}{Temperature}\right), \quad \text{[Eq. 41]}$$

By using these regressions, the activation energies can be extracted from each sample type. The slope of the linear fit for each sample type is equal to the activation energy divided by k, the Boltzmann constant. Table 4 shows the activation energies obtained from the linear fits of both the insulin controls and the 500mM arginine samples.

Table 4. Table of experimentally determined slopes and activation energies for insulin control samples and 500mM arginine samples.

Sample Type	Slope of fit (E/k)	Energy (kcal/mol)
Insulin control	8800 ± 2600	17 ± 5.2
Arginine (500mM)	18000 ± 1400	26 ± 2.8

5. Discussion

The kinetics of protein aggregation are relevant and important in the fields of protein science and medicine because of its implications in the research and treatment of amyloid diseases such as Alzheimer's, Parkinson's, and Huntington's disease. Insulin, which falls in a class of amyloid proteins, is used as a model system because it has been well researched due to its use in treating diabetes. The aggregation pathway of insulin has been studied and outlined in previous research (Mauro, et al., 2007). Insulin progresses from small monomeric and dimeric structures to oligomeric structures. These oligomeric structures then aggregate further into long fibrils and finally into bundles or clusters of fibrils. This aggregation pathway has many areas in which modifications can be made in order to influence the overall aggregation process.

As illustrated in Figure 11 & Figure 12, insulin aggregation is highly temperature dependent. There is a negative correlation between sample temperature and the lag time before fibril formation in the sample. This relationship has been documented previously using DLS (Mahmood, 2008). This increase in reaction rate, as seen through a calculated decrease in lag time, is most likely due to the increase in the average kinetic energy of the smaller monomeric and dimeric insulin structures, which would increase the rate at which oligomeric

structures would form. The faster this nucleation process, the shorter the lag time becomes. Thus, heat would decrease the lag time by hastening both nucleation and elongation of the insulin fibrils.

Arginine is widely used as an additive for suppressing protein aggregation, assisting in protein refolding, and enhancing protein solubility (Lyutova, Kasakov, & Gurvits, 2007). This research has focused on arginine's effect on insulin aggregation. Figure 13 & Figure 14 show the effect of different concentrations of arginine on the aggregation of samples of 1mM insulin at a fixed temperature of 70°C. Arginine concentrations of 10mM, 20mM, 50mM, 100mM, 200mM, and 500mM were added and observed. Generally, the trend of lag time on arginine concentration was linear and positive. Due to the instability of high concentration stock solutions of arginine in the standard insulin solvent (20% (v/v) acetic acid with 0.1M NaCl), it was difficult to consistently reproduce the general trend. Since reproducibility was an issue, the procedure was modified a few times to try to obtain more reproducible data but was still unsuccessful. Instead of trying to reproduce samples at different arginine concentrations, one concentration was chosen and samples were run at different temperatures. For the temperature runs, a new and successful procedure was developed and used.

A constant concentration of 500mM arginine was used for the temperature dependence experiments. The temperature dependence of aggregation in the presence of arginine appears similar visually to the temperature dependence seen in the insulin controls when viewing the count rate regressions (Figure 15). However, further analysis of the lag times from the arginine samples showed trend that fit an exponential decay (Eq. 33) rather than a negative linear relationship as seen by the insulin control samples (Figure 23). The exponentially decaying ratio between the lag time of the arginine samples and that of the control samples emphasizes the exponential influence of arginine on the aggregation process. Based on these results, arginine

seems to be the most effective at the lower bounds of the temperature range used in these experiments (between 60°C and 75°C). At higher temperatures, the lag times of samples with and without arginine were very similar. As stated previously, the temperature limits used in these experiments were due to both slow aggregation kinetics (at lower temperatures) and mechanical limitations (at higher temperatures). Using the regression of the arginine samples and the ratio of the lag times, the lag time for the aggregation of insulin at body temperature (approximately 37°C) is in the order of 4 to 5 days with the addition of arginine as compared to a 1 or 2 days for a control sample. While extrapolations to 37°C may not be accurate, the trends shown between 60°C and 85°C do suggest that there could be a significant difference between the aggregation rate of insulin with and without arginine.

Arginine is used as an excipient in protein-based biopharmaceuticals because of its ability to suppress protein aggregation (Ghosh, Sharma, & Chattopadhyay, 2009, p. 1135). An excipient is an inactive substance used as a carrier for an active substance. Excipients are used because of the immunogenic nature of aggregated proteins— protein aggregates trigger undesired immune responses in humans. By preventing the aggregation of the active protein, arginine maintains high drug effectiveness and helps keep material costs from rising. However, the actual process of suppression is currently unknown.

Some studies suggest that arginine inhibits aggregation by slowing down the protein-protein interactions (Ghosh, Sharma, & Chattopadhyay, 2009, p. 1135). This slowing of protein interactions is predicted by gap effect theory. Gap effect theory suggests that additives that are larger than water increase the activation energy of protein-protein associations while keeping the energy of isolated proteins unchanged (Ghosh, Sharma, & Chattopadhyay, 2009, p. 1135). In order to observe this possible change in activation energy, an Arrhenius analysis was done for both the control samples and the arginine samples at 500mM (Figure 18). As seen in Equations

38 & 39, the slopes of the linear fits for both Arrhenius plots were different. The slopes of the Arrhenius plots represent the activation energy of their respective reactions. The control samples had a slope of -8800, which translates into 17 ± 5 kcal/mol, while the arginine samples had a slope of-18000, which translates into 26 ± 3 kcal/mol. This increase in activation energy for the arginine samples lends credence to the gap effect theory. Since the only difference between the two sample types is the presence of 500mM arginine, then the presence of arginine is the only explanation for this apparent difference in the activation energies. As seen in Table 5, the activation energies of chemical bonds and associations can vary greatly.

Table 5. Table of bond activation energies for different types of bonds.

Bond Type	Activation Energy (kcal/mol)
Hydrogen bonding in ice	6.7
Peptide bonding (pair)	7.8
Carbon-carbon bond	82
Beta-sheet transition	~0.5/residue (~25 for monomeric insulin)

The final energies of both sample types are on the same order as the activation energy of monomeric insulin in β-sheet formation; this is expected since insulin aggregation is the result of insulin forming β-sheets structures. The change in activation energy between the control and the arginine group is approximately 9 kcal/mol. Since the only difference between the sample types was the presence of arginine, the 9 kcal/mol difference may characterize the aggregation-inhibiting activity of arginine. The increase in the activation energy of the arginine sample suggests that the delay in aggregation may be because of the higher energy requirements necessary for the insulin to undergo β-sheet formation and further aggregation. Since insulin requires more energy to form its normal aggregate structures, the overall aggregation is delayed—as seen by the increase in the lag time of arginine samples when compared to their control counterparts.

In conclusion, this research observed the aggregation process of bovine insulin with and without the presence of L-Arginine over a range of temperatures from 60°C to 85°C. Without the presence of arginine, insulin followed a positive linear relationship between reaction rate and temperature, which was shown by a negative linear relationship between lag time and temperature. In the presence of 500mM arginine, the lag time for insulin aggregation decreased exponentially with increasing temperature. Finally, the Arrhenius analysis of the samples suggests that arginine, by increasing the activation energy of protein-protein interactions, does slow down or delay the lag time of the insulin aggregation.

Works Cited

Ahmad, A., Uversky, V. N., Hong, D., & Fink, A. L. (2005). Early Events in the Fibrillation of Monomeric Insulin. *Journal of Biological Chemistry , 280*, 42669-42675.

Berne, B., & Pecora, R. (1976). *Dynamic Light Scattering.* Hoboken, New Jersey: John Wiley & Sons Inc.

Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M., & Robinson, C. V. (2000). Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Science , 9*, 1960-1967.

Cummins, H. (1974). Applications of Light Beating Spectroscopy to Biology. In H. Cummins, & E. Pike (Eds.), *Photon Correlation and Light Beating Spectroscopy.* Dordrecht, Holland: D. Reidel Publishing Company.

Dahneke, B. (1983). *Measurement of Suspended Particles by Quasielastic Light Scattering.* Hoboken, New Jersey: John Wiley & Sons Inc.

Ghosh, R., Sharma, S., & Chattopadhyay, K. (2009). Effects of Arginine on Protein Aggregation Studied by Fluorescence Correlation Spectroscopy and Other Biophysical Methods. *Biochemistry , 48*, 1135-1143.

Hua, Q.-x., & Weiss, M. A. (2004). Mechanism of Insulin Fibrillation. *The Journal of Biological Chemistry , 279* (20), 21449-21460.

Jaimohan, S., Naresh, M., & Mandal, A. (n.d.). *Crystal Structure of Bovine Insulin- Hexameric Form.* Retrieved November 20, 2010, from Protein Data Bank: http://www.pdb.org/pdb/explore/explore.do?structureId=2ZP6

Krebs, M. R., MacPhee, C. E., Miller, A. F., Dunlop, I. E., Dobson, C. M., & Donald, A. M. (2004). The formation of spherulites by amyloid fibrils of bovine insulin. *PNAS , 101* (40), 14420-14424.

Lyutova, E. M., Kasakov, A. S., & Gurvits, B. Y. (2007). Effects of Arginine on Kinetics of Protein Aggregation Studied by Dynamic Light Scattering and Turbidimetry Techniques. *Biotechnology Progress , 6*, 1411-1416.

Mahmood, B. (2008, June). Using Dynamic Light Scattering to Study the Early Kinetics of Insulin Aggregation. Union College Department of Physics and Astronomy.

Mauro, M., Craparo, E. F., Podesta, A., Bulone, D., Carrotta, R., Martorana, V., et al. (2007). Kinetics of Different Processes in Human Insulin Amyloid Formation. *Journal of Molecular Biology , 366*, 258-274.

Merlini, G., & Bellotti, V. (2003). Molecular mechanisms of amyloidosis. *The New England Journal of Medicine , 349* (6), 583-596.

Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., et al. (2001). Effects of Environmental Factors on the Kinetics of Insulin Fibril Formation: Elucidation of the Molecular Mechanism. *Biochemistry , 40*, 6036-6046.

Shiraki, K., Kudou, M., Fujiwara, S., Imanaka, T., & Takagi, M. (2002). Biophysical Effect of Amino Acids on the Prevention of Protein Aggregation. *Jounal of Biochemistry , 132*, 591-595.

Smith, M., Sharp, J., & Roberts, C. (2008). Insulin Fibril Nucleation: The Role of Prefibrillar Aggregates. *Biophysical Journal , 95*, 300-3406.

Thompson, L. K. (2003). Unraveling the secrets of Alzheimer's amyloid fibrils. *PNAS , 100* (2), 383-385.

Timofeev, V., Baidus, A., Kislitsyn, Y., & Kuranova, I. (n.d.). *Structure of Human Insulin.* Retrieved November 20, 2010, from Protein Data Bank: http://www.pdb.org/pdb/explore/explore.do?structureId=3E7Y

Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Philo, J. S., & Arakawa, T. (2004). Role of Arginine in Protein Refolding, Solubilization, and Purification. *Biotechnology Progress , 20*, 1301-1308.