NAB-MQP6

Protein Dynamics: Probing Biological Processes at the Molecular Level using Hemoglobin

A Major Qualifying Project

Submitted to the Faculty

of the

Worcester Polytechnic Institute

In Partial Fulfillment of the Requirements for the

Degrees of Bachelor of Science in

Biophysics & Physics by:

David Pesce

Jeffrey Sanders

March 12, 2007

In Collaboration with:

Erika Holzbaur, PhD

University of Pennsylvania, Department of Physiology

&

William Royer Jr., PhD

University of Massachusetts Medical School, Department of Biochemistry and Molecular Pharmacology

Nancy Burnham, PhD
Department of Physics
Worcester Polytechnic Institute
Project Advisor

Abstract

Proteins are essential for all cellular processes. Hemoglobin is one of the most well studied proteins and is the standard model for allosteric regulation. Allostery is an important, yet poorly understood, component of biochemistry. Using Normal Mode Analysis (NMA), the dynamics of scapharca dimeric hemoglobin, an allosteric protein, have been simulated and correlate with previous crystallographic data. This indicates that NMA may be a useful tool for predicting early intermediate structures and explaining allosteric regulation at the molecular level.

Table of Contents

Acknowledgements

We would like to first thank Dr. Nancy Burnham for agreeing to work on our biophysics MQP. We would also like to thank Deli Lui for instrumentation help, Erika Holzbaur for the dynein samples, Bill Royer for his extensive knowledge of hemoglobin proteins and Doug Tischer for volunteering, and the AFM Lab for dealing with all of our messes.

Introduction

Hemoglobin is one of the most commonly known proteins. It was one of the first proteins whose molecular structure was solved by x-ray crystallography and among the first proteins to have its amino acid sequence determined. It has been studied for over a century now and has become the model protein for allosteric regulation. Allosteric regulation, or simply allostery, is defined as the regulation of a protein by cooperative interactions in binding ligands to sites that are distant from each other in the protein structure. A ligand is any small molecule that binds a specific site of a protein. In the case of hemoglobin the physiological ligand is oxygen and the binding sites are the heme rings [1, 2].

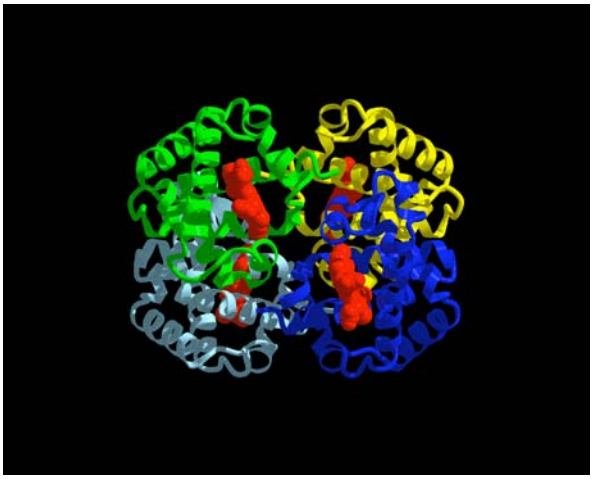

Figure 1. Crystal structure of human hemoglobin. Hemoglobin is a homodimer of heterodimers consisting of alpha and beta monomers. The two alpha monomers are colored light blue and blue while the two beta monomers are colored green and yellow. The heme rings that bind oxygen are labeled red. The heme rings are responsible for giving blood its red color (http://www.psc.edu/science/Ho/Ho-hemoglobin.html).

While the structure and function of hemoglobin are very well characterized, the mechanism of allosteric regulation is still poorly understood. To date there is exist no single model that can account for the allosteric events that occur during oxygen binding. Two competing models are the concerted model and the sequential model. The concerted

model postulates that the subunits of a protein are tightly coupled and a change in one area of the protein will confer a change in all the other subunits. Thus all subunits must be in the same conformation. The model also states that in the absence of a ligand the protein will favor one of the two states and there will always been an equilibrium between the R and T states. The sequential model, on the other hand, states that the subunits are connected but do not adopt the same conformation. It also states that a ligand binds to the protein by an induced fit. Induced fit refers to the active site of the protein having a specific geometry to bind a ligand [3, 4].

To determine if either of these models is valid, one would have to study a simpler system than human hemoglobin to follow the conformational changes during ligation at the molecular level. A simpler system would consist of only two binding sites, so the communication between only the two sites can be followed. A solution to this problem is presented in an invertebrate mollusk hemoglobin protein. Scapharca inaequivalis hemoglobin (HbI) forms a dimer with only two heme rings that can bind oxygen. The location of the heme rings are close to one another, making the conformational changes small intermolecular motions instead of global domain motions [5].

Scapharca hemoglobin has been studied for several years, but questions remain to understanding its allosteric function. While biologists are able to follow the ligation and subsequent conformational changes at atomic resolution, they are still unable to elucidate a mechanism for explaining allostery. Only recently with the incorporation of molecular dynamics simulations have researchers been able to probe protein structure and dynamics in high detail where experimentation falls short. Normal mode analysis, an elastic rod-modeling program, has been applied to several proteins to predict domain motions.

Normal mode analysis has already been used to predict the motions of several proteins, and has been found in many cases to be accurate with experimental data. If HbI's motions can be predicted by normal mode analysis, the results may be able to provide a better understanding of allostery at the molecular level [6, 7].

1.1 Allosteric Regulation

Allostery, which comes from the Greek words *allos* meaning other and *stereos* meaning shapes, is the regulation of a protein by binding a ligand. If one site of the protein binds a ligand, this will cause a conformational change that will affect the binding at another site on the protein. Allosteric regulation can either cause positive feedback control or negative inhibition. This is a vital process for cell survival as metabolism has been shown to be regulated by allostery [4].

Cooperativity can also be described quantitatively. Since proteins need to bind and then release certain molecules, it is important to understand how fast these reactions occur. If the protein has *n* binding sites then the reversible binding can be described by an equilibrium expression

$$P + nL \leftrightarrow PL_n,$$

and the equilibrium constant for this expression would be

$$K_a = \frac{[PL_n]}{[P][L]^n},$$

where K_a M^{-1} has units of inverse molarity (M^{-1}). This association constant can be used to measure the affinity of a particular ligand for a protein. Since the amount of ligand that binds to the protein is very small the concentration of unbound ligand is close to the total concentration of ligand in solution [8].

This allows the equilibrium to be written in terms of the fraction of binding sites occupied over the total binding sites. The fraction (θ) can be expressed as the following

$$\theta = \frac{[L]}{[L]^n + K_d},$$

where K_d is the dissociation constant of protein and ligand. If this expression is rearranged and the natural log is taken of each side, the resulting expression yields

$$\log\left(\frac{\theta}{1-\theta}\right) = \frac{[L]^n}{K_d}.$$

This is called the Hill equation. The slope of the Hill plot (n_H) reflects the degree of interaction between binding sites. If n_H equals one, then the ligand binding is not cooperative. If it is greater than one this indicates there is positive cooperative binding and a value of less than one indicates negative cooperativity. Negative cooperativity impedes the ability of ligands to bind after the first one is bound. To adapt this equation to hemoglobin, the ligand [L] must be replaced by pO_2, the partial pressure of oxygen and K_d is replaced by P_{50}^n:

$$\log\left(\frac{\theta}{1-\theta}\right) = n\log|pO_2| - n\log\left|P_{50}^n\right|.$$

Figure 2 shows the hill plots for the binding of oxygen to hemoglobin and myoglobin [8, 9].

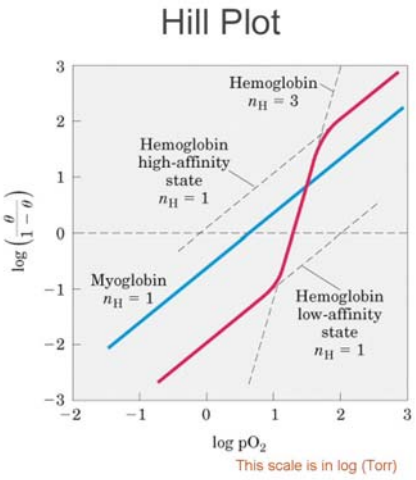

Figure 2. Hill plots for hemoglobin and myoglobin. The hill plot for myglobin gives a hill coefficient value of 1, indicating there is no cooperativity in binding oxygen. Hemoglobin, which has a value greater than one, does show positive cooperativity. It is important to note that the hill coefficient for hemoglobin is less than the number of binding sites for oxygen. This is a normal feature for a protein that exhibits allosteric regulation [2].

1.2 Hemoglobin: History and Characterization

Hemoglobin is a necessary oxygen carrier protein found in vertebrates and invertebrates. While there are many hemoglobin and heme proteins, their structures are not conserved and their functions can be different depending on the species. The classical role of hemoglobin is to bind molecular oxygen and deliver it to tissues; also to bind carbon dioxide and bring it to the lungs to be removed from tissue. Hemoglobin proteins are able to bind oxygen because of the heme group found in the interior of the protein. The heme group consists of an iron atom surrounded by a porphyrin ring. Both oxygen and carbon can reversibly bind the iron in the heme ring [1].

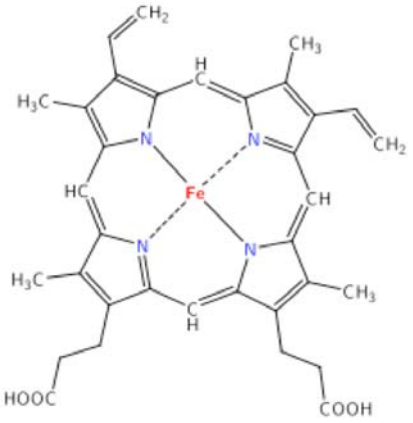

Figure 3. Heme ring structure. The heme is a prosthetic group that consists of an iron atom in the center of a porphyrin ring. This ring is responsible for binding oxygen and carbon monoxide in hemoglobin and myoglobin proteins (wikipedia.com).

All hemoglobin proteins share a common tertiary structure, which suggests that they are evolutionarily related. Hemoglobin residues are designated by their homologous helical or corner position in the sperm whale myoglobin, which was the first protein structure determined. The only point of ligation between the subunit and the heme group is the coordination between the heme iron and proximal histidine at position F8.

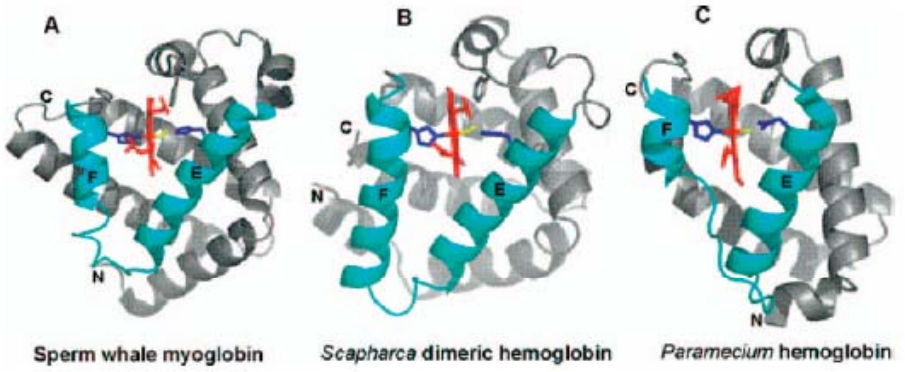

Figure 4. Heme position and fold in sperm whale myoglobin. Scapharca and Paramecium hemoglobin proteins. The heme group in each of these proteins, colored red, is shown interacting with the E and F helices, which are colored cyan and blue respectively. While the structures of each protein are different, the heme group binds to the proximal histidine F8, distal histidine E7, or the glutamine E7 and the highly conserved phenylalanine CD1 [10].

Ligands bind to the heme ring on the opposite, or distal, side of the heme group adjacent to the E helix. Comparison studies have shown that the different configurations alter oxygen affinity using three broad mechanisms. Stereo chemical effects in the proximal pocket of the protein can impact the reactivity of the iron atom and can lower affinity by limiting accessibility or can increase affinity by providing favorable electrostatic interactions [11, 12].

The most characterized protein of these hemoglobin proteins is the mammalian hemoglobin. Mammalian hemoglobin is assembled into a tetramer from two dimers containing the α and β subunits. Both of these subunits are evolutionarily related to each other and to myoglobin. The discovery of the protein was very important in 20th century molecular biology. Max Perutz obtained the crystal structure in 1968 and for the first time revealed structural transitions that underlie allosteric regulation in proteins [11].

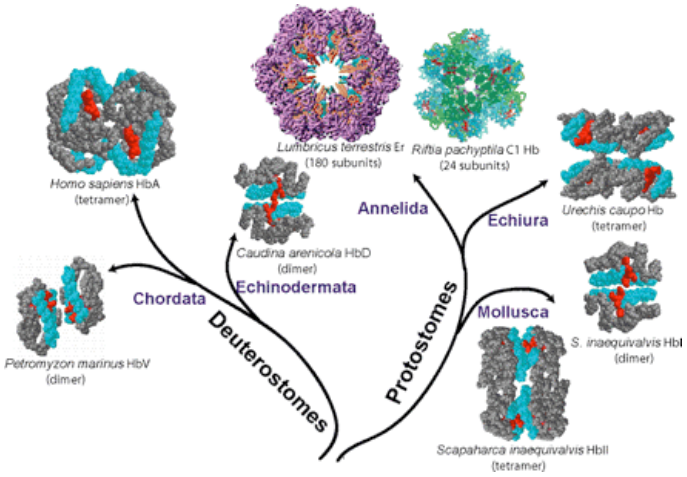

Figure 5. Pholgenetic tree of known hemoglobin protein structures. The tetrameric and dimeric hemoglobin proteins are shown as van der Waals spheres for the main chain and heme group atoms with the heme shown in red, the E and F helices in cyan and the rest of the main chain in gray. The Riftia C1 hemoglobin is depicted with a main chain trace with the green and blue corresponding to the two different subunit types and the heme groups are in red. The Lumbricus is shown in a 5.5 Å density map with the hemoglobin subunits colored magenta and the linker regions colored blue and gold. Of all these hemoglobin proteins only the Urechis hemoglobin does not exhibit cooperative binding [10].

Christian Bohr, another scientist, measured the hemoglobin oxygenation that showed binding was sigmoidal. This curve indicated that the binding of oxygen cooperative. He also discovered that carbon monoxide competes with oxygen when binding to hemoglobin. Later a physiologist named Gilbert Adair discovered that mammalian hemoglobin has four binding sites, and he calculated affinities for each site. He found that binding increased at each successive site with the last binding site having a much greater affinity than the first [3, 8].

Linus Pauling suggested the first structural model for cooperative oxygen binding. He discovered that he could reproduce the oxygen binding curves with a two-parameter model. In this model the binding event of oxygen to a heme group caused an interaction with a neighboring heme group. The first parameter was taken as an interaction and the other as an intrinsic binding constant. Since curve fitting only required one parameter, Pauling concluded that hemoglobin contains either a tetrahedral or square symmetric arrangement of heme groups. This was in 1935, well before Perutz solved the crystal structure of hemoglobin [3].

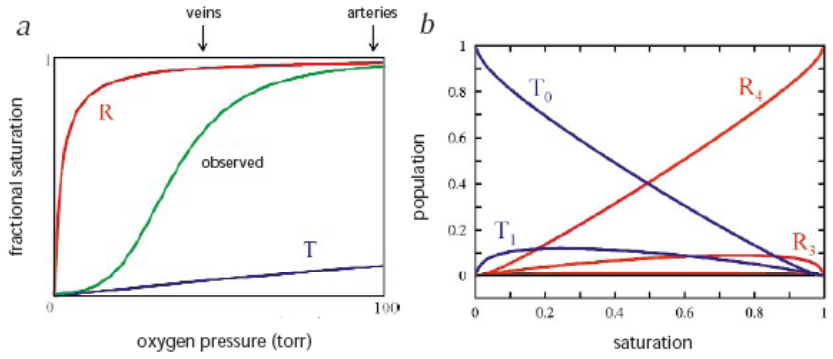

Figure 6. Hemoglobin oxygenation curves according to the MWC model. a) shows the observed binding curve from hemoglobin and the binding curves for the two noncooperative states (R and T). At low oxygen levels hemoglobin exists in the T state, as the oxygen pressure rises the equilibrium shifts to the R state. b) shows 10 populations of hemoglobin as a function of ligand saturation for the MWC model with four identical binding sites. Each curve is labeled as R or T state with the subscript corresponding to the number of ligands bound [3].

The first model to describe this cooperative binding was proposed by Jacques Monod and Jean-Pierre Changeux. They had studied enzymes and their activation/inhibition by substrates and ligands. They noticed that many enzymes were activated or inhibited by ligands in a cooperative manner and that these proteins contained multiple subunits. Together with Jeffries Wyman they developed a theoretical model for all multi-subunit proteins called the MWC model. In this model cooperativity arises from an equilibrium between two structures having different structural arrangements of subunits. The two structures are referred to as tense and relaxed states. The tense state, or T state, has a low affinity for ligand binding while the relaxed state, or R state, has a high affinity for a ligand. The model explains the shift from the T state to the R state with increasing oxygen pressure as required by Le Chatlier's principle. This model explained the sigmoidal curve for hemoglobin oxygenation that Bohr had found [13-15].

Q.H. Gibson furthered the understanding of protein cooperativity by studying hemoglobin kinetics. He discovered that photodissociation of carbon monoxide and hemoglobin increased the rebinding twenty times faster then the initial rate of mixing deoxy-hemoglobin and carbon monoxide. These results suggested that the number of already bound ligands does not determine the rate of ligand binding. This observation agrees with MWC model but not with the stereochemical model. Gibson also discovered that the optical spectrum of deoxy-hemoglobin after a light flash is different from the spectrum of deoxy-hemoglobin at equilibrium. He suggested this fast reacting state was a different structure [16].

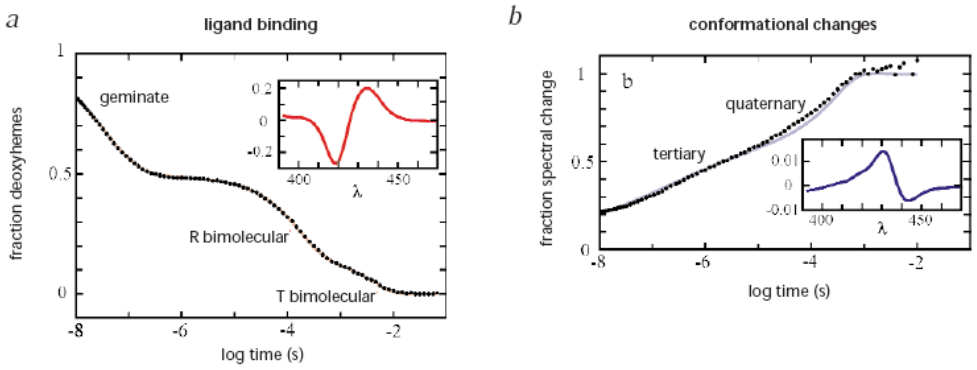

Figure 7. Kinetics of hemoglobin. Kinetics of hemoglobin following nanosecond photodissociation of carbon monoxide complex40. a) Ligand rebinding kinetics obtained from the average deoxy minus carbonmonoxy difference spectrum (inset). Geminate rebinding (rebinding of dissociated ligands before they escape from the protein) to R is followed by bimolecular rebinding to R and then to molecules that have switched from R to T. b) Protein conformational changes obtained from the change in the deoxyhemoglobin spectrum (inset). In both (*a*) and (*b*) the points are experimental, and the dotted curves are calculated from the extended MWC model. Because the deoxyheme spectral change is more than ten-fold smaller than the spectral change due to ligand binding (as indicated by the relative amplitudes in the insets), the deviations between the data and the fit in (*b*) represent less than 1% of the total spectral amplitude, and may not be significant [3].

1.3 Scapharca Dimeric Hemoglobin: Structure and Function

One of the simplest models for allosteric regulation is one with two chemically identical binding sites. The traditional mammalian hemoglobin has been extremely useful in gaining more insight into cooperativity, but still leaves some questions unanswered. To truly understand allosteric regulation at a molecular level, as simpler model has been used. The homodimeric hemoglobin from arid clams (HBI) assembles into a fold similar to myoglobin, but is assembled completely differently from mammalian hemoglobin proteins. The contacting region formed by the E and F helices is very extensive, which brings the iron atoms closer in proximity than found in mammalian hemoglobin. It suggests the route for communication between the two subunits is short, making HBI a good model for allosteric regulation [5].

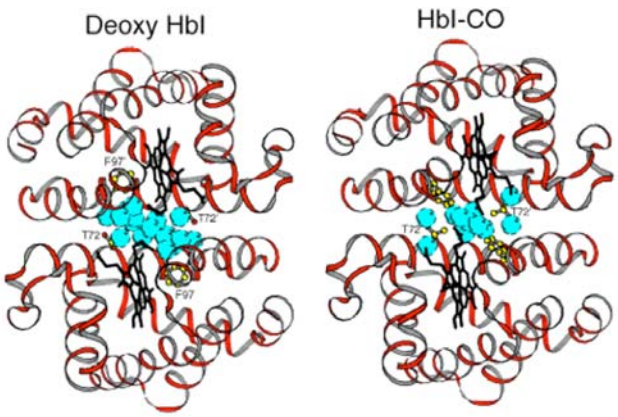

Figure 8. Scapharca Hemoglobin in the deoxy and CO bound forms. The transition from the T to R state has shown that the water molecules are disrupted at the interface upon ligand binding. It has been found that water play an important role in the communication between subunits (http://www.umassmed.edu/bmp/faculty/royer).

Upon binding a ligand HbI undergoes several structural transitions. These changes were first noted when high resolution structures of both the ligated HbI and the deoxy HbI were solved. To quantitatively describe the differences between the two states, the root mean square deviations (R.M.S.D.) of atomic positions were determined by superimposing one structure onto another. Generally residues with higher B-factors will also have larger R.M.S.D. values. The B-factor of an atom is the measure of how much an atom oscillates around the position specified in a model. The R.M.S.D. for each residue indicated that there were many residues involved in this ligation event. The differences in subunits provided the first information about effects of ligand binding, specifically at the interface between heme groups in the two subunits [5].

The degree of similarity in the two structures varies by subunit. The residues in the E-helix (residues 67-82) are almost completely unaffected by ligation of any region of the subunit. The F helix (residues 88-104), however, does show some significant conformational changes when ligation occurs. Amino acid residues within proximity to the heme ring are highly dynamic; specifically arginine 53(Arg53), lysine 96(Lys96), phenylalanine 97(Phe97), asparagine 100(Asn100), histidine 101(His101) and arginine

104(Arg104). These residues are involved in hydrogen bonding of the heme propionate groups. When comparing the R and T states, these residues adopt distinctly different conformations [5, 11].

In 2006 Knapp et al. presented time resolved x-ray crystallography results following the cascade of events of ligand binding in HBI. They were able to construct a sequence of structural events that mediate cooperativity after the unbinding of CO. They found an intermediate that formed 5 ns after ligand dissociation and were able to characterize the movement of the heme ring and the proximal side chains of the F helix and C-D loop. The work also demonstrated the importance of water molecules in the shifting between the R and T states. The conclusion of the time resolved studies was that these changes are responsible for the concerted transition that occurs at ~ 1 μs [17].

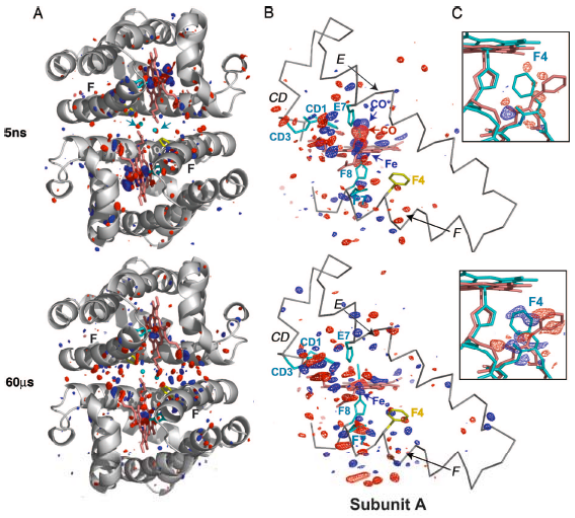

Figure 9. Difference Fourier map of Scapharca dimeric hemoglobin. HBI structural transitions were captured using Laue diffraction after laser excitation. The laser pulse serves to excite the electrons interacting between the heme ring and the CO. This causes the CO to dissociate and HBI to undergo a transition from R to T state. Figure 9A. Shows the difference map at 5 ns and 60 us after a laser pulse. The blue density corresponds to positive density that was not present in the dark map; the red corresponds to negative density that has was present in the dark map. 9B. Shows an alpha carbon trace of the CD region with the E and F helix and the heme group. 9C illustrates the flipping of the Phe97 residue [17].

1.4 Normal Mode Analysis

Normal mode theory is based on the harmonic approximation of the potential energy function around an energy minimum. A normal mode in physical terms is any motion in which all *n* coordinates oscillate sinusoidally with the same frequency. If a deformable system is perturbed then it will oscillate at a particular frequency. Normal modes have several characteristics when applied to chemical systems: each mode will behave like a simple harmonic oscillator; a normal mode is a concerted motion of many atoms, the center of mass does not move, all atoms pass through their equilibrium position at the same time and all normal modes are independent [18].

The approximation of the potential energy function allows one to solve analytically for the equation of motions by diagonalizing the Hessian matrix.

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Figure 10. A Hessian matrix Given a real valued function $f(x_1, x_2, ..., x_n)$, if all the second partial derivatives of *f* exist then the Hessian matrix of *f* would be $H(f)_{ij}(x) = D_i D_j f(x)$ [19].

The eigenvectors can be found from the Hessian matrix and comprise the normal modes. The corresponding eigenvalues are the squares of the frequency. What one finds when dealing with two more coupled oscillators is that they have several normal frequencies and that the general motion is a combination of vibrations at all different normal frequencies [18].

To illustrate the properties of normal mode analysis, take a simple system with four equal masses, *m,* two equal springs and two equal strings both with effective spring constants of *k*. They make up a system in which the two upper and two lower masses are constrained to move along rigid, frictionless bar separated by a distance *d.* The masses are labeled one through four and their movements away from the equilibrium positions are given by x_1, x_2, x_3 and x_4.

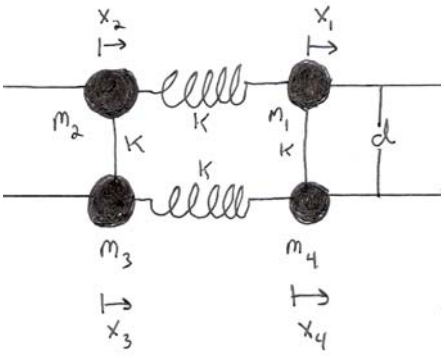

Figure 11. System of coupled oscillators. In this system two sets of point masses are connected by two springs and two strings all with spring constants equal to *k.* The length of the string is denoted as *d.* x_1, x_2, x_3 and x_4 correspond to the movements of the point masses away from their equilibrium positions.

To determine the normal frequencies and normal modes a system of equations containing the forces on all four point masses must be derived. The four equations are

$$\begin{aligned}
x_1 &: m_1 x_1'' = -k(x_1 - x_2) + k(x_4 - x_1) = -k(2x_1 - x_2 - x_4) \\
x_2 &: m_2 x_2'' = -k(x_2 - x_3) + k(x_1 - x_2) = -k(-x_1 + 2x_2 - x_3) \\
x_3 &: m_3 x_3'' = -k(x_2 - x_3) + k(x_3 - x_4) = -k(-x_2 + 2x_3 - x_4) \\
x_4 &: m_4 x_4'' = -k(x_4 - x_1) + k(x_3 - x_4) = -k(-x_1 - x_3 + 2x_4)
\end{aligned},$$

putting these linear equations in matrix form yields

$$\begin{pmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & m_3 & \\ 0 & & & m_4 \end{pmatrix} \begin{pmatrix} x_1'' \\ x_2'' \\ x_3'' \\ x_4'' \end{pmatrix} = -k \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix},$$

or in compact form $\vec{M}\vec{x}'' = -\vec{K}\vec{x}$. From this matrix the normal frequencies can be determined by solving

$$\det\left(\vec{K} - \omega^2 \vec{M}\right) = 0.$$

The resulting roots of this determinant are

$$\omega^2 = 0, \frac{2k}{m}, \frac{2k}{m}, \frac{4k}{m}.$$

These are the squares of the normal frequencies of this system. Now that we know the normal frequencies we must solve the eigenvalue equation

$$\left(\vec{K} - \omega^2 \vec{M}\right)\vec{a} = 0,$$

to obtain the normal modes. The normal modes for $\omega^2 = 0$ would be

$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = 0,$$

or

$$a_1 = \frac{a_2 + a_4}{2}, a_2 = \frac{a_1 + a_3}{2}, a_3 = \frac{a_2 + a_4}{2}, a_4 = \frac{a_1 + a_3}{2}.$$

In this case all the masses move with the same speed to the right or the left without any vibrating (no strings or springs are stretched). To obtain eigenvalues for the other modes, the eiqenvalue equation would have to be solved by inserting the appropriate ω value. When $\omega^2 = \frac{2k}{m}$ the top two and bottom masses will move in phase with each other, but out of phase with respect to the other two masses. The strings in this case are stretched. For the other $\omega^2 = \frac{2k}{m}$ solution, the left and right pairs of masses are in phase when moving. Again the bottom and the top sets are out of phase. The springs, in this solution,

will be stretched. The last solution, $\omega^2 = \frac{4k}{m}$, involves the stretching of all strings and springs as all masses are now out of phase with respect to each other.

Normal Mode Analysis has many applications outside of springs and classical mechanics. Recently NMA has been applied to other computational areas. It has proven to be useful in studying the molecular deformations in macromolecules seen in dynamic events. Biological macromolecules and supramolecular complexes can be efficiently studied this way. Using normal mode coordinates one can predict protein dynamics at high resolution. Over 1700 proteins have been analyzed using NMA and their motions approximated by applying a perturbation in the direction of the low frequency modes [7].

One of the major applications of normal modes in biophysics is the identification of conformational changes of enzymes upon ligand binding. This method is useful for studying protein cooperativity and allosteric regulation. This method has already been successfully applied to the study of membrane channel opening, structural movements of the ribosome, dynamic properties of DNA polymerase and viral capsid maturation. NMA is usually used to predict the manner in which a protein will function and shown to be accurate for these cases. Another major application that is beginning to be explored is using NMA in crystallography for data phasing [20-22].

The motion of a protein can be represented as a superposition of *n* normal modes. These normal modes fluctuate around an energy minimum. The displacement of *i* atomic coordinates can be expressed as:

$$r_i(t) = \frac{1}{\sqrt{m_i}} \sum_{k}^{3N} C_k a_{ik} \cos(\omega_k t + \phi_k),$$

where *m* is the mass of atom *i*, C_k and ϕ_k are the amplitude and phase of mode *k*, the vibrational frequency is $v_k = \omega_k / 2\pi$ and a_{ik} is the *i*th coordinate of the eigenvector *k*. Since there are large number of coordinates in the system a simplified harmonic potential is solved to obtain the normal modes. In the simplified harmonic potential the potential energy function used in an all atom force field is replaced by single parameter potential:

$$E_p = \sum_{d_{ij}^0 < R_c} c(d_{ij} - d_{ij}^0)^2 ,$$

where d_{ij} is the distance between two atoms *i* and *j*, d_{ij}^0 is the distance between these atoms in the three dimensional structure, *c* is half of the spring constant for the potential and is assumed to be the same for all interacting pairs, and R_c is an arbitrary cutoff parameter beyond which interactions are not taken into account. Using this simplified potential the stage of energy minimization is not required; the starting structure is used only as a reference structure. From these distance fluctuations measured between residues *i* and *j* in mode *k*, a map can be generated for all amino acid residues in the protein structure. In this map the amino acid residues behave as rigid bodies [23].

Using this distance fluctuation map two important quantities can be measured: the overlap and the collectivity. Overlap, denoted as I_k, measures the degree of similarity between the direction of a chosen conformational change $\Delta\vec{r}$ and the direction given by the normal mode *k*. If the overlap is equal to one then the directions are identical.

$$I_k = \frac{\left| \sum^{3N} a_{ik} \Delta \mathbf{r}_i \right|}{\left| \sum^{3N} a_{ik}^2 \sum^{3N} \Delta r_i^2 \right|^{1/2}} .$$

Collectivity, denoted as κ_k, is the degree of protein motion in the mode *k.* It corresponds to the number of atoms that are significantly affected in k. The value is calculated as

$$\kappa_k = \frac{1}{N}\exp\left(-\sum^{3N} \alpha\Delta A_{ik}^2 \log \alpha\Delta A_{ik}^2\right),$$

where A_{ik} is the amplitude of the displacement of atom *i* in *k* and α is the normalization factor chosen such that $\sum^{N} \alpha\Delta A_{ik}^2 = 1$. The conformational change of the protein is at a collective maximum for a value of one. Small intermolecular motion, on the other hand, involves only a few atoms. This reduces κ to its minimum ($\kappa = 1/N$) [23].

2 Methods

Protein Purification for HbI

The expression of scapharca hemoglobin was performed as previously described by [24]. Protein purification was performed by first pelleting the lysed cells at 30,000g for 30 minutes. The supernatant was then removed and saturated with CO. After CO saturation, a 50% ammonium sulfate precipitation was performed. Any precipitated protein was removed by centrifugating the sample at 14,000g. Another ammonium sulfate cut was performed, this time 95% saturated ammonium sulfate was added to the supernatant to precipitate the target hemoglobin proteins. The pellet containing protein was then resuspended in 40 mM CHES buffer at pH 9.0. The sample was then pH adjusted to the isoelectric point (pI) of HbI by dialyzing overnight in 40 mM CHES pH 9.0.

The next day the sample was centrifuged at 20,000g to remove any precipitated protein by using a 10kDa MW cutoff centriprep concentrator to a final volume of 25-30 ml. The sample was then run over a 50 ml DEAE 650S column that was pre-equilibrated with 40 mM CHES at pH 9.0. The HbI was monitored visually and eluted sample was collected from the column flowthrough. Recovered protein was then dialyzed overnight in 40mM HEPES buffer at pH 7.0. The next morning the sample was concentrated to a final volume of 15-25 ml. The sample was then loaded onto a pre-equilibrated CM Biogel-A cation exchange column. The column was cleaned by flowing 400 ml of 0.5 M NaCl solution then 200 ml of ddH2O through the column, which was then equilibrated by flowing in 40mM HEPES at pH 7.0. HbI tends to stick to the bottom of the column;

elution of HbI was obtained with 350 ml of 50mM NaCl in 40mM HEPES at pH 7.0. Purified protein was then checked for concentration and frozen at -80 C[24].

Photolysis Experiments

Photolysis is an analytical method that is used to determine kinetics of fast chemical reactions. In this process a short light flash sends out photons to a chemical sample. The molecules in the sample can absorb these photons and reach an excited state. For hemoglobin, excitation leads to the loss of bound CO, which can be monitored by absorption changes at 434 nm. The absorption is recorded by using a continuous light source and a photodetector.

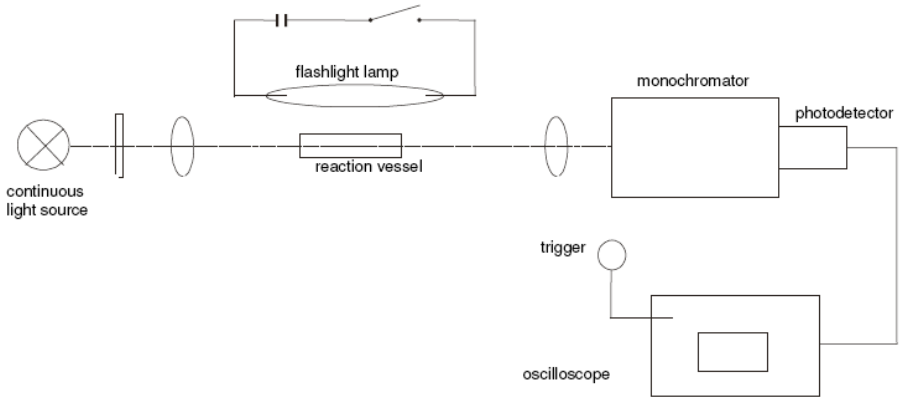

Figure 12. Photolysis experimental setup. The set up for a photolysis experiment requires a reaction vessel that light can easily transmit through. The continuous light source is set up in the direction of the photodiode and the pulsing light source perpendicular to them. The photodiode is connected to a photomultiplier tube that amplifies the signal and then sends it to an oscilloscope. The trigger is usually connected to the pulsing light source (http://www.chemie.uni-hamburg.de/studium/praktika/pc_v/Flash.pdf).

To determine the reaction kinetics the temporal gradient of the concentration of the sample has to be measured. The photolysis takes advantage of a temporal function of the voltage. A photomultiplier tube is used to detect the light that is emitted from the sample. Within the linear range of the photomultiplier tube its signal and the vertical

amplitude of the oscilloscope are proportional to the light intensity. The transmittance (T(t)) of a sample as a function of time can be described as:

$$T(t) = \frac{I(t)}{I_o} = \frac{V(t)}{V_o},$$

where I(t) is the time dependent intensity of the light, I_o is the intensity before the light flash, V(t) is the time dependent vertical amplitude, and V_o is the baseline amplitude. From the transmittance the optical density can be determined by Beer's law:

$$O.D.(t) = \varepsilon \cdot c(t) \cdot d = \log \frac{I_o}{I(t)} = \log \frac{V_o}{V(t)},$$

where ε is the extinction coefficient, d is the width of the sample holder and c(t) is the concentration of the sample as a function of time [25].

To measure the reaction rates of CO with HBI, a photolysis experiment was set up in this manner. The pulsing light source used was a Sunpack auto544 thryristor. The flash emitted from the Sunpack was focused using a plexiglass light guide. To block light of wavelengths less than absorption values of interest, a yellow filter was positioned after the Sunpack flash head. The continuous light source was a 150W halogen bulb with blue filters used on either side of the sample; this was done to limit the observing light to the 300-500 nm range. The signal from a Bauch and Lomb monochrometer was collected using a photomultiplier tube and sent to a Tektronix TDS 620 oscilloscope. The data were collected by a computer using IGOR Pro v.4 for analysis [26].

Normal Mode Analysis of HbI

The normal mode analysis was performed by submitting protein data base (PDB) files to the ElNemo server (www.elnemo.com). Ten models for the first five nontrivial modes were computed for the CO-bound HbI structure (PDB code: 3SDH). Since HbI

exists in both the R and T states, the deoxy HbI structure (PDB code: 4SDH) was also submitted to contribute to each normal mode of a possible confirmation change. The amplitude minimum (DQMIN) was set to -100 and the amplitude maximum (DQMAX) was set to 100 and the step size was 20. These parameters are used to compute the structural models for a given normal mode. In addition to the first five nontrivial modes, 18 additional modes were calculated. The R.M.S.D. values for each mode were then compared to the R.M.S.D. values found previously in Royer et al., 1994. All simulations were viewed with Visual Molecular Dynamics (VMD) and Pymol. R.M.S.D. values were then compared with crystallographic data of the R and T states of HbI, including the time-resolved x-ray crystallographic data.

Results

Flash Photolysis Results

Flash photolysis is an excellent tool to probe for cooperative binding in hemoglobin. In the case of HbI, flash photolysis is used to remove CO and monitor the rebinding via changes in absorbance. As the flash intensity is adjusted, differences in rates of rebinding to fully unliganded and partially unliganded can be compared. Kinetic parameters can then be determined and cooperativity can be described in a quantitative fashion.

The rebinding of a single CO molecule to the HbI solution can be modeled using a single exponential. The Igor program fits raw data to this exponential using the following equation:

$$O.D.(t) = \left(K_o + K_1\right)e^{-K_2 t},$$

where K_o is the baseline, K_1 is the amplitude and K_2 is the rebinding rate of CO to HbI. The K_2 is useful as it tells you the time constant for CO to rebind to HbI. If HbI is cooperative, then K_2 should be larger when only one CO dissociates than if both CO molecules dissociated.

Setting the Sunpack to 1/64 flash provides a low enough intensity such that rebinding primarily occurs to singly liganded HbI molecules. The five flash traces were collected every five seconds for the 1/64 flash experiment and then averaged. The curve fit yielded a K_2 value of 2.687e+02, and a K_1 of 0.1636. The experiment was then repeated using a ¼ flash. The energy for a ¼ flash is intense enough to remove all bound CO to HbI. The expected result would be a lower K_2 value since HBI would be in the T state when CO started to rebind and the transition to R would require more time. The K_2

obtained from the ¼ flash average was 1.784e+02, and K_1 was 0.7465. This agrees with our claim that ¼ knocks both COs off while rebinding following a 1/64th occurs primarily to singly liganded species. These data provide an indication of cooperative binding of CO to HbI.

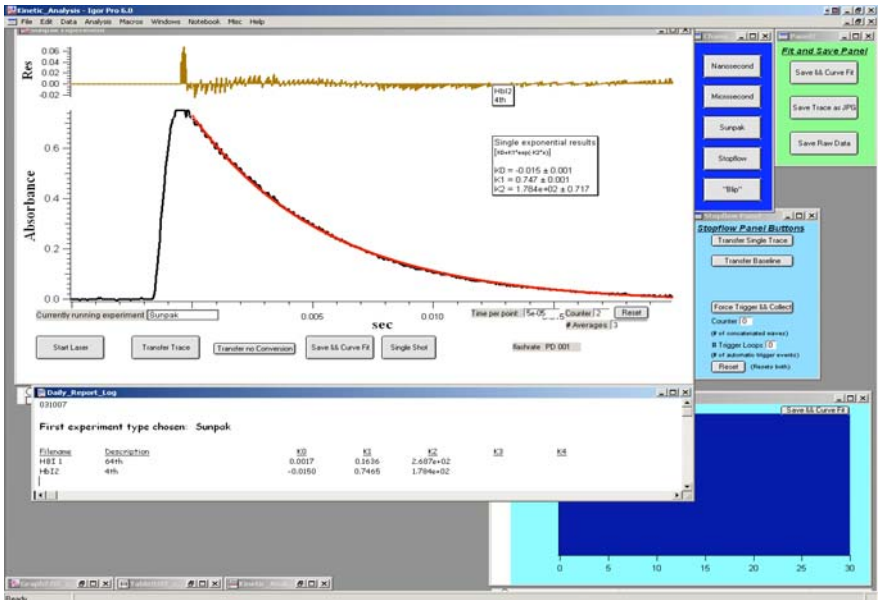

Figure 13. Kinetic experiment analysis using Igor Pro. Igor is a useful program for graphing waves and curve fitting. Kinetics can be measured on the nanosecond time scale using a laser and slower kinetics using light flashes. This figure shows the interface for a fitted curve and the calculated rate constants. This experiment was done on the wild-type HbI. The rate constants for 1/4th and 1/64th flashes are displayed in the daily report log.

Normal Mode Analysis on HbI

From the normal analysis 31 modes in all were computed. ElNemo automatically throws out the first five modes as they are just translations and rotations of the macromolecule and trivial for analyzing actual motion. When computing the normal modes, ElNemo calculates B factors for every atom in the protein structure based on the first 100 modes. The calculated B factors are then compared to the B factors from the crystallographic data in the PDB file. A high correlation between the two sets indicates that normal modes accurately capture the overall flexibility features of the model. Differences between B factors from crystallographic data and the normal mode analysis

can also indicate where protein flexibility is modified due to crystal packing. The correlation for HbI was 0.692 for the 290 alpha carbons in the protein.

To determine if the B factor prediction was accurate for HbI, another B factor prediction for a well-characterized protein by NMA was calculated and compared to HbI. X-ray scattering and NMA have previously been used to study Taq polymerase, a DNA polymerase involved in DNA replication. The correlation for Taq was 0.699 for 807 alpha carbon atoms. This value is very close to the correlation of HbI B factors, which suggests NMA is applicable for HbI.

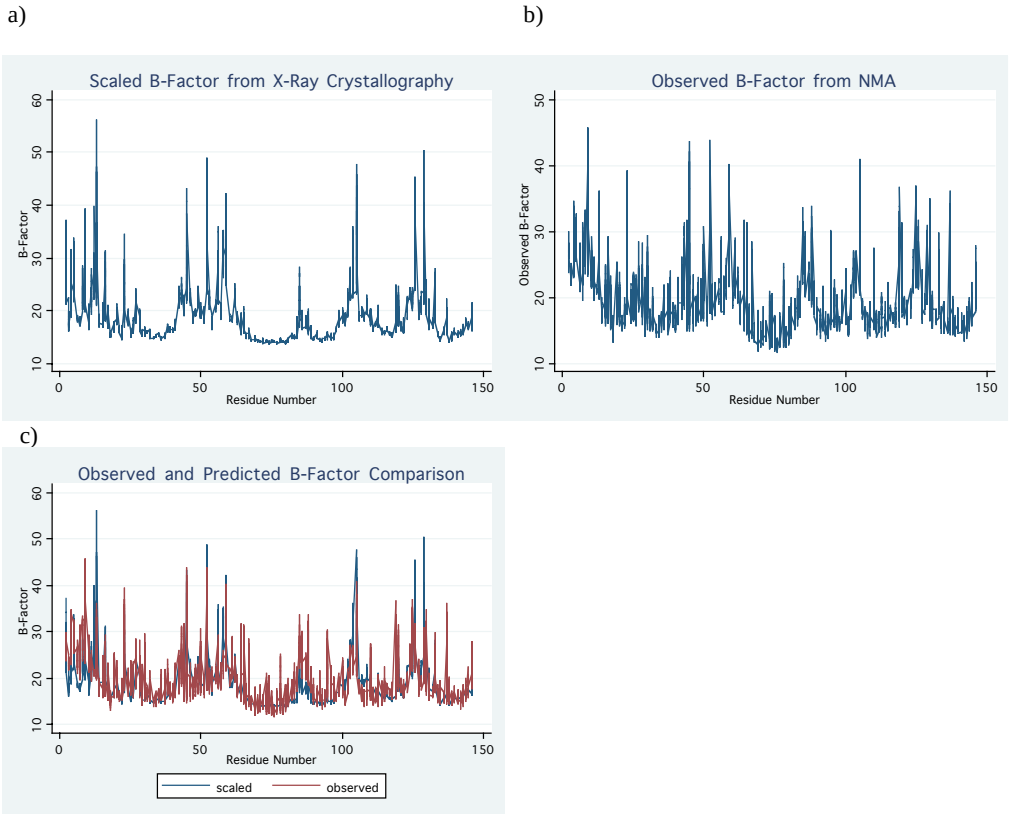

Figure 14. B-factors of each residue of HBI from x-ray crystallography and normal mode analysis. The B-factors are plotted for the main chain atoms of the protein. (a) The B-factors from x-ray crystallographic data[5] , (b) the B-factors calculated from NMA and (c) show the overlap between the two sets of B-factors. The amount overlap indicates the NMA is accurately predicting the possible motions of the atoms in HbI.

To determine which modes of the NMA represented actual conformational changes of the ligation event, R.M.S.D. values were compared for all modes to the

crystallographic data from [5]. It is expected that there are a few very subtle changes, mainly occurring in the F helix residues. The mode that best fit the crystallographic data was mode 20. The largest R.M.S.D. values in mode 20 occur in residues 99-107 of both subunits. These residues are located in the F helix of the protein. Crystallographic data shows the largest ligand-linked movements in the crystal structure occur in the F helix (residues 88-104). This suggests mode 20 best captures the actual movement of the protein during CO dissociation.

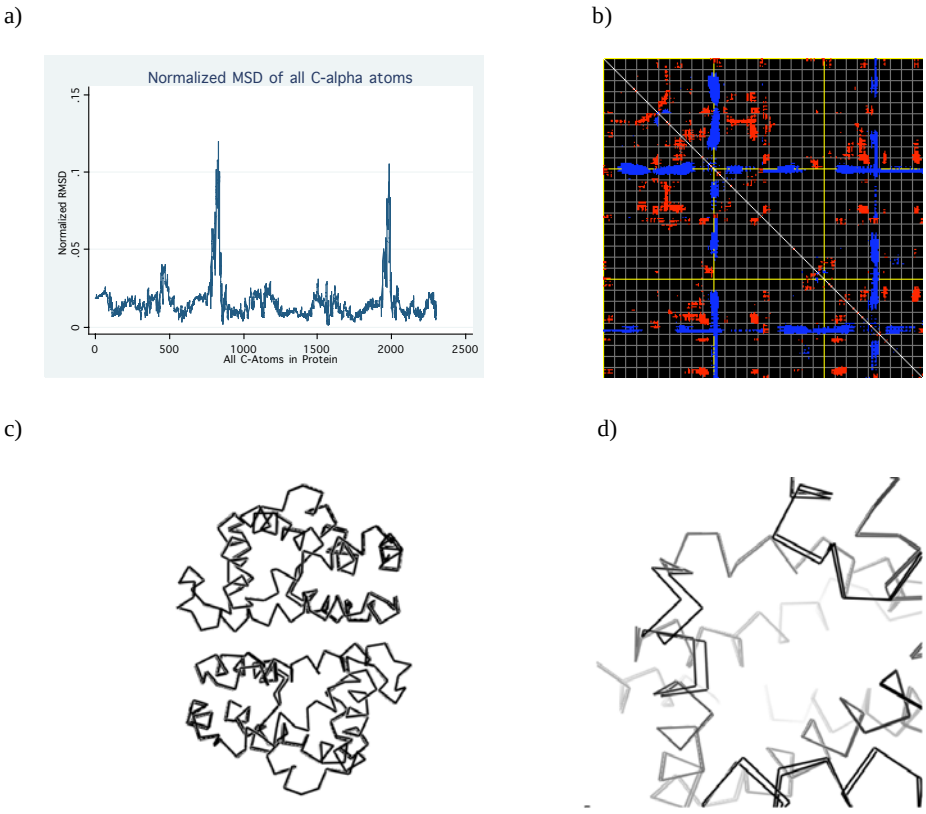

Figure 15. R.M.S.D. values for each residue predicted by NMA. a) The graph plots the normalized mean square displacement for all the alpha carbon atoms in the protein. The two sharp peaks correspond to the section of the F helix that is involved the ligation event. b) shows the alpha carbon distance fluctuation map for HbI. This matrix displays the maximum distance fluctuations between all pairs of alpha carbon atoms and between the two extreme conformations that were computed for mode 20 (DQMIN/DQMAX). Distance increases are plotted in blue and decreases in red for the strongest 10% of the residue pair distance changes. Every pixel corresponds to a single residue. Grey lines are drawn every 10 residues, yellow lines every 100 residues. c) Highlights the overlap of the two static states in mode 20. There is very little displacement of whole subunits with the exception of the F helix in d).

Discussion and Future Work

Scapharca dimeric hemoglobin is a fully cooperative protein and a perfect system for studying allosteric regulation at the molecular level. Time-resolved crystallographic studies have shown the conformational changes upon ligand binding at high resolution. Normal mode analysis has successfully been able to predict the movement of the main chain atoms when HBI undergoes its conformational change. The B-factors and R.M.S.D. values calculated in the most relevant normal mode correlate with the values from the crystallographic study.

The fact that the computational results agree with experimental data suggests that NMA may be useful to determine the mechanism for allosteric regulation for other proteins. There are many other hemoglobin proteins whose crystal structures are not known for both states. NMA may be able to predict the conformational changes of these proteins if a crystal structure in one state in unattainable. The next step in studying HbI would be to use normal mode analysis to predict the motions of mutated HbI. This could provide a simple computational method for determining the residues that are necessary for allosteric regulation. It may also be able to predict early intermediate structures that occur too rapidly for time resolved x-ray crystallography to observe.

Motor proteins, another group of allosteric proteins are less understood than hemoglobin, could also be studied with normal mode analysis. Unlike hemoglobin proteins, motor proteins undergo very large conformational changes and exhibit large domain motions when binding ATP. The structures of motor proteins are much bigger and more complex; making it hard to study their dynamics using x-ray crystallography or

other spectroscopic methods. Perhaps molecular dynamics, including NMA, could solve this problem computationally [8, 27, 28] (see appendix A).

Dynein is one such protein that would be an excellent candidate for such computational methods. Dynein is a large cytoplasmic protein that is involved in many cellular processes. There is currently no crystal structure of any of the dynein sub-units with the exception of one light chain. A normal mode analysis done by a homologous fit of dynein subunits has already shown structural changes that agree with comparative studies with other proteins. The reports so far have helped biologists gained some insight, but there are still many questions about dynein structure and dynamic properties [28, 29].

The most studied aspect of dynein has been force production. The ability to convert chemical energy to purely mechanical energy is something biologists have been trying to explain since the first studies on muscle proteins were done. Today optical tweezers is the primary tool for measuring step-size, force generation and maximum force output of single molecule motor proteins. While this method is very accurate at measuring these parameters, different groups have reported different values for the same experiments with the same proteins. Controversy in this field has become widespread. One resolution to this problem would be to measure these parameters using a different technique [30-32].

Atomic force microscopy (AFM) could be one possible solution to this problem. AFM is a very useful tool to perform accurate force measurements. Force measurements have been done on several proteins and many more are now being studied with AFM. We have taken advantage of the force measuring capabilities of the AFM to design experiments for measuring forces generated by motor proteins, specifically dynein. We

are currently attempting to test our design to see if the data matches the previous optical tweezers experiments (see appendix B).

Proteins are highly dynamic macromolecules. Many proteins are regulated by allsotery, an essential component of cell survival. Hemoglobin, the classical model of allostery, has been studied great detail in attempts to find a model that explains this phenomena. Scapharca dimeric hemoglobin is one of the simplest model for studying allosteric regulation. Time-resolved crystallography has been used to track conformational changes in HbI on the nanosecond time scale. In this study we have used normal mode analysis, a computational model for predicting protein motions, to successfully predict the conformational changes seen in crystallographic studies. This suggests NMA may be able to predict conformational changes for other, much larger proteins whose structural information is not be well known. It also suggests that NMA may be a useful tool in studying allosteric regulation in both large-scale motions and small intermolecular motions.

References

1. Dickerson, R.E. and I. Geis, *Hemoglobin : structure, function, evolution, and pathology*. 1983, Menlo Park, Calif.: Benjamin/Cummings Pub. Co. 176 p.
2. Lehninger, A.L., D.L. Nelson, and M.M. Cox, *Lehninger principles of biochemistry*. 4th ed. 2005, New York: W.H. Freeman. xxv, 1119, [91] p.
3. Eaton, W.A., et al., *Is cooperative oxygen binding by hemoglobin really understood?* Nat Struct Biol, 1999. **6**(4): p. 351-8.
4. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins, 2004. **57**(3): p. 433-43.
5. Royer, W.E., Jr., *High-resolution crystallographic analysis of a co-operative dimeric hemoglobin.* J Mol Biol, 1994. **235**(2): p. 657-81.
6. Trylska, J., et al., *Ribosome motions modulate electrostatic properties.* Biopolymers, 2004. **74**(6): p. 423-31.
7. Suhre, K. and Y.H. Sanejouand, *ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W610-4.
8. Voet, D., J.G. Voet, and C.W. Pratt, *Fundamentals of biochemistry : life at the molecular level*. 2nd ed. 2006, New York: Wiley. 1 v. (various pagings).
9. Saroff, H.A. and A.P. Minton, *The Hill plot and the energy of interaction in hemoglobin.* Science, 1972. **175**(27): p. 1253-5.
10. Royer, W.E., Jr., et al., *Allosteric hemoglobin assembly: diversity and similarity.* J Biol Chem, 2005. **280**(30): p. 27477-80.
11. Royer, W.E., Jr., et al., *Cooperative hemoglobins: conserved fold, diverse quaternary assemblies and allosteric mechanisms.* Trends Biochem Sci, 2001. **26**(5): p. 297-304.
12. Bolognesi, M., et al., *Nonvertebrate hemoglobins: structural bases for reactivity.* Prog Biophys Mol Biol, 1997. **68**(1): p. 29-68.
13. Perutz, M.F. and L.F. TenEyck, *Stereochemistry of cooperative effects in hemoglobin.* Cold Spring Harb Symp Quant Biol, 1972. **36**: p. 295-310.
14. Ackers, G.K. and M.L. Johnson, *Linked functions in allosteric proteins. Extension of the concerted (MWC) model for ligand-linked subunit assembly and its application to human hemoglobins.* J Mol Biol, 1981. **147**(4): p. 559-82.
15. Changeux, J.P. and S.J. Edelstein, *Allosteric mechanisms of signal transduction.* Science, 2005. **308**(5727): p. 1424-8.
16. Gibson, Q.H., *The photochemical formation of a quickly reacting form of haemoglobin.* Biochem J, 1959. **71**(2): p. 293-303.
17. Knapp, J.E., et al., *Allosteric action in real time: time-resolved crystallographic studies of a cooperative dimeric hemoglobin.* Proc Natl Acad Sci U S A, 2006. **103**(20): p. 7649-54.
18. Taylor, J.R., *Classical mechanics*. 2005, Sausalito, Calif.: University Science Books. xiv, 786 p.
19. Dacorogna, B., *Introduction to the calculus of variations*. 2004, London Singapore ; Hackensack, N.J.: Imperial College Press ; Distributed by World Scientific. xii, 228 p.

20. Delarue, M. and Y.H. Sanejouand, *Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model.* J Mol Biol, 2002. **320**(5): p. 1011-24.
21. Tama, F., et al., *Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9319-23.
22. Krebs, W.G., et al., *Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic.* Proteins, 2002. **48**(4): p. 682-95.
23. Valadie, H., et al., *Dynamical properties of the MscL of Escherichia coli: a normal mode analysis.* J Mol Biol, 2003. **332**(3): p. 657-74.
24. Summerford, C.M., et al., *Bacterial expression of Scapharca dimeric hemoglobin: a simple model system for investigating protein cooperatively.* Protein Eng, 1995. **8**(6): p. 593-9.
25. Atkins, P.W. and J. De Paula, *Physical chemistry*. 7th ed. 2002, New York: W.H. Freeman. xxi, 1139 p.
26. Knapp, J.E., et al., *Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin.* Biochemistry, 2005. **44**(44): p. 14419-30.
27. Fletcher, D.A. and J.A. Theriot, *An introduction to cell motility for the physical scientist.* Phys Biol, 2004. **1**(1-2): p. T1-10.
28. Serohijos, A.W., et al., *A structural model reveals energy transduction in dynein.* Proc Natl Acad Sci U S A, 2006. **103**(49): p. 18540-5.
29. Williams, J.C., H. Xie, and W.A. Hendrickson, *Crystal structure of dynein light chain TcTex-1.* J Biol Chem, 2005. **280**(23): p. 21981-6.
30. Kastrikin, N.F., *Force generation and ATP hydrolysis in muscle contraction.* J Theor Biol, 1980. **84**(2): p. 387-400.
31. Knight, A.E., G. Mashanov, and J.E. Molloy, *Single molecule measurements and biological motors.* Eur Biophys J, 2005. **35**(1): p. 89.
32. Mehta, A.D., et al., *Single-molecule biomechanics with optical methods.* Science, 1999. **283**(5408): p. 1689-95.

Appendix A. Motor Proteins

Motor proteins are enzymes that convert chemical energy from ATP into mechanical work to drive cellular processes and cell motility. These motors harness energy released by ATP hydrolysis. Nearly one hundred motor proteins exist in a eukaryotic cell, which is a complex cell where the genetic material is protected by a membrane-bound nucleus They differ by the type of filament they bind to, the direction that they move on their specific filament, and the various cargos that they carry. Axonemal, dynein, kinesin, and muscle myosin were the first discovered members of related motor families. These motor families are defined by similarities in the genetic sequences in their motor domains. Comparisons of these sequences and functional assays provide insight into a motor's mechanism for movement [1, 2].

Motor proteins travel along specific filaments to carry their cargo. The two filaments that are associated with motor proteins are actin and microtubules. These protein filaments form the cytoskeleton of a cell. The building block proteins, actin and tubulin, have been studied in great detail. Crystal structures exist for actin, alpha-tubulin, and beta-tubulin. The atomic structures have been used to build the models of these macromolecular assemblies. There are also several proteins that are related to these filaments. There are over a dozen actin-related proteins (Arps) and five classes of tubulin proteins [1].

Dysfunction in a motor protein can lead to motor neuron degeneration and disease. Several diseases have been linked to dysfunctional motor proteins including: spinal muscular atrophy (SMA), amyotrophic lateral sclerosis (ALS), and smooth brain disease (lissencephaly). While many studies have been done on motor proteins, many

questions still remain unanswered. The functionality of all motor proteins is still not entirely understood. Myosin and kinesin motor families have been studied in great detail; several structures have been solved and mechanisms for motility have been proposed. Dynein's structure and mechanism for movement, on the other hand, have yet to be determined. Without a crystal or NMR structure, it has been difficult to determine the structural details of movement [3, 4].

Dynein

Dynein is a minus-end-directed microtubule motor that is necessary for many diverse cellular processes. Dynein belongs to the AAA+ (**A**TPase **a**ssociated with various cellular **a**ctivities) ATPase family and is very different from kinesin or myosin motor proteins. Dynein is the largest of the molecular motors; its size is 1.2 MDa. Dynein plays a vital role in mitotic spindle formation, chromosome segregation, and transport of various cargoes. The dynein family has two major branches: cytoplasmic dynein and axonemal dynein [2, 3].

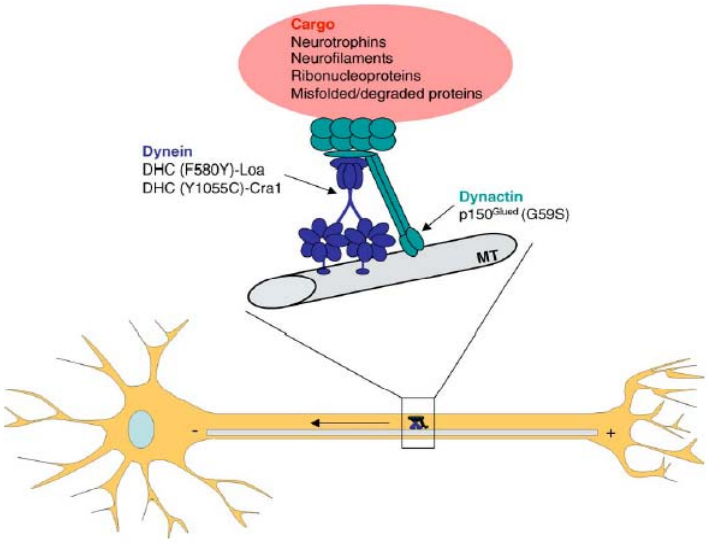

Figure A1. Cytoplasmic dynein and dynactin and its role in retrograde transport. Dynein and dynactin are necessary proteins to carry various cargo along axons in brain [3].

Both dynein and its activator dynactin are necessary for various cellular activities. In neurons they transport neurtrophins, neurofilaments, ribonucleoproteins and material targeted for degradation. This cargo is moved from the distal end of the neuron to the cell body. The transport along the microtubule is oriented so that dynein moves from the + end to the – end of the cell. Many mutations in any of the subunits will render dynein nonfunctional, which can cause severe deleterious effects on the cell [4].

Cytoplasmic dynein is thought to be found in all eukaryotic cells and is important for vesicle trafficking and for the localization of the Golgi apparatus near the center of the cell. Cytoplasmic dynein is typically a homodimer with two large motor domains as heads. It also contains several other subunits called intermediate, light intermediate and light chains. Axonemal dynein, the other branch of the dynein super-family, is found as a monomer, heterodimer, or heterotrimer, containing one, two, or three motor domain heads. Axonemal dyneins are localized in the outer doublet microtubules in the axoneme of cilia flagella. These doublet microtubules form projections called dynein arms, which contains many dynein molecules. Axonemal dynein, in this arrangement, produces bending motions in cilia flagella by the generation of torque and oscillation [3, 5].

Unlike other motor families like kinesin and myosin, whose structure and mechanism for motility are well understood, little is known about the mechanism and structural details of dynein. To date no crystal structure exists for an entire dynein molecule. Some structural basis for dynein has been determined by looking at other AAA+ proteins. Dynein's motor domain is composed of six to seven AAA+ domains, which are contained in a single polypeptide chain. Four of the domains have been found to bind ATP, AAA1 being the primary site of ATP hydrolysis. A coiled-coiled stalk

appears after the AAA4 domain and the microtubule-binding domain is located at the tip of this stalk. The N-terminus of the six to seven AAA+ domains is the linker domain that has been suggested to interact with the ring and is possibly involved in force generation. The N-terminus of the heavy chain has been found to be involved in dimerization, binding to dynein associated proteins like dynactin [6-8].

Dynactin is a multi-subunit protein that is required for most of cytoplasmic dynein's functional activity. It is required for mitosis, making it a vital protein in the cell. Dyanctin contains eleven different subunits and a total of twenty subunits weighing a total 1.2 MDa. Dynactin aids dynein by binding directly to it and allows the motor to move long distances over microtubules. Dynactin's largest subunit, p150Glued, binds to dynein and enhances motor processivity [9].

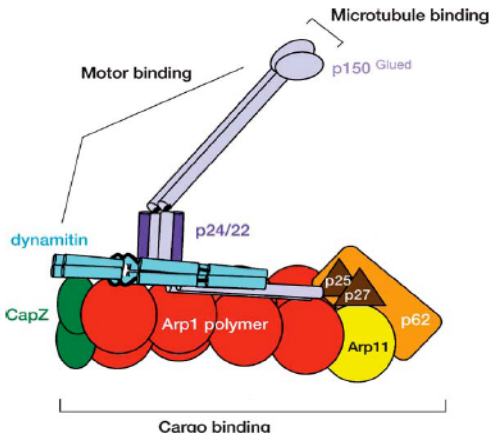

Figure A2. **Structure of Dynactin** with appropriate subunits labeled. No crystal structure exists for dynactin; all structural information is inferred from electron microscopy data [9].

Processivity and Stepping Behavior

While motor proteins may have some structural similarities, they have very different functions. Functional details of the movement of kinesin and myosin have been studied and their movement is now well understood. Dynein, on the other hand, is less

understood. Its structural details and functions for movement may not be similar to those of myosin and kinesin. Several models have been formulated to describe dynein's processivity (a motor protein's processivity is defined as the movement of that motor protein), but there are still many questions about the structure-function relationship of dynein [10].

To understand functional differences between motor proteins, one must look at the motor's duty ratio. The duty ratio was developed to combine the fact that motor proteins have both chemical and mechanical cycles. The duty ratio, r, can be defined by the fraction of time a head spends in the attached phase to its particular filament

$$r = \frac{\tau_{on}}{\tau_{on} + \tau_{off}}.$$

The differences in this ratio can be useful in determining whether a motor is processive. The minimum number of heads necessary for continuous movement can be related to the duty ratio

$$r \cong \frac{1}{N_{\min}}.$$

This guarantees that at least one head will be bound to its filament at a given time. For a processive motor, one expects that its duty ratio will be at least 0.5. Kinesin and cytoplasmic dynein, both having two heads, must have a ratio of ≥ 0.5. If there is a one-to-one relationship between mechanical cycles and chemical cycles, then the speed of a motor [11] should be equal to

$$v = k_{ATPase}\Delta,$$

where k_{ATPase} is the rate at which a single head hydrolyzes one molecule of ATP and Δ is the distance traveled by each head relative to the filament per mechanical cycle. In-vivo cytoplasmic dynein's speed has been experimentally determined from native bovine brain at 30° C as $v_{dynein(in-vitro)} = -1250$ nm/s (Paschal, King et al. 1987). The in vivo speed of dynein was determined by studying retrograde transport in squid axoplasm at room temperature. It was found to be $v_{dynein(in-vivo)} = -1110$ nm/s [1, 12].

To determine a motor's step size, the working distance first needs to be calculated. Working distance can be calculated from the duty ratio as

$$r = \frac{\tau_{on}}{\tau_{on} + \tau_{off}} = \frac{\delta \cdot k_{ATPase}}{v} = \frac{\delta}{\Delta},$$

where δ is the working distance. To determine these parameters, single molecule techniques are usually employed. Modified motility assays can be combined with force transducers to measure force generation and step size from a single motor on a bead. Atomic force microscopes and optical tweezers are some of the common tools used to determine these parameters Several studies have indicated varying step sizes for dynein with and without cargo. One group has reported that dynein proteins take large steps at low load (24-32 nm), but 8 nm steps when multiple dynein molecules interact with a microtubule. Another study reported 8 nm steps of cytoplasmic dynein carrying beads with or without quantum dots. To this date dynein's step size and load dependence is still controversial [1, 10, 13, 14].

Biological Applications of AFM

Recent innovations have allowed for a greater understanding of the force interactions of biological molecules; most importantly, cantilevers with tip radius of

curvatures in the nanometer range. Having such a small tip allows for the measurement of single molecules at near atomic resolution. Most AFMs can measure forces between 0.01 and 100 nN, which fit within the scales of most biological processes. Some biological applications of this force microscopy include protein-ligand interactions, antigen-antibody pairs, and protein-membrane interactions [14].

Measurements made by AFMs can be used to determine intramolecular and intermolecular force interactions. One of the first biological force interaction experiments was performed on the modular protein titin, which is "comprised of multiple tandem repeats of immunoglobulin and fibronectin III domains, each possessing a β-sandwich structure" [15]. The titin force measurements taken showed a periodic data pattern which corresponded to unfolding events for the protein. The AFM data for titin was confirmed by optical trapping force-extension experiments, showing that AFM force measurements were accurate. Single molecule force measurements of other biological molecules also show the possibility of force-induced conformation, or structural, changes. Such conformational changes in a molecule could help provide information on their mechanism of motility [15].

The use of fluid-tapping AFM can be used to create topographical images of biological molecules. Information about these molecules can be extracted from images in real time, which are able to reveal processes occurring in the molecule. This technique has been applied to the study of *Escherichia coli* RNA polymerase (RNAP) holoenzymes. Images revealed that there was a reduced DNA contour length, which led to the proposal that "DNA wraps around RNAP in the open motor complex" [16]. Topographical images of DNA bending induced by similar methods is shown below in Fig. A3.

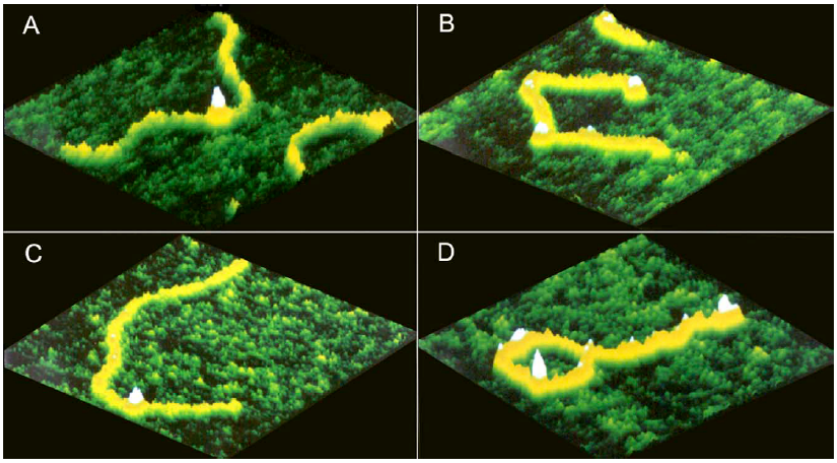

Figure A3. DNA bending induced by Cro bound at specific sites (A and B) and non-specific sites (C and D). "The DNA is a 1-kb double-strand fragment containing the k OR region to which Cro binds located between 370 and 440 bp from one end of the DNA. The scan sizes are 250 nm" [16].

Fluid-tapping AFM has also been used to image single *Escherichia coli* RNA polymerase (RNAP) molecules. RNA polymerases have been found to adhere to gold surfaces. In this experiment gold was deposited into mica samples. The top surface of mica was cleaved, revealing gold deposits. This gold surface acted as an atomically ultra-flat mica surface. Multiple techniques were used to prepare the gold-mica surface, as can be seen in Figure A4. Self-assembled monolayers of ω–functionalized alkanethiols were then formed on the gold surface, allowing RNAP to bind more tightly to the surface. This strong binding force allowed for better imaging quality and reliability and also minimized distortion of the sample by the AFM tip dragging the RNAP on the surface. In addition, this experimental setup made it possible to measure interaction forces between the RNAP and the EG-thiol [17].

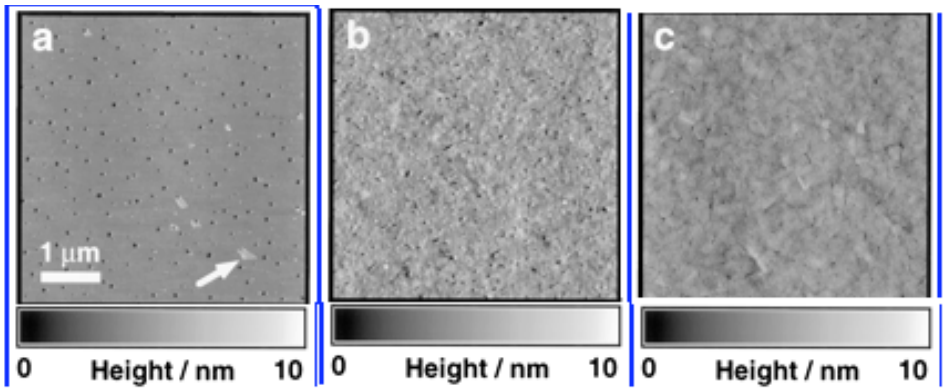

Figure A4. **AFM images obtained using tapping-mode in air of ultra flat gold-mica surfaces, created using different methods**. "(*a*) gold sputtered onto mica heated to ~300°C in an argon atmosphere at 50 millitorr, (*b*) gold evaporated onto mica at room temperature with a deposition rate of 0.13 nm/s and pressure below 2×10^6 torr, and (*c*) gold evaporated onto mica heated to ~350°C with a deposition rate of ~15 nm/s at a pressure between 5×10^6 and 7×10^6 torr" [17].

AFM can also be used to manipulate single large macromolecular assemblies like actin filaments. Forces that are applied to the actin cause an attached micro needle to bend until a restoring force is matched, and then the strength of the applied force can be measured (shown in Fig. A5a). It is also possible to measure mechanical properties of the actin filament, such as tensile strength and stiffness. This is done by attaching one end of the filament to a flexible micro needle and the other end to a very stiff micro needle. Force is then applied to the stiff micro needle and the displacement of the flexible one is measured at the time of the destruction of the actin filament. The setup for this can be seen in Fig. A5b [18].

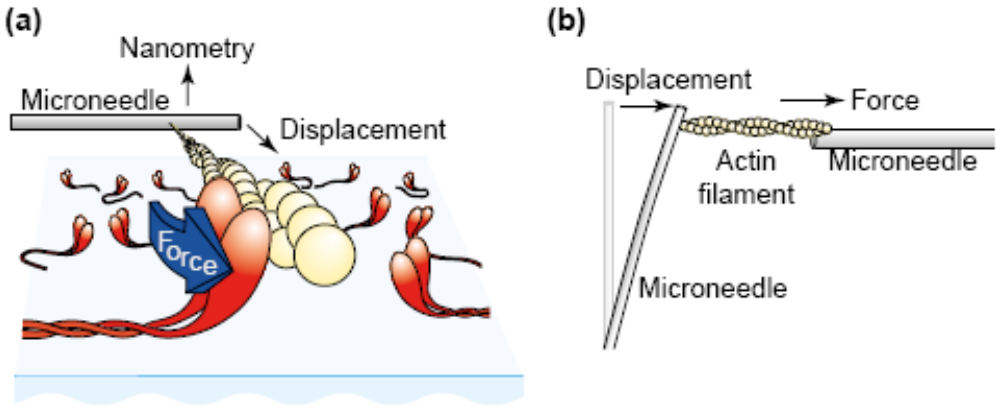

Figure A5. Force spectroscopy on actin filaments. Figure 5a shows an actin filament attached to a micro needle, which is able to measure the applied force on the actin filament by measuring the displacement of the micro needle. Figure 5b shows an actin filament attached between a stiff (right) and flexible (left) micro needle. This can be used to measure mechanical properties such as tensile strength and stiffness [18].

Measuring absolute forces of systems becomes difficult as these forces are directly dependant on the sample preparation techniques and conditions. Consequently, it is unreliable to compare direct results from one research group to another, as these values may be skewed from one another. In order to properly compare data, the measuring of the change of forces throughout a process becomes a much more reliable way for the community as a whole to check figures against other research groups. Even with this, there currently exists no universal standardized method for preparing biological samples for force microscopy, so this has limiting effects on the data output [16].

AFM is currently able to unzip single molecules *in vitro* and dissect supramolecular complexes; these uses allow for the study and measurement of the forces between interacting molecules, and also for the forces measured within a single molecule interacting with itself. Future AFM instruments are predicted to "allow similar measurements to be performed with organelles and living cells"[19]. These capabilities will give researchers information about "cellular trafficking and interactions" [19]. Also, AFM could be combined with optical microscopy in order to be able to measure multiple

signals at one time, which will allow for the study of cellular networks in greater detail than currently possible [19].

References

1. Howard, J., *Mechanics of motor proteins and the cytoskeleton*. 2001, Sunderland, Mass.: Sinauer Associates. xvi, 367 p.
2. Spudich, J.A., *Molecular motors take tension in stride.* Cell, 2006. **126**(2): p. 242-4.
3. Alberts, B., *Molecular biology of the cell*. 4th ed. 2002, New York: Garland Science. xxxiv, 1463, [86] p.
4. Hirokawa, N. and R. Takemura, *Kinesin superfamily proteins and their various functions and dynamics.* Exp Cell Res, 2004. **301**(1): p. 50-9.
5. Oiwa, K. and H. Sakakibara, *Recent progress in dynein structure and mechanism.* Curr Opin Cell Biol, 2005. **17**(1): p. 98-103.
6. Gibbons, I.R., et al., *Multiple nucleotide-binding sites in the sequence of dynein beta heavy chain.* Nature, 1991. **352**(6336): p. 640-3.
7. Vale, R.D. and Y.Y. Toyoshima, *Rotation and translocation of microtubules in vitro induced by dyneins from Tetrahymena cilia.* Cell, 1988. **52**(3): p. 459-69.
8. Reck-Peterson, S.L. and R.D. Vale, *Molecular dissection of the roles of nucleotide binding and hydrolysis in dynein's AAA domains in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2004. **101**(39): p. 14305.
9. Schroer, T.A., *Dynactin.* Annu Rev Cell Dev Biol, 2004. **20**: p. 759-79.
10. Reck-Peterson, S.L., et al., *Single-molecule analysis of dynein processivity and stepping behavior.* Cell, 2006. **126**(2): p. 335-48.
11. Uyeda, T.Q., S.J. Kron, and J.A. Spudich, *Myosin step size. Estimation from slow sliding movement of actin over low densities of heavy meromyosin.* J Mol Biol, 1990. **214**(3): p. 699-710.
12. Brady, S.T., K.K. Pfister, and G.S. Bloom, *A monoclonal antibody against kinesin inhibits both anterograde and retrograde fast axonal transport in squid axoplasm.* Proc Natl Acad Sci U S A, 1990. **87**(3): p. 1061-5.
13. Mallik, R., et al., *Cytoplasmic dynein functions as a gear in response to load.* Nature, 2004. **427**(6975): p. 649-52.
14. Toba, S., et al., *Overlapping hand-over-hand mechanism of single molecular motility of cytoplasmic dynein.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5741-5.
15. Allen, S., et al., *Measuring and visualizing single molecular interactions in biology.* Biochem Soc Trans, 2003. **31**(Pt 5): p. 1052-7.
16. Yang, Y., H. Wang, and D.A. Erie, *Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy.* Methods, 2003. **29**(2): p. 175-87.
17. Thomson, N.H., et al., *Oriented, active Escherichia coli RNA polymerase: an atomic force microscope study.* Biophys J, 1999. **76**(2): p. 1024-33.
18. Ishii, Y., A. Ishijima, and T. Yanagida, *Single molecule nanomanipulation of biomolecules.* Trends Biotechnol, 2001. **19**(6): p. 211-6.
19. Engel, A., Y. Lyubchenko, and D. Muller, *Atomic force microscopy: a powerful tool to observe biomolecules at work.* Trends Cell Biol, 1999. **9**(2): p. 77-80.

Appendix B. Methodology

To study the interaction forces between the microtubule-binding domain of the dynein heads and a microtubule, several preparations must be made in order to acquire high quality data that are interpretable. First the protein samples must be pure to perform the experiments. High quality images of the proteins must be obtained in order to determine the proteins' confirmations on mica surfaces. Force measurements must be made in the absence of proteins. High quality images of the proteins must be done in order to determine if complexes have formed. These force measurements can be compared to the actual force produced by protein-protein interactions and then quantified. This section describes the methods for purifying and setting up the biological component of the project, the AFM instrumentation necessary for imaging biological samples using fluid tapping mode, force spectroscopy, and the molecular recognition force microscopy (MRFM) experiments.

Purification of Dynein/Dynactin

To perform any biochemical assay, protein samples must be purified to remove any unwanted contaminants or interacting molecules. Protein purification varies from protein to protein; some can be quite complicated while others are relatively easy. Dynein and dynactin can be found in large concentrations in the brain and are usually taken from mouse or cow brain. Cytoplasmic dynein and dynactin for our experiments were co-purified from bovine brain by microtubule-affinity ATP extraction and sucrose gradient centrifugation as previously described Karki and Holzbaur 1995. The proteins were flash frozen in 25% sucrose in liquid nitrogen and stored at -80 °C. The protein samples were prepared at the University of Pennsylvania by the Holzbaur lab.

Microtubules

For the dynein-microtubule experiments microtubules must be prepared. Fresh stocks of microtubules are necessary, as they do not last for a long period of time due to rapid depolymerization. Lyophilized tubulin (>99% pure) from bovine brain was purchased from Cytoskeleton Inc. (Denver, CO). The tubulin sample was resuspended in 200 µl of ice-cold G-PEM buffer solution to make a 5mg/ml solution. PEM buffer consists of 100mM Na-PIPES, 1mM MgSO4 and 1mM EGTA at a pH of 6.8 obtained with KOH or HCL. PEM buffer was sterile filtered and stored at 4 C. G-PEM consists of a stock PEM buffer solution with 1mM GTP stored in -20 C. The tubulin solution was spun in an ultra-centrifugation machine for 10 minutes at 90k RPM at 4 C to remove any non-specific aggregates. After centrifugation GTP was added to make a concentration of 1mM. The solution was then incubated at 37 C for 20 minutes. Taxol is then added to make a concentration of 50µM, in order to stabilize the microtubule assemblies. Taxol solution consists of hydrate taxol powder in 100 µl of DMSO to 2mM. After taxol was added microtubules were incubated again at 37 C for 20 minutes. All microtubule samples were stored at room temperature in a dark area.

Buffers

Biological buffers must be prepared in order for biological molecules to remain in the folded, native form. Specific buffers must be prepared for individual proteins. Careful attention must be paid to pH, since proteins will unfold under non-native conditions. The main buffer used for all dynein experiments was the motility assay buffer (MAB). A 1L solution of MAB was prepared and contained the following: 10mM Na-PIPES (3.35g/L), 50mM potassium acetate (4.9g/L), 4mM MgSO4 (4ml of a 1M stock), and 1mM EGTA

(0.38g/L). The pH was adjusted to 7.0 using KOH/HCl. All buffers used in our experiments were filtered with a 0.22 μm filter from Millipore to remove any particulates in solution.

Mica Surfaces

Flat surfaces are required to visualize single molecules with an AFM. Mica is a very easy solution to this problem because it is extremely flat and is a charged surface. Mica samples were purchased from SPI supplies (2spi.com). The mica was freshly cleaved by removing a layer with adhesive tape. Freshly cleaved mica surfaces were secured to metal sample disks with non-conducting double-sided tape.

Fluid tapping imaging of dynein/dynactin and microtubules

To observe single molecules in their native form they must be in a buffer that best matches true physiological conditions. Atomic force microscopy is the only tool that can image single biological molecules in solution at near-atomic resolution. Sub-molecular features such as protein dynamics and structure-functional properties can be studied by AFM because of the high signal to noise ratio for single molecules [1].

Imaging in fluid is essentially the same process as using tapping mode or contact mode in air, however a fluid cell is required for containing the liquid and mounting a probe. Fine adjustments must also be made to the AFM to correct for refractive effects as the laser goes through the liquid-air boundary. The cantilevers will behave slightly differently in liquid due to damping. Noise levels can be higher in liquid as a result of cantilever dampening and molecules freely diffusing in the liquid.

To image samples and surfaces in liquid, a fluid cell must be used. A fluid cell consists of a small metal assembly with a glass window, rubber skirt, and holder for an

AFM probe. The glass window provides a level surface for the laser beam to pass into the liquid. Without this level surface the laser can be distorted or scattered by unstable liquid surfaces. The fluid cell also has openings for fluid inlet and outlet. Fluid can be injected into the fluid cell by using plastic tubing and a clean syringe. This is useful for buffer exchange in many biological experiments.

Figure B1. Fluid cell for the Dimension 3100 Microscope. Fluid cells allow for imaging in both contact and non-contact mode in the presence of liquid. This cell has a glass cantilever holder for positioning the AFM probe. The o-ring is used to prevent liquid from leaking out the cell (Vecco.com).

Before using the fluid cell it must be clean to prevent any unwanted contamination. The fluid can be cleaned by soaking it in water with soap. While the fluid cell is wet, the glass surface can be cleaned with a Q-tip or Kim-wipe. Careful attention must be paid when cleaning the glass surface because it can be easily scratched. Scratches on the surface can distort the laser and cause problems with aligning the laser. The fluid cell can be aired dry or blown dry with nitrogen. This process can also be done with ethanol or methanol, but the skirt must be removed and dried quickly as these chemicals can dissolve the glue holding the glass surface in place.

Cantilever selection for use in fluid tapping depends on the sample. Cantilevers with low stiffness have been known to work well with biological samples. Before the

cantilevers were mounted, they were exposed to UV light for 15-20 minutes to remove any contaminants. Clean cantilevers were then mounted on the fluid cell for imaging. Samples also have to be prepared in advance. It is necessary to incubate the sample on the mica surface to allow the proteins to stick to the mica surface. For the dynein and microtubules, incubation times of 20-30 minutes were used.

Once the cantilever has been mounted on the fluid cell, clean tubing was positioned on the fluid cell with two syringes at each end. Before the AFM scan head is placed on the stage, the fluid cell should be pre-wetted with the fluid used for the experiment. Fluid should also be put into the injecting syringe. Once placed on the stage, a small of liquid should be injected to remove air bubbles from the line. This can be done until a small drop is suspended on the fluid cell, held in place by surface tension. At this point the fluid cell can be lowered onto the sample surface. Once the fluid cell has made contact with the surface more fluid can be injected into the cell. This is done to remove air bubbles and inject buffers for experiments. If air bubbles still exist in the fluid cell, they can be removed by quickly injecting more fluid. If too much liquid is injected, it may come out the top of the liquid cell and leak onto the top of the glass surface. This liquid should be removed immediately or it will distort the laser, preventing proper alignment. The process may be slightly more complicated when using buffer solutions as opposed to water.

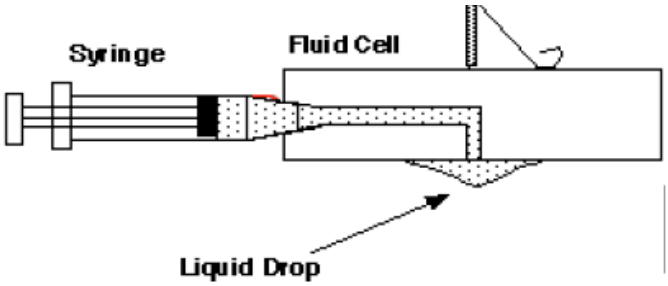

Figure B2. Fluid cell with pre-wet surface a small amount of liquid is injected before lowering the fluid cell to the surface. This removes air in the tubing line and also helps prevent air bubbles when the cell makes contact with the sample surface (veeco.com).

Once the fluid cell skirt is touching the surface and the cantilever is close to the surface, the laser can be adjusted and a frequency calibration should be done. This is done at this point so that the cantilever's measured resonant frequency will be accurate in the fluid cell environment near the sample. Measuring the frequency in air is not necessary, but it is useful to get an idea of the magnitude of the frequency. Frequencies in air tend to create a single, sharp peak. For tapping mode the drive frequency is selected as a bit to the left of the maxima; this is not usually the case in fluid. Dampening and motion from the liquid can cause multiple peaks to form; this can be overcome by increasing the drive amplitude percentage. If a range of peaks exists, the cleanest and tallest peak was selected.

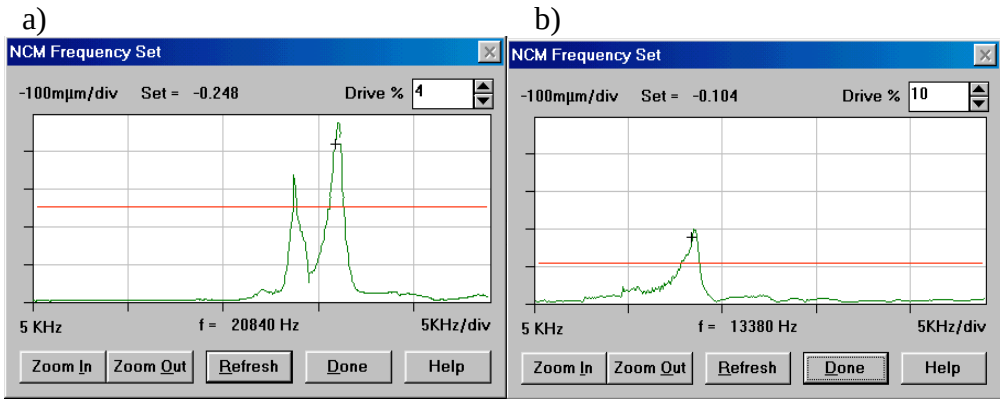

Figure B3. Frequency sets in air and liquid. Figure a) shows a frequency sweep done in a liquid cell in the absence of fluid. A single resonance peak can be seen at low drive frequencies. Figure b) shows a frequency sweep of the same cantilever in a liquid cell in the presence of water. Higher drive frequencies are necessary in order to resolve a sharp, distinguishable peak.

After the laser has been adjusted and the frequency in the fluid environment has been obtained, the sample is now ready to be imaged. Optimization of the imaging process in fluid can be difficult because of noise levels. The set point must not be raised too high or the tip can alter or damage the protein. Scan rates higher than 1.2 Hz can miss small molecules and/or features of interest in protein samples, so the average rates are about 1 Hz.

Once the images are obtained they must be processed to remove noise, increase resolution and obtained topographical information. Imaging software can be used to clean up images using methods such as Fast Fourier Transforms (FFTs), rank filters and other image refining methods. It can also be used to generate histograms for sample surfaces. This is necessary when obtaining near-atomic resolution images of protein samples.

Molecular Recognition Force Microscopy (MRFM)

AFMs are excellent tools to determine adhesive forces between molecules. By attaching a biological molecule to an AFM tip, interactive forces can be measure by

bringing the tip close to the substrate surface then retracting the tip. Previous work has been done on DNA, RNA and several proteins [1]. Molecular Recognition Force Microscopy (MRFM) is useful extension of biophysical experiments. By attaching specific biological molecules to the tip, intramolecular forces can be measured between the molecule and a particular substrate of interest. Previous AFM work has been done on the folding mechanism of titin, a protein necessary for in the contraction of striated muscle tissues [2]. More recent experiments have been done with force recognition with simultaneous imaging with lyzosyme and surface immobilized antigens [3].

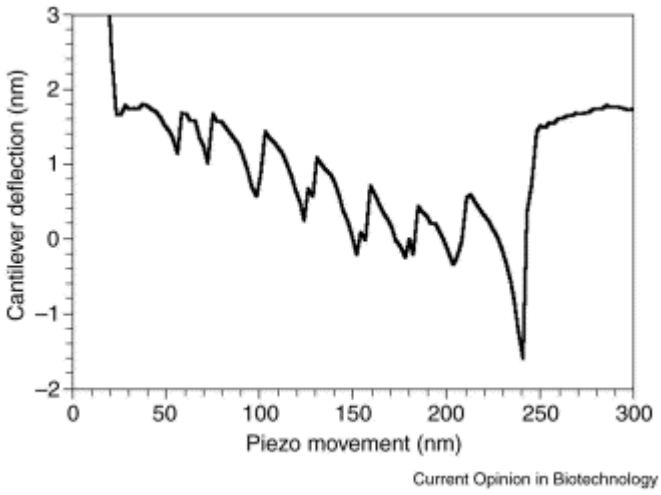

Figure B4. A Force spectroscopy curve of a titin molecule. An AFM tip was pressed on a gold surface covered with a low concentration of titin molecules. The cantilever was lowered for 0.5s then retracted for 1.2 mm/s. Each minimum on the force curve corresponds to a force maximum where an individual domain of a titin molecule was unraveled. The distance between the peaks corresponds to the length of the unraveled domain [2].

Many proteins generate forces in the cell. Motor proteins exert forces to drive various cellular processes. They are required for cellular mitosis and the movement of proteins and large complexes [4]. AFMs are excellent tools to measure such forces on the molecular level. Extensive work has been done to measure force production of single kinesin molecules using optical trapping [5]. Recently single molecule optical trapping

experiments have also been applied to dynein [6]. However, the mechanism of movement remains controversial [7]. Optical trapping has been the conventional method for measuring force production for motor proteins. AFM offers an alternate method for acquiring force measurements that can help confirm data from optical trapping experiments. An AFM can be used in similar manner to optical trapping to obtain force measurements, replacing the bead to a fixed cantilever.

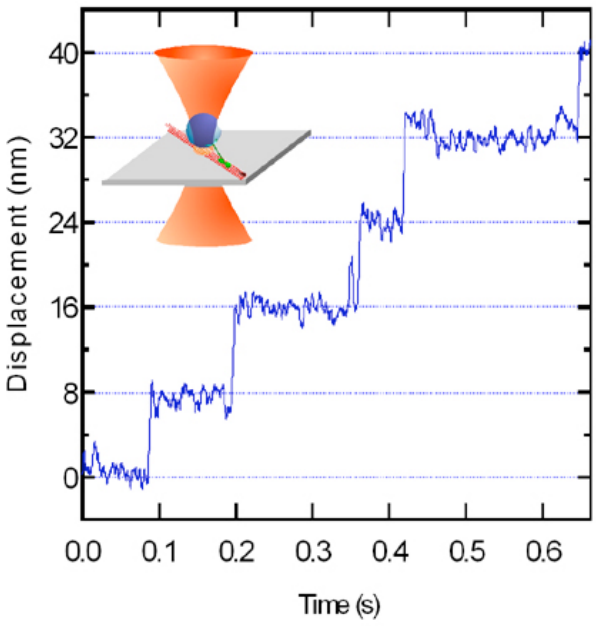

Figure B5. Optical trapping experiments done on kinesin to determine step size. Kinesin molecules were coated on a polystyrene bead and the laser was focused on the bead. Changes in momentum of the bead were measured and the step size was determined (http://www.stanford.edu/group/blocklab/Optical%20Tweezers%20Introduction.htm).

To measure force production by a single dynein molecule, dynein must be coated onto beaded cantilevers and lowered onto a mica surface containing microtubules. Tipless cantilevers from (spmtips.com) are necessary so that beads can be mounted on the tip. Once the cantilever with dynein fixed on the bead comes within a short distance of a microtubule, the dynein protein will bind tightly to the microtubule. The force of

adhesion of a single dynein molecule can be measured by retracting the cantilever until the dynein protein breaks free. This can be troublesome if the ΔG of unfolding is less than that of the binding between the dynein head-microtubule binding site and the microtubule. Careful attention must be paid to interpreting the force curves.

References

1. Engel, A., Y. Lyubchenko, and D. Muller, *Atomic force microscopy: a powerful tool to observe biomolecules at work.* Trends Cell Biol, 1999. **9**(2): p. 77-80.
2. Tskhovrebova, L. and J. Trinick, *Extensibility in the titin molecule and its relation to muscle elasticity.* Adv Exp Med Biol, 2000. **481**: p. 163-73; discussion 174-8.
3. Allison, D.P., P. Hinterdorfer, and W. Han, *Biomolecular force measurements and the atomic force microscope.* Curr Opin Biotechnol, 2002. **13**(1): p. 47-51.
4. Howard, J., *Mechanics of Motor Proteins and the Cytoskeleton*. Vol. Sinaure Associates, Inc. 2001, Sudnerland, Massachusetss.
5. Kuo, S.C. and M.P. Sheetz, *Force of single kinesin molecules measured with optical tweezers.* Science, 1993. **260**(5105): p. 232-4.
6. Toba, S., et al., *Overlapping hand-over-hand mechanism of single molecular motility of cytoplasmic dynein.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5741-5.
7. Reck-Peterson, S.L., et al., *Single-molecule analysis of dynein processivity and stepping behavior.* Cell, 2006. **126**(2): p. 335-48.