Computational Protein Design

Melissa Beswick
beswickm12@gmail.com

Bhavin Gala
bhavinpgala@gmail.com

Rebecca Sheng
rebecca01682@yahoo.com.tw

Anne Kim
annekim94@gmail.com

Sanjana Matta
sanmat75@yahoo.com

1 Abstract

Heat shock proteins, abbreviated as Hsp's, are chaperone proteins that protect, fold, and transport substrate proteins under stressful conditions, such as heat, acidity, and aggregation of proteins within the cell. By protecting proteins in tumor cells and allowing them to continue multiplying, some Hsp's can prolong the lives of cancer cells [5]. As a result, researchers are trying to inhibit the ATP binding site in Hsp's to prevent them from gaining energy. In order to understand the inhibition of Hsp, we investigated the structure of Hsp's and hypothesized that there are analogous structures in all Hsp's that allow them to resist heat. Using computational tools like PyMOL, the Protein Data Bank, Jalview, and Visual Molecular Dynamics simulations, we found that some Hsp's, like Hsp110 and Hsp70, had some similarities in structure. However, our data suggested that most members of the Hsp family have different structures, which determine their various functions. Our hypothesis was not supported by our data, but this may not be entirely conclusive because computational tools lack the accuracy and efficiency of experiments in real laboratories. In future work with heat shock proteins, virtual data is suggested only as a resource for conjecture rather than conclusion.

2 Background

A protein is "a functional biological molecule consisting of one or more polypeptides folded and coiled into a specific three-dimensional structure" [4]. These macromolecules are essential to life because they are involved in nearly all cellular functions. Some proteins are catalytic enzymes that speed up chemical reactions. Other proteins provide structural support, storage, transport, cellular communications, movement, and defense against foreign substances [5]. From protecting extreme thermophiles from high temperatures to making up the human brain, proteins support many functions of life due to their varying structures [4].

Proteins are made up of chains of amino acids called polypeptides. There are 20 different amino acids in existence, which are identified by their carboxyl and amino groups. A finite portion of an organism's DNA codes for a specific chain of amino acids. The chains vary in sequence depending on the function of the protein. Once the chains are created, they fold into their necessary secondary structures, which include α helices, β pleated sheets, and random coils [4].

If a polypeptide chain does not fold properly or if there is a mutation that results in an inaccurate amino acid sequence, a protein will not be able to serve its intended purpose. This type of mistake can lead to many fatal problems. Fortunately, there are chaperone proteins that facilitate the folding of proteins. These chaperones prevent protein aggregation and provide a location for client proteins to fold or refold. Although this does not always work as intended,

chaperone proteins increase the rate of accurate protein synthesis in cells [5].

Within the chaperone family, there is a unique group classified as “heat shock” proteins. When proteins are exposed to heat or other environmental stresses, they can denature and lose their ability to function [8]. Some effects that induce the synthesis of these stress proteins include heat shock, cold shock, ultraviolet radiation, amino acid analogues, alcohol, excess oxidation, excess reduction, a deficit of ATP, and osmotic shock. Rather than denaturing under stressful conditions, heat shock proteins are stimulated by stressful conditions and are triggered to protect and refold proteins [5]. There are many different sizes and varieties of heat shock proteins, so their functions vary. However, they do have a few main contributions to cells: they prevent the aggregation of proteins, assist in protein folding and re-folding, stabilize proteins and maintain their state, and promote the proper formation of disulfide bridges. Overall, they help maintain homeostasis within cells [8].

3 Introduction

As the prevalence of cancer continues to increase in society, scientists are attempting to develop new cancer treatments and cures, one of which is heat shock protein inhibition. When cancer cells develop in the body, heat shock proteins are unable to detect the potential harm that these cells can cause. Instead, they respond to the stress that tumor cells are experiencing, such as oxygen and glucose deprivation. Various cancer treatments, like radiation therapy and chemotheraphy, also increase heat shock protein induction in response to the stress forced upon cells; the increased number of heat shock proteins gives the tumor cells a great chance of survival.

Heat shock protein 90 (Hsp90) is one of the chaperones that protects proteins within cancer cells when the cancer cells are undergoing treatment and increased stress. Researchers have been working to prevent these proteins from getting the energy they need to function. On a molecular level, adenosine triphosphate (ATP) provides energy. Hsp90 has an ATP binding site which researchers have been trying to block. Recently, they have been working with geldanamycin, a chemical that competitively binds to the ATP binding site of Hsp90. Instead of allowing ATP to fit into the binding site and provide energy to the Hsp90, scientists have been trying to block this site with geldanamycin. This research will hopefully bring scientists closer to finding a cure for cancer.

The goal of this project was to determine whether or not heat shock proteins have specific structures that make them resistant to heat shock and other environmental stresses. If heat shock proteins do have specific heat resistant structures, it would be possible to find ways to target them and prevent them from protecting cells, such as tumor cells, that are detrimental to the survival of an organism.

The hypothesis for this lab was that heat shock proteins have specific structures in common that allow them to perform similar functions. By focusing on Hsp31, Hsp33, Hsp70, Hsp90, and Hsp110, we predicted that these proteins would share analogous structures; the number following each Hsp indicates the protein’s molecular weight in kilodaltons.

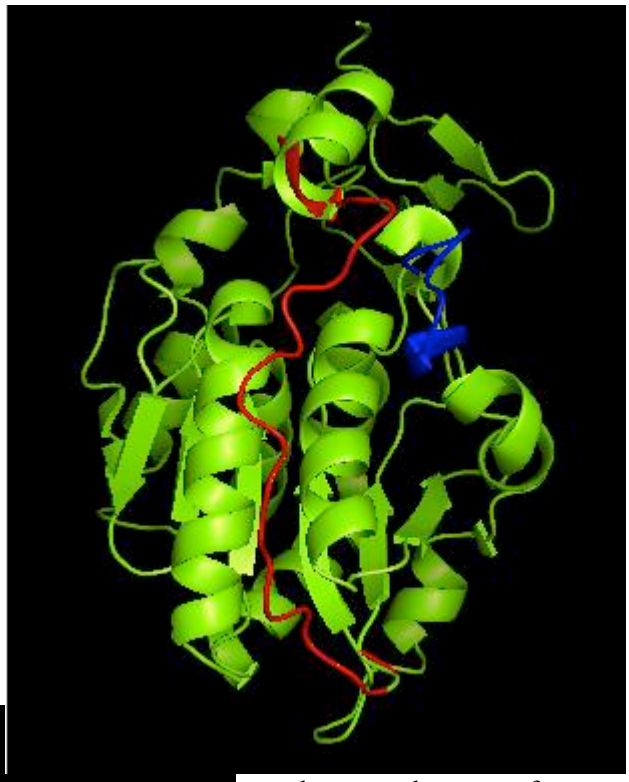
4 Literature Review

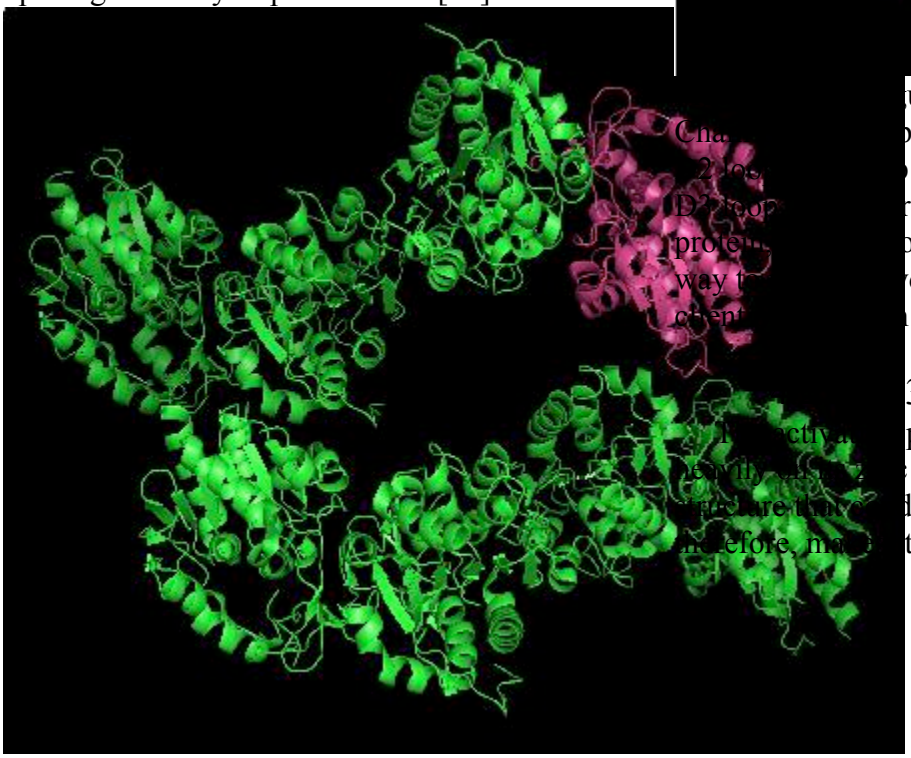
4.1 Hsp31

Hsp31 members are distinguished by differences in their catalytic triads and quaternary structure. A catalytic triad includes the three amino acids that are located within the active site of specific enzymes. Quaternary structure is the structure of a protein that contains two or

more polypeptide chains folded into their necessary shape [5]. EcHsp31, the specific Hsp31 focused on in this report, interacts with denaturing proteins in the event of extreme heat and releases the proteins back into the cell once the temperature cools [4].

Flexibility is extremely important to the success of EcHsp31; when there is heat, specific loops on the protein move around so that it can perform its proper function as a chaperone. The most prominent flexible loops are the D2 and D3 loops [12]. When the D2 loop moves under heat shock conditions, it exposes the residues that create a hydrophobic portion of the protein. This hydrophobic portion can possibly be used for protein intermediates to fold properly. D3 is influenced by D2's movement and moves when D2 does, completing the opening of the hydrophobic bowl [12].





ure shows a close up of
highlighted in red is the
highlighted in blue is the
ess is applied to this
loops move out of the
drophobic patch to which
bind.

3

process of Hsp33 relies
finger motif, a unique
denses the protein and,
t more stable. [9]

Figure 1: This image shows a complete Hsp31. Hsp31 has 8 separate chains that combine to make up its quaternary structure. The chain that is highlighted in red is Chain H. Chain H was the most complete chain when the protein was crystallized.

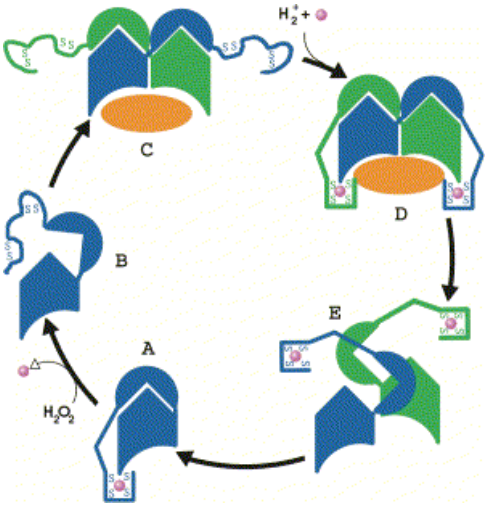


Figure 3 **The Hsp33 Cycle**: Stage A indicates the monomeric state of Hsp33. Stage B shows the loosening of the monomer. Stage C shows the dimerization of two Hsp33 monomers. Stage D illustrates the Hsp33 dimer. The orange portion of stage C and stage D represents the area where the substrate protein binds for protection and transportation. Stage E shows the dissociation of the Hsp33 monomers [9].

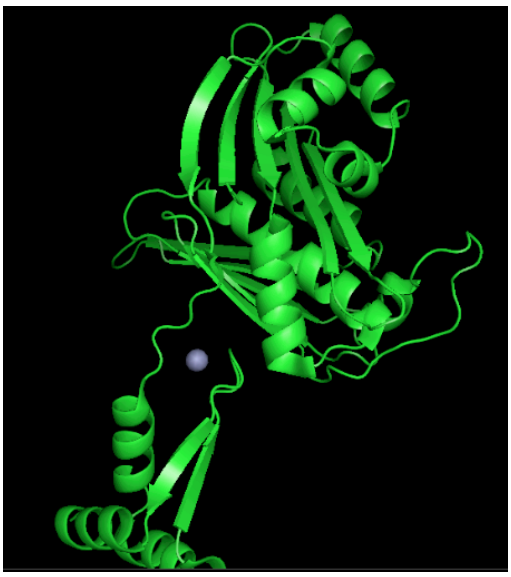


Figure 4: The green molecule is an Hsp33 monomer; the gray sphere indicates zinc.

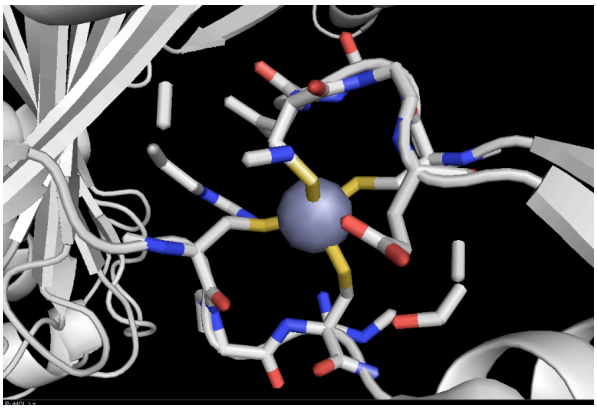


Figure 5: Close view of Zinc on an Hsp33 monomer with nearby cysteine residues

Stage A:

As shown in Figure 4, zinc, colored gray, is located in the end pocket of the Hsp33 monomer. Figure 5 shows four cysteine residues. The yellow portion of the residue indicates the element sulfur. The four cysteines, Cys268 and Cys271, Cys235 and Cys237, form two strong disulfide bonds along with a network of other hydrogen bonds. These bonds are responsible for stabilizing Hsp33 [9].

Stage B:

Once exposed to stressful environments, the monomeric Hsp33 releases zinc, which breaks the disulfide bonds and hydrogen bonds. This allows the structure of Hsp33 to become loose [9].

Stage C:

Once Hsp33 monomer is loose, it structurally combines with another Hsp33 [9].

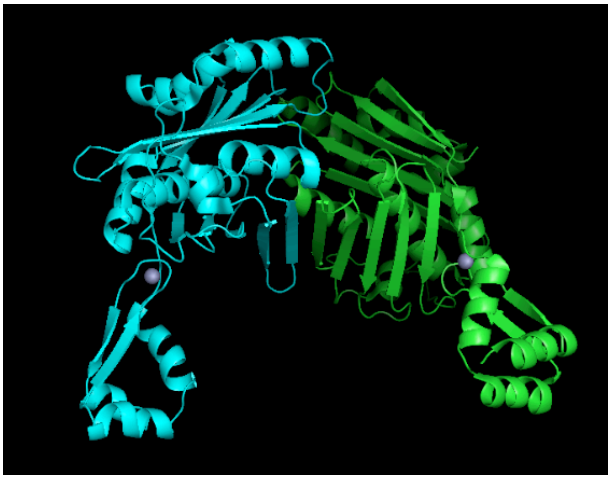


Figure 6: Hsp33 Dimer

Stage D:

The zinc of each Hsp33 then reappears and causes its nearby residues, cysteines, to re-form disulfide bonds and hydrogen bonds. In other words, zinc comes back and locks the two monomers together, forming a dimer, as shown in Figure 6. This dimerization process allows Hsp33 to become active [9]. In its active state, Hsp33 acts as chaperone, monitoring protein folding processes and preventing the interactions of misfolded proteins with other normal proteins [8].

Stage E:

After Hsp33 performs its functions, the dimer dissociates; each Hsp33 monomer becomes inactive, completing the Hsp33 cycle [9].

4.3 Hsp70

4.3.1 Structure and Function

Hsp70 is a chaperone protein that involves in several roles throughout the cells of eukaryotes, eubacteria, and archaea [11]. Apart from conducting the traditional tasks of a chaperone, such as transporting client proteins across membranes, protecting client proteins from unfavorable environments, and regulating protein folding, Hsp70 also degrades misfolded proteins and prevents apoptosis, or cell death. Moreover, this chaperone folds proteins as soon as they are formed at the ribosome by detecting any hydrophobic regions of the emerging proteins and assisting them to fold into a conformation that promotes stability.

As in all proteins, Hsp70's ability to perform these duties is defined by the domains and terminals of its protein structure as well as its interactions with substrates at locations called activity or binding sites. Hsp70 remains active at the N-terminal, C-terminal, J-domain, peptide-binding domain (PBD), and nucleotide-binding domain (NBD). Proteins from the J-domain of Hsp70, such as Hsp40, are involved in the catalysis of protein activity and the "recruitment" of substrates to the PBD [11]. The PBD is where client proteins bind to Hsp70. Activity at the NBD, where ATP, ADP, and other nucleotides attach to Hsp70, determines Hsp70's ability to attach to client proteins at the PBD. This regulation of the attachment of substrates to binding sites is called an allostery, and it is facilitated by a flexible linker region between the NBD and PBD (Figure 7A).

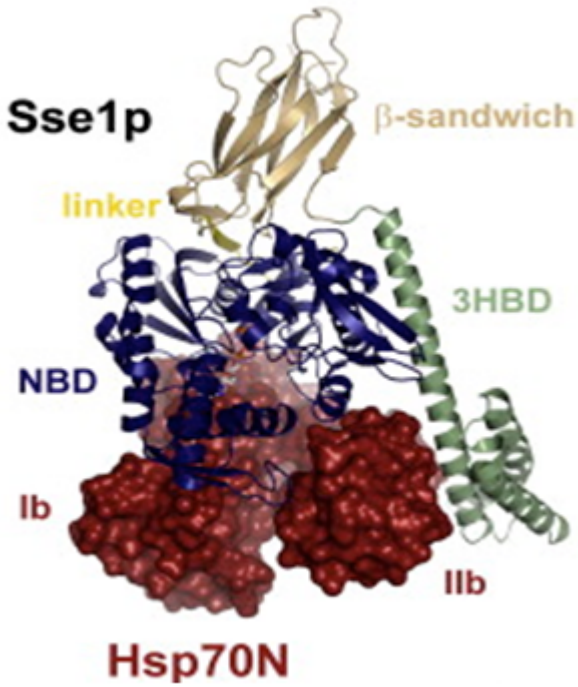
The ATP-ADP cycle in Hsp70 regulates the heat shock protein's binding with client proteins. When ATP is attached to the NBD, Hsp70 is free to attach to client proteins since the lid of Hsp70 is open. The enzyme ATPase then hydrolyzes ATP, which is the addition of water to turn the ATP nucleotide into an ADP molecule and an inorganic phosphate group that is later released. During this hydrolysis, the now ADP-bound Hsp70 tightens around the PBD to such an extent that it cannot bind to or release any molecules, including the client protein. In order to open the Hsp70 binding site and thus reactivate Hsp70, ADP must be

exchanged for ATP. The ATP-ADP cycle repeats from here [6].

4.3.2 Hsp70-Hsp110 Complex

Since Hsp70 hydrolyzes ATP too slowly, it needs the help of both ATPase and the J-domain protein Hsp40 as catalysts of this reaction. As ATPase and Hsp40 hydrolyze the ATP bound in the NBD of Hsp70, Hsp70 folds the client protein in the PBD into its native form. Because ATP changes into ADP as a result of hydrolysis, the Hsp70 tightens around the client protein (Figure 7B). Only the binding of ATP-bound Hsp110, which exchanges the used ADP on Hsp70 into ATP, can reverse the tight, conformational change of Hsp70, allowing Hsp70 to release the properly folded client protein (Figure 8) [6].

A



B

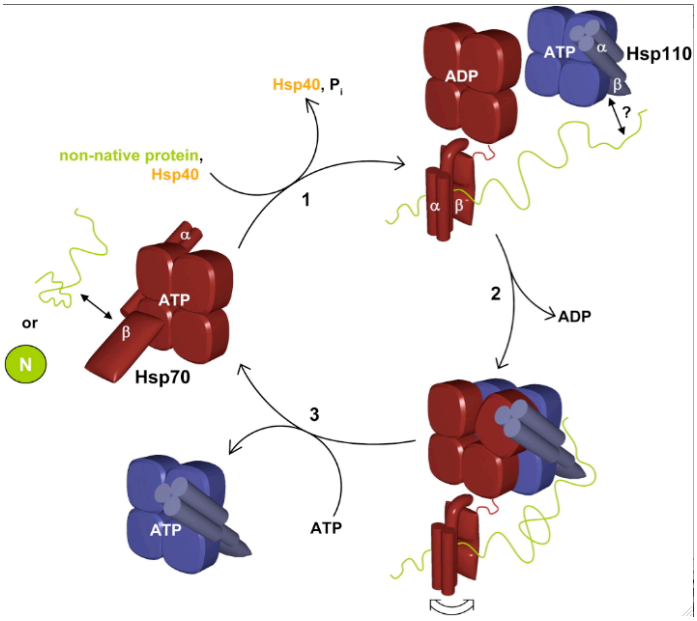


Figure 7: **(A)** Attachment of Hsp70N (red), which is the nucleotide binding site (NBD) of Hsp70, to the NBD of Sse1p, which is a nucleotide exchange factor (NEF) of Hsp110. The linker region between the NBD (in blue) and PBD (beta sandwich) of Sse1p is colored in yellow. The three helix bundle domain (3HBD) is colored in green [10].

(B) The figure above shows the process that Hsp110 and Hsp70 undergo to properly fold a substrate protein as well as the ATP-ADP cycle occurring throughout the nucleotide exchange process[10].

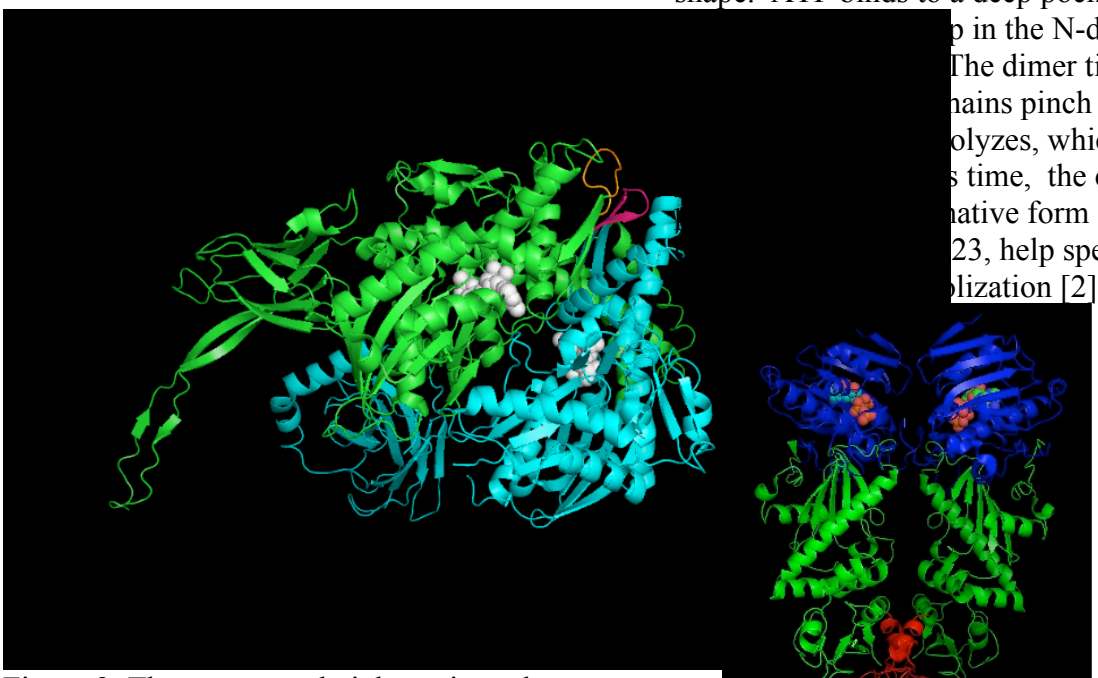


Figure 8: The orange and pink sections show the linker regions of Hsp110 and Hsc70, respectively [12]. The white spheres show ADP.

4.4 Hsp90

Hsp90 is a chaperone protein that is present in both eukaryotic and prokaryotic organisms. Its primary functions are to support protein maturation and signal transduction. Hsp90 exists as a dimer in the cytoplasm of cells. The levels of Hsp90 increase within the membrance when cells are under stress. Hsp90 is divided into three domains: the N-domain, the M-domain, and the C-domain (Figure 9). The N-domain contains an ATP binding site. The M-domain, or middle domain, is where the client protein binds. The C-domain is the dimerization domain. The C-domain experiences conformational changes which open and close the entire Hsp 90 dimer [14].

Hsp90 folds around proteins in order to complete its job and assist its clients. Hsp90 goes through an ATP cycle. In the beginning, the Hsp 90 dimer is in a "V" shape. ATP binds to a deep pocket in the N-

p in the N-domain to
The dimer tilts 20 Å
ains pinch together.
olyzes, which takes
s time, the client
ative form [14]. Co-
23, help speed up the
lization [2].

Figure 9: The blue section is the N-domain, the green section is the M-domain, and the red section is the C-domain. The ATP molecules are in in the N-domain.

5 Methods

5.1 Protein Databases

Resources such as the Protein Data Bank and the Universal Protein Resource consolidate thousands of proteins into their own databases. These organize information from previous scientific research papers into categorized information tables. The literature published by scientists contains the protein's taxonomy and amino acid sequences. The databases proved to be significant sources of information when researching the different chaperone proteins.

On the PDB website, each protein's page contains a visual representation of the

protein's structure. In order to determine the crystal structure of a protein, it must be crystallized out of a solution and undergo x-ray diffraction. In order to determine the crystal structure of a protein, the protein must be crystallized out of a solution and undergo x-ray diffraction. The results of this procedure are published as secondary structures. Each of the proteins on the PDB has a unique code, making it easy to access any one of the over 70,000 proteins in this resource.

5.2 PyMOL

5.2.1 General Features

PyMOL is a molecular graphics software that allows for the virtual manipulation of proteins that are "fetched" from the Protein Data Bank using the PDB codes. Through PyMOL, users have the ability to manipulate the structures of obtained proteins for analysis. Proteins are imported in the default form of "sticks," which represent the bonds between every atom in the molecule and can be changed to show elements, three-dimensional secondary structures, and other forms that allow for more efficient structural analysis. Any differences between areas of proteins were shown by color-coding the structure (Figure 10). All such operations allow the user to conveniently visualize different areas of the protein and understand the connections and interactions within its structure.

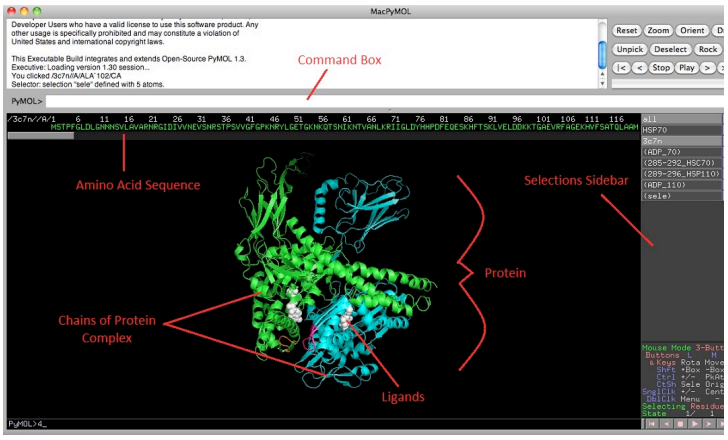


Figure 10: Screenshot of PyMOL software. The PDB codes are entered in the command box to display a protein in the viewing window. The protein here is shown in "cartoon" form as secondary structures and colored by chain. Ligands are shown as spheres in white. The sidebar allows for the manipulation of all the proteins being used in that PyMOL session.

5.2.2 Surface Manipulation

In this research project, PyMOL was used to determine the defining features in each heat shock protein's structure in order to understand its overall function in the cell. For instance, the surface charges, hydrophobic regions, and hydrophilic regions of the protein were studied to indicate the different interactions a protein could have with its surroundings. This information could also be used to understand the manner in which a protein folds (Figure 11). Scripts for these operations were obtained from the website PyMOLWiki, which contains a script library capable of emphasizing certain features on virtual proteins.

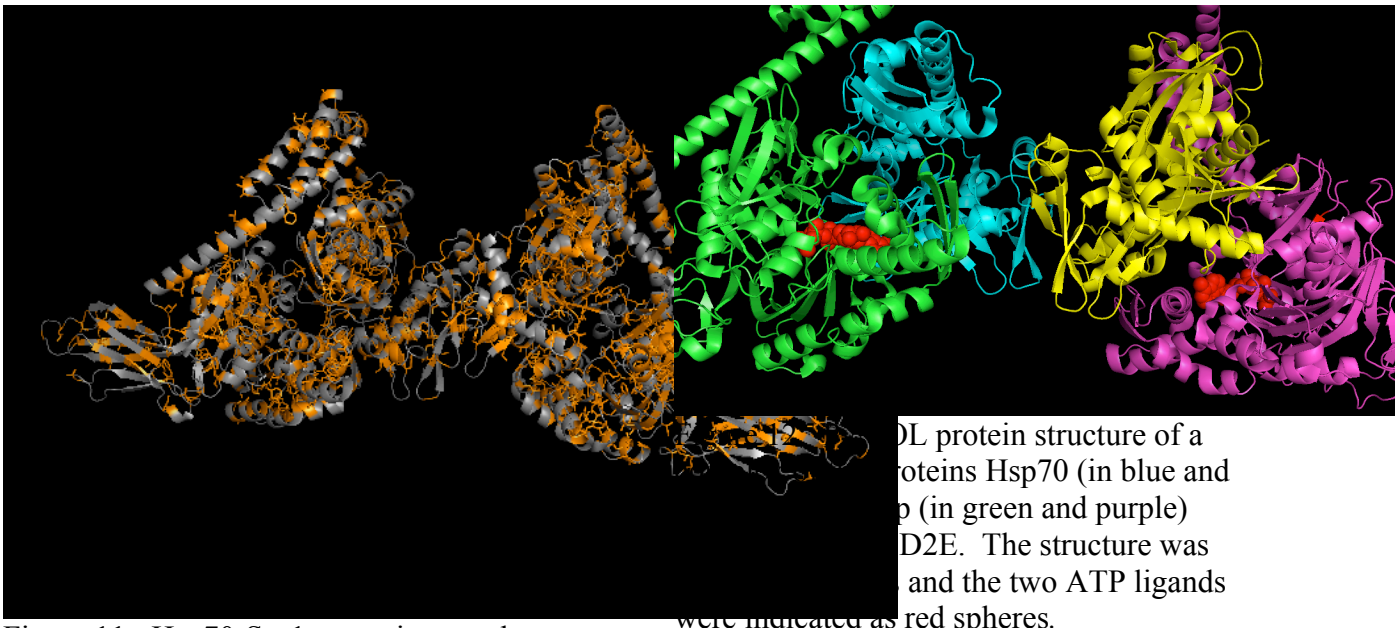


Figure 11: Hsp70-Sse1p protein complex (PDB code 3d2e) colored orange in hydrophobic areas.

5.2.3 Ligands

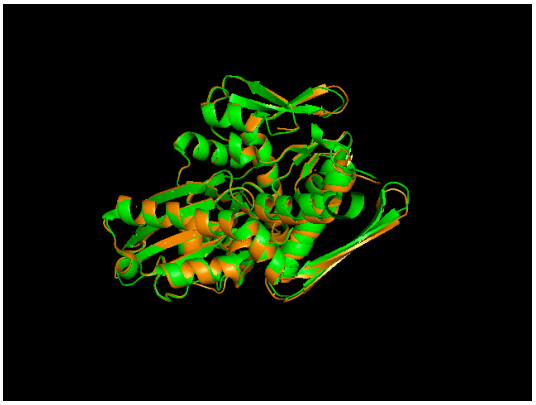
A major portion of the understanding of a protein comes from the protein's interactions with other molecules and substrates at its active sites. Active sites are where molecules attach to the protein in order to affect its performance. By viewing the amino acid sequence displayed on the PyMOL viewing window, the ligands, which are the molecules attached to the protein's active sites, were located and then emphasized through spherical representations (Figure 12). This process indicated the presence and location of binding sites on the chaperone, which are the main centers of activity in a protein. Therefore, we predicted that any changes made to these locations would severely alter the functionality of the chaperones. This was the basis for most of our computational experiments on each protein.

)L protein structure of a roteins Hsp70 (in blue and p (in green and purple) D2E. The structure was and the two ATP ligands were indicated as red spheres.

5.2.4 Alignment

By virtually depicting similarities in protein structures, PyMOL can align proteins in terms of amino acid sequences. Every similar structure and sequence is matched with the others through a line. A shorter line indicates greater similarity between two regions, and any outliers in the alignment of amino acids are rejected (Figure 13). Therefore, the final representation of the alignment between two proteins in PyMOL is actually their conservation, which is the presence of recurring similarities in the proteins. This feature of PyMOL was used to understand whether or not the heat shock proteins had any sequential or structural similarities.

A



B

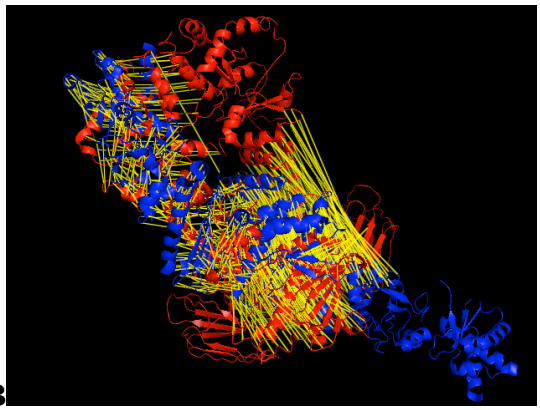


Figure 13: **(A)** This image shows alignment of Hsp70 proteins, PDB codes 3qfp (orange) and 3qfu (green). Notice the minimal alignment lengths, indicating significant similarity between the proteins.
(B) This image shows the alignment of an Hsp90 closed chaperone complex (PDB code 2cg9, red) and the same Hsp90 as an open chaperone complex (PDB code 2ioq, blue). Notice the relatively large alignment lengths, indicating large differences in structure between the two proteins due to different conformations.

5.3 Jalview

Jalview aligns multiple sequences at a time. It is written in Java. Jalview shows how sequences align and provides a variety of functions such as comparing the sequence identities and color coding particular sections. Jalview was used in this project to align Hsp31, Hsp33, Hsp70, and Hsp90.

5.4 Simulation

Using Amber script format, a program for Hsp31 and Hsp33 was written to simulate behavior under different temperatures and with mutations to amino acids. The scripts dictated the size of the simulation, the protein observed, ions in the virtual solution, and the explicit solvents. Although the specificity in the script makes the simulation take a long time to run, it ultimately results in a more accurate simulation.

An advanced ten core processor and a graphics processing unit were used to run the Virtual Molecular Dynamics (VMD) program, which is a molecular visualization program. Even with this faster computer, the VMD takes about a day to yield 5 nanoseconds of simulation data. It takes this long because the program must keep track of each atom's movement while also considering the medium, temperature, mutations, and ion interactions. Once the initial coordinates of each atom are inputted into the program, the program calculates the force of the atom using Newton's second law of motion, generating a "force field" that predicts the next movement of the atom. Once the new position is calculated, the process is repeated. These calculations cause the simulations to take a very long time as they show the movement of the atoms.

6 Results and Discussion

6.1 Jalview Multiple Sequence Alignment

Jalview was used to align Hsp31, Hsp33, Hsp70, and Hsp90. There were 0%

similarities between Hsp70 and Hsp33, Hsp31 and Hsp70, Hsp90 and Hsp33, Hsp90 and Hsp70, and Hsp90 and Hsp31. However, there was a 34.78% similarity in the sequences of Hsp31 and Hsp33. This data can be seen in Table 1. They only have similar structures because they are both small heat shock proteins and are part of the same heat shock protein family.

Since there were not many significant similarities in the sequences of the heat shock proteins, the focus of this project was changed. The initial intent was to compare heat shock protein structures to each other. It then changed to the analysis of individual heat shock proteins themselves.

Hsp Pairs	**Percentage ID**
70 and 33	0%
31 and 33	34.78%
31 and 70	0%
90 and 33	0%
90 and 70	0%
90 and 31	0%

Table 1:
This table shows the separate heat shock protein pairs and their pairwise identifications. All the pairs were 0% similar except for Hsp31 and Hsp33. Hsp31 and Hsp33 had a 34.78% similarity.

6.2 Simulation of Hsp31 Under Different Temperatures

The simulation using VMD technology tested chain H of Hsp31 at 300 K for 10 nanoseconds and at 400 K for 5 nanoseconds. At 300 K, the protein jostled due to the presence of water molecules, but the temperature was not high enough to actually see the loops move. At 400 K, the protein fluctuated between a looser and tighter state at a faster pace. There were also times when the D2 loop and D3 loops moved. The D2 loop moved dramatically to the right, exposing hydrophobic beta pleated sheets below. The D3 loop also moved, exposing the same region at the center of the protein. The temperature was not high enough for the heat shock protein to demonstrate its full function, but the simulation did show this protein's relative stability under stressful conditions and its ability to expose certain hydrophobic regions.

Figure 14: On the left, this image shows chain H of Hsp31 moving in 300 K. The D2 and D3 loops are still relatively in place. The D2 loop is red and the D3 loop is green. On the right, the image shown is of the same chain at 400 K. The temperature difference causes the loops to move. The D2 loop move significantly to the left. The loops move to expose a hydrophobic patch at the center of the protein. Client proteins can then bind to this patch.

6.3 Simulation of Hsp33 Under Different Temperatures

The Hsp33 simulation tested the zinc finger structure under 300 K for 10

nanoseconds, 400 K for 10 nanoseconds, and 600 K for 9 nanoseconds. The simulation took out zinc from the Hsp33 structure and mutated its nearby residues, which were made of cysteine, to alanine. The purpose of this was to explore the importance of zinc finger. Zinc, along with disulfide bonds and a network of hydrogen bonds, stabilize the structure of Hsp33[9]. By mutating cysteines to alanine, we predicted that strong disulfide bonds would not form and that the structure would fall apart.

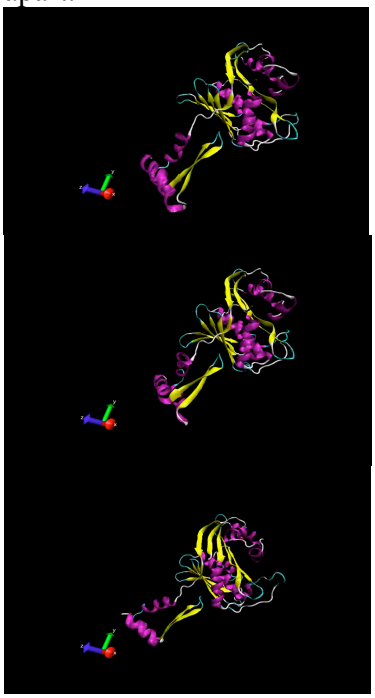


Figure 15: Hsp33 Simulation at 300K

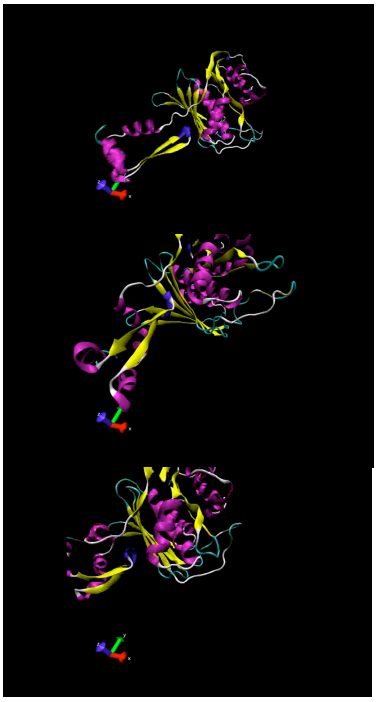


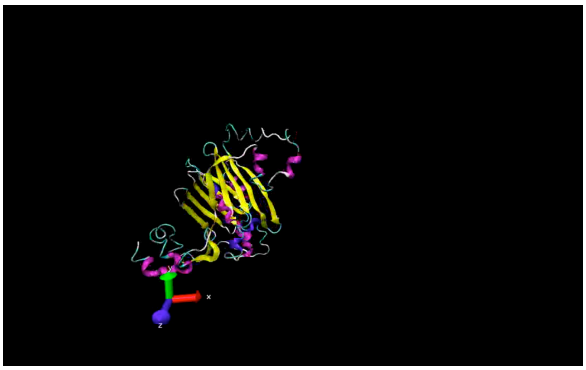
Figure 16: Hsp33 Simulation at 400K

Under 300 K, the simulation showed that Hsp33 fluctuated; however, the wing of the pocket without the zinc and cysteines did not unfold (as shown through the consistent structure in Figure 15). After the temperature was increased to 400K, Hsp33 fluctuated even more. The pocket did not completely unfold, but it did stretch out more than it did in the initial simulation (Figure 16). Although the chaperones did not behave exactly as predicted in the simulations, we are able to make some observations about how the molecules reacted under stressful conditions. From 300K to 400K, we noticed that Hsp31 fluctuated more vigorously in response to the increase in stress. The reason that the wing did not unfold could be because of alanine's hydrophobic nature. ince alanine is hydrophobic, it naturally is attracted to other hydrophobic regions. During the 10 nanoseconds of the simulation, the wing fluctuated more compared to the rest of the protein structure, even if it did not completely unfold. This demonstrates that the wing structure is innately loose. From

300 K to 400 K, the increased fluctuation suggests that an further increased temperature would cause Hsp33 to fluctuate even more, possibly even to the point of the wing pocket unfolding completely.

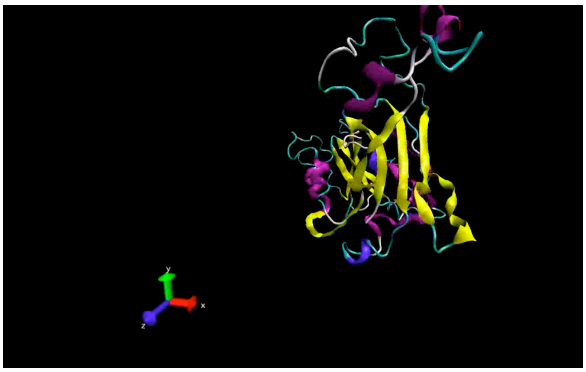
Another source of the discrepancy between our prediction and the simulation is the fact that the Amber script used to run the simulation was probably not as accurate as an experiment in a laboratory. The lack of real proteins is one major origin of error in research involving solely computational observations.

A



B

C



D

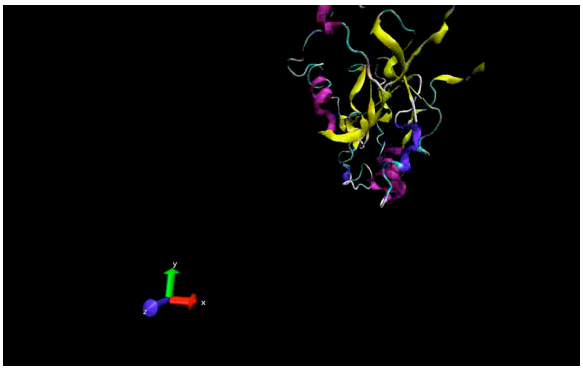


Figure 17: Hsp33 Simulation at 600K

Under 600K, the entire Hsp33 fluctuates more compared to the simulation under 400K. The overwhelming fluctuation eventually unfolds Hsp33 monomer; the short term prediction was confirmed. Even though this heat shock protein did denature under these extremely hot conditions, this does not mean it is not still resistant to heat. Any protein would denature at 600 K.

As shown from A through D in Figure 17, Hsp33 gradually unfolds within nine nanoseconds of the simulation. In Figure 17 D, Hsp33 completely loses its structure. The simulation shows that under extremely stressful environments, in this case the high temperature, Hsp33 structure will not be stable without its zinc finger; therefore, the zinc finger is the key structure that stabilizes Hsp33.

6.4 PyMOL Alignment of Hsp70 and Hsp110

The smooth cooperation between Hsp110 and Hsp70 is a result of similarity in structure. Although the two chaperones have a molecular mass difference of 40 kilodaltons, they share a 30% amino acid identity, which facilitates the smooth binding bridge between the two nucleotide binding domains [1].

After aligning the two chaperones in PyMOL, the similarities were even more

striking. The lengths of the yellow lines that appeared in the PyMOL alignment of Hsp70 and Hsp110 represented the difference between analogous structures of the two chaperones; the length of the yellow lines was inversely proportional to the similarity in structure (Figure 18). Since PyMOL showed that the NBDs around the ADPs were so similar, this confirmed that the 30% similarity in amino acid identity directly correlated to the similarity in three dimensional structure [1].

Figure 18: The green chain is Hsp110 while the light blue chain is Hsp70. The similarity between the aligned chaperones is represented by the length of the yellow lines between analogous structures of the two chaperones. The shorter yellow lines around the red ADP and blue ADP show the similarity in structure around that particular area.

7 Conclusion

Although there are some heat shock proteins that work together and have similarities in structure, such as Hsp70 and Hsp110, most Hsp's do not share this similarity in structure or function (Figures 18). Heat shock proteins are induced by heat and perform a variety of functions, like folding, protecting, and transporting proteins. Heat shock proteins have different structures to facilitate these processes, including zinc fingers, D loops, N-domain structures, and nucleotide binding domains [5] (Figures 2, 5, 8, 9). Since they have such differing structures, they rarely align as illustrated by Hsp110 and Hsp90 below (Figure 19).

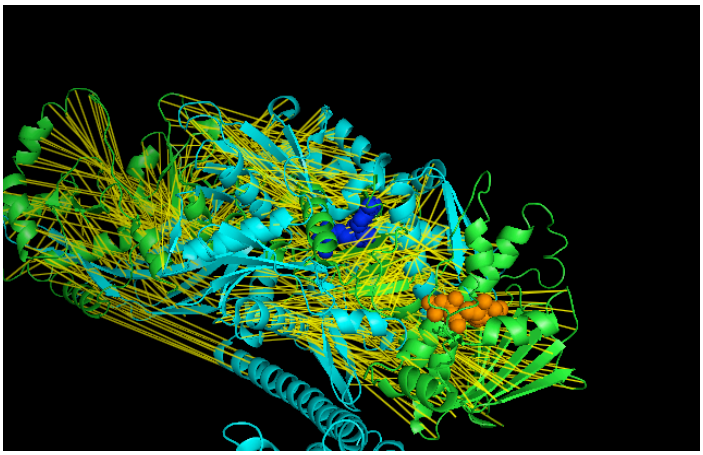


Figure 19: The green molecule is Hsp110. The light blue molecule is Hsp90. The large mess of long yellow lines represents the difference in structure between the two chaperones. This difference is further emphasized by the malignment of ATP molecules (dark blue and orange) of the two proteins.

Though our computational data in PyMOL and VMD simulations do not fully support our hypothesis, it must be emphasized that virtual means of research and experimentation are neither the most efficient nor the most accurate. PyMOL and VMD scripts can take hours longer to write and run compared to other lab procedures. In addition to script-writing time, VMD simulations can take days to yield a few nanoseconds of data. This can limit the scope of results, which may explain our inconclusive data with the Hsp33 simulation.

The results recorded by computational tools may not even be correct. PyMOL and

VMD only project tentative predictions of what would happen in a wet laboratory in which the virtual processes would actually be tested. The computer programs do not consider the subtle reactions between the elements in particular structures, which makes wet labs the best tool for decisive research; there may even be novel interactions that have not yet been discovered, so they cannot be programmed into these computational tools. Computational tools like PyMOL, VMD, and Jalview should be used sparingly in research to provide some semblance of data before devoting time and expensive laboratory resources to actual experimentation.

8 Acknowledgements

We would like to thank our Mentor John Khan, Program Coordinator Jean Patrick Antoine, Group Counselor Stoyan Lazarov, Head Counselor Daniel Cobar, and Counselor Namrata Kulkarni for providing their guidance in our project.

We would also like to thank Rutgers University, the Rutgers University School of Engineering, Morgan Stanley, the State of New Jersey, Lockheed Martin, Automated Control Concepts Inc., Silver Line Building Products, Sharon Ma and Nan Yao (parents of Johnathan Yao, GSET 2010), Tomasetta family and Laura Overdeck for their support of the Governor's School of Engineering and Technology.

9 References

[1] "Anti-Hsp110." *Antibodyworld*. Stressmarq. Web. 17 July 2011. <http://www.biomol.de/antibody_Anti-Hsp110.html?aRelated=SPC-195C&Suchbegriff=HSP&suchWeise=antibodies&hersteller=Epitomics&gruppenbezeichnung=1&1=3275-1&show_all=yes&sort_direction=ASC&sort_order=bezeichnung1>.

[2] Ali, Aruf M. U. "Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex." *Nature* (2006): 1013-1017. Web. 16 Jul 2011. <http://www.nature.com/nature/journal/v440/n7087/full/nature04716.html>.

[3] "The Chaperoning Activity of Hsp110." *The Journal of Biological Chemistry*. The Department of Molecular and Cellular Biophysics, Roswell Park Cancer Instituted, 28 May 1999. Web. 18 July 2011. <http://www.jbc.org/content/274/22/15712.full.pdf+html>.

[4] Campbell, Neil, and Jane Reece. *AP Edition Biology* . 7. San Francisco: Pearson , 2005. Print.

[5] Csermely, Peter. "Heat Shock Proteins." 17 Jul 2002: 67-75. Print.

[6] Cyr, Douglas. "Swapping Nucleotides, Tuning Hsp70." 133.6 13 Jun 2008. 945-947. *Science Direct*. Web. 16 Jul 2011. <http://www.sciencedirect.com/science/article/pii/S0092867408006983>.

[7] Goodsell, David. "Hsp90." *Molecule of the Month* Dec 2008. n. pag. *Protein Data Bank*. Web. 16 Jul 2011. <http://www.pdb.org/pdb/101/motm.do?momID=108>.

[8] "Heat Shock Proteins- Structure and Overview." *Stedwards*. N.p., n.d. Web. 16 Jul 2011. <http://www.cs.stedwards.edu/chem/Chemistry/CHEM43/CHEM43/HSP/STRUCTURE.HTML>.

[9] Izabela, Janda. United States. *Crystal Structure of the Reduced, Zn2+-Bound Form of the B. Subtilis Hsp33 Chaperone and Its Implications for the Activation Mechanism*. Charlottesville: Department of Molecular Physiology and Biological Physics, 2004. Web. 16 Jul 2011. <http://www.ncbi.nlm.nih.gov/pubmed/15458638>.

[10] Oh, Hyun Ju. "The Chaperoning Activity of Hsp110." *American Society for Biochemistry and Molecular Biology* 274.22 (1999): 15712-15718. Web. 16 Jul 2011. <http://www.jbc.org/content/274/22/15712.full.pdf+html>.
[11] Polier, Sigrun. "Structural Basis for the Cooperation of Hsp70 and Hsp110 Chaperones in Protein Folding." 133.6 13 Jun 2008. 1068-1079. *Science Direct*. Web. 16 Jul 2011. <http://www.sciencedirect.com/science/article/pii/S0092867408006788>.
[12] Quigley, Paulene. "A New Native EcHsp31 Structure Suggests a Key Role of Structural Flexibility for Chaperone Function." *Protein Society* 13.1 1 Jan 2009. 269-277. *Wiley Online Library*. Web. 16 Jul 2011. <http://onlinelibrary.wiley.com/doi/10.1110/ps.03399604/abstract>.
[13] Schuermann, Jonathan. "Structure of the Hsp110: Hsc70 Nucleotide Exchange Machine." 31.2 25 Jul 2008. 232-243. *Science Direct*. Web. 16 Jul 2011. <http://www.sciencedirect.com/science/article/pii/S1097276508003390>.
[14] Wandinger, Sebastian. "The Hsp90 Chaperone Machinery." *American Society for Biochemistry and Molecular Biology, Inc.* 283.27 (2008): 18473-18477. Web. 16 Jul 2011. <http://www.jbc.org/content/283/27/18473.full.pdf+html>.

10 Appendix

10.1 PyMol Codes

Showing Residue Charges:

```
mstop
dss
hide all
#zoom all
#orient
show cartoon,all
color gray,all

select pos,(resn arg+lys+his)
show sticks, (pos and !name c+n+o)
color marine,pos
disable pos
select neg,(resn glu+asp)
show sticks, (neg and !name c+n+o)
color red,neg
disable neg
set cartoon_smooth_loops,0
```

http://www.pymolwiki.org/index.php/Show_charged

Showing Hydrophilic Residues:

```
mstop
dss
hide all
#zoom all
#orient
show cartoon,all
color gray,all

select hydrophilics,(resn arg+lys+his+glu+asp+asn+gln+thr+ser+cys)
show sticks, (hydrophilics and !name c+n+o)
color green,hydrophilics
disable hydrophilics
set cartoon_smooth_loops,0
```

http://www.pymolwiki.org/index.php/Show_hydrophilic

Showing Hydrophobic Residues:

```
mstop
dss
hide all
#zoom all
#orient
show cartoon,all
color gray,all

select hydrophobes,(resn ala+gly+val+ile+leu+phe+met)
show sticks, (hydrophobes and (!name c+n+o))
color orange,hydrophobes
disable hydrophobes
set cartoon_smooth_loops,0
```

http://www.pymolwiki.org/index.php/Show_hydrophobics

10.2 Amber Codes

10.2.1

```
#!/bin/bash
file_pdb_name=chainH_1pv2
rad=8

cat <<eof> file_inputs

source leaprc.ff99SB

protein = loadpdb $file_pdb_name.ini.pdb

check protein
saveamberparm protein
$file_pdb_name.sw.prmtop temp

addions arah1 Cl- 0.0
addions arah1 Na+ 0.0
solvateoct arah1 TIP3PBOX $rad
saveamberparm protein
```

```
$file_pdb_name.prmtop
$file_pdb_name.inpcrd
savepdb protein $file_pdb_name.pdb
quit
eof
tleap -s -f file_inputs
rm file_inputs temp
```

10.2.2

```
#!/bin/bash

 &cntrl
  imin=1,
  maxcyc=1000,
  ncyc=500,
  ntb=1,
  ntr=1,
  cut=10.0
 /

500.0
RES 1 281
END
END


 &cntrl
  imin=1,
  maxcyc=2500,
  ncyc=1000,
  ntb=1,
  ntr=0,
  cut=10.0
 /


 &cntrl
  imin=0,
  irest= 0,
  ntx=1,
  ntb=1,
  cut=10.0,
  ntr=1,
  ntc=2,
  ntf=2,
  tempi=0.0,
  temp0=300.0,
  ig=-1,
  ntt=3,
  gamma_ln=1.0,
  nstlim=10000,
  dt=0.002,
  ntpr=100,
  ntwx=100,
  ntwr=1000
 /

10.0
RES 1 281
END
END

&cntrl
  imin=0,
  irest=1,
  ntx=7,
  ntb=2,
  pres0=1.0,
  ntp=1,
  taup=2.0,
  cut=10.0,
  ntr=0,
  ntc=2,
  ntf=2,
  tempi=300.0,
  temp0=300.0,
  ntt=3,
  gamma_ln=1.0,
  nstlim=50000,
  dt=0.002,
  ntpr=100,
  ntwx=100,
  ntwr=1000
 /
```

10.2.3

```
#!/bin/bash

mpirun -np 8
```

```
$AMBERHOME/exe/sander.MPI -O -i
min1.in -o min1.out -p prmtop -c inpcrd -r
min1.rst -ref inpcrd < /dev/null

mpirun -np 8
$AMBERHOME/exe/sander.MPI -O -i
min2.in -o min2.out -p prmtop -c min1.rst -r
min2.rst < /dev/null

$AMBERHOME/exe/pmemd.cuda -O -i
md1.in -o md1.out -p prmtop -c min2.rst -r
md1.rst -x md1.mdcrd -ref min2.rst

$AMBERHOME/exe/pmemd.cuda -O -i
md2.in -o md2.out -p prmtop -c md1.rst -r
md2.rst -x md2.mdcrd

$AMBERHOME/exe/pmemd.cuda -O -i
md1CUDA.in -o md1CUDA.out -p prmtop -
c md2.rst -r md1CUDA.rst -x
md1CUDA.mdcrd
```

10.3 PDB Codes

Hsp31 — 1pv2
Hsp33 — 1vzy
Hsp70 and Hsp110 — 3d2e, 3c7n
Hsp90 — 2cg9

10.4 UniProt Accession Numbers

Hsp31—P31658
Hsp33—P37565
Hsp70—P32589
Hsp90—P02829