Computational Prediction of the Aggregated Structure of Denatured Lysozyme

Pongsathorn Chotikasemsri
Western Kentucky University, pongsathorn.chotikasemsri@wku.edu

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Amino Acids, Peptides, and Proteins Commons, Cellular and Molecular Physiology Commons, Molecular Biology Commons, and the Molecular, Genetic, and Biochemical Nutrition Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/120

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
COMPUTATIONAL PREDICTION OF THE AGREGATED STRUCTURE OF DENATURED LYSOZYME

A Thesis

Presented to

The Faculty of the Department of Biology

Western Kentucky University

Bowling Green, Kentucky

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Pongsathorn Chotikasemsri

December 2009
COMPUTATIONAL PREDICTION OF THE AGREGATED STRUCTURE OF DENATURED LYSOZYME

Date Recommended 9/23/2009

Claire A. Rinehart
Director of Thesis

Sigrid Jacobshagen

Di Wu

Dean, Graduate Studies and Research Date
ACKNOWLEDGEMENTS

I am so thankful to my wonderful thesis committee: Dr. Claire Rinehart who always encouraged me and suggested many ways to solve the problems; Dr. Sigrid Jacobshagen who gave me advice about biochemical techniques; Dr. Di Wu who challenged me with many ideas to make this project more acceptable; Dr. James Gary who allowed me to use the Beowulf cluster to test my hypothesis for this project; Dr. John Andersland who directed me in the use of the electron microscope and suggested other techniques to help solve my problems; Daniel Fitzpatrick, Brian Kaddour, and Dexter who helped me solve many networking and grid computing problems. Thank you so much for the time that you have given to me during this project.

Most importantly, I am thankful to my parents and my wife who always encouraged me to study with happiness.
# TABLE OF CONTENTS

Chapter One: Introduction 3
   Overview 3
   Lysozyme 6
   ROSETTA++ 9
   Radius of Gyration 10
   Energy Score 11
   Root Mean Square Deviation 11
   Aggregated Protein Models 12

Chapter Two: Materials and Methods 13
   Computational Protein Prediction 13
   Lysozyme Stock Preparation 15
   Lysozyme Denaturation 15
   Transmission Electron Microscopy 16

Chapter Three: Results 17
   Computational Optimization 17
   Lysozyme *Ab Initio* Results 19
   Lysozyme Homodimer Docking Results 21
   RMSD Comparisons between predicted homodimers 25
   RMSD Comparisons between predicted homotrimers 38
   Electron Microscopy 59
   Predicted Lysozyme Aggregated Structure 62
Chapter Four: Discussions

Machine and software performance 65
Random aggregates and fibril aggregates 66
Conclusions 71

References 74
LISTS OF ILLUSTRATIONS

Figure 1.1 DnaK-DnaJ-GprE refolding mechanism  

Figure 1.2 Native structure of hen egg white lysozyme  

Figure 1.3 Fibril structure of hen egg white lysozyme  

Figure 1.4 Different stages of white hen egg lysozyme aggregates  

Figure 1.5 Radius of Gyration  

Figure 3.1 Rosetta++ performance optimization  

Figure 3.2 Energy score and radius of gyration dot graph of

Lysozyme \textit{Ab Initio} results  

Figure 3.3 Energy score and radius of gyration 3D density graph of predicted

Lysozyme homodimers  

Figure 3.4 Structure comparison of homodimer lysozyme sorted by

radius of gyration  

Figure 3.5 Structure comparison of homodimer lysozyme sorted by energy score  

Figure 3.6 RMSD can group different interfaces for structural models  

Figure 3.7 RMSD heat map of LLt515 comparisons between homodimers

sorted by radius of gyration
Figure 3.8 RMSD heat map of LLt515 comparisons between homodimers
sorted by energy score

Figure 3.9 RMSD heat map of LLt998 comparisons between homodimers
sorted by radius of gyration

Figure 3.10 RMSD heat map of LLt998 comparisons between homodimers
sorted by energy scores

Figure 3.11 RMSD heat map of comparisons between homodimers D1 - D15

Figure 3.12 RMSD heat map of LLt515 comparisons between homotrimers
sorted by radius of gyration

Figure 3.13 RMSD heat map of LLt515 comparisons between homotrimers
sorted by energy scores

Figure 3.14 RMSD heat map of LLt687 comparisons between homotrimers
sorted by radius of gyration

Figure 3.15 RMSD heat map of LLt687 comparisons between homotrimers
sorted by energy scores

Figure 3.16 RMSD heat map of LLt998 comparisons between homotrimers
sorted by radius of gyration
Figure 3.17 RMSD heat map of LLt998 comparisons between homotrimers

sorted by energy scores 44

Figure 3.18 RMSD heat map of comparisons between homotrimers T1 - T30 57

Figure 3.19 Three families of homodimers 58

Figure 3.20 Five families of homotimers 58

Figure 3.21 Aggregated lysozyme structures under TEM 60

Figure 3.22 Randomly aggregated lysozyme structures under TEM 61

Figure 3.23 Predicted lysozyme aggregated structures comparing with the random aggregated structure observed from TEM. 63

Figure 3.24 Helical aggregate prediction model 64

Figure 4.1 LLt515 and Native Lysozyme structures 69

Figure 4.2 Homodimers D4 and D7 with highlighted aggregate “cores” 70

Figure 4.3 Three subunits from the aggregate model with highlighted “cores” 71

Figure 4.4 Helical aggregate prediction model 72
LISTS OF TABLES

Table 3.1 Summary of unique homodimer structures in the lowest 1% of the energy scores or radius of gyration plots 32

Table 3.2 Summary of unique homotrimer structures in the lowest 1% of the energy scores or radius of gyration plots 46
Mis-folded proteins and their associated aggregates are a contributing factor in some human diseases. In this study we used the protein lysozyme as a model to define aggregation structures under denaturing conditions. Sasahara et al. (2007), Frare et al. (2009, 2006), and Rubin et al. (2008) observed conditions where heat denatured lysozyme formed fibril structures that were observed to be 8-17 nanometers in diameter under the electron microscope. Even though the crystal structure of lysozyme is known, the denatured form of this protein is still unknown. Therefore, we used Rosetta++ protein folding and blind docking software to create *in silico* models of the protein at denaturing temperatures and subsequently docked them into aggregates. Here we compare those structures and select forms consistent with the fibril structure from the previous papers. The next step is to be able to use the predicted models of the fibrilar forms of denatured lysozyme to help us understand the exact conformation of fibril structures. This will let us confirm the docking interactions during the fibril aggregation process. The ultimate goal is to use the validated denatured structures to model interactions with heat shock proteins during the dis-aggregation process.
CHAPTER ONE

Introduction

OVERVIEW

Alzheimer’s, transthyretin familial amyloid polyneuropathy, and senile systemic amyloidosis are diseases caused by protein aggregation (Pain, 2000). According to several studies by Mogk and Bukau (2004) and De Marco et al. (2005), protein aggregation can be induced when the native environment is changed in temperature, chemical concentration or composition, pH, or salinity. Many scientists have tried to understand the structure of aggregates in order to cure these diseases. For example, by heating lysozyme to 100 °C and then cooling the lysozyme with phosphate buffer and 1 M NaCl to 20 °C at a rate of 0.01 degree per second, lysozyme aggregates can become fibril structures (Sasahara et al., 2007). Protein modeling is another way to understand 3D protein structures so that predicted structural properties can be tested by methods such as chemical cross-linking, hybrid formation, fluorescence energy transfer, and cryo-electron microscopy (Pain, 2000). Many bench experiments have been tested on aggregated forms of lysozyme for almost a decade (Mogk et al., 2003), yet scientists still do not clearly understand the exact aggregated structures. Therefore, an alternative approach is to do computational protein structure prediction on the denatured form of lysozyme in order to gain further insights into the structure. For computational protein modeling, Rosetta++, a CASP award-winning prediction program, was run to predict protein structures from amino acid sequences. Lysozyme is a good candidate to run in Rosetta++ to predict its aggregate structures, because the availability of experimented data and the sequence and native structure are known. Also, for bench experiments on protein denaturation and aggregation, lysozyme is commercially available in abundance and cheap.
The Rosetta++ is a freely distributed suite of programs that can not only fold proteins \textit{ab initio} by minimizing the energy of the protein structure, but it can also be used to define optimal subunit docking interactions that are useful in developing models of aggregate forms. The folding environment can be defined at elevated temperatures sufficient to simulate denaturing conditions. Since seed numbers are given to Rosetta++ for each folding attempt, a diverse set of models are generated, covering the range of possible conformational structures, for denatured lysozyme. The best denatured models will be those of minimal energy and radius of gyration that dock together into aggregates that meet the denatured fibril dimensions.

In 2002, Walter and Bunchner proposed a dis-aggregating mechanism where chaperones help aggregates refold to active protein structures as shown in figure 1.1. If the Rosetta++ program can accurately predict the lysozyme aggregate structures, it will help us not only understand how aggregates form but will let us test models of how chaperones are able to dis-aggregate these proteins.

Chaperones are proteins that can assist other proteins to properly fold into their native structures. Many chaperones are heat shock proteins whose expression is increased at elevated temperatures or other cellular stresses. Although most newly synthesized proteins can properly fold without chaperones, some proteins strictly require them (Pain, 2000). In addition, chaperones can help some aggregated proteins or some misfolded aggregate proteins to disaggregate and refold into their proper structures (Pain, 2000). However, in mitochondria and endoplasmic reticulum, non-heat shock protein chaperones aid in the transport of protein across the membranes by maintaining the proteins in a denatured state (Pain, 2000).
DnaK-DnaJ-GrpE chaperones are important components in a complex that assists in the dis-aggregation mechanism. They can be found in chloroplast, mitochondria, endoplasmic reticulum, and cytoplasm. However, each organelle has unique differences in its chaperone system (Walsh et al., 2004).

According to research by Mogk and Schlieker (2003), refolding and disaggregating proteins cannot occur with only one kind of chaperone, but they need to work together as a chaperone network. Also, ClpB (Hsp100) and DnaK (Hsp70) can disaggregate and refold the aggregated luciferase proteins better than ClpB and other chaperones do alone. On the other hand, other aggregates may require additional factors in order to disaggregate a particular aggregate protein. Additionally, there are other factors which can contribute to the refolding and the disaggregating mechanisms. These are defined as “Osmolytes” which refers to the cellular concentration of viscous organic compounds surrounding the protein (Rosgen et al., 2005). Therefore, the details of these disaggregating mechanisms involve complicated intermolecular networks between chaperones and other molecules inside a cell.

A summary of the disaggregation process (figure 1.1) shows the DnaJ-DnaK and GrpE refolding model which was described by Walter and Bunchner (2002). First, DnaJ binds to non-native protein and presents it to DnaK. Then, DnaK refolds it using the energy from ATP hydrolysis. The ADP bound to DnaK is then displaced by GrpE. GrpE and the native protein are released upon the binding of ATP to the DnaK protein. A second model (not shown) uses ClpB to refold the aggregates. ClpB uses the ATP binding domain, AAA+, to trap the aggregates along with ATP. Hydrolysis of ATP
drives the refolding process and results in the release of the native form of the protein (Zolkiewski, 2006).

Figure 1.1. DnaK-DnaJ-GprE refolding mechanism (Walter and Bunchner, 2002 used with permission).

Step one: DnaJ binds to non-native protein and presents it to DnaK. Step two: DnaJ disassociates from DnaK, and DnaK refolds the non-native protein using the energy from ATP hydrolysis. Step three: the ADP bound to DnaK is displaced by GrpE. Step four: GrpE and the native protein are released upon the binding of ATP to the DnaK protein.

LYSOZYME

Lysozyme is a 147 amino acid protein with a molecular weight of around 14.4 kilodalton and is found in hen egg whites, tears, saliva, and mucus. However, based upon the Protein Data Bank (PDB), there are only 129 amino acids in native structural forms because some amino acids are cut during the protein maturation process. Lysozyme serves as a defense against bacteria by cleaving the glycosidic bond on peptidoglycans, especially in the cell wall of Gram-Positive bacteria. In 2002, Huang et al. reported that a “Trigger factor”, defined as a peptidyl-prolyl cis–trans isomerase found associated with
functional 50S ribosomal subunits, can affect the chaperone function in the folding process of lysozyme and can decelerate lysozyme activity.

The native structure of lysozyme is composed of $\alpha$-helix and $\beta$-sheet structures that are cross-linked by four disulfide bonds (figure 1.2). Lysozyme aggregates can be formed by mutation in the protein sequences, or induced by a harsh environment. For example, at a pH of 1.6-2 and a temperature of 57-70 °C, the amyloid fibril structure of lysozyme can be formed (Sasahara et al., 2007)

Figure 1.2. Native structure of hen egg white lysozyme (MMDB: 14790, PDB: 1E8L).
Conformation of single native lysozyme peptide from hen egg whites. The amino acid residues are numbered from the terminal amino group (1) to the terminal carboxyl group (129). Orange lines indicate the four disulfide bridges. Alpha-helices are depicted as green arrow tubes, Beta-sheets as brown flat arrows. The two pictures represent the lysozyme structure at different angles.

Figure 1.3. Fibril structure of hen egg white lysozyme lysozyme (Rubin et al., 2008, used with permission).
Helical-like fibril lysozyme structure as observed under Cryo-SEM. Its diameter is around 15 - 17 nm.
Many papers have reported the diameter of lysozyme aggregates. Aggregates can form both spherical or fibril structures depending on the incubation time. The formation of spherical aggregates (called random aggregates) occurs first, followed by the formation of fibril structures (called mature aggregates). The diameter of the spherical form is around 8 - 17 nm, and fibril aggregates are around 13-17 nm as shown in figure 1.3 and 1.4 (Frare et al., 2009 and 2006, Rubin et al., 2008).

Additionally, there are many other theories about lysozyme aggregates and proposed ideas explaining how they could occur. For example, Frare et al. (2006) studied the core structure of the lysozyme amyloid fibrils by a combination of limited proteolysis and Fourier Transform Infrared (FTIR) spectroscopy. They found that lysozyme fibrils contain extensive β-sheet structures, which are tolerant to pepsin at low pH, and that random or spherical (early state) aggregates contained non-β-sheet structures. In addition, they identified the sequence between amino acid 32 and 108, including the β-sheet and helix C of the native protein, to be the highly aggregation-prone region which was
resistant to protease and probably unfolded to form aggregates when it was heated. In addition, Trexler et al. (2007) confirmed that the residues from 32 to 108 formed the core of lysozyme fibrils.

Xu et al., (2007) also confirmed that the partial unfolding of lysozyme initiated aggregate formation through two-states. Transition from a native to an intermediate form required more energy from cooperativity with other proteins than when changing from an intermediate to an aggregate form. During the transition from native lysozyme to an intermediate state, the presence of helical structures decreased from 32 percent to 6 percent (Xu et al., 2005).

In other evidence from Frare et al. (2009), they denatured human lysozyme with heat at 60 °C, pH 3.0. Denatured samples were combined with the fluorescent probes thioflavin T and 1-anilino-naphthalene-8-sulfonate and examined under TEM and by FTIR spectroscopy. They found that the denatured lysozyme formed as spherical particles and interacted through exposed hydrophobic patches that contained significant elements of the β-sheet structure that is characteristic of the mature fibrils. They also stated that the spherical form is the initial phases of aggregation.

ROSETTA++

The Rosetta++ program is freely distributed software that predicts possible protein conformations by calculating the folds based on the favorable energy levels of the protein sequences for each structural segment. Moreover, Rosetta++ has received awards from Critical Assessment of Techniques for Protein Structure Prediction (CASP) competition for four consecutive years: CASP3, CASP4, CASP5, and CASP6. This program was used to explore the possible conformations of lysozyme, which are not
easily found in traditional approaches, such as NMR and X-ray crystallography. NMR and X-ray crystallography are the techniques used to study one stage of protein conformation under a single condition, despite the fact that a protein can fold independently into many different forms inside the cell. Moreover, each conformation may have distinctive functions which may barely be detected or observed by in NMR or crystallographic substrates. Thus, the limitations of these experimental approaches can be overcome by creating dynamic computer models of these proteins from known or similar crystal structures. From these models, we can predict denatured and aggregated forms of proteins that will help identify the interactions occurred during aggregation and subsequently with chaperone proteins during disaggregation. These structural predictions can then be tested by specific biophysical methods.

In addition to the protein folding software, the Rosetta++ Docking program is used to model protein binding mechanisms which can allow us to identify and predict protein-protein interactions leading to aggregation and protein-chaperone interactions. Such models will allow us to narrow down the possible mechanisms of aggregation and dis-aggregation and will provide a basis for formulating bench experiments.

RADIUS OF GYRATION

Radius of Gyration is used to calculate the size of molecule by summing the distance of each backbone carbon atom to the center of mass shown in figure 1.5. Here is the formula for the radius of gyration, $R_g$:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} r_i^2},$$

where $r_i$ is the distance of each carbon atom, i, measured from the center of mass and n is the total number of carbon atoms.
ENERGY SCORE

The three-dimensional structure of a protein is the one in which the Gibbs free energy for the whole system in which it resides, is lowest, in other words it has the most negative ΔG value. The energy score is therefore a value of ΔG that is used to determine the likelihood of formation and stability for the protein structure. Rosetta++, will calculate an energy score of each of the predicted structures.

ROOT MEAN SQUARE DEVIATION

For this experiment, Root Mean Square Deviation or Root Mean Square Distance (RMSD) is a mathematical formula used to calculate the differences between any two molecules in 3D space by measuring the average distance between the backbones of superimposed proteins. Therefore, this can be used to determine how structurally similar two molecules are to each other. Here is the formula of RMSD:

\[ RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2} \]

, where \( \delta_i \) is the distance on X, Y, Z axis between N pairs of equivalent atoms.

In this study, the RMSD was used to determine the relative orientation of homodimer or homotrimer that were aligned on a plane that defined the interface between
subunits. When pairs of dimers were compared to each other, low RMSD numbers indicate that the two dimer models interact at similar subunit interfaces. High RMSD numbers indicate that two dimer models interact at very different subunit interfaces. The same approach was used to distinguish trimer models with different interacting subunit interfaces.

AGGREGATE PROTEIN MODELS

The purpose of this study was to develop a model of the aggregate form of lysozyme that fit the observations reported in the literature. This was done by folding lysozyme structures, *Ab Initio*, under denaturing conditions. The resulting models were then docked with each other to form homodimers and homotrimers. These models were then compared to each other in order to find families of multimers with similar structures. The lowest energy representatives of these families were then examined to see if representatives from different families could, together, form higher order aggregates. Several homodimer and homotrimer families were identified and combinations of these families were put together to form higher order structures that matched the dimensions of observed aggregates.
CHAPTER TWO
Materials and Methods

COMPUTATIONAL PROTEIN PREDICTION

The following machines were used to run Rosetta++ 2.2.0

- 24 node Beowulf cluster computer machine (2 Dual core Xeon 2.0 Ghz with 2Gb RAM and 2 Dual core Xeon 2.4 Ghz with RAM 1Gb on Fedora core 5 and gcc 3.4)
- 7 node cluster computer machine (2 G5 with 2Gb RAM and on Mac OS 10.5)
- 6 Desktop computer (Pentium 4 with Hyper threading, 1 Gb RAM on Windows OS)

Rosetta++ software version 2.2.0 was downloaded from:

http://www.rosettacommons.org/software/

The rosetta Ab Initio program was used to generate the predicted structures of denatured lysozyme and the docking program was used to generate the possible dimers and trimers. Before submitting the amino acid file to the Ab Initio program, a fragment library was generated by first submitting the sequence to the online Robetta server (http://robetta.bakerlab.org/fragmentsubmit.jsp). The Robetta server accepted the FASTA format for the lysozyme sequence and generated the fragment library. The library files were then run locally in the Ab Initio program.

A generic script to run in Ab Initio mode is

./rosetta.gcc --series xx --protein t000 --chain _ -nstruct 4000
The temperature parameter was set to the default. So, no more option commands or flags were needed.

From the above script, the flag “series” was used to specify the result name with any of two alphanumeric characters. The flag “protein” was used to specify the dot DAT format file name from the protein fragment library. In this case, the protein name was set to t000 because the file name of the protein is t000.dat. The flag “chain” was used to specify the chain name of protein, which also needs to be matched with the file name, if there is more than one chain to run. In this case, the chain was set to _ because the file name of the protein is t000.dat. The flag “nstruct” was used to specify the total number of pdb formatted results. In this case, it was asked to generate 4,000 results. When the program finished running, it will place all 4,000 ab initio results into pdb format files. So, the result file names will be XXt0000001.pdb to XXt0004000.pdb.

According to the Rosetta++ manual, all unknown protein structures have to be first run in Ab initio mode to get all possible folds before running in Docking mode. Therefore, the lysozyme sequence number NP_990612.1 was taken from the NCBI database and was run in Ab initio mode for one thousand folds. The Docking mode was used to run each fold to generate all possible homodimers and selected homotrimers.

A generic script was used to run the Docking mode:

```
./rosetta.gcc a1 t000 _ -dock -symmetry -n_monomers 2 -nstruct 9999 -s LLt0000001
```

From the above script, a1 t000 _ is a short form of the “series”, “protein”, and “chain” flags. The flag “dock-symmetry” was used to activate the symmetrical docking
function of this program. The flag “n_monomers” was used to specify the total number of monomers to dock together. In this case, two monomer subunits will dock together. The flag “nstruct” was used to specify the total number of the pdb formatted results. In this case, it was asked to generate 9,999 results. The flag “s” was used to specify the file name of \textit{ab initio} pdb file to run. So, in this case the result file names will be a1t0000001.pdb to a1t0009999.pdb.

To compare the similarity between two protein structures, the pdb results were aligned and visualized in MacPymol version 1.0 by this command: Align ‘first structure name’, ‘second structure name’. For example, Align a1t0002546.pdb a7t0005896.pdb.

Once the aggregate predictions were finished, structure similarity was calculated as RMSD values and summary graphs were visualized with Mathematica 6.

**LYSOZYME STOCK PREPARATION**

Hen egg-white lysozyme, that had been lypholized and crystallized three times, was purchased from Sigma and used without further purification. A 6 mg/ml lysozyme stock was prepared with nano pure water and 1M NaCl with phosphate buffer. HCl and NaOH were used to adjust the pH of the solutions to 6 and 12.2. The solutions were then agitated at 310 rpm for 24 hours at a room temperature (adapted from Sasahara et al., 2007. and Kumar et al., 2008).

**LYSOZYME DENATURATION**

The lysozyme stock was loaded into 200 µl PCR tubes. One hundred times diluted samples were heated in a thermocycler from 20 to 77 °C and from 20 to 100 °C at
the rate of 0.01 °C per second. The samples were then cooled down at the same rate (adapted from Sasahara et al., 2007).

TRANSMISSION ELECTRON MICROSCOPY

The 6 mg/ml denatured lysozyme solution was diluted 3 and 10-fold with nano pure water, and 5 µl aliquots of the diluted samples were dropped onto carbon-coated collodion copper grid and incubated for 60 seconds. The excess liquid was removed by blotting with filter-paper and the grid was negatively stained with 2 percent (w/v) uranyl acetate for 60 seconds. The liquid on the grid was blotted against filter-paper and the grid was air-dried. All images were taken using a JEOL JEM 100CX transmission microscope with an acceleration voltage of 100 kV. The magnification was 66,000 X and 100,000X (adapted from Sasahara et al., 2007)
CHAPTER THREE

Results

COMPUTATIONAL OPTIMIZATION

To optimize the processing of data in this project, it was important to find the best way to set up the programs and commands for each machine so that they would run at maximum performance. If the performance was not maximized, the volume of results for this project would not have been finished in time. Thus, testing was done with different compile optimizations and varying numbers of Rosetta++ programs running simultaneously with a small IbpB test molecule. Two classes of computers were used in this project, those with 32-bit processors and those with 64-bit processors. Most machines had dual processing cores and many had multiple chips.

One Rosetta++ option was to compile the program to take advantage of a single processor or multiple processors. The Rosetta++ program was therefore compiled for a single processor (J1 in Figure 3.1) or four processors (J4 in Figure 3.1). When the J1 and J4 programs were run with a single job, neither took advantage of multiple processors nor multiple cores. Therefore, one to four programs (TW1 - 4 in Figure 3.1) were run simultaneously in order to optimize use of all of the processors. Figure 3.1 shows the performance for compilation options J1 and J4 combined with one to four simultaneous programs running from a single copy of the Rosetta++ program. Four copies of the Rosetta++ program were stored separately on the hard drive and single instances of programs were simultaneously run from each to see if there was an advantage to having separate copies of the program (J1 Fs4 Tws4 in Figure 3.1).
Figure 3.1. Rosetta++ performance optimization. Summary and speed comparison of IbpB prediction for 4,000 folds for different numbers of compilations and terminal windows on 32 bit and 64 bit processors. The total computing time divided by the number of processes (TW) gave the processing time / TW as shown on the Y-axis. The labels on the x-axis encode the compilation options and number of processes running simultaneously. The number following the letter “J” indicates the number of processors for which compilation was optimized, the numbers following the letters “TW” indicates the number of Terminal windows or processes that were simultaneously running and the letter “Fs” indicates the number of Rosetta++ programs that were installed on the hard drive. Each time bar is the average of three replications where the variation between replications was not greater than 100 seconds.

Figure 3.1 shows speed comparison results from eight machines. The graph shows time in hours on the Y axis and different conditions for the executed jobs on X axis. The fastest condition took 25.7 hours on a 64 bit machine which was compiled for one processor and run with 4 terminal windows from 4 separate copies of the Rosetta++ program. The fastest 32 bit machine had Rosetta++ compiled for 4 processors and was run with 4 terminal windows simultaneously. It took 73.1 hours to finish 4,000 folds. The
slowest conditions on both 32 bits and 64 bits machines were the J1 TW1 and J4 TW1 conditions, which used a single processor, and took 148.5 hours and 102.2 hours respectively.

LYSOZYME \textit{AB INITIO} RESULTS

The lysozyme fragment file, containing 147 amino acids, was run in Rosetta++ \textit{Ab Initio} mode for 1,000 folds at a denaturing temperature in order to predict representative number of possible denatured lysozyme folds. The reason why this molecule needed to be run in \textit{Ab Initio} mode was that a denatured lysozyme model for the fibril structure described by Sasahara et al. (2007) was not available. Before submitting the amino acid file to the \textit{Ab Initio} mode, a fragment library was generated by first submitting the sequence to the online Robetta server (http://robetta.bakerlab.org/fragmentsubmit.jsp). This Robetta site was very easy to use because it has a graphical interface; therefore, there was no need for a command line to run the molecule. The fragment lysozyme library, named LLt000_03_05.200_v1_3 and LLt000_09_05.200_v1_3, was submitted to the Rosetta++ program as a rosetta database by specifying the path name to the fragment lysozyme library along with the following commands to run lysozyme in \textit{Ab Initio} mode:

```
./rosetta.gcc64 -series LL -protein t000 -chain _ - constant_seed -nstruct 1000
```

See the materials and methods for an explanation of the meaning of each command.

The \textit{Ab Initio} fold produced a characteristic fold energy and radius of gyration for each of the 1000 models. These results are shown as a plot of energy score vs radius of gyration in Figure 3.2. It shows that the radius of gyration for the models is distributed
along a range from 14 to 18.5 Angstrom (Å) and the energy scores range from −90 to almost 30 kJ/mol. Fold models with the lowest energy scores are usually the most stable and are most likely to be found as actual structures. But since the object of this project was to determine the most probable denatured aggregates, exclusion of high energy forms cannot be done on monomers but must be done at the aggregate state since their conformations, energy scores, and sizes may change. Therefore, each of the 1000 folded models was docked to themselves to form higher aggregate forms.

![Ab initio results of Lysozyme 1,000 folds](image)

Figure 3.2. Energy score and radius of gyration dot graph of denatured lysozyme *Ab Initio* results.

Energy score and radius of gyration values from 1,000 *Ab Initio* folds of lysozyme were used to plot the distribution.
LYSOZYME HOMODIMER DOCKING RESULTS

Models for the aggregates of denatured lysozyme were created by using the Rosetta++ docking mode to create 10,000 homodimer models for each of the 1000 Ab Initio folded denatured lysozymes labeled LLt####. The following commands were used to run each Lysozyme Ab Initio fold in docking mode as homodimers to generate 9,999 docked models:

```
./rosetta.gcc64 a1 t000 -dock -symmetry -n_monomers 2 -nstruct 9999 -s
LLt0000001
...
```

These computations, 9,565,206 homodimer structure models were recovered because 433,794 result files were corrupted while the programs were running. All docked structure energy scores and radius of gyrations were plotted as the 3D density graph shown in Figure 3.3. According to Frare et al. (2009), the size of the lysozyme fibril aggregates should not be larger than approximately 17 nm. Although the radius of gyration for all these molecules fell into this range, I could not run all ten million homodimer molecules in the Docking mode to create higher order structures because of computing power and time constraints. Therefore, I chose to focus on the models with a radius of gyration between 18 and 20 angstroms (see Figure 3.3). The lowest six energy score homodimers and the lowest six radius of gyration homodimer structures were
further examined in detail with MacPyMOL. Surprisingly, all twelve of these homodimer molecules were found to be derivatives of the LLt515, LLt687, and LLt998 *Ab Initio* folds.

Figure 3.3. Energy score and radius of gyration 3D density graph of predicted Lysozyme homodimers. The 3D density graph represents the number of homodimers at the intersection of each Energy score and Radius of gyration value and the orange box represents the area of graph limited to 18 and 20 Angstrom (Å). The density color scale is descending from white, brown, green and blue color (from 6000 to 0 representatives).

MacPymol showed the comparative details for model 3D structures. Unlike the six lowest energy score homodimer structures (Figure 3.5b), the six lowest radius of
gyration homodimer (Figure 3.4b) are similar in structure. Therefore, the radius of
gyration was used to group or cluster the homodimer results together.

Figure 3.4. Structural comparison of lysozyme homodimers.
(a) 3D structure of the lowest radius of gyration of lysozyme homodimer model;
(b) Superimposed 3D structures of the six lowest radius of gyration lysozyme
homodimer models.

Figure 3.4 clearly shows that the six lowest radius of gyration models have almost
the same in structural conformation and their RMSD (Root Mean Square Distance) is less
than 0.01 Å. The average diameter of these homodimers is around 56 to 57 Å.

Surprisingly, all six lowest radius of gyration homodimers structures came from the
single lysozyme Ab initio fold, LLt515. In contrast, the six lowest energy score
homodimers structures came from two Ab initio folds; five structures came from LLt687
and the other from LLt998. When the six lowest radius of gyration homodimers and the
native lysozyme structure were compared, the RMSD value was 11.917 Å. But, the six
lowest energy score homodimer structures compared to native lysozyme gave RMSD
values ranging from 12.757 Å to 16.261 Å. This showed that the six lowest energy score
homodimer structures are not similar in structural conformation, and that their RMSD
diffs between these six molecules by up to 0.20 Angstrom. The average total size of
these homodimers was around 72 to 77 Å. These RMSD comparisons confirmed that the six lowest radius of gyration homodimers have more structural similarity than the six lowest energy score homodimers.

Figure 3.5. Structure comparison of lysozyme homodimers with the lowest energy score.
(a) 3D Structure of the lowest homodimer structure; (b) Superimposed 3D structures of six lowest score lysozyme homodimer structures.
RMSD COMPARISIONS BETWEEN PREDICTED HOMODIMERS

The docking of the 1000 predicted lysozyme structures into 10,000 dimers each created almost 10 million predicted structures. Higher order structural associations, defined by docking dimers, were not possible because of computer memory limitations. However, because many of the lowest energy and radius of gyration models were derived from just a few folds, it was possible to explore the lower energy domains of the dimer distribution to define dimer pairs that could further associate into dimer/dimer structures. This was done by comparing 100 dimer structures from the lowest energy or radius of gyration models to each other by RMSD (Root Mean Square Distance). As mentioned in the introduction, RMSD is a mathematical formula used to calculate the differences in structure between any two molecules in 3D space. Low RMSD numbers identified dimers that had similar structures and docking faces while high RMSD usually identified complementary interfaces that could possibly dock together to form higher order dimer/dimer structures in proteins with similar folds (see Figure 3.6). Therefore, RMSD was a good approach to group similar structural homodimer proteins together by their docking interfaces.

Models with the lowest energy scores are usually the most probable and stable model but may not be an absolute indicator of the best model. Since the literature provided size estimates of denatured lysozyme the lowest radius of gyration was also used as a selective criteria. Since the six of the smallest radius of gyration homodimers came from the single lysozyme Ab initio fold, LLt515 and the six lowest energy score homodimers structures came from two Ab initio folds LLt687 and LLt998.
Figure 3.6 RMSD can group different interfaces for structural models. RMSD was used to group proteins which have the same interface and structural similarity. (A) On left, two molecules that have low RMSD and almost completely overlap when superimposed them on each other (right). (B) On left, two molecules that have high RMSD and do not completely overlap when superimposed them on each other (right).

It was assumed that further examination of the lowest 100 (1%) of the homodimer structures from these *Ab initio* folds would give a good representation of the real structures expected under denaturing conditions. The structure of each homodimer was compared to every other homodimer within each *Ab initio* fold and the RMSD results are represented in the heat diagrams found in Figure 3.7 through Figure 3.10. However, most
of LLt687 homodimer files were corrupted; therefore, I could not use their files to run a RSMD comparison.

Figure 3.7 RMSD heat map of LLt515 comparisons between homodimers sorted by radius of gyration (lowest at 1). Numbers d1 through d3 represent the location of homodimer structures with the lowest radius of gyration representing the three conserved structural families within the heat map.
Figure 3.8 RMSD heat map of LLt515 comparisons between homodimers sorted by energy score (lowest at 1). Numbers d4 through d7 represent the location of homodimer structures with the lowest energy scores representing the four conserved structural families within the heat map.
Figure 3.9 RMSD heat map of LLt998 comparisons between homodimers sorted by radius of gyration (lowest at 1). Numbers d8 through d12 represent the location of homodimer structures with the lowest radius of gyration representing the five conserved structural families within the heat map.
Figure 3.10 RMSD heat map of LLt998 comparisons between homodimers sorted by energy scores (lowest at 1). Numbers d13 through d15 represent the location of homodimer structures with the lowest radius of gyration representing the three conserved structural families within the heat map.
The homodimers labeled d1 through d15 represent the lowest energy or radius of gyration representatives of homodimer structures that differ from the other homodimer structures within the same *Ab initio* fold. These representatives are the most likely candidates for forming higher order dimer-dimer structures. For comparison purposes, the structures, energy scores and radius of gyration have been summarized in Table 3.1. In addition, d1 to d15 were compared and plotted by their RMSD values as shown in Figures 3.11.
Table 3.1. Summary of unique homodimer structures in the lowest 1% of the energy scores or radius of gyration plots.
Structures are aligned along their two fold axis of interaction for comparison purposes. The (sort) designations under the Primary Fold header refers to: ES = sorted by Energy Score and RG = sorted by Radius of Gyration. The residues 1-20 (Blue), residues 21-40 (Marine Blue), residues 41-70 (Green), residues 71-100 (Yellow), residues 101-119 (Orange), residues 125-147 (Red).

<table>
<thead>
<tr>
<th>Class</th>
<th>Structure</th>
<th>Energy Score (\Delta G) (kJ/Mol)</th>
<th>Radius of Gyration (nm)</th>
<th>Primary Fold (Sort)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td><img src="image1" alt="Structure Image" /></td>
<td>-42.4</td>
<td>18.00</td>
<td>Llt515 (RG)</td>
</tr>
<tr>
<td>D2</td>
<td><img src="image2" alt="Structure Image" /></td>
<td>-41.43</td>
<td>18.01</td>
<td>Llt515 (RG)</td>
</tr>
<tr>
<td>D3</td>
<td><img src="image3" alt="Structure Image" /></td>
<td>-42.2</td>
<td>18.07</td>
<td>Llt515 (RG)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>D4</strong></td>
<td>-47.57</td>
<td>18.78</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td><strong>D5</strong></td>
<td>-46.35</td>
<td>18.42</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td><strong>D6</strong></td>
<td>-45.99</td>
<td>19.04</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td><strong>D7</strong></td>
<td>-45.90</td>
<td>20.44</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Interaction Energy</td>
<td>Relative Displacement</td>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>D8</td>
<td>-60.21</td>
<td>19.05</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>D9</td>
<td>-55.73</td>
<td>19.19</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>D10</td>
<td>-62.33</td>
<td>19.24</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>D11</td>
<td><img src="image1.png" alt="Image" /></td>
<td>-65.51</td>
<td>19.29</td>
<td>LLt998 (RG)</td>
</tr>
<tr>
<td>D12</td>
<td><img src="image2.png" alt="Image" /></td>
<td>-63.15</td>
<td>19.37</td>
<td>LLt998 (RG)</td>
</tr>
<tr>
<td>D13</td>
<td><img src="image3.png" alt="Image" /></td>
<td>-67.79</td>
<td>19.65</td>
<td>LLt998 (ES)</td>
</tr>
<tr>
<td>D14</td>
<td><img src="image4.png" alt="Image" /></td>
<td>-65.84</td>
<td>19.58</td>
<td>LLt998 (ES)</td>
</tr>
<tr>
<td>D15</td>
<td>-65.08</td>
<td>21.02</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
</tbody>
</table>

[Image of a molecular structure]
Figure 3.11 RMSD heat map of comparisons between homodimers D1 through D15 (lowest at 1). D1-D15 orderly correspond with 1-15 in the plot on X and Y axis.
RMSD COMPARISONS BETWEEN PREDICTED HOMOTRIMERS

The primary folds, LLt515, LLt687, and LLt998 were run again in Docking mode to generate homotrimer structures since it might be possible that these molecules can form as homotrimers whose radii are still under 17 nm. The following commands were used to run Rosetta for homotrimers docking:

```shell
./rosetta.gcc64 a1 t000 -dock -symmetry -n_monomers 3 -nstruct 9999 -s LLt0000515

./rosetta.gcc64 a2 t000 -dock -symmetry -n_monomers 3 -nstruct 9999 -s LLt0000687

./rosetta.gcc64 a3 t000 -dock -symmetry -n_monomers 3 -nstruct 9999 -s LLt0000998
```

See the materials and methods for an explanation of the meaning of each command.

Subsequently, the homotrimer results were sorted by energy score and radius of gyration. The lowest 100 structures from each sort were taken and each homotrimer was compared to every other homotrimer within each Ab initio fold. The RMSD results from these comparisons are represented in the heat diagrams found in Figure 3.12 through Figure 3.17.
Figure 3.12 RMSD heat map of LLt515 comparisons between homotrimers sorted by radius of gyration (lowest at 1). Numbers t1 through d4 represent the location of homotrimer structures with the lowest radius of gyration representing the four conserved structural families within the heat map.
Figure 3.13 RMSD heat map of LLt515 comparisons between homotrimers sorted by energy scores (lowest at 1). Numbers t5 through t8 represent the location of homotrimers structures with the lowest energy scores representing the four conserved structural families within the heat map.
Figure 3.14 RMSD heat map of LLt687 comparisons between homotrimers sorted by radius of gyration (lowest at 1). Numbers t9 through t15 represent the location of homotrimer structures with the lowest radius of gyration representing the seven conserved structural families within the heat map.
Figure 3.15 RMSD heat map of LLt687 comparisons between homotrimers sorted by energy scores (lowest at 1).
Numbers t16 through t19 represent the location of homotrimers structures with the lowest energy scores representing the four conserved structural families within the heat map.
Figure 3.16 RMSD heat map of LLt998 comparisons between homotrimers sorted by radius of gyration (lowest at 1).
Numbers t20 through t24 represent the location of homotrimer structures with the lowest radius of gyration representing the five conserved structural families within the heat map.
Figure 3.17 RMSD heat map of LLt998 comparisons between homotrimers sorted by energy scores (lowest at 1). Numbers t25 through t30 represent the location of homotrimers structures with the lowest energy score representing the six conserved structural families within the heat map.
The homotrimers labeled t1 through t30 represent the lowest energy or radius of gyration representatives for homotrimer structures that differ from the other homotrimer structures within the same *Ab initio* fold. These representatives are the most likely candidates for forming higher order trimer-trimer structures. For comparison purposes, the structures, energy scores and radius of gyration have been summarized in Table 3.2. In addition, t1 to t30 were compared and plotted by their RMSD values shown in Figure 3.18.
Table 3.2. Summary of unique homotrimer structures in the lowest 1% of the energy scores or radius of gyration plots.
Structures are aligned along their two fold axis of interaction for comparison purposes. The (sort) designations under the Primary Fold header refers to: ES = sorted by Energy Score and RG = sorted by Radius of Gyration. The residues 1-20 (Blue), residues 21-40 (Marine Blue), residues 41-70 (Green), residues 71-100 (Yellow), residues 101-119 (Orange), residues 125-147 (Red).

<table>
<thead>
<tr>
<th>Class</th>
<th>Structure</th>
<th>Energy Score($\Delta G$) (kJ/Mol)</th>
<th>Radius of Gyration (nm)</th>
<th>Primary Fold (sort)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td><img src="image" alt="T1 Structure" /></td>
<td>-51.08</td>
<td>21.69</td>
<td>LLt515 (RG)</td>
</tr>
<tr>
<td>T2</td>
<td><img src="image" alt="T2 Structure" /></td>
<td>-51.40</td>
<td>21.95</td>
<td>LLt515 (RG)</td>
</tr>
<tr>
<td>T3</td>
<td>-47.19</td>
<td>22.00</td>
<td>LLt515 (RG)</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>-55.91</td>
<td>22.03</td>
<td>LLt515 (RG)</td>
<td></td>
</tr>
<tr>
<td>T5</td>
<td>-62.81</td>
<td>23.46</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td>-62.05</td>
<td>22.49</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>-62.49</td>
<td>23.83</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td>T8</td>
<td>-61.88</td>
<td>23.59</td>
<td>LLt515 (ES)</td>
<td></td>
</tr>
<tr>
<td>T9</td>
<td>-81.13</td>
<td>22.66</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>-88.42</td>
<td>23.00</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>T11</td>
<td>-90.27</td>
<td>23.13</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>T12</td>
<td>-95.40</td>
<td>23.2</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T13</td>
<td>-94.94</td>
<td>23.29</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>T14</td>
<td>-95.55</td>
<td>23.31</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>T15</td>
<td>-75.75</td>
<td>23.39</td>
<td>LLt687 (RG)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>T16</td>
<td>-99.54</td>
<td>23.57</td>
<td>LLt687 (ES)</td>
<td></td>
</tr>
<tr>
<td>T17</td>
<td>-99.34</td>
<td>23.95</td>
<td>LLt687 (ES)</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>-99.10</td>
<td>23.97</td>
<td>LLt687 (ES)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>T19</td>
<td>-99.02</td>
<td>23.85</td>
<td>LLt687 (ES)</td>
<td></td>
</tr>
<tr>
<td>T20</td>
<td>-80.39</td>
<td>22.26</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>T21</td>
<td>-90.83</td>
<td>22.38</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T22</td>
<td>-91.31</td>
<td>22.46</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>T23</td>
<td>-89.77</td>
<td>22.66</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>T24</td>
<td>-90.69</td>
<td>22.71</td>
<td>LLt998 (RG)</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T25</td>
<td>-99.17</td>
<td>23.89</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>T26</td>
<td>-97.24</td>
<td>24.16</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>T27</td>
<td>-96.44</td>
<td>23.07</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>T28</td>
<td>-94.55</td>
<td>22.52</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>T29</td>
<td>-93.81</td>
<td>25.26</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
<tr>
<td>T30</td>
<td>-93.55</td>
<td>25.98</td>
<td>LLt998 (ES)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.18 RMSD heat map of comparisons between homotrimers T1 through T30 (lowest at 1). T1-T30 orderly correspond with 1-30 in the plot on X and Y axis.

According to the summary of RMSD heat maps of D1 to D15 and T1 to T30 (Figure 3.11 & 3.18) for the homodimers and homotrimers, it was concluded that there were three major structural family interfaces for homodimers and five structural family interfaces for homotrimers.
Figure 3.19 Three families of homodimers. These molecules are the representative structures from D1 to D15, shown in a superimposed format in order to illustrate their similarity. (Family A) D1, D2, & D3. (Family B) D4 and D5. (Family C) D9, D10, D12, and D14.

Figure 3.20 Five families of homotimers. These molecules are the representative structures from T1 to T30 shown in a superimposed format in order to illustrate their family similarity. (Family A) T1, T2, and T3. (Family B) T4 and T6. (Family C) T10, T11, T12 T13, T14, T16, T17, T18, and T19 (Family D) T21, T23, T24, and T28. (Family E) T22 and T27.

In summary, although Figure 3.19 and Figure 3.20 shown three families of homodimer structure and five families of homotrimer structures, there are additional unique structures that are not included in these families. For instance, as shown in Table
3.1, D6, D7, D8, D11, D13 and D15 are all unique from each of the other homodimers. Also, for the homotrimers, T5, T7, T8, T9, T15, T20, T25, T26, T29 and T30 are unique from other homotrimers. Therefore, there are a total of 9 unique homodimer structures and 15 unique homotrimer structures.

Once a defined set of unique dimer and trimer models were identified, these were docked together to get higher order aggregates and a prediction model for the denatured lysozyme aggregate.

ELECTRON MICROSCOPY

Lysozyme samples were prepared and denatured using the procedures from Sasahara et al. (2007) and Kumar et al. (2008). When the samples were observed under the Transmission Electron Microscope (TEM), they showed aggregated structures (Figure 3.21). The observed structures did not directly correspond to the extended fibril structures observed by Sasahara et al. (2007) or Rubin et al. (2008) however, they did form large but irregular aggregates that approached 17 nm in diameter for the fibers between the holes. When lysozyme was heated to 77 instead of 100 °C the aggregates formed the simpler ginger root-like-structures shown in Figure 3.22. Even though the 8-17 nm filaments described by previous papers could not be reproduced experimentally, two different aggregate forms were obtained.
Figure 3.21 Aggregated lysozyme structures under TEM

Lysozyme aggregated structures were formed by heating from 20 to 100 °C with a 0.01 degree/second temperature gradient in pH12.2 phosphate buffer. The sample was held at 77 °C for 2 minutes and cooled down at the same rate to a final temperature of 20 °C. The bar represents 30 nm.
Figure 3.22 Randomly aggregated lysozyme structures under TEM. Lysozyme randomly aggregated structures were formed by heating from 20 to 77 °C with a 0.01 degree/second temperature gradient in pH12.2 phosphate buffer. The sample was held at 77 °C for 2 minutes and cooled down at the same rate to a final temperature of 20 °C. The bar represents 10 nm.

The branch structure pictured in Figure 3.22 suggests that two or more types of docking faces were interacting to initiate the branch structure. It is possible that these structures formed from heterodimer or heterotrimer structures, which were not considered in this study because of computational constraints, but this study focused on building these higher order structures from the nine unique homodimer and fifteen unique homotrimer models. In order to check which combinations of docking faces could form higher order structures such as those pictured in Figure 3.22, single monomers were superimposed on one another in MacPymol and then combinations which overlapped their companion dimer or trimer subunits were eliminated. This allowed the selection of
dimer and trimer combinations that could share monomers and create larger aggregated structures. Using this approach, aggregated models were constructed to fit the structures observed under the electron microscope. The results are presented in the following section.

PREDICTED LYSOZYME AGGREGATED STRUCTURE

The goal of this work was to create a model whose diameter matches with the size of denatured lysozyme aggregates reported in the literature or observed locally under the TEM. Therefore, I tried to dock all combination of the unique faces from D1-D15 and T1-T30 in order to fit the size of aggregates from the evidence. Figure 3.23 shows the best model results for the TEM picture in Figure 3.22. It shows the best homotrimer structure from models T6 and T7 derived from fold LLt515, which best fits the aggregate pictured in Figure 3.22. When the energy scores for the aggregated structure were compared, T6 = -62.05 kJ/mol and T7 = -62.49 kJ/mol, these two structures should associate at about the same time and nucleation of the aggregate would allow it to immediately grow. Because these energies are very close together there is probably not a preferred hierarchal assembly scheme for these molecules. Since the size of this predicted molecule is exactly the same dimensions as the TEM picture in Figure 3.22, it strongly supports this predicted model as a candidate structure.
Figure 3.23 Predicted lysozyme aggregated structures comparing with the random aggregated structure observed from TEM.

Left, the picture from Figure 3.22. The bar represents 10 nm. (Right) the predicted lysozyme aggregated structures composed of four trimers which match in size and angle with the lysozyme random aggregated structure (Left) from TEM. The models on the right came from fold LLt515. The central model upon which the other three models were super imposed is model T6. The other three overlapping homotrimers are from model T7. The size markers in this molecule are 92.8 Å (line A), 58.8 Å (line B), 61.2 Å (line C), and 56.6Å (line D).

In addition, the best model to match the cryo-SEM picture from Rubin (2008) is shown in Figure 3.24. It is made from alternating dimer interactions defined in D4 and D6. The free energy for D4 and D6 are –47.57 and –45.99 kJ/Mol respectively. These values are close to each other but much lower than the random aggregate forms shown above.
Figure 3.24 Helical aggregate prediction model.
Helical aggregate structure derived from the interaction of two homodimer faces. 31 molecules of alternating D4 (red) and D7 (gold) interact at two different interfaces to form a helical aggregate that closely matches the structures reported by Rubin (2008). The model on the left is the topview of the figure on the right. The exterior diameter, (a), is 16.2 nm. and the interior diameter, (b), is 9.2 nm. The thickness of the thread, (c), is around 6 nm., the rise per turn, (d), is 14.5 nm.
CHAPTER FOUR

Discussion

MACHINE AND SOFTWARE PERFORMANCES

In order to perform Ab initio folding in an efficient and timely manner, it was important to optimize the Rosetta++ performance. The best way to run Rosetta++ on 32 bit and 64 bit machines was to use one set of Rosetta++ files with the compilation option set to four processors and with four terminal windows running four processes. From the computational analysis results of IbpB prediction in Ab Initio mode (Figure 3.1), it was possible to derive a formula to predict the time required for a set of calculations on 64 bit and 32 bit computers housing up to 4 computational cores. The terminal windows correspond to the number of simultaneous programs being executed.

Total time usage, in seconds, for a 64 bit operating system machine =

\[(0.0803125)(\text{Length of sequence to be folded})(\text{Number of folds to perform on each terminal window})(\text{speed of CPU in GHz})(\text{Number of Terminal Windows}) \pm 100 \text{ seconds}\]

Total time usage, in seconds, for 32 bits operating system machine =

\[(0.190364583)(\text{Length of sequence})(\text{Number of fold to perform on each terminal window})(\text{speed of CPU in GHz})(\text{Number of Terminal Windows}) \pm 100 \text{ seconds}\]

Therefore, any molecules that run on 64 bit machines should complete a job in 2.37 times faster than those running on 32 bit machines. However, these ideal performance estimates started flattening out when the Rosetta++ was run with three or more terminal windows at the same time.
When running a single terminal window to fold a modestly sized protein on the quad core computers, usually a little over 1 processor were used. However, when larger proteins, such as ClpBC (a heat shock protein), were folded, a little over 2 processors were used. The Rosetta home website (http://www.rosettacommons.org/) suggested that running Rosetta++ under the control of the condor program will help improve the use of all available processors without launching multiple terminal windows since it monitors the usage of each processor and distributes jobs to idle processors.

Although running Rosetta++ from four separate files showed the best performance on 64 bit machines, it is not really practical because the size of the Rosetta++ program itself is around 2 GB, which takes up much hard drive space. Also, the performance gain is not that much more than for a single copy of Rosetta++.

For the Docking mode in Rosetta++, the symmetrical docking option was used in this experiment and limiting the docking to “head to head” and “tail to tail” orientations. Therefore, it prevents “head to tail” forms from being considered in this experiment.

RANDOM AGGREGATES AND FIBRILL AGGREGATES

The 147 amino acids form of hen-egg white lysozyme was folded and docked in this study instead of the 129 amino acid form. Processing of the 147 amino acid lysozyme to the 129 amino acid form during entry into the endoplasmic reticulum removes the N-terminal 18 amino acid signal peptide. These 18 amino acids did not seem to interfere with the folding of “core” structures that were mentioned by Frare et al. (2006) and Trexler et al. (2007), however, they should be removed in future studies in order to represent the secreted form of the protein.
The goal of this research was to derive a structural model of denatured lysozyme aggregates that could be used to explore the dis-aggregation process facilitated by chaperone proteins. Sasahara et al. (2007), Frare et al. (2009, 2006), and Rubin et al. (2008) demonstrated that denatured lysozyme could form fibril aggregates that were about 8-17 nm in diameter. Several attempts to reproduce these fibers failed to create the same fibril aggregate structures, probably because the protein that was visualized in the TEM soon after denature did not have time to aggregate into the fibril form. Obviously there were some conditions used by Sasahara et al. (2007) that were not reproduced in our hands. For example, recently Kumar et al. (2008) described how surfactants and DTT affect the size, dynamics, activity and growth of soluble lysozyme aggregates. A closer look at the ideal conditions for fibril growth is needed beyond this study.

Almost 10 million homodimer and homotrimer structural models were constructed with Rosetta++ folding and docking programs. An examination of the lowest energy models for homodimers showed that they were derived from two folds, LLt687 and LLt998. Examination of homodimer models with the smallest radius of gyration showed that they came from a single fold, LLt515. Therefore, these folds became the focus for deriving the aggregate models. The major structural families of homodimer and homotrimer representatives that were derived from these folds are summarized in Tables 3.1 and 3.2. Family members were easily identified from heat maps created by comparing each structure to all other structures in a pairwise manner and plotting their RMSDs. The models that were sorted by radius of gyration often gave structures with energy scores significantly higher than those sorted by energy score. It was also observed that several
duplicate structures were found between models sorted by the radius of gyration and the energy score.

The magnitude of the RMSD derived by comparing two models seemed to be an effective way of distinguishing homodimer and homotrimer models that interacted at different interfaces. For example, if the RMSD between two dimer models from the same fold were greater than 200, then there was a good chance that they were bound together at different faces. By taking two models with different interaction faces and alternately superimposing them, higher order structures were assembled whose diameter matched the range of 8-17 nm as presented in Figure 3.23. Similarly, the helical model derived by assembling the homodimers D4 and D7 (Figure 3.24) matched the diameter of the fibrils reported by Ruben et al., (2008).

The aggregate models presented in this study match the physical dimensions of fibril aggregates reported in the literature and the random aggregates observed in this study under the TEM. However, comparison of LLt515, which was used to create the denatured aggregate models, to the native lysozyme structure (Figure 4.1) shows a decrease in the number of alpha helices from 53.74% for the native form to 51.70% for LLt515. In contrast, Xu et al. (2005) reported that aggregates showed a decrease in helix structures from 30% to 6% compared to the native lysozyme form. Therefore, this proposed aggregate model does not fit all of the literature and additional models should be considered.
Frare et al. (2006) and Trexler et al. (2007) reported that the aggregate “cores” included residues 32 – 108. These residues were acid resistant and most likely found at the interacting interface or buried in the denatured structure. Figure 4.2 shows the D4 and D7 homodimers with these residues highlighted in white and Figure 4.3 shows three monomers aggregated together with these same residues highlighted in the central monomer. The core structures are seen buried in the aggregates as they participate in the docking interfaces. In addition, alpha helices located in the D4 core structure may facilitate D4-D4 homodimer formation (Figure 4.2A), and β-sheets located in the D7 core structure facilitate D7-D7 homodimer formation (Figure 4.2B); thus, each contributing to the overall higher order lysozyme aggregate model shown in Figure 4.3.
Figure 4.2 Homodimers D4 and D7 with highlighted aggregate “cores”.

(A) D4 shows core structure labeled in white. (B) D7 shows cores structure labeled in white. They both show the molecule from 2 different angles with the right model rotated vertically from the left model.
Figure 4.3 Three subunits from the aggregate model with highlighted “cores”.

(Left) Two D4 molecules and one D7 molecules aggregated together with core structures labeled in white. (Right) Schematic picture shows how the D4 (Red circles) and D7 (Black circle) homodimer models fit together to form the higher order aggregate.

Additional aggregate forms, different from those suggested by this study, have also been suggested in the literature. For example, Sophianopoulos (1969) and Holloday and Sophianopoulos (1972) suggested that lysozyme may form head to tail aggregates. This study did not consider such asymmetric docking structures. Also, microenvironments, such as acidic phospholipids surfaces, may contribute to the association of lysozyme into aggregate structures (Gorbenko et al., 2007). In this acidic membrane environment, Gorbenko et al. (2007) found evidence from fluorescent probes suggesting that tetramers may be an optimal aggregate form.

CONCLUSIONS

In conclusion, the aggregate model constructed in this study matched the fibril diameter observed by a number of investigators, but did not conform to a reduction in the percent of alpha helical structures. The constructed model did support the observation
that the core set of residues from 32 through 108 were buried in the denatured aggregate as observed by Frare et al. (2006).

The homodimer and homotrimer models identified in this study do not represent all of the possible interactive forms of dimer and trimer structures since the symmetry option in the docking was invoked and only identical surface interactions were considered (for example: top-top or bottom-bottom). More computing resources will be needed for a more comprehensive model of the denatured aggregate structure of lysozyme, including head to tail docking structures. However, this study has defined the use of methods, such as RMSD, for identifying sets of unique dimer and trimer structures that can be used to build aggregate forms. Defining a reduced number of these unique sets is key in reducing the number of comparisons that needs to be made in future dimer and trimer docking studies. The refolding of dimer and trimer structures may also help define the conversion of alpha to beta structures that may be key in the transition of aggregate structures to more stable forms. Also, studies considering just the 129 amino acid form of lysozyme or even the core residues may also help to better understand the aggregate forms and their transitions better.
Figure 4.4 Helical aggregate prediction model.
Helical aggregate structure derived from the interaction of two homodimer faces. 31 molecules of alternating D4 and D7 interact at two different interfaces to form a helical aggregate that closely matches the structures reported by Rubin et al. (2008). The model on the left is the topview of the figure on the right. The exterior diameter, (a), is 16.2 nm. and the interior diameter (b) is 9.2 nm. The thickness of the thread, (c), is around 6 nm., the rise per turn, (d), is 14.5 nm.
REFERENCES


PDB (Protein Data Bank) http://www.pdb.org/.


