STRUCTURAL MODELLING AND FUNCTIONAL ANALYSIS OF THE
ENGINEERED SMALL HEAT-SHOCK PROTEIN, TPV-HSP14.3 FROM
THERMOPLASMA VOLCANIUM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ILIR SHERAJ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

MAY 2014

Approval of the thesis

STRUCTURAL MODELLING AND FUNCTIONAL ANALYSIS OF THE ENGINEERED SMALL HEAT-SHOCK PROTEIN, TPV-HSP14.3 FROM *THERMOPLASMA VOLCANIUM*

submitted by **ILIR SHERAJ** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**   _____

Prof. Dr. Orhan Adalı
Head of Department, **Biology**   _____

Prof. Dr. Semra Kocabıyık
Supervisor, **Biology Dept., METU**   _____

Examining Committee Members:

Prof. Dr. Emel Arınç
Biology Dept., METU   _____

Prof. Dr. Semra Kocabiyik
Biology Dept., METU   _____

Prof. Dr. Fatih İzgü
Biology Dept., METU   _____

Prof. Dr. Candan Gürakan
Food Engineering Dept., METU   _____

Assist. Prof. Dr. Mehmet Somel
Biology Dept., METU   _____

                                           Date: 06.05.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Ilir SHERAJ

Signature:

ABSTRACT

STRUCTURAL MODELLING AND FUNCTIONAL ANALYSIS OF THE ENGINEERED SMALL HEAT-SHOCK PROTEIN, TPV-HSP14.3 FROM *THERMOPLASMA VOLCANIUM*

Sheraj, Ilir

M.Sc., Department of Biology

Supervisor: Prof. Dr. Semra Kocabıyık

May 2014, 256 pages

In this study, a small heat shock protein tpv-Hsp14.3 from a thermoacidophilic Archaeon, *Thermoplasma volcanium* was studied. sHSPs are low molecular weight proteins involved in different stress responses to protect the cellular proteome and enhance survival of the host organism. For structure-function analysis of the protein, both experimental and computational tools were used. Sequence alignments with closely-related sHSPs showed high sequence conservation at the core α-Crystalline domain and a V/I/L-X-V/I/L motif in the middle of the C-terminus. Computational results for secondary structure prediction showed a canonical class II sHSP with nine β-strands, which constitute the core α-Crystalline domain, flanked by an N-terminal α-helix and a short coiled coil secondary structure at the C-Terminus. These results were confirmed by the predicted three-dimensional (3D) of the tpv-Hsp14.3. 3D model analysis showed that sHSP protein monomers are interacting by means of β6-strand swapping to dimerize, and then form higher order oligomers which interact and protect the protein substrates under stress conditions.

Sequence alignments at different levels and structure comparison with the sHSPs studies so far were carried out to determine important residues involved in dimer formation and chaperoning function. From these studies, three single-residue, and a double-residue were determined for mutagenesis. They were: R69, to be changed into R69K, R69E, and R69M; R81 to be changed into R81K, R81E and R81M; K87 to be changed into K87R, K87E and K87I; and QR(80-81) to be changed into QR(80-81)EL. To this end, site-directed mutagenesis was performed to change the selected residues into analogs, negatively-charged, and hydrophobic residues. Then the mutants and the wild-type proteins were expressed in *E.coli* and purified by affinity column chromatography. To see the change in the chaperoning function of the mutants, pig heart Citrate Synthase (CS) was used as model enzyme. CS was incubated with both, wild-type and each one of the sHSP mutant variants at denaturing temperature (47°C). The initial activity then was measured at 35°C which is the optimum temperature for CS. The enzyme assay results showed that the enzyme activities of mutant variants R69E, R69K, K87R and K87I, increased by 1.9-, 1.8-, 2.6- and 2.7-fold respectively, compared to CS activity incubated with the wild-type tpv-Hsp14.3. K87E and QR(80-81)EL mutants had a slight increased CS activity (1.2- and 1.3-fold respectively) as compared to the wild-type tpv-Hsp14.3. On the other hand, CS activities when incubated with R69M and R81K mutant variants were 0.9- and 0.6-fold lower, respecively than CS incubated with the wild-type chaperone. CS activity when the enzyme was incubated with R81E and R81M mutants did not significantly change as compared to CS activity when the enzyme was incubated alone at 47°C.

Finally, Dynamic Light Scattering (DLS) Spectroscopy was used to study the oligomer dynamics of wild-type and mutant variants of tpv-Hsp14.3 in solution. DLS was carried out under two different conditions: In the first, the wild-type tpv-Hsp14.3 was incubated alone and with pig heart CS or Bovine Glutamate Dehydrogenase (GDH) substrates enzymes at temperatures where the two enzymes lose their activities rapidly. In the second case, the wild-type tpv-Hsp14.3 and the mutant variants were incubated at 20°C, 35°C, 45°C and 60°C for 10 minutes separately and

DLS measurements were taken. Then, these data were converted into hydration radius (R_h) values which represent sizes of the oligomers in the solution and their distribution were recorded as peaks. According to these results, there was a general trend of decreasing in R_h size with increasing the temperature reducing in the number of peaks towards a single main peak.

Keywords: Small Heat Shock Proteins (sHSPs), Tpv-Hsp14.3, *Thermoplasma volcanium*, 3D structure modeling, Structure-function analysis.

ÖZ

THERMOPLASMA VOLCANIUM'UN MODİFİYE EDİLMİŞ KÜÇÜK ISI ŞOKU PROTEİNİ TPV-HSP14.3 YAPI MODELLEMESİ VE İŞLEVSEL ANALİZİ

Sheraj, Ilir

Yüksek Lisans, Biyoloji Bölümü

Tez Yöneticisi: Prof. Dr. Semra Kocabıyık

Mayis 2014, 256 sayfa

Bu çalışmada, termoasidofilik bir Arkeabakteri olan *Thermoplasma volcanium*'un küçük ısı şok proteini (sHSP) incelenmiştir. sHSP proteinler çeşitli stress koşullarında konakçı organizmanın hücresel proteomunu korumak ve hayatta kalmasını sağlamakta etkin olan düşük moleküler ağırlıklı proteinlerdir. Bu proteinin yapı-fonksyon analizi için deneysel yaklasim ve hesaplamalı yöntemler kullanılmıştır. Tpv-Hsp14.3 ile onunla yakından ilgili sHSP'ler ile yapılan dizi hizalamaları bir çekirdek α-kristalin alanı ve C-termina ucun ortasındaki bir V/I/L-X-V/I/L motifin yüksek oranda korunduğunu gösterilmiştir. Hesaplamalı ikincil yapı tahmini sonuçları, kanonik sınıf II Shsp'ler de olduğu gibi α-heliks yapısında bir N-terminal ve kısa bir ikincil yapısı olan C-ucu arasında sekiz β-iplikçikten oluşan bir çekirdek α-kristalin alanı bulunduğunu göstermiştir. Bu sonuçlar, tpv-Hsp14.3'ün hesaplamalı yöntemlerle tahmin edilen 3 boyutlu yapısı ile de doğrulanmıştır. 3D yapı modeli analizi tpv-Hsp14.3 monomerlerinin β6 ipliklerinin takası yolu ile dimerleştiğini göstermiş olup, bunlar daha sonra etkileşerek oligomerler oluşturmakta ve protein substratları bu yolla stresten korumaktadırlar.

Bu güne dek çalışılmış sHSP lerle farklı düzeylerde dizi hizalamaları ve yapı karşılaştırma çalışmaları yolu ile dimer oluşumu ve şaperon fonksyonuna katkıda bulunan önemli gruplar belirlenmiştir. Bunlar, R69 (R69K, R69E ve R69M değiştirilmesi için), R81 (R81K, R81E ve R81M olarak değiştirilmesi için), K87 (K87R, K87E ve K87I olarak değiştirilmesi için), ve QR(80-81) [(QR(80-81)EL olarak değiştirilek için)] guruplarıdır. Seçilen grupların kendi benzerleri, negative yüklü olanlar ve hidrofobik olanlarla yer değiştirmesi, bölgeye özel mutagenezle yapılmıştır. Daha sonra, mutantlar ve yaban tip protein, *E.coli*'de ifade edilmiş ve afinite kolon kromatografisi ile saflaştırılmıştır. Mutantların şaperon fonksiyonundaki değişikliği görmek için, domuz kalbi Sitrat Sentaz (CS) enzimi model olarak kulanılmıştır. CS, yaban tip ve sHSP mutant varyantları ile ayrı ayrı denaturasyon sıcaklığında (47°C) inkübe edilmiştir. Enzim aktivitesi CS için optimum sıcaklık olan 35°C'de ölçülmüştür. Enzim aktivite deney sonuçları R69E, R69K, K87R ve K87I mutant sHSP'lerin şaperon aktivitelerinin yaban tipe göre sırası ile 1.9-, 1.8-, 2.6- ve 2.7-kat daha yüksek olduğunu göstermiştir. Yaban tip tpv-Hsp14.3 ile karşılastırdığında K87E ve QR(80-81)EL mutantları'nın CS için şaperon aktiviteleri sırasıyla 1.2- ve 1.3-kat gibi az bir artış göstermiştir. Diğer taraftan, R69M ve R81K mutant sHSP ile inkübe edilen CS enziminin aktivitesi, yaban tiple inkübasyona göre sırası ile 0.9 ve 0.6 kat daha az bulunmuştur. CS R81E ve R81M ile inkübe edildiğinde, yabani tip tpv-Hsp14.3 ile inkübasyona göre aktivitesi önemli ölçüde değişmiştir.

Son olarak, dinamik ışık saçılımı (DLS) Spektroskopisi ile çözelti içinde tpv-Hsp14.3 yabani tip ve mutant varyantlarının oligomer dinamikleri çalışılmıştır. Bunun için, DLS iki farklı koşulda gerçekleştirilmiştir: biricisinde yaban tip tpv-Hsp14.3 tek başına ve domuz kalbi CS enzimi ya da sığır glutamat dehidrojenaz (GDH) enzimi ile bu enzimlerin aktivitelerini kaybettikleri sıcaklıklarda inkübe edilmiştir. İkinci durumda, yaban tip tpv-Hsp14.3 ve mutant varyantları, 20°C, 35°C, 45°C ve 60°C'de 10 dakika inkübe edilmiş ve DLS ölçümleri alınmıştır. Daha sonra, bu veriler oligomerlerin çözelti içindeki boyutlarını temsil eden hidrasyon yarıçaplarına (R_h) dönüştürülmüş ve bunların dağılımı tepe noktaları olarak

kaydedilmiştir. Bu sonuçlara göre, R_h değerleri sıcaklık artışı ile giderek azalma ve tepe nokta sayısının tek bir esas ana tepe noktasında indirgenme şekline genel eğilim gözlenmiştir.

Anahtar Kelimeler: Küçük Isı Şok Protein (sHSP), tpv-Hsp14.3, *Thermoplasma volcanium*, 3D yapı modelleme, yapı-işlev analizi.

To HSJ

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional structure
3D	Three-dimensional structure
ACD	Alpha-Crystalline domain
AGE	Agarose gel electrophoresis
Cryo-EM	Cryo-electron microscopy
CS	Citrate synthase
DLS	Dynamic light scattering
EM	Electron microscopy
GDH	Glutamate dehydrogenase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sHSP	Small hit shock protein
Tpv-Hsp14.3	*Thermoplasma volcanium* heat shock protein 14.3
UPS	Ubiquitin proteasome system
UPR	Unfolded protein response

CHAPTER 1

INTRODUCTION

1.1. Basics of protein Structure, Folding, and Molecular Chaperones

1.1.1. Protein Structure Characteristics

All the living organisms are made up of cells, which are the basic units of life. Cells on the other hand are made of macromolecules of Nucleic Acids, Proteins, Lipids and Carbohydrates. Proteins are responsible for cellular structural organization and functional maintenance. The basic building blocks of proteins are the 20 natural amino acids of the L-α-configuration which are linked together via peptide bond. The bond is formed by an amine nucleophile transferring a pair of free electrons to the to the carboxyl group of the previous amino acid to increase the chain length by one amino acid. As a result of this reaction, a water molecule is lost, and the incorporated amino acids are called "residues". These reactions are catalyzed in the cells by a complex cellular machinery called ribosome, but they can be also be carried out in laboratory by chemical synthesis under highly controlled reaction conditions. (Hughes, 2009)

The peptide bond is not a single bond which can rotate freely 360°, but rather it has restrictions owing to its partial double bond character resulting from electron distribution between amine and Carbonyl groups. Another further restriction to the polypeptide backbone rotation is imposed by the amino acids' R-groups. With the

exception of Glycine, which has only a Hydrogen atom as the R-Group, all the other amino acids have different groups, ranging from Alanine's methyl, to Tryptophan's large aromatic group. As a result, only two bonds can rotate in a polypeptide, and their allowed angles are represented by Ramachandran Plot (Ramachandran *et al.*, 1963). These angles are called φ between N and C_α, and ψ between C_α and carbonyl group as shown in figure1.1.
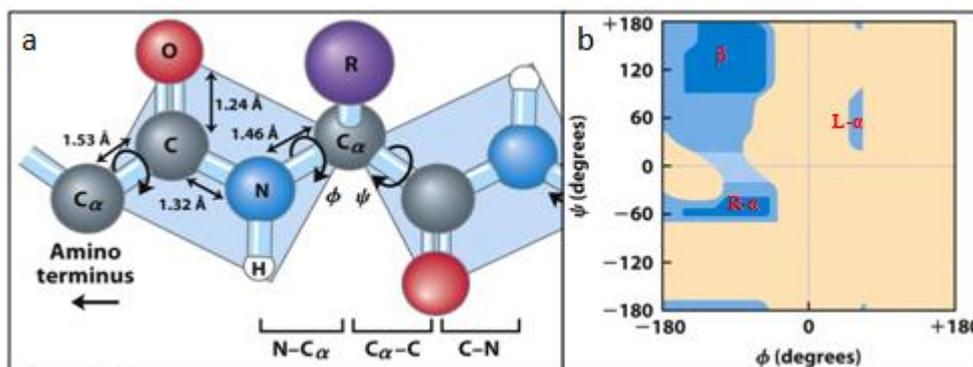
Figure 1.1: Polypeptide Bond Angles and Their Distribution in Ramachandran Plot. (a) The three bonds forming the polypeptide backbone. φ and ψ angles, together with bond lengths are shown. (b) Ramachandran Plot with the three most populated areas: β-strand region, the right-handed α-helix, and the left-handed α-helix, denoted by R-α and L-α respectively. [Nelson at al. Lehninger Principles of Biochemistry, 5th edition (W.H. Freedman and Company, 2008)]

The linear arrangement of amino acids is called the Primary Structure of the Protein, which according to Anfinsens Hypothesis carries all the information the polypeptide needs to fold (Anfinsen, 1973). This linear chain folds further to form the Secondary Structure of the Protein, which is represented in three main forms, alpha-helix, beta-sheet, and coiled-coil (or loop) structures. They were first predicted by Linus Pauling and Robert Corey (Pauling *et al.*, 1951), and were then experimentally proven by Max Perutz who obtained protein crystal diffractions (Perutz *et al.*, 1960). The main

forces forming and keeping the secondary structure are the hydrogen bonds between N-H group and the C=O group.

The first form of secondary structure to be theoretically predicted was the alpha helix (Pauling *et al.*, 1951). It's architecture is formed by hydrogen bonds between C=O group of nth residue, which acts a s a hydrogen bond acceptor, and the N-H group of $(n+4)^{th}$ residue, which acts as a hydrogen bond donor. One turn of the helix is made of 3.6 residues and it has a right-hand conformation. Since the H-bonds are in the same direction in the whole helical structure, this causes the helix to have a dipole moment. It is slightly positive at the N-terminus and slightly negative at the C-terminus. This has important implications in a protein's function and interaction, especially with ligands.

After alpha helix, Pauling and Corey (1951) predicted the beta-sheet secondary structure, which is the most populated region in the Ramachandran plot. This is because of the favorable φ and ψ angles. Beta-sheets are made of beta-strands, which are 5-10 residues long with broad angles. They are linked to each other through hydrogen bonds formed between the backbone N-H and C=O groups. Beta strands arrange themselves in two ways to form a beta sheet: parallel, where all of them run in the same direction, and antiparallel, where they run in opposite direction to each-other. There is also a mixed beta sheet, where the strands are both parallel and aniparallel, but this is not favored due to energy constraints. All beta strands are right-handedly twisted due to the favorable hydrogen bonding.

The third type of secondary structure is the coiled-coil, or more commonly known as loops. They are not regular structures like alpha-helices and beta-strands, but their length varies drastically. They are generally found on protein surfaces and participate in protein ligand interactions and enzyme active sites. Loops vary mostly within members of a protein family, and this is thought to be due to their capacity to tolerate mutations, since they will not cause a drastic change in the general three dimensional structure of the protein (Branden & Tooze, 1999).

The next level on the protein organization is the tertiary or three-dimensional (3D) Structure, also known as the Protein Fold. It is the functional form of the protein representing the spatial arrangement of the amino acid residues in three dimensional space. The main determinants of the 3D structures are the functional groups and the interactions they make with each-other. The major forces stabilizing a protein's 3D are the electrostatic forces, or salt bridges between positively- and negatively-charged groups, hydrogen bonds, and van der Waals interactions. A protein fold may be composed of many domains, which are defined as a part of the protein that, if separated from the main body, would maintain the function and 3D structure independently. Another part of the fold is the Motif, also known as Supersecondary Structure. They are made of a few secondary structural elements, but are not structurally independent.

The last level in the protein structure classification is the Quaternary Structure, which is made of several polypeptide chains associated to form a multimeric structure. They could be either heteromers, where different polypeptides come together, or homomers, where the same polypeptide chains are associated together. The different folds that form the quaternary structure can function independently, as stated before, or cooperatively, where the function of one subunit depends on that of the others. In this way, the activity of enzymes, or their interactions with ligands or other proteins are changed (Nelson and Cox, 2008; Branden and Tooze, 1999). All the levels of protein structure are shown in figure 1.2.

1.1.2. Protein Folding Theory

Protein folding process is one of several mysteries of molecular biology that has fascinated the scientists for more than half a century. It started with the elucidation of the first protein structure, that of Myoglobin by Perutz and Kendrew in 1958 (Kendrew *et al.* 1958). The cornerstone of protein folding is the Anfinsen's Hypothesis which states that: "The three-Dimensional structure of a native protein is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment…"(Frangsmyr, 1993).
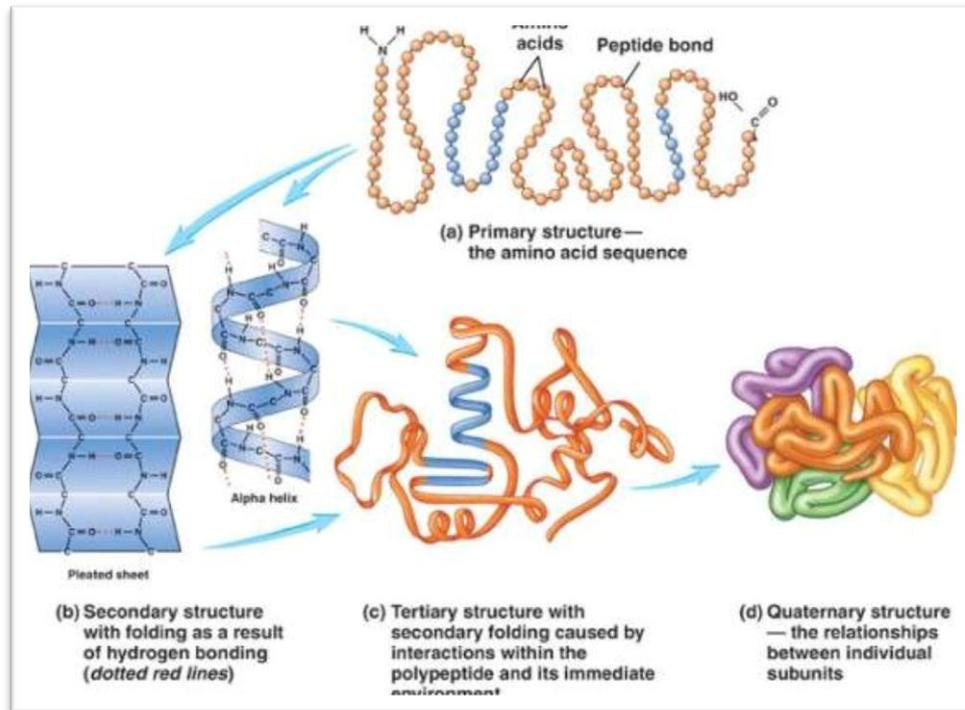
Figure 1.2: The Structural Organization of Proteins: The structural hierarchy is labeled in the figure. (Lodish *et al.*, Molecular Cell Biology, 5th Edition)

Protein folding can be defined as the process by which polypeptide chains self-assemble in a spontaneous and rapid way to form a 3D conformation that has a biological function (Jahn *et al.* 2008). The folding process is represented by landscape model as shown in figure 1.3.

At the top of the figure, the protein is in a state of high energy and high entropy. As the folding progresses down the path, entropy and energy will both decrease, forming more favorable interactions between the side chain and protein backbone atoms. More details regarding the processes will be given in the next subsections. Protein folding is complicated not only by very high number of possible space configurations of the residues, but also by inter-residue biophysical interactions, and the interactions of residues with their surrounding environment.
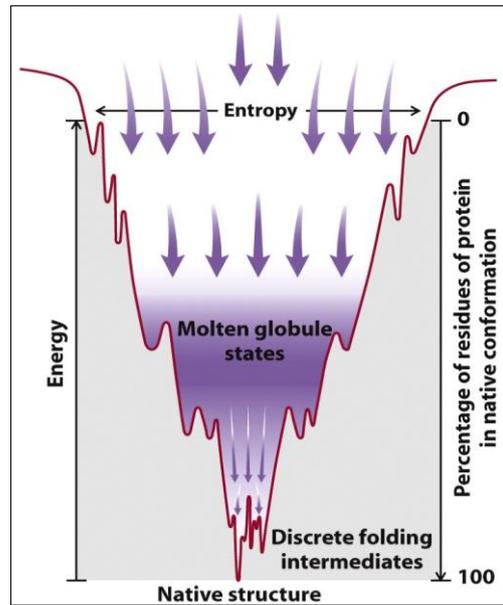
Figure 1.3: The Landscape Model of Protein Folding. (Nelson at al. Lehninger Principles of Biochemistry, 5th ed., W.H. Freedman and Company, 2008, p. 143)

There are many such interactions, but the most important ones are represented by the formula below (Conn, 2008).

$$E = E_{ss} + E_{es} + E_{hb} + E_{dip} + E_{rep} + E_{ps} + E_{pc}$$

E_{ss} is the side chain - side chain energy which may be attractive or repulsive; E_{es} is the side chain electrostatic (salt bridge) energy; E_{hb} are the hydrogen bond interactions between the backbones; E_{dip} is the backbone dipole interaction energy; E_{rep} is the repulsion energy when two atoms come very close to each other; E_{ps} represents the probability that some residues will form individual secondary structures, lowering the energy; and E_{pc} is the cooperative lowering of the energy if two adjacent amino acids assume the same secondary structure. To explain how all these complicated energy interactions form the final native structures, two major views on the folding process have been proposed: The "Classical View", and the "New View" (Hart *et al.* 2009).

6

1.1.3. The Classical Model

Classical Model was based on Levinthal Paradox, where Levinthal argued that since the number of the possible conformations for a protein are astronomical, then proteins must fold by specific "Folding Pathways" (Levinthal, 1968; Dill, 1997). This was because of the contradicting phenomena of the global energy minimum and the speed of protein folding, since for a 100-residue protein to fold on its own it would take 1.6×10^{27} years. In other words, a protein which has so many conformational possibilities has to "find" its correct conformation, and do so in a fraction of a millisecond. To make these two ideas compatible, Levinthal introduced the protein folding pathway in the folding landscape, according to which there must be a "well-defined sequence of events following one another" (Levinthal, 1968). According to the model, in the folding landscape there is a Kinetic Control and a Thermodymanic Control over protein folding. Thermodynamic control means that the final native structure of the protein is independent of the pathway, since thermodynamics is concerned only about the initial and the final energy states and not the pathway, but this process is very slow. Kinetic control on the other hand means that folding process is quick because it is pathway-dependent, but the protein can reach only local minimal energy. The combination of the two controls will give global energy minima, and a very quick folding. This model proposes that for large proteins there are multiple intermediates, while for small proteins there is only one intermediate state, as shown in the equations below, where N is the native state, X is the completely unfolded state, and I are the intermediate states.

Small Proteins: $X \rightleftharpoons I_1 \rightleftharpoons N$

Large Proteins: $X \rightleftharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons \ldots \rightleftharpoons N$

When the folding of the proteins is shown in the energy funnel, large proteins are shown to fold according to a pathway, while small proteins fold in a smooth funnel due to their folding being single-step (Figure 1.4 a and b).
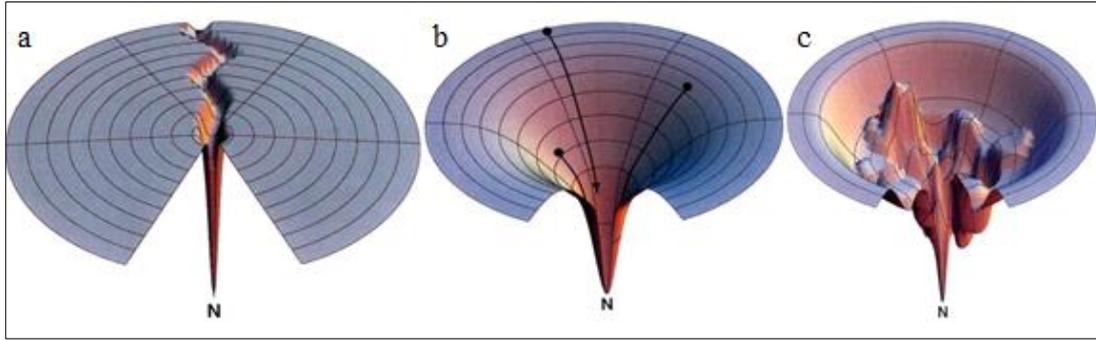
Figure 1.4: The Protein's Folding Energy Funnels. (a) Levinthal "folding pathway" for large proteins. (b) The smooth energy funnel for small proteins according to both the classical and the new view. (c) The rugged funnel for large proteins according to the new view (Dill *et al.*,1997).

1.1.4. The "New View"

According to the new view, folding is seen as a diffusion-like process with asynchronous individual chain movements which are continuously hit by the Brownian forces of other sequences (Bryngelson *et al.* 1995). They will all together find the way to the final native structure through the energy landscape. The new view is based mostly on statistical models. The folding process is analogue to water trickles flowing through mountain sides, and getting down to a main river. As the protein folds and goes down the funnel, it lowers both energy and the entropy, causing the narrowing of the funnel space. This narrowing could be the result of two different phenomena: An initial collapse of the extended polypeptide chain bringing distant sequences together and causing the formation of an initial tertiary structure. This causes a significant lowering of entropy because it decreases the freedom of movement of large chains. It is then followed by specific secondary structure elements formation and the more specific 3D native contacts. The other proposed mechanism is by forming the secondary structure before the 3D, which will lower the chain entropy and narrow the funnel width by restriction of individual residue movements. Then the 3D native structure formation is followed by the correct arrangement of the secondary structure chains. According to this view, a small

protein will fold quickly and it will have a smooth landscape (Figure 1.4.b), while large proteins will fold more slowly and their landscape will be rough (figure 1.4.c). In the landscape (figure 1.4.c) the hills represent energetic barriers and the deep ravines represent metastable but not native 3D structure of the protein.

After the "Classical View" was proposed by Levinthal, major research was directed towards the aim of finding the folding pathways of the proteins, but with no success. The latest research backed by the latest physical methods and computational biology favor the "New View". Especially in the last decade scientists found out that for the protein to fold, it is not left on its own, but rather the cell has a whole protein machinery dedicated to this task, that of Molecular Chaperones (Hartl, 1996; Hartl *et al.* 2009).

1.2.Molecular Chaperones and Proteostasis

1.2.1. The Important Role of Molecular Chaperones in Protein Folding.

Molecular chaperones are essential for protein folding and maintenance in the cell for many reasons. As it was explained in the section of protein folding theory, the folding process is not an easy matter, and in the highly crowded cellular environment it becomes even more complicated (Herbst *et al.* 1997; Kubelka *et al.* 2004). Firstly, although it was shown that folding is helped in a crowded environment because of concentration drive, the intracellular environment is mostly colloidal, made of possibly millions of different biological macromolecules. The total protein concentration in the cell is 300-400g/l. This means that unwanted interactions are highly probable if the estimation that a mammalian cell expresses more than 10.000 different polypeptides is taken into account (Ellis *et al.* 2006). To further complicate the problem, recent studies have found a new class of proteins called Intrinsically Disordered Proteins (IDPs) which in physiological conditions are highly dynamic ensembles rapidly interconverting from one form of secondary structure to another. Sometimes they even convert from a disordered molten globule to an ordered secondary structures to carry out their functions. Findings in Eukaryotes suggest that

9

25-30% of their proteins are mostly disordered, more than 50% have a long region of disorder, and that more than 70% of proteins involved in cellular signaling and cancer have long disordered regions (Parry *et al.* 2005; Xie *et al.* 2007; Xie *et al.* 2007; Vucetic *et al.* 2007). These proteins are involved in almost all biological functions, especially in regulation, control and signaling pathways in which low affinity but high specificity interactions are required, and also in post-translational modification and proteolytic cleavage sites (Dunker *et al.* 2005). These proteins are bound by molecular chaperones to maintain their structure and interact with their substrates, after which they assume their functional 3D conformation. The best example are the steroid receptors, which are bound by molecular chaperones, but when the ligands are available, they cause the dissociation of molecular chaperones and bind themselves to the receptors' hydrophobic pockets, causing receptor dimerization or multimerization, and as a result making possible their binding to the target gene promoters (Galigniana *et al.* 2010). Research in recent times has also shown that many polypeptides require molecular chaperones to fold in the cell (Dunker *et al.* 2008).

Secondly, the rate of protein synthesis in a bacterial ribosome is 20 amino acids per second and in Eukaryotic one 5-9 amino acids per second. The exit channel of the large ribosomal subunit has a length of 80-100Å, with a diameter of 10 Å at the narrowest and 20 Å at the widest region. This would prevent protein folding inside the ribosome channel except for alpha-helices formation. As a result, 30-60 amino acids, depending whether they are in an extended conformation or have formed alpha helixes at the C-terminal region, cannot form long-range interactions which are essential for protein folding. Given this information, it is understandable that the translation of a 300 residue protein, which is the mean polypeptide length would take 15-75 seconds. During this time, the partially folded nascent chains would be exposed to the cellular environment. For larger multi-domain proteins, the time is substantial. That would make them prone to non-native interactions in the cellular environment and as a result aggregation. To prevent such harmful effects, molecular chaperones interact with the nascent chain, rescuing it from unwanted interactions
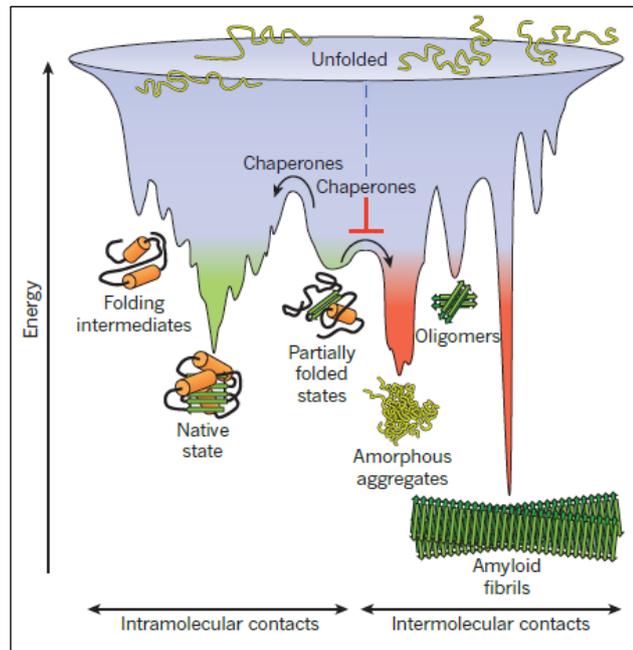
Figure 1.5: The Role of Molecular Chaperones in Protein Folding Funnels. (Hartl at al. 2011)

until the correct secondary structures and the long-range interactions are formed (Elcock, 2006; Kaiser, 2006).

The last but not least is the case when the chaperone system fails due to mutations or ageing. In such cases, many protein aggregation disorders start to show up. They have different clinical manifestations, from mild disease states, to even lethal ones. These will be explained later in their relevant sections (Morimoto, 2008; Balch *et al.* 2008). The role of molecular chaperones in the protein folding is shown in a simplified manner in figure 1.5.

As was stated above, the folding funnels are rugged for large proteins where there are kinetic barriers shown by hills separating the deep valleys. The valleys are metastable conformations of the proteins, but not the native structure. When proteins are entrapped in such valleys they need external energy to get uphill, and this energy is provided by molecular chaperones which interact with the entrapped protein and

bring it uphill to enter the funnel again on its route to the native conformation. These hills can be compared to the activation energy of chemical reactions. Another implication is that of protein aggregation that occurs as a result of mutations or some external stress causing protein misfolding and exposure of hydrophobic chains. These proteins will tend to aggregate, forming again metastable structures such as Amorphous Aggregates or Amyloid Fibrils which are toxic to the cell. Molecular Chaperones bind to these proteins before aggregation and tend to fold them in their correct native structure, and in case they cannot, they are sent to the degradation pathways (Hartl, 1996; Hartl *et al.* 2009; Mymrikov *et al.* 2011).

1.2.2. The Major Classes of Molecular Chaperones and their Properties

A Molecular Chaperone is defined as any protein that interacts with another protein, stabilizing it's native structure or help it to acquire its correct 3D conformation, without being part of the final structure (Hartl, 1996; Hartl *et al.* 2009). As this definition implies, the number of molecular chaperones should be vast, and indeed it was found to be the case. Such proteins are involved in almost all cellular functions such as *de novo* folding, protein structure maintenance, protein trafficking, signal transduction, cytoskeleton maintenance, and response to all kinds of cellular stress to mention just a few (Mymrikov *et al.* 2011). Their Classification is based on their molecular weight, and so the main classes are Hsp40, Hsp60, Hsp70, Hsp90, Hsp100 and the small HSPs (sHSP). HSP stands for Heat-Shock Proteins, because they were first identified in *D.melanongaster* and were correlated with increased expression after a mild increase in body temperature (Tissieres *et al.* 1974). Later they were found to be involved in all types of stress, but the conventional definition of Heat-Shock Proteins was not changed (Mymrikov *et al.* 2011). All these chaperone classes except sHSPs have an ATPases subunit which uses ATP to change the structure of the chaperone which then will be used to refold the misfolded or newly synthesized proteins. On the other hand, sHSPs are ATP-independent, and act generally as "Holdases" of misfolded proteins, providing the cell with a buffering capacity against the misfolded and potentially toxic proteins until they are correctly refolded by the ATP-dependent chaperones or sent to the degradation pathways (Frydman *et*

al. 1994; Hartl, 1996; Hartl *et al.* 2009). To have a better appreciation of protein folding in vivo, it is better to take a look at the whole process of a protein's life, from its biogenesis in ribosome, up to its degradation.

1.2.2.1.Protein Folding in Ribosome

As was stated in section 1.2.1., protein folding may start co-translationally in the ribosome exit channel, but this is not enough due to the low speed of translation and the crowded environment. So there are chaperones that interact directly with the ribosome exit channel to guide the newly-synthesized protein towards its native conformation. There are differences between Prokaryotes and Eukaryotes on the mechanism of interaction at this stage, and Archaea seems to be more closely-related to Eukaryotes than Bacteria. Bacteria use the Chaperone Trigger Factor (TF), while Eukaryotes use the Hsp70-J-protein system together with heterodimeric Nascent Polypeptide-Associated Complex (NAC). The best understood system is the Bacterial TF, and its action mechanism is shown in figure 1.6 (Kramer *et al.* 2002; Maier *et al.* 2003; Kaiser *et al.* 2006; Hartl *et al.* 2009; Kim *et al.* 2013).
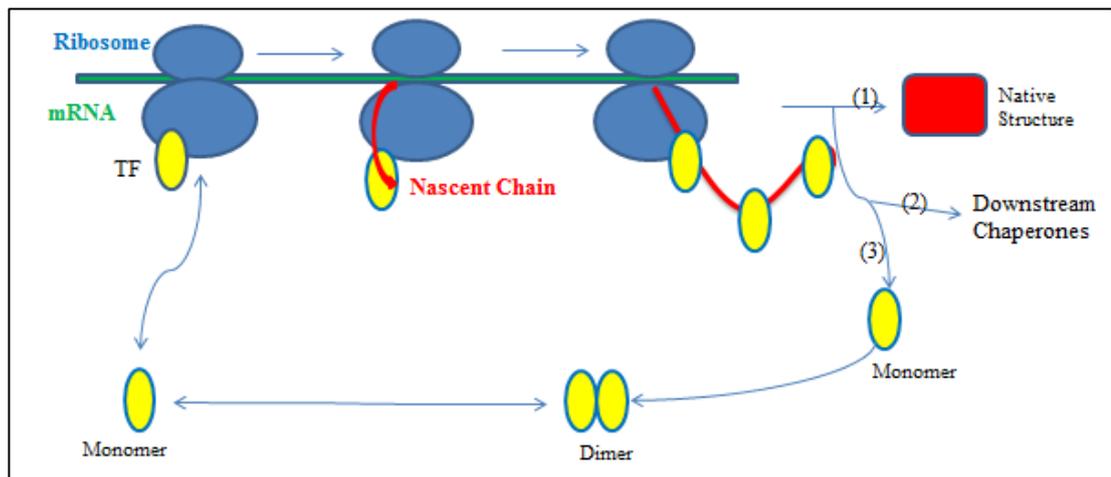
Figure 1.6: Trigger Factor Action Mechanism. The names of the structure are shown by the same color code.

Briefly, TF is in equilibrium between monomer and dimmer. When a nascent chain gets out of the ribosome, TF dissociates into monomer and binds the new protein. When the chain is elongated further, a second TF may bind to it. The possible number of TFs that bind to a polypeptide chain is not known, and to date there is not found any correlation between TF number and polypeptide chain length. If the protein acquires its native structure, then TF dissociates from it and is re-assembled into its dimeric form to bind another protein.

If the protein doesn't acquire its native structure, then it is transferred to the downstream ATP-dependent chaperone network to help it fold correctly. It is worth noting that about 70% of bacterial proteins after passing by TF acquire their native structure, and under normal conditions need no more help by the chaperones, while the rest 30% of larger size require the help of other chaperone classes, (Kaiser *et al.* 2006; Kim *et al.* 2013).

On the other hand, Eukaryotic and Archaeal systems are closely related to each other. The model system for Eukaryotes is Yeast, which has a system made of Ssb/Ssz/Zuotin, and NAC. Another component is the Prefoldin, which is mostly involved in the folding of Actin and Tubulin. Experiments have shown that deletion of any of the components of NAC/ Ssb/Ssz/Zuotin system causes cellular defects, meaning that both of them are essential (Gautschi *et al.* 2002; Hundley *et al.* 2002; Wegrzyn *et al.* 2005). The Eukaryotic system is still poorly understood mechanistically, and major efforts are under way for its elucidation. Recently a major interest has arisen for the Archaeal system also, which is the least understood of all (Hundley *et al.* 2002; Kim *et al.* 2013).

1.2.2.2.Hsp40/Hsp70 System

Hsp70 proteins are constitutively expressed. In bacteria they are also called DnaK, and they function in close association with Hsp40 (DnaJ). They work also in cooperation with a Nucleotide Exchange Factor (NEF) which catalyzes the ADP/ATP exchange to make the proteins available for additional cycles. The binding and release by Hsp70 is regulated by allosteric coupling of a conserved N-terminal

ATPase domain with a C-terminal Peptide-Binding Domain, which contains a beta-sandwich subdomain and an alpha-helical lid. Beta sandwich recognizes ~7 amino acid hydrophobic-enriched polypeptide segments, which occur every 50-100 residues in proteins. The alpha-helical lid and conformational changes in beta-sandwich regulate the protein's affinity for the target in an ATP-dependent manner (Zhu, 1996; Rudiger *et al.* 1997; Mayer *et al.* 2000; Young *et al.* 2003). So, starting from TF (figure 1.6), when the protein is unable to fold, either because of its length, or mutation, or some stress conditions, it is transferred to DnaK, which will then make it enter the cycle shown in figure 1.7.

First, Hsp40 binds the protein and delivers it to the ATP-bound Hsp70. Another function of Hsp40 is to accelerate the hydrolysis of ATP to ADP, causing the alpha-helical lid closure and as a result tight binding of the substrate to Hsp70. After this Hsp40 is phosphorylated and dissociates. Then NEF will cause the dissociation of ADP bound to Hsp70, and that position will then be occupied by ATP, which will open the lid again and as a result the substrate will be released. At this stage, the substrate may be folded to its native state. If it is not folded correctly, it will be recycled once again in the Hsp70 cycle, and if after a certain number of cycles the number of which is not known it cannot gain its native structure, then it is sent to Chaperonin system (Hartl *et al.* 2009; Kim *et al.*, 2013).

Bukau *et al.* (2006) proposed a Kinetic Partition process for the way Hsp70 prevents aggregation: Its binding to non-native substrates prevents aggregation by shielding the hydrophobic regions transiently, and at the same time reducing the aggregate-prone substrates concentration. The release will then permit fast-folding substrates to bury the hydrophobic region, whereas those that need longer folding time will be rebound by Hsp70 once again to prevent aggregation (Langer *et al.* 1992; Schiene-Fischer *et al.* 2002).
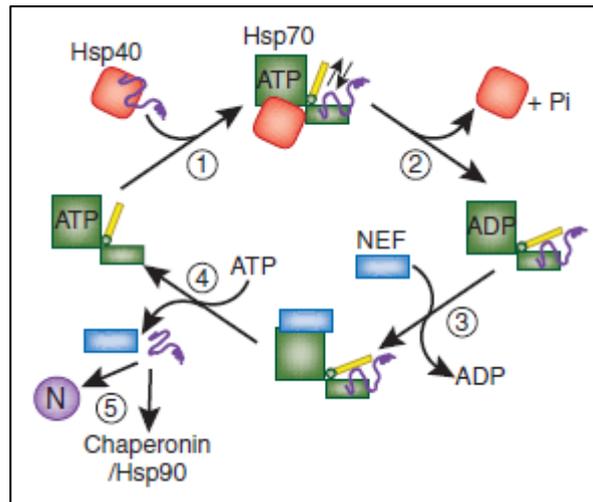
Figure1. 7: Mechanism of Hsp40/Hsp70 System. (Hartl at al. 2009)

1.2.2.3.The Chaperonins

Chaperonins are large double-ring protein complexes of 800-900 kDa molecular weight. They form a big cage inside which target proteins are enclosed and isolated from the surrounding environment. Chaperonins are divided in two groups (Horwich *et al.* 2007). Members of group I also called Hsp60 are found in Bacteria, Mitochondria and Chloroplasts. In bacteria they are called GroEL and are made of 7-membered rings. To function, they interact with Hsp10, in bacteria known as GroES. Their action mechanism is very complex, but here a summarized version is given (figure 1.8). First, the substrate protein is transferred to GroEL/GroES complex from DnaJ/DnaK after the latter's failure to re-fold it. ATP is used to move the domains of GroEL causing stretching of bound substrate and make it undergo local unfolding. Then GroES binds and encapsulates the GroEL, to form the complex GroEL/GroES. The substrate is allowed to fold in the chaperonine cage for some time. At the end of the cycle, if the protein is folded correctly it is released from the chaperonine cage, otherwise it is resent once again for another cycle of refolding inside the cage.
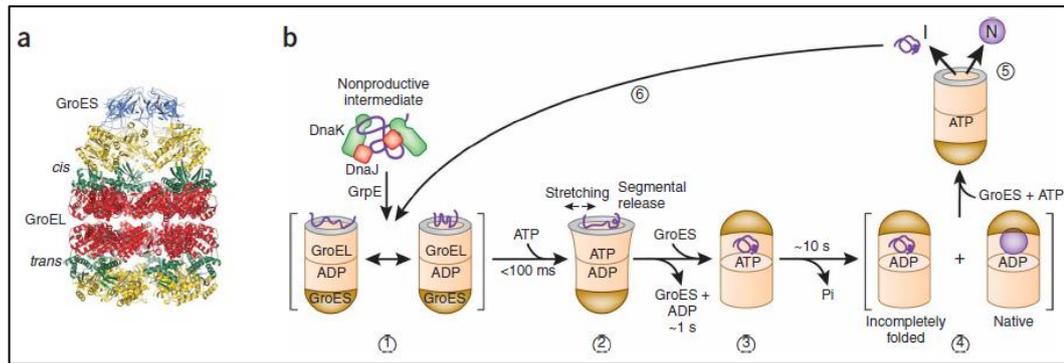
Figure 1.8: (a) Chaperonine Crystal Structure and (b) Action Mechanism. (Hartl *et al.* 2009)

Group II chaperonins are present in Eukaryotes and Archaea, and they are called as TRiC/CCT and Thermosomes, respectively. They consist usually of eight-membered rings, but sometimes it may be nine as well. In contrast to group I they are independent of Hps10 because the lid function of this chaperone is intrinsically built in some alpha-helical extensions in the chaperonin rings. Their mechanism is still a matter of debate and not much is known about their kinetics (Hartl *et al.* 2009).

In *E.coli*, GroEL interacts and help to fold correctly more than 250 cytosolic proteins, most of them 20-50 kDa of molecular weight, and complex α/β or α+β domain topologies, like the TIM Barrel fold (Gromiha *et al.* 2004). These proteins are stabilized by long range interactions, and as a result, populate kinetically-trapped folding intermediates, making it difficult to achieve the native structure. Enclosing inside the cage is thought to be beneficial for folding since it would isolate the substrate from the outside environment. Studies have shown that enclosure of the substrate increases the folding rate by 10-fold compared to spontaneous folding. Structural analyses have also shown that inside the cage are some essential polar residues. Molecular Dynamics simulations have suggested that the polar residues recruit and order water molecules around themselves, creating a polar local environment, and as a result forcing the substrate to bury its hydrophobic regions (England *et al.* 2008).

17

1.2.2.4. HSP90 Chaperone System

HSP90 is at the center of Protein homeostasis network (Proteostasis) involved in controlling many signaling pathways, such as Apoptosis, Telomere maintenance, signal transduction, and especially targeted protein degradation pathways (Taipale *et al.* 2010). It acts downstream of the Hsp40/Hsp70 complex, and is involved mostly in regulation of many signal transduction proteins like kinases and steroid receptors. At the same time, the action mechanism and kinetics of Hsp90 are highly regulated by co-factors. The crystal structure was determined recently for both the main protein and some of its regulatory factors (Ali *et al.* 2006; Shiau *et al.* 2006; Wandinger *et al.* 2008). In contrast to other chaperones, the kinetics of Hsp90 are not well understood, but progress is continuing towards that goal, especially after its suspected important role in cancer (Neckers, 2007). Based on experimental data, an action mechanism is proposed for it, and summarized in figure 1.9.
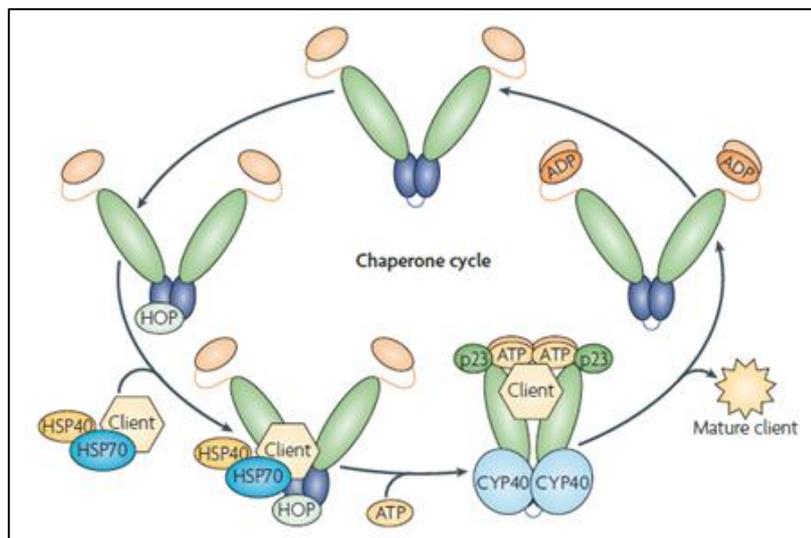
Figure 1.9: Hsp90 Action Mechanism. (Taipale *et al.* 2010)

First, the Hsp40/Hsp70 system transfers its bound client protein to Hsp90 by the help of HOP regulatory protein. ATP binds to the ATPase domain (orange color), causing the closure of the lid, and as a result Hsp90 dimerization occurs. Then ATP is hydrolyzed, causing the opening of the lid, and as a result of structural changes of Hsp90, the client protein bound to it mostly by means of hydrophobic regions is destabilized and then allowed to refold correctly. After this cycle, the substrate is released, and Hsp90 is once again free to bind another substrate.

1.2.2.5. Hsp100 Chaperones

Hsp100 is the Eukaryotic name for the bacterial homolog Clp. This is a whole protein family, which is part of the AAA+ ATPase Superfamily. The distinguishing feature of them is the presence of a basic core of 200-250 amino acids which forms an alpha-helical domain and a Walker-type Nucleotide Binding Domain. Hsp100/Clp is divided into two groups (figure 1.10) (Hanson *et al.*, 2005).
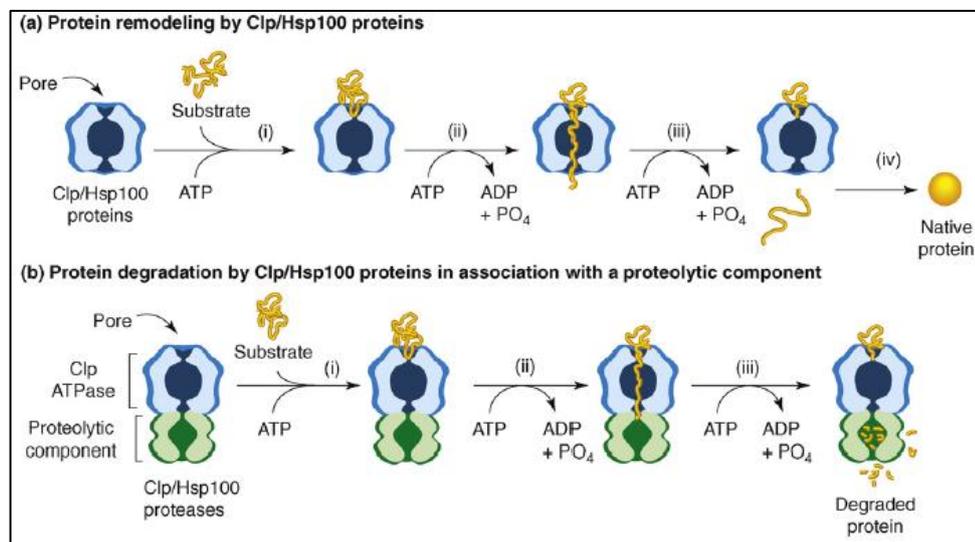
Figure 1.10: Action Mechanism of Hsp100. (a) Hsp100/Clp binds to the aggregated substrate and causes its unfolding by passing through its central channel so that it can be refolded again. (b) Hsp100/Clp in conjuction with a Protease component forces the aggregated substrate through its channel, unfolding it, and then it is degraded in the proteolytic component. (Doyle at al. 2008)

Group 1 consists of proteins with two AAA+ modules, and its members include Hsp104, ClpB, mitochondrial Hsp78 and plant Hsp101. Group 2 consist of protein with one AAA+ module, such as ClpX and HslU. These chaperones have the ability to solubilize any protein that is aggregated after severe cellular stress condition by using ATP in conjunction with Hsp40/Hsp70 system (Wickner *et al.* 1994). The mechanism is still not settled, but there appear to be two action mechanism, one for substrate refolding (figure 1.10.a), and the second for substrate degradation (figure 1.10.b), depending whether Hsp100/Clb interacts with a proteolytic component or not (Hanson *et al.* 2005). The mechanism proposes that exposed hydrophobic patches of aggregated protein are recognized by the system and are unfolded by passing through the channel of the complex in a reaction dependent on ATP hydrolysis. Then the polypeptides are released to refold again, and if they cannot, they are sent to Hsp40/Hsp70 system. If the polypeptides are bound by Hsp100/Clp which is interacting with a protease component, the protein is passed through the channel and this time it is degraded. The way of choosing one of these pathways is not known (Sauer *et al.* 2004; Doyle *et al.* 2008).

1.2.2.6.Small Heat Shock Proteins (sHSPs).

Small Heat Shock Proteins are members of the chaperone Superfamily, but they are ATP-independent. These are small proteins varying in molecular weight from 12-43 kDa. Their defining feature is the Alpha-Csystallin Domain (ACD) which is flanked by an N- and C-terminus of varying length between different organisms. sHSPs are accepted not to be involved directly in protein folding, though there are reports for such occurrences (Basha *et al.* 2011; Mchaurab *et al.* 2009). They act mainly as "Reservoirs" of mislfolded proteins, binding to them and keeping them in that state under stress, preventing aggregation. Then they interact with Hsp70 system, which uses ATP to refold the misfolded proteins (Narberhaus, 2002). More detailed information for sHSPs will be given in the next section.

1.2.2.7. Protein Degradation in Proteasome and Autophagy

Although Proteasomes and Autophagy are not related to Molecular Chaperones functionally, they are a part of the Protein Quality Control (PQC) system and the Proteostasis Network, which maintains the proteome of the cell (Tyedmers *et al.* 2010; Winkler *et al.* 2012). As a critical component of this network, proteasomes are large molecular machines involved in protein degradation. It is found in all three domains of life, but the basic understanding of its mechanism was elucidated by using the Archaeal *Thermoplasma acidophilum* as a model. Subsequently, detailed structural and assembly studies were carried out in Bacteria and Eukaryotes. The exact mechanism of how proteins are directed to proteasome, especially in Bacteria and Archaea that have no ubiquitin system is not known. In Eukaryotes, proteins to be degraded are tagged by a certain code of Ubiquitin, which destines them for degradation (Hershko *et al.* 1998; Schubert *et al.* 2000; Finley, 2009).

Autophagy is another degradation pathway that is less selective than proteasome but more bulky (Lee *et al.* 2013). It is a lysosomal degradation pathway essential for survival, differentiation, development, and homeostasis. In autophagy, membraneous structures called Autophagosomes are formed around organelles or protein aggregates, engulfing them into membraneous structures which latter fuse with cellular lysosomes to form an Autolysosome. The cargo inside the autolysosome will be degraded as in any other lysosomal degradation (Boya *et al.*, 2013). In addition to damaged cellular organelles, clearing of protein aggregates by autophagy has turned out to be very essential, and the cross-talk between autophagy and proteasome systems is extensive (Kon *et al.* 2010; Arias *et al.* 2011), so that when one is blocked, the others efficiency is increased to deal with the accumulating damaged proteins (Kopito, 2000; Kaganovich *et al.* 2008; Tyedmers *et al.* 2010; Arndt *et al.* 2010). Now that almost all major components of the Proteostasis network, were described, a brief explanation of their integration and cooperation in maintaining the cellular physiology can be made.

1.2.2.8. The Proteostasis Network.

PQC acts like a "check-up" system that the proteins are folded correctly. If some proteins fail to be folded, or are misfolded as a result of stress or mutations, this system is activated. It is best understood in Rough Endoplasmic Reticulum (RER) where a large part of protein synthesis takes place in Eukaryotes. It is an elaborate system of molecular chaperone and signal transduction pathways (Buchberger *et al.* 2010). Proteostasis is the Proteome Homeostasis, or proteome maintenance system of the cell (Kim *et al.* 2013). It is made of approximately 1300 cellular proteins in mammals, of which 400 are involved in protein biogenesis, 300 in protein conformational maintenance, and about 700 in protein degradation pathways, with about 100 of them being common to more than one of these systems (Haynes *et al.* 2010). This system integrates molecular chaperones for correct protein folding and transport, with the machinery of disaggregation i.e., oligomeric AAA+ proteins, and degradation of misfolded proteins, i.e., Ubiquitin-Proteasome System (UPS) and Autophagy. The huge number of components shows the importance of this system in maintaining the cellular physiology. Integrated in the network pathways are the Cytosolic Heat-Shock Response, Unfolded Protein Response (UPR) in Endoplasmic Reticulum, UPR in Mitochondria, and other cellular stress pathways such as oxidative stress, inflammation and nutrient starvation, including cellular ageing and longeivity pathways (Morley *et al.* 2004; Cohen *et al.* 2006; Ben-Zvi *et al.* 2009; Cohen *et al.* 2009; David *et al.* 2010). About 700 proteins involved in degradation include the components of Ubiquitin-Proteasome System (UPS) and components of Autophagy (Walter, 2011). A summarized version is shown in figure 1.11.

Recent studies have shown the importance of the proteostasis network by demonstrating that in Eukaryotes, approximately half of the proteins synthesized in ER do not satisfy the requirements of Protein Quality Control System, and for some others, the folding success rate is even lower (Schubert *et al.* 2000). On the other hand, deficiencies in proteostasis, either as a result of mutations, or age-related decline in its capacity to deal with proteome control (Morley *et al.* 2002), have been shown to be related to various diseases such as neurodegeneration, lysosomal storage
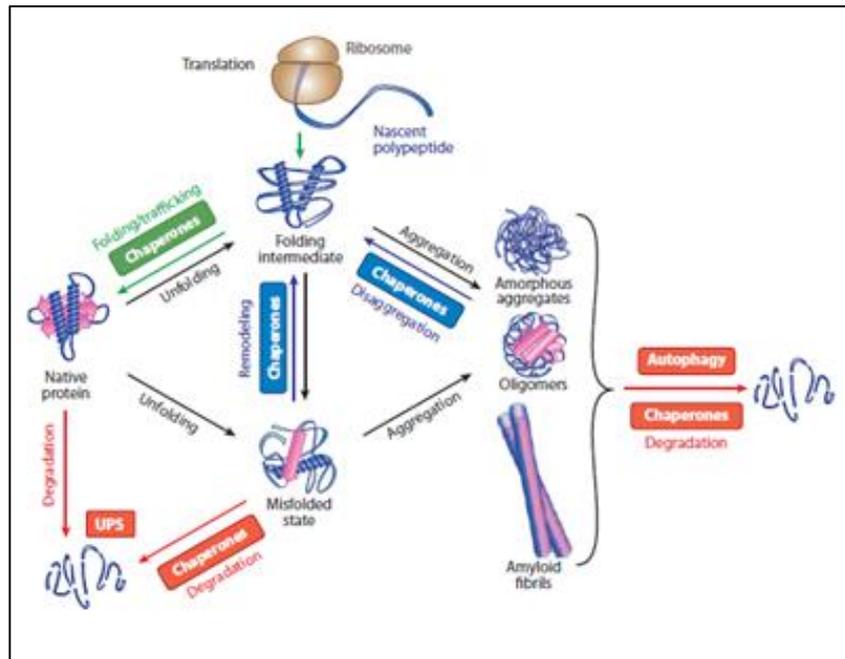
Figure 1.11: Proteostasis Network. (Kim *et al.* 2013)

disease, cystic fibrosis, cancer and cardiovascular diseases (Forman *et al.* 2004; Olzscha *et al.* 2011).

1.3.The Small Heat Shock Proteins

1.3.1. General Characteristics of small HSPs.

Small HSPs are a versatile class of molecular chaperones first detected to be expressed after *Drosophila* was subjected to mild heat stress (Tissieres *et al.* 1974). They are organized into multimers to carry out their actions, but this is not always the case. The molecular weight of a sHSP monomer ranges from 12-43 kDa, and the majority are between 14-27 kDa (Haslebeck *et al.* 2008; Mchaourab *et al.* 2009, Basha *et al.* 2011). The sequence homology of these proteins between organisms, and even between different classes of the same organism is very low (Waedick *et al.* 2009). Despite this, structure determination of different sHSP classes from different

domains of life has shown a general structural conservation of the α-Crystallin Domain (ACD), which is the "flagship" of the sHSPs (Basha *et al.* 2011).

The ACD is located C-terminally and is made up of an average of 90 amino acid residues. It is contains seven antiparallel β-strands forming a β-sandwich resembling an immunoglobulin fold. The ACD is found between an N-terminal and a C-terminal region. The N-terminus is a relatively long sequence compared to the C-terminus, and both its length and sequence conservation is very low. It is generally amphipathic in nature. The C-terminus is far shorter, and its sequence identity is low as well (Haslebeck *et al.* 2008; Mchaourab *et al.* 2009, Clark *et al.* 2011, Basha *et al.* 2011). The only conserved region of it, in at least 95% of sHSPs, is the I/L-X-I/L motif near the end (Waedick *et al.* 2009). Moreover, in some archaeal species, a tenth β-strand is found at the middle of C-terminus, thought to be important in multimer formation (Bertz *et al.* 2010).

Structurally, sHSPs may be divided into three categories, depending on their sequence lengths and chaperone activity: The first category is the scope of this study. The ACD of these sHSPs is in the middle flanked by N- and C-termini and forms higher oligomeric structures with chaperone activity (Haslebeck *et al.* 2008). The second category is that of some sHSPs found in *C.elegans* which have the shortest N-terminal region and are generally in dimeric form. They have no chaperone activity and were found to be induced in a development-dependent manner (Kokke *et al.* 1998: Kokke *et al.* 2001). The third category is that of Tapeworms, which have two ACDs in a single polypeptide, joined by a modified C- and N-Terminus in between. They are found to be tetramers under oxidizing conditions and dimmers under reducing conditions. They have chaperone function, and their interaction is different from the classical chaperones (Kappe *et al.* 2004).

1.3.2. Small Heat Shock Proteins in the Three Domains of Life

Eukaryotes are the organisms containing the highest number of sHSPs, and this is not surprising given the complexity of multicellular organisms and their demands for maintenance of the cellular and tissue physiology (Waedick *et al.* 2009). The

organism with the lowest sHSPs number is *S.cerevisiae,* a unicellular eukaryote, which has only two sHSPs: Hsp26 (Franzman *et al.* 2005; Franzman *et al.* 2008; Benesch *et al.* 2010) and Hsp42, while the organisms with the highest number of sHSP genes are plants. The genome of *A.thaliana* which is the best studied plant species, encodes 19 sHSPs. The land plants contain 11 families of sHSPs, and each family has a number of paralogs and orthologs. Of those, six families encode proteins targeted for cellular organelles like nucleus, chloroplasts, mitochondria, peroxisomes and endoplasmic reticulum, while five stay in cytosol. Such organelle-targeted sHSPs are found uniquely in plants with the exception of *D.melanogaster*, which contains one sHSP targeted for mitochondria (Waedick *et al.* 2009).

Humans on the other hand have 10 sHSP genes, named HSPB1-HSPB10 (Mymrikov *et al.* 2011). Their molecular mass varies between 17-28.4 kDa. Some of them are expressed in all organs, like HSPB1, HSPB5, HSPB6 and HSPB8, while the rest are expressed in certain tissue types. The best known among them are HSPB4, which is the standard name of αA-Crystallin, and HSPB5, which is the standard name of αB-Crystallin. All members of the HSPB have chaperone activity, and some of them seem to even have specific substrates. In addition to their chaperoning function during physiological stress, they are involved in a range of physiological functions, such as maintenance of cytoskeleton, muscle contraction, platelet function, insulin-dependent metabolism regulation in muscles, cell survival and apoptosis, carcinogenesis, and even in the immunity response (Horowitz, 2000; Akerfelt *et al.* 2010; Mymrikov *et al.* 2011; Clark *et al.* 2012).

Small HSPs are also present in almost all the bacteria studied to date with few exceptions. The general trend is having one or two sHSP genes only, but there are cases where there is none, or cases with 3, 4 or even 10 sHSP genes (Narberhaus, 2002). A question arises as to what makes some bacteria have none, or so many sHSP genes. The answer is not known, but certain speculations can be made. Some microorganisms with no sHSP genes have a common characteristics of possessing a very small genome, because such organisms devote their genetic information to essential functions only (Fraser *et al.* 1995). However, rather than genome size, the

lifestyle of each microorganism appears to be the more important determinant of the presence or absence of sHSPs (Shigenobu *et al.* 2000). This is supported by the observation that some microorganisms have sHSP genes although they have small genomes. These organisms live in insects whose body temperature fluctuates with that of the ambient, and so they are essential to deal with heat/cold shock. On the other hand, microorganisms of similar genome size that live as parasites in humans or other homeothermic animals have no such genes, because the temperature of the host is constant (Read *et al.* 2000). There is also the other group of organisms that contain 4-10 sHSP genes. They are generally plant-root inhabitant species, and to date no plausible explanation can be suggested for the presence of such a high number of sHSP genes in their genomes (Natera *et al.* 2000).

According to the present data, all the archaea domain species contain one or two and sometimes three sHSP genes with the exception of *Halobacterium sp*. strain *NRC-1* which has five such genes. Given the extreme conditions in which Archaea live it seems puzzling for them to have such proteins, especially for some who also lack homologs of Hsp40/70, Hsp90 and Hsp100, which in other organisms seem to be more important in protein folding. However, recent research suggests that the chaperonin system (also known as thermosome), has a more essential role in Archea than in other organisms (Large *et al*. 2009). Or alternatively, there may be other unexplored chaperone systems unique to Archaea (Lund, 2011).

1.3.3. Transcriptional Regulation of the Heat Shock Response.

It is very important for the cell to control the heat shock response at the transcriptional level, otherwise higher or low activation of this response would compromise the protein homeostasis and the energetic balance of the cell (Spriggs *et al.* 2010; de Nadal *et al.* 2011). The best understood regulatory mechanisms are those of Eukaryotes. Invertebrates have only one Heat Shock Factor (HSF) which acts as the transcription factor activating the cellular response to stress, while the vertebrates and plants have four of them, HSF1-HSF4 (Akerfelt *et al.* 2010). The master regulator is HSF-1, while the others are involved also in other cellular functions like

development and cellular lifespan. HSP transcription is controlled by Heat Shock Elements (HSEs) found upstream of the *hsp* genes in multiple copies. Under normal physiological conditions, HSF-1 is kept in a monomeric form by interacting with Hsp90, and according to recent evidences also with Hsp40/70 system of chaperones (Guo *et al.* 2001). It is thought to be phosphorylated in the monomeric form, or after trimerization on the DNA (Shamowsky *et al.* 2006). When cells are under stress conditions, Hsp90 chaperones release the HSFs because they are needed to deal with the unfolded proteins that accumulate in the cell, so HSFs are free to form trimer (Shi *et al.* 1998). There is also another contributing factor of a Ribonucleoprotein complex made of eukaryotic Elongation Factor-1A (eEF1A) and Heat-Shock RNA-1 (HSR-1), a noncoding RNA species with a thermosensing capacity, which upon heat stress undergoes a conformational change, and together with eEF1A facilitates HSF1 trimerization (Shamowsky *et al.* 2006). After trimerization, HSF1 DNA-Binding Domain (DBD) is exposed, it enters the nucleus, and binds to HSEs. There it causes the release of a preinitiated paused RNA Polymerase II complex, causing a rapid cellular response to the damaged proteins in the cell. When the cellular response to stress is finished, it is bound again by Hsp90 and it is also acetylated in certain residues (Petesch *et al.* 2008). The mechanism is shown in figure 1.12.

Once the heat shock response is activated, there are major changes in the cell to cope with the danger of unfolded and aggregated proteins (Spriggs *et al.* 2010). The main effectors are the Eukaryotic Translation Initiation Factor 4E (eIF4E), eIF2a, eIF4G family members and the stress granules, which are cytoplasmic RNA-protein complexes containing non-translating mRNAs, translation initiation components and other additional proteins that affect mRNA function (Buchan *et al.* 2009). They are activated by stress and affect mRNA translation and stability. Translation Initiation Factors are phosphorylated in certain residues, and so protein synthesis is halted, while the stress granules are activated. The net result is a transient decrease in the biogenesis of growth-related proteins, and an increase in the synthesis of stress-related proteins (Lopez-Maury *et al.* 2008; de Nadal *et al.* 2010; de Nadal *et al.* 2011).
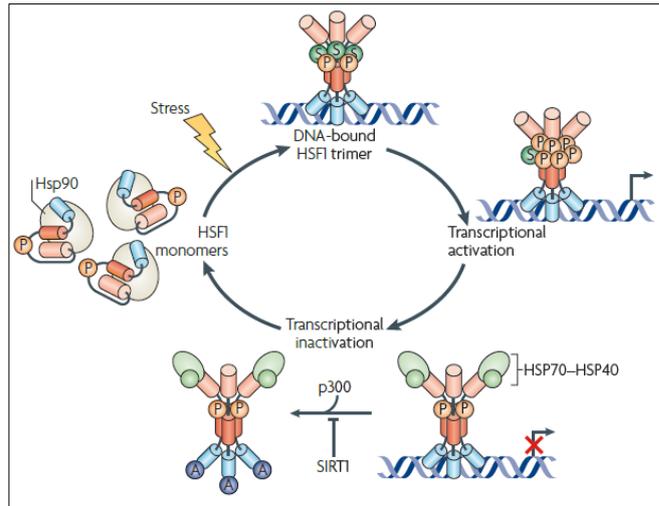
Figure 1.12: HSF1 Activation. (Akerfelt at al. 2010)

There are indications that pre-formed transcription complexes, and translation of existing transcripts in stress granules, generates proteins more rapidly than would *de novo* transcriptional induction, making it crucial for the initial adaptive response to stress. Simultaneously with the induction of HSP gene transcription, there is an increase in the efficiency of HSP mRNA translation, but a decrease in the global rates of protein synthesis (de Nadal *et al.* 2011).

Transcriptional regulation of Heat Shock Response in Bacteria is done by at least two different mechanisms: The first one is Positive Regulation, in which a specific Sigma Factors, in *E.coli* particularly Sigma Factor σ^{32} (RpoH), upregulates the expression of its target genes during heat shock. σ^{32} factor is negatively regulated by DnaJ/K machinery which recruits it in normal physiological conditions. During heat shock, or other stress conditions, the misfolded proteins bind to DnaJ/K complex so σ^{32} factor is released and it will direct RNA Polymerase to at least 30 of its target genes involved in heat shock response and express them. The σ^{32} factor upregulation is done by increased transcription rate of *rpoH* mRNA and by stabilization of the transcript. After the stress is over, σ^{32} factor is again bound by DnaJ/K complex and it is sent to FtsH protease for degradation (Gross, 1996; Yura *et al.* 2000).

28

The second mechanism is the Negative Regulation, which depends on a repressor protein or a cis-acting element binding to the promoter region of the target genes (Narberhaus, 1999). When temperature is increased, the DNA-Repressor interaction is abolished, and the promoter site is open for interaction with RNA Polymerase, and as a result the heat shock genes are induced. One example is the Controlling Inverted Repeat of Chaperone Expression (CIRCE), which acts as a binding site for a repressor protein HrcA, and also promotes rapid mRNA turnover by forming a destabilizing secondary structure (Zuber *et al.* 1994). HrcA depends on GroEL/ES chaperonins to maintain it in the correct conformation, and during heat shock, the chaperonin is recruited to help correct folding of other proteins. This deprives HrcA from attaining its correct 3D structure, and as a result its release from the promoter regions of the target genes and bound mRNAs, taking off the brakes for heat shock response (Babst *et al.* 1996; Mogk *et al.* 1997).

There are also other emerging mechanisms involving short mRNAs, especially in the *Rhizobium* genus. These short RNAs attain a secondary structure and bind to the mRNA's 5'-UTR region, masking the Ribosome Binding Site (RBS) and as a result blocking translation. During heat shock, the secondary structure of this short RNA is destabilized, and as a result it is released, relieving the heat shock response genes (Naberhaus *et al.* 1996; Naberhaus *et al.* 1998; Nocker *et al.* 2001). There are also other mechanisms, but these three, and a combination of positive and negative repression are the best understood to date (Narberhaus, 2002).

The least understood mechanisms of Heat Shock Response are those of Archaea. The main problem is that Archaea have transcriptional machinery homologous to that of eukaryotes, including the TATA-Binding Protein (TBP), TFIIB, named TFB in Archaea, and an RNA Polymerase II (Bell *et al.* 1998; Soppa, 1999). On the other hand, the transcriptional regulators are homologous to the bacterial ones. Since they do not have any σ-like factors similar to bacteria, or Heat Shock Factors (HSFs) and Heat Shock Elements (HSEs) like Eukaryotes, their heat shock response mechanisms are not clear (Thompson *et al.* 1998; Thompson *et al.* 1999).

One of the best studied mechanisms is the heat shock response of *Pyrococcus furiosus* Archaeon. These are hyperthermophilic organisms growing optimally at about 100°C. Their heat shock response is regulated by the heat shock regulator Phr. In a study to find out the minimal promoter sequence of HSR in *Pyrococcus furiosus,* they found that the promoter of AAA+ ATPase and other heat shock proteins had a typical TATA-box at -25 followed by a purine-rich BRE element, while the cis-acting elements are poorly defined. A consensus sequence was defined where Phr regulator binds. The way in which heat response is activated however is different from Eukaryotes and Prokaryotes: Phr binding to this consensus sequence prevents RNA polymerase recruitment at target gene promoters under optimal growth conditions. During stress response, the protein is released and as a result transcription is initiated (Keese *et al.* 2010).

In another study in *Haloferax* and *Halobacteria sp. NRC-1,* Lu *et al*. (2008) studied the transcription mechanism of the heat-induced *hsp5* gene. They found the essential heat-shock response elements to be confined at the region between TATA box and BRE. However, the way transcription is activated is different from that of *Pyrococcus furiosus*. In these species, under heat shock, some heat-inducible general transcription factors such as TFB2 and TFBb associate with different TBPs in an analogous way to the Bacteria's alternative δ-factors to modulate the heat-shock response.

Hyperthermophiles, in addition to induction of HSP genes, accumulate compatible solutes such as di-myo-inositol phosphate (DIP) as much as 20-folds or terhalose (Santos *et al.* 2002). These are what are currently known for the Archaea heat shock response, and it is clear that major research is required in this field.

1.3.4. Structural Characteristics of Small Heat Shock Proteins

As stated in the previous sections, the Small Heat Shock Proteins are very diverse in their primary sequence and their oligomerization. In figure 1.13 a sequence alignment of the proteins whose structures are determined is shown.

```
HSPB1_HUMAN      ---------------------------------------MTERRVPFSLLRGPSW---        16
HSPB6_RAT        -----------------------------------------MEIRVPVQPSWLRR          14
CRYAB_HUMAN      -------------------------------------------MDIAIHHPWIRR          12
CRYAA_BOVINE     -------------------------------------------MDIAIQHPWFKR          12
Q8UUZ6 ZEBRAFISH -------------------------------------------MDIAIQHPWFRR          12
HSPS METJA       --------------------------------------------------MFGRD           5
Q970D9 SULTO     --------------------------------------------------MYYLG           5
HSP26_YEAST      MSFNSPFFDFFDNINNEVDAFNRLLGEGGLRGYAPRRQLANTPAKDSTGKEVARPNNYAG     60
HS16B_WHEAT      ----------------------------------------------MSIVRRTNVFD-        11
Q8PNC2_XANAC     ---------------------------------------------MNIVRYPQFPTH         12

HSPB1_HUMAN      --DPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPGYVRPLPPAAIESPAVAAPAYS     74
HSPB6_RAT        ASAPLPGFSTPGRLFDQRFGEGLLEAELASLCPAAIAPYYLR--------------APS     59
CRYAB_HUMAN      PFF---PFHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRP--------------PSF     54
CRYAA_BOVINE     TLG---PF-YPSRLFDQFFGEGLFEYDLLPFLSSTISPYYRQ--------------SLF     53
Q8UUZ6  ZEBRAFISH TLG------YPTRLFDQFFGEGLFDYDLFPFTTSTVSPYYRH--------------SLF    51
HSPS_METJA       PFD-----SLFERMFKEFFATPMTGTTMIQ--SSTGIQISGK------------------     40
Q970D9 SULTO     KEL-----QKRSEELS-------------------RGFYEL------------------     22
HSP26 YEAST      ALY-----DPRDETLDDWFDNDLSLFPSGFGFP-------RS-----------------     90
HS16B_WHEAT      PFA-----DLWADPFDTFR----SIVPAISGGGSET---AAF-----------------     41
Q8PNC2_XANAC     ALQ-----NEIKHVFDRFF----EQ----NGDTDESAVVTAQ-----------------     41
                                      :.

                              β2        β3              β4           β5
HSPB1_HUMAN      RALSRQLSSGVSEIRHTADRWRVSLDVNHFAPD---ELTVKTKDG-VVEITGKHEE----     126
HSPB6_RAT        VAL------PTAQVPTDPGYFSVLLDVKHFSPE---EISVKVVGD-HVEVHARHEE----     105
CRYAB_HUMAN      LRAPSWFDTGLSEMRLEKDRFSVNLDVKHFSPE---ELKVKVLGD-VIEVHGKHEE----     106
CRYAA_BOVINE     RTVLD---SGISEVRSDRDKFVIFLDVKHFSPE---DLTVKVQED-FVEIHGKHNE----     102
Q8UUZ6 ZEBRAFISH RNILDSSNSGVSEVRSDREKFTVYLDVKHFSPD---ELSVKVTDD-YVEIQGKHGE----     103
HSPS METJA       ------G-FMPISIIEGDQHIKVIAWLPGVNKE---DIILNAVGD-TLEIRAKRSPL-MI      88
Q970D9 SULTO     ------V-YPPVDMYEGGYLVVVADLAGFNKE---KIKARVSGQNELIIEAEREITEP-     71
HSP26_YEAST      ------V-AVPVDILDHDNNYELKVVPGVKSKKDIDIEYHQNKNQ-ILVSGEIPSTLNE     142
HS16B_WHEAT      ------A-NARMDWKETPEAHVFKADLPGVKKE---EVKVEVEDGNVLVVSGERTK-EKE      90
Q8PNC2_XANAC     ------W-VPRVDIKEEVNHFVLYADLPGIDPS---QIEVQMDKG-ILSIRGERKS-ESS      89
                        .      .    :  .    .      :   *  .

                          β6         β7                β8         β9
HSPB1_HUMAN      -RQDEHGYIS----RCFTRKYTLPPG--VDPTQVSSSLSPEGTLTVEAPMPKLA--TQSN     177
HSPB6_RAT        -RPDEHGFIA----REFHRRYRLPPG--VDPAAVTSALSPEGVLSIQATPASAQAS----     154
CRYAB HUMAN      -RQDEHGFIS----REFHRKYRIPAD--VDPLTITSSLSSDGVLTVNGPRKQ----VSGP     155
CRYAA BOVINE     -RQDDHGYIS----REFHRRYRLPSN--VDQSALSCSLSADGMLTFSGPKIPSGVDAGHS     155
Q8UUZ6  ZEBRAFISH -RQDDHGYIS----REFHRRYRLPSN--VDQSAITCTLSADGLLTLCGPKT-SGIDAGRG    155
HSPS_METJA       TESERIIYSEIPEEEIYRTIKLPAT--VKEENASAKFEN-GVLSVILPKAESSIKKG--     143
Q970D9 SULTO     ----GVKYLTQ-R-PKYVRKVIRLPYNVAKDAEISGKYEN-GVLTIRIPIAGTSV-----     119
HSP26_YEAST      ESKDKVKVKES-SSGKFKRVITLPDYPGVDADNIKADYAN-GVLTLTVPKLKPQKDGKNH     200
HS16B_WHEAT      DKNDKWHRVER-SSGKFVRRFRLLEDA--KVEEVKAGLEN-GVLTVTVPKAEVKKPEVKA     146
Q8PNC2_XANAC     TETERFSRIER-RYGSFHRRFALPDSA--DADGITAAGRN-GVLEIRIPKRPAATPRR--     143
                          *           .           : * *  .

HSPB1_HUMAN      EIQTIPVTFESRAQLGGPEAAKSDETAAK        205
HSPB6_RAT        --LPSPPAAK----------------        162
CRYAB_HUMAN      ERTIPITREEKPAVTAAPKK--------        175
CRYAA_BOVINE     ERAIPVSREEKPSSAPSS----------        173
Q8UUZ6 ZEBRAFISH DRTIPVTREDKSNSGSSS----------        173
HSPS_METJA       --INIE--------------------        147
Q970D9 SULTO     --IKI---------------------        123
HSP26_YEAST      VKKIEVSSQESW--GN------------        214
HS16B_WHEAT      --IQISG-------------------        151
Q8PNC2_XANAC     --IQV-G----NGQN-TS--GSTVQ--        158
```

Figure 1.13: Sequence Alignment of Structurally Resolved sHSPs. Residues comprising β-strands are in colored. The ACD comprises β2– β9 strands. It is marked by the line. The conserved I/L–X–I/L motif in the C-terminal extension is inside the box. Abbreviations for organism names: Human HSPB1 and HSPB6; Rat, *Rattus rattus;* Bovine, *Bos taurus* ; Zebrafish, *Danio rerio*; Metja, *Methanocaldococcus janaschii*; Sulto, *Sulfolobus tokodaii*; Yeast, *Saccharomyces.cerevisiae*; Wheat, *Triticum aestivum*; Xanac, *Xanthomonas axonopodis*. Alignment was generated using ClustalW, and then the C-terminal region was adjusted to align the I/L–X–I/L motif for some sequences.

The flagship of sHSPs is the ACD, whose topology is conserved with some minor differences. The human ACD contains seven β-strands forming the β-sheet, while the invertebrate sHSPs have eight such strands. The difference is as a result of the human's β-6 and β-7 coming together and forming a single strand called β6+7, which forms an interface for dimerization (Laganowsky *et al.* 2010; Bagneris *et al.* 2009; Clark *et al.* 2011; Baranova *et al.* 2011; Jehle *et al.* 2010). On the other hand, the invertebrates have two distinct β-strands, and their dimerization mode is different. Instead of a dimer interface, the monomers are held together by the interaction of β-6 strand of one monomer mainly with β-2 and β-3 of the other monomer. There are also other interactions involved, but these are the best established to date (Kim *et al.* 1998; van Montfort *et al.* 2001, Takeda *et al.* 2011; Basha *et al.* 2011; Mchaurab *et al.* 2009). A whole list of interactions was published by Kim *et al.* (1998) for the MjHsp16.5 protein, and they can be extended to others as well. Representative structures for the two modes of dimerization are shown in figure 1.14.

There are some small differences apparent from figure 1.14. One of them is that MjHsp16.5 contains an additional β-strand, called β-1. It is still a matter of debate whether this strand is part of the N-terminus or part of α-Crystallin (Kim *et al.*, 1998, Haslbeck *et al.* 2008, Machourab *et al.* 2008, Bertz *et al.* 2010, Mchourab *et al.* 2012). Since MjHsp16.5 was the first resolved structure, and in the latter resolved sHSPs structures no β1 strand was found, the conventional naming of the strands starts from β2 (Horwitz, 2009, Machourab *et al.* 2009). Another difference is that StHsp14.0 has in one monomer a short 3-residue β10 strand, and also a short other strand between β7 and β8. It is though that the later is an artifact of experimental conditions, while for the former there are indications that it is be real due to secondary structure predictions (Takeda *et al.*, 2011; Hanazono *et al.* 2012).

The human αB-Crystalline structure does not have a β2 strand according to the current experimental data even though a β-strand secondary structure is predicted (Jehle *et al.* 2011; Braun *et al.* 2011). The reason for not being able to resolve it is the high degrees of freedom of the residues in the region.
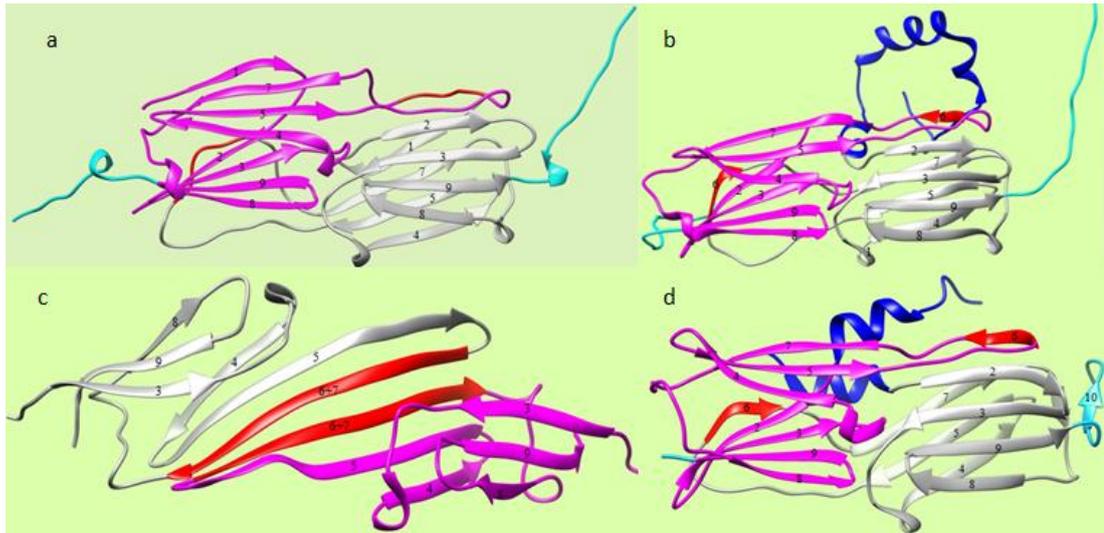
Figure 1.14: The Two Main Modes of sHSP Dimerization. (a) *M.janaschii* MjHsp16.5 dimer, PDB 1SHS (Kim *et al.*, 1998). (b) *T.aestivum* WtHsp16.9, PDB 1GME (van Montfort *et al.,* 2001). (c) Human αB-Crystallin, PDB 2KLR (Jehle *et al.*, 2010). (d) *S.tokodaii* StHsp14.0, PDB 3AAB (Takeda *et al.*, 2011). N-terminal region when present is colored dark blue. The C-terminal region is colored cyan. α-Cryltallin is colored magneta and dark grey for each monomer. The β-6 strand though it is part of α-Cryltallin it is colored red for clarity. Note that is not shown in (a) due to the structure being unrefined in PDB, though it is stated in the original article (Kim *et al.*, 1998; Quinlan *et al.* 2013), while in (c) it is β6+7. All the images were generated and modified by UCSF Chimera.

Recently Jehle *et al.* (2011) and used structural data together with modeling to determine the secondary structure and even the 3D structure of the N-terminal region of human αB-Cryltallin. They showed that it is composed of two α-helices and two β-strands. Up to date, two modes of dimerization have been observed in the resolved structures of sHSPs, and based on this, sHSPs are divided into two classes. Class I includes the mammals, specifically humans (Laganowsky *et al.* 2010a; Jehle *et al.* 2010), bovine (Laganowsky *et al.* 2010b), rat (Bagneris *et al.* 2009), and also zebrafish which is a Eukaryote (Laganowsky *et al.* 2010b). As stated above, they have an extended β-strand including β6 and β7, and so it is called β6+7. These extended β6+7 strands run in antiparallel directions in the two monomers. There are two important salt bridges formed between residues that hold the monomers together in human αB-Crystallin between R120-D109 and D80-R107 from each monomer

(Jehle *et al.* 2010). Since the orientation is antiparallel, four such salt bridges are found in each dimer interface. These interactions are conserved among mammalain sHSPs. These proteins are polydisperse in nature and for this reason there are no full 3D structures of them. Despite this, Mass Spectrometry and Cryo-Electron Microscopy (Cryo-EM) have shown many structural details and possible action mechanisms (McDonald *et al.* 2012; Stengel *et al.* 2010; Jaya *et al.* 2009; Benesch *et al.* 2010), as will be explained in the coming sections.

Class II includes the archaeal (Kim *et al.* 1998; Takeda *et al.* 2011; Hanazono *et al.* 2012), wheat (van Montfort *et al.* 2001), and yeast sHSPs, though the last one's structure is not solved neither by X-Ray nor by NMR, but it is relayed on structure prediction and Cryo-EM data (White *et al.* 2006; Benesch *et al.* 2010). Their dimmerization, as shown in figure 1.14 is different. It occurs mainly through swapping of the β6 strand of one monomer to an interface created mostly by β2 and β3 of the other monomer. There are also additional interactions detected by NMR for the WtHsp16.9 (van Montfort *et al.*, 2001) like the salt bridge between E100 and R108, which are the counterparts of Human R120 and D109, though the studies on this area are not complete. This is thought to be the main contributor of the regular monodisperse structure for these sHSPs (Haslbeck *et al.* 2008), making it possible for them to be crystallized and resolved in multimeric forms. Another support for it are the Cryo-EM images obtained when these sHSPs are alone or with substrates, and in both cases they are observed as round regular structures (Haslbeck *et al.* 1999; Shi *et al.* 2006; Haslbeck *et al.* 2008; Benesch *et al.* 2010).

1.3.5. Packing of the Dimmers to Form Higher Oligomeric Assemblies

The only assemblies which have been studied by high-resolution methods are the ones for monodisperse, class II sHSPs. For the other class no high-resolution, multimeric assembly structures exist to date except for EM and Cryo-EM images (Lee *et al.* 1997; Stengel *et al.* 2010; Jehle *et al.* 2009). The two best understood monodisperse structures are MjHsp16.5 (Kim *et al.*, 1998) and WtHsp16.9 (van Montfort *et al.*, 2001), which despite similarities have also differences. They have in

common the interaction of a conserved motif at the edge of C-terminal tail made of I/L-X-I/L residues, which interact with a groove formed by β4 and β8 strands at the edge of α-Cryltallin of the adjacent dimer. This motif is conserved in 95% of sHSPs studied to date. What is thought to form different assemblies is the angle of a hinge linking the C-terminus to α-Cryltallin domain. A 30° angle difference is thought to make MjHsp16.5 form 24-mer spherical oligomers while WtHsp16.9 forms two hexameric discs to assemble in a final dodecameric structure (Kim *et al.*, 1998; van Montfort *et al.* 2001). Another contributor to the oligomeric assembly is the N-terminus, but not in all of them. In WtHsp16.9 this region seems to be essential, since the two hexameric discs are knotted to each other through this particular region. In MjHsp16.5 deletion of this region does not change the size of oligomers. This shows that each class of sHSPs have their own specific oligomerization properties (Bertz *et al.* 2010; Mchaurab *et al.* 2010; Basha *et al.* 2011).

As for the polydisperse sHSPs, there are no full high-resolution structures (Basha *et al.* 2011). By using a combination of methods such as NMR, EM, SAXS and Molecular Modeling they have started to elucidate important interactions and build plausible models of how these proteins come together and function (Jehle *et al.* 2011; Braun *et al.* 2011). Jehle *et al.* (2010) described the first such model of αB-Csystallin. By using homology modeling and data from Solid-State NMR they determined the missing 65 N-terminal residues from the high resolution structure. The proposed secondary structure is made up of two α-helixes and two antiparallel β-strands. Further investigation showed that three dimers come together to form a hexamer, through interactions of the C-terminus with the groove formed by β4/8 of one of the monomers in the adjacent dimer. Than four of such triangular hexamers were fitted into an αB-Crystallin EM density map, forming four three-fold axis and three two-fold axis (Figure 1.15.a). Then they carried out further NMR studies to determine the distance between certain residues in the N-terminal region so that their 3D arrangement could be decided. After that the predicted structure of 65 residues was fitted in the EM density map together with the rest of the resolved αB-Crystallin (Figure 1.15.b).
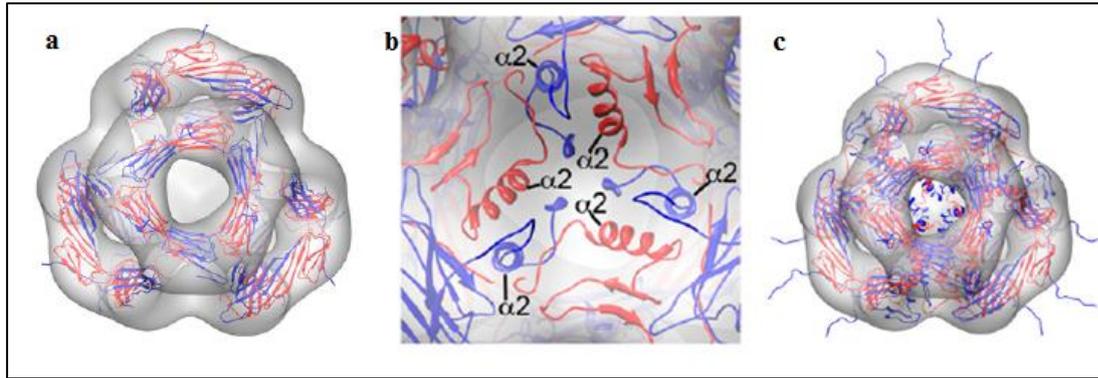
Figure 1.15: The Construction of Polydisperse Human αB-Crystallin. (a) An EM map and the dimmers resolved by NMR fit inside the shell. (b) Addition of modeled 1-65 residues to the N-terminus of the structure and filling of the empty cavity in the middle of the shell. (c) Addittion of C-terminal residues with half of them being oriented to the outside. (Jehle at al. 2011)

The Cryo-EM is an empty shell, and this empty region was filled by inserting the predicted region of the first 65 residues. This is in agreement with other Cryo-EM models from other sHSP proteins, but due to the hight flexibility of the α-chains at the N-terminus, no such density could be obtained for the human sHSPs (Peschek *et al.* 2009). Based on mutagenesis studies, it was found that αB-Crystallins lacking residues 1-68 fail to form oligomers, while those lacking the first 35 residues form large oligomers comparable to the wild-type protein (Augusteyn, 1998; Kundu *et al.* 2004). This led the authors conclude that the second half of the N-terminus is the essential region for oligomerization (Jehle *et al.* 2011). Further investigation of the C-terminal residues showed that as a result of the different orientation of the ACD which is reflected to the C-terminus, half of the monomeric subunits have C-terminus facing the inside of the shell, and the other half facing the outside of it (Figure 1.15.c). Moreover, residues 1-44 are very flexible, showing that they are highly dynamic due to the exchange of the subunits among the multimeric ensembles. Then they increased the number of multimeric subunits by fitting dimmers in the "windows" formed from the hexamers, forming higher oligomeric structures.

This model was the first study in such details of a polydisperse sHSP and it found some very important interactions involved in every level of sHSPs multimerization. Thus, ACD are responsible for dimer formation, the C-terminus is responsible for defining the hexameric units, which are the base to form higher oligomers, which in turn interact together by second half of the N-terminus. In a later study by Mchaourab et al. (2012) proposed that the N-terminus may actually be a determinant of polydispesity. They base their hypothesis on experiments where the middle domain of the N-terminus from Hsp27, a human polydisperse sHSP, was transferred to the N-terminus of the monodisperse MjHsp16.5 and caused this protein to form oligomers of 48 subunits.

1.3.6. Chaperone Activity Assays

The most widely used assay for chaperone activity is substrate protection from aggregation measured by light scattering. In this assay, the substrate together with sHSPs is heated at a temperature it normally aggregates, and in this way the protection efficiency of sHSPs is measured. They have found that different ratios of substrate/sHSP are sufficient to prevent aggregation, and this depends on the substrate. The suppression of light scattering reflects the competition of the substrate with itself resulting in aggregation, and the chaperone which will protect it from aggregating (Das et al. 1999; Ehrnsperger et al. 1997; Lee et al. 1997; Stromer et al. 2003; Shashidharamurthy et al. 2005; Haslbeck et al. 2008; Basha et al. 2011). What has turned out to be the rule is that the larger the molecular weight of the substrate, the higher molar ratio of sHSPs is needed to prevent aggregation (Stromer et al. 2003; Hasleck et al. 2008; Machaurab et al. 2009; Basha et al. 2011). Another method is Size-Exclussion Chromatography (SEC), where the oligomers size can be observed, either alone or with substrate. It is used widely to measure the size and distribution of oligomers depending on the substrate/sHSP molar ratio, or on temperature. The regularity observed is that complexes become smaller as the ratio of sHSPs to substrate increases, which makes sense because higher sHSP concentration practically would prevent substrate aggregation (Lee et al. 1997;

Berengian *et al.* 1999; Haslbeck *et al.* 1999; Giese *et al.* 2002; Stromer *et al.* 2003; Shi *et al.* 2006; Cheng *et al.* 2008; Jaya *et al.* 2009; Takeda *et al.* 2011). A third method is substrate activity protection assay. Here, the substrate's activity is measured at higher temperatures than its optimal, generally at denaturing temperatures together with sHSPs at different molar ratios. Then enzyme activity is measured in the classical way, and in this manner, the protection efficiency of sHSPs is measured (Kocabiyik *et al.* 2012; Quinlan *et al.* 2013). This method is the most effective one to measure the chaperone activity, but is not very widespread due to the difficulty and the stringent conditions to carry out enzyme assay.

1.3.7. Chaperone Action Mechanisms of sHSPs.

Although the importance of sHSPs in proteostasis is well established, their exact role and action mechanism are still unknown. One model proposes that under extreme stress conditions, sHSPs provide an energy-independent mechanism to buffer the unfolding proteins (Koteiche *et al.* 2006). This energy-independence is crucial since during stress the energy reserves of the cell are generally depleted (Nadal *et al.* 2011). So they act as reservoirs for non-native proteins to be refolded back to their native state by the other ATP-dependent chaperones or degraded to avoid their toxic effects (Koteiche *et al.* 2006; Hasslbeck *et al.* 2008; Mchaourab *et al.* 2009; Basha *et al.* 2011).

The mechanism by which sHSPs interact with the substrate has been a long dispute, and two models are proposed for it: the model that proposes sHSP oligomer dissociation during stress, here called "Dissociation Model", and the other model proposes that sHSP oligomer binds to the substrate without dissociation, here called "Non-Dissociation Model".

The recent research with Mass Spectrometry and FRET seem to favor the dissociation model (Giese *et al.* 2002; Giese *et al.* 2004; Shashidharamurthy *et al.* 2005; Mchaourab *et al.* 2009; Stengel *et al.* 2009; Jehle *et al.* 2010; Benesch *et al.* 2010). According to it, the original oligomeric structure, for example 12-mer for wheat and 24-mer for *M.janaschii* serve as reserves of sHSPs and at the same time

protects them from unwanted interactions with native proteins. When temperature is increased, or a protein is destabilized by stress or mutations, sHSPs will be activated. In the case of high temperature, chaperone activity can be explained by thermo-sensitivity of sHSPs and as a result their dissociation to bind the non-native proteins and re-associate back into big multimeric units. However, in the case of a substrate protein destabilized by mutation, the way in which sHSPs are activated is not known (Shashidharamurthy *et al.* 2005; Koteiche *et al.* 2006; Mchaourab *et al.* 2009).

To explain this, two mechanisms are proposed: One mechanism proposes that that the "windows" in the shell of the oligomer, or the C-termini facing the outside of the shell (figure 1.15c) would leave open hydrophobic pockets, especially the $\beta4/\beta8$ groove, which in turn will be used by the substrate to bind and induce sHSPs activation and subsequent dissociation (Jehle *et al.* 2011; Braun *et al.* 2011). The second mechanism proposes that in normal physiological conditions, sHSPs are not all in a regular multimeric form, but rather they are in a dynamic equilibrium where multimers of different subunits associate and dissociate in a continuous manner. These free units serve as sensors of destabilized proteins, bind them, and then associate into big oligomeric subunit with the substrate to protect them from aggregation. Then by a concentration-dependent mechanism, the majority "reserve" sHSPs will dissociate and bind more of the destabilized substrate proteins to protect them from aggregation, shifting the equilibrium towards dissociation of unbound chaperones and then re-association with the substrate (Giese *et al.* 2004; Shashidharamurthy *et al.* 2005) (Figure 1.16). The first model is based on results from Size-Exclussion Chromatography (SEC), where a single peak is observed, however this peak is very broad, suggesting a range of oligomeric species are present in the cell (Lee *et al.* 1997; Horwitz, 2003; Stromer *et al.* 2003; Kappe *et al.* 2004; Cheng *et al.* 2008). The second mechanism is based on results from fluorescence label experiment that monitor the subunit exchange of oligomers during physiological and stress temperature.
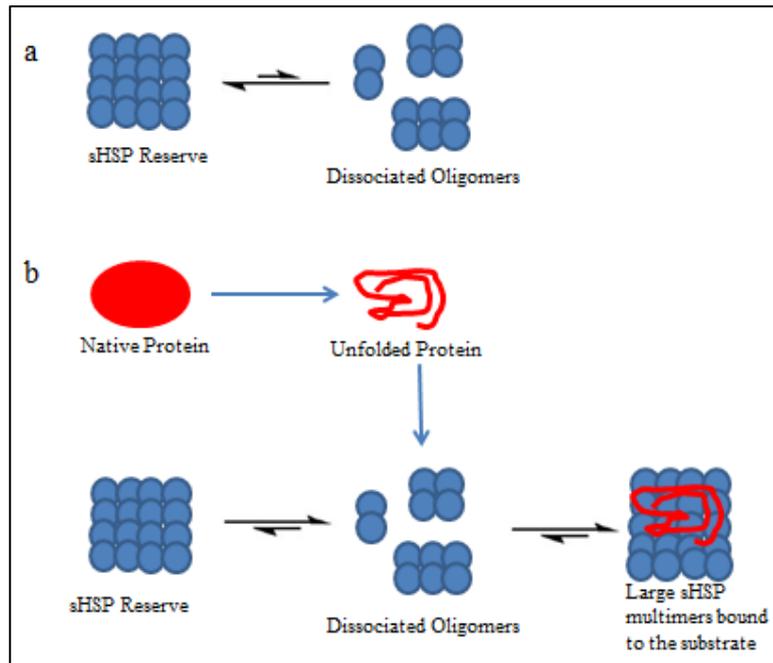
Figure 1.16: Proposed Mechanism of sHSPs Binding to Their Substrates. (a) sHSPs during normal conditions. There is a sHSP reserve in high oligomeric form, and a small percentage of it is dissociated into smaller oligomers, starting from dimers and more. (b) During stress, the equilibrium shifts to the right, with the "reserve" sHSPs dissociating, and then associating with the unfolded substrate to form large sHSP-Substrate complexes.

Experimental evidence suggests that the exchange speed increases when temperature is increased, but exchange continues in a steady state even in physiological conditions (Franzman *et al.* 2005; Benesch *et al.* 2010). Another support for the second model is from experimetns performed with Mass Spectrometry, where a range of subunits are detected in the cell and quantified. Results suggested that such species are present in the cell, but their low concentration and the low resolution of mass spectrometry prevent their quantification. They are proposed to be the sensors of destabilized proteins in the cell (Benesch *et al.* 2010; Stengel *et al.* 2009; Jehle *et al.* 2011). Research by Ghosh *et al.* (2005) based on computational studies and pin array assays to determine different regions of the human αB-Crystalline involved in interactions with different substrate proteins also supports the second model. According to their data, different regions of the sHSP protein interact with different

substrates with different affinities, explaining at the same time the differential expression of sHSPs in different organs in higher eukaryotes (Ghosh *et al.* 2005; Ghosh *et al.* 2006). This substrate preference for buried regions of the αB-Crystalline is difficult to be reconciled with the multimer binding the destabilized substrate without exposing buried hydrophobic regions of the chaperone. Dissociation of the sHSP multimer is required to expose these substrate binding sites. Whatever the substrate-binding mechanism, the sHSPs after binding the substrate will try to keep it from aggregating, and sent it for refolding to the chaperones that use ATP, especially Hsp40/70 system. If these systems cannot refold the misfolded protein, then it is sent to the degradation pathway.

The non-dissociation model was proposed by Franzmann *at al*. (2005) for yeast ScHSP26. They found that this sHSP showed chaperone activity even after chemical cross-linking, meaning that it is not necessary for the sHSP to dissociate for binding and protecting the substrate. Further investigation revealed a very different mechanism from the dissociation model discussed above: Yeast Hsp26 cannot act as a chaperone unless it is activated by heat, and moreover they found a region at the N-terminus that served as a thermosensor. The activity of this sHSP is dependent on global changes in the secondary and tertiary structure of the protein, while its oligomeric assembly is left intact. In this protein, subunit exchange is independent of the chaperone activity, and even though it is continuously occurring in the cell, unless the cells are exposed to heat treatment, there is no interaction with the substrate, even in cases when it is heat-denatured before incubated with the Hsp26 (Franzman *et al*. 2008). However, the latest reports by Benesch *at al*. (2010) based on high resolution Mass Spectrometry (MS) cast some doubts on it. They used MS to quantify the oligomeric distribution of yeast HSP26, and found that at 25°C the main assembly is the 24-mer, while with the increasing of the temperature higher dispersed oligomers are formed, the majority of them being 40-mers. This would imply temperature-dependent dissociation of the reserve 24-mer and re-association into larger 40-mers together with the substrate (Benesch *et al*. 2010). But still a distinction of ScHsp26 is its inability to bind heat-denatured proteins unless heat-

treated, making it unfeasible to bind and protect proteins destabilized by mutations, because its thermosensor domain in the N-terminus is activated only by heat (Franzman *et al.* 2008). The way by which yeast deals with proteins destabilized by mutation is thought to be by using the second sHSP, Hsp42, which is not characterized yet (Benesch *et al.* 2010). This new mechanism shows once more the structural and mechanistic versatility of sHSPs and supports the idea that there cannot be a unifying principle for all of them.

Researchers have also tried to figure out the kinetics of sHSP-substrate interactions, and to date many mechanisms were proposed, but the most plausible seems to be the Minimalist Model proposed by Koteiche *et al.* (2006). Before explaining it, it is worth mentioning the methodology. They did not do the experiments in high temperatures because such extreme *in vitro* conditions may compromise the function of sHSPs under study, and also create a pool of substrate misfolded states, making a precise mechanism impossible. In other words, when a substrate protein is heated, there will be a whole spectrum of misfolded states, and the results obtained will be only the average of all of them (Koteiche *et al.* 2006). Instead they used mutant variants of the T4 lysoszyme (T4L) that had been studied before for their stability and whose structures were determined by x-ray crystallography. The variants used had different internal stabilities, meaning some were more stable than the others at a specific temperature, but their 3D structures were the same (Mathews, 1995; Mathews 1996). Based on this, the logic of the experiments was that the less stable T4L variant will bind with a higher affinity to sHSP proteins, and this binding energy could be calculated. Then based on it, a mechanism could be proposed. The whole model is based on three equations, which take into account both the substrate and the chaperone:

(1) $N \rightleftharpoons I_1 \rightleftharpoons \ \ldots\ldots \ \rightleftharpoons I_i \rightleftharpoons U$

(2) $sHSP \rightleftharpoons (sHSP)_a$

(3) $(sHSP)_a \rightleftharpoons \ {}^{I1}_{U} \ \rightleftharpoons \ {}^{C1}_{Ci}$
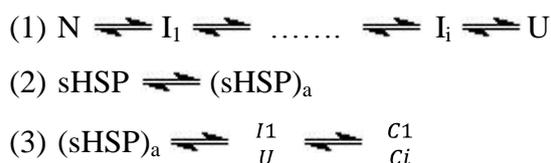
Equation (1) describes the states of substrate, from native (N), to different levels of

partial unfoding ($I_1 \rightarrow I_i$), to the unfolded state (U). The second equation describes the state of sHSPs, from their inactive condition (sHSP) to high affinity and/or high capacity active state (sHSP)$_a$. The third equation stands for sHSP and unfolded states of the substrate (Koteiche *et al.* 2006). As explained above for the mechanisms of sHSP-substrate interaction, experimental data have shown that the sHSPs of some organisms dissociate into smaller oligomers or even mono- or dimers, while those of some other organisms are proposed to stay in their complex structure without dissociating when they bind the misfolded or denatured substrates (Franzman *et al.* 2005; Haslbeck *et al.* 2008; Franzman *et al.* 2008; Mchaourab *et al.* 2009; Stengel *et al.* 2009; Benesch *et al.* 2010; Bertz *et al.* 2010; McDonald *et al.* 2012). The importance of the Minimalist Model is that it can explain both of the cases, because according to it the sHSPs in the multimeric form do not need to dissociate, but only be activated. This activation can occur by exposing their substrate binding domains in the environment. As a result, equations (1) and (2) are coupled together, where the activated sHSPs bind to partially unfolded or completely unfolded substartes. Under steady-state conditions, binding reflects an energy preference of the unfolded substrate to interact with chaperones and form different substrate-sHSP complexes (C1-Ci). The importance of this mechanism, in contrast to the ones proposed in experiments where the chaperone is first heat-activated and then its substrate protection is measured by light scattering (Franzman *et al.* 2008) lies in the fact that for a substrate to bind to sHSPs, it is enough for it to be destabilized even by mutations, which is mostly the case *in vivo* and disease conditions, and the high temperature factor to activate the chaperone does not have to be present (Koteiche *et al.* 2006; Mchaourab *et al.* 2009).

1.3.8. Structural Rearrangements of the Bound Substrate

Another research area of the sHSP-substrate interaction focuses on the manner the substrate binds to the sHSP oligomers and is protected from aggregation. This was studied by focusing on the structural rearrangement that the substrate undergoes when it is bound to sHSPs. There are reports showing that a whole range of substrate states from destabilized to less misfolded to molten globule bind to sHSPs (Carver *et*

al. 2002; Stromer *et al.* 2003; Claxton *et al.* 2008; Cheng *et al.* 2008). The question is whether a model can be proposed for the manner of sHSP oligomer-substrate interaction. There is no agreement in the results of experiments that followed different approaches. Based on spin-label experiments and secondary-structure studies, it was found that α-Lactalbumin has extensive loss in the secondary structure suggesting extensive unfolding of the protein when it was bound to α-Crystallin (Bettelheim *et al.* 1999; Carver *et al.* 2002). In contrast, the same experimental approach with β- and γ-Crystallins and rhodanese substarates showed them to retain most of the secondary structure when bound to α-Crystallin (Das *et al.* 1996; Das *et al.* 1999). In another experiment performed by monitoring substrate Hydrogen-Deuterium exchange for Malate Dehydrogenase, it was found that this exchange was limited. It means that the substrate when bound to TaHsp16.9 (Wheat) and PsHsp18.1 (*P.sativum*) is either partially folded or some of its regions are protected in an unfolded way inside the chaperone complex, making them inaccessible to deuterium in solution (Cheng *et al.* 2008). The most accurate studies in this area seem to be those of Claxton *et al.* (2008) who used the unstable variant of T4 Lysozyme (T4L) mentioned in the section above together with spin-labeling and energy calculations to determine the level of substrate unfolding when bound to sHSP. Energy calculations based on thermodynamic studies showed two energetically different populations of the substrate, suggesting that either the T4L binds the sHSP complex in two structurally different forms, or it is bound in two thermodynamically different environments. To find out which of these possible interactions is true, spin label was used to measure distances between certain residues. It was found that the distances of specific residues between domains and those at the active sites had increased. At the same time, an increase of the characteristic distance of the residues in the α-helix was observed, suggesting a loss in the substrate native structure. Further investigation showed that C-terminus of T4L which is hydrophobic was buried in the low-accessibility region of the chaperone complex, while the N-terminal region is exposed to the environment and has high-backbone dynamics. These results suggest that rather than being two different

conformations of T4L, the two energetically different populations are associated with topologically different binding sites in the sHSP complex (Claxton *et al.* 2008).

Based on the results of the experiments above, two models of sHSP-substrate interactions were proposed: the Low Affinity and the High Affinity models (Claxton *et al.* 2008). According to the former, the substrate binds on the surface of the sHSP complex in specific substrate-binding sites, especially the β4/8 groove where the L/I-X-L/I motif binds. According to the later, the substrate is buried deep inside the sHSP oligomer. This binding may be substrate-specific, as shown above, and as a result unfolding was observed, or condition specific, with less misfolded substrate binding in the low-affinity mode, and the more misfolded one binding in the high-affinity mode (Figure 1.17.a). Cryo-EM studies also support the proposal of two substrate binding environments because they have observed two main oligomeric states of the sHSP complexes (White *et al.* 2006; Quinlan *et al.* 2013).
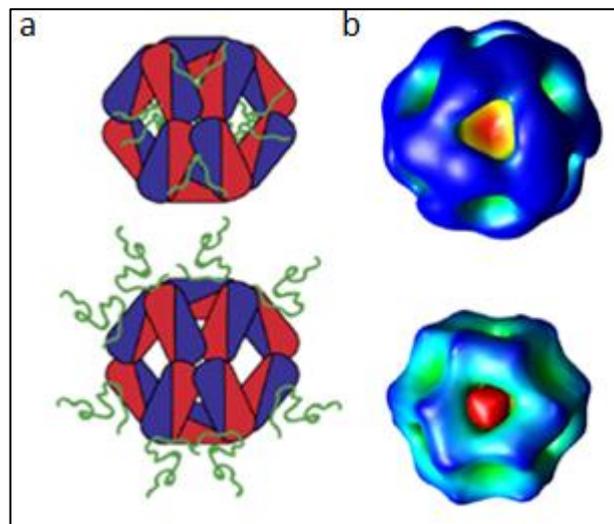
Figure 1.17: The Substrate-sHSP Interaction Mode and sHSPs Oligomers. (a): the two proposed binding modes are shown, the upper figure for high affinity, and the lower lower figure for low affinity model. (b): the two structural models observed by Cryo-EM. The upper structure is the predominant form at stress temperature, while the lower structure is the predominant form at physiologic conditions. ((a) Claxton at al. 2008; (b) Haslbeck at al. 2008).

According to the studies of *M.janaschii* Hsp16.5, *A.fulgidus* Hsp20.2 (Haslbeck *et al.* 2008) and Yeast Hsp26 (White *et al.* 2006), there are two different complexes of the sHSPs: One is a complex of smaller diameter but with protein density at the center, proposed to be the N-terminus of the sHSPs, while the second complex is bigger in size with no protein density at the center, but rather an empty shell (Figure 1.17.b).

However data from Quinlan *et al.*, (2013) contradict the findings of Haslbeck *et al.* (2008) on the difference in particle size for the two different complexes of *M.janaschii* Hsp16.5, finding them equal in size, but proving that one model has a protein density at the center while the other is empty. Haslbeck *at al.* (2008) proposed that the two multimeric forms of sHSPs observed by Cryo-EM represent two different populations of the chaperone during physiological and stress conditions, based on quantification studies of the two structures in physiological and stress conditions, but the results are not conclusive (Haslbeck *et al.* 2008).

Regarding the sites of substrate binding to the molecular chaperone oligomer, there is no consensus to date. The first approach to find such sites was used with ANS, and latter its derivative Bis-ANS and UV-Crosslinking, first in GroEL/GroES system and then in sHSPs (Das *et al.* 1996) Both of them are hydrophobic probes that bind to hydrophobic patches in the proteins and as a result their emission spectra changes. This change can be monitored by fluorescence spectroscopy. Making use of this fact, Seale *et al.* (1995) used Bis-ANS and then did UV-Crosslinking experiments to form a covalent bond of the probe to the residues it had bound when sHSPs were incubated at high temperature. Then the proteins were sequenced and the binding sites found. Latter chemical cross-linking experiments of sHSPs and substrate proteins performed by other researchers produced different sites, especially in α-Crystallin domain and the N-terminus (Ahrman *et al.* 2007; Ghosh et al; 2006; Jaya *et al.* 2009). A consensus seems to be formed about the N-Terminus based on site-directed mutagenesis and deletion. When the complete N-terminal region is deleted in yeast Hsp26, the chaperone activity is abolished altogether, but when shorter patches of it are removed, the chaperone activity falls accordingly (Haslbeck *et al.* 1999; Stromer *et al.* 2003). In other studies of *M.janaschii* Hsp16.5, when the N-

terminal region is deleted, the chaperone activity falls accordingly with the region removed, but when it is removed in its entirety, still there is chaperone activity, in contrast to yeast Hsp26 (Usui *et al.* 2004). Other studies with human α-crystallins have found similar results: When the N-terminus is extended, chaperone activity increases, and when it is deleted, decreases (Shashidharamurthy *et al.* 2005). In yet another study they have found a correlation of chaperone activity and the hydrophobicity of the N-terminus introduced by site-directed mutagenesis, giving further proof that this region is essential for binding the substrates (McDonald *et al.* 2012). There are also reports suggesting that the β4/8 groove where the C-terminal tail I/L-X-I/L motif binds, is also an important substrate-binding site. A competition mechanism is proposed between the substrate and the I/L-X-I/L motif which causes the oligomer dissociation and as a result its binding to the unfolded proteins (Pasta *et al.* 2004; Jehle *et al.* 2010; Takeda *et al.* 2011; Quinlan *et al.* 2013). If everything is put into perspective, the emerging picture is that there is not a single substrate-binding site. This is supported by the experiments of Ghosh *et al.* (2005) as explained above, where they found different regions of the α-Crystallin interacting with different substrate proteins. Even though to date the N-terminus and α-Crystallin seem to be the major substrate binding sites, for each molecular chaperone of different organisms there are differences and there doesn't seem to be a single substrate binding site (Basha *et al.* 2011).

1.3.9. The Importance of sHSPs in Diseases

Small HSPs are indispensable for proper cellular function (Bukau *et al.* 2006; Hartl *et al.* 2011; Kim *et al.* 2013). Many cellular proteins are targets of sHSPs for their proper folding and function, including cytoskeleton, amyloid fibrils, growth factors and steroid receptors to mention a few (Basha *et al.* 2011). sHSPs are involved in a number of diseases. The best and most famous example is lens cataract, where mutations in the α-Crystallins or post-translation modifications lead to lens crystalline and cytoskeleton aggregation, destroying lens opacity and causing cataract (Clark *et al.* 2000; Graw, 2009; Horwitz, 2009; Clark *et al.* 2012). Hsp27 and αB-Crystalline are associated with many neurodegenerative diseases in human

Central Nervous System. These include Parkinson, Alzheimer, Familial Amyloidotic disease, Creutzfeldt-jacob, Amyloid Laterla Schlerosis, Charcot-Marie-Tooth, and Alexander Diseases (Irobi *et al.* 2004; Selkoe *et al.* 2004; Simon *et al.* 2007; Muchowsky *et al.* 2005; Chaundry *et al.* 2006; Brown *et al.* 2007; Wang *et al.* 2007; Mymrikov *et al.* 2011).

Small HSPs are also involved in cancer, whose role is thought to be mostly protective against chemotherapy and also against hypoxic stress during the growth of tumor cells (Kamada *et al.* 2007). Hsp27 was found to be an important regulator of cellular proliferation and apoptosis, making it a strong candidate for carcinogenesis. Its expression is elevated in a number of tumors like astrocytic brain tumors (Assimakopoulou *et al.* 1997) and Breast cancer (Ciocca *et al.* 1992). Similar protective results were reported for Hsp22, whose overexpression in Stomach tumors, and proliferating keratinocytes, as well as rat pheochromocytoma (PC12) cells protected them against chemotherapy damage and its overexpression is correlated with aggressive tumorogenesis (Gober *et al.* 2005), however, in Melanoma, Prostate cancer and Ewing Sarcoma, overexpression of this sHSP induced apoptosis, and as a result protection from tumorogenesis (Gober *et al.* 2003). Similarly, α-Crystallin acted as a negative regulator of tumor in pancreatic cancer (Deng *et al.* 2010). These results indicate that different classes of sHSPs have different effects from each other, and even the same sHSP class has different effects in different tissues, showing that the effect of sHSPs in tumorogenesis depend largely on the environment they are found (Mymrikov *et al.* 2011).

Small HSPs are also thought to be involved in diabetes as a result of Hyperglycemic stress, but the exact link is not found yet. In a recent study in diabetic rat, an upregualtion of αA-, αB-Crystallins and Hsp22 were found in the model organism's retina, and also phosphorylation of specific residues linked to chaperone activation (Reddy *et al.* 2013). Last but not least is the protective role of sHSPs in cardiac tissue, especially after ischemic stress. When the hearts of Hsp20 Transgenic mice are put under ischemic stress *ex vivo*, they show decreased myocyte cell death, smaller infarct size, and improved contraction compared to the hearts of wild-type

organisms. This appears to be the result of the important role Hsp20 has in autophagy activation, which is critical in dealing with ischemic injury (Fan *et al.* 2005; Qian *et al.* 2009). Similar results were obtained for Hsp22 and αb-Crystalline transgenic mice (Ray *et al.* 2001; Depre *et al.* 2006; Sui *et al.* 2009). When mice models lacking αB-Crystalline and Hsp27 were challenged with ischemic stress, they showed 2-fold decrease of contractile recovery, and at the same time an increase in apoptosis and necrosis (Ray *et al.* 2001). However they develop normally under non-stress conditions, indicating that αB-Crystalline and Hsp27 are not essential for heart development, but rather, they are essential for heart protection during stress (Morrison *et al.* 2004).

1.4.Aim of the Study

The aim of this study was to carry out site-directed mutagenesis and characterize the Archeal small heat shock protein Hsp14.3 from *Thermoplasma volcanium.*

Firstly, bioinformatics tools were used to characterize the primary structure of Tpv-Hsp14.3 protein by comparative sequence analysis and alignment. The protein's secondary and three-dimensional structures were also predicted. After that, based on structure comparison with other sHSPs whose structures were resolved by NMR and X-Ray Crystallography, certain amino acid residues were selected for mutation.

The positions, their counterparts in orthologs, and possible interactions they may be involved in were deduced by superimposition of the predicted models of Tpv-Hsp14.3 with resolved 3D structures of different sHSPs.

Then site-directed mutagenesis studies were carried out to make amino acid substitutions at the target sites. The changes introduced by mutagenesis were later confirmed by DNA sequencing.

This was followed by expression of each protein variant in *E.coli* and purification.

Finally, enzyme activity protection assay was carried out to study the chaperone function of each mutant sHSP as compared to the wild-type protein. Dynamic Light

Scattering spectroscopy was used to observe the alterations in oligomere dynamics at different temperatures.

CHAPTER 2

MATERIALS AND METHODS

2.1. Materials

2.1.1. Chemicals, Enzymes and Kits

All the chemicals used in these experiments were of molecular biology grade and highest purity possible.

Agarose gel, Ampicllin, Ammonium Persulfate, Ammonium chloride (NH_4Cl), Tetracycline, Ethylenediaminetetraacetic Acid (EDTA), Ethidium Bromide (Et-Br), Citrate Synthase (Pig Heart), Glutamate Dehydrogenase (Bovine liver) and Tetramethylethylenediamine (TEMED), purchased from Sigma Chemical Company, Spruce St. St. Louis, Missouri, USA.

Agar was purchased from Acumedia, Baltimore, USA.

Yeast extract was purchased from Difco, Detroit, USA.

Absolute ethanol was from Reidel de Häen, Germany.

TRIS, sodium chloride (NaCl), sodium dodecyl sulfate (SDS), β-mercaptaethanol, sodium hydroxide (NaOH), α-D-Glucose, sodium dihydrogenphosphate (NaH_2PO_4) magnesium sulfate ($MgSO_4$), potassium dihydrogen phosphate (KH_2PO_4), and tryptone were purchased from Merck, Darmstadt, Germany.

Primers used in RT-PCR experiments were synthesized by TIB MOLBIOL GmbH, Berlin, Germany.

Restriction endonucleases *SalI* and *PstI*, *HindIII, BamHI, SmaI* and their buffers were purchased from Fermentas UAB, Vilnuis, Lithuania.

MboII and *BamHI* were purchased from New England Biolabs Inc. US.

Wizard® Plus SV Minipreps DNA Purification Kit was purchased from Promega Coorperation, Madison, WI, USA.

QIAprep Spin Miniprep Kit and QIAexpressionist kit for protein purification kits were purchased from Qiagen, Hilden, Germany.

Transformer^{TM} Site-Directed Mutagenesis Kit was purchased from Clontech, Takara, Japan.

Stratagene QuikChange® Site-Directed Mutagenesis Kit was purchased from Agilent Technologies, USA.

2.1.2 Buffers and Solutions

The compositions of the buffers and solutions used in these experiments are given in Appendix A. All buffers and solutions preparations in the lab were done in ultrapure water and sterilized in autoclave (ALP Co. Ltd., Tokyo, Japan).

2.1.3 Molecular Size Marker and Ladders

The molecular size markers for DNA were λ DNA/ EcoRI+HindIII Marker; GeneRuler DNA Ladder Mix, Ready-to-Use (100-10,000 bp) and GeneRuler 50 bp DNA Ladder. Molecular size markers for protein was Ready-to-Use PageRuler Prestained Protein Ladder. All of them were purchased from Fermentas UAB, Vilnuis, Lithuania. They are shown in Appendix B.

2.2. Media and Strains

2.2.1. Bacterial Strains

The plasmid source for all purification, mutagenesis, transformation and sequencing experiments was the recombinant *E. coli* pQE-31/775 strain, previously constructed in our laboratory (Kocabiyik *et al*. 2012).

The bacterium strain *E. coli* BMH 71-18 *mut*S was used in mutagenesis experiments by Transformer[TM] Site-Directed Mutagenesis Kit. (Clontech Takata, Japan)

The bacterial strain *E.coli XL1*-Blue Supercompetent cells were used in mutagenesis experiments by QuikChange® Site-Directed Mutagenesis Kit. (Startagene Agilent Companies, Germany)

2.2.2. Cultures and Media

The recombinant *E.coli* pQE-31/775, and *E. coli* BMH 71-18 *mut*S cells were grown on LB agar medium (1.5%) with (100 μg/ml) ampicillin by incubating overnight at 37°C, and then kept at 4°C. The recombinant *E.coli XL1*-Blue Supercompetent cells were grown on LB agar (2%) initially, and then their stocks were maintained at 1.5% LB agar medium. The renewal of media was done regularly every month.

2.3. Experimental Methods

2.3.1. Plasmid Isolation from Recombinant *E.coli* pQE-31/775 cells

The recombinant *pQE-3/775* vector containing *Thermoplasma volcanium GSS1 tvn0775* gene was isolated from recombinant *E.coli* pQE-3/775 cells. For this purpose, cells were picked from the plate and transferred into bottles containing 10 ml LB broth with 100μg/ml Ampicillin antibiotic. They were incubated overnight by shaking at 207rpm and 37°C in Heidolph Unimax 1010 Shaking Incubator (Heidolph Instruments GmbH, Kelheim, Germany). The following day, cells were harvested by centrifugation for 15 minutes at 10.000g in an IEC Cliniclal Centrifuge (Damon/IEC Division, USA), and the pellet was used for plasmid isolation by two different kits.

2.3.1.1. Plasmid Isolation by Wizard® Plus SV Minipreps DNA Purification System Kit

To check the purity of the plasmid, Wizard® Plus SV Minipreps DNA Purification System Protocol was used. The cell pellet was re-suspended in 250μl of Cell Resuspension Solution and then vortexed until all of the pellet was re-suspended. After that, 250μl of Cell Lysis Solution was added and the tubes were inverted 4-6 times each. Then 10μl of Alkaline Protease Solution was added and the tubes were inverted another 4-6 times. After addition of 350μl of Neutralization Solution the tubes were again inverted 4-6 times. Then the tubes were centrifuged for 10 minutes in Eppendorph Centrifuge at 13000 rpm. The clear cell lysate was transferred in the spin column and was centrifuged at the maximum speed for 1 minute, so that the DNA would bind to the column. The flow-through was discarded and the column was reinserted into the collection tube. Then 750μl of Wash Solution was added. The spin column in the tube was centrifuged at top speed for 1 minute, the flowthrough discarded and the column was inserted into the same collection tube. Then 250μl of Wash Solution was added for a second time and it was centrifuged at the maximum speed for 2 minutes to wash the column thoroughly. Finally, the spin column was transferred into a sterile 1.5 ml microcentrifuge tube, and 100μl of Nuclease-Free Water was added to it. It was then centrifuged at maximum speed for 1 minute at room temperature. Finally, the column was discarded and the eluted DNA sample was stored at -20°C for further use.

2.3.1.2. Plasmid Isolation by QIAprep Spin Miniprep Kit

The plasmids isolated with this Kit were used for mutagenesis experiments and sequencing. The cell pellet was resuspended in 250μl P1 buffer, into which RNase and LyseBlue were added according to the Kit instructions, and then it was transferred in a microcentrifuge tube. 250μl of P2 buffer was added, and tubes were mixed by inverting them 4-6 times. They were then incubated at room temperature for 5 minutes and the color of the lysate turned into homogenous blue. After that, 350μl N3 buffer was added and the tubes were mixed again by inverting them 4-6 times and incubated in ice for 2 minutes. By the end of this time the blue color had dissapeared. The tubes were then centrifuged for 10 minutes at 13000rpm in Eppendorph Centrifuge. Then 850μl of supernatant was transferred into QIAprep

spin column by pipetting and centrifuged for 1 minute at 13000rpm in a micro-centrifuge. The flow-through was discarded and 500μl PB buffer was added to wash the column and it was centrifuged again 1 minute at 13000rpm. Again, the flow-through discarded. Then the spin column was washed with 750μl of PE buffer by centrifugation for 1 minute at 13000 rpm and the flow-through was discarded. Centrifugation was repeated once more at 13000 rpm for 1 minute to remove any remaining residual buffer. Then the column was transferred into a sterile 1.5 ml micro-centrifuge tube and DNA was eluted by adding 50μl EB buffer and centrifugation for 1 minute at 13000 rpm. The flow-through was then stored at -20°C for further use.

2.3.2. Plasmid Digestion with Restriction Endonucleases

The recombinant pQE-3/775 plasmid purified by one of the methods explained in section 2.3. was digested with restriction enzymes to check for presence of the cloned gene and its characterization. Single digestions were performed with *HindIII* or SmaI, and double digestions with *SalI* and *PstI*. All enzymes were from Fermentas and digestions were done according to the Manufacturer's protocol. The total volume of a digestion solution was 20-30μl. The incubations were done in a Multi-block heater (Lab-Line), for 2 hours for single digests, and 3 hours for the double digests.

2.3.3. Agarose Gel Electrophoresis

For plasmid DNA characterization agarose gel electrophoresis (AGE) was used. A submarine agarose gel Mini Sub™ DNA Cell BioRad Apparatus (BioRad, Richmond, CA, U.S.A.) was used to run the samples. The gel was prepared with 1X TAE buffer (pH 8.20) and 0.8% agarose with 0.5μg/ml Ethidium Bromide. Then samples were loaded in volumes between 10-20μl. Electrophoresis was performed at 70 volts, 70mAmp currency using a BioRad Power Supply 200/2.0 (BioRad, Richmond, CA, U.S.A.). After electrophoresis, visualization and photographing were done in a Vilber Lourmat TFP-M/WL (Marne La Vallee, Cedex 1, France).

2.3.4. Site Directed Mutagenesis

2.3.4.1. Mutagenesis by Transformer^{TM} Site-Directed Mutagenesis Kit

The principle behind this kit is using two different primers: a selection primer, by which a restriction enzyme cut site in the recombinant plasmid is changed, so that it will not be recognized in the newly synthesized plasmid *in vitro* and as a result the plasmid will not be linearized, and a Mutagenic Primer, which is designed to change the desired site in the clonned gene. After the plasmid is amplified *in vitro*, it is then digested with the restriction enzyme whose site is changed by the mutagenic primer. As a result, the parental plasmids will be cut and linearized, but the mutagenic ones will not. This has two advantages: The linear plasmid has a transformation efficiency of about 1/100 of the parental plasmid, and the probability of having both mutations simultaneously is very high. Thus, the chances of transforming the cells with the mutant plasmids increase. For mutagenesis, the protocol in the Kit's manual was followed with slight modifications. Since the procedure is elaborate and involves many steps simultaneously, it will be broken down into subsections.

Preparation of competent cells: First, the cells were prepared: *E. coli* BMH 71-18 *mut*S cells were grown on a 1.5% LB agar plate containing 50 μg/ml tetracycline antibiotic. They were incubated overnight at 37°C in the plate. The next day, a single isolated colony was picked up and transferred in a bottle containing 5 ml of LB broth provided with 50 μg/ml tetracycline and grown overnight by shaking at 220rpm and 37°C temperature in a Heidolph Unimax 1010 Shaking Incubator (Heidolph Instruments GmbH, Kelheim, Germany). The next day, 500μl of these cells were transferred into a 250 ml flask containing 50 ml of LB medium without antibiotic. They were incubated by shaking at 220 rpm for about 3 hours, and the absorbance was taken in regular intervals at OD_{600} in a Shimadzu 1601 UV Visible Spectrophotometer (Shimadzu Analytical Co., Kyoto, Japan) until they reached an absorbance of 0.5 ± 0.03 as instructed by the Kit. When this value was reached, the cells were chilled in ice for 20 minutes and transferred in centrifuge tubes to be centrifugated for 5 minutes at 4°C temperature at 1200g in a Sigma 3K30 Centrifuge

(Sigma, Germany). The pellet was resuspended in TSS solution and distributed into eppendorph tubes in aliquots. Some of the cells were kept at -80°C for further use.

Preparation of Plasmids and Primers: Both, selection and mutagenic primers were prepared by dilution to 0.5-0.6 μg/μl final concentration. The wild-type plasmid samples were diluted to a final concentration of 0.6-0.7 μg/μl. The Primers were designed according to the kits instructions, with a length of 18-20 nucleotides for single mutations, and 28 nucleotides for double mutation. The sequence length of selection primers was 26 nucleotides long. Their sequences are shown in the table 2.1. The mutagenic site is bold and underlined.

Plasmid Denaturation and Primer Annealing to the Plasmid: The aim of this step is to seperate the strands of the plasmid and to anneal the primers to their desired sites. The preparations were done as specified in the Kit with appropriate volumes and dilutions. The 10X Annealing buffer (provided by the Kit), Plasmid DNA, the

Table 2.1: Primer Sequences for Transformer Site-Directed Mutagenesis Kit

No.	Primer Name	Primer Sequence
Selection Primers		5'-CTGCAGCC**AAGATT**AATTAGCTGAGC-3'
1	**HindIII**	
2	**SmaI**	5'-GCTCGGTAC**CCCAGG**TCGACCTGCAG-3'
Mutagenic Primers		5'-CGATCAG**AAG**GTTGAC-3'
3	**R81K**	
4	**R81E**	5'-CGATCAG**GAG**GTTGAC-3'
5	**R81M**	5'-CGATCAG**ATG**GTTGAC-3'
6	**R69K**	5'-GTGCGGAG**AAG**AAGAGAG-3'
7	**R89E**	5'-GTGCGGAG**GAG**AAGAGAG-3'
8	**R69M**	5'-GTGCGGAG**ATG**AAGAGAG-3'
9	**QR(80-81)EL**	5'-GTATATATCGAT**GAGCTC**GTTGACAAAG-3'
10	**K87R**	5'-GTGTAT**AGA**GTAGTTAAGC-3'
11	**K87E**	5'-GTGTAT**GAA**GTAGTTAAGC-3'
12	**K87I**	5'-GTGTAT**ATA**GTAGTTAAGC-3'

Selection Primer for restriction enzyme cut site, and Mutagenic Primer for changing the desired site were mixed to form a total volume of 20μl, which was then incubated at 100°C pre-boiling water for 3 min. Then, it was immediately chilled in ice for 5 minutes and then briefly centrifuged to collect the solution at the bottom of the tube.

Synthesis of the Mutant DNA Strand: The aim of this step is to synthesize the mutant strand of the plasmid by extending the two primers annealed to the plasmid. To the reaction solution formed previously, Synthesis Buffer, T4 DNA Polymerase and T4 DNA Ligase (all of them provided by the kit), were added. Lastly the final volume was made up to 30μl by adding 5μl double-distilled water. After mixing and brief centrifugation, the tubes were incubated at 37°C for 2 hours in a heat-block. Then, the reaction was stopped by heating the tube containing the solution at 70°C for 5 minutes to inactivate the enzymes.

The Primary Selection by Restriction Digestion: The aim of this step is to linearize the plasmids which were not mutated at the restriction enzyme cut sites, since with such plasmids the transformation efficiency will be 100X low in comparison to the circular plasmid. For digestion, 2μl HindIII (10 units/μl, Roche) was added to the solution described in the "Synthesis of the Mutant DNA Strand" and incubated for an additional 2 hours so that digestion shall take place. After that time period, the reaction was stopped by transferring it into water bath at 70°C to deactivate the restriction enzyme for 5 minutes.

The First Transformation in Bacteria: The aim of this step is to transfer the plasmids into bacteria for *in vivo* amplification. In each eppendorph, 100μl competent cells prepared as described in "Preparation of competent cells" section and 10μl of digestion solution were added, and then chilled on ice inside refrigerator for 20 minutes. Then they were transferred for 1 min at 42ºC heat block, and immediately transferred into glass tubes containing 1ml LB medium with no antibiotic. They were incubated for 1 hour at 37ºC by shaking at 220rpm.

Cell Culture and Colony Counting: The last step of the procedure was that of culturing and colony counting. An aliquot (100μl) from each tube containing

transformed cells were transferred it into LB agar plates with ampicilin (100 μg/ml) and tetracycline (50 μg/ml). The plates were incubated overnight at 37ºC and the colonies were counted the next day to calculate CFU and transformation efficiency by the formula:

Number of colonies /μg of transferred DNA

Then, from randomly selected colonies, plasmids were isolated as described in section 2.3.1.2 and digested with specific restriction enzymes (section 2.3.2) to find out whether the intended sites had been changed by AGE (Section 2.3.3).

Characterization of Mutants by Restriction Digestion: For characterization of putative mutants created by Transformer^{TM} Site-Directed Mutagenesis Kit, the plasmids were purified by QIAprep Spin Miniprep Kit and a Microcentrifuge, to be used in restriction digestion. After changing the target sites, newly-created restriction sites for *BamHI, MboII* and *BseRI* were identified. *BamHI* digestion was done according to the instructions of the Manufacturer (Fermentas), while digestion with MboII and BamHI was done according to New England Biolabs instructions.

2.3.4.2. Transformation by QuikChange® Site-Directed Mutagenesis Kit

This kit makes use of the methylation properties of Bacterial DNA. Bacterial and plasmid DNA in *E.coli* are methylated at every G nucleotide of 3'-GACT-5' sequence and as a result they are recognized as self and protected from digested by the restriction enzymes of the host. On the other hand, *DpnI* restriction enzyme recognizes and cuts the fully- or hemi-methylated DNA, while it does not digest the unmethylated one. Taking advantage of this enzymes property, the kit selects the mutant plasmid DNA which is synthesized *in vitro* in the presence of the primers, PfuTurbo DNA polymerase, and DNA ligase. The non-mutant wild-type (wt) plasmid DNA is methylated when it is isolated from bacteria. After cycles of *in vitro* amplification there will be both parental and mutant plasmids in reaction mixture. When the amplification products are treated with *DpnI* enzyme, the enzyme will digest the methylated parental DNA and leave the mutant plasmids intact. The latter

are the desired plasmids for transformation, in this way increasing the transformation efficciancy drastically.

All the experiments were performed according to the kit instructions. The primers used in these experiments are shown in table 2 with the mutagenic sites being bold and underlined.

The reaction setup was done accroding to the kit instructions. Briefly, primers at a concentration of 50ng/μl and double-strand wt DNA plasmid containing the gene of interest at a concentration 15-16ng/μl were used. The reaction mixture contained 5μl of 10X reaction buffer, 2μl of double-stranded plasmid DNA, 2.5μl of forward and

Table 2.2: Primer Sequences for QuikChange® Site-Directed Mutagenesis Kit

No.	Primer Name	Primer Sequence
1	**R69M(f)**	5'-GTGCGGAG**ATG**AAGAGAG - 3'
2	**R69M(r)**	5'-CTCTCTT**CAT**CTCCGCAC - 3'
3	**R69E(f)**	5'-GTGCGGAG**GAG**AAGAGAG - 3'
4	**R69E(r)**	5'-CTCTCTT**CTC**CTCCGCAC - 3'
5	**R69K(f)**	5'-GTGCGGAG**AAG**AAGAGAG - 3'
6	**R69K(r)**	5'-CTCTCTT**CTT**CTCCGCAC - 3'
7	**R81K(f)**	5'-GTATATATCGATCAG**AAG**GTTGACAAAGTG-3'
8	**R81K(r)**	5'-CACTTTGTCAAC**CTT**CTGATCGATATATAC-3'
9	**R81E(f)**	5'-GTATATATCGATCAG**GAG**GTTGACAAAGTG-3'
10	**R81E(r)**	5'-CACTTTGTCAAC**CAC**CTGATCGATATATAC-3'
11	**R81M(f)**	5'-GTATATATCGATCAG**ATG**GTTGACAAAGTG-3'
12	**R81M(r)**	5'-CACTTTGTCAAC**CAT**CTGATCGATATATAC-3'
13	**K87I(f)**	5'-CAAAGTGTAT**ATA**GTAGTTAAGCTGCCCG-3'
14	**K87I(r)**	5'-CGGGCAGCTTAACTAC**TAT**ATACACTTTG-3'
15	**K87R(f)**	5'-CAAAGTGTAT**AGA**GTAGTTAAGCTGCCCG-3'
16	**K87R(r)**	5'-CGGGCAGCTTAACTAC**TCT**ATACACTTTG-3'
17	**K87E(f)**	5'-CAAAGTGTAT**GAA**GTAGTTAAGCTGCCCG-3'
18	**K87E(r)**	5'-CGGGCAGCTTAACTAC**TTC**ATACACTTTG-3'
19	**QR(80-81)EL(f)**	5'-GTATATATCGAT**GAGCTC**GTTGACAAAG-3'
20	**QR(80-81)EL(r)**	5'-CTTTGTCAAC**GAGCTC**ATCGATATATAC-3'

Table 2.3: PCR Parameters for Site-Directed Mutagenesis.

Segment	Cycles	Temperature	Time
1	1	95°C	30 sec.
2	16	95°C	30 sec.
		55°C	1 min.
		68°C	4 min.

reverse primers, 1μl of dNTP mix and 37μl of double-distilled H_2O in a final volume of 50μl. After addittion of 1μl of PfuTurbo DNA polymerase (2.5U/μl), PCR amplification was performed in a conventional Techgene thermal cycler (Techne Inc. NJ, USA) with the parameters given in table 2.3. Then the samples were incubated in ice for 2 minutes to cool to room temperature and 1μl (10U/μl) of DpnI enzyme provided by the Kit was added into the reaction mixture, and incubated for 1 hour at 37°C for digestion of parental DNA.

Then, the transformation step was carried out. *E.coli* XL1-Blue supercompetent cells were gently thawed on ice, and 50μl aliquots were added to each tube prechilled in ice. The 14-ml BD Falcon polypropylene round-bottom tubes were used for this purpose. Then, 1μl of Dpn I-treated DNA from each sample reaction was added to aliquots of supercompetent cells. The cells were incubated for 30 minutes in ice and then pulse-heated at 42ºC for 45 seconds followed by incubation in ice for an additional 2 minutes. Then 0.5ml supplemented NZY^+ Broth was added and incubated at 37ºC by shaking at 250rpm for 1 hour. Finally, the transformed cells were spread onto LB Agar plates and incubated overnight at 37ºC in incubator. The next day, colonies were counted and transformation efficiency was calculated as explained in section 2.3.4.1. For each mutation site, 10-20 colonies were picked up and their stock cultures were prepared for further characterization.

2.3.5. DNA Sequencing

To confirm the nucleotide substitutions at the target sites, plasmids were isolated from 2-4 mutant colonies for each mutation for DNA sequencing by Microsynth

Company. Plasmid isolation was performed as described in section 2.3.1.2 and the concentrations were adjusted as required by the company in our laboratory.

2.3.6. Protein Purification

2.3.6.1. Cell Lysate Preparation

Cell Lysate was prepared according to the previously optimized conditions in our laboratory (Kocabiyik at al. 2012). Protein from mutants obtained by QuikChange® Site-Directed Mutagenesis Kit and confirmed by sequencing were purified. The cell cultures were prepared in 100 ml LB media provided with 100μg/ml ampicillin in 500ml flasks and grown for 105-140 min at 37ºC with vigorous shaking at 200 rpm until they reached OD_{600} 0.6±0.05 as measured by spectroscopy (Shimadzu Analytical Co. Kyoto, Japan). When the intended OD values were reached, cultures were induced with IPTG provided by the kit and allowed to continue growing for an additional 4-5 hours in the same conditions. Then the cells were then harvested by centrifugation at 4000g for 20 minutes in a Sigma 3K30 Centrifuge (Sigma, Germany). The supernatant was discarded and the pellet was resuspended in 1/20 volume Lysis buffer prepared according to the kit instructions. Then, 1mg/ml final concentration of Lysozyme (Sigma) was added and the cells were incubated in ice for cell lysis to occur. After that, cells were sonicated in a Sonicator VC 100 (Sonics and Materials, CT, U.S.A) on ice and the cell lysate was centrifuged at 10.000g in a Sigma 3K30 Centrifuge for 20 minutes. Then the supernatant was stored at -20ºC for further studies.

2.3.6.2. Protein Purification

Protein purification was carried out by Qia Expressionist Kit (QIAGEN Inc. Valencia, USA.). This was based on QIAGEN technology that uses Nickel-Nitrilotriacetic Acid (Ni-NTA) agarose binding to the 6-His tag proteins. The procedure followed was as described in the Kit manual. The lysate–Ni-NTA mixture was loaded into the column, and the flow-through was collected in eppendorph tubes, approximately equally based on judgment. Then the column was washed twice with

wash buffer and the wash fractions were collected in eppendorph tubes. The 6-His tagged protein elution was achieved by loading 4 times elution buffer (0.5 ml each time), and elution fractions collected in 4 different tubes. After that, the columns were washed, and kept at 4 °C for further use as instructed by the Manufacturer. The collected samples were stored at -20°C for further analysis.

We also carried out purification for one mutant tpv-Hsp14.3 using Fast-Ni-NTA column. The columns were pre-filled with resin, which was resuspended, and then 5ml of protein extract was loaded and allowed to settle together with the resin in the column. Than the flow-through fractions were collected in eppendorph tubes. After that, the column was washed twice with wash buffer and collected into eppendorph tubes. Finally elution was acheived by loading 1 ml elution buffer two times and collecting the elution fractions in two separate eppendorph tubes.

2.3.7. Heat Treatment of the Cellular Extract

The cell extract was thawed on ice and then incubated at 65°C in a Kottermann water bath (Labortechnik, Germany) for 10 minutes. Then the tubes were incubated in ice for the denatured proteins to precipitate. Then the pellet was removed by centrifugation in a Sigma 3K30 Centrifuge (Sigma, Germany) at 12000 rpm and 4°C for 90 minutes. The supernatant was stored at -20°C to be used in purification as described in section 2.3.6.2 or run directly into SDS-Polyacrylamide gel.

2.3.8. Protein Ultrafiltration and Concentration

For this purpose, a method developed in our lab before and optimized for the tpv-Hsp14.3 protein was used. First, the concentration tubes with filters were loaded with 900-1000µl sample and centrifuged at 4000g in a Clinical Centrifuge (Labofuge 200, Heraeus Sepatech, Germany) until about half of the solution had passed the filter. Centrifugation was repeated at least two times, each time by adding 500µl lysis buffer. The filtrate of the last centrifugation was discarded and the protein sample in the concentration tube was saved and stored at -20°C until used.

2.3.9. SDS- Polyacrylamide Gel Electrophoresis

SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the Laemeli procedure. 5 % stacking gel and 12.5 % separating gels were prepared according to the optimized conditions. As a molecular marker PageRuler™ Pre-stained Protein Ladder from Fermentas (UAB, Vilnuis, Lithuania) was used. The running system was that of Bio-Rad (Bio-Rad, Richmond, CA, U.S.A) and the apparatus assembly and running conditions as described by the Instruction Manual.

The procedure of sample preparation was: 10μl sample mixed with 2X tracking dye containing 5% SDS and incubated in boiling water bath for 5 minutes to denature the proteins. Then the protein sample was kept in ice for about 2 minutes. Besides protein samples, 7μl of standard marker was also loaded. After the electrophoresis process, the gel was stained with Coomasie brilliant blue and then destained with destaining solution. The photo was taken in Vilber Laurmat Gel Imaging and Analysis System (Marne La Vallee Cedex 1, France.Vilber Lourmat) under white light.

2.3.10. Determination of Chaperone Activity by Enzyme Protection Assay

In these experiments pig heart Citrate Synthase (CS) was used as the model enzyme (Sigma). It catalyzes the reaction between Acetyl-CoenzymeA (acetyl-CoA) and Oxaloacetic acid (OAA) to form their product Citric Acid (Srere *et al.* 1963).

The hydrolysis of the thioester of acetyl CoA results in the formation of CoA with a thiol group (CoA-SH). The thiol reacts with the DTNB in the reaction mixture to form 5-thio-2-nitrobenzoic acid (TNB). This yellow product (TNB) is observed spectrophotometrically by measuring absorbance at 412 nm. The reaction is summarized below:

(1) **Acetyl CoA + Oxaloacetate → Citrate + CoA-SH + H$^+$ + H$_2$O**
(2) **CoA-SH + DTNB → TNB + CoA-S-S-TNB**
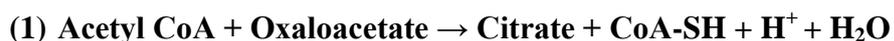
The Activity Protection Assay was performed based on previous studies in the laboratory (Kocabiyik *et al.* 2012). The experimental design was as follows: CS alone at 35°C (its optimal temperature), CS incubated in water bath for 10 minutes at 47°C (it loses its activity at this temperature), and CS incubated in water bath for 10 minutes at 47°C in the presence of the wild-type and each mutant chaperone variants. In the reactions the same amount of each chaperone variant and CS were used. First, CS in 99μl assay buffer was incubated at 35 °C when its activity was measured at optimal conditions or at 47°C when its activity under increased temperature was measured for 10 minutes. Then it was added into a quartz cuvvete which contained DTNB in Assay Buffer pre-incubated at 35°C for 3min. After that oxalacetate and Acetyl-CoA were added and the absorbance was measured at 412nm to determine the enzyme activity at optimal (positive control) and suboptimal temperatures (negative control). For CS heat protection assays, the enzyme activity was measured after preincubation of the enzyme with the purified chaperone at 47°C for 10 minutes. Afterwards, the the enzyme activity was measured as described above.

2.3.11. Studying Chaperone Dynamics by Dynamic Light Scattering

To determine the dynamics of molecular chaperone complex formation as a function of temperature and also the differences between mutant variants in this respect, Dynamic Light Scattering (DLS) Spectroscopy was used. Before describing the experimental procedure, some details of this method will be given. This technique measures particle sizes from nanometers to microns by analyzing fluctuations of scattering data caused by Brwonian motions (random motions of particles in fluid resulting from their random collisions with the molecules of the liquid) of the particles in the solution. It measures the light scattering of a solution as a function of time at a certain angle, than the data are analyzed by the Autocorrelation Function (ACF). The mathematical principles of the method are explained by Berne & Pecora (2006). Here some simple practical principles will be given only (Figure 2.1). There is a laser source which sends the light to the sample and the scattered light is collected by a detector at a certain wavelength, in this case at 632.8 nm (Figure 2.1.a.). The moving particles in the solution create time versus intensity fluctuations

in the order on micro-seconds (Figure 2.1.b.). Then the time/intensity fluctuation data are processed by the software using autocorrelation, which gives the ACF (Figure 2.1.c.). Finally, the processing of the ACF gives the diffusion coefficient (D).

Then, using the Einstein-Stokes Equation, the Hydrodynamic Radius (R_h) of the particles is found. Finally, by means of Laplace Transformation Equation, the distribution of particle size can be determined, as shown in figure 2.1.d. It should be kept in mind that R_h is not a real radius, since many factors can affect its size, like the shape of the protein being globular or extended, the ionic strength of the medium, other proteins that interact with it, the stability of the protein, and the crowdness of the solution. Also water molecules which interact with the protein form a hydration shield around it, like in the case of ions in solution, and in DLS the particles are assumed to be perfect spheres. All these factors generally make the particles measured by DLS appear bigger than their real size.
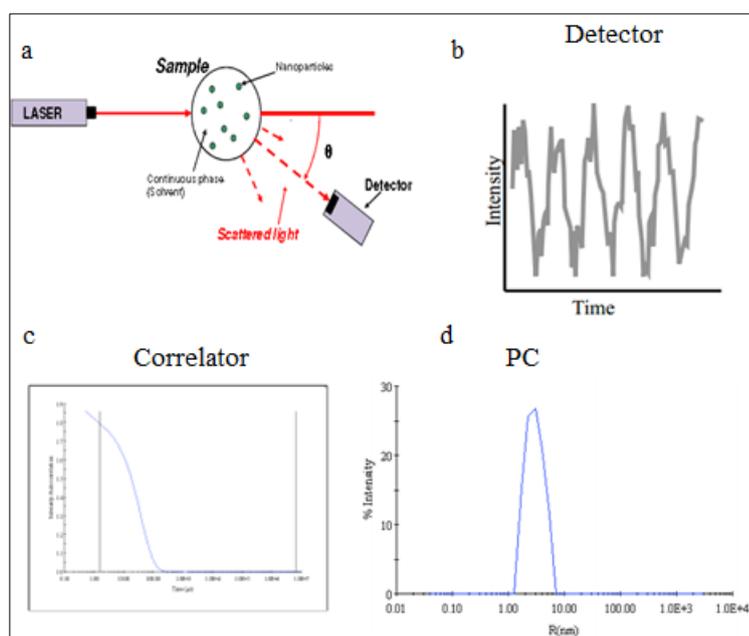
Figure 2.1: Principles of Dynamic Light Scattering. (Retreived from http://lnbio.cnpem.br/spectroscopyandcalorimetry/files/2012/08/LightScattering.pdf on January 29th, 2013).

In DLS experiments, the proteins purified by the method described in section 2.3.6.2 were used. The two best elution samples as judged by SDS Gel photo were combined and their concentration was measured by absorbance spectroscopy by the formula Conc. = $[OD_{235} - OD_{280}]/2.51$ X Dilution Factor (Whitaker *et al.* 1980). Then the samples were further diluted to reach the same protein concentration for each experiment so that comparison would be possible. Experiments were carried out at different temperatures with 10 minute incubation and light scattering measurement. The instrument was Light Scattering Electronics & Multiple Tau Digital Correlator ALV/LSE-5003 with ALV/CGS-3 Compact Ganiometer System (Malvern Technologies). The parameters used were the viscosity of the buffer 0.6473, the refractive index 1.332 (we measured it by Schmidt + Haensch ATRW2 refractometer), the absorbance angle: 90°, and the absorbance wavelength 632.8 nm.

We set up two more experiments where the wild type chaperone was incubated with pig heart Citrate Synthase (Sigma) or bovine Glutamate Dehydorgenase (Sigma) for 10 minutes at 45°C and 51°C respectively, at which temperatures the enzymes loose their activity and aggregate. Then measurements were done for the chaperone itself or for chaperone plus substrate as described above.

2.4. Computation Methods

2.4.1. Sequence Alignments

Sequence alignments, both local and global were performed with Clustal Omega program available at the European Bioinformatics Institute website (http://www.ebi.ac.uk/). The default parameters were used for all alignments with three iterations. Local pairwise sequence alignments were performed with ClustalW program available at the same website with the default parameters.

2.4.2. Secondary Structure Determination

With the advancement of technology and knowledge of protein structure, the power of protein secondary structure prediction increased. The recommended approach for the current state is to combine the results of different approaches on what is called

the Meta-Servers (King *et al.* 2000). Such combination includes Advanced Machine Learning approaches such as Artificial Neural Networks (ANN), Decision Trees, Voting, Linear Discrimination or Consensus (Albrecht *et al.* 2003). Such an approach was first successfully implemented in JPRED metaserver (Cuff *et al.* 1998). It combines the results of different "second generation" secondary structure approaches such as PHD, NNSSP, DSC and PREDATOR, all of them using different prediction strategies and emphasizing different features of secondary structure. Moreover, the server predicts the Solvent Accessibility based on J-NET approach (Cuff et al, 2000) besides transmembrane helices and coiled-coil regions of the protein (Cole *et al.* 2008).

In this study, to predict the secondary structure of the protein tpv-Hsp14.3, J-PRED program was used. This was shown by the Critical Assessment of Techniques for Protein Structure Prediction (CASP) studies (http://predictioncenter.org/) to be the best program for secondary structure prediction. What follows explains the basis of this method in brief details: the server starts first by implementing an algorithm that generates several MSA, then it generates a consensus sequence for all. The server uses also experimental data from protein databanks plus statistical methods to generate a secondary structure of the query. When all the data are averaged, an "optimal" secondary structure with approximately 80% accuracy is predicted (Cuff et al, 2000).

2.4.3. Three-Dimenssional Structure Prediction

The protein's three dimensional structure (3D) is it's biologically functional form. As technology developed and computation power increased, coupled with the increase in the number of proteins with resolved 3D structures in the protein structure database, predicting the 3D structures of the proteins has become one of the most important goals of computational biology.For this there are three main approaches:

2.4.3.1. Comparative Modeling

It is a fairly accurate approach, reaching an accuracy of approximately 80% for small proteins. It is based on two assumptions: first, that proteins with similar sequence

have similar structures, and second, that the protein 3D structure is better conserved than its sequence. It employs homology search and sequence alignment of the query to already resolved proteins. The steps of the method are: Fold Recognition (FR), where the amino acid sequence of the query is compared in protein databases for similar folds. Then comes the step of Template Identification and Selection, where the best templates are selected, to go in the third step, that of Target-Template Alignment. Then the model is built upon the selected templates, followed by refinement to reduce the global potential energy of the model, as proposed by Anfinsen Hypothesis. After these steps, the Loop and Side-chain Modeling can be done for certain region, and finally Model Quality Assessment step is performed to give the most accurate model based on energy minimizations. A summary of these steps is shown in figure 2.2.
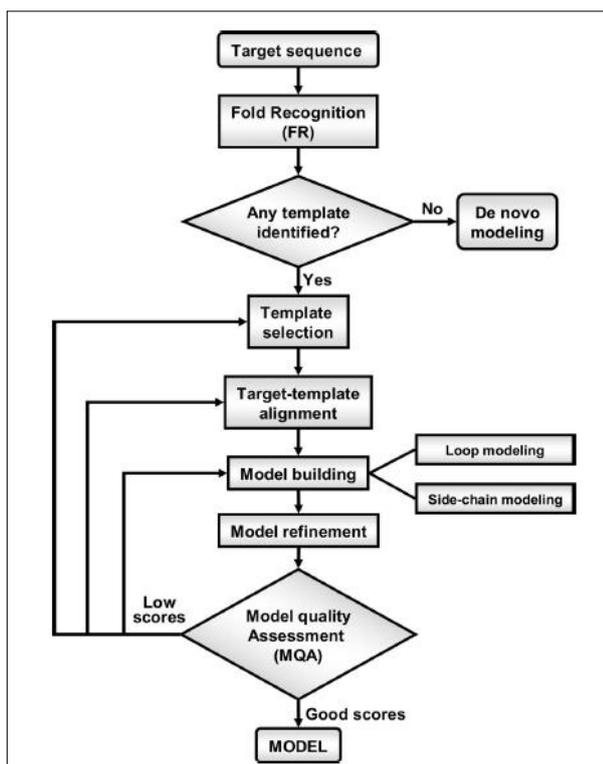
Figure 2.2: Steps of Homology Modeling. (Bujnicki, 2009)

For each of these steps there are certain programs which are beyond the scope of this study (Pawloski *et al*. 2008). Although the method seems very straight, there are certain requirements that must be considered always so that the accuracy of the model will be reasonable. The first is the conventional threshold sequence identity required for accurate models, and that has a cut-off value of 30%. According to the studies, identities lower than this value give inaccurate results (Ohlson *et al.*, 2004, Soding, 2005).

The other important point where caution must be taken is modeling the gap left from sequence alignment, especially in the loop regions, and for this database search has to be performed to fill it from similar sequences of short peptide three-dimensional structures found in the relevant databases (Wallner *et al.*, 2005, Xiang *et al.*, 2007). To be more sure for the accuracy of the model, Molecular Dynamics, Conformation Space Annealing and Monte Carlo Simulations are highly preferred to simulate the model and reduce the global energy further (Levitt *et al.*, 1975, Wales *et al.*, 1999, Li *et al.*, 1987). Another advanvantage of the simulation programs is that they increase the chances of getting the lowest potential energy, and at the same time decrease the probability of the model being stuck in some metastable states of the energy funnel, as explained in section 1.1.2. and shown in figure 1.5 (Rohl *et al.* 2004; Bradley *et al.* 2005; Summa *et al.* 2007).

2.4.3.2.Template-Free Modeling

This approach is also called *Ab Initio* modeling, and it is used to predict the 3D protein structure directly from the amino acid linear sequence without any reference structure. It is based on Anfinsen's Hypothesis as explained in the protein folding section 1.1.2. To achieve reliable models in a reasonable time scale, the approach employs different algorithms and inter-atomic forces. It uses Quantum Mechanics, Empirical Force Fields, Knowledge-Based Force Fields, and other calculations which are complex statistical and mathematical models beyond the scope of this thesis (Bujnicki, 2009). Then it uses Molecular Dynamics and Monte Carlo Simulations to reduce the total energy and refine the model (Levitt *et al.*, 1975, Wales *et al.*, 1999,

70

Li *et al.*, 1987). The accuracy of *De Novo* modeling is not as high as that of Homology Modeling, but it is a good substitute for structure prediction when no homologue structures have been resolved for a query. One major disadvantage of *Ab Initio* is the size limitation of the proteins due to the high conformational space and computation limit (Bujnicki, 2009).

2.4.3.3.Protein Threading (Fold Recognition)

This category is the composite of *Ab Initio* and Homology Modeling, thus giving the best result of both approaches (Bujnicki, 2009). When the sequence identity of a query to one or some targets whose structures have been resolved is very low, this approach uses the sequence homology and *de novo* prediction for the absent region to form a backbone structure. Literally the sequence of the query is threaded on a consensus backbone generated by the target structures, and then the absent regions are modeled by using rotamer databases as it was done in comparative modeling method for loop and gap modeling steps (Wallner *et al.*, 2005, Xiang *et al.*, 2007). Finally, Molecular Dynamics and Monte Carlo Simulations are used to lower the total energy and give the final structure. This approach is also very accurate, sometimes even more accurate than homology modeling since it overcomes the size-limitation of *Ab Initio*, and the problem of gaps and loops of Homology Modeling (Kolinski *et al.*, 2005).

In this study, to predict the three-dimensional structure of *T.volcanium* Hsp14.3, I-TASSER Metaserver (Zhang, 2007), which has ranked as the best structure predictor according to Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments (http://predictioncenter.org/), and Modeller version 9.2 (Eswar, 2006) were used. The first is a meta-server using a protein threading approach, and the second uses homology modeling approach. In the case of I-TASSER, the sequence of each variant was loaded to the database, and the results were obtained after some time, depending on the load of the server. A simplified flow protocol of I-TASSER is shown in figure 2.3.
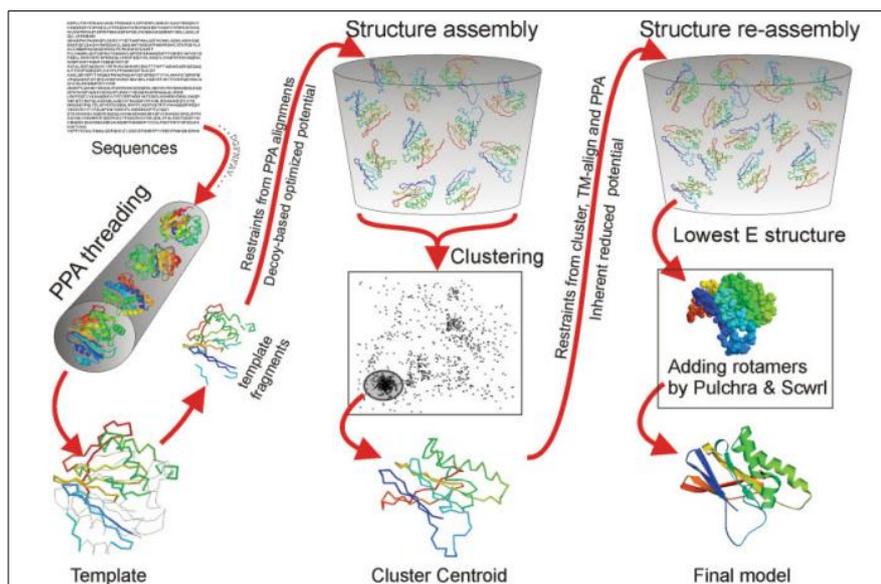
Figure 2.3: A Simplified I-TASSER protocol. (Zhang, 2007)

The program starts first with loading of the query sequence, which is then threaded on already resolved protein structures in databases based on sequence homology. After that the fragments of the protein whose structures were predicted based on threading are assembled and the gaps are predicted by *Ab Initio* modeling. At the same time, secondary structure prediction is performed by different programs to give a consensus secondary structure. Then a pool of structures is formed which are simulated by Monte Carlo simulations and clustered.

Finally, a cluster centroid is obtained which is simulated once again to remove any structural restraint and give the lowest possible E-score. The best structures were selected for further studies, as judged by their scores. The confidence scores (C-Scores) of the models varied from 0.32 to 0.89. The C-scores are typically in the range of [-5,2]. A C-score of higher value means a high confidence for the model. Based on this, the models with the highest scores were chosen while the rest four were discarded. Another way we used to generate the 3D models was using Modeller program (Eswar, 2006), which is structured on homology approach only. But after

further examination, it was found that the models generated by Modeller had many interatomic clashes, which was not experienced in the models generated by I-TASSER. For this reason, we preferred I-TASSER throughout the study for 3D structure predictions. The models' energies were further lowered by using the Chimera's energy minimization option but no differences were observed in RMSD values, meaning that the models are quite accurate, as judged by the available computation power.

CHAPTER 3

RESULTS

3.1. Bioinformatics Analysis and Structure Prediction

3.1.1. Multiple Sequence Alignments

For mutagenesis studies a top-down approach was followed, meaning that the residues to be changed were deduced after many different types of sequence alignments and protein structure predictions (Stotz *et al.* 2006). Based on the structure modeling, the important residues to be mutated in *T.volcanium* Tpv-Hsp14.3 were deduced by comparing their positions to the already studied residues in structurally determined proteins. The best way to find homologs for proteins whose sequences are not conserved but their structures are, such as sHSPs, is to use the Position Specific Iteration Basic Local Alignment Search Tool (PSI-BLAST), because the algorithm it uses is based on the principle of iteration of sequence alignment several times until the "optimal" alignment is found. This makes PSI-BLAST a perfect tool for less homologous proteins (Altschul *et al.* 1997). PSI-BLAST was re-iterated three times using the automatically adjusted parameters until no more new sequence appeared in the screen for the best one hundred matching sequences to the query. Moreover, the hypothetical and redundant sequences were discarded, except in the case of strain differences. From these results, proteins with 30% sequence identity and more were selected, making up a total number of 64 sequences.

The reason of choosing percentage identity as a criterion is that the scores differ for different matrices used, while sequence identity makes more biological sense (Altschul *et al.* 1990). All of the sequences are from species of Bacteria and Archaea, and only one among them has a resolved 3D structure, which is that of *S.tokadii* (accession number Q970D9). With the exception of the Archaeon *T.acidophilum* (accession number NP_394323.1) which had a sequence identity of 76%, the identity of the rest of the sequences dropped to 46% and below, with most of them between 30-35%. The species' names corresponding to each accession number are given in the figure legend (figure 3.1). After Selection of the sequences, Multiple Sequence Alignment (MSA) was performed using Clustal Omega program (Thompson *et al.* 1997) version 1.2.0, and to increase the accuracy of the alignment, three iterations were performed (Thompson *et al.* 2000). The sequence coloring was automatically done by Clustal Omega program.

As is obvious from the MSA (figure 3.1), the conservation of the residues varies from identical to less conserved. However, there is a difference in the sequence length between Bacterial and Archaeal proteins by 30-40 residues, especially at the N- and C-termini, and also at the loops of α-crystallin. Based on this observation, it would make more sense to align only the Archaeal sequences for better results, since the main purpose of the MSA algorithm is to reduce the number of gaps due to the high gap penalty, and prefer residue mismatches to give the highest possible alignment score (Thompson *et al.* 2000).

In figure 3.2 alignment results for the Archaeal sHSP sequences are shown.The alignment parameters were the same as the ones for figure 3.1 and coloring was done automatically by Clustal Omega. This sequence alignment yielded more "highly-conserved" residues than the alignmentof sHSPs sequences from different groups of organisms (figure 3.1). Among the most conserved residues are Q80, R81, P92, Y104 and G107, by taking as reference the *T.volcanium* sHSP sequence.

```
gi|NP_111294.1|        ------------------------------------------------MYTPIKFFTNEMI 13
gi|NP_394323.1|        ------------------------------------------------MYTPVRFFTNEML 13
gi|YP_008142169.1|     ------------------------------------------------MYRPLKYYSDEFM 13
gi|YP_023517.1|        ------------------------------------------------MYRPLKFYSNEFI 13
gi|WP_009070347.1|     ----------------------------------------------------MEIILKEIS 9
gi|YP_002841880.1|     ---------------------------------------------------MMNVIMREIG 10
gi|NP_343935.1|        ---------------------------------------------------MMNVIMREIG 10
gi|Q970D9|             ----------------------------------------------------MYYLGKELQ 9
gi|YP_256268.1|        -------------------------------------------------MSFTYVIEKELG 12
gi|YP_876436.1|        -------------------------------------------------MGLIKSMAREMA 12
gi|ABZ07832.1|         -------------------------------------------------MGLVKHVAKEFF 12
gi|YP_006776550.1|     -----------------------------------------------------MGLVKEVI 8
gi|YP_006774607.1|     -----------------------------------------------------MGLVKEVI 8
gi|YP_001583042.1|     -----------------------------------------------------MGLVKEVI 8
gi|WP_010189564.1|     -------------------------------------------------MGLVKSLAKEMI 12
gi|WP_007402545.1|     -------------------------------------------------MGLVKSLAKEMI 12
gi|WP_007551556.1|     -------------------------------------------------MGLVKSMAKEMI 12
gi|WP_008298667.1|     -------------------------------------------------MGLVKSMAKEMI 12
gi|YP_006863216.1|     -------------------------------------------------MGIGGYVARSIA 12
gi|ABB88979.1|         -------------------------------------------------MGVGQFMAKEFI 12
gi|YP_004809151.1|     ----------------------------------------MRTNTFDDIDRLFDRMNRFAG 21
gi|YP_003357413.1|     -------------------------------------MAEFTWDIQEELRRIEDRMNRMFG 24
gi|YP_005063435.1|     -------------------MKYYVTY----------------NHQNPVSNFESLFN 21
gi|WP_007563476.1|     -------------------MSVRDLIPWNRTTGTPAPTALSNEAANPFLTLHREVNRLFD 41
gi|YP_001568567.1|     ------------------------------MLERRRDDRDDREDLLSPFDFFNREFE 27
gi|YP_002335217.1|     ---------------------------------MLARRNYFDPFVELQREIDRLFE 23
gi|YP_001244140.1|     -------------------------------MLLGRREDIFRPFRELQREIDRLFD 25
gi|WP_008193529.1|     -------------------------------MLLGRREDIFRPFRELQREIDRLFD 25
gi|NP_228185.1|        -------------------------------MLLGRREDIFRPFRELQREIDRLFD 25
gi|WP_010524295.1|     ---------------------------------------MATPMTRLDPMSLMDDVFR 19
gi|WP_017599436.1|     -------------------------------MTRKKFGNPFHGVMDMITEMNRISE 25
gi|WP_017604807.1|     -------------------------------MAPRKFNNPFHGVVDMITEMNRMSD 25
gi|WP_017584932.1|     -------------------------------MTPKRFNNPFHGVVDMITEMNRISD 25
gi|WP_017571483.1|     -------------------------------MAPKRFNNPFHGVVDMITEMNRISD 25
gi|YP_003681484.1|     -------------------------------MASRRFHNPFHGVVDMITEMNRISD 25
gi|WP_019607970.1|     -------------------------------MASRRFHNPFHGVVDMITEMNRISD 25
gi|WP_017566186.1|     -------------------------------MAPKKFHNPFHGVVDMITEMNRISD 25
gi|NP_395730.1|        ----------------------------------------MPLPTGSTSSWLQN 14
gi|YP_001900983.1|     ------------------------------MSDLFFGTDLLGEFDRLQRQMATLFA 26
gi|WP_004628761.1|     ------------------------------MSDLFFGTDLLGEFDRLQRQMATLFA 26
gi|YP_466070.1|        ----MAKPMKHDHGKEPKERGRLAALRE-----S-MPALPP---WM------ERIDELVG 41
gi|YP_002248210.1|     ---------------MTRKETKDVVRV-----E-PSVLSP---FEEIERLFDEVMRRPF 35
gi|YP_004112624.1|     -------------------------------MRYVDIFDEMDNLMRGFGNLTP 22
gi|CAJ73249.1|         ---------------------------------------MMMPLSRTLETLLGLQEALDLTRD 24
gi|YP_006446048.1|     --------------------MFDLISR-----G--ENEREIRPSDRIRKLYQEMDAFIS 32
gi|WP_006633026.1|     -------------------------------MLMRYWQPFTEIETIREQLDKVFD 24
gi|WP_006633768.1|     -------------------------------MLMRYWQPFTEIETIREQLDKVFD 24
gi|WP_007354249.1|     -------------------------------MLMRSWQPFAEIETIRQQLDKAFD 24
gi|WP_019488126.1|     -------------------------------MLMRSWQPFAEIETIRQQLDKAFD 24
gi|YP_007000777.1|     ----------------------MTLVRW-----NPLREIERWEPFPEMGIIRQQMDRLFE 33
gi|WP_002767035.1|     -------------------------------MALIRWEPFREVESLQKEMNRLLD 24
gi|WP_004160648.1|     -------------------------------MALIRWEPFREVESLQKEMNRLFD 24
gi|WP_009370651.1|     --------------------------------MLVKWNPFQELERTLDWWGAQPF 23
gi|YP_003197589.1|     -----------------------------------MVIDFSSFYDLPRSMEQLFD 20
gi|WP_004469872.1|     -------------------------------------MSLIKW--------RGNDMWPDFN 16
gi|YP_004861811.1|     -------------------------------MTLQRFDPFRDIMNLRNQMDALFN 24
gi|YP_001953023.1|     -------------------------------MALVKYNPLRELRSMQDQMDRLLN 24
gi|YP_006721940.1|     -------------------------------MAIVKYTPFGDLRNLQEQMNRLLD 24
gi|NP_951596.1|        -------------------------------MAIVKYSPFRDMMNMQEQMNRLLD 24
gi|WP_006421149.1|     ------------------------------M-----E-MRELAPWSRPGDVIGLRNEMNTLFD 27
gi|WP_006421140.1|     ------------------------------M-----K-MRDLVPWSRPGDIVTLRREMDSLFD 27
gi|YP_003431999.1|     ------------------------------M-RRSIALWNPFAELERIRREFDRLIE 26
gi|WP_006981010.1|     MALSKLDKETEETKNQLHQFQPMSIIRY-----Q-FPQVPNWSSSDRLASLRDEVNRLFD 54
gi|WP_007418376.1|     -------------------MNNLTRW-----Q-RPDLSAWPTFGKLFGLRNELERLFD 33
```

Figure 3.1: Multiple Sequence Alignment of TvShsp14.3 with Proteins Having 30% or More Sequence Identity.

```
gi|NP_111294.1|      KNVS----------NTVKEVSSFIYPPVTLYQDSSDLVLEAEMAGFDK-KNIKVSVNK- 60
gi|NP_394323.1|      KNVS----------NTVKEMSSFIYPPITMYQDGTDLVLEAEMPGFDK-KDIKITVEK- 60
gi|YP_008142169.1|   KNIN----------NRAKEIMTFMYPPVTMYEDNGYIGIEADLPGFAR-DDIKVTLEK- 60
gi|YP_023517.1|      KNIN----------DRAQEIISFLYPPMTMYQENGYIYIDLDMPGFKK-DNISVTLEK- 60
gi|WP_009070347.1|   RQVN----------ELTKEFYERILPPVDIYEQGNTLLVLVDLPGFDK-KDISVRLTSE 57
gi|YP_002841880.1|   KKLD----------ELSRELYESVFPPIDMYEEGGELVIVADLAGFNK-DKISVRLSAQ 58
gi|NP_343935.1|      KKLD----------ELSREFYESVIPPIDMYEEGGELVVVADLAGFNK-DKISVRLSAQ 58
gi|Q970D9|           KRSE----------ELSRGFYELVYPPVDMYEEGGYLVVVADLAGFNK-EKIKARVSGQ 57
gi|YP_256268.1|      RRIE----------QLNRGFYELVYPPVDMYEEGGYLIVVADLAGFNK-DKIKARISGQ 60
gi|YP_876436.1|      REFG----------SKSREFYEFVLPPVDIYMGPSELKVLIDIPGFKK-DEIRLSISR- 59
gi|ABZ07832.1|       KEVD----------DKSREFFEFVLPPIDLHEENDNLIVTVDIPGFDK-SDIKVTMNG- 59
gi|YP_006776550.1|   KEIG----------NKSREFYEFVLPPIDMYLYDDNLKIIIDIPGFAK-KDIELSLCG- 55
gi|YP_006774607.1|   KEIG----------NKSREFYEFVLPPIDMYLNSDNLKVIIDIPGFSK-KDIKLTLCG- 55
gi|YP_001583042.1|   KEIG----------NKSREFYEFVLPPIDMYLNDDNLKVVIDIPGFSK-KDIKLTLCG- 55
gi|WP_010189564.1|   KEIG----------NKSREFYEFVLPPVDIHLENDKLIVLIDIPGFTK-KDIKLSLDG- 59
gi|WP_007402545.1|   KEIG----------NKSREFYEFVLPPVDIHLENDKLILLIDLPGFTK-NDIKLSLDG- 59
gi|WP_007551556.1|   KEIG----------NKSREFYEFVLPPVDMYLDNDKLILIVDMPGFTK-KDIKLSLDG- 59
gi|WP_008298667.1|   KEIG----------NKSREFYEFVLPPIDMHLDDDKLTLLVDIPGFMK-KDIKLSLNG- 59
gi|YP_006863216.1|   KELD----------NRSREFYEFVMPAIDMVEDGSDLVVTIDLPGFAK-KDINLWITG- 59
gi|ABB88979.1|       KEIG----------NRSREFYEFVMPPVDVYEEGSELIVVIDLAGFQK-KDIHLSIYK- 59
gi|YP_004809151.1|   FDDGVG-S----------DSMQSSTMKVDVSDHGDELVVVADLPGFDR-EEIDLSVDD- 67
gi|YP_003357413.1|   EGQGRRPG-----------WQRNSEVPNVDVQEHGNDVIVTADMPGVDK-SDIKINVRNG 72
gi|YP_005063435.1|   DIWSDW------------GVSSSKIPPVDIFETSKAYVLEAELAGYKK-DEVQVNVDK- 66
gi|WP_007563476.1|   DVFTGFGS-------VPSLANRSFGWPNVELVEADGGLRLSAELPGLDE-KDVELLVED- 92
gi|YP_001568567.1|   DFFRSL------------PFGTTSRGEMDVYETDDDYIVECELPGLNK-KDIKVQLNN- 72
gi|YP_002335217.1|   DFVRPS------------RFDTTHFPKVDVYETDKEVVIEAELPGLKK-DDVKITIED- 68
gi|YP_001244140.1|   DFFRTEV----------RPAKEFFAPDMDVYETDDEVVIEVEIPGIDR-KDVKITVEE- 72
gi|WP_008193529.1|   DFFRTEV----------RPAKEFFAPEMDVYETDDEVVIEVEIPGIDR-KDVKITVEE- 72
gi|NP_228185.1|      DFFRTEV----------RPAKEFFAPDMDVYETDDEVVIEVEIPGIDR-KDVKITVEE- 72
gi|WP_010524295.1|   SLRSPLLGSSALAGAGVGGDRESGFVPAVDARRDGEDLVLRADLPGIDPEKDVNVELSG- 78
gi|WP_017599436.1|   TMSSSEISHA----GERERGFADAWSPPTDILARGDDLVIRCEMPGVFE-EDVTVSLSQ- 79
gi|WP_017604807.1|   SLSSFETSQA----GERERGFSDAWSPPTDILARGNDLVIRCEISGVFE-EDVAVNLNH- 79
gi|WP_017584932.1|   TMSSIETSQA----GERERGFADAWSPPTDILARGKDLVIRCEVPGVYE-EDVSVSLNH- 79
gi|WP_017571483.1|   SMSSMETGNA----GERERGHADAWSPPTDILARGDDLVIRCELPGVQE-HEVAVSLNN- 79
gi|YP_003681484.1|   SMSSMETGNA----GERERGHADAWSPPTDILARGNDLVVRCEVPGVQE-QDVAVSLNN- 79
gi|WP_019607970.1|   SMSSMETGNA----GERERGHADAWSPPTDILARGNDLIVRCEVPGVQE-QDVAVSLNN- 79
gi|WP_017566186.1|   SMSSMETGNA----GERERGHADAWSPPTDILARGDDLVVRCEVPGVQE-QDVSVSLNN- 79
gi|NP_395730.1|      SGFPSRL--------------FETGRNDYELYEEDDEFVLSVEMPGFDP-EEITVSWDE- 58
gi|YP_001900983.1|   GAPASLR------------ATRLGTFPPVNIGSTDDSVEIVAFAPGLDP-AQIDVSIDK- 72
gi|WP_004628761.1|   GAPASLR------------ATRLGTFPPVNIGSTDDSVEIVAFAPGLDP-AQIDVSIDK- 72
gi|YP_466070.1|      ERWSAFWP-V-----LRLGEELAFKVPPVDVYEDGRDLVLKAELPGLRR-EDIKVEITG- 93
gi|YP_002248210.1|   SLFRSFV--------PRLREEAEIVSPAVDIYEEGDDLVVKAELPGINK-EDIEVKITD- 85
gi|YP_004112624.1|   MAVPASRR-----------SGSPHGYPALNIWEDSNSFHVDVACPGVRK-EDIDISVNQ- 69
gi|CAJ73249.1|       AGFFENG------------TASSGVYPPVNIFEKNGDLVLVAELPGVKK-EDLNIEVKE- 70
gi|YP_006446048.1|   RFKEFGI------------ASAEGFFPAVDVSETEKEMIVRAELPGMTC-EELDLSVSS- 78
gi|WP_006634026.1|   QRATTRD------------NSEATWMPALELADAGDNFVLKAQLPGIDP-KDIDVQVTR- 70
gi|WP_006633768.1|   QRATTRD------------NSEAAWMPALELVDAGDNFVLKAQLPGIDP-KDIDVQVTR- 70
gi|WP_007354249.1|   QLATTRD------------NSEVAWMPALELADAGDNFVLKAQLPGIDP-KDVDVQVTR- 70
gi|WP_019488126.1|   QLATTRD------------NSEVAWMPALELADAGDNFVLKAQLPGIDP-KDVDVQVTR- 70
gi|YP_007000777.1|   QLLPPDA----------VERVGLTFIPPAEIAETDSDLKLKVEIPGLDA-KDLDVEVTP- 81
gi|WP_002767035.1|   RIVPTDVG--------NGEKVGLSFIPAAEMTETPEAVQLKLEIPGMEA-KDLNVEVTA- 74
gi|WP_004160648.1|   RLVPTDVG--------NGEKMGLSFIPAVEMTETPEAVQLKLEIPGMEA-KDLNVEVTA- 74
gi|WP_009370651.1|   RRPQWDDP----------QASTVNWVPPVNVYEDTEHLFVEAQLPGIDM-KDVKISVTD- 71
gi|YP_003197589.1|   NFWQPSSF-----------PQRRQAFPPLNISEDSENVYVRAEMPGLHV-EDVDLTLTD- 67
gi|YP_004469872.1|   FDFNLPA------------LSNIFARPRIDIMESETEITATAELPGVDK-KDIEISVHD- 62
gi|YP_004861811.1|   EIGMGLLPRT-----AGQTEAAATWSPAVDIYETDKEIVLKAELPDIKQ-EDIRVSVDN- 77
gi|YP_001953023.1|   LSWGGGEY-------PGEDIKEGIWQPAVDIYETADSIVIKAELPDVDQ-KDIDVRIED- 75
gi|YP_006721940.1|   MAWSRES---------GEELREGVWQPPVDIFEDENGVVIKAELPGIDQ-KDIEVKIED- 73
gi|NP_951596.1|      LAWSKQG---------GEELREGAWQPPVDIFEDENAVIIKAELPGIDQ-KDIEVRIED- 73
gi|WP_006421149.1|   RFFDWRPF--------SAGSGVSVWRPALDVSETPKEVLVRAELPGMDP-KEIDISLHD- 77
gi|WP_006421140.1|   RFFDWRPF--------GGGTEVSVWSPALDVSETPKEVLVRAELSGMDP-KEIEINLHD- 77
gi|YP_003431999.1|   EMWPRE------------EVERAFAPAVEMYETDNEIVVKAELPGVKK-ENIEVSIKD- 71
gi|WP_006981010.1|   FSWPSR----------D-SGLFSGWSPALDVFDDKDNLVVKVELPGLNK-DEINISLDK- 101
gi|WP_007418376.1|   VPFSELAQ-------G-SNLLSIWNPAIDVYEDKDNVTVKAELPGMKK-EEIEVSLHD- 82
                                                  . :
```

Figure 3.1: Multiple Sequence Alignment of TvShsp14.3 with Proteins Having 30% or More Sequence Identity. (Continued)

```
gi|NP_111294.1|        NVLTISAERKREY-------STVYIDQRVDKVYKVVKLPVEIEQ-QDI--SAKYSEGILT 110
gi|NP_394323.1|        NVLTLRAERKREY-------RSVYIDQRVDKVFKVVRLPVDVDQ-ASI--SAKYQDGLLT 110
gi|YP_008142169.1|     NAIVIRADREIKP------EGTVFENQRPEKVFKRMSLPMEVDTEQDF--SAKYNDGVLS 112
gi|YP_023517.1|        SYVVINASREINK------GGTVFENQRPSKVFKRIQLPGEPDKNADV--SAKYEDGVLH 112
gi|WP_009070347.1|     GVLRIEARRDVEQ------AGIKHVTQRPSRLSREFKLPVKVPKDAEV--EGKYENGVLT 109
gi|YP_002841880.1|     NELIINAEREIQY------IGTKYTTQRPLKIHKVIRLPVKIKRDSQV--TAKYDNGVLT 110
gi|NP_343935.1|        NELIINAEREIQY------IGTKYATQRPLKIHKVIRLPVKVKRDSQV--TAKYENGVLT 110
gi|Q970D9|             NELIIEAEREITE------PGVKYLTQRPKYVRKVIRLPYNVAKDAEI--SGKYENGVLT 109
gi|YP_256268.1|        NELIIEAEREISE------PGIKYLTQRPKYIRRTIKLPINVAKDAEV--AGKYENGVLT 112
gi|YP_876436.1|        NILHISAERRAC-------HGRAVCAQRPKLIDKSIRLPIYVKRDEEPVAPAKYEDGVLT 112
gi|ABZ07832.1|         NVLSINAEKKDAV------NGRTIMKQRPRVIDKKMRLPITVKEGEEKVNSAKYSDGVLT 113
gi|YP_006776550.1|     DILSINAKKIIDE----KESEFLISNQRPNVIDKKIRLPIEIKQGEEKIKSAKYENGILT 111
gi|YP_006774607.1|     DILSIQSCKKVDE----KESESLILKQRPNIIDKKIRLPIDIKQGEEKIDSAKYEDGVLT 111
gi|YP_001583042.1|     DILSIQACKEVDE----KESESLISKQRPNIIDKKIRLPIDIKQGEEKIDSAKYEDGVLT 111
gi|WP_010189564.1|     NILSIQACKEISD----EKNHNMICNQRPNIIDKKMKLPVEIDDGG-EVSSAKYVDGVLT 114
gi|WP_007402545.1|     NILSIQACKEISD----EKNHNMICNQRPNIIDKKMRLPVEIDDAE-KVNSAKYVDGVLT 114
gi|WP_007551556.1|     NILSINASKEISE----EKNRNIICNQRPNVIDKKIRLPIELREGENGMSSAKYDNGVLT 115
gi|WP_008298667.1|     NILSIQACKEISE----D-KHNMICNQRPNIIDKKIRLPIELKDGEESVSSAKYENGVLK 114
gi|YP_006863216.1|     NVLSISAKRGQEE-----ALGTVYYRHRPVRIEKKVLLPISVKEEDKVVGKATYENGVVT 114
gi|ABB88979.1|         DILSIKAKRTAEG----LDFTTVHYMQRPMQVEKRIPLPISITDEENINSKATYVNGVVT 115
gi|YP_004809151.1|     DQLTIAASHQVDAEETDEERSYVHRERASRTVRRSISLPVEVDAE-GA--SAAYSNGVLT 124
gi|YP_003357413.1|     NILEISAQKRTEME--KREEGFIRHERGYTGYYRSITLPAPVDRS-GA--SARYNNGVLE 127
gi|YP_005063435.1|     HVLRLTSSKKSEK--TPEGKKALVRERYYKEFERSFSLPEDIDEE-AI--EGEFADGILT 121
gi|WP_007563476.1|     GVLTLRGEKRAEN--TD--KARGYSERSYGRFERVIALPFPVEEE-KV--EASFRNGVLT 145
gi|YP_001568567.1|     DLLTISAEKKESD--EVKRGNVYRRERYFGRIERTIRLPEYIDKD-KI--KAEYENGVLK 127
gi|YP_002335217.1|     NVLTIKGERKFNR--EDKGKNYKIIERAEGYFERSFGLPEYVDVE-KI--KAKFNDGVLT 123
gi|YP_001244140.1|     NILKISGEKKVER--EQKGKNYYYVERSAGKFERAIRLPDYVDAE-KI--KAEYKNGVLT 127
gi|WP_008193529.1|     NILKISGEKKLER--EQKGKNYYYVERSAGKFERAIRLPDYVDVE-KI--KAEYKNGVLT 127
gi|NP_228185.1|        NILKISGEKKLER--EQKGKNYYYVERSAGKFERAIRLPDYVDVE-KI--KAEYKNGVLT 127
gi|WP_010524295.1|     RTLTVSGERRAEQ------EAEGLREVRYGSFHRTVTLPQEVSSE-AI--RADYDAGVLT 129
gi|WP_017599436.1|     NVLTISGQRRRDE----DDVVYYSSERYMGAFRRDISLPENIGDE-DV--EATYGEGLLE 132
gi|WP_017604807.1|     GVLTISGERRRDE----EDVVYYSAERFMGTFRREISLPDKVQEE-DI--EAGYGEGLLE 132
gi|WP_017584932.1|     GILTISGERRRDE----DDVVYYSSERFMGTFRREISLPDDVDEE-HI--EASYGEGLLV 132
gi|WP_017571483.1|     GILTVSGERRRDD----DDVIYYSSERFMGTFRREISLPEAVDED-DI--EASYGEGLLE 132
gi|YP_003681484.1|     GILTISGERRRDG----DDVVYYSSERFMGTFRREISLPEAVDEE-DI--EASYGEGLLE 132
gi|WP_019607970.1|     GILTISGERRRDG----DDVVYYSSERFMGTFRREISLPEAVDEE-DI--EASYGEGLLE 132
gi|WP_017566186.1|     GILTISGERRRDG----DDVVYYSSERFMGTFRREISLPEAVDEE-DI--EASYGEGLLE 132
gi|NP_395730.1|        GVLNIAAEHEDET-----------RSQRKTYHRRFRFPKNVEDD-DI--EAQYNNGILE 103
gi|YP_001900983.1|     GLLTISGERKRPDFEMTDDTRVYADERFTGAFRRVIELPQDVDPD-KV--NARYVNGCLT 129
gi|WP_004628761.1|     GLLTISGERKRPDFEMTDDTRVYADERFTGAFRRVIELPQDVDPD-KV--NARYINGCLT 129
gi|YP_466070.1|        DLVTLSGKKEKE--RIERKDYHRVERATGAFTRTVRLPVEIAVD-RV--TAKFTEGVLE 148
gi|YP_002248210.1|     DYLTISGEKKKEE--KVEKKDYRYERSYGSFSRTFRLPVDVQTD-KA--KAKFEKGVLE 140
gi|YP_004112624.1|     NMLTLEFERKQLQ---GDSLEYSRIESRYGKFKRNVALKADVEID-AI--AASYEDGILS 123
gi|CAJ73249.1|         NILRLSGTRTIDY---GKNVSYHRIERNFSEFDRTLRLPFNIESE-KV--QAEYKEGLLV 124
gi|YP_006446048.1|     NVLTIRGEKKAEH--MENQEDVRMNECFFGRFSRSVTLPSQIDQD-AV--EAEYKQGVLT 133
gi|WP_006633026.1|     EAISISGERRYEN--TDEKPRYVRSEFRYGKFHRVLPLPAHIEND-SV--QAEYKDGILT 125
gi|WP_006633768.1|     EAISISGERRYEN--TEEKPRYVRSEFRYGKFHRVLPLPAHIQND-SV--QAEYKDGILT 125
gi|WP_007354249.1|     EAISISGERRYEN--IEAKSGYVRSEFRYGKFHRVLPLPARIQND-SV--QAEYKDGILT 125
gi|WP_019488126.1|     EAISISGERRYEN--IEAKSGYVRSEFRYGKFHRVLPLPARIQND-SV--QAEYKDGILT 125
gi|YP_007000777.1|     ESVSISGERKSET--TTEAEGLTRSEFRYGKFQRVIALPAVVDNE-KV--AAEYKDGILY 136
gi|WP_002767035.1|     DSLTISGERKSEI--KTEEEGFTRTEFRYGKFHRVIPLPVQVDNN-NV--TAEYKDGILN 129
gi|WP_004160648.1|     DSLTINGERKSEI--KTEEEGITRTEFRYGKFHRVIPLPVQVDNN-NV--TAEYKDGILN 129
gi|WP_009370651.1|     NTLQLHGERKEER--TEDRDNYHFREAQYGAFARSFRLPGYVKPD-LA--TARYDKGVLT 126
gi|YP_003197589.1|     NSLIIKGERVQEE------GKYFRQERAAGVFQRLVNLNVPVQRD-NI--SATMRNGILE 118
gi|YP_004469872.1|     DVLEIKGQTSKET--ERKNQSYYLNERYYGSFERRIGLPAEVDSE-RT--TAKFENGILT 117
gi|YP_004861811.1|     NRLSITGERKFES--EVKRENYHRIERSYGTFARTFTLPPTVDQD-NI--RAEYKQGVLT 132
gi|YP_001953023.1|     NLLTIKGERKHES--EVKKENYHRIERYFGSFQRSFKLPATVEQE-KV--AASCEKGVLT 130
gi|YP_006721940.1|     NTLTIRGERKHDQ--EVKKENYHRVERYYGSFMRSFSLPTTIDRD-TV--KAVCDKGILT 128
gi|NP_951596.1|        NTLTIRGERKHEE--EVKRENYHRVERYYGSFQRSFSIPATIDQE-KV--RASSDKGVLT 128
gi|WP_006421149.1|     NVLTVKGERKQEK--EEKDENYHRVERSYGSFVRSFRLPAEVESE-KV--GASYKDGILT 132
gi|WP_006421140.1|     NVLTVRGERKQEK--EDKEENYHRVERSYGSFVRSLRLPAEVESD-NV--DATYKDGILM 132
gi|YP_003431999.1|     NTLHIRGEKKEER--EEKTETIHRLERVYGKFERVLTLPVDVKAE-EV--KAEYKDGILE 126
gi|WP_006981010.1|     GVLTVSGERKQEH--ESKEGESFRSERYFGKFHRSVTLPATVDST-KV--SASYKDGILT 156
gi|WP_007418376.1|     GALVISGERKSEE--KFENAETYRAERFVGRFHRTVTLPSSVKGD-QV--KAQYKDGILT 137
                       : :                            : . :              .   * :
```

Figure 3.1: Multiple Sequence Alignment of TvShsp14.3 with Proteins Having 30% or More Sequence Identity. (Continued)

```
gi|NP_111294.1|          VRMKTKNIK-N---VEIE---------------- 124
gi|NP_394323.1|          VRMKAKDIK-T---VEIE---------------- 124
gi|YP_008142169.1|       LKIPVKNVK-T---VKIE---------------- 126
gi|YP_023517.1|          LSIPAKDVK-S---IKVE---------------- 126
gi|WP_009070347.1|       LKIPIEGAA-K---VKIE---------------- 123
gi|YP_002841880.1|       IRIPVDGSV-S---IKIE---------------- 124
gi|NP_343935.1|          IRIPVEGSV-S---IRIE---------------- 124
gi|Q970D9|               IRIPIAGTS-V---IKIE---------------- 123
gi|YP_256268.1|          IRIPIAGVS-A---IKIE---------------- 126
gi|YP_876436.1|          LIFPVTGRG-K--SISIE---------------- 127
gi|ABZ07832.1|           IEIPITKKG-K--EISLD---------------- 128
gi|YP_006776550.1|       LIIPVIKKG-K--DIPIE---------------- 126
gi|YP_006774607.1|       LVIPVTKKG-K--DISIE---------------- 126
gi|YP_001583042.1|       LVIPVTKKG-K--DISIE---------------- 126
gi|WP_010189564.1|       VTIPVKKHG-K--DISIE---------------- 129
gi|WP_007402545.1|       VIIPVKKHG-K--DISIE---------------- 129
gi|WP_007551556.1|       IIILVKKHG-K--DIPIE---------------- 130
gi|WP_008298667.1|       IIIPVQKQG-K--DITIE---------------- 129
gi|YP_006863216.1|       LRVPIPKTT-N---IPVT---------------- 128
gi|ABB88979.1|           LKIPLPHTS-K---IPVT---------------- 129
gi|YP_004809151.1|       VTLPTLAES-SGHRIDID---------------- 141
gi|YP_003357413.1|       ITLPMTKEA-TENIIPVS---------------- 144
gi|YP_005063435.1|       VTLPKKELV-KPKQIEVKIKENV----------- 143
gi|WP_007563476.1|       VTLPRSEKA-PERGRRIAINSGNNGGQKTAH----- 175
gi|YP_001568567.1|       LTIPKVETA-KGEGKEIKIE-------------- 146
gi|YP_002335217.1|       IELPKKETK-DKKVIDIQVE-------------- 142
gi|YP_001244140.1|       IRVPKKEER-KRKVIEVEVQE------------- 147
gi|WP_008193529.1|       IRVPKKEER-KKKVIEVEVQE------------- 147
gi|NP_228185.1|          IRVPKKEER-KKKVIEVEVQE------------- 147
gi|WP_010524295.1|       VTVQGVYAQEAPQRIQVTRGESGQDS-AQQIES--- 161
gi|WP_017599436.1|       VTVHGAANAAVPTTISISRRRR------------ 154
gi|WP_017604807.1|       IVVHGAAYQRGPAQISITRRQRR----------- 155
gi|WP_017584932.1|       VTVHGAANQRGPKRISIARRKRT----------- 155
gi|WP_017571483.1|       VVVHGAANVRKPRPITIRKRSRGRS---------- 157
gi|YP_003681484.1|       VVVHGAANARGPRRIAIRKRTPKKS---------- 157
gi|WP_019607970.1|       VVVHGAANARGPRRIAIRKRTPKRS---------- 157
gi|WP_017566186.1|       VVVHGAANARGPRRITIRKRTPKRS---------- 157
gi|NP_395730.1|          VRLPVLTGA-TTRGKQIEVQA------------- 123
gi|YP_001900983.1|       ISVGKSEAS-KPRQITVQG--------------- 147
gi|WP_004628761.1|       ISVGKSEAS-KPRQITVQG--------------- 147
gi|YP_466070.1|          IRAPMAGEG-KGHGRKIEVT-------------- 167
gi|YP_002248210.1|       IRIPKTEEA-KKKERKLQIE-------------- 159
gi|YP_004112624.1|       IAIPKAAHA-KPRKIEISA--------------- 141
gi|CAJ73249.1|           VSLPRAETD-KPKKIAIQ---------------- 141
gi|YP_006446048.1|       VKLLKSGAS-EQGVSQITVQ-------------- 152
gi|WP_006634026.1|       LTLPKVTEA-RNKVVKINLAQVAGTTANPTLEQANQ 160
gi|WP_006633768.1|       LTLPKVTEA-RNKVVKINLAQVAGAPANPALEQAN- 159
gi|WP_007354249.1|       KEWLCLME------------------------- 133
gi|WP_019488126.1|       LTLPKVAEA-RNKVVKINLAEIAGASRNPALEQANQ 160
gi|YP_007000777.1|       LTIPKAESE-KHKAVKVRL--------------- 154
gi|WP_002767035.1|       LTLPKAEEE-KNKVVKVSISPTIAQ---------- 153
gi|WP_004160648.1|       LTLPKAEEE-KNKVVKVSISPAIAQ---------- 153
gi|WP_009370651.1|       VQVPKREET-KPRSIPIDVKG------------- 146
gi|YP_003197589.1|       IRLPKSEEI-KPKKISIDVG-------------- 137
gi|YP_004469872.1|       IVMPKLHPD-KPKGRKIDIQ-------------- 136
gi|YP_004861811.1|       VSLPKREVA-QGRNIAIQVN-------------- 151
gi|YP_001953023.1|       VTLPKKEEV-KPKQINVEVK-------------- 149
gi|YP_006721940.1|       ITLPRREET-KPKQINVEVK-------------- 147
gi|NP_951596.1|          ITLPKREEV-KPKQITVEVT-------------- 147
gi|WP_006421149.1|       IRLKKSAKV-AKKKIEIKTT-------------- 151
gi|WP_006421140.1|       IKLKKSEKI-AQRKIEIKAG-------------- 151
gi|YP_003431999.1|       IRLPKSEVS-KEKKIEIK---------------- 143
gi|WP_006981010.1|       VDLPKAEEA-KPKQIAVNVA-------------- 175
gi|WP_007418376.1|       ITLPKAEEA-KPKQIEVNIA-------------- 156
```

Figure 3.1: Multiple Sequence Alignment of TvShsp14.3 with Proteins Having 30% or More Sequence Identity. (Continued). The color code of the sequences is: Red for small hydrophobic residues, Blue for acidic residues, Magneta for basic residues, green for Hydroxyl-, sulfhydryl-, amine-containing groups and Glycine. The species names corresponding to each accession number are as shown below, while a comprehensive table with species name and protein information is shown in the Appendix E: NP_111294.1 *Thermoplasma volcanium GSS1*, NP_394323.1 *Thermoplasma acidophilum DSM 1728*, YP_008142169.1 *Ferroplasma acidarmanus*, NP_395730.1 *Halobacterium sp. NRC-1*, YP_023517.1 *Picrophilus torridus DSM 9790,* YP_004809151.1 *halophilic archaeon DL31,* YP_466070.1 *Anaeromyxobacter dehalogenans 2CP-C,* Q970D9 *Sulpholobous Tokadii ,*

ABZ07832.1 uncultured marine *crenarchaeote HF4000_ANIW141J13*, WP_017599436.1 *Nocardiopsis lucentensis*, YP_002841880.1 *Sulfolobus islandicus Y.N.15.51*, WP_010524295.1 *Nesterenkonia sp. F*, WP_007551556.1 *Candidatus Nitrosoarchaeum koreensis*, WP_010189564.1 *Candidatus Nitrosoarchaeum limnia*, YP_003681484.1 *Nocardiopsis dassonvillei*, WP_017566186.1 *Nocardiopsis synnemataformans*, WP_008298667.1 *Candidatus Nitrosopumilus salaria*, WP_006981010.1 *Chthoniobacter flavus*, YP_006863216.1 *Candidatus Nitrososphaera gargensis*, WP_007402545.1 *Candidatus Nitrosoarchaeum limnia*, YP_001953023.1 *Geobacter lovleyi*, NP_343935.1 *Sulfolobus solfataricus*, WP_019607970.1 *Nocardiopsis sp. CNS639*, WP_017604807.1 *Nocardiopsis alkaliphila*, WP_017571483.1 *Nocardiopsis halotolerans*, WP_017584932.1 *Nocardiopsis ganjiahuensis*, YP_001568567.1 *Petrotoga mobilis SJ95*, YP_003431999.1 *Hydrogenobacter thermophilus TK-6*, YP_005063435.1 *Sphaerochaeta pleomorpha*, YP_006721940.1 *Geobacter metallireducens GS-15*, WP_006421149.1 *delta proteobacterium NaphS2*, YP_001900983.1 *Ralstonia pickettii 12J*, WP_004628761.1 *Ralstonia pickettii*, NP_951596.1 *Geobacter sulfurreducens PCA*, YP_006776550.1 *Candidatus Nitrosopumilus sp. AR2*, WP_006634026.1 *Microcoleus vaginatus*, YP_002335217.1 *Thermosipho africanus TCF52B*, ABB88979.1 *uncultured crenarchaeote*, YP_006774607.1 *Candidatus Nitrosopumilus koreensis AR1*, CAJ73249.1 *Candidatus Kuenenia stuttgartiensis*, YP_007000777.1 *Anabaena sp. 90*, WP_002767035.1 *Microcystis aeruginosa*, WP_006633768.1 *Microcoleus vaginatus*, WP_007354249.1 *Oscillatoria sp. PCC 6506*, YP_003357413.1 *Methanocella paludicola SANAE*, YP_876436.1 *Cenarchaeum symbiosum A*, YP_001583042.1 *Nitrosopumilus maritimus SCM1*, WP_007418376.1 *Pedosphaera parvula*, YP_003197589.1 *Desulfohalobium retbaense DSM 5692*, YP_004112624.1 *Desulfurispirillum indicum S5*, WP_019488126.1 *Oscillatoria formosa*, WP_006421140.1 *delta proteobacterium NaphS2*, WP_007563476.1 *Methylobacterium sp. GXF4*, YP_006446048.1 *Desulfomonile tiedjei DSM 6799*, WP_009370651.1 *Candidatus Poribacteria sp. WGA-A3*, YP_001244140.1 *Thermotoga petrophila RKU-1*, WP_008193529.1 *Thermotoga sp. EMP*, NP_228185.1 *Thermotoga maritima MSB8*, WP_004160648.1 *Microcystis aeruginosa*, WP_009070347.1 *Metallosphaera yellowstonensis*, YP_256268.1 *Sulfolobus acidocaldarius DSM 639*, YP_004469872.1 *Thermoanaerobacterium xylanolyticum*, YP_002248210.1 *Thermodesulfovibrio yellowstonii*, YP_004861811.1 *Candidatus Chloracidobacterium thermophilum B.*

In the figure 3.2 the classical I/L–X–I/L motif in the C-terminal extension is also shown in box. In contrary to the previously published literature (Waedick *et al.* 2009), considering many newly added sequences, the variation in this conserved motif is actually V/I/L-X-V/I/L, and all of these amino acids, are hydrophobic. Residue colors also show that the substitutions are mainly for analogous residues, and mostly they are found at the N-terminus and at the very end of the C-Terminus.

```
gi|NP_111294.1|      MYTPIKFFTNEMIKNVSNTVKEVSSFIYPPVTLYQDSSDLVLEAEMAGFDKKNIKVSVNK    60
gi|NP_394323.1|      MYTPVRFFTNEMLKNVSNTVKEMSSFIYPPITMYQDGTDLVLEAEMPGFDKKDIKITVEK    60
gi|YP_008142169.1|   MYRPLKYYSDEFMKNINNRAKEIMTFMYPPVTMYEDNGYIGIEADLPGFARDDIKVTLEK    60
gi|NP_395730.1|      -MPLPTGSTSSWLQNSGFPS-RLFETGRNDYELYEEDDEFVLSVEMPGFDPEEITVSWDE    58
gi|YP_023517.1|      MYRPLKFYSNEFIKNINDRAQEIISFLYPPMTMYQENGYIYIDLDMPGFKKDNISVTLEK    60
gi|Q970D9|           ----MYYLGKELQKRSEELSRGFYELVYPPVDMYEEGGYLVVVADLAGFNKEKIKARVSG    56
gi|ABZ07832.1|       -MGLVKHVAKEFFKEVDDKSREFFEFVLPPIDLHEENDNLIVTVDIPGFDKSDIKVTMNG    59
gi|YP_002841880.1|   ---MMNVIMREIGKKLDELSRELYESVFPPIDMYEEGGELVIVADLAGFNKDKISVRLSA    57
gi|WP_007551556.1|   -MGLVKSMAKEMIKEIGNKSREFYEFVLPPVDMYLDNDKLILIVDMPGFTKKDIKLSLDG    59
gi|WP_010189564.1|   -MGLVKSLAKEMIKEIGNKSREFYEFVLPPVDIHLENDKLIVLIDIPGFTKKDIKLSLDG    59
gi|WP_008298667.1|   -MGLVKSMAKEMIKEIGNKSREFYEFVLPPIDMHLDDDKLTLLVDIPGFMKKDIKLSLNG    59
gi|YP_006863216.1|   -MGIGGYVARSIAKELDNRSREFYEFVMPAIDMVEDGSDLVVTIDLPGFAKKDINLWITG    59
gi|WP_007402545.1|   -MGLVKSLAKEMIKEIGNKSREFYEFVLPPVDIHLENDKLILLIDLPGFTKNDIKLSLDG    59
gi|NP_343935.1|      ---MMNVIMREIGKKLDELSRESVIPPIDMYEEGGELVIVADLAGFNKDKISVRLSA    57
gi|YP_006776550.1|   -----MGLVKEVIKEIGNKSREFYEFVLPPIDMYLYDDNLKIIIDIPGFAKKDIELSLCG    55
gi|ABB88979.1|       -MGVGQFMAKEFIKEIGNRSREFYEFVMPPVDVYEEGSELIVVIDLAGFQKKDIHLSIYK    59
gi|YP_006774607.1|   -----MGLVKEVIKEIGNKSREFYEFVLPPIDMYLNSDNLKVIIDIPGFSKKDIKLTLCG    55
gi|YP_876436.1|      -MGLIKSMAREMAREFGSKSREFYEFVLPPVDIYMGPSELKVLIDIPGFKKDEIRLSISR    59
gi|YP_001583042.1|   -----MGLVKEVIKEIGNKSREFYEFVLPPIDMYLNDDNLKVVIDIPGFSKKDIKLTLCG    55
gi|WP_009070347.1|   ----MEIILKEISRQVNELTKEFYERILPPVDIYEQGNTLLVLVDLPGFDKKDISVRLTS    56
gi|YP_256268.1|      -MSFTYVIEKELGRRIEQLNRGFYELVYPPVDMYEEGGYLIVVADLAGFNKDKIKARISG    59
                          .  :.          .        :         : :  :: **   ..*


gi|NP_111294.1|      -NVLTISAERKRE---YSTVYIDQRVDKVYKVVKLPVEIEQ-QDI--SAKYSEGILTVRM   113
gi|NP_394323.1|      -NVLTLRAERKRE---YRSVYIDQRVDKVFKVVRLPVDVDQ-ASI--SAKYQDGLLTVRM   113
gi|YP_008142169.1|   -NAIVIRADREIKP--EGTVFENQRPEKVFKRMSLPMEVDTEQDF--SAKYNDGVLSLKI   115
gi|NP_395730.1|      -GVLNIAAEHEDET-------RSQR-KTYHRRFRFPKNVEDD-DI--EAQYNNGILEVRL   106
gi|YP_023517.1|      -SYVVINASREINK--GGTVFENQRPSKVFKRIQLPGEPDKNADV--SAKYEDGVLHLSI   115
gi|Q970D9|           QNELIIEAEREITE--PGVKYLTQRPKYVRKVIRLPYNVAKDAEI--SGKYENGVLTIRI   112
gi|ABZ07832.1|       -NVLSINAEKKDAV--NGRTIMKQRPRVIDKKMRLPITVKEGEEKVNSAKYSDGVLTIEI   116
gi|YP_002841880.1|   QNELIINAEREIQY--IGTKYTTQRPLKIHKVIRLPVKIKRDSQV--TAKYDNGVLTIRI   113
gi|WP_007551556.1|   -NILSINASKEISEEKNRNIICNQRPNVIDKKIRLPIELREGENGMSSAKYDNGVLTIII   118
gi|WP_010189564.1|   -NILSIQACKEISDEKNHNMICNQRPNIIDKKMKLPVEIDDGGE-VSSAKYVDGVLTVTI   117
gi|WP_008298667.1|   -NILSIQACKEISED-KHNMICNQRPNIIDKKIRLPIELKDGEESVSSAKYENGVLKIII   117
gi|YP_006863216.1|   -NVLSISAKRGQEE-ALGTVYYRHRPVRIEKKVLLPISVKEEDKVVGKATYENGVVTLRV   117
gi|WP_007402545.1|   -NILSIQACKEISDEKNHNMICNQRPNIIDKKMRLPVEIDDAEK-VNSAKYVDGVLTVII   117
gi|NP_343935.1|      QNELIINAEREIQY--IGTKYATQRPLKIHKVIRLPVKVKRDSQV--TAKYENGVLTIRI   113
gi|YP_006776550.1|   -DILSINAKKIIDEKESEFLISNQRPNVIDKPIEIKQGEEKIKSAKYENGILTLII   114
gi|ABB88979.1|       -DILSIKAKRTAEGLDFTTVHYMQRPMQVEKRIPLPISITDEENINSKATYVNGVVTLKI   118
gi|YP_006774607.1|   -DILSIQSCKKVDEKESESLILKQRPNIIDKKIRLPIDIKQGEEKIDSAKYEDGVLTLVI   114
gi|YP_876436.1|      -NILHISAERRAC---HGRAVCAQRPKLIDKSIRLPIYVKRDEEPVAPAKYEDGVLTLIF   115
gi|YP_001583042.1|   -DILSIQACKEVDEKESESLISKQRPNIIDKKIRLPIDIKQGEEKIDSAKYEDGVLTLVI   114
gi|WP_009070347.1|   EGVLRIEARRDVEQ--AGIKHVTQRPSRLSREFKLPVKVPKDAEV--EGKYENGVLTLKI   112
gi|YP_256268.1|      QNELIIEAEREISE--PGIKYLTQRPKYIRRTIKLPINVAKDAEV--AGKYENGVLTIRI   115
                          :  :  :          :*         :  . :*        .      . *  :*::  : .


gi|NP_111294.1|      KTKNI----K-NVEIE-    124
gi|NP_394323.1|      KAKDI----K-TVEIE-    124
gi|YP_008142169.1|   PVKNV----K-TVKIE-    126
gi|NP_395730.1|      PVLTGATTRGKQIEVQA    123
gi|YP_023517.1|      PAKDV----K-SIKVE-    126
gi|Q970D9|           PIAGT----S-VIKIE-    123
gi|ABZ07832.1|       PITKK----GKEISLD-    128
gi|YP_002841880.1|   PVDGS----V-SIKIE-    124
gi|WP_007551556.1|   LVKKH----GKDIPIE-    130
gi|WP_010189564.1|   PVKKH----GKDISIE-    129
gi|WP_008298667.1|   PVQKQ----GKDITIE-    129
gi|YP_006863216.1|   PIPKT----T-NIPVT-    128
gi|WP_007402545.1|   PVKKH----GKDISIE-    129
gi|NP_343935.1|      PVEGS----V-SIRIE-    124
gi|YP_006776550.1|   PVIKK----GKDIPIE-    126
gi|ABB88979.1|       PLPHT----S-KIPVT-    129
gi|YP_006774607.1|   PVTKK----GKDISIE-    126
gi|YP_876436.1|      PVTGR----GKSISIE-    127
gi|YP_001583042.1|   PVTKK----GKDISIE-    126
gi|WP_009070347.1|   PIEGA----A-KVKIE-    123
gi|YP_256268.1|      PIAGV----S-ATKIE-    126
                          :  :
```

Figure 3.2: MSA of Archaeal sHSP Species Showing 30% Amino Acid Sequence Identity or More. Accession numbers and their respective species: NP_111294.1 *Thermoplasma volcanium GSS1*, NP_394323.1 *Thermoplasma acidophilum DSM 1728*, YP_008142169.1 *Ferroplasma acidarmanus*, NP_395730.1 *Halobacterium sp. NRC-1,* YP_023517.1 *Picrophilus torridus DSM 9790,* Q970D9 *Sulpholobous Tokadii,* ABZ07832.1 marine *crenarchaeote HF4000_ANIW141J13,* YP_002841880.1 *Sulfolobus islandicus Y.N.15.51,* WP_007551556.1 *Candidatus Nitrosoarchaeum koreensis,* WP_010189564.1 *Candidatus Nitrosoarchaeum limnia,* WP_008298667.1 *Candidatus Nitrosopumilus salaria,*

YP_006863216.1 *Candidatus Nitrososphaera gargensis,* WP_007402545.1 *Candidatus Nitrosoarchaeum limnia,* NP_343935.1 *Sulfolobus solfataricus,* YP_006774607.1 *Candidatus Nitrosopumilus koreensis AR,* YP_876436.1 *Cenarchaeum symbiosum A,* YP_001583042.1 *Nitrosopumilus maritimus SCM1,* WP_009070347.1 *Metallosphaera yellowstonensis,* YP_256268.1 *Sulfolobus acidocaldarius DSM 639.*

In the previous studies, it was also claimed that the residues at the core of α-crystalline domain, and more specifically those making up the beta strands are more conserved than the rest (Basha *et al.* 2011, van Montfort *et al.* 2001). These conclusions are in agreement with our results. Another striking observation in figure 3.2 was the occurrence of most of the changes in the loop regions linking one beta strand to the other. Moreover, these loops lengths vary very much in distant species. To further confirm this notion, secondary structure prediction of the protein of interest was carried out.

3.1.2. Secondary Structure Prediction

The predicted secondary structure by J-PRED is shown in the figure 3.3. As is obvious from the figure, a classical α-crystalline fold with eight β-strands is generated, flanked by an N-terminal region with α-helical secondary structure and a short irregular C-terminus. The numbering of beta strands start from β2 for historical reasons, so they are named as β2-10 (Kim *et al.* 1998). Jnetconf values which stand for the reliability of the prediction are also the highest for the secondary structures involving regular alpha-helix and beta strands secondary structures, while for the loops they are lower. This confirms the structural and also sequence alignment studies that the most conserved regions are the regular secondary structures, while the loops length and sequences are highly divergent, even between organisms of the same class.
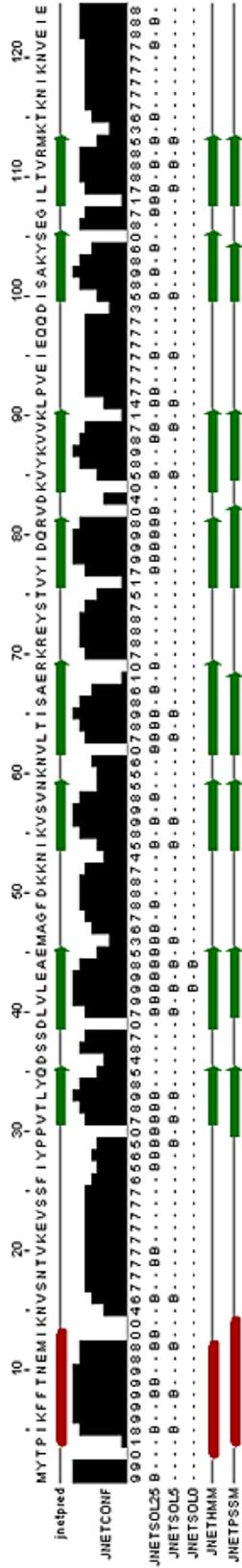
Figure 3.3: Secondary Structure Prediction of Tpv-Hsp14.3: The Abbreviations in the figure stand for: Jnetpred: Final secondary structure prediction for the query. JnetCONF: Jnet prediction reliability with accuracy range 0-9, bigger number means higher accuracy. JNETSOL25: Jnet prediction of residue burial, less than 25% solvent accessibility. JNETSOL5: Jnet prediction of residue burial, less than 5% solvent accessibility. JNETSOL0: Jnet prediction of residue burial, 0% expossure. JNETHMM: Jnet Hidden Markov Model (HMM) profile. JNETPSSM: Jnet PSI-BLAST Position-Specific Scoring Matrix (PSSM).

This predicted secondary structure is used as the reference in all the other studies involving sequence alignment and secondary structure annotation in the coming sections. The most reliable results of J-PRED are the secondary structure prediction, while the prediction of the buried residues by this program is not that much reliable, for which validation of the data with other algorithms will be required (Cuff et al, 2000). The better results for such residues are gained by 3D structure prediction, as will be discussed in later section.

3.1.3. Tpv-Shsp14.3 Sequence Alignment with well-characterized sHSPs.

Since the aim of the study is to find well-conserved and potential functionally important residues, the query must be compared with structurally and functionally characterized homologous proteins. A search in the database revealed ten such proteins. Nine of them have X-Ray or NMR resolved structures with different resolutions, while one (yeast sHSP) has been studied by electron microscopy. However, structural modeling of yeast Hsp26 based on the structure of Wheat Hsp16.9 and fitting into the shell generated by Cryo-Electron microscopy and substrate-binding studies have shown some important residues involved in protein-substrate interaction (White *et al.*, 2006). The proteins with resolved 3D structures are: Human αA-Crytstallin (Laganowsky *et al.*, 2010), Human αB-Crystallin (Jehle *et al.*, 2010), Rat Hsp20 (Bagneris *et al.*, 2009), Wheat Hsp16.9 (van Montfort *et al.*, 2001), Bovine αA-Crytstallin (Laganowsky *et al.*, 2010), Zebrafish αA-Crytstallin (Laganowsky *et al.*, 2010b), *Xanthomonas* HspA (Hilario *et al.*, 2011), *M.janaschii* Hsp16.5 (Kim *et al.*, 1998), *S.tokadii* Hsp14.0 (Takeda *et al.*, 2011). Though at first glance this seems straight forward, aligning these sequences is challenging. The big challenge lies in the sequence length difference between Archaeal sHSPs which are approximately of 120 amino acids length, and eukaryotic ones (the yeast sHSP is composed of 214 residues and the others are about 170 residues). This length discrepancy reduces the power of sequence alignment because the aim of any algorithm is to give the maximum score by introducing the least number of gaps (Thompson *et al.* 2000). If the sequence length is very significant, then gaps must be

introduced, but to keep their number as small as possible, many sequence mismatches are permitted by the algorithm. By using different Multiple Sequence Alignment programs, like T-COFFEE, MUSCLE, ClustalW2 and Clustal Omega in European Bioinformatics Institute (EBI) webpage (http://www.ebi.ac.uk/), the alignment with the least number of gaps at the region of alpha crystalline was obtained by Clustal Omega program (figure 3.4) (Thompson *et al.* 1990). This choice was based on the already published literature which has proven that the greatest sequence divergence between sHSPs is in the N- and C-terminal regions and in the loops between the beta strands of the α-crystalline.

In figure 3.4, the α-crystalline domain is the region under the line. Human HSPB, HSPB5 and Rat Hsp20 do not have a β2 strand revealed by structural studies to date. However, it is disputable and may be an artifact of experimental conditions for structure determination (Laganowsky *et al.*, 2010, Jehle *et al.*, 2010, Bagneris *et al.*, 2009). Moreover, human, bovine, rat and zebrafish sHSPs have a single, long β6+7 strand that forms an interface with the same strand of the other monomer to form the dimer as shown in figure 1.14. This is thought to be the dimerization mode in higher eukaryotes, especially the vertebrates (Basha *et al.*, 2011). The sHSP sequences of archaea and plants have two separate β6 and β7 strands and their mode of monomer interaction to form dimers is shown in figure 1.14 and is explained in details in section 1.3.4. In the sequence alignment it is also apparent that the V/I/L-X-V/I/L motif (inside the box) is found in most of the sHSPs with the exception of rat protein. The ultimate aim of the strategy is to find the most conserved residues of the sequences mentioned above, and this can be best acheived by local sequence alignment. The reason is that local sequence alignment programs concentrate on the most conserved "seed" regions and discard the highly divergent sequences (Thompson *et al.* 2000). Since the focus is α-crystalline, and it is the most conserved domain, the results would be more reliable by this method. There is a debate regarding the bounrdies of the α-crystalline with no consensus reached yet. However, in published literature the researchers use certain boundries, and in this study, it was relied upon them. (Basha *et al.* 2011).

```
 HSPB1_HUMAN      ----------------------------------------MTERRVPFSLLRGPSW--- 16
 HSPB6_RAT        ----------------------------------------MEIRVPVQPSWLRR 14
 CRYAB_HUMAN      ----------------------------------------MDIAIHHPWIRR 12
 CRYAA_BOVINE     ----------------------------------------MDIAIQHPWFKR 12
 Q8UUZ6_Z.FISH    ----------------------------------------MDIAIQHPWFRR 12
 HSPS_M.JANASCHII -------------------------------------------MFGRD 5
 Q970D9_S.TOKADII ------------------------------------------------------- 0
 Q97AM1_T.VOL     ------------------------------------------------------- 0
 HSP26_YEAST      MSFNSPFFDFFDNINNEVDAFNRLLGEGGLRGYAPRRQLANTPAKDSTGKEVARPNNYAG 60
 HS16B_WHEAT      ---------------------------------------------MSIVRRTNVFD- 11
 Q8PNC2_XANTHOM   --------------------------------------------MNIVRYPQFPTH 12

 HSPB1_HUMAN      --DPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPGYVRPLPPAAIESPAVAAPAYS 74
 HSPB6_RAT        ASAPLPGFSTPGRLFDQRFGEGLLEAELASLCPAAIAPYYLR---------------APS 59
 CRYAB_HUMAN      PFF---PFHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRP--------------PSF 54
 CRYAA_BOVINE     TLG---PF-YPSRLFDQFFGEGLFEYDLLPFLSSTISPYYRQ--------------SLF 53
 Q8UUZ6_Z.FISH    TLG------YPTRLFDQFFGGLFDYDLFPFTTSTVSPYYRH--------------SLF 51
 HSPS_M.JANASCHII PFD-----SLFERMFKEFFATPMTGTTMI--QSSTGIQISGK----------------- 40
 Q970D9_S.TOKADII ----------M-----YYLGKELQK-----RSEELSRGFYEL----------------- 22
 Q97AM1_T.VOL     -MY-----TPI-----KFFTNEMIK-----NVSNTVKEVSSF----------------- 26
 HSP26_YEAST      ALY-----DPRDETLDDWFDNDLSLFPSGFGFP-------RS----------------- 90
 HS16B_WHEAT      PFA-----DLWADPFDTFR----SIVPAISGGGSET---AAF----------------- 41
 Q8PNC2_XANTHOM   ALQ-----NEIKHVFDRFF----EQ----NGDTDESAVVTAQ---------------- 41

                                β2          β3            β4        β5
 HSPB1_HUMAN      RALSRQLSSGVSEIRHTADRWRVSLDVNHFAPD---ELTVKTKDG-VVEITGKHEERQDE 130
 HSPB6_RAT        VAL------PTAQVPTDPGYFSVLLDVKHFSPE---EISVKVVGD-HVEVHARHEERPDE 109
 CRYAB_HUMAN      LRAPSWFDTGLSEMRLEKDRFSVNLDVKHFSPE---ELKVKVLGD-VIEVHGKHEERQDE 110
 CRYAA_BOVINE     RTVLD---SGISEVRSDRDKFVIFLDVKHFSPE---DLTVKVQED-FVEIHGKHNERQDD 106
 Q8UUZ6_Z.FISH    RNILDSSNSGVSEVRSDREKFTVYLDVKHFSPD---ELSVKVTDD-YVEIQGKHGERQDD 107
 HSPS_M.JANASCHII ------GF-MPISIIEGDQHIKVIAWLPGVNKE---DIILNAVGD-TLEIRAKRSP-LMI 88
 Q970D9_S.TOKADII ------VY-PPVDMYEEGGYLVVVADLAGFNKE---KIKARVSGQNELIIEAEREITEPG 72
 Q97AM1_T.VOL     ------IY-PPVTLYQDSSDLVLEAEMAGFDKK---NIKVSVNKN-VLTISAER-KREYS 74
 HSP26_YEAST      ------VA-VPVDILDHDNNYELKVVVPGVKSKKDIDIEYHQNKNQ-ILVSGEIPSTLNE 142
 HS16B_WHEAT      ------AN-ARMDWKETPEAHVFKADLPGVKKE---EVKVEVEDGNVLVVSGERTK-EKE 90
 Q8PNC2_XANTHOM   ------WV-PRVDIKEEVNHFVLYADLPGIDPS---QIEVQMDKG-ILSIRGERKS-ESS 89
                              .  :  .    .:           .:        : :  ..

                                β6          β7            β8       β9
 HSPB1_HUMAN      -----HGYIS----RCFTRKYTLPPGV--DPTQVSSSLSPEGTLTVEAPMPKLA--TQSN 177
 HSPB6_RAT        -----HGFIA----REFHRRYRLPPGV--DPAAVTSALSPEGVLSIQATPASAQAS---- 154
 CRYAB_HUMAN      -----HGFIS----REFHRKYRIPADV--DPLTITSSLSSDGVLTVNGPRKQ----VSGP 155
 CRYAA_BOVINE     -----HGYIS----REFHRRYRLPSNV--DQSALSCSLSADGMLTFSGPKIPSGVDAGHS 155
 Q8UUZ6_Z.FISH    -----HGYIS----REFHRRYRLPSNV--DQSAITCTLSADGLLTLCGPKT-SGIDAGRG 155
 HSPS_M.JANASCHII TESERIIYSEIPEEEEIYRTIKLPATV--KEENAS-AKFENGVLSVILPKAESSIKKG-- 143
 Q970D9_S.TOKADII -----VKYLT-QRPKYVRKVIRLPYNVA-KDAEIS-GKYENGVLTIRIPIAGTSV----- 119
 Q97AM1_T.VOL     T-----VYID-QRVDKVYKVVKLPVEIE--QQDIS-AKYSEGILTVRMKTKNIKN----- 120
 HSP26_YEAST      ESKDKVKVKE-SSSGKFKRVITLPDYPGVDADNIK-ADYANGVLTLTVPKLKPQKDGKNH 200
 HS16B_WHEAT      DKNDKWHRVE-RSSGKFVRRFRLLEDA--KVEEVK-AGLENGVLTVTVPKAEVKKPEVKA 146
 Q8PNC2_XANTHOM   TETERFSRIE-RRYGSFHRRFALPDSA--DADGIT-AAGRNGVLEIRIPKRPAATPRR-- 143
                              .  :  :            .         :* * .

 HSPB1_HUMAN      EITIPVTFESRAQLGGPEAAKSDETAAK    205
 HSPB6_RAT        ---LPSPPAAK----------------    162
 CRYAB_HUMAN      ERTIPITREEKPAVTAAPKK--------    175
 CRYAA_BOVINE     ERAIPVSREEKPSSAPSS----------    173
 Q8UUZ6_Z.FISH    DRTIPVTREDKSNSGSSS----------    173
 HSPS_M.JANASCHII ---INIE--------------------    147
 Q970D9_S.TOKADII ---IKI---------------------    123
 Q97AM1_T.VOL     ---VEIE--------------------    124
 HSP26_YEAST      VKKIEVSSQESW--GN-----------    214
 HS16B_WHEAT      ---IQISG-------------------    151
 Q8PNC2_XANTHOM   ---IQV-G----NGQN-TS--GSTVQ--    158
```

Figure 3.4: MSA of Tpv-Hsp14.3 and Characterized sHSPs. Sequences by order: Human HSPB1 (Hsp27), Rat HSPB6 (Hsp20), Human HSPB5 (αB-Crystalline, Hsp20), CRYAA Bovine (αA-Crystalline), Z.fish (Zebrafish αA-Crystalline), *M.janaschii* (MjShsp16.5), *S.tokadii* (StShsp14.0), T.vol (*T.volcanium* TvShsp14.3), Yeast (Hsp26), Wheat (WtHsp16.9), Xanthom. (*Xanthomonas* Hsp18). Except for *M.janaschii*, all the rest have 9β strands. TvHsp14.3 secondary structure was based on JPRED prediction. MSA generated by Clustal Omega (EBI).

As it is seen in figure 3.5, the accuracy of sequence alignment when only the α-crystalline domains of the proteins are aligned increases. In the published literature, there are no extensive functional analysis on the residues involved in sHSPs dimerization, but structural analysis have indicated certain important amino acids (Basha *et al.*, 2011, Mchaurab *et al.*, 2009).

According to the published literature for human αB-Crystalline (CRYAB) 3D structure, the important residues involved in dimer formation are as follows: D80, D109, R116, F118 and R120 residues from one β6+7 strand interact with the residues of the opposite β6+7 strand to form a dimer. D109 forms a salt bridge with R120 and R107 forms a salt bridge with D80 in both sides of the dimer (Jehle *et al.*, 2010, Baldwin *et al.*, 2011, Jehle *et al.*, 2009, Clark *et al.*, 2011, Baldwin *et al.*, 2011). (These interactions are shown in the figure 3.8). As for the other residues, the experimental data are not conclusive, and moreover no hydrogen bond could be computed by Chimera software. Regarding the residues involved in higher oligomeric structure Jehle *et al.* (2010) studied the changes in the oligomer interaction coupled to changes in pH in human αB-Crystalline. As a result they found that the residues of the I/L-X-I/L motif at C-terminus of one dimer may interact with a hydrophobic pocket formed between β4 and β8 strands from another dimer in different geometries to give different higher oligomeric structures such as 12-mers or 24-mers. The residues forming this pocket are T132, I133 and T134 from β4 and T158, I159, I161 and T162 from β8 (Jehle *et al.*, 2010).

The loops between the beta strands change the position of the residues in sequence alignment and reduce the alignment accuracy. To increase the accuracy of the query residues matching critical amino acids possibly involved in function, the next step has been to determine their position in secondary structure, or to predict the three dimensional structure of tpv-Hsp14.3 query protein and then superimpose it to the already resolved sHSP structures. This approach is explained in the next section.

Figure 3.5: MSA for Alpha Crystalline Domain of Tpv-Hsp14.3 and the Characterized sHSPs. Sequences by order: Rat HSPB6 (Hsp20), Human HSPB1 (Hsp27), Human HSPB5 (αB-Crystalline, Hsp20), CRYAA Bovine (αA-Crystalline), Z.fish (Zebrafish αA-Crystalline), Yeast (Hsp26), *Xanthomonas* (sHsp18), Wheat (WtHsp16.9), *M.janaschii* (MjShsp16.5), *S.tokadii* (StShsp14.0), *T.volcanium* (TvShsp14.3), TvHsp14.3 secondary structure was based on JPRED prediction. MSA generated by Clustal Omega (EBI).

89

3.1.4. The Three-Dimensional Structure of Wild Type Protein

The 3D structure generated for the wild type tpv-Hsp14.3 by I-TASSER is shown in figure 3.6. The tpv-Hsp14.3 models generated had all the signatures of the conventional α-Crystalline domain, with the exception of β6 strand which was not shown in the graphic form. This strand was predicted by the secondary structure prediciton program JPRED as shown in figure 3.3 and also by I-TASSER. The reason why it was not shown graphically by Chimera and PhyMol (not shown here) is not known.

However, it can be speculated that the reason why β6 was not shown graphically may be the fact that it is not a stable strand in Class II sHSPs. Studies in other sHSPs suggest that it is transient state, i.e., the secondary structure changing from β-strand to loop according to the conditions of substrate binding or multimerization (Kim *et al.* 1998; Basha *et al.* 2011).

Another important feature of the predicted Tpv-Hsp14.3 structure is that it dimerizes not like the vertebrate sHSPs, but like the Class II sHSPs, which are found and characterized in plants and Archea. Such a dimerization mode is shown in figure 1.14. A possible dimerization model of tpv-Hsp14.3 is shown in figure 3.7. It should be noted that this is only a predicted dimeric structure formed by superimposing the tpv-Hsp14.3 monomers one at a time on StHsp14.0 structure of *S.tokadii*, which is the closest in sequence identity to *T.volcanium*. The dimeric structure could not be refined to reduce the global energy due to computational constraints. However it serves the purpose of showing how the Tpv-Hsp14.3 monomers may interact with each other and form the basic functional unit of the sHSPs which is the dimer
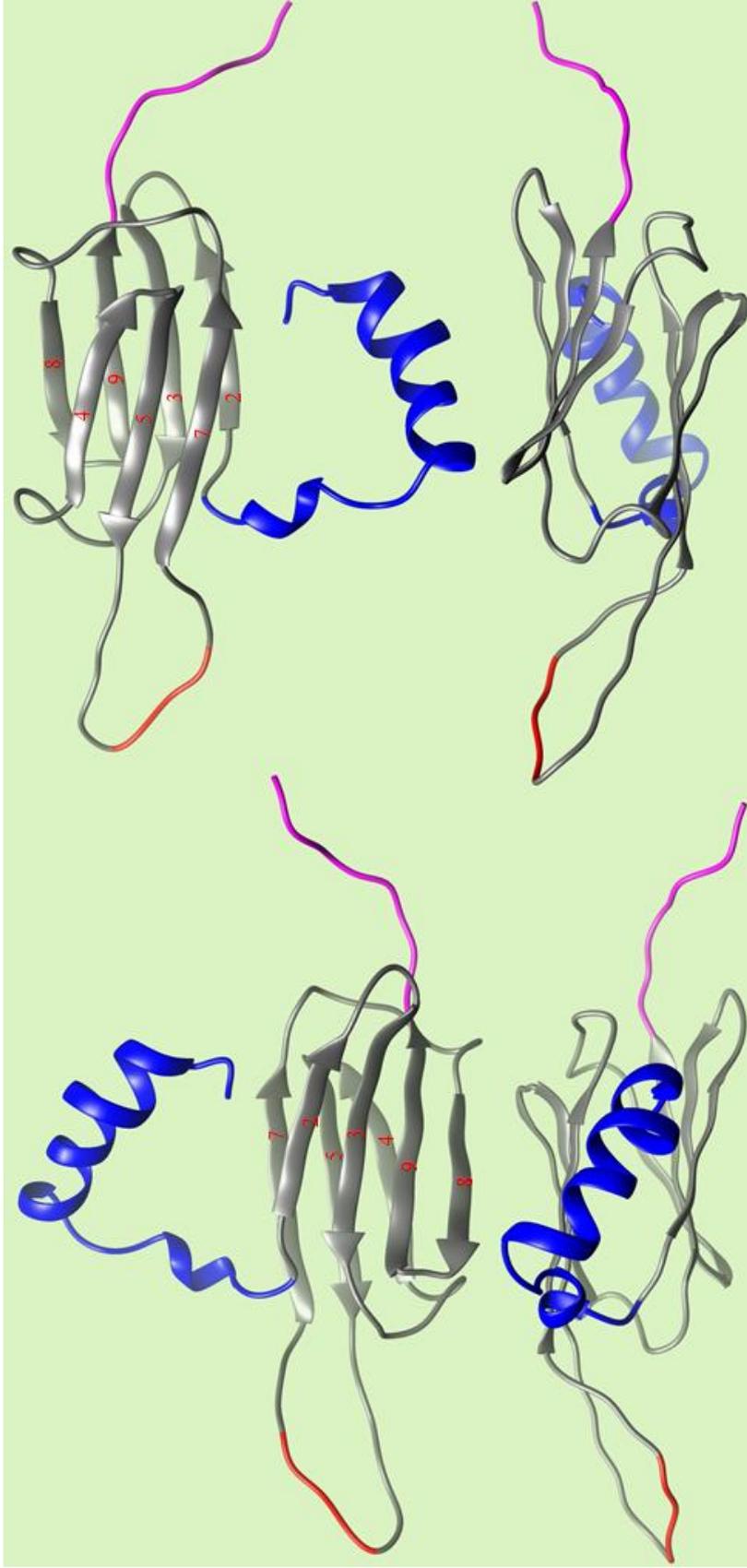
Figure 3.6: The Tpv-Hsp14.3 Three-Dimensional Structures Generated by I-TASSER Metaserver. The protein model is shown from four different angles. Blue color represents N-Terminal region, Grey and red colors represent the α-crystalline, and the pink color represents the C-terminus. Beta strands are labeled according to the conventional way, starting from β2. The predicted β6 was not represented in the model generated by I-TASSER, but it was predicted as a secondary structure and is present in the other homologs, so it is highlighted with red color. Images generated by UCSF Chimera.
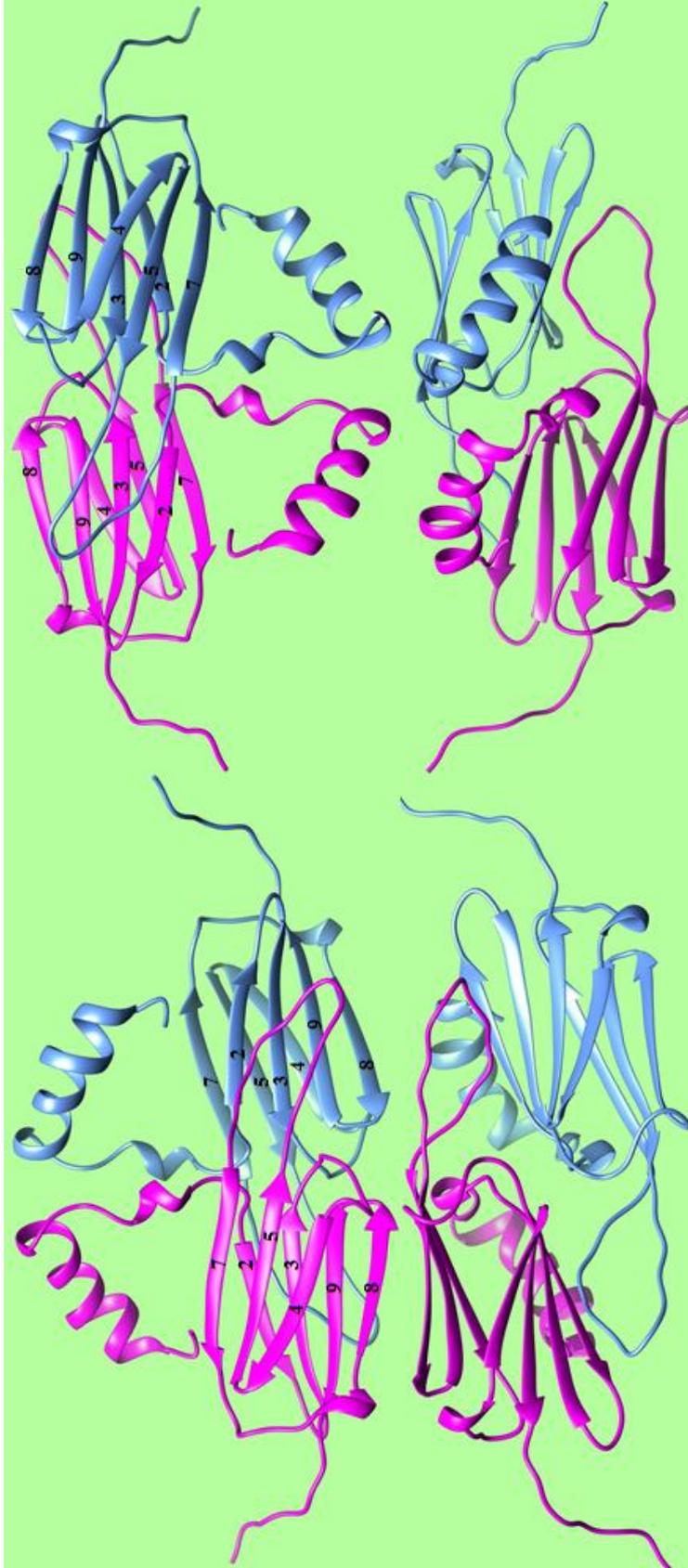
Figure 3.7: A Possible Dimerization Model of Tpv-Hsp14.3. The dimer complex is shown from four different angles. Each monomer is shown in magneta and cyan colors respectively. The numbers of β-strands are given according to the conventional way of sHSPs starting from β2. β6 strand is not shown for the reasons explained previously. Images generated by UCSF Chimera software

3.1.5. Comparison of Tpv-Hsp14.3 Model Structure with Already Resolved sHSP Structures.

After structure prediction, the models can be superimposed on already resolved structures of the proteins mentioned in section 3.1.3. The best characterized sHSP to date is the Human αB-Crystalline. However, there is a drawback in the structure superposition with this protein because Human sHSP dimerization mode is different from the query since the former is Class I and the latter Class II sHSPs. A structure of this protein and its important residues involved in dimer formation are shown in figure 3.8 (Jehle *et al.*, 2010). These residues include: D80, D109, R116, F118, R120 and R123. To see which residue of Tpv-Shsp14.3 is the counterpart of any one of these residues in human αB-Crystalline, local pair-wise sequence alignment was performed by ClustalW (figure 3.9). The reason of choosing local sequence alignment is the fact that it will give better results by getting rid of the unnecessary gaps introduced as a result of longer sequences of human, as expalined in section 3.1.
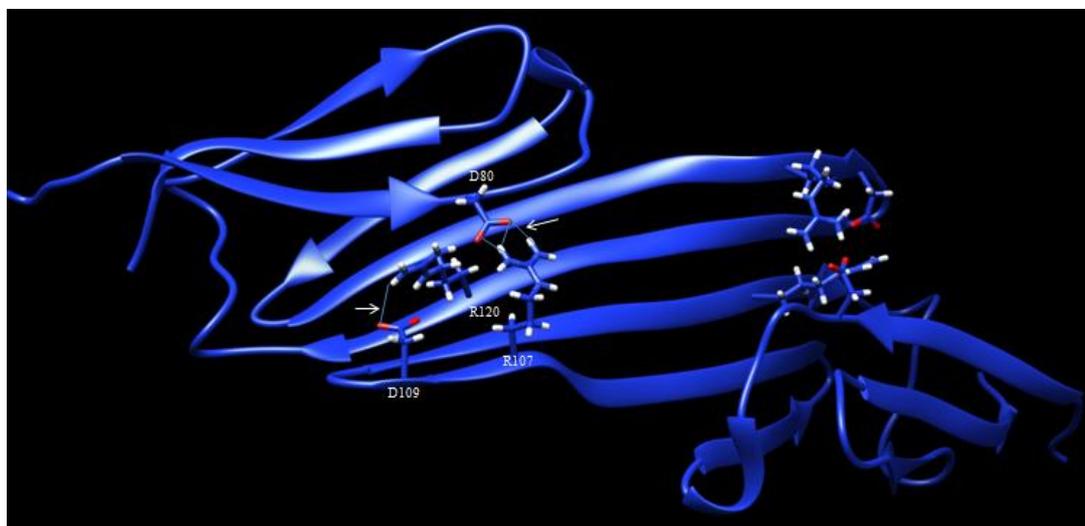
Figure 3.8: Human αB-Crystallin Dimer Interaction Sites. All the residues mentioned above are not shown for reasons of clarity. Hydrogen bonds are shown and marked by arrows. The same bonds are formed between the same residues on the other side of the protein but they were not shown again for reasons of clarity. Image generated by Chimera. Human structure accession number is 2KLR (Jehle *et al.*, 2010).

```
THEVO   6 KFFTNEMIK-NVSNTVKEVSSF-IYPPVTLYQDS------SDLVLE----  43
          :||...::: ::..|...:|.| :.||..|...|       |::.||
HUMAN  26 QFFGEHLLESDLFPTSTSLSPFYLRPPSFLRAPSWFDTGLSEMRLEKDRF  75


THEVO  44 ---AEMAGFDKKNIKVSVNKNVLTISA---ERKREYSTVYIDQRVDKV-YK  87
             .:::..|..:.:||.|..:|:.:..   ||:.|:.  :|. |...:.|:
HUMAN  76 SVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHG--FIS-REFHRKYR 123


THEVO  88 VVKLPVEIEQQDI-SAKYSEGILTVRMKTKNIKNVE  122
          :    |.::::...| |:..|:|:|||....|.:...|
HUMAN 124 I---PADVDPLTITSSLSSDGVLTVNGPRKQVSGPE  156
```

Figure 3.9: Pairwise Local Sequence Alignment Between Tpv-Hsp14.3 and Human αB-Crystallin. Important residues in human protein and their similar or identical counterparts in *T.volcanium* are bold and underlined. Sequence alignment generated by ClustalW.

The local sequence alignment shows four residues of tpv-Hsp14.3 matching four of the important residues of human αB-Crystalline. They are E45 matching with D80, R69 matching with R107, I78 with I114, R81 with R120, and K87 with R123. Glutamic Acid matching with Aspartic Acid and Lysine with Arginine are not surprising given their similar biochemical properties and their interchangeable nature.

The next step to refine the results even more and to obtain position-specific sequence alignment, is to superimpose the human sHSP and *T.volcanium* sHSP structures on each other, and detect the positions of the query that match with the functionally important residues of the human protein. This is done by UCSF Chimera imposing the predicted PDB query structure on the PDB structure found in the protein databank. When the structure of tpv-Hsp14.3 model was superimposed on human αB-Crystalline structure, the results turned out to be different as shown in figure 3.10. Except for tpv-Hsp14.3 E45 positions matching human D80, all the rest of highlighted residues in figure 3.9 were different from the matches in pair-wise sequence alignment. Moreover, a new position turned out to match the critical human R120 which is involved in dimerization and whose mutation causes many diseases (Clark *et al.*, 2011). It matches with *T.volcanium* K87 position. The reason for this is the "greedy" nature of sequence alignment to reduce the number of gaps which
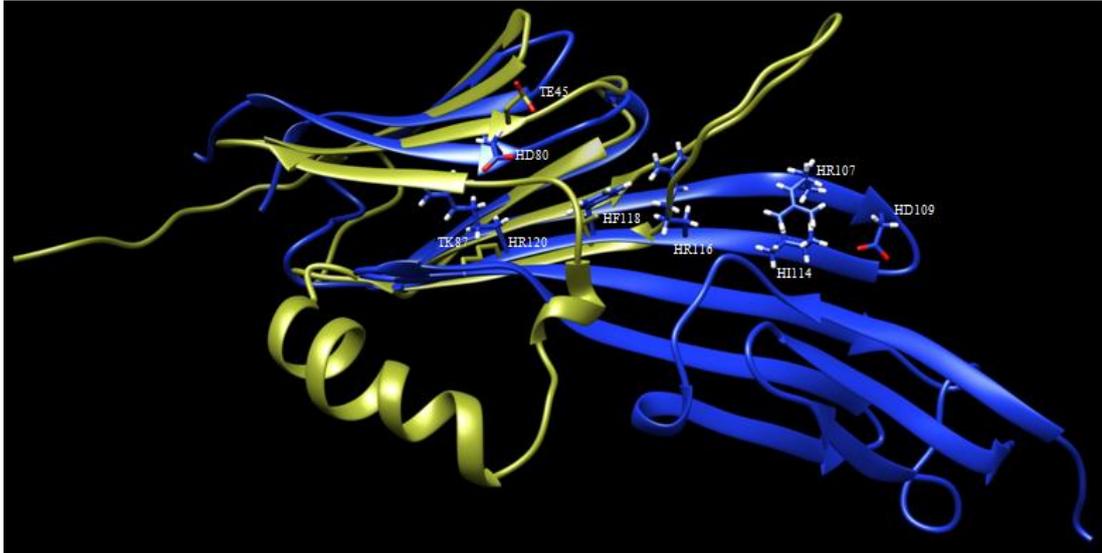
Figure 3.10: Tpv-Hsp14.3 Superimposed on Human αB-Crystalline. The query model structure is shown in orange and human structure in blue color. The residues are also labeled, with H standing for human, and T standing for T.volcanium before each amino acid residue. Image generated by UCSF Chimera. Human protein accession number is 2KLR (Jehle *et al.*, 2010)

would be many when T.volcanium and Human proteins are compared due to their differences in loop lengths.

Keeping in mind that the dimerization modes of human and *T.volcanium* proteins are very different from each other, attention was then turned to Wheat sHSP16.9 structure. It is the only fully-determined structure to date but few potential residues involved in dimer formation are studied experimentally (von Montfort *et al.* 2001). One of them is R108 which is the counterpart of human R120. It is proposed to interact with E100, but such an interaction is not conserved in other sHSPs. Two other important interactions for dimer formation based on computational studies with Chimera are R101 and R85 forming two salt bridges with D60 (figure 3.11).

Figure 3.11: Important Interactions for Wheat sHSP16.9 Dimerization. The bonds are marked by arrows and the residues are named. Image generated by UCSF Chimera. Accession number of the structure is 1GME.

Following the same logic as with the human αB-Crystalline, first local sequence alignment between wheat and *T.volcanium* sHSP sequences was carried out. The results are shown in figure 3.12.

```
THEVO   12 MIKNVSNTVKEVSSFIYPPVTLYQDSSDLVLEAEMAGFDKKNIKVSVNK- 60
              ::..:|....|.::|...:.:..:.....|.:|::.|..|::||.|..
WHEAT   27 IVPAISGGGSETAAFANARMDWKETPEAHVFKADLPGVKKEEVKVEVEDG 76

THEVO   61 NVLTISAERKREYSTVY-----IDQRVDKVYKVVKLPVEIEQQDISAKYS 105
              |||.:|.||.:|.....      :::...|..:.:|.:.:.:::.|...
WHEAT   77 NVLVVSGERTKEKEDKNDKWHRVERSSGKFVRRFRLLEDAKVEEVKAGLE 126

THEVO  106 EGILTVRMKTKNIKNVEIE   124
              .|:|||.:....:|..|::
WHEAT  127 NGVLTVTVPKAEVKKPEVK   145
```

Figure 3.12: Pairwise Local Sequence Alignment between Tpv-Hsp14.3 and Wheat sHSP16.9. Important residues in wheat protein and their similar or identical counterparts in *T.volcanium* are bold and underlined. Sequence alignment generated by ClustalW.

The local alignment shows matches between R69 with R85, R81 with R101, and K87 with R108 from *T.volcanium* and Wheat sHSPs respectively. *T.volcanium* R81 is not an exact match for wheat R101 but it can be assumed to be given the nature of sequence alignment. Moreover, E45 from *T.volcanium* matches D60 from wheat sHSPs, and these two residues have similar chemical properties.

Continuing with the same methodology as the one used for human sHSP structural comparison, a similar comparison was done with wheat, where the *T.volcanium* and wheat sHSP structures were superimposed on each other to see what residues of the query matched with the wheat sHSP16.9 residues proposed to be important in dimer formation. These structures superpositions are shown in figures 3.13 and 3.14.
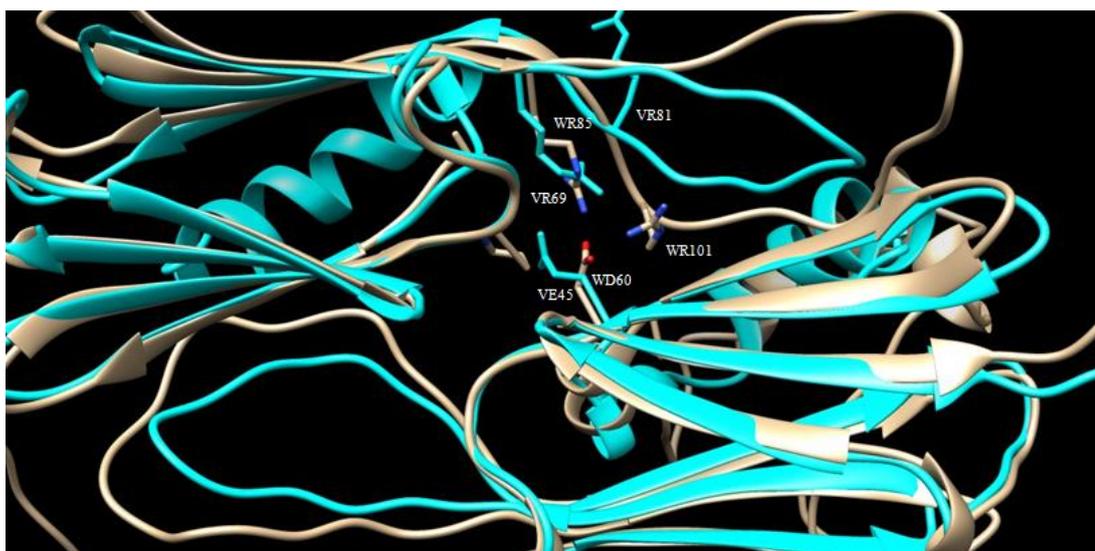
Figure 3.13: *T.volcanium* and Wheat sHSPs Structure Superposition. Wheat structure is grey and *T.volcanium* cyan. Wheat residues are named with letter W followed by their one-letter code and number, while *T.volcanium* residues by V followed by their one letter code and number Image generated by UCSF Chimera. Accession number of the wheat sHSP16.9 structure is 1GME.
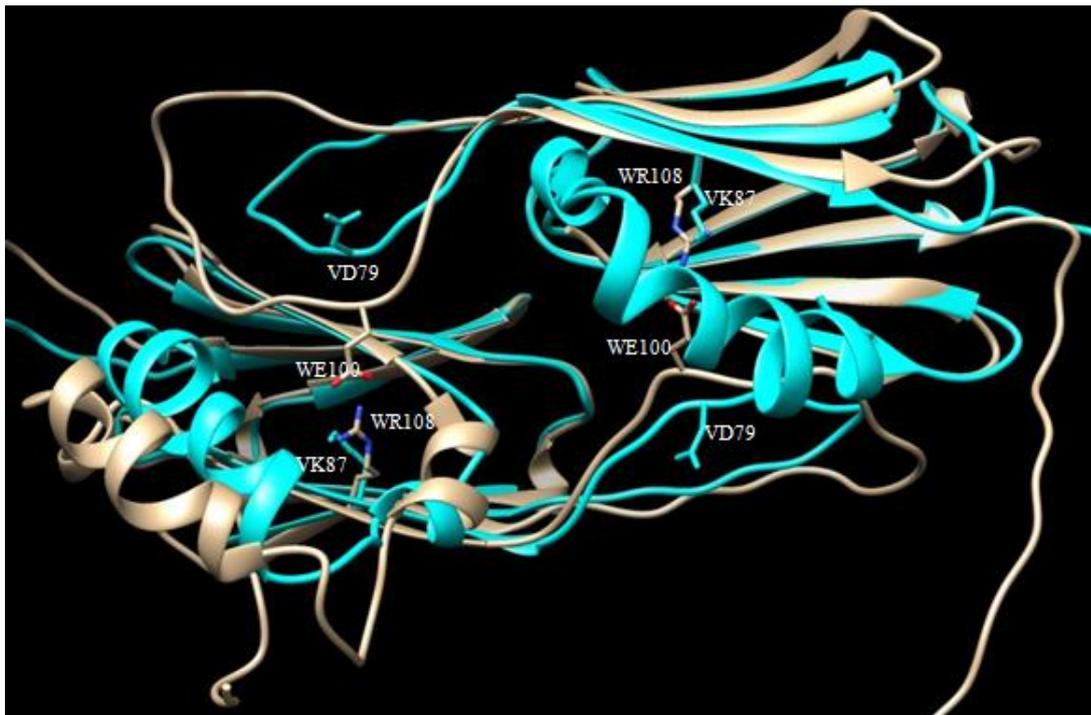
Figure 3.14: Model Structure Superposition on Wheat WtShsp16.9. Wheat structure is grey and T.volcanium is cyan. Wheat residues are named with letter W followed by their one-letter code and number, while T.volcanium by V followed by their one letter code and number Image generated by UCSF Chimera. Accession number of the wheat Shsp16.9 structure is 1GME.

Figure 3.13 shows that *T.volcanium* sHSP14.3 residues match well with some of the wheat conserved residues. Of them, VR69 matches WR85, VE45 matches WD60 and VR81 matches WR101, but the orientations of the last pair are opposite in 3D space.

Figure3.14 shows that VK87 matches with WR108 and VD79 with WE100. Again the orientation in 3D space of the second residue is not as in wheat, but it may be a structure prediction artifact. So to summarize, VE45 matches with WD60, VR69 with WR81, VD79 with WE100, VR81 with WR101, and VK87 with WR108. These results also confirmed the results of local sequence alignment in figure 3.12.

Another superimposition is between the structures of *S.tokadii* sHSP (closely related to Tpv-Hsp14.3) and wheat sHSP structure. This is done to find similar residues in *S.tokadii* sHSP protein by comparing it to better characterized wheat sHSP because no functional studies exist to date for the former except its 3D structure (Takeda *et al.* 2011). As figure 3.15 shows, there is not a perfect fit of both monomers when *S.tokadii* and wheat 3D structures are compared with each other. This is as a result of a difference in dimerization angle that the monomers of each protein form when they come together to form a dimer.

This angle is proposed to be important in the number of monomers involved in the formation of the multimeric structure, as explained in section 1.3.5. For this reason, one pair of monomers from each protein match very well but the other pair does not fit.
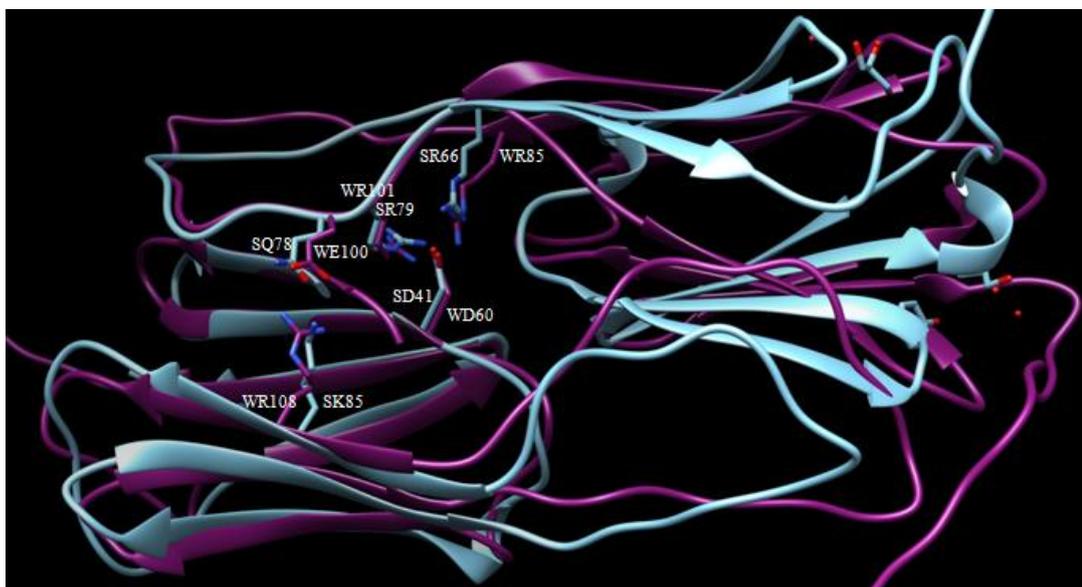
Figure 3.15: Wheat and *S.tokadii* sHSPs 3D Structure Superpositions. Wheat structure is pink and S.tokadii is cyan. Wheat residues are named with letter W followed by their one-letter code and number, while S.tokadii by S followed by their one letter code and number Image generated by UCSF Chimera. Accession number of *S.tokadii* Shsp14.0 structure is 3AAB and that of wheat Shsp16.9 id 1GME.

The figure shows β-sandwich which is the core of α-Crystalline domain fits perfectly between the two proteins, with the exceptions of the wheat loops which are very long. The residues SR66 matches with WR85, SR79 with WR101, SD41 with WE60, SK85 with WR108 and SQ78 matches with WE100. The residues are almost exactly the same as the matches of *T.volcanium*, with the exception of SQ78 instead of VD79, and SD41 instead of VE45, but these two residues have similar properties and therefore exhangable.

Then, pairwise local sequence alignment was performed between *T.volcanium* and *S.tokadii* sHSPs to see whether these residues matches can be observed at sequence level.These results in figure 3.16 show a perfect match between them.

```
THEVO     7 FFTNEMIKNVSNTVKEVSSFIYPPVTLYQDSSDLVLEAEMAGFDKKNIKV 56
            :...|:.|......:.....:||||.:|::...||:.|::|||:|:.||.
SULTO     3 YLGKELQKRSEELSRGFYELVYPPVDMYEEGGYLVVVADLAGFNKEKIKA 52

THEVO    57 SVN-KNVLTISAERK-REYSTVYIDQRVDKVYKVVKLPVEIEQQ-DISAK 103
            .|: :|.|.|.|||: .|....|:.||...|.||::||..:.:. :||.|
SULTO    53 RVSGQNELIIEAEREITEPGVKYLTQRPKYVRKVIRLPYNVAKDAEISGK 102

THEVO   104 YSEGILTVRMKTKNIKNVEIE 124
               |..|:||:|:......::||
SULTO   103 YENGVLTIRIPIAGTSVIKIE 123
```

Figure 3.16: Pairwise Local Sequence Alignment between *T.volcanium* and *S.tokadii* sHSPs. Important residues in *S.tokadii* protein and their similar or identical counterparts in *T.volcanium* are bold and underlined. Sequence alignment generated by ClustalW.

Next, structure comparison between tpv-Hsp14.3 model and *S.tokadii* sHSP14.0 was performed by structure superposition and checked wheather the above residues match with each other or not. For clarity, a structure of *S.tokadii* sHSP14.0 is shown alone in figure 3.17.
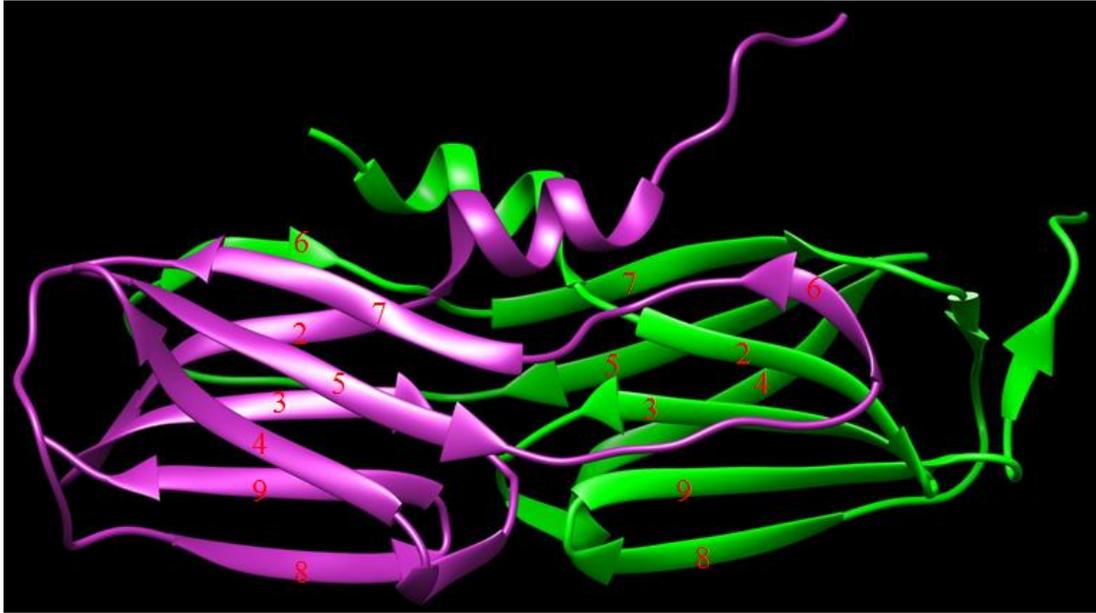
Figure 3.17: 3D Structure of *S.tokadii* sHSP14.0. β-strands are numbered according to the conventional way. Image was generated by Chimera.

This figure shows the structure of *S.tokadii* sHSP14.0 which resembles a conventional Class II sHSPs with an N-terminal region, the α-Crystalline domain made of eight β-strands, and a short C-terminus, in which a shorter 9th strand called β10 strand is found. However, whether this β10 is unique to *S.tokadii* sHSP14.0 or its an experimental artefact is a matter of dispute and it was discussed in section 1.3. The structure is shorter than the original protein because it was not possible to obtain crystal data from the full-length protein and as a result, some residues of the N-terminal region are not shown (Takeda *et al.* 2011).

Figure 3.18 shows the structure superposition of *S.tokadii* and *T.volcanium* sHSPs, and the results are in agreement with local sequence alignment in figure 3.16. VE45 matches with SD41, and Glu and Asp have similar biochemical properties, both being acidic and able to form hydrogen bonds. SR66 matches with VR69, SQ78 with VQ80, SR79 with VR81, and SK85 with VK87. The positions of the residues of *T.volcanium* protein in 3D space do not fit with *S.tokadii* residues due to the fact that
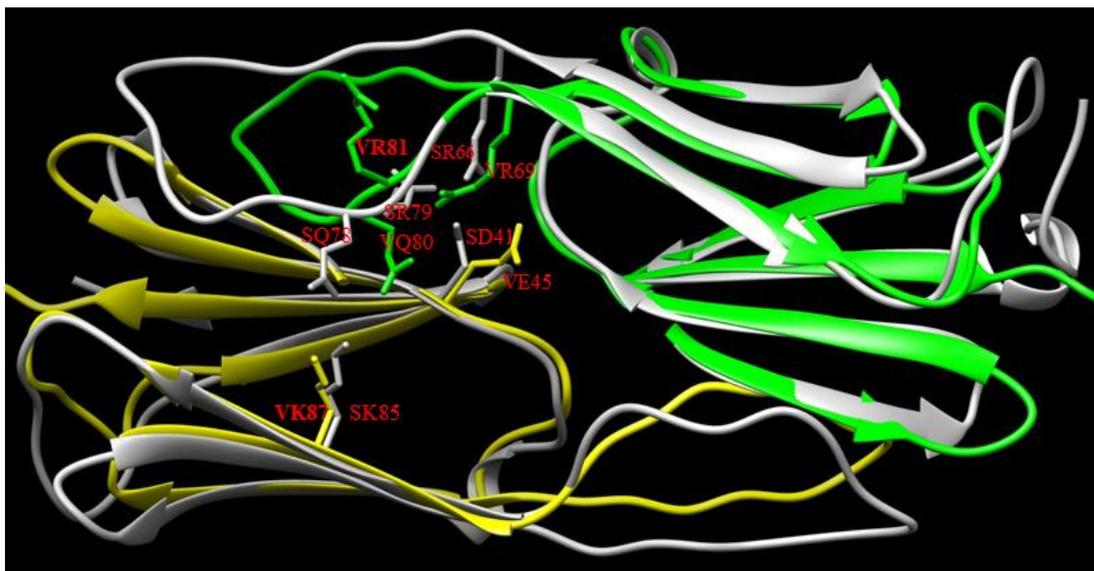
Figure 3.18: Structure Superposition of Tpv-Hsp14.3 Model on S.tokadii sHSP14.0. *S.tokadii* structure dimer is grey, while for the model, one monomer is yellow and the second green. *S.tokadii* residues are named with letter S followed by their one-letter code and number, while *T.volcanium* by V followed by their one letter code and number. The N-terminus of the proteins were removed to make the view clear. Image generated by UCSF Chimera. Accession number of *S.tokadii* sHSP14.0 structure is 3AAB.

the protein is a model, but their location in similar positions is a strong argument in the favor of their importance in sHSPs dimerization.

The last structure to be compared is that of another archeon, *M.Janaschii*. A figure of the dimeric structure is shown in figure 3.19. This is also a class II sHSP, making it's structure very relevant to this particular study. However, structural studies have shown that it has ten β-strands, not nine like most of Class I chaperons. For this reason the numbering of β-strands starts from 2 in sHSPs because this was the first sHSP to be crystallized and whose structure was resolved (Kim *et al.* 1998).
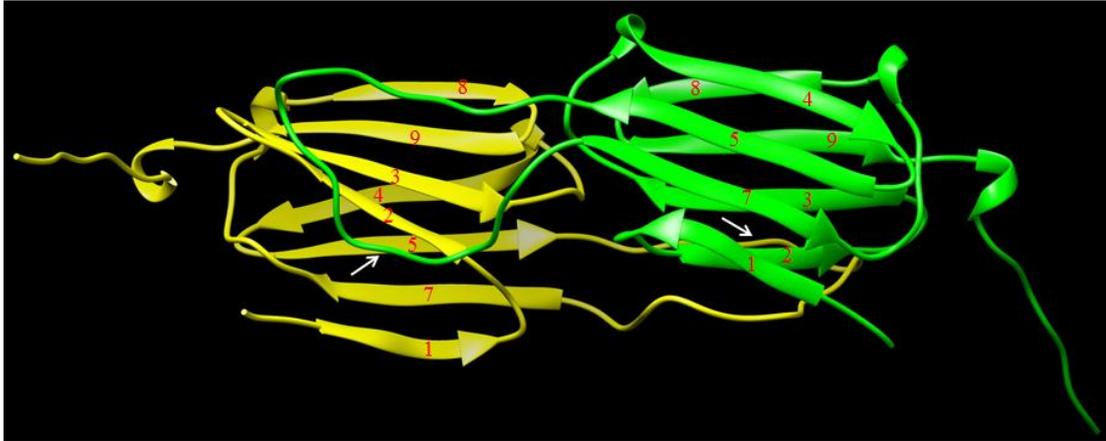
Figure 3.19: 3D Structure of *Mjanaschii* sHSP16.5. Arrows show the locations of β6 strands. Accession code is 1SHS. Image was generated by Chimera.

A structural comparison between *M.janaschii* sHSP16.5 and wheat sHSP16.9 revealed common residues like WR85 and MR83, WE100 and ME98, WR108 and MR107, but there were also differences in MW41 as equivalent of WD60, and MI99 as equivalent of WR101 (figure 3.20). These last two residues are very different from each other. Searching for hydrogen bonds in MjHsp16.9 by chimera didn't reveal any interactions between the residues shown in figure 3.20. This means that a different set of interactions are involved in *M.janaschii*. Structural studies by Kim *et al.* (1998) who crystallized this protein revealed interaction for multimer formation but were not concerned with interactions involved in dimer formation.

Keeping in line with the previous comparisons, pairwise local sequence alignement for *M.janaschii* and *T.volcanium* sHSPs sequences was performed and results shown in figure 3.21.

Figure 3.20: Wheat sHSP16.9 and *M.janaschii* Shsp16.5 Structure Superposition. Wheat structure is cyan and *M.janaschii* is yellow color. Wheat residues are named with letter W followed by their one-letter code and number, while *M.janaschii* by M followed by their one letter code and number. Some parts of the wheat N-terminus was deleted for clarity. Image generated by UCSF Chimera. Accession number of *M.Janaschii* sHSP16.5 structure is 1SHS and wheat sHSP16.9 is 1GME.

```
THEVO   17  SNTVKEVSSFIYPPVTLYQDSSDLVLEAEMAGFDKKNIKVSVNKNVLTIS     66
            |:|..::|...:.|:::.:....:.:.|.:.|.:|::|.:..::.|.|.
METJA   31  SSTGIQISGKGFMPISIIEGDQHIKVIAWLPGVNKEDIILNAVGDTLEIR     80


THEVO   67  AERKREYST-----VYID-QRVDKVYKVVKLPVEIEQQDISAKYSEGILTV  111
            |:|.....|       :|.: ...::|:.:|||..::::::.|||:...|:|:|
METJA   81  AKRSPLMITESERIIYSEIPEEEEIYRTIKLPATVKEENASAKFENGVLSV  131
```

Figure 3.21: Pairwise Local Sequence Alignment between *T.volcanium* and *M.janaschii* sHSPs. Residues in *M.janaschii* protein matching with the residues designated as important in *T.volcanium* are bold and underlined. Sequence alignment generated by ClustalW.

In this case, VE45 matched with MW53, R69 with MR83, and VK87 with MR107. The region VQ80 and VR81 did not match, and these results are in agreement with structure comparison of wheat and *M.janashcii* sHSPs shown above. To refine the results, *T.volcanium* model structure was superposed on M.janaschii sHSP structure as shown in figure 3.22.

As in the case of wheat, there is MI99 in instead of VR81 and MW41 instead of VE45, while VK87 is the counterpart of MR107 and VQ80 the counterpart of ME98. These results are also in agreement with pairwise sequence alignment results, except for residues VQ80 and VR81.
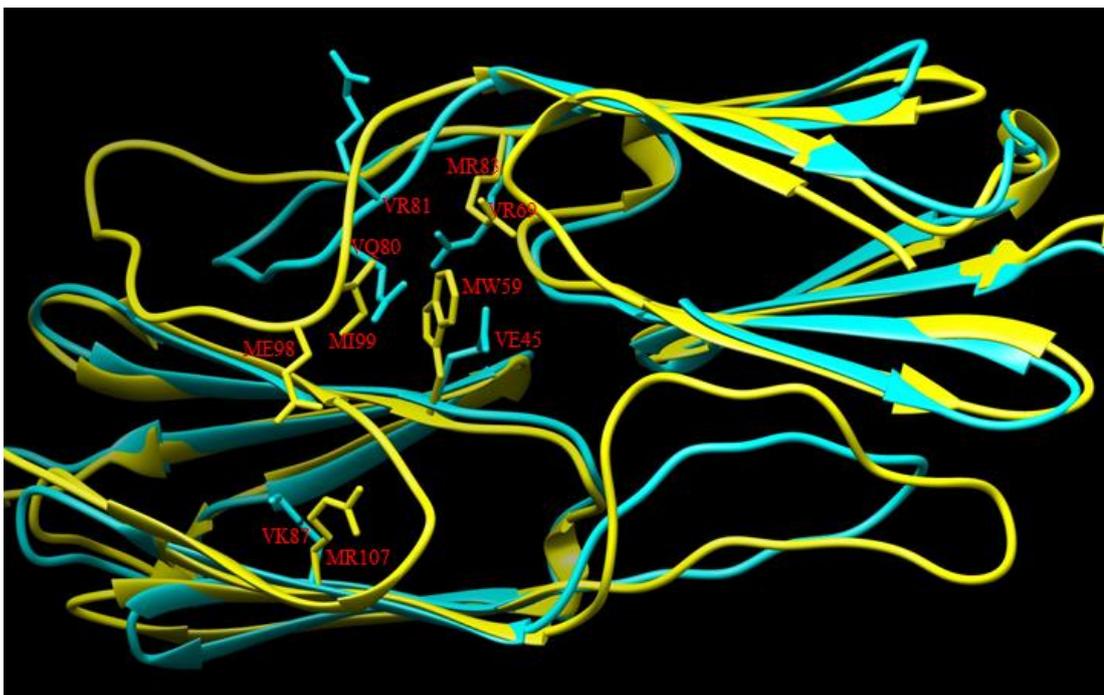
Figure 3.22: Structure Superposition of Tpv-Hsp14.3 Model on M.janaschii sHSP14.5. *M.janaschii* sHSP16.5 is yellow while the *T.volcanium* sHSP14.3 is cyan color. T.volcanium resiues are named with letter V followed by their one-letter code and number, while M.janaschii by M followed by their one letter code and number. Image generated by UCSF Chimera. Accession number of *M.Janaschii* sHSP16.5 structure is 1SHS.

After these studies, it was decided to change four sites in the α-crystalline domain of *T.volcanium* tpv-Hsp14.3 sequence. The sites were decided by relying mostly on local sequence alignments and verified by structure comparison. These sites are R69, R81 and K87 as single mutations, and QR80-81 as a double mutation. Since all the sites were positively charged, with the exception of Q80, which is polar positive but uncharged amino acid, it was reasoned that the best way was to change each single site by an analogous residue, a negatively charged residue, and by a hydrophobic residue. So R69 was changed into R69K, R69E and R69M. R81 was changed into R91K, R81E and R81M. K87 was changed into K87R, K87E and K87I, and the double mutant QR80-81 was changed into QR(80-81)EL.

For all the mutant variants structures, were predicted as well in the same way as the wild type. Based on the principles of structure prediction by threading, there is not expected any significant change in the overall structure of any of the mutants as compared to the wild-type. To further prove this assumption, the structures of wild-type and each mutant variants were superposed on each other and the root-mean-square deviation (RMSD) values were calculated by Chimera software. RMSD is a standard measurement used in protein structure evaluation. It measures the average distance between the backbone atoms of superposed proteins (Maiorov *et al.* 1994). RMSD values for one-to-one structure comparisons of wild-type and each mutant were 0.6-0.9. In another case, the 3D structures all mutant variants for a mutated position were compared to each other and the wild-type structure and their RMSD values were calculated. These values were between 0.9-1.2. All these values are below the threshold RMSD value of 2, which is the standard value for a good fit between protein structures. Based on these values, it can be claimed there are no gross changes in the structures when a residue or two are mutated.

3.2. Plasmid Isolation and Restriction Digestion

The recombinant *tvn0775* gene was previously expressed heterologously in *E.coli* cells in our laboratory (Kocabiyik *et al.* 2012). Digestion of recombinant plasmid *PQE31/775* with *SmaI* or *HindIII* linearizes the vector, while double digestion with
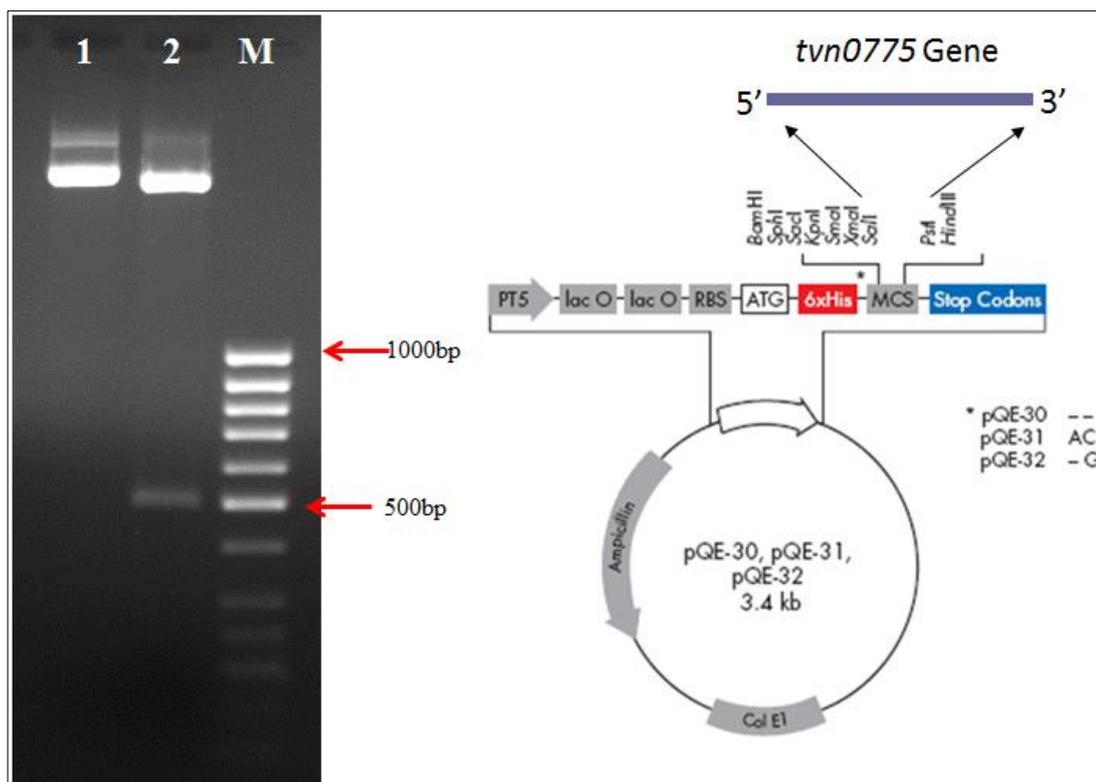
Figure 3.23: Agarose Gel Electrophoresis of the Recombinant *PQE31/775* vector. Single and double digestions are shown on the left. Vector digested with (1) SmaI, (2) SalI and PstI (3) GeneRuler 50bp DNA Ladder (Fermentas, Lithuania). In the right, schematic diagram of the expression vector construct and inserted gene position.

SalI and *PstI* excises the gene. The original sequence of the gene is 372 nucleotides, but together with addittionl sequences from the multiple cloning site (MCS) of the vector, the size of the subcloned fragment became 447 nucleotides. The expression vector size with the gene is about 3.9 kb: 3463b plasmid plus 447 b of the inserted gene. This vector has a lac Operon and ampicillin resistance gene as selection marker. The results of single and double digestions carried out to confirm the presence of the cloned gene are shown in figure 3.23.

In the figure, both recombinant plasmid digestion and vector construct are shown. When the plasmid is digested with SalI only, it is linearized, and when it is digested with SmaI/PstI enzymes, the gene is excised as shown by the lower band in the

second well with a length of approximately 500 base pairs. This confirms the presence of the gene in the recombinant plasmid.

3.3. Site Directed Mutagenesis

Five target sites for mutagenesis were selected based on sequence alignment analysis explained in section 3.1. The logic of the mutagenesis was to change the sites in three different ways, an analogue of the original residue with similar physico-chemical properties, and two others would be very different: one would be acidic instead of the original basic residue, and the other would be hydrophobic, since all the selected residues except Gln were charged. So, Arginine-69 (R69), a basic residue, was converted it into its analogue Lysine (K) (R69K), into Glutamic Acid (E) (R69E), which is an acidic residue, and into Methionine, a hydrophobic residue. The next selected site was Arg81, which was again changed into Lys (R81K), Glu (R81E) and the bulky hydrophobic residue Met (R81M). Another residue was Lys87 which was changed into Arg (K87R), Glu (K87E) and Ile (K87I). The last mutagenesis was a double mutation at positions Gln80 and Arg81. These two amino acids were simultaneously changed into Glu80 and Leu81, respectively. The original and mutated sequences for each case are shown in the figures 3.24-3.26.

```
gtactcactataagtgcggagagaaagagagaatactctaccgtatatatcgatcagcgc  243 Wild-type
 V   L   T   I   S   A   E   R   K   R   E   Y   S   T   V   Y   I   D   Q   R    81

gtactcactataagtgcggagaTGaagagagaatactctaccgtatatatcgatcagcgc  243 R69M
 V   L   T   I   S   A   E   M   K   R   E   Y   S   T   V   Y   I   D   Q   R    81

gtactcactataagtgcggagGAaaagagagaatactctaccgtatatatcgatcagcgc  243 R69E
 V   L   T   I   S   A   E   E   K   R   E   Y   S   T   V   Y   I   D   Q   R    81

gtactcactataagtgcggagaAGaagagagaatactctaccgtatatatcgatcagcgc  243 R69K
 V   L   T   I   S   A   E   K   K   R   E   Y   S   T   V   Y   I   D   Q   R    81
```

Figure 3.24: Part of Tpv-Hsp14.3 Gene and its Rrotein Sequence for R69 Mutants. Both DNA and protein sequences show the flanking regions of R69 residue (colored). The changed nucleotides in the codons are capitalized.

108

```
accgtatatatcgatcagcgcgttgacaaagtgtataaagtagttaagctgcccgtagag  282  Wild-type
 T   V   Y   I   D   Q   R   V   D   K   V   Y   K   V   V   K   L   P   V   E   94


accgtatatatcgatcagAAGgttgacaaagtgtataaagtagttaagctgcccgtagag  282  R81K
 T   V   Y   I   D   Q   K   V   D   K   V   Y   K   V   V   K   L   P   V   E   94


accgtatatatcgatcagATGgttgacaaagtgtataaagtagttaagctgcccgtagag  282  R81M
 T   V   Y   I   D   Q   M   V   D   K   V   Y   K   V   V   K   L   P   V   E   94


accgtatatatcgatcagGAGgttgacaaagtgtataaagtagttaagctgcccgtagag  282  R81E
 T   V   Y   I   D   Q   E   V   D   K   V   Y   K   V   V   K   L   P   V   E   94


accgtatatatcgatGaCcTcgttgacaaagtgtataaagtagttaagctgcccgtagag  282  QR(80-
 T   V   Y   I   D   E   L   V   D   K   V   Y   K   V   V   K   L   P   V   E   94  81)EL
```

Figure 3.25: Part of Tpv-Hsp14.3 Gene and its Protein Sequence for R81 and QR(80-81)EL Mutants. Both DNA and protein sequences show the flanking regions of residues 80-81 (colored). The changed nucleotides are capitalized.

```
Accgtatatatcgatcagcgcgttgacaaagtgtataaagtagttaagctgcccgtagag  282  Wild-type
 T   V   Y   I   D   Q   R   V   D   K   V   Y   K   V   V   K   L   P   V   E   94


AccgtatatatcgatcagcgcgttgacaaagtgtataTagtagttaagctgcccgtagag  282  K87I
 T   V   Y   I   D   Q   R   V   D   K   V   Y   T   V   V   K   L   P   V   E   94


AccgtatatatcgatcagcgcgttgacaaagtgtataGagtagttaagctgcccgtagag  282  K87R
 T   V   Y   I   D   Q   R   V   D   K   V   Y   R   V   V   K   L   P   V   E   94


AccgtatatatcgatcagcgcgttgacaaagtgtatGaagtagttaagctgcccgtagag  282  K87E
 T   V   Y   I   D   Q   R   V   D   K   V   Y   E   V   V   K   L   P   V   E   94
```

Figure 3.26: Part of Tpv-Hsp14.3 Gene and its Protein Sequence for K87 Mutants. Both DNA and protein sequences show the flanking regions of residue K87 (colored). The changed nucleotides are capitalized.

3.3.1. Mutagenesis with Transformer^{TM} Site-Directed Mutagenesis Kit

First, mutagenesis experiments were carried out with Transformer^{TM} Site-Directed Mutagenesis Kit (Clontech, Takara Japan), and Kit's instructions were followed. After the first transformation with *pUC19M* vector, (a 2686 bp plasmid), the transformation efficiency was $3x10^5$ and the percentage of blue cells was in the range of 65%, both of them lower than the expected $1x10^6$ and 70-90%. The transformation efficiency when the amount of plasmid DNA was doubled turned out to be $1x10^5$. This might be related to the competence state of the host cells.

For the mutagenesis experiments with *pQE321/775* plasmid, the first two sites selected were R69M and R69E. The primers used were as described in the chapter of "Materials and Methods", and the amount of plasmid was 0.1μg in all mutation reactions. Selection was done by cutting the plasmid in mutagenesis mixture with 2μl of 10U/μl *HindIII* enzyme (Roche, Zwitzerland). Transformation efficiency for R69E mutation was $1x10^3$, while that of R69M was $3x10^3$. When primers concentration was quadrupled, the transformation efficiency for R69E was $4x10^3$ and that of R69M was $3x10^3$. These efficiencies are 100-fold lower than the experiments performed with control plasmid.

To further elaborate on the cause of the low efficiency, the same cells under the same conditions were transformed with pUC19 vector, which has a size of 2686 bp. When the cells were transformed with 0.25μg of this vector, the efficiency was $1x10^5$ that is close to the one obtained with test plasmid. When the cells were transformed with wild-type *PQE31/775* recombinant vector, the number of cells transformed with linear plasmid was 3-4% (higher than expected) of those transformed with undigested plasmid. This result indicated that the transformation efficiency with the recombinant plasmid would be low. According to the troubleshoting section of the Kit, this high ratio of cells transformed with linear plasmid as compared to the circular one may be due to inefficient digestion, but it was shown that HindIII enzyme digests the plasmid completely (figure 3.40). The only explanation for this seems to be the size of the plasmid, since the PUC vectors with a size of ~2.7kb transformed with a higher efficiency than our vector with a size of ~3.9kb. The claim that low plasmid purity decreases the transformation efficiency is also rejected, because the same plasmid was used with the second kit for mutagenesis, and it showed very good results. Additionally, the same method was used to isolate plasmid for DNA sequencing, prooving that the plasmid is pure.

3.1.1.1. Characterization of mutants by Agarose Gel Electrophoresis after Restriction Digestion

When the sequences were changed according to the desired mutations, the insert within the recombinant vector were scanned for new or lost restriction sites due to

Table 3.1: Restriction Enzyme Cut Sites after Mutation at Specific Sites.

Site	RE	Recognition	Freq.	Cut Positions	Fragment Length
WT	MboII	GAAGA	8	1283, 1896, 2687, 2758, 3513, 3591, 3700, 3896	**1295**, 613, 791, 71, 755, 78, 109, 196
R69M	MboII	GAAGA	9	397, 1283, 1896, 2687, 2758, 3513, 3591, 3700, 3896	**409, 886**, 613, 791, 71, 755, 78, 109, 196
R69E	MboII	GAAGA	9	397, 1283, 1896, 2687, 2758, 3513, 3591, 3700, 3896	**409, 886**, 613, 791, 71, 755, 78, 109, 196
R69K	MboII	GAAGA	10	394, 397, 1283, 1896, 2687, 2758, 3513, 3591, 3700, 3896	**3, 409, 886**, 613, 791, 71, 755, 78, 109, 196
WT	BseRI	GAGGAG	1	115	Linear 3890
R69E	BseRI	GAGGAG	2	115, 395	**280**, 3610
WT	BccI	CCATC	7	137, 143, 702, 1377, 2962, 3086, 3373	6, **559**, 390, 675, 1585, 124, 287, 674
R81M	BccI	CCATC	8	137, 143, 412, 702, 1377, 2962, 3086, 3373	6, **269, 290**, 390, 675, 1585, 124, 287, 674

mutations. After the scan, the sites were determined and the fragment length calculated. Restriction sequence scan was done by RestrictionMapper Version 3 and the new sites formed after vector insertions are shown in table 3.1.

No restriction sites were found to be lost after mutation. The sites of class I restriction Enzymes were excluded due to their restriction site ambiguity. From the enzymes in table 3.1 only *MboII* and *BseRI* were used for verification of the mutants. The highlighted fragments are the longer fragments in the wild-type gene and the two new sub-fragments were yielded by digestion with the selected restriction enzymes after mutations. Plasmid isolation for the cells mutated in positions R69M and R69E was done by QIAprep Spin Miniprep Kit (Qiagen) as described in section 2.3.1.2. The first thing to be observed was whether the *HindIII* restriction site had been really changed or not, since it was expected to be mutated by the selection primer in the recombinant plasmid used in this study. *HindIII* digested plasmids from selected putative mutant colonies were run in 1.5% AGE as described in section 2.3.3 and the gel photo is shown in figure 3.27.
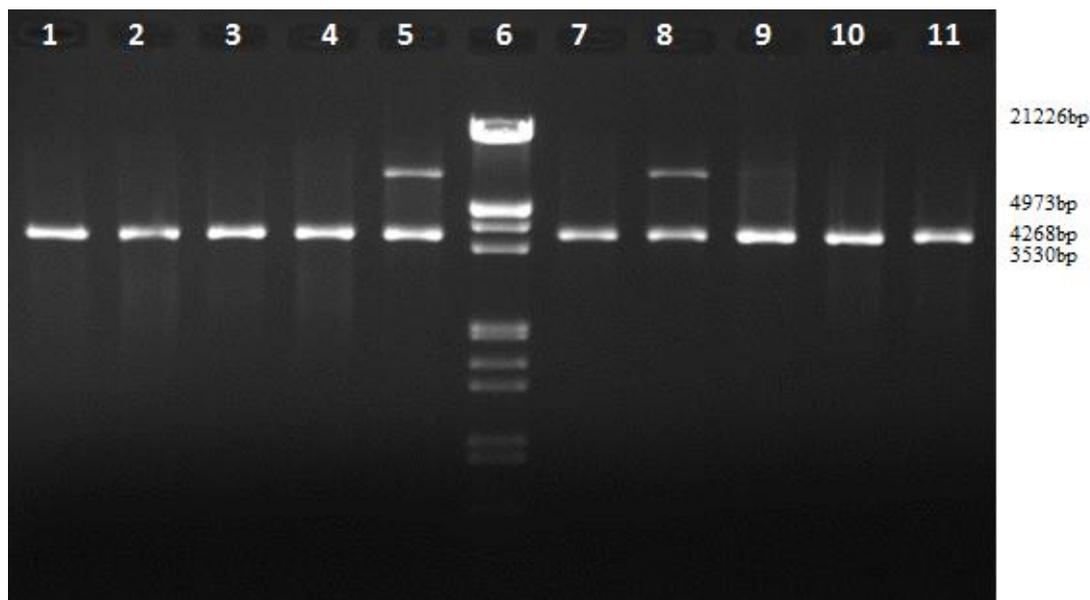
Figure 3.27: AGE of Plasmid after Mutagenesis Digested with *HindIII*. The samples were run in 1.5% AGE. Sample order is: Lanes (1)-(4): plasmid isolated from colonies 2-5 for the site R69M. (5): colony 10 of R89M. (6): Lambda DNA/*EcoRI+HindIII* marker (Fermentas, Lithuania). (7): Wild-type plasmid. (8) and (9): plasmids from colonies M11 and M14 respectively. (10) and (11): plasmids from colonies 1 and 4 of R69E mutants. The sizes for the four bands of the marker are shown in the right of the figure.

All the plasmids digested with *HindIII* enzyme at one site were linearized, meaning that *HindIII* cut site was not mutated. However, plasmids from colonies 8 and 11 (lanes 5 and 8) were not digested complitely, or may be that two plasmids fused to form a concatamer. This result shows the inefficiency of mutation at the selection site.

3.3.1.2. Plasmid Isolation from Mutant Cells and AGE Characterization.

Since the mutation of *HindIII* enzyme cut site was independent of the mutation at R69 site, among the putative mutant colonies some were randomly selected for characterization by restriction enzyme digestion targeting the mutant forms of the vector when R69 residue was changed into R69M and R69E. The expected result for

112

R69M is the loss of a long 1295bp band because it would be cut by *MboII* into two bands of lengths 409 and 886 base pairs (table 3.1). As for R69E, the plasmid is supposed to give two bands after digesting it with *BseRI* enzyme of lenghts 280bp and 3610bp respectively. On the other hand, the wild-type would only be linearized. The gel photo for R69M putative mutant colonies is shown in figure 3.28.
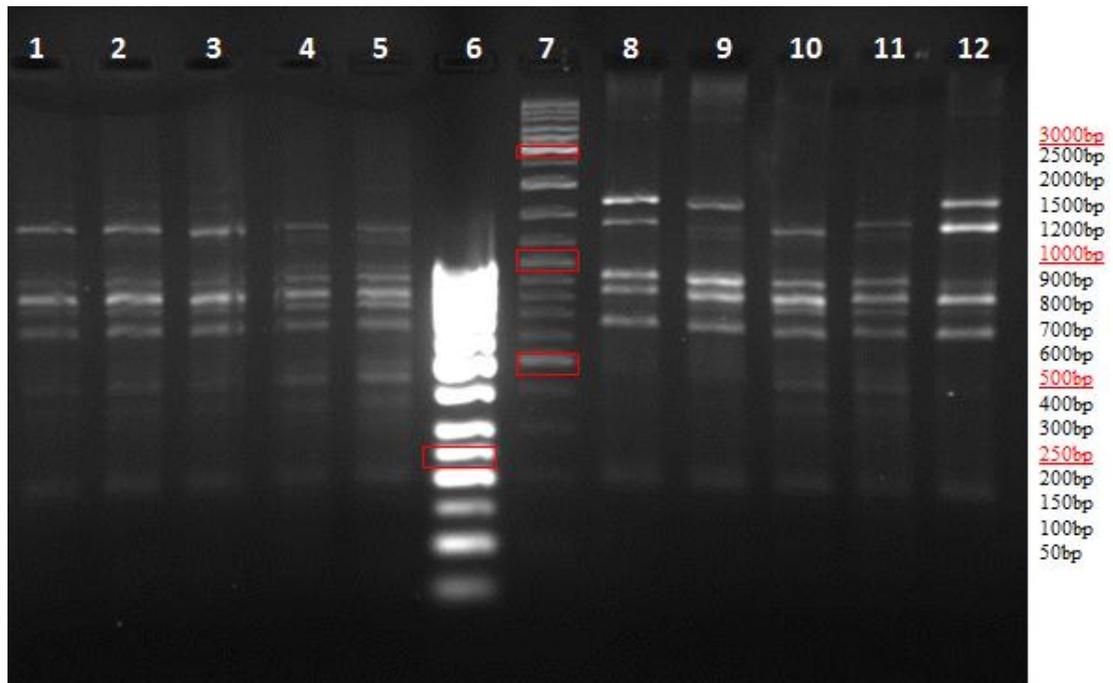
Figure 3.28: AGE of Plasmid Isolated from R69M Colonies. 1.5% Agarose Gel Electrophoresis run for 10 different plasmids isolated from single cell colonies. Lanes (1)-(5): colonies 1-5. (6): GeneRuler 50bp DNA Ladder(Fermentas, Lithuania). (7): GeneRuler DNA Ladder Mix 10000-100 (Fermentas, Lithuania). (8): colony 7. (9): colony 8. (10): colony 10. (11): colony 11. (12): Colony 14. These colonies are all from the first transformation. The boxed bands in the marker lanes stand for the underlined fragment length to the right of the photo.

The large number of fragments formed by the many cut sites of *MboII* restriction endonuclease makes the discretion of the bands almost impossible. The volume of the plasmid loaded was also increased from 10 to 15µl with the intention of increasing band intensity, but still the results did not change (figure 3.29).
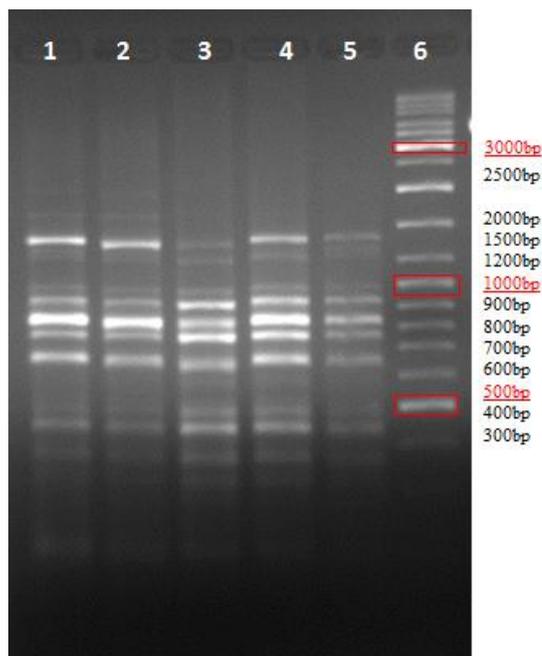
Figure 3.29: AGE of Plasmid Isolated from R69M Colonies. 1.5% Agarose Gel Electrophoresis run for 5 selected putative mutant R69M colonies. Lanes (1): colony 2. (2): Colony 5. (3): colony 10. (4): colony 14. (5): wild-tupe. (6) GeneRuler DNA Ladder Mix 10000-100 (Fermentas, Lithuania). These colonies are all from the first transformation. The boxed bands in the marker lanes stand for the underlined fragment length to the right of the photo.

Discretion of the mutant R69E would be relatively easier, since two bands are expected to be formed instead of a linear wild-type recombinant vector. But there is only one band after digestion, meaning that the sire R69 was not mutated (figure 3.30).
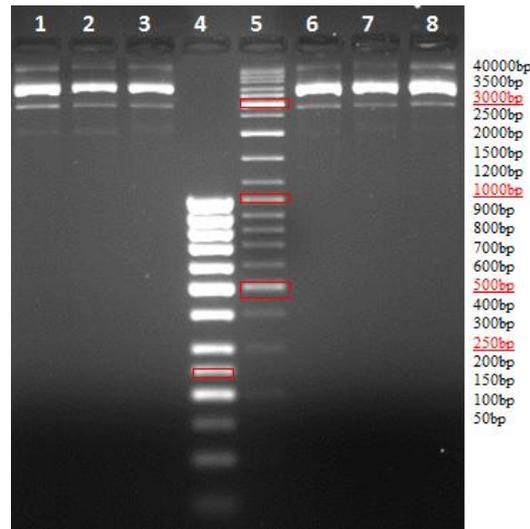
Figure 3.30: AGE of Plasmid Isolated from R69E Colonies. 1.5% Agarose Gel Electrophoresis run from 6 different R69E plasmids isolated from putative mutant single cell colonies. Lanes (1)-(3): Colonies 1-3. (4): GeneRuler 50bp DNA Ladder (Fermentas, Lithuania). (5): GeneRuler DNA Ladder Mix 10000-100 (Fermentas, Lithuania). (6)-(8): 4-6 colonies. These colonies are all from the first transformation

Another mutagenesis experiment performed for R69E and R69M substitutions yielded three promising mutant colonies, namely colonies 3 and 4 (R69M 1/3 and R69M 1/4) for R69M mutant, shown in lanes 3 and 4 in figure 3.31, and colony 3 for R69E mutant from the second transformation (R69E 2/3) (figure 3.32). Since bands of R69M 1/3 and R69M 1/4 were different than the rest, these two plasmids from R69M mutants and R69E 2/3 were sent for sequencing.

Some plasmids have bigger size than the original vector, and some bands are bigger than expected, as shown in figures 3.27 and 3.31. This may be due to mutations or plasmids forming concatamers. The sequences were shown to be mutated where there is a deletion in the plasmid, at least in the sequenced region (Appendix C), because it was not sequenced completely.
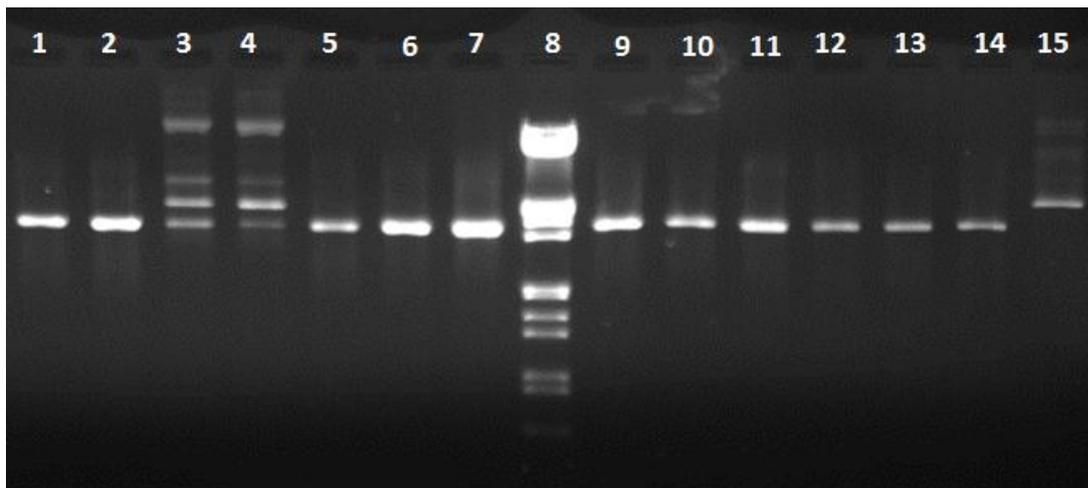
Figure 3.31: AGE of Plasmid Isolated from R69M Colonies. 1.5% Agarose Gel Electrophoresis run for 14 different plasmids isolated from single cell colonies. Lane (1): R69M 1/1 colony. (2) R69M 1/2. (3) R69M 1/3. (4) R69M 1/4. (5) R69M 2/2. (6) R69M 2/3. (7): R69M2/4 (8) Lambda DNA/*EcoRI+HindIII* marker (Fermentas, Lithuania). (9): R69M 2/5. (10): R69M 2/9. (11): R69M 2/10 (12): R69E 1/1. (13): R69E 1/3. (14): R69E 1/5. (15) wild-type undigeste plasmid. (Colonies with numbers 1/ are from the first transformation, and colonies with number 2/ from the second transformation).
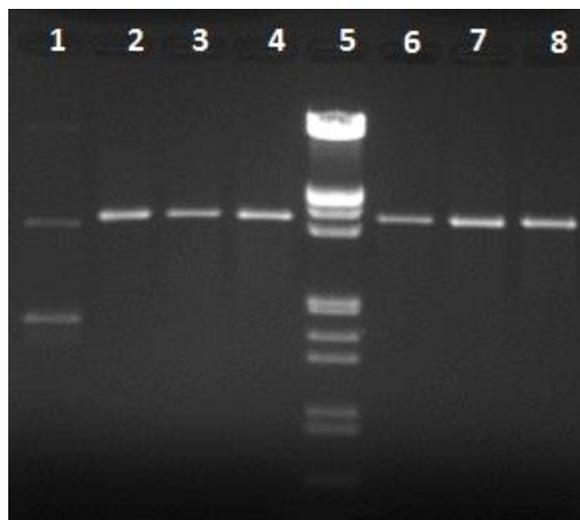
Figure 3.32: AGE of Plasmid Isolated from R69E Colonies. 1.5% Agarose Gel Electrophoresis run from 7 different R69E plasmids isolated from putative mutant single cell colonies. Lane (1): R69E1/6 colony, (2) R69E1/7. (3): R69E2/1. (4): R69E2/3. (5) Lambda DNA/*EcoRI+HindIII* marker (Fermentas, Lithuania). (6): R69E2/5. (7): R69E2/7. (8): R69E2/8. (Colonies with numbers 1/ are from the first transformation, and colonies with number 2/ from the second transformation).

Based on the low transformation efficiency, and in the inability to change the sites into the desired mutation as shown by AGE in figures 3.40-3.45, this kit was seen as unsuitable and the mutagenesis was not continued any more with it. Whatever the reason of the failure may be, three samples R69E 2/3, R69M 1/3 and R69M 1/4 were sent for sequencing as described in section 2.3.5 and the results showed an internal deletion in one sequence, no change in another one, and a mutated sequence in the third sample (Appendix C). However, when the flanking sequences outside the gene region were analyzed, there were several other mutations, including deletions. Since the Kit did not work with the plasmid system used in this study, it was decided to continue with another mutagenesis kit.

3.3.2. Site-directed Mutagenesis by QuikChange® Site-Directed Mutagenesis Kit

The optimization experiments were carried out with the *pWhitescript* 4.5bp control plasmid as described in the protocol, and the transformation efficiency obtained was $1x10^8$ CFU/μg DNA. The blue colonies made up 98.7% of the total colonies that grew after transformation. These are the required results according to the kit manual, meaning that the procedure is working fine. The mutagenesis experiments were carried out with *pQE31/775* recombinant plasmid. The best results were obtained by using a total of ~31ng plasmid DNA for transformation, with efficiencies in the order of 10^6-10^7 CFU/μg DNA, while with a total amount of 20ng DNA no colonies were obtained. When the total plasmid amount was increased to 36ng and 40ng the efficiencies were of the orders of 10^4 CFU/μg DNA. So based on these rsults, all the experiments were carried as described in the section 2.3.4.2 with a plasmid amount of 31 ng. The transformation efficiencies were calculated by the formula of CFU/μg of DNA and the results are shown in the table 3.2. Among the putative mutant colonies, 2-4 of them were selected for each mutation and the plasmids were isolated from them as described in section 2.3.1.2. Purified plasmids were then sent for sequencing, and after obtaining the results, the sequences were analyzed for the mutated sites. To check whether they had any mutation in the flanking regions, pair-wise sequence alignments with ClustalW software of EBI were performed for the

Table 3.2: Transformation Efficiencies in the Mutagenesis Experiments.

No.	Mutation	µl plated	Total Col. No.	CFU	Trans. Eff.
1	R69K	550µl	914	1662	3×10^6
2	R69E	545µl	620	1138	2×10^6
3	R69M	540µl	2260	4185	7×10^6
4	R81K	520µl	332	638	1×10^6
5	R81E	500µl	239	478	8×10^5
6	R81M	550µl	1548	2815	5×10^6
7	K87R	550µl	3646	7012	1×10^7
8	K87E	550µl	3268	5942	1×10^7
9	K87I	520µl	3768	7246	1×10^7
10	QR(80-81)EL	500µl	633	1266	2×10^6

available sequence region. Such sequence analysis for each mutation revealed that all the sites had changed, except three cases: In two cases, the expected K87R was K87I and vice versa. This may also be due to a labeling error. In the third case the expected R69K was changed into Asparagine instead of Lysine codon. All the sequences and their sequence alignments with the wild type gene are shown in Appendix C.

3.4. Expression of the Engineered sHSPs and Their Purification.

All the mutant *sHSP* genes in *pQE31* expression vector were expressed under the control of lacZ promoter/operator upon IPTG induction. After obtaining the cellular extracts, for each variant protein, two samples were prepared for examination by gel electrophoresis, one heat-treated and the other not treated. Since *T.volcanium* is a thermophilic Archaeon, it was expected that by heat treatment at 65°C, its proteins will not be affected, while those of *E.coli* will be denatured and thus can be precipitated by centrifugation at high speed. The benefit of this is the increased purity of the extract as most of the proteins will be removed by centrifugation. Then the rest when applied to column will be better purified and used for biochemical characterization.
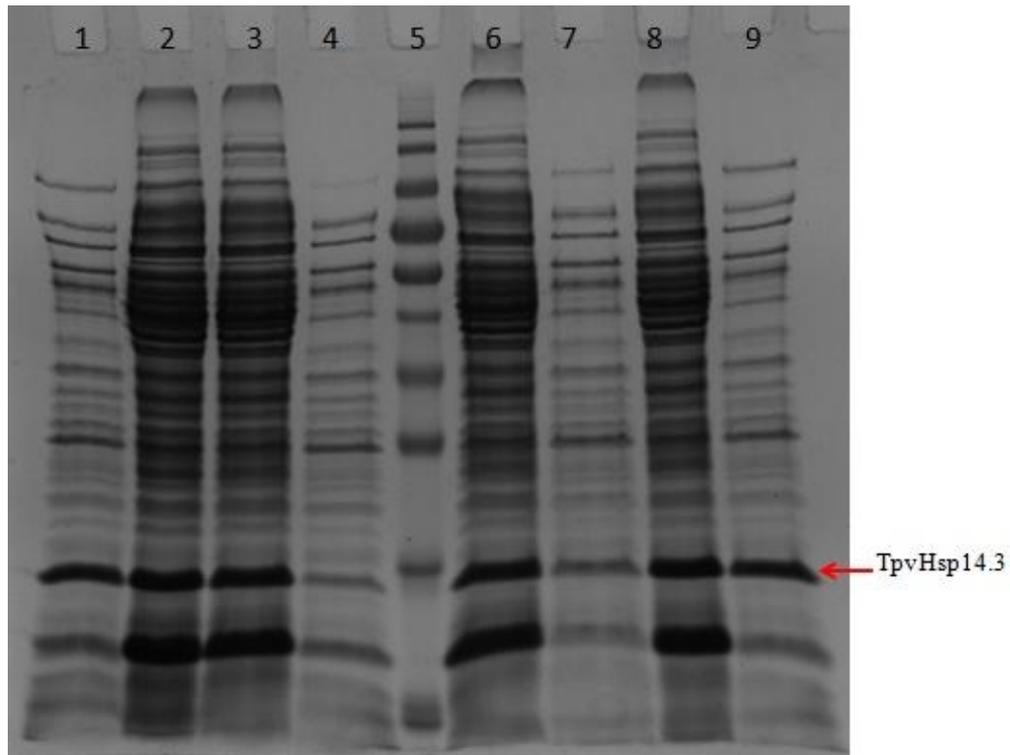
Figure 3.33: SDS-PAGE of Cell Extracts of the Wild-Type and Mutant *E.coli* Cells.
Lanes: (1) R69K mutant after heat treatment. (2) R69K mutant not heat treated. (3) R69E mutant not heat treated. (4) R69E mutant after heat treatment. (5) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (6) R69M mutant not heat treated. (7) R69M mutant after heat treatment. (8) Wild-type not heat treated. (9) Wild type after heat treatment.

As figure 3.33 shows, after heat treatment most of the bands for mutant Tpv-Hsp14.3 get weaker except the band corresponding to the wild-type Tpv-Hsp14.3 and R69K mutant. The loss in band intensity was significant for R69M and R69E. This may be explained by decreased heat stability of these variants as a result of specific mutations.

Figure 3.34 shows that the mutant variants of tpv-Hsp14.3 at position K87 are more stable than those of R69 as judged by band loss after heat treatment. The change of Lys for Arg and Glu at position 87 does not cause any change in the band strength after heat treatment. The double mutant QR(80-81)EL is also not affected much,

while the K87I mutant is very unstable. For the latter, even the band of unheated cellular extract is quite weak, and after heating, it is almost lost. This suggests that K87I variant may be very unstable.
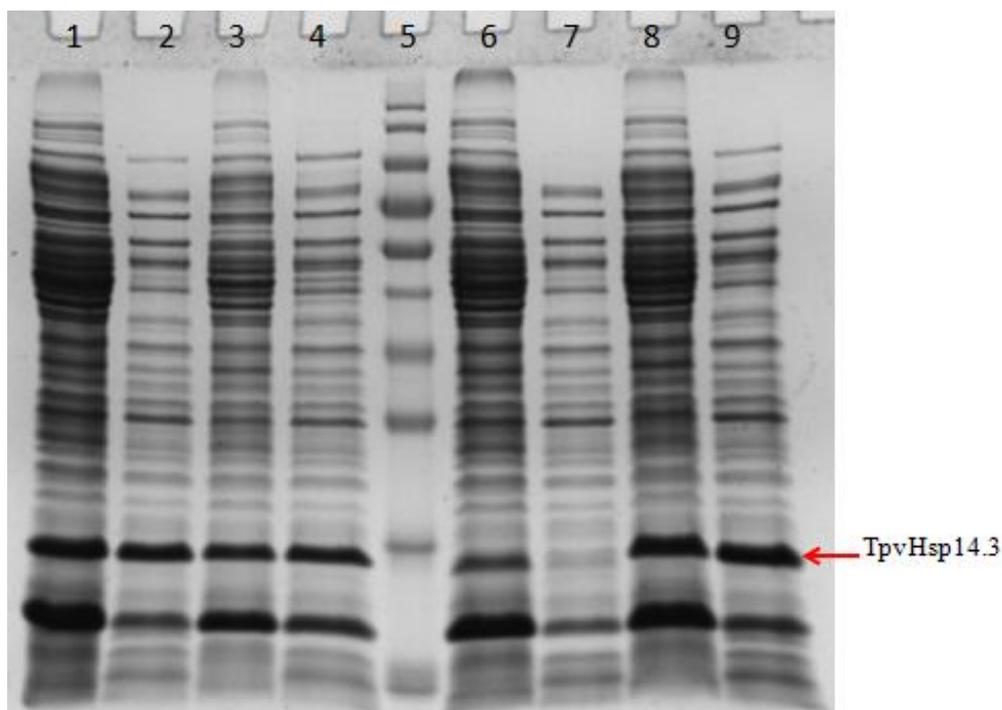
Figure 3.34: SDS-PAGE of Cell Extracts for Mutant *E.coli* Cells. Lanes: (1) K87E mutant untreated. (2) K87E mutant after heat treatment. (3) K87R mutant not heat treated. (4) K87R mutant after heat treatment. (5) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (6) K87I mutant not heat treated. (7) K87I mutant after heat treatment. (8) QR(80-81)EL mutant not heat treated. (9) QR(80-81)EL mutant after heat treatment.

Figure 3.35 shows that changing of R81 for Lys and Met resulted in no significant effect on the band strength after heat treatment, but when it is changed for Glu, some loss is observed.
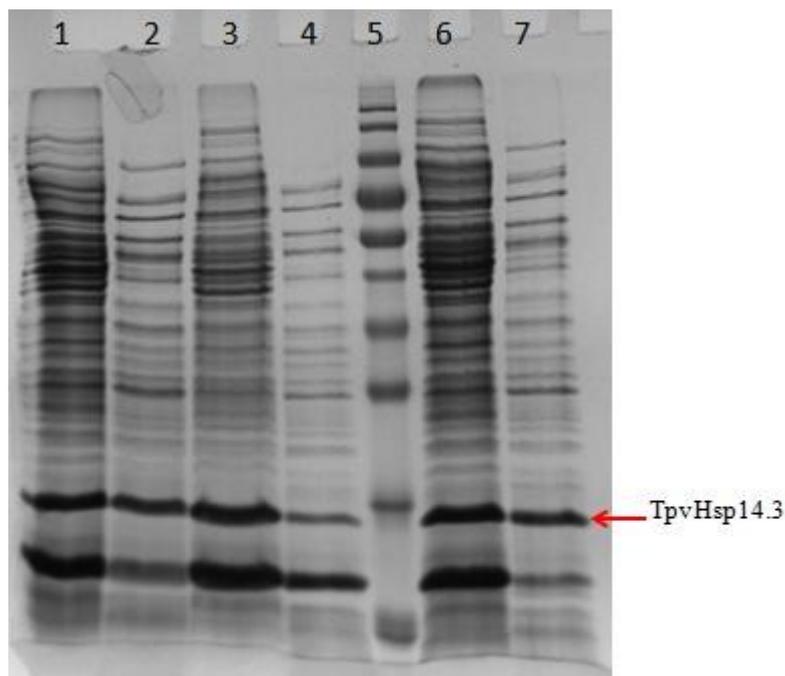
Figure 3.35: SDS-PAGE of Cell Extracts for Mutant *E.coli* Cells. Lanes: (1) R81K mutant not heat treated. (2) R81K mutant after heat treatment. (3) R81E mutant not heat treated. (4) R81E mutant after heat treatment. (5) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (6) R81M mutant not heat treated. (7) R81M mutant after heat treatment.

3.5. Protein Column Purification

All of the eleven protein variants (ten mutants and one wild-type) were purified by Ni-NTA affinity column chromatography under native conditions, except one mutant protein which was purified using Fast-Ni-NTA column. For Ni-NTA column chromatography, either heat-treated or untreated cell extracts were used depending on sHSPs concentration judged by band strength shown in figures 3.36-3.47.

Figures 3.49-3.60 show the purification steps for wild-type and mutant variants of tpv-Hsp14.3 protein. In all purifications, usually second and third elutions samples yielded unique and discrete bands of the sHSP. These samples were used for enzyme assay and DLS experiments because of their high purity.
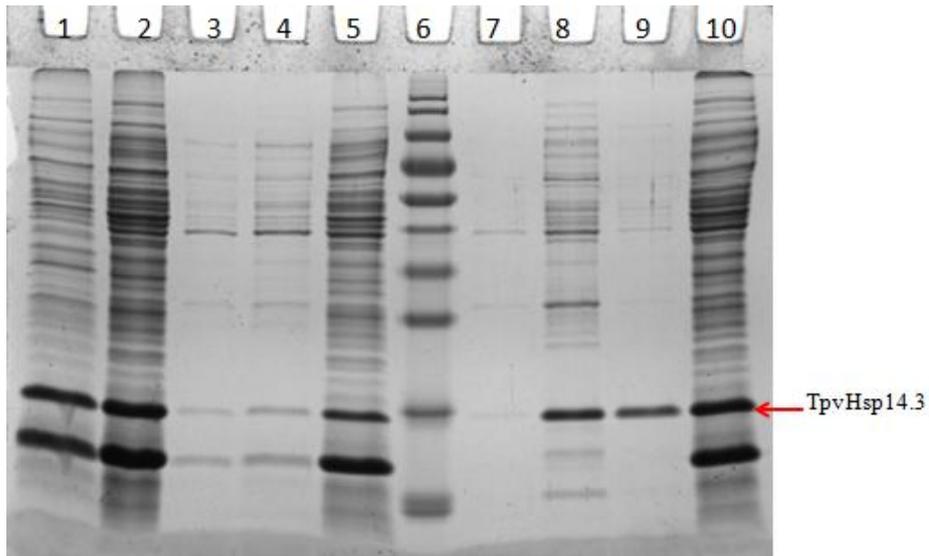
Figure 3.36: SDS-PAGE of R81E Mutant Variant of *Tpv*-Hsp14.3 Purified by Fast-Ni-NTA Column Chromatography. Lanes: (1) FT-5. (2) FT-2. (3) W2-1. (4) W1-4. (5) W1-1. (6) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (7) W2-4. (8) E1. (9) E2. (10) Cellular extract. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution)
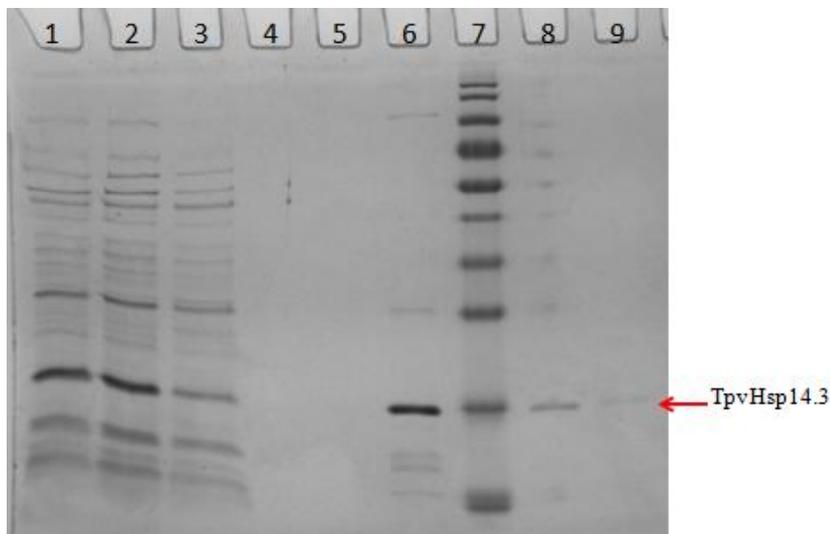
Figure 3.37: SDS-PAGE of Wild-Type *Tpv*-Hsp14.3 Protein Purified By Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) E1. (6) E2. (7) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (8) E3. (9) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution)
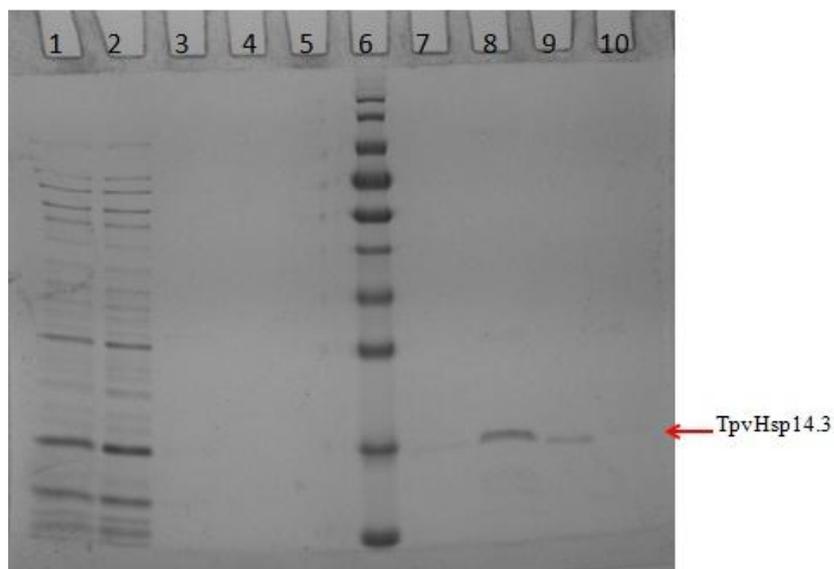
Figure 3.38: SDS-PAGE of R69K Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) W2-4. (6) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (7) E1. (8) E2. (9) E3. (10) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution)
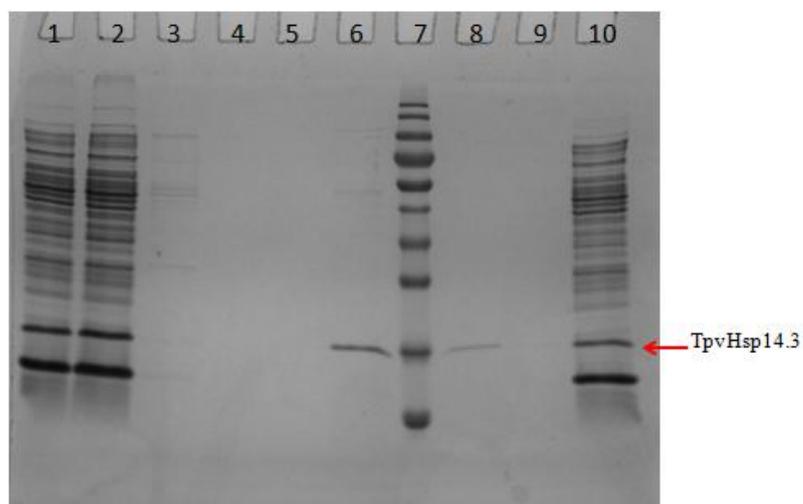
Figure 3.39: SDS-PAGE of R69E Mutant Variant Of *Tpv*-Hsp14.3 Purified By Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W2-4. (5) E1. (6) E2. (7) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (8) E3. (9) E4. (10) W1-1. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution)
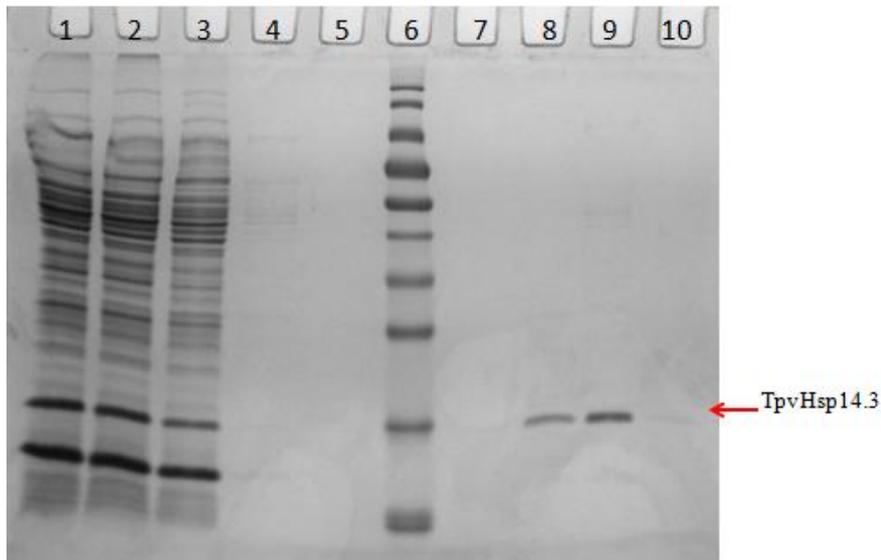
Figure 3.40: SDS-PAGE of R69M Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) W2-4. (6) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (7) E1. (8) E2. (9) E3. (10) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution)
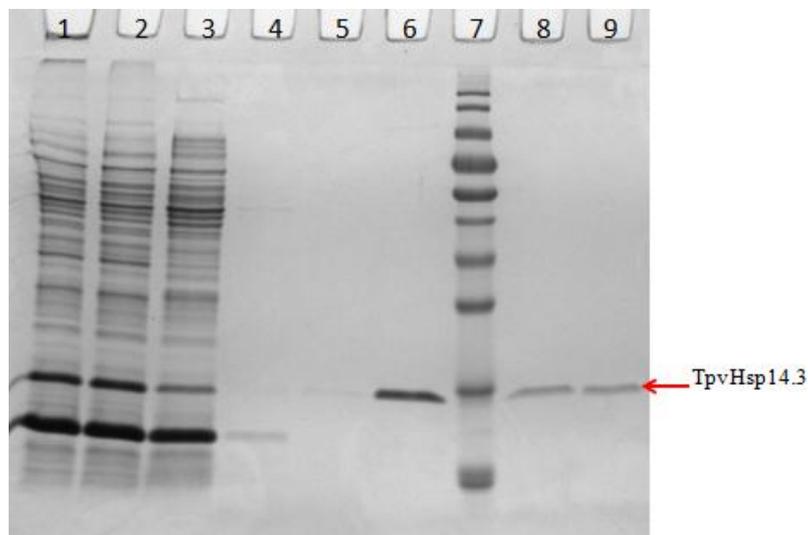
Figure 3.41: SDS-PAGE of R81E Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) E1. (6) E2. (7) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (8) E3. (9) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).
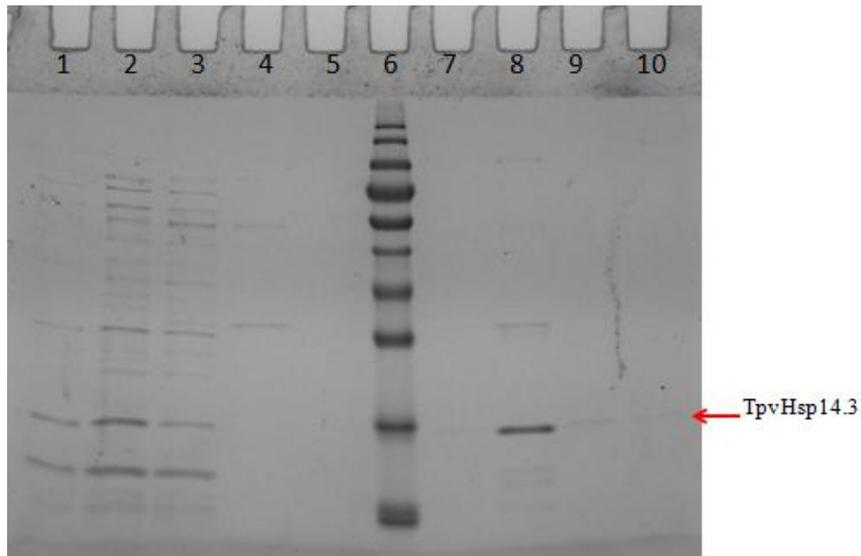
Figure 3.42: SDS-PAGE of R81K Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) W2-4. (6) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (7) E1. (8) E2. (9) E3. (10) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).
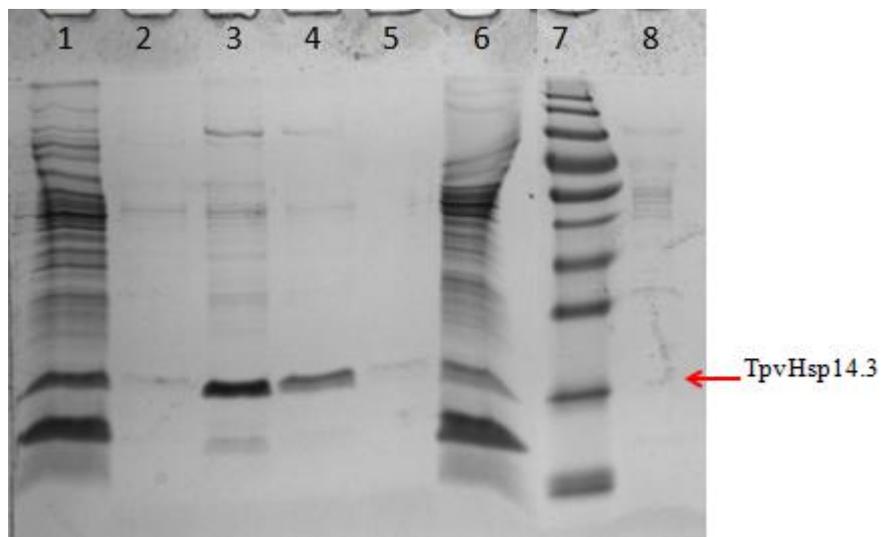
Figure 3.43: SDS-PAGE of R81M Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-4. (2) E1. (3) E2. (4) E3. (5) E4. (6) W1-1 (7) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (8) W1-4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).

125
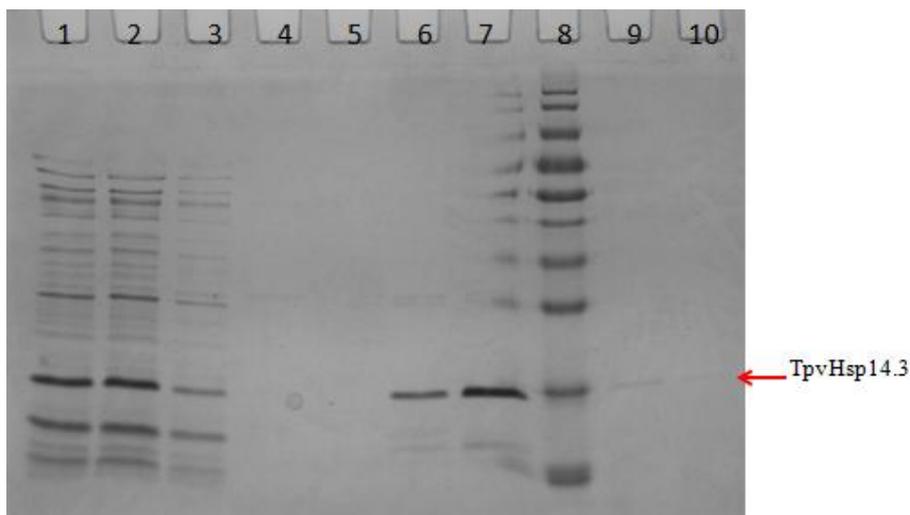
Figure 3.44: SDS-PAGE of QR(80-81)EL Double-Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) W1-4. (5) W2-4. (6) E1. (7) E2. (8) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (9) E3. (10) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).
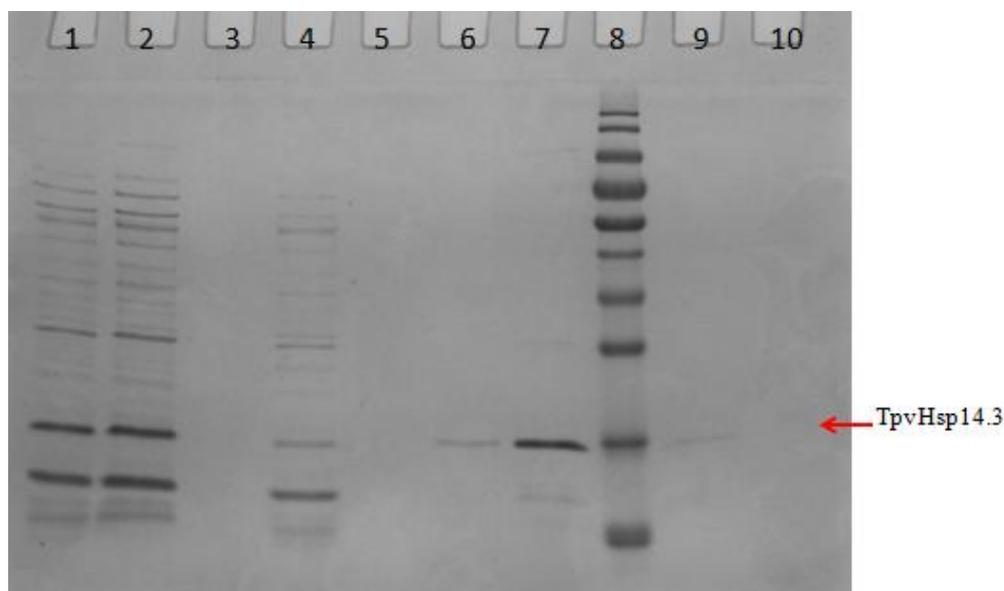
Figure 3.45: SDS-PAGE of K87R Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-4. (4) W1-1. (5) W2-4. (6) E1. (7) E2. (8) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (9) E3. (10) E4. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).
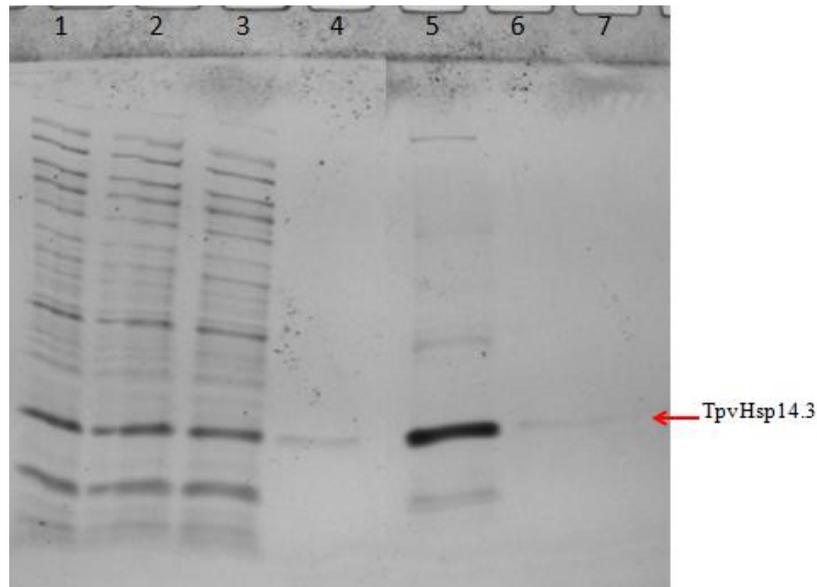
Figure 3.46: SDS-PAGE of K87E Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Lanes: (1) FT-1. (2) FT-4. (3) W1-1. (4) E1. (5) E2. (6) E3. (7) E4. There was no marker used in this gel due to some technical problems. (The abbreviations used are; FT: Flow Through; W1: First Wash; W2: Second Wash; E: Elution).
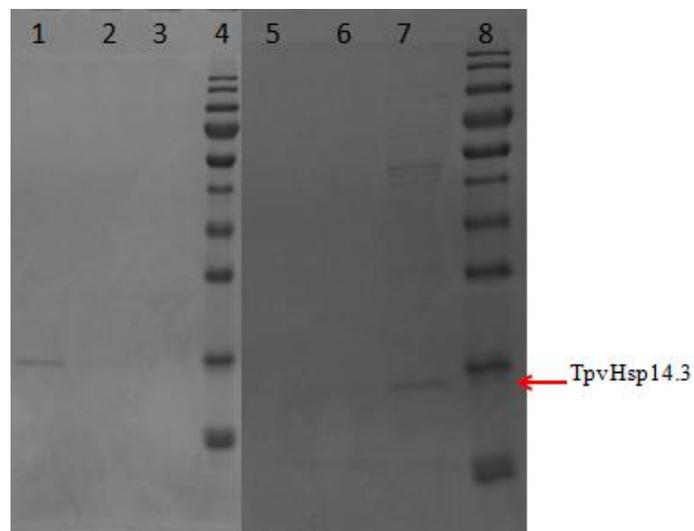
Figure 3.47: SDS-PAGE of K87I Mutant Variant of *Tpv*-Hsp14.3 Purified by Ni-NTA Column Chromatography. Elution purification from two experimsnts. Lanes: (1) E2. (2) E3. (3) E4. (4) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). (5) E4. (6) E3. (7) E2. (8) Protein ruler PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania). Lanes (1)-(3) from first purification, and lanes (5)-(7) from second purification. E stands for Elution.

3.6. Chaperone Activity of the Engineered Tpv-Hsp14.3

The model enzyme used in these experiments was pig heart Citrate Synthase (CS) (Sigma, Alrdich) which carries the reaction described in section 2.3.10. Enzyme activity was measured under optimal and denaturing temperature in four different settings: First, the model enzyme itself in the reaction mixture at 35°C, which is its optimal temperature. Second, CS activity in the same reaction mixture but at the temperature of 47°C. At this temperature CS loses its activity (negative control). Third, activity measurement when CS was incubated with the wild-type tpv-Hsp14.3 in the reaction mixture at 47°C, which serves as a positive control to compare the protection activity of all the mutants. And finally, activty measurement when CS was incubated with each mutants one at a time at 47°C. The protein amount of all the sHSP variants was kept at 34-35µg. The experiments were run in replicates. Figure 3.48 shows the absorbance spectra of CS alone and in the presence of the wild-type tpv-Hsp14.3 chaperone.
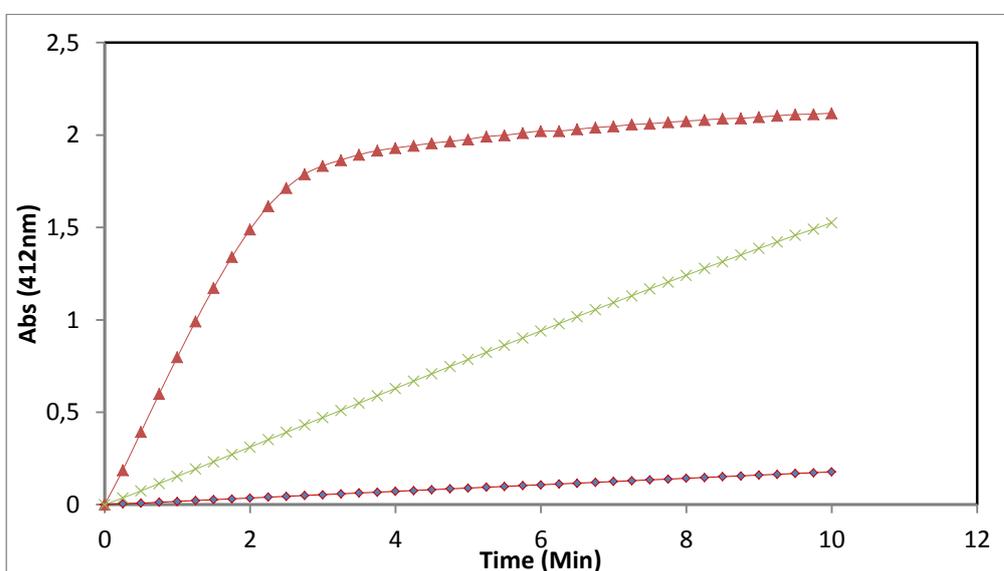
Figure 3.48: Citrate Synthase Activity Measured Under Different Conditions: (——) Citrate synthase alone incubated at 47°C. (——) Citrate synthase alone incubated at 35°C. (——) Citrate Synthase incubated at 47°C in the presence of the wild-type molecular Chaperone.

As the figure shows, when CS was incubated alone at 35°C, its activity was 823±1.27ΔmAbs/min. When CS was incubated alone at 47°C, its activity was 18.1±0.1ΔmAbs/min, a 45-fold fall in as compared to its activity at optimal temperature. However, when CS was incubated with the Wild-Type tpv-Hsp14.3, its activity was 160.1± 0.032 ΔmAbs/min, a 5-fold decrease as compared to CS activity at 35°C, but 9-fold higher compared to CS activity at 47°C. These results indicate an efficient heat protective activity of the wild-type tpv-Hsp14.3 chaperone.

Heat protection effects of the R69 mutants are shown in figures 3.49.
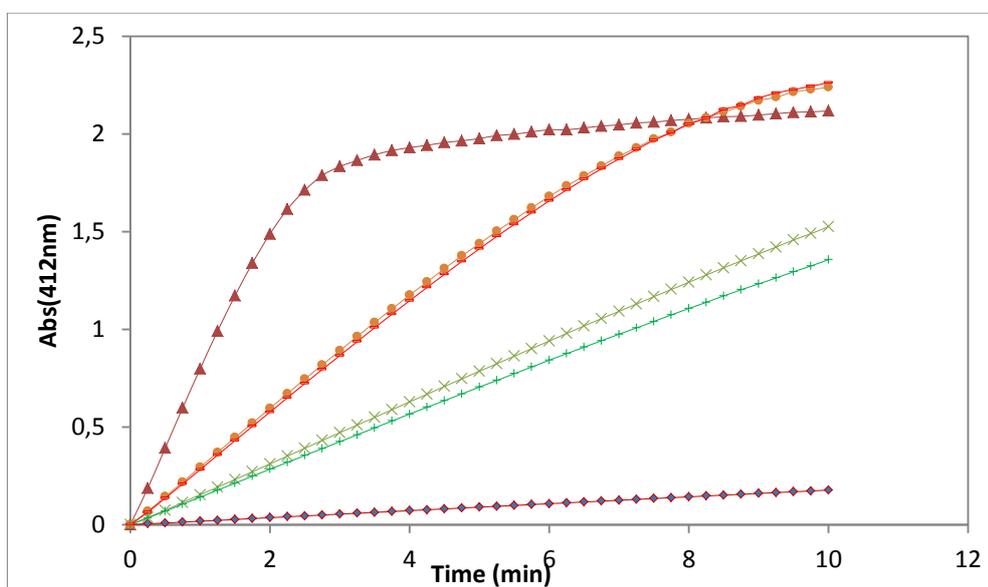
Figure 3.49: Citrate Synthase Activity Measured Under Different Conditions: (▬♦▬) Citrate synthase alone incubated at 47°C. (▬▲▬) Citrate synthase alone incubated at 35°C. (▬✳▬) Citrate Synthase incubated at 47°C in the presence of the wild-type molecular Chaperone. (▬●▬) Citrate Synthase incubated at 47°C in the presence of R69E molecular Chaperone variant. (▬+▬) Citrate Synthase incubated at 47°C in the presence of R69M molecular Chaperone variant. (▬▬) Citrate Synthase incubated at 47°C in the presence of R69K molecular Chaperone variant.

The activity of CS when incubated at 47°C togather with R69E mutant of tpv-Hsp14.3 is 302.4±0.9ΔmAbs/min, 1.9-fold higher than CS activity when incubated with the wild-type sHSP under the same conditions. Similarly, the presence of R69K mutant sHSP, CS activity was 293.5± 0.2ΔmAbs/min, a 1.8-fold higher as compared to its activity when the enzyme was incubated with the wild type sHSP at 47°C. When CS was incubated with the R69M variant its activity was 141.3±0.02, showing a 1.1-fold decrease as compared to CS activity when incubated with the wild type sHSP CS at 47°C.

Heat protection effects of the R81 mutants are shown in figures 3.50.
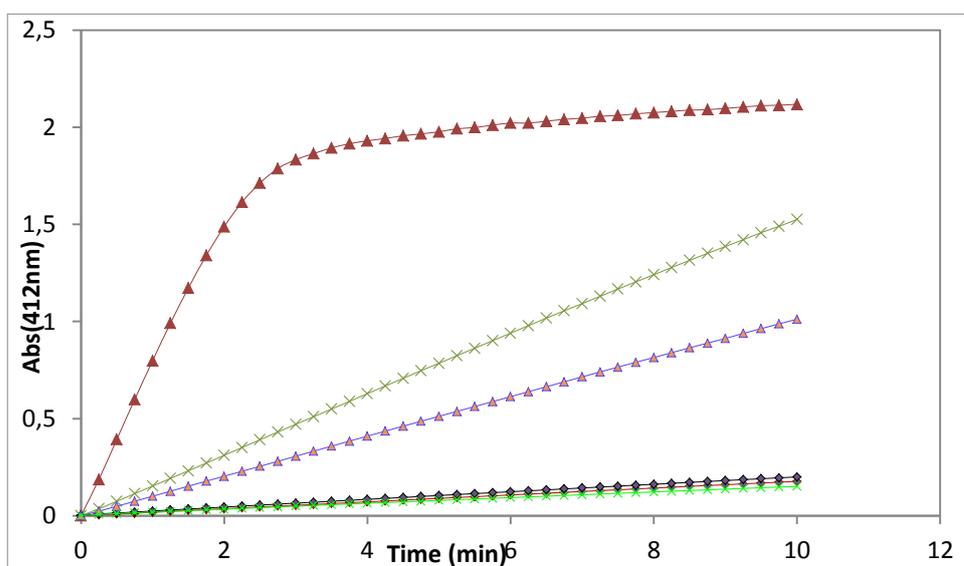
Figure 3.50: Citrate Synthase Activity Measured Under Different Conditions: (——) Citrate synthase alone incubated at 47°C. (——) Citrate synthase alone incubated at 35°C. (——) Citrate Synthase incubated at 47°C in the presence of the wild-type molecular Chaperone. (——) Citrate Synthase incubated at 47°C in the presence of R81E molecular Chaperone variant. (——) Citrate Synthase incubated at 47°C in the presence of R81M molecular Chaperone variant. (——) Citrate Synthase incubated at 47°C in the presence of R81K molecular Chaperone variant.

As the figure shows, heat protection activity of tpv-Hsp14.3 decreases most drastically in the mutants of R81 position. The activity of CS incubated with R81E mutant variant at 47°C is 20±0.01ΔmAbs/min, an 8-fold decrease as compared to its activity when incubated with the wild-type sHSP. The activity of CS incubated with R81K mutant variant at 47°C is 103.3±0.3ΔmAbs/min, a 1.5-fold decrease as compared to its activity when incubated with the wild-type sHSP under the same conditions. The activity of CS incubated with R81M mutant variant at 47°C is 15.5±0.2ΔmAbs/min, a 10.3-fold decrease as compared to its activity when incubated with the wild-type sHSP under same conditions. As these results show, R81K variant has still chaperoning activity, although slightly less than of the wild type sHSP. On the other hand, R81E and R81M mutants have lost their chaperoning activity completely.

Heat protection effects of the double mutant QR(80-81)EL are shown in figures 3.51. As the figure shows, the activity of CS incubated with QR(80-81)EL double mutant at 47°C is 194±0.1ΔmAbs/min, a 1.2-fold higher as compared to its activity when incubated with the wild-type tpv-Hsp14.3, although single mutants of position R81 generally exhibited lowered chaperone activity.

The original residues Gln and Arg were exchanged with Glu and Leu respectively. A double mutant exchanging one partially-positive (Gln) and one basic residue into an acidic (Glu) and hydrophobic (Leu) residues would be expected to have drastic effects, but it was not the case. This could be explained by the fact that the new sequence is changed almost to the sequence of *M.janaschii* sHSP16.5 in this particular region, which is E98 and I99 as shown by structure superposition in figure 3.22. Since Leu is an analogue of Ile, not much difference between the two variants can be expected.
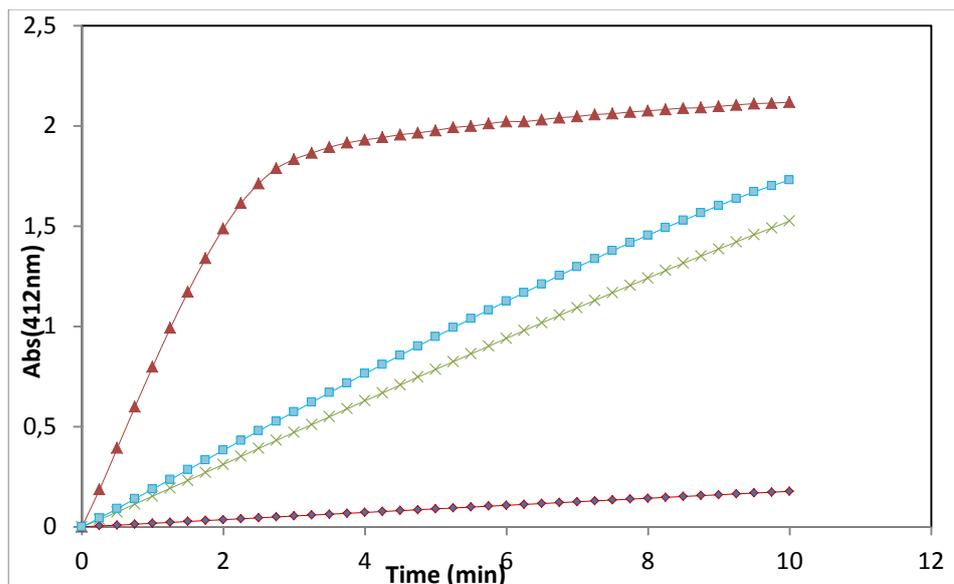
Figure 3.51: Citrate Synthase Activity Measured Under Different Conditions: (——◆——) Citrate synthase alone incubated at 47°C. (——▲——) Citrate synthase alone incubated at 35°C. (——✳——) Citrate Synthase incubated at 47°C in the presence of the wild-type molecular Chaperone. (——■——) Citrate Synthase incubated at 47°C in the presence of the QR80-81EL molecular Chaperone. The activity for each case is given in the chart.

Heat protection effects of the K87 mutants are shown in figures 3.52. All the mutations at position K87 caused an increase in CS activity when they were incubated with the enzyme at 47°C as compared to the wild-type sHSP under the same conditions. The activity of CS incubated with K87R mutant variant at 47°C is 412.1±0.2ΔmAbs/min, a 2.5-fold higher as compared to its activity when the enzyme was incubated with the wild type sHSP at 47°C. The activity of CS incubated with K87E mutant variant at 47°C is 212.3±0.05ΔmAbs/min, a 1.3-fold higher as compared to its activity when the enzyme was incubated with the wild type sHSP at 47°C. Lastly, the activity of CS incubated with K87I mutant variant at 47°C is 429.1±0.2ΔmAbs/min, a 2.7-fold higher as compared to its activity when the enzyme was incubated with the wild type sHSP at 47°C.

132
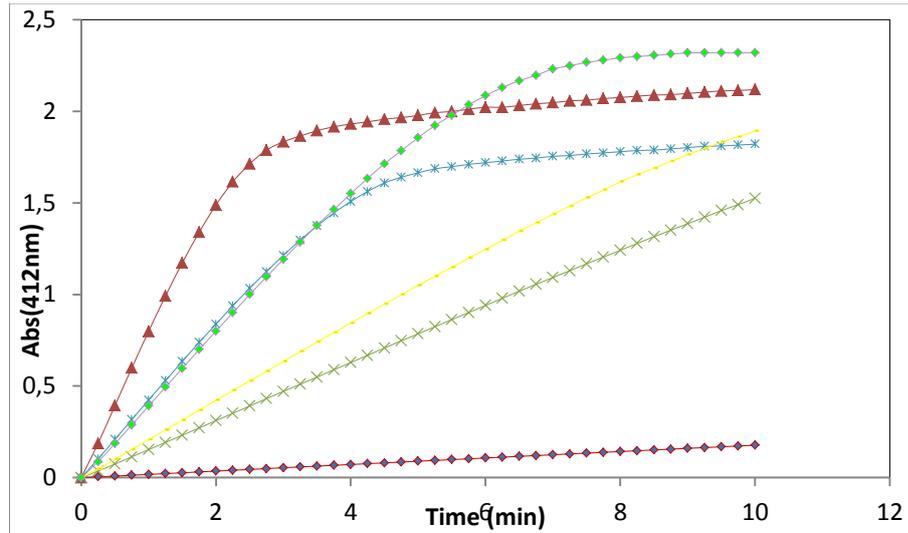
Figure 3.52: Citrate Synthase Activity Measured Under Different Conditions: (——)
Citrate synthase alone incubated at 47°C. (——) Citrate synthase alone incubated at 35°C.
(——) Citrate Synthase incubated at 47°C in the presence of the wild-type molecular
Chaperone. (——) Citrate Synthase incubated at 47°C in the presence of K87I molecular
Chaperone variant. (——) Citrate Synthase incubated at 47°C in the presence of K87E
molecular Chaperone variant. (——) Citrate Synthase incubated at 47°C in the presence of
K87R molecular Chaperone variant.

The combined results for heat protection of pig heart CS by the wild-type and each
one of the mutant tpv-Hsp14.3 at 47°C are shown in figure 3.53.
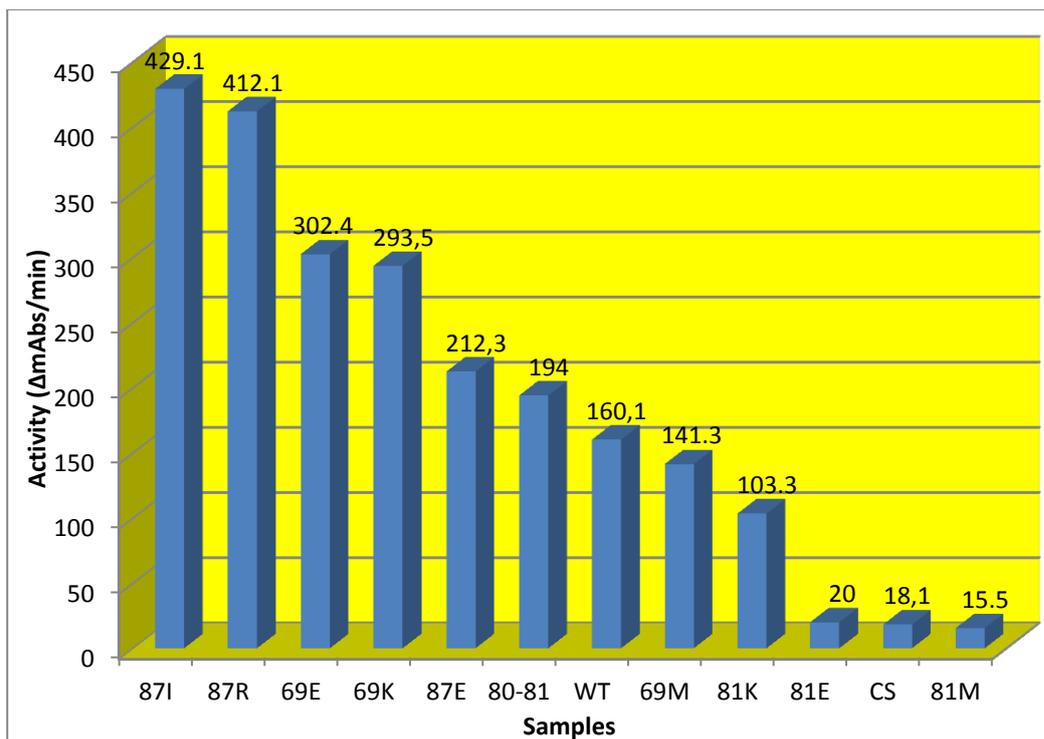
Figure 3.53: CS Activity Protection in the Presence of Wild-Type and Mutant Variants of Tpv-Hsp14.3. Summary of CS activities incubated with different tpv-Hsp14.3 variants. In the sample axis, the number stand for the position and the letter for the residue introduced. WT stand for the wild-type tpv-Hsp14.3, and CS stand for CS enzyme at 47°C. The activity values are shown on top of each bar.

3.7. Dynamic Light Scattering Studies.

The original experimental data obtained in DLS studies are given in appendix E. In this section, the data are represented by charts. Some data that had very small values can not be shown in the charts, but they are found in Appendix D. The data shown in these charts are of the Simple Fit Distribution for 10 seconds. In DLS, the particle size is assumed by the software to be the diameter of a "perfect sphere", and this sphere represents a multimer of sHSPs. The monomers interact with each other to form large protein assemblies.

3.7.1. Optimization of DLS measurement.

In our study, early DLS experiments were run mostly for optimization purposes. DLS measurements at different temperatures for 30 seconds give single peaks and maximum peak values are shown in figure 3.54.
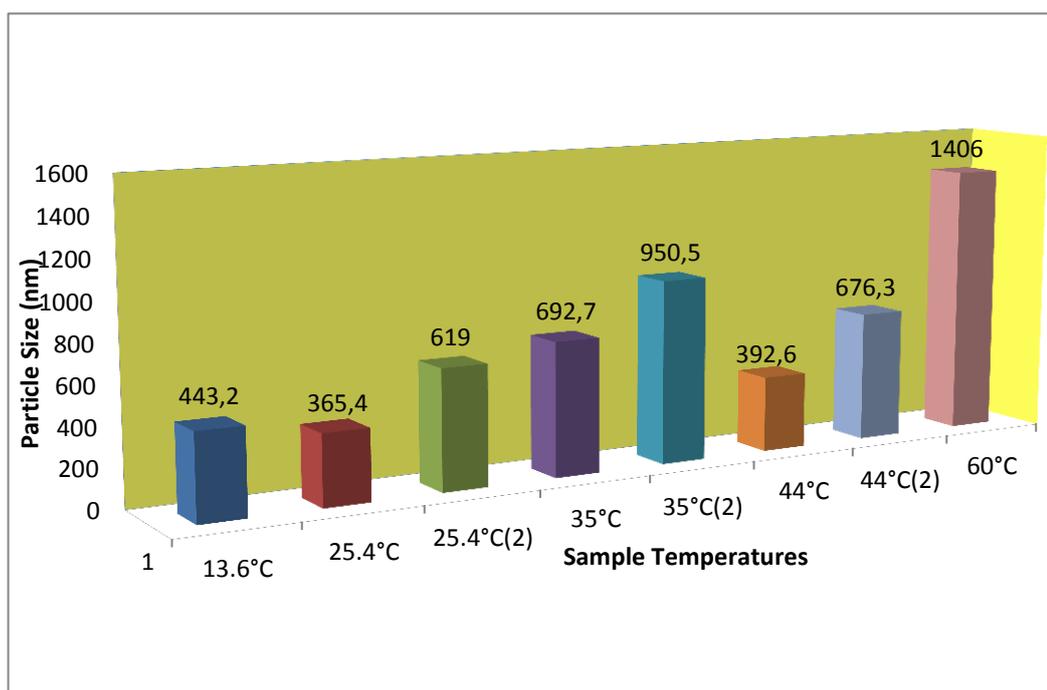
Figure 3.54: DLS Measurements of the Wild-Type Tpv-Hsp14.3 during 30 Seconds. The amount of sHSP protein was 1.88 mg. The peak size for each temperature and are given in the chart. The number 2 in parenthesis after temperature stands for the second measurement of light scattering at the same temperature.

In contrast, measurements for 10 seconds give more peaks (figure 3.55), which was judged to be more realistic for sHSPs given their dynamic nature. This dynamism is emphasised even more at higher temperatures, and for this reason, the rest of experiments were done for 10 second measuring time. Oligomeric size distribution for the wild-type tpv-Hsp14.3 at different temperatures are: at 13.6°C 98.9% of the

particles had size of 1985 nm diameter, 0.4% 34.5 nm and 0.7% 0.1 nm. At 35°C 94.9% of the particles had size of 1121 nm, 4.6% 65.6 nm and 0.5% 0.1 nm. At 44°C 99.8% of particles' size was 1406 nm and the remaining 0.2% was 0.9 nm. At 60°C, 89.3% of the particles had size of 764.2 nm, while 9.8% 10910 nm and 0.9% were at a diameter of 0.2 nm.
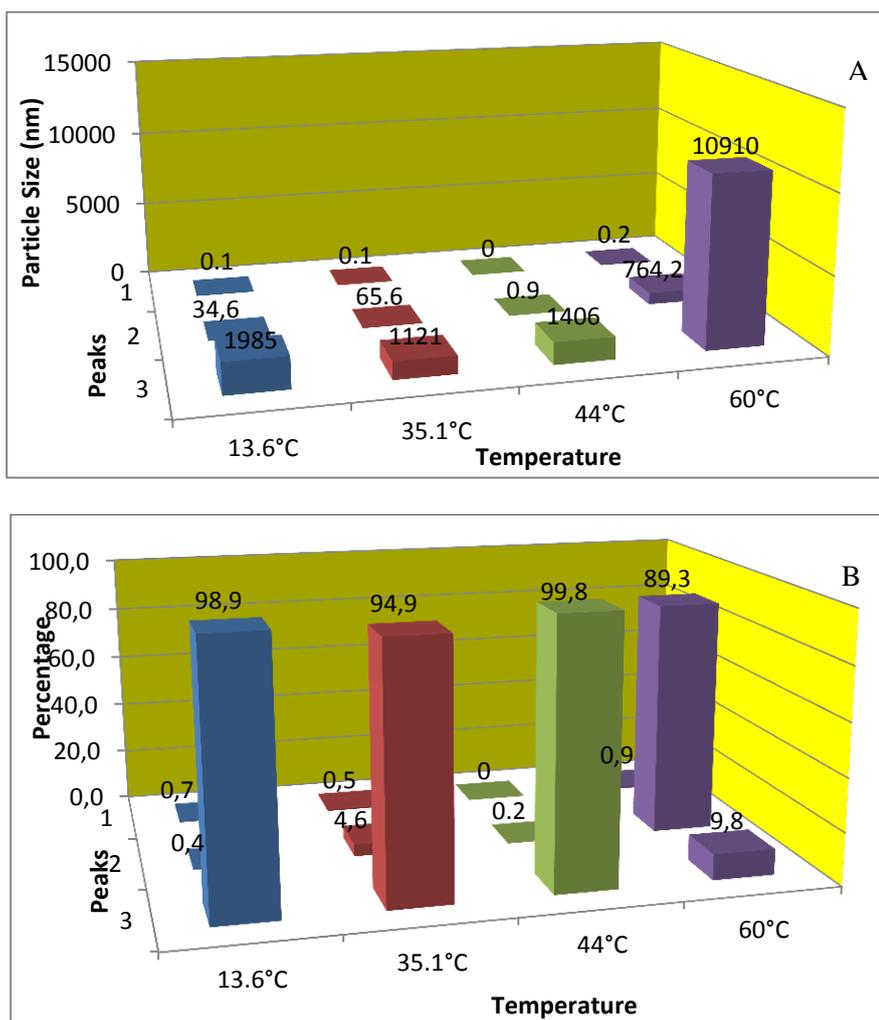
Figure 3.55: DLS Analysis of the Wild-Type Chaperone Measured for 10 Seconds.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A for different temperatures. The sHSP amount in this DLS experiment was 1.88 mg. Measurements were taken for 10 seconds at the temperatures shown in the bar diagram.

3.7.2. DLS Measurements for the Wild-Type Tpv-Hsp14.3 in the Presence and Absence of Protein Substrate.

The first set of experiments was performed with pig heart Citrate Synthase incubated with tpv-Hsp14.3 (figure 3.56). When 1mg of wild type tpv-Hsp14.3 protein was incubated at 47°C for 10 minutes, 46.6% of the protein is found in complexes at a size of 8000 nm diameter, 51.5% of 519 nm, and 1.8% of 0.14 nm diamter. When 15µl CS is added and incubated with the chaperone at 47°C, 38.2% of the particles have size of 2660 nm diameter, 61.6% of 218.2 nm, and 0.2% of 0.14 nm diameter. The first peak may be an experimental artifact.

The second set of experiments were performed with Bovine Glutamate Dehydrogenase (GDH) enzyme incubated with 0.85mg tpv-Hsp14.3. GDH (Sigma, Aldrich) is a mesophilic enzyme, whose optimal assay activity is at 25°C, and its activity is lost rapidly at 56°C. As the bars in figure 3.57 show, at 53°C 99.2% of the particles are at a size of 627.5 nm diameter and 0.8% of 0.16 nm diameter. When GDH was added and incubated for 10 min at 53°C, the complex gave a single peak representing particles with size 587.3 nm diameter.
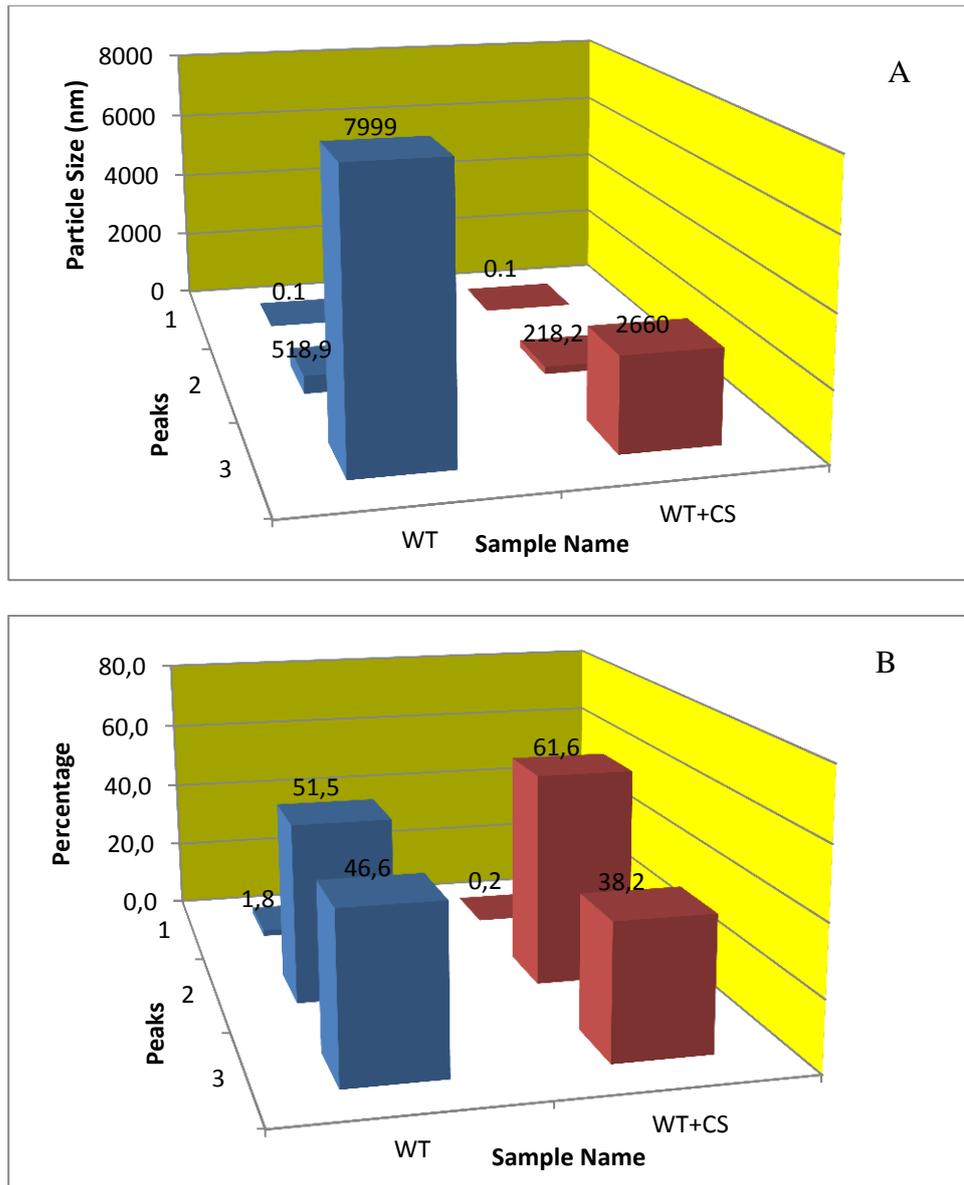
Figure 3.56: DLS Data of the Wild-Type Tpv-Hsp14.3 Alone and With CS. Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A. WT stands for wild-type sHSP, and WT+CS for wild-type sHSP plus 15μl CS. Incubations were done for 10 minutes at 47°C.

138
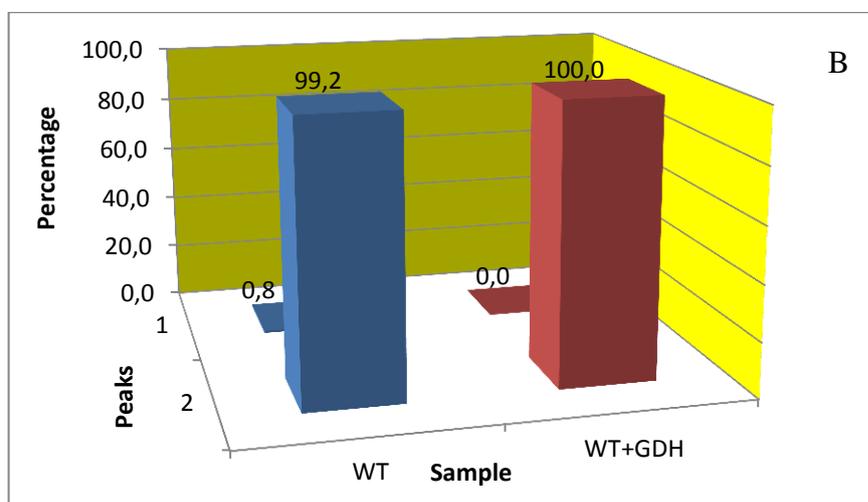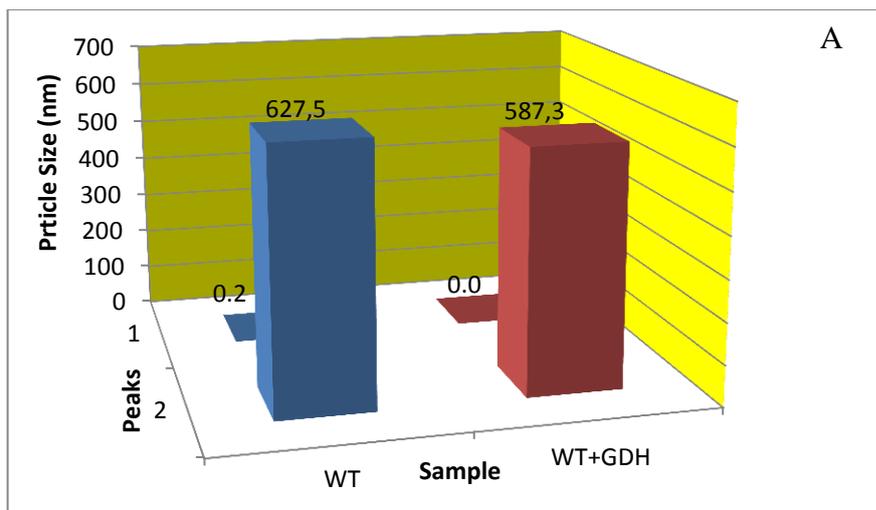
Figure 3.57: DLS data of the Wild-Type Tpv-Hsp14.3 Alone and with GDH. Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A. WT stands for 0.85mg wild-type sHSP, and WT+GDH for 0.85mg wild-type sHSP plus 15µl GDH enzyme. Incubations were done for 10 minutes at 53°C.

3.7.3. DLS Measurements for the Different Mutant Variants of Tpv-Hsp14.3 Chaperone

The concentrations of sHSPs in figures 3.58-3.68 are 0.7mg/ml. The Oligomer size distribution of wild-type tpv-Hsp14.3 is shown in figure 3.58
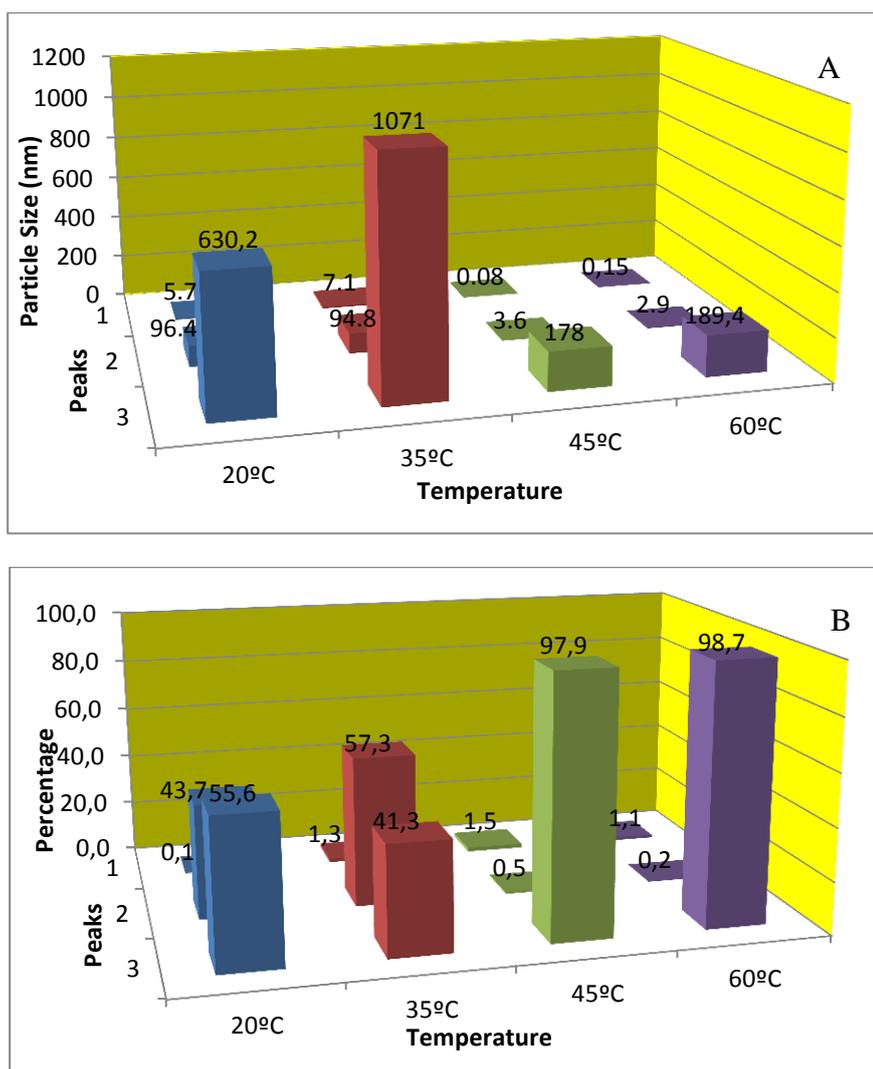
Figure 3.58: Wild-type Tpv-Hsp14.3 Oligomer Size and their Percent Distribution. Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.

The figure shows that at 20°C, 55.6% of the particles have a size of 630.2 nm diameter, and 43.7% of 96.4 nm. and 0.1% of size 5.7 nm diameter. At 35°C 41.3% are of a size of 1071 nm diameter, 57.3% of 94.8 nm and 1.3% have a size of 7.1 nm diameter. At 45°C there is a dramatic shift, with 97.9% of particles having a size of 178 nm diameter, and 0.5% a size of 3.6 nm diameter. The last size of 0.1 nm diameter is irrelevant and ignored. At 60°C 98.7% are of a size of 189.4 nm diameter, and 0.2% have a diameter of 2.9 nm. The 0.15 nm size is also discarded.

Oligomeric size distribution of R69K mutant of tpv-Hsp14.3 is given in figure 3.59. The figure shows that at 20°C almost all the particles have a size of 2730 nm diameter, because the 0.1 nm particle size can be ignored. At 35°C 91.2% have size of 2053 nm diameter, 1.6% have size of 40.3 nm and a small 0.1% have size of 1 nm diameter. At 45°C the main peak of 87.2% has a size of 1555 nm diamter, 4.1% a size of 68.9 nm and 8.6% 0.1 nm diameter. At 60°C 83.8% have a size of 857.6 nm diameter, 1% 33 nm, 0.4% 3.1 nm, and 14.8% had a size of 0.1 nm diameter, not shown in the bar because such a small size is irrelevant for even a monomer of a chaperone.

Oligomeric size distribution of R69E mutant of tpv-Hsp14.3 is given in figure 3.60. The figure shows that at 23°C 95.6% of the particles are of size 3014 nm diameter, 1.3% size of 38.9 nm and a 2.1% size of 0.6 nm diameter. At 35°C the 97.7% were 944 nm diameter, 9.2% 0.8 nm and the rest were negligible. At 45°C 90.8% were of a size of 863 nm diameter and 9.2% 0.5 nm. Finally at 60°C 92.3% had size of 1074 nm diameter, 2.7% 1.1 nm and a 5% 0.1 nm diameter.
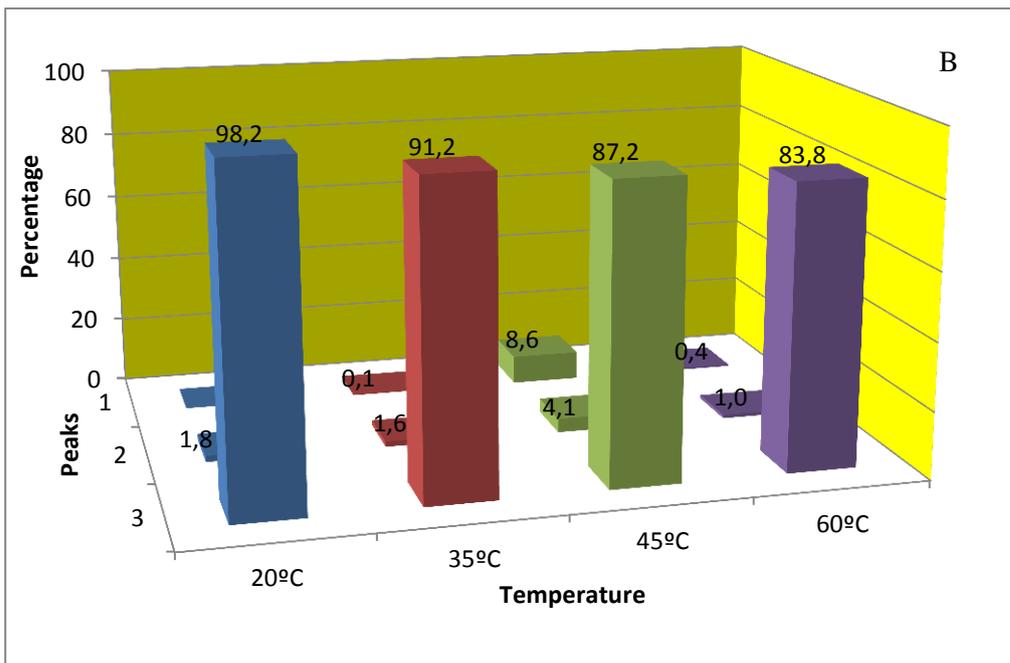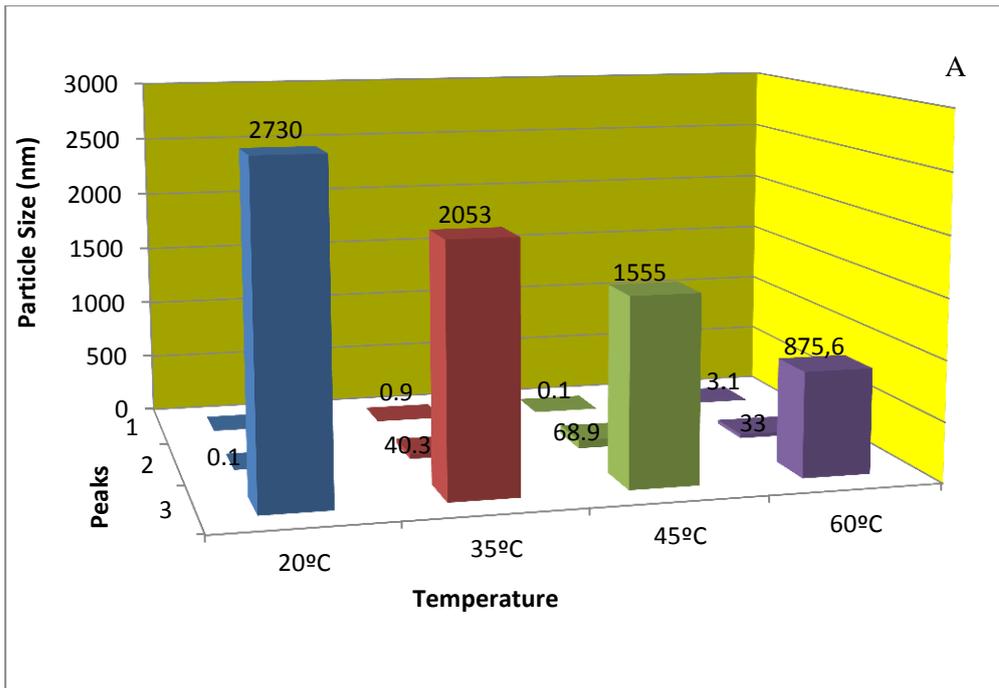
Figure 3.59: R69K Mutant Tpv-Hsp14.3 Oligomer Size and their Percent DistributioN.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at
pannel A vs. temperature. The values are shown at the top of each bar.
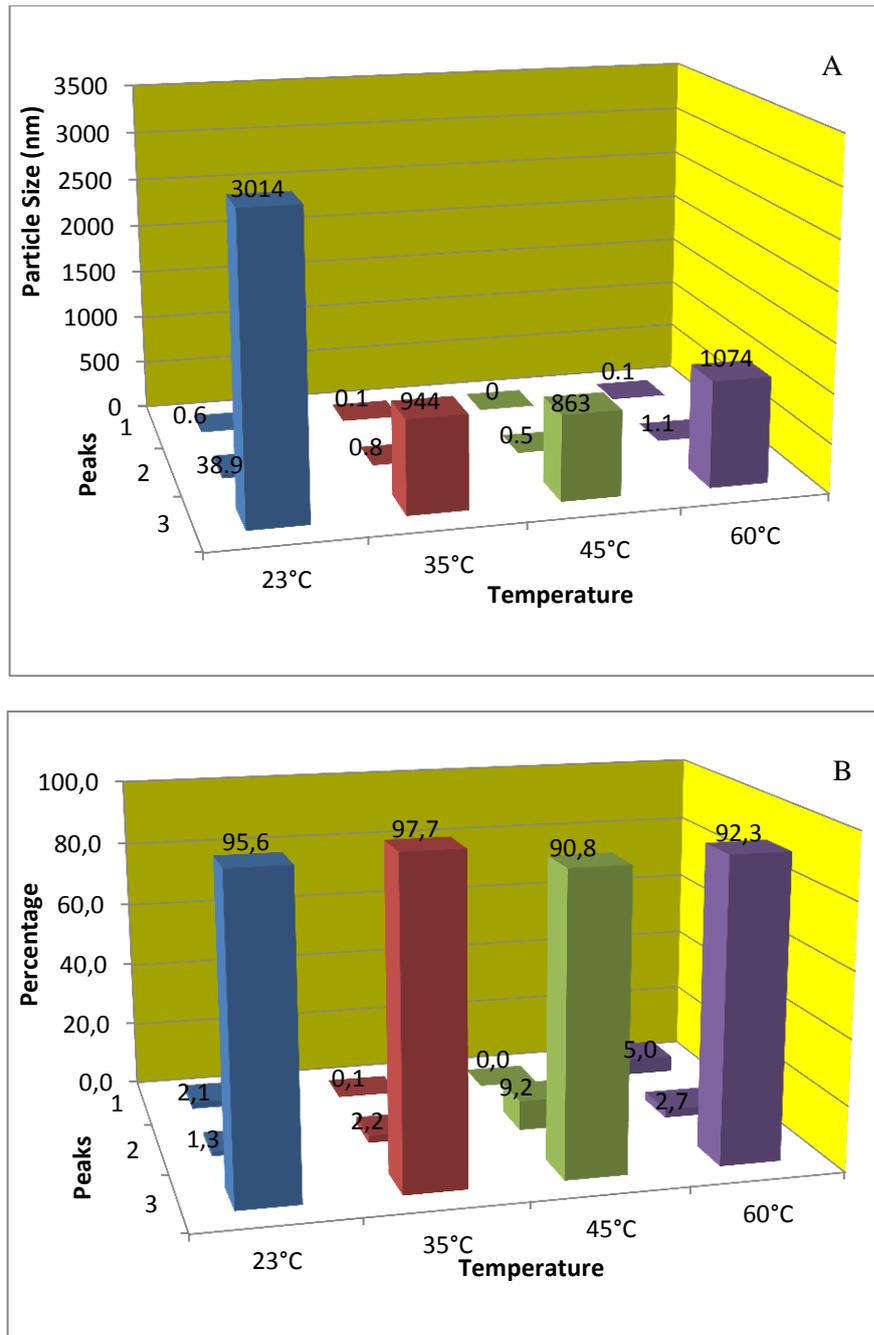
Figure 3.60: R69E Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.

Oligomer size distribution of R69M mutant of tpv-Hsp14.3 is given in figure 3.61. The figure shows that at 23°C 97.2% of particles are of size of 3576 nm diameter, and 0.4% 32.4 nm and 2.3% have size of 0.1 nm diameter. At 35°C 95.1% are of size 2920 nm diameter, 2.7% 165.4 nm and 2.2% 1.4 nm diameter. At 45°C 96.1% are of size1434 nm diameter, 0.8% of 8.7 nm and 3.2% 0.2 nm diameter. Finally at 60°C 93.5% are of size 1353 nm diameter, 1.7% 41.5 nm and the rest 4.9% have size 0.4 nm diameter.

Oligomer size distribution of R81K mutant of tpv-Hsp14.3 is given in figure 3.62. The figure shows that at 23°C 95.6% of the particles have diameter of 3026 nm, 2.1% 95.5 nm and 0.2% 1.5 nm diameter. At 35°C 97.9% have a size of 2218 nm diameter, 1% 61 nm and the rest 1% 0.4 nm diameter. At 45°C 92.9% were 2373 nm diameter, 5.4% 188.1 nm and the rest 1.7% 0.1 nm diameter. Finally at 60°C 93.2% had a size of 1975 nm diameter, 4.8% 151.4 and the rest 0.1% 0.7 nm diameter.
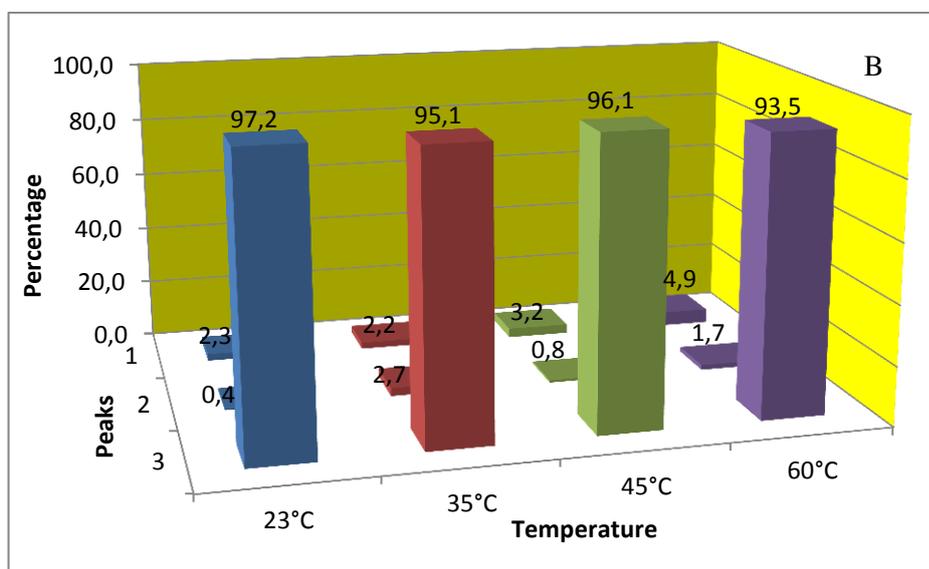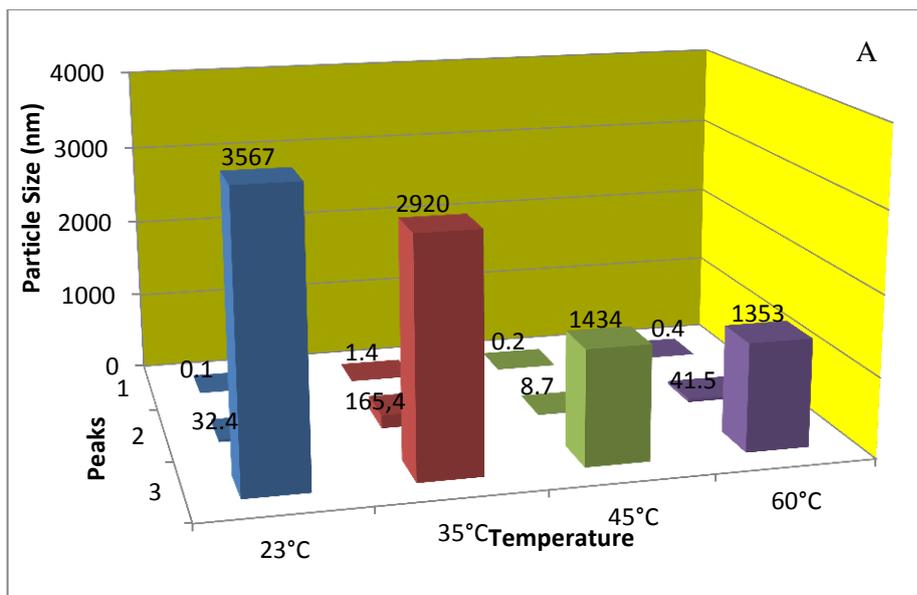
Figure 3.61: R69M Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.
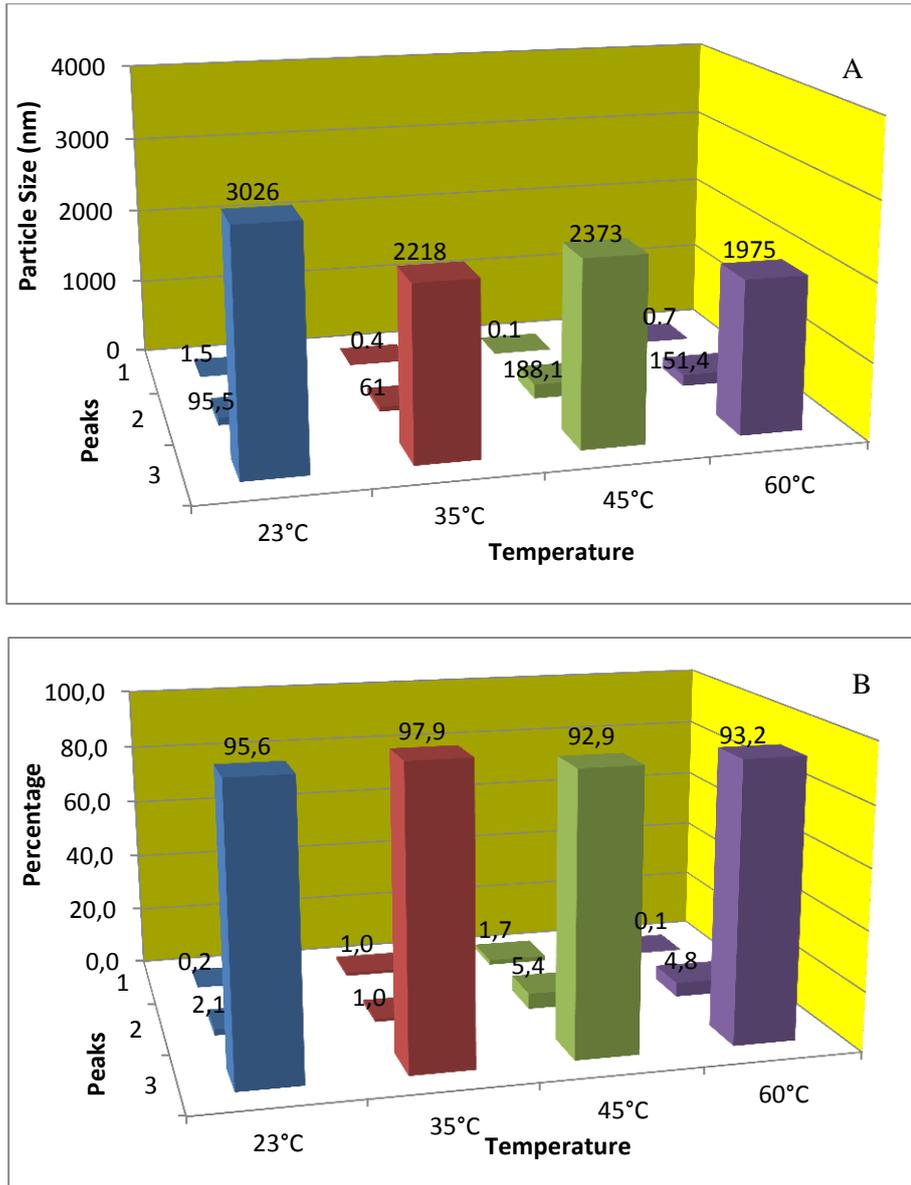
Figure 3.62: R81K Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.

Oligomer size distribution of R81E mutant of tpv-Hsp14.3 is given in figure 3.63. The figure shows that at 20°C 99.6% of the particles are of size 3842 nm diameter and the rest 0.4% 0.17 nm diameter. At 35°C 93.5% had size of 1528 nm diameter, 1.4% 7.4 nm and the rest 5.1% 0.1 nm diameter. At 45°C 92.6% had size of 1445 nm diameter, and 7.3% 0.1 nm diameter. At 60°C 86.5% had size of 942.4 nm diameter, 1.1% 3 nm, and 12.4% 0.1 nm diameter.

Oligomer size distribution of R81M mutant of tpv-Hsp14.3 is given in figure 3.64. The figure shows that at 20°C 99.7% of the particles have size of 567.7 nm diameter, 0.2% 7.6 nm, and 0.1% 0.1 nm diameter. At 35°C 37.6% had size of 1932 nm diameter, 56.5% 148.7 nm, and 0.5% 19.7 nm diameter. At 45°C 36.4% had a size of 567.6 nm diameter, 61.6% 73.5 nm, and 1.1% 5.7 nm diameter. Finally at 60°C 54.4% had size of 535.1 nm diameter, 44.6% 52 nm, and 0.2% 0.3 nm diameter. The first peak at 35°C may be an experimental artifact, as such cases occur in DLS for short measurement times.
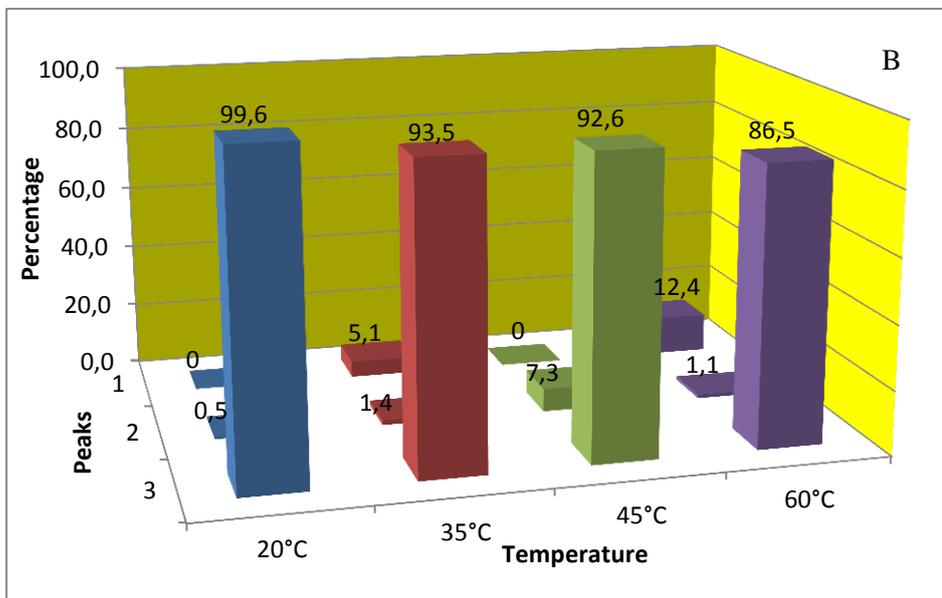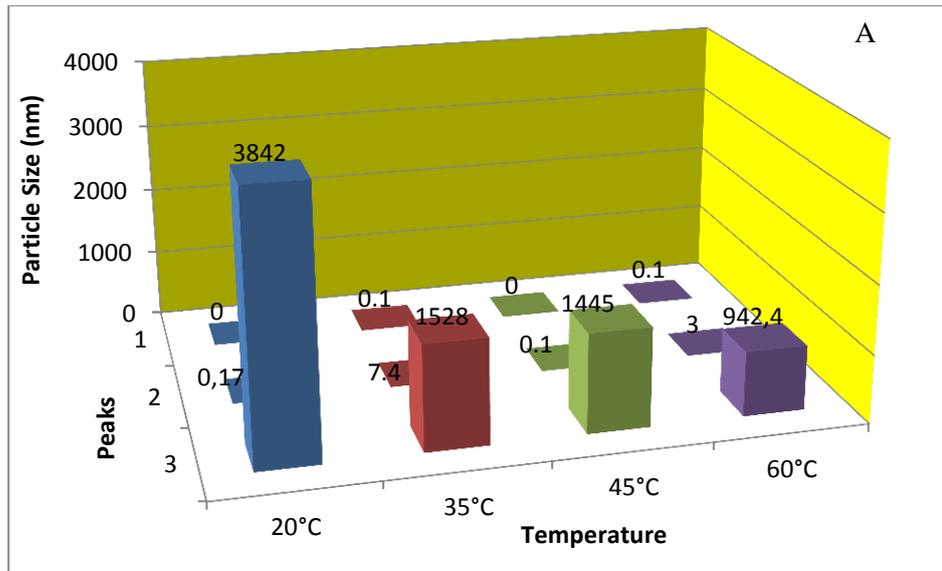
Figure 3.63: R81E Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.
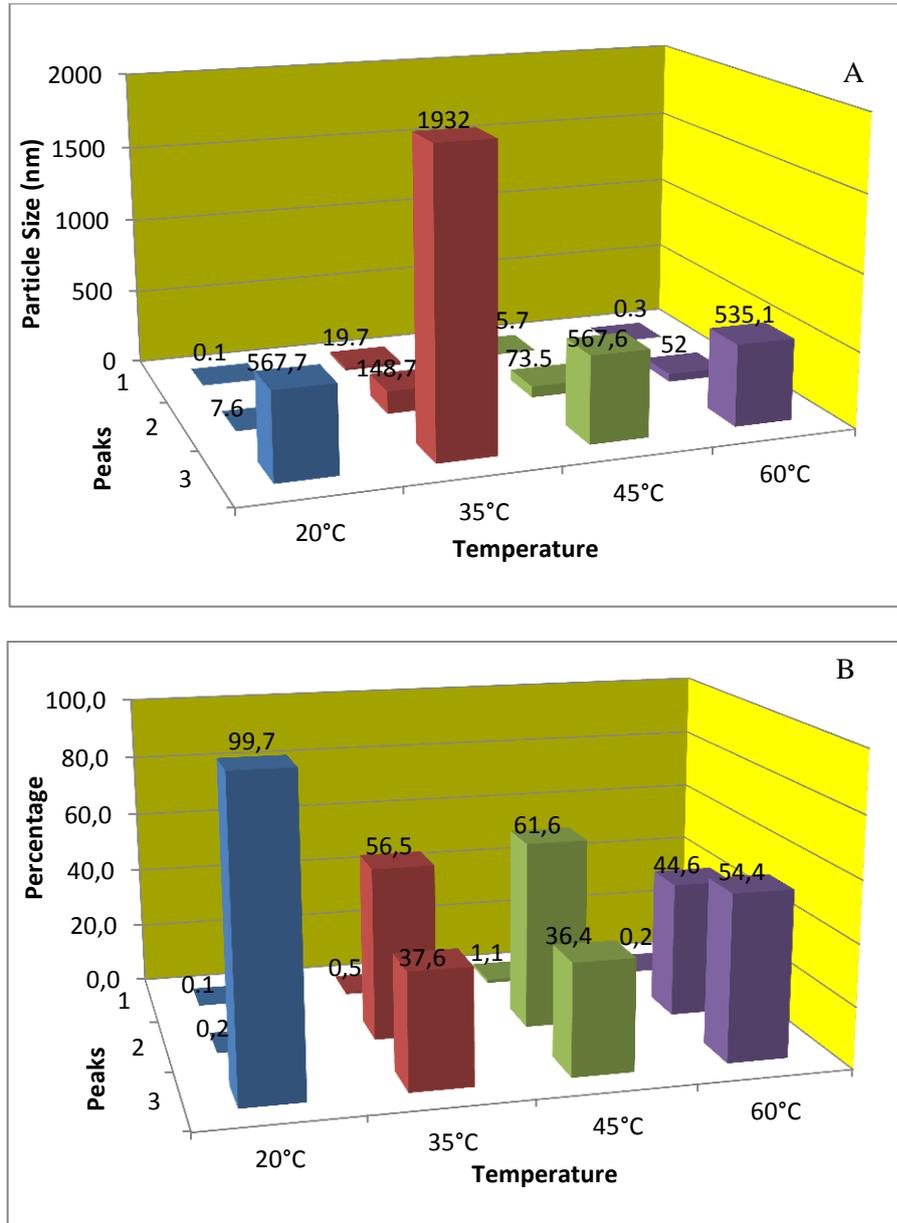
Figure 3.64: R81M Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at
pannel A vs. temperature. The values are shown at the top of each bar.

Oligomer size distribution of QR(80-81)EL mutant tpv-Hsp14.3 show in figure 3.65. The figure shows that at 20°C 66.6% of the particles have size of 1816 nm diameter, 28.7% 197 nm, 3.6% 31.6 nm diameter, and a small percentage was at the range of 10^{-2} nm. At 35°C 67.8% had a size of 1264 nm diameter, 29.4% 107 nm, and 2.8% 16.3 nm diameter. At 45°C there is a shift, with 99% of the proteins having size of 420 nm diameter, and the rest 1% 1.5 nm. Finally at 60°C 98.9% of the proteins had size of 287.5 nm diameter, and 1.1% a size of 0.4 nm diameter.

Oligomer size distribution of K87R mutant of tpv-Hsp14.3 is given in figure 3.66. The figure shows that at 20°C 88.7% of particles were of a size 1230 nm diameter, 11% of size 82.2 nm and 0.3% of size 0.1 nm diameter. At 35°C 75.9% had size 1241 nm diameter, 22.7 size of 121.8 nm, and 1.1% size of 15.8 nm diameter. At 45°C there is a shift towards a single pick with 99.5% of proteins having size 575.3 nm diameter, 0.2% size of 1.3 nm and 0.3% 0.2 nm diameter. At 60°C 98.3% had size of 362.4 nm diameter, 0.2% 0.3 nm, and 1.5%0.1 nm diameter.
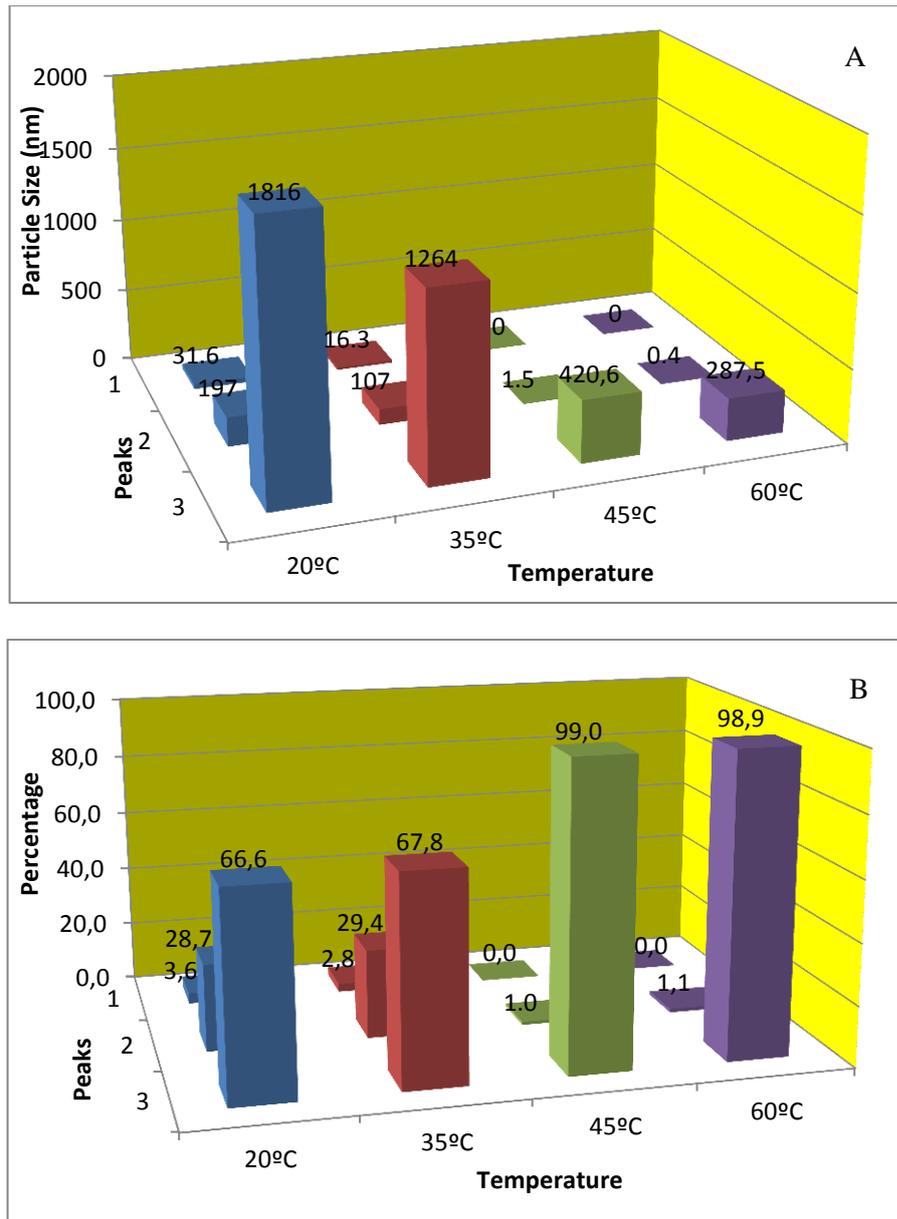
Figure 3.65: QR(80-81)EL Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution. Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.
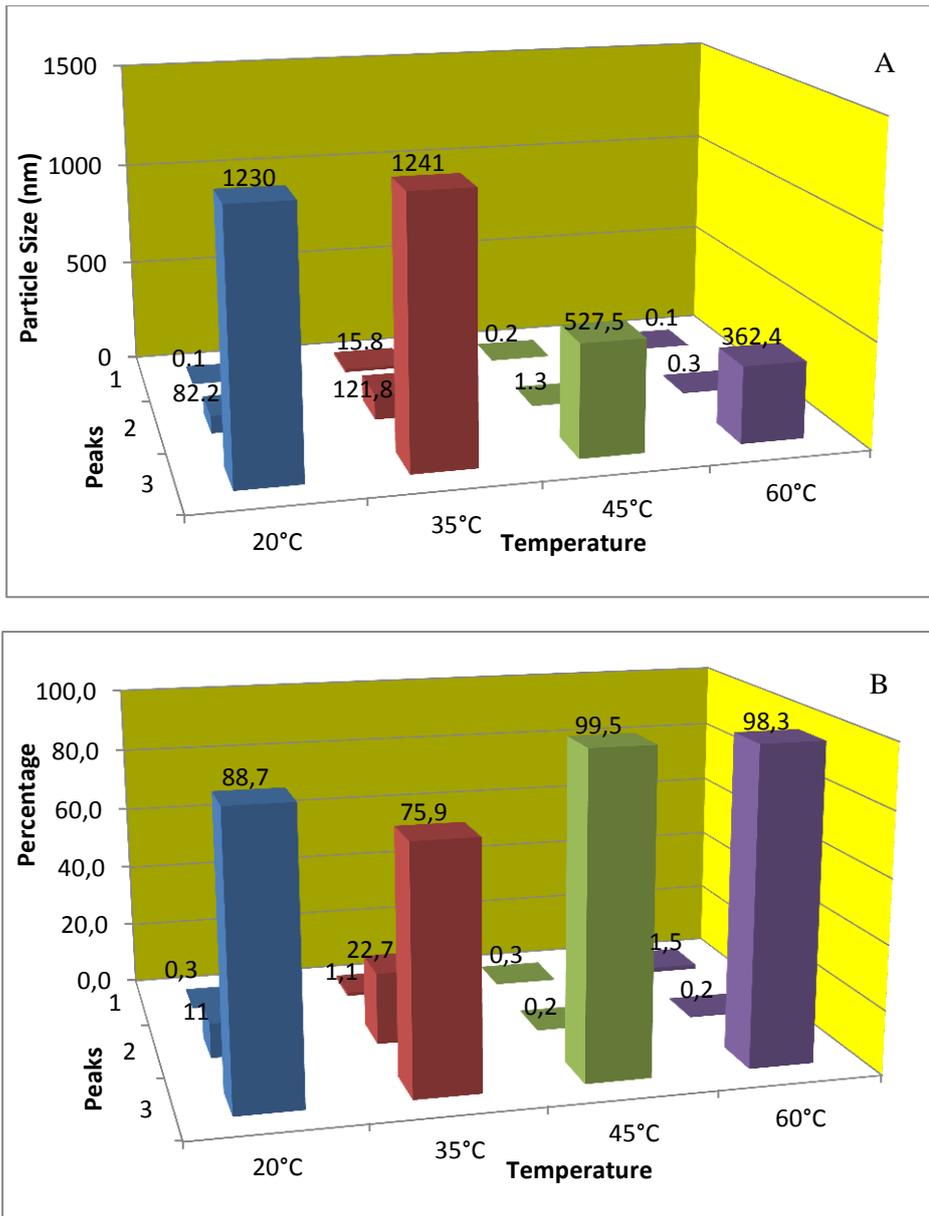
Figure 3.66: K87R Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at
pannel A vs. temperature. The values are shown at the top of each bar.

Oligomer size distribution of K87E mutant of tpv-Hsp14.3 is given in figure 3.67. The figure shows that at 20°C 65.6% of particles were of a size of 1878 nm diameter, 33.9% with size 120.2 nm and 0.5% with size 0.1 nm diameter. At 35°C 96.7% had size of 327.3 nm diameter, 2.3% 8.7 nm, and 0.1% 0.1 nm diameter. At 45°C 98.7% had size of 254.9 nm diameter, 0.3% 2.2 nm, and 1.1% 0.1 nm diameter. An finally at 60°C 99% had size of 185.5 nm diameter, 0.1% 1.2 nm and 0.9% 0.3 nm diameter.

Oligomer size distribution of K87I mutant of tpv-Hsp14.3 is given in figure 3.68. The figure shows that at 23°C the size distribution of particles was 62.6% of size 4272 nm diameter, 33.1% 258.3 nm, 3.3% 44.1 nm and the rest 1% were of a size of 0.07 nm diameter. At 35°C 68% had size of 1431 nm diameter, 30.7% size of 116.9 nm, and 0.4% size of 7.6 nm diameter. At 45°C 98.4% of the proteins had size of 441.1 nm diameter, 1.1% size of 4.4 nm and 0.6% size of 0.1 nm diameter. Finally, at 60°C 99.9% were of a size of 406.7 nm diameter and the rest 0.1% of size 0.1 nm diameter.
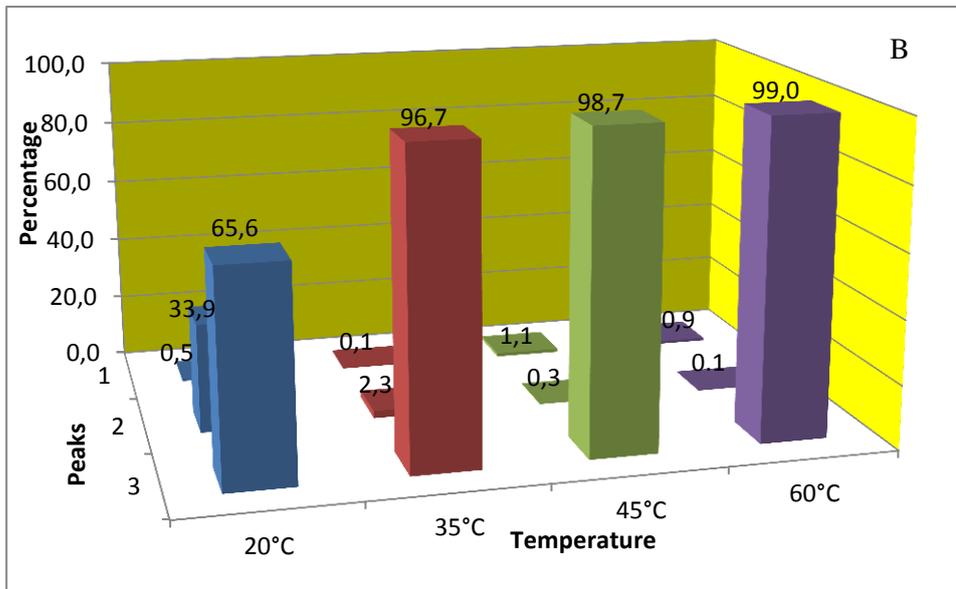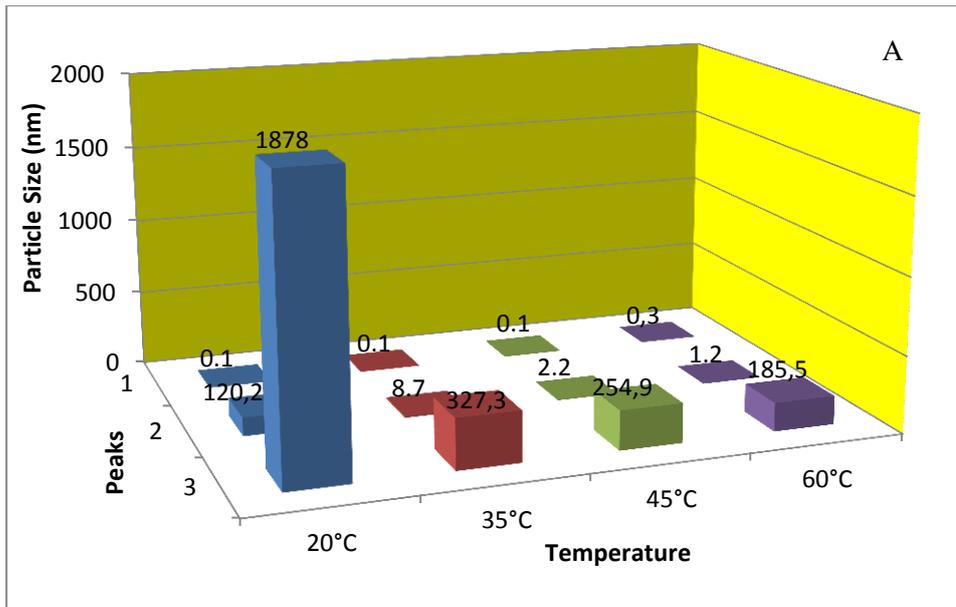
Figure 3.67: K87E Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at
pannel A vs. temperature. The values are shown at the top of each bar.
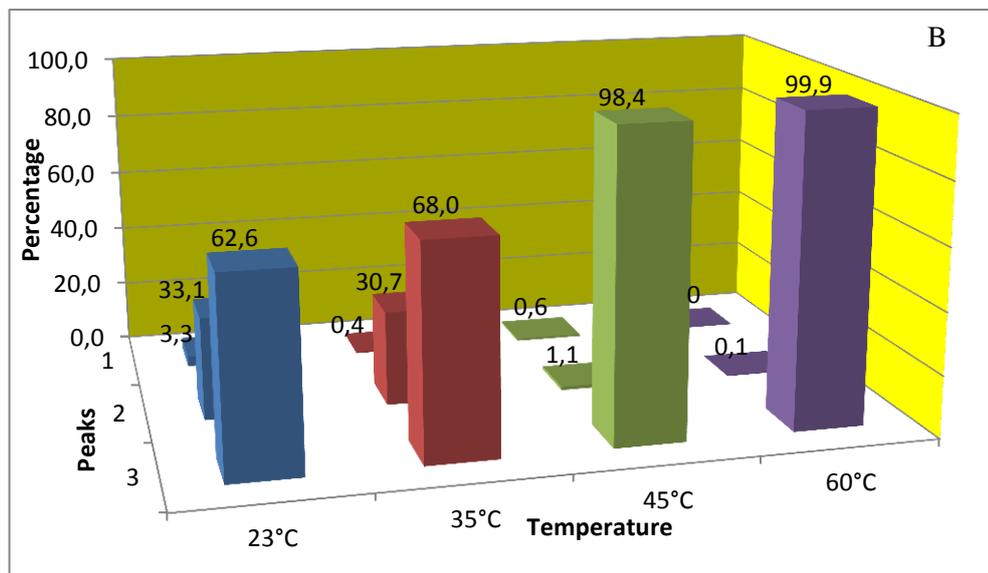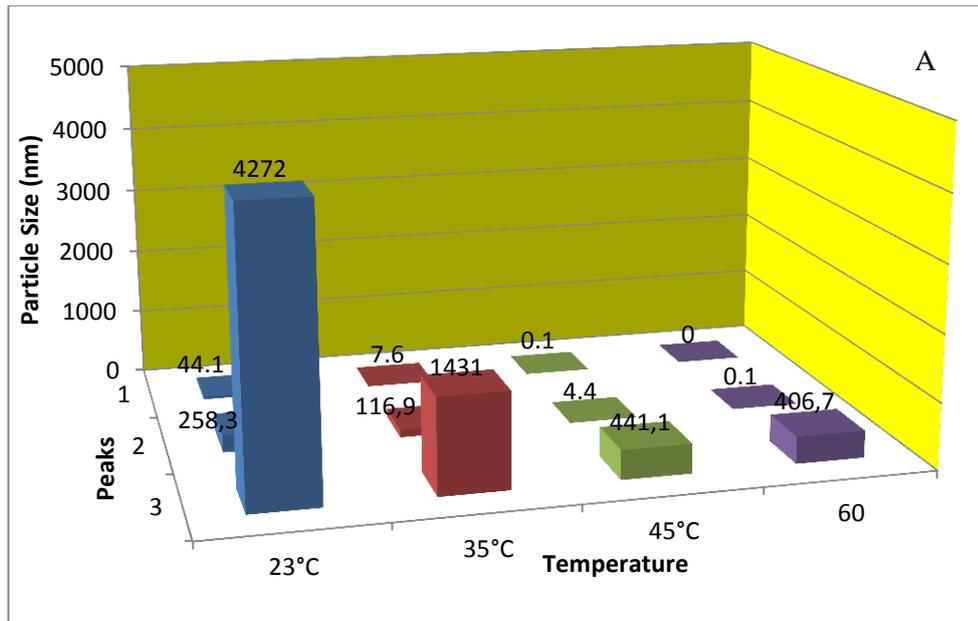
Figure 3.68: K87I Mutant Tpv-Hsp14.3 Oligomer Size and their Percent Distribution.
Pannel A: Multimer size vs. temperature. Pannel B: Percent size distribution of oligomers at pannel A vs. temperature. The values are shown at the top of each bar.

CHAPTER 4

DISCUSSION

Molecular chaperones are indispensable for maintaining cellular homeostasis. They are found in all domains of life, from unicellular organisms to humans, but there are differences in the number of genes, as well as the classes not only between different domains of life, but even between different classes of organisms (Waedick *et al.* 2009; Kim *et al.* 2013). They are involved in protein folding during synthesis, maintenance of proteins in a folded state during cellular stress, refolding after misfolding by stress or mutations, and directing protein aggregates to degradation to save the cell from the toxic effects of misfolded and aggregated proteins (Hartl *et al.* 2011). Their importance was noticed by the scientists due to the strong correlations found between mutations in chaperone genes and the diseases. This led to an explosion of research in this area, and to date it is found that molecular chaperones are involved practically in almost all the diseases, from hereditary diseases, to cancers, cardiovascular diseases, diabetes and neurodegeneration (Simon *et al.* 2007; Mymrikov *et al.* 2011; Kim *et al.* 2013). Indispensible components of this molecular chaperone network are the small heat shock proteins (sHSPs), for which less is known regarding the function and action mechanisms as compared to the other chaperone variants (Bukau *et al.* 2006; Kim *et al.* 2013).

In this study, the sHSP of a moderate thermophilic Archeon *Thermoplasma volcanium* Tpv-Hsp14.3 was investigated by a post-genomic approach. First, the proteins sequence retrieved from protein databank was used for sequence alignment to find well-conserved residues. These selected residues were further proven to be

important after secondary and tertiary structure predictions of the query protein. The 3D structure of the query was superposed on already resolved 3D structures from other organisms sHSPs and the locations of residues matching important amino acids involved in the function of sHSPs in other organisms were found. The mutated proteins were expressed in *E.coli*, and then purified for further investigations. After that, wild type and mutant Tpv-Shsp14.3 variants' enzyme activity protection assay was carried out to compare their chaperoning capacity. Finally, Dynamic Light Scattering (DLS) Spectroscopy was used to observe the oligomer sizes formed at different temperatures for both wild-type and mutant variants of the protein.

To date, only two proteins of hyperthermophilic Archaea have been thoroughly characterized, with their structures being determined by high resolution X-ray Crystallography, *M.janaschii* (Kim *et al.*, 1998) and *S.tokadii* (Takeda *et al.*, 2011). All other structurally and functionally characterized sHSPs are from Eukaryotes (van Montfort *et al.* 2001; Bagneris *et al.*, 2009; Jehle *et al.*, 2010; Laganowsky *et al.*, 2010; Laganowsky *et al.*, 2010; Hilario *et al.*, 2011). This study is unique for a number of reasons: First, Tpv-Hsp14.3 is not only the first moderate thermophilic Archaeal sHSP protein mutated and characterized to date, but also the first sHSP systematically mutated in the α-Crystalline domain. Second, a number of diverse methods were used for the protein characterization, especially the innovative methods of bioinformatics and computation biology to deduce and then mutate functionally important sites. Third, enzyme protection activity assay was used to assess the chaperoning function of the protein variants. This is more reliable than the most commonly used method of light scattering (Basha *et al.* 2011), since the latter method assumes that the protein is aggregated and has lost its function. But this may not be the case, as shown by Quinlan *et al.* (2013) who incubated Citrate Synthase (CS) with *M.janaschii* MjShsp16.5 R107G mutant, (the equivalent of human R120D mutant involved in cataract) and no significant change in enzyme activity was observed, although light scattering spectroscopy was indicating substrate aggregation. Last but not least, DLS Spectroscopy was used because it gives more reliable results regarding the dynamics of the chaperones in solution as compared to chromatographic methods. (Haslebeck *et al.* 2008).

Protein sequence analysis are in agreement with previous results, where a high sequence divergence is observed at the N- and C-termini, while the α-Crystalline domain is better conserved. Another important result is the conserved I/L-X-I/L motif at the C-terminal end (Waedick *et al.* 2009), which is defined this protein as V-X-I, and the same sequence is found in many other hyperthermophilic Archaea. The motif actually is I/L/V/-X/I/L/V as yielded by a comprehensive sequence alignment. At the beginning only eukaryotic and mainly bacterial sequences were used in sHSPs studies, and as a result the motif was mainly defined as I/L-X-I/L. For this reason, the conventional name is I/L-X-I/L. These three amino acids are all hydrophobic and interchangeable so the variation is expected. On the other hand, the possibility of valine being the residue of choice for hyperthermophiles should not be ruled out. Whether it is peculiar to hyperthermophiles should be investigated further as more sHSP sequence data added to gene and protein data banks for Archaea.

Our protein's secondary and tertiary structure prediction results are in agreement with the general structure of class II sHSPs (Haslebeck *et al.* 2008; Mchaourab *et al.* 2009; Basha *et al.* 2011). Eight β-strands, four on one side and three on the other side constituting a sandwich form an immunoglobulin-like domain. One strand called β6 is found in the middle of a long loop linking these two β-sheets together to form the "α-Crystalline domain" (ACD), the "flagship" of the sHSPs. ACD is flanked by a relatively long N-terminal region, mostly comprised of an α-helical structure and a shorter C-terminus which is coiled-coil.

After the 3D structures of the wild-type and mutant tpv-Hsp14.3 were predicted with I-TASSER and validated, they were compared to observe whether there was any gross structural change or not. We observed no significant change. This was expected since the modelling program only predicts the structure of the query based on the templates backbone, so it is not possible to predict structural changes for point mutations (Zhang, 2007). Then the structures of each mutant were compared to the wild-type protein structure by superposing them on each other. The RMSD values for these pairwise structure comparisons were 0.6-0.8, meaning there are no significant overall structural differences between the structures since the

conventional cut-off RMSD value is 2.0. RMSD values <2 show no significant differences between the structures being compared, and bigger values show significant structural differences (Maiorov *et al.* 1994). Next, another comparison was made for all three mutants and the wild type protein structures for a specific mutated residue, and the RMSD values varied from 0.9-1.2. In this case, the differences grow, but still they are smaller than 2.0.

The plasmid carrying the gene *tvn0775* encoding Tpv-Hsp14.3 with a 6-His Tag fusion was constructed previously in our lab and transformed to *E.coli* (Kocabiyik *et al.* 2012). The change of the intended sites was carried out successfully with QuickChange Site Directed Mutagenesis Kit, giving efficiencies of 10^6-10^7, while our attempt for mutagenesis with Transformer Kit (Chlontech) failed. The main reason seems to be the length of our plasmid, since in experiments with *pUC* vectors the efficiency was high. The replacement of the target nucleotides by the desired ones were further confirmed by DNA sequencing. Sequencing results showed that QuickChange kit had very good mutation efficiency, while the putative mutants from Transformer Kit showed internal deletions and changes of unintended sites.

Heat treatment of the cellular extract at 65°C and then running them in SDS-PAGE yielded some hints of the difference of the variants' stabilities. We observed roughly two classes of sHSP bands on the gel: In one class, there was no significant loss of the tpv-Hsp14.3 band as compared to that of samples not heated. In the second class, there was a significant loss of the tpv-Hsp14.3 band as compared to that of untreated sample. The only exception was K87I mutant whose band was very weak even before heat treatment and almost lost after that. These results may indicate that mutant sHSPs whose bands were weak after heating may be very unstable, but physical methods must be used to confirm this. The column purification of the chaperones from cellular extract was done according to these results. The samples whose bands were weakened by heat treatment were not treated by heat at 65°C before column purification, and these included R69E, R69M, R81M, R81E and K87I. The rest of the chaperone samples were purified after the cellular extract was heated to aggregate and precipitate most of the *E.coli* proteome by centrifugation.

Enzyme activity protection assay is an accurate method to measure the chaperoning activity of sHSPs (Kocabiyik *et al.* 2012). Several model enzymes have been used to carry out *in vitro* activity assays. Their activity was measured at elevated temperatures, in which it would be lost if the model protein was alone. After addition of sHSP in the solution and incubation for a certain time, the enzyme activity was protected. The elevated temperature in the experiments with Glutamate Dehydrogenase was performed at 56°C, the temperature at which the enzyme loses its activity very rapidly. In the presence of *Pyrococcus furiosus* sHSP, the enzyme was kept in a soluble form for an extended period of time, and its activity was protected, but it showed a decline with time (Laksanalamai *et al.* 2001). In another study, experiments were carried out with Citrate Synthase (CS) and *Thermococcus* sHSPs protected CS activity at 45°C. There were no direct evidences for this conclusion which was decided by quantifying the solubility of the substrate after heating (Usui et. al., 2001).

In this study we performed heat protection assays with sHSPs purified by column chromatography and checked by SDS-PAGE. Pig heart CS enzyme was used as the chaperone substrate. The enzyme's activity was measured first, under its optimal temperature without chaperone. Then, activity measurements were done at the temperature of 47°C both in absence and in the presence of each chaperone mutant. As the results in table 4.1 indicate, the activity of enzyme differed for each residue substitution. In position 69, when Arg was changed to its equivalent Lys, the activity was almost twice higher. The same result was obtained for R69E mutant, despite Glu being negatively charged while Arg positive. When Met, a hydrophobic residue, was the substituting amino acid, the activity was about the same as the wild-type.

The results obtained by X-Ray Crystallography and NMR show an important hydrogen bond between the counterparts of tpv-Hsp14.3 R69-E45 pair in other organisms: R107 interacts with D80 in humans (Jehle *et al.* 2010), R85 with D60 in wheat (van Montfort *et al.* 2001), and R66 interacts with D41 in *S.tokadii* (the interaction of the first two rely on experimental data, while the last relies on computational studies by Chimera) (Takeda *et al.* 2011).

161

Table 4.1: The Change in CS Activity in Presence of Tpv-Hsp14.3 Mutant Variants with Respect to the Activity Under Optimal Conditions and in the Presence of the Wild-Type sHSP.

Name	Sample Activity	CS Activity at 35°C/ CS+ sHSP Activity	Activity/WT sHSP + CS Activity
CS_35°C	823	1	5.0
CS_47°C	18.1	45	0.1
WT+ CS_47°C	160.1	5	1.0
R69E+ CS_47°C	302.412	2.7	1.9
R69K+ CS_47°C	293.5	2.9	1.8
R69M+ CS_47°C	141.3	5.8	0.9
R81E+ CS_47°C	20.01	41	0.1
R81K+ CS_47°C	103.345	8	0.6
R81M+ CS_47°C	15.47	53	0.1
QR(80-81)EL+ CS_47°C	194	4.2	1.2
K87E+ CS_47°C	212.3	3.9	1.3
K87R+ CS_47°C	412.108	2	2.6
K87I+ CS_47°C	429.135	1.9	2.7

The results here indicate that the chaperone activity is not affected even from very drastic mutations, such as replacing Arg with a hydrophobic residue in this position. This may be because of two reasons: either such a bond is not as important in *T.volcanium* sHSP as in the other organisms, which may be quite possible given the fact that such a bond is not found in *M.janaschii*, and/or such an amino acid exchange may cause a local structural disturbance which enhances the chaperone activity.

When position R81 is mutated in the same manner as position R69, the results are quite different: R81E and R81M have significantly reduced chaperoning activity as compared to the wild type. On the other hand, when Arg is changed into its equivalent Lys (R81K), the activity drops to almost half of the wild-type. It is important to note that R81 of *T.volcanium* is the equivalent of Wheat R101 which in turn is thought to be important in dimer formation by forming a hydrogen bond with D60 of the other monomer (van Montfort *et al.* 2001). R81 is also the equivalent of

R116 in human αB-Crystalline, which is suggested to be one of the important residues in dimer formation by interacting with F118 (Laganowsky *et al.* 2010). It is also the equivalent of *S.tokadii* R79 which according to hydrogen bond computations by Chimera software interacts with D41. In wheat and *S.tokadii* the equivalents of *T.volcanium* R81 in (R101 in wheat and R79 in S.tokadii) and *T.volcanium* R69 (R85 in wheat and R66 in S.tokadii) interact with D60 in wheat and D41 in *S.tokadii*, which are the equivalents of E45 in *T.volcanium*. Since Asp and Glu are equivalent residues with very similar biochemical properties, and found at the same positions in 3D structures of the three organisms, the idea of *T.volcanium* E45 being the counterpart of wheat D60 and *S.tokadii* D41 is plausible. But why changing R69 did not cause such a drastic decrease in chaperoning activity as caused the changes in R81 is not known, and it needs more detailed structural studies. The results of the double mutant showed an increase in chaperoning activity. It was changed at positions 80 and 81 from Gln to Glu and Arg to Leu, respectively. This is a drastic sequence change, but the results can be explained by the similarity of the mutant sequence at these positions to that of *M.janaschii.* According to 3D structure superposition, the sequence of *M.janaschii* sHSP at this position is EI(98-99) (Kim *et al.* 1998). *T.volcanium* E80 is the equivalent of *M.janaschii* E98 and *T.volcanium* L81 is equivalent of *M.janaschii* I99, although less hydrophobic.

The best studied position to date is K87, whose equivalents in other organisms are found to be important residues in chaperone function and structure. Its equivalent in human is R120, in wheat R108, in *M.janaschii* R107 and in *S.tokadii* K85. Most of the studies, both functional and structural were carried out with human αB-Crystalline by which R120 was found to form a network of hydrogen bonds, and the mutant form of it is involved in many diseases (Clark *et al.* 2000; Graw, 2009; Horwitz, 2009; Clark *et al.* 2012). High resolution studies have shown that the R120G mutant has an altered crystal structure enclosing the groove in the α-Crystalline domain since Gly cannot form hydrogen bonds with any residue (Clark *et al.* 2011; Baranova *et al.* 2011). However, when the same mutation was introduced to *M.janaschii* MjsHsp16.5 the crystal structure of the mutant R107G did not show any

difference from the wild type (Quinlan *et al.* 2013). Also protection of CS activity by wild-type and mutant MjShsp16.5 proteins was almost same, but the mutant did not prevent β-crystalline aggregation at elevated temperature. This result may be explained as follows: First, the sHSP may show substrate preference, protecting some better than others (Das *et al.* 1996; Bettelheim *et al.* 1999; Das *et al.* 1999; Carver *et al.* 2002), or it may bind the substrate protein with a higher affinity, and it falsely appears to be aggregating by light scattering spectroscopy (Claxton *et al.* 2008). But to date there are no studies where the enzyme is tested for its activity after light scattering to show whether the apparent aggregated enzyme substrate has activity or not. The Cryo-EM studies of MjShsp16.5 showed a difference in the oligomerization of R107G as compared to wild-type protein at elevated temperature. Since the oligomeric structure is the one that binds and protects the substrate, this may be reflected in the chaperones function and disease cases (Quinlan *et al.* 2013). The reason why there is no structural changes in MjHsp16.5 R107G as in human R120G mutant may be because of their differences in dimerization, with the latter forming an interface between β6+7 strands, while the former dimerizing mostly by β6 strand swapping as discussed before (Quinlan *et al.* 2013).

The enzyme assay results Tpv-Hsp14.3 at position K87 were quite unexpected. All the substitutions resulted in an increased chaperoning activity as compared to wild-type. K87E had a 1.3-fold higher activity, as the replacing amino acid has different biochemical properties than Lys. On the other hand, K87R and K87I had similar chaperone activity, despite the significant chemical differences between Arg and Ile, the former basic and the latter highly hydrophobic. The higher activity of K87R may be explained by the fact that Arg is the residue in this position for all the chaperones whose structure have been determined except *S.tokadii* StHsp14.5. This may be attributed to a higher chaperoning activity of the sHSPs with Arg instead of Lys, but there is no study comparing enzyme activity protection by different sHSPs under the same conditions to date. As for K87I, the explanation doesn't seem very straight forward, but one explanation may be a higher substrate binding capacity. The explanation of increased hydrophobicity being correlated with increased chaperone

activity seems tempting. However, in the other positions where hydrophobic residues replaced the original basic residues chaperone activity of the sHSP variants was decreased. Moreover K87R mutant had similar activity with K87I mutant, even though Arg is a positively charged amino acid. Since mutation at equivalent positions in human and *M.janaschii* showed very different results, it may be possible for sHSPs of different organisms, mutations at this position may have different effects, making it impossible to make plausible inferences by comparisons. A good way to observe what happens to the multimeric form of this sHSP when mutated at this position is a detailed structural analysis by techniques such as Cryo-EM.

Dynamic Light Scattering (DLS) Spectroscopy has been used previously to monitor the prevention of protein aggregation by chaperones. In one study Bettelheim *et al.* (1999) monitored Dithiothreitol-denatured α-Lactalbumin that was complexed complex with α-Crystalline in a period of 24 hours. They were able to see discrete bands by DLS, which showed different chaperone-substrate complexes. DLS is the method of choice to monitor protein aggregation (Schuler *et al.* 1999; Kumar *et al.* 2007). But there are very few reports on monitoring sHSPs dynamics by DLS to date (Haslebeck *et al.* 2008).

The idea behind using DLS is that as the temperature of the solution is increased, the chaperones will form higher oligomeric complexes, increasing particle size (Basha *et al.* 2011). This change in particle size is simultaneously quantified by the spectroscopy, which gives both, particle size in diameter, and their distribution in the entire solution. The time of measurement was chosen to be 10 seconds, meaning the light scattering was measured during a 10 seconds period by sending the light from the laser and measuring it's scattering. This is reasonable regarding sHSPs rapid association/dissociation properties (Benesch *et al.* 2010; Haslebeck *et al.* 2008; Mchaourab *et al.* 2009). When DLS measurement was taken for 30 seconds, there was a single peak of a broad diameter range from 600-1400nm (Fig.3.54). The principle of DLS is to measure the light scattering and transform it into averaged quantitative data, and this result show that the predominant form of the chaperones in solution is in large protein complexes. When DLS measurement was taken for 10

seconds, the results showed a decrease in complex diameter with the increasing temperature (Fig. 3.55) which was not always smooth, but taken into account the principles of DLS explained in section 2.3.11 it can be assumed to be a regular trend of tpv-Hsp14.3 as will be shown latter.

To study sHSP oligomerization dynamics in the presence of substrate protein we used two model enzymes, pig heart citrate synthase (CS) and bovine Glutamate Dehydrogenase (GDH). The enzymes were incubated with tpv-Hsp14.3 at temperatures they are normally denatured. As is expected from the sHSPs, with increased temperature they will associate with themselves and the substrate to give larger complexes as compared to low temperatures. In figure 3.56 the trend is not exactly like this, but there is a shift in the peak sizes from very big ones to smaller peaks. This can be explained in the following way: At 47°C sHSPs are associated into big complexes with each other giving two main peaks. The large peak size is mainly because of the high sHSP concentration in solution and also the "perfect sphere" assumption of DLS. When CS is added, it will disturb the equilibrium of sHSPs in solution, forcing them to dissociate and re-associate again with the enzyme substrate under stress conditions, and free sHSPs in solution. But since the CS concentration compared to sHSPs is very low, not all sHSPs are associated with the enzyme. sHSPs will form regular complexes binding strongly to the enzyme, giving smaller peaks. In the case of GDH (Fig. 3.57) there is no significant change in peak size after incubation with the enzyme, but in this case the sHSPs concentration is lower as compared to the previous case. It may also reflect substrate preferences, as reported for other sHSPs (Claxton *et al*. 2008).

The last DLS measurements were carried out for all the mutant variants of tpv-Hsp14.3 and the wild-type under the same temperature and the same concentration. Generally 3-4 peaks are obtained by DLS, but only the three peaks that have a percentage of at least 0.1% of the entire solution are shown. Of these only two peaks have diameters that theoretically represent at least a sHSP dimmer, which is the basic unit of sHSPs (Benesch *et al.* 2010; Basha *et al.* 2011).

The general pattern for oligomere size distribution in all sHSP variants was a decrease in diameter size (Panels A) and shifting of the particle size distribution toward a single main peak as shown by the percent distribution (Panels B) in figures 3.58-3.68. Given the DLS assumptions, it is impossible to see minor size differences between the complexes, and sometimes they may be of very close size, but if the general pattern of the peaks is judged, a decrease in particle diameter with increasing temperature is observed. The only exception from this are the peaks for R81M mutant (Fig. 3.64), which generally shows two main peaks of almost equal distribution, and a general pattern of decreasing oligomere size was not observed. SDS results for this mutant protein showed no difference between heat-treated and untreated samples, as judged by band intensities, but chaperone activity was very low.

Since DLS measurements were done between 20-60°C, it can be argued that bigger size at 20°C and subsequent decrease towards 60°C is as a result of cold shock. This is what some researchers have argued (Haslebeck *et al.* 2008), since cold is also a physiological stress condition for organisms other than psychrophiles. *T.volcanium* grows optimally at 60°C, and its growth range is 33-67°C, fitting this explanation. Haslebeck *et al* (2008) used Cryo-EM to determine the multimeric structure MjHsp16.5 and quantified the results. They found that the high-binding mode (shown in figure figure 1.17) of MjHsp16.5 was distributed almost equally in both heat and cold-stress temperatures, by being the major component of sHSP complexes, while the low-binding mode of the chaperone was found to be the major component in optimal temperatures.

Given the short time of scattering measurement, the dynamics of sHSPs in solution and the way DLS assumes a perfect sphere, the DLS results of Tpv-Hsp14.3 can be explained as shown in figure 4.1.
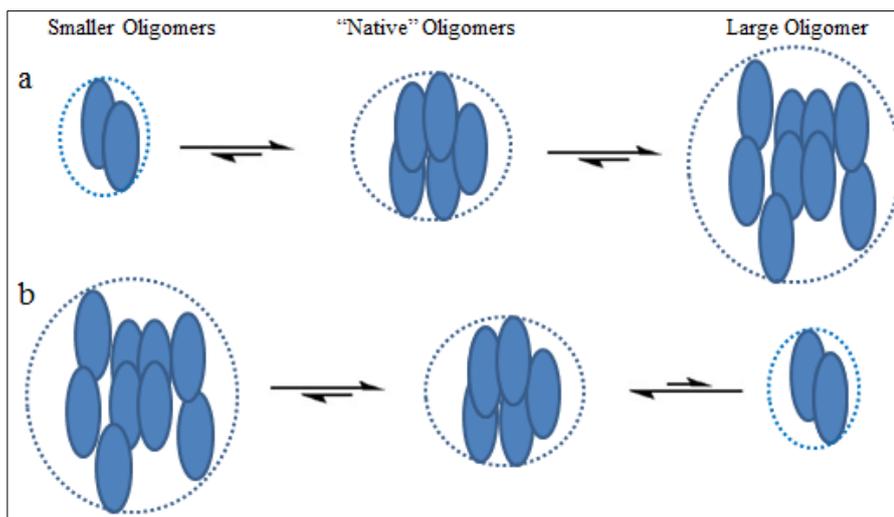
Figure 4.1: DLS Peak Distribution Related to Temperature. (a) sHSPs dynamics during cold/heat shock. (b) sHSPs behavior as the temperature approaches the organisms optimal growth conditions. The oval shapes do not represent sHSP monomers or dimmers, but they are shown only for visual purpose. The dashed circle shows the perfect sphere assumed by DLS for any particle in solution. Arrows show the direction of sHSP association /dissociations.

Since scattering is measured in 10 seconds time periods, many random interactions are measured and averaged by the software to show them as peaks. Given also the assumption of DLS that the particle is an ideal sphere, loose protein interactions are detected by the spectrometry to form real complexes of such big sizes, as may be the case with some large peaks in figures 3.55, 3.58 and 3.64. This is the possible explanation for the very large diameters of the protein complexes.

At low temperatures, which can be considered as stress for *T.volcanium* cells, the equilibrium is shifted towards large complexes which represent the stress mode of sHSPs (Figure 4.1.a). As the temperature is increased, the solution approaches more to the optimal conditions for *T.volcanium* cells, so the chaperones shift from stress mode into physiological mode. This phenomenon is accompanied by an equilibrium shift towards the "native" multimeric structure of the sHSPs, so the peaks decrease in size and accumulate in a single major peak. To accomplish this, the large sHSP

168

complexes will dissociate and re-associate into multimers resembling the complex in the physiological state of the organism (figure 4.1.b). However, given the high protein concentration in the solution still random interactions cannot be avoided. Adding here the hydration effect, this may explain the smaller size of particles with increasing temperature. However, for the reasons given above, particles diameters are bigger than their real sizes.

REFERENCES

Ahrman, E., Lambert, W., Aquilina, J.A., Robinson, C.V., Emanuelsson, C.S. (2007) Chemical cross-linking of the chloroplast localized small heat-shock protein, Hsp21, and the model substrate citrate synthase. *Protein Sci.* 7, 1464-78.

Åkerfelt, M, Morimoto, R., Richard, Sistonen (2010) Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell. Biol.* 11, 545-555.

Albrecht, M., Tosatto, S.C., Lengauer, T., Valle, G. (2003) Simple consensus procedures are effective and sufficient in secondary structure prediction. *Protein Eng*. **16**, 459-462.

Ali, M.M.U., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W., Prodromou, C., Pearl, L.H. (2006) Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* **440,** 1013−1017.

Altschul, S.F., Gish,W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped blast and Psi-blast: A new generation of protein database search programs, *Nucleic Acids Res*, **25**, 3389–3402.

Anfinsen, C.B., (1973) Principles that govern the folding of protein chains. *Science* **181** (4096): 223–230

Arias, E., Cuervo, A.M. (2011) Chaperone-mediated autophagy in protein quality control. *Curr. Opin.Cell Biol.* 23:184–189.

Arndt, V., Dick, N., Tawo, R., Dreiseidler, M., Wenzel, D., Hesse, M., Furst, D.O., Saftiq, P., Saint, R., Fleischmann, B.K., Hoch, M., Hohfeld, J. (2010) Chaperone-assisted selective autophagy is essential for muscle maintenance. *Curr. Biol.* 20:143–148.

Assimakopoulou, M., Sotiropoulou-Bonikou, G., Maraziotis, T., Varakis, I. (1997) Prognostic significance of Hsp-27 in astrocytic brain tumors: an immunohistochemical study. *Anticancer Res* 17: 2677–2682.

Augusteyn, R.C. (1998) alpha-Crystallin polymers and polymerization: The view from down under. Int J Biol Macromol 22:253–262.

Babst, M., Hennecke, H., Fischer, H.M. (1996) Two different mechanisms are involved in the heat-shock regulation of chaperonin gene expression in *Bradyrhizobium japonicum. Mol. Microbiol*. **19:**827–839.

Bagneris, C., Bateman, O.A., Naylor, C.E., Cronin, N., Boelens, W.C., Keep, N.H., Slingsby, C. (2009) Crystal structures of a-crystallin domain dimers of aB-crystallin and Hsp20. *J. Mol. Biol*. 392, 1242–1252.

Balch, W.E., Morimoto, R.I., Dillin, A., Kelly, J.W. (2008) Adapting proteostasis for disease intervention. *Science* **319,** 916−919.

Baldwin, A.J., Hilton, G.R., Lioe, H., Bagnéris, C., Benesch, J.L., Kay, L.E. (2011) Quaternary dynamics of aB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J. Mol. Biol*. 413, 310–320.

Baldwin, A.J., Lioe, H., Hilton, G.R., Baker, L.A., Rubinstein, J.L., Kay, L.E., Benesch, J.L. (2011) The polydispersity of aB-crystallin is rationalised by an inter-converting polyhedral architecture. *Structure* 19(12):1855-1863.

Baranova, E.V., Weeks, S.D., Beelen, S., Bukach, O.V., Gusev, N.B., Strelkov, S.V. (2011) Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J Mol Biol*. 411(1):110-122.

Basha, E., O'Neill, H., Vierling, E. (2011). Small heat shock proteins and a-crystallins: dynamic proteins with flexible functions. *Trends in Biochemical Sci*. 37(3): 106–117.

Bell, S.D., Jackson, S.P. (1998) Transcription and translation in Archaea: a mosaic of eukarya and bacterial features. *Trends Microbiol*. **6:**222–228.

Benesch, J.L., Aquilina, J.A., Baldwin, A.J., Rekas, A., Stengel, F., Lindner, R.A., Basha, E., Devlin, G.L., Horwitz, J., Vierling, E., Carver, J.A., Robinson, C.V. (2010) The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated. *Chemistry & Biology.* 17, 1008–1017.

Ben-Zvi, A., Miller, E.A., Morimoto, R.I. (2009) Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *PNAS.* 106:14914–14919.

Berengian, A.R., Parfenova, M., Mchaourab, H.S. (1999) Site-directed spin labeling study of subunit interactions in the alpha-crystallin domain of small heat-shock proteins. Comparison of the oligomer symmetry in alphaA-crystallin, HSP 27, and HSP 16.3. *J. Biol. Chem*. 274, 6305–6314.

Berne, B.J., Pecora, R. (2000). *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*. Dover Publications.

Bertz, M., Chen, J., Feige, M.J., Franzmann, T.M., Buchner, J., Rief, M. (2010) Structural and mechanical hierarchies in the alpha-crystallin domain dimer of the hyperthermophilic small heat shock protein Hsp16.5. *J Mol Biol.* 30;400(5):1046-56.

Bettelheim, F.A., Ansari, R., Cheng,Q. F., Zigler, J.S. (1999) The mode of chaperoning of dithiothreitol-denatured R-lactalbumin by R-crystallin. *Biochem. Biophys. Res. Commun*. 261, 292–297.

Boya, P., Reggiori, F., Codogno, P. (2013) Emerging regulation and functions of autophagy. *Nat Cell Biol.* 15(7):713-20.

Bradley, P., Misura, K.M., Baker, D. (2005) Toward high-resolution de novo structure prediction for small proteins, *Science*, **309**, 1868–1871.

Branden, T. & Tooze, J., (2000). *Introduxtion to protein Structure.* Garland Science; 2nd Edition.

Braun, N., Zacharias, M., Peschek, J., Kastenmüller, A., Zou, J., Hanzlik, M., Haslbeck, M., Rappsilber, J., Buchner, J., Weinkauf, S. (2011) Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach. *PNAS*. 20;108(51):20491-20496.

Brown, I.R. (2007) Heat shock proteins and protection of the nervous system. Ann N Y Acad Sci. 1113:147-58.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G., (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct. Func. Gent.* **3**, 167-95.

Buchan, R.J., & Parker, R. (2009) Eukaryotic stress granules: The ins and outs of translation. *Mol. Cell* 36, 932–941.

Buchberger A, Bukau B, Sommer T. (2010) Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol Cell.* 22;40(2):238-352.

Bujnicki J. (2009) *Prediction of Protein Structure, functions and interactions*. John Wiley and Sons.

Bukau, B., Weissman, J., Horwich, A. (2006) Molecular Chaperones and Protein Quality Control. *Cell*. 5, 125(3):443-451.

Carver, J.A., Lindner, R.A., Lyon, C., Canet, D., Hernandez, H., Dobson, C.M., Redfield, C. (2002) The interaction of the molecular chaperone alpha-crystallin with unfolding R-lactalbumin: A structural and kinetic spectroscopic study. *J. Mol. Biol.* 318, 815–827.

Chaudhuri, T.K., Paul, S. (2006) Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J*. 273, 1331-1349.

Cheng, G., Basha, E., Wysocki, V. H., Vierling, E. (2008) Insights into small heat shock protein and substrate structure during chaperone action derived from hydrogen/deuterium exchange and mass spectrometry. *J. Biol. Chem*. 283, 26634–26642.

Ciocca, D.R., Fuqua, S.A., Lock-Lim, S., Toft, D.O., Welch, W.J., McGuire, W.L. (1992) Response of human breast cancer cells to heat shock and chemotherapeutic drugs. *Cancer Res* 52: 3648-3654.

Clark, A.R., Lubsen, N.H., Slingsby, C. (2012) sHSP in the eye lens: crystallin mutations, cataract and proteostasis. *Int J Biochem Cell Biol*. 44(10):1687-1697.

Clark, A.R., Naylor, C.E., Bagnéris, C., Keep, N.H., Slingsby, C. (2011) Crystal structure of R120G Disease mutant of human alphaB-crystallin domain dimer shows closure of a groove. *J. Mol. Biol.* 408, 118–134.

Clark, J.I., Muchowski, P.J. (2000) Small heat-shock proteins and their potential role in human disease. *Curr. Opin. Struct. Biol*. 10, 52–59.

Claxton, D.P., Zou, P., McHaourab, H. S. (2008) Structure and orientation of T4 lysozyme bound to the small heat shock protein alpha-crystallin. *J. Mol. Biol.* 375, 1026–1039.

Cohen, E., Bieschke, J., Perciavalle, R.M., Kelly, J.W., Dillin, A. (2006) Opposing activities protect against age-onset proteotoxicity. *Science* 313:1604–1610.

Cohen, E., Paulsson, J.F., Blinder, P., Burstyn-Cohen, T., Du, D., Estepa, G., Adame, A., Pham, H.M., Holzenberger, M., Kelly, J.W., Masliah, E., Dillin, A. (2009) Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell* 139:1157–1169.

Cole, C., Barber, J.D., Barton, G.J. (2008) The Jpred 3 secondary structure prediction server, *Nucleic Acids Res.* 36 (suppl. 2): W197-W201.

Conn, P. M. (Ed). (2008). *Progress in molecular biology and translational science*, *Vol. 84*. Elsevier.

Cuff, J.A., Barton, G.J. (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **40**, 502–511.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., Barton, G.J. (1998) Jpred:Aconsensus secondary structure prediction server. *Bioinformatics* **14**, 892–893.

Das, K. P., Choo-Smith, L. P., Petrash, J. M., Surewicz, W. K. (1999) Insight into the secondary structure of non-native proteins bound to a molecular chaperone R-crystallin. An isotope-edited infrared spectroscopic study. *J. Biol. Chem.* 274, 33209–33212.

Das, K. P., Petrash, J.M., and Surewicz, W.K. (1996) Conformational properties of substrate proteins bound to a molecular chaperone alpha-crystallin. *J. Biol. Chem.* 271, 10449–10452.

David, D.C., Ollikainen, N., Trinidad, J.C., Cary, M.P., Burlingame, A.L., Kenyon, C. (2010) Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol.* 8:e1000450.

De Nadal, E., Ammerer, G., Posas, F., (2011) Controlling Gene expression in response to stress. *Nature Rev. Gnet.* 12, 833-845.

De Nadal, E., Posas, F. (2010) Multilayered control of gene expression by stress-activated protein kinases. *EMBO J.* 29, 4-13.

Deng, M., Chen, P.C., Xie, S., Zhao, J., Gong, L., Liu, J., Zhang, L., Sun, S., Liu, J., Ma, H., Batra, S.K., Li, D.W. (2010) The small heat shock protein alphaA-crystallin is expressed in pancreas and acts as a negative regulator of carcinogenesis. *Biochim. Biophys. Acta.* 1802, 621–631.

Depre, C., Wang, L., Sui, X., Qiu, H., Hong, C., Hedhli, N., Ginion, A., Shah, A., Pelat, M., Bertrand, L., Wagner, T., Gaussin, V., Vatner, S.F. (2006) H11 kinase prevents myocardial infarction by preemptive preconditioning of the heart. *Circ Res.* 98, 280 –288.

Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol*. 4, 10-19.

Doyle, S.M., Wickner, S. (2008) Hsp104 and ClpB: protein disaggregating machines. *Trends in Biochemical Sciences*. 34, 40-48.

Dunker, A. K., Silman, I., Uversky, V. N., Sussman, J. L. (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18,** 756−764.

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., Uversky, V.N. (2005) Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, *272*, 5129-5148.

Ehrnsperger, M., Graber, S., Gaestel, M., Buchner, J. (1997) Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J*. 16, 221–229.

Elcock, A.H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLOS Comput. Biol.*2, e98.

Ellis, R. J. & Minton, A. P. (2006) Protein aggregation in crowded environments. *Biol. Chem.* **387,** 485−497.

England, J.L., Lucent, D., Pande, V.S. (2008) A role for confined water in chaperonin function. *J. Am. Chem. Soc.* 130, 11838–11839.

Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A. (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.

Fan, G.C., Ren, X., Qian, J., Yuan, Q., Nicolaou, P., Wang, Y., Jones, W.K., Chu, G., Kranias, E.G. (2005) Novel cardioprotective role of a small heat-shock protein, Hsp20, against ischemia/reperfusion injury. *Circulation*. 111, 1792–1799.

Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu.Rev. Biochem.* 78:477–513.

Forman, M.S., Trojanowski, J.Q., Lee, V.M. (2004) Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nature Medicine.* 10; 1055-1063.

Frangsmyr , T. (Ed). (1993). *Nobel Lectures, Chemistry 1971–1980.* World Scientific Publishing Co., Singapore.

Franzmann, T. M., Menhorn, P., Walter, S., Buchner, J. (2008) Activation of the chaperone Hsp26 is controlled by the rearrangement of its thermosensor domain. *Mol. Cell.* 29, 207–216.

Franzmann, T. M., Wuhr, M., Richter, K., Walter, S., Buchner, J. (2005) The activation mechanism of Hsp26 does not require dissociation of the oligomer. *J. Mol. Biol.* 350, 1083–1093.

Fraser, C.M., Gocayne, J., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.L., Kerlavage, A.R., Sutton, G.G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A.,Venter, J.C. (1995) The minimal gene complement of Mycoplasma genitalium. *Science* 270:397–403.

Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F. U. (1994) Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370,** 111−117.

Galigniana, M.D., Echeverría, P.C., Erlejman, A.G., Piwien-Pilipuk, G. (2010) Role of molecular chaperones and TPR-domain proteins in the cytoplasmic transport of steroid receptors and their passage through the nuclear pore. *Nucleus* 1:4, 299-308.

Gautschi, M., Mun, A., Ross, S., Rospert, S. (2002) A functional chaperone triad on the yeast ribosome. *PNAS*. 99, 4209–4214.

Ghosh, J. G., Estrada, M. R., Clark, J. I. (2005) Interactive domains for chaperone activity in the small heat shock protein, human RB crystallin. *Biochemistry* 44, 14854–14869.

Ghosh, J. G., Shenoy, A. K.Jr., and Clark, J. I. (2006) N- and C-Terminal motifs in human alphaB-crystallin play an important role in the recognition, selection, and solubilization of substrates. *Biochemistry* 45, 13847–13854.

Giese, K. C., Vierling, E. (2002) Changes in oligomerization are essential for the chaperone activity of a small heat shock protein in vivo and in vitro. *J. Biol. Chem*. 277, 46310–46318.

Giese, K. C., Vierling, E. (2004) Mutants in a small heat shock protein that affect the oligomeric state. Analysis and allele-specific suppression. *J. Biol. Chem*. 279, 32674–32683.

Gober, M.D., Smith, C.C., Ueda, K., Toretsky, J.A., Aurelian, L. (2003) Forced expression of the H11 heat shock protein can be regulated by DNA methylation and trigger apoptosis in human cells. *J Biol Chem* 278: 37600–37609.

Gober, M.D., Wales, S.Q., Aurelian, L. (2005) Herpes simplex virus type 2 encodes a heat shock protein homologue with apoptosis regulatory functions. *Front Biosci* 10: 2788–2803.

Graw, J. (2009) Genetics of crystallins: cataract and beyond. *Exp. Eye Res*. 88, 173–189.

Gromiha, M.M., Selvaraj, S. (2004) Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol*. 86, 235–277.

Gross, C.A. (1996) Function and regulation of the heat shock proteins, p. 1382–1399. *In* F. C. Neidhardt *et al.* (ed.), Escherichia coli and Salmonella: cellular and molecular biology. American Society for Microbiology, Washington, D.C.

Guo, Y., Guettouche, T., Fenna, M., Boellmann, F., Pratt, W.B., Toft, D.O., Smith, D.F., Voellmy, R. (2001) Evidence for a mechanism of repression of heat shock factor 1 transcriptional activity by a multichaperone complex. *J. Biol. Chem.* **276**, 45791–45799 (2001).

Hanazono, Y., Takeda, K., Yohda, M., Miki, K. (2012) Structural studies on the oligomeric transition of a small heat shock protein, StHsp14.0. *J Mol Biol*. 422(1):100-108.

Hanson, P.I., Whiteheart, S.W. (2005) AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol*. 6, 519–529.

Hartl, F. U. & Hayer-Hartl, M. (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Struct. Mol. Biol.* **16,** 574−581.

Hartl, F. U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381,** 571−580.

Hartl, F.U., Bracher, A., Hart, M.H., (2011) Molecular chaperones in protein folding and proteostasis. Nature. 475, 324-332.

Haslbeck, M., Kastenmüller, A., Buchner, J., Weinkauf, S., Braun, N. (2008) Structural Dynamics of Archaeal Small Heat Shock Proteins. *J. Mol. Biol.* 378, 362–374.

Haslbeck, M., Walke, S., Stromer, T., Ehrnsperger, M., White, H.E., Chen, S., Saibil, H.R., Buchner, J. (1999). Hsp26: a temperature-regulated chaperone. *EMBO J*. 18, 6744–6751.

Haynes, C.M., Ron, D. (2010) The mitochondrial UPR-protecting organelle protein homeostasis. *J. Cell Sci.* 123:3849–3855.

Herbst, R., Schafer, U., Seckler, R. (1997) Equilibrium intermediates in the reversible unfolding of firefly (*Photinus pyralis*) luciferase. *J. Biol. Chem.* **272,** 7099−7105.

Hershko, A., Ciechanover, A. (1998) The ubiquitin system. *Annu. Rev. Biochem.* 67:425–79.

Hilario, E., Martin, F.J., Bertolini, M.C., Fan, L. (2011) Crystal structures of xanthomonas small heat shock protein provide a structural basis for an active molecular chaperone oligomer. *J. Mol. Biol.* 408, 74–86.

Horwich, A.L., Fenton, W.A., Chapman, E., Farr, G.W. (2007) Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell Dev. Biol.* 23, 115–145.

Horwitz, J. (2000) The function of α-crystallin in vision. *Semin. Cell Dev. Biol.* 11, 53–60.

Horwitz, J. (2009) alpha crystallin: the quest for a homogeneous quaternary structure. *Exp. Eye Res.* 88, 190–194.

Hughes, A.B. (Ed.). (2009) *Amino Acids, Peptides and Proteins in Organic Chemistry, Volume 1, Origins and Synthesis of Amino Acids.* New York: Wiley-VCH.

Hundley, H., Eisenman, H., Walter, W., Evans, T., Hotokezaka, Y., Wiedman, M., Craig, E. (2002) The in vivo function of the ribosome-associated Hsp70, Ssz1, does not require its putative peptide-binding domain. *PNAS.* 99, 4203–4208.

Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., Michalik, A., De Vriendt, E., Jacobs, A., Van Gerwen, V., Vennekens, K., Mazanec, R., Tournev, I., Hilton-Jones, D., Talbot, K., Kremensky, I., Van Den Bosch, L., Robberecht, W., Van Vandekerckhove, J., Van Broeckhoven, C., Gettemans, J., De Jonghe, P., Timmerman, V. (2004) Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat Genet.* 36(6), 597-601.

Jahn T.R., Radford S.E., (2008) Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* 469:100–117.

Jaya, N., Garcia, V., Vierling, E. (2009) Substrate binding site flexibility of the small heat shock protein molecular chaperones. *PNAS.* 106(37), 15604-15609.

Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S. Kühne, R., Stout, J.R., Higman1, V.A., Klevit, R.E., Rossum, B., Oschkinat, H. (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nat. Str. Mol. Biol.* 17 (9): 1037-1043.

Jehle, S., Vollmara, B.S., Bardiauxb, B., Dovea, K.K., Rajagopala, P., Gonena, P., Oschkinatb, H., Klevita, R.E. (2011). N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *PNAS* 108(16): 6409-6414.

Kaganovich, D., Kopito, R., Frydman, J. (2008) Misfolded proteins partition between two distinct quality control compartments. *Nature* 454:1088–1095.

Kaiser, C.M., Chang, H.C., Agashe, V.R., Lakshmipathy, S.K., Etchellas, S.A., Hayer-Hartl, M., Hartl, F.U., Barral., J.M. (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* 444, 455–460.

Kamada, M., So, A., Muramaki, M., Rocchi, P., Beraldi, E., Gleave, M. (2007) Hsp27 knockdown using nucleotide-based therapies inhibit tumor growth and enhance chemotherapy in human bladder cancer cells. *Mol. Cancer Ther.* 6, 299–308.

Kappe, G., Aquilina, A., Wunderink, L., Kamps, B., Robinson, C.V., Garate, T., Boelens, W.C., de Jong, W.W. (2004) Tsp36, a Tapeworm Small Heat-Shock Protein with a Duplicated α-Crystallin Domain, Forms Dimers and Tetramers with Good Chaperone-like Activity. *PROTEINS: Structure, Function, and Bioinformatics* 57:109–117.

Keese, A.M., Schut, G.J., Ouhammouch, M., Adams M.W.W., Thomm, M. (2010). Genome-Wide Identification of Targets for the Archaeal Heat Shock Regulator Phr by Cell-Free Transcription of Genomic DNA. *J. Bacteriol.* 192(5):1292-1298.

Kendrew, J.C., Bodo, G,. Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181 (4610): 662-666.

Kim, K.K., Kim, R., and Kim, S.H. (1998) Crystal structure of a small heat shock protein. *Nature* 394, 595–599.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., Hartl, F.U. (2013) Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* 82, 323–355.

King, R.D., Ouali, M., Strong, A.T., Aly, A., Elmaghraby, A., Kantardzic, M., Page, D. (2000) Is it better to combine predictions? *Protein Eng*. **13**, 15–19 (2000).

Kocabıyık, S. Aygar, S. (2012) Improvement of protein stability and enzyme recovery under stress conditions by using a small HSP (tpv-HSP 14.3) from *Thermoplasma volcanium. Process Biochemistry.* 47, 1676–1683.

Kokke, B.P., Boelens, W.C., de Jong, W.W. (2001) The lack of chaperonelike activity of Caenorhabditis elegans Hsp12.2 cannot be restored by domain swapping with human alphaB-crystallin. *Cell Stress Chaperones.* 6(4):360-7.

Kokke, B.P., Leroux, M.R., Candido, E.P., Boelens, W.C., de Jong, W.W. (1998) Caenorhabditis elegans small heat-shock proteins Hsp12.2 and Hsp12.3 form tetramers and have no chaperone-like activity. *FEBS Lett.* 21;433(3):228-32.

Kolinski, A., Bujnicki, J.M. (2005) Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins* **61 Suppl 7**, 84–90.

Kon, M., Cuervo, A.M. (2010) Chaperone-mediated autophagy in health and disease. *FEBS Lett.* 584:1399–404.

Kopito, R.R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* 10:524–530.

Koteiche, H.A., McHaourab, H.S. (2006) Mechanism of a Hereditary Cataract Phenotype: Mutations in RA-crystallin activate substrate binding. *J. Biol. Chem*. 281, 14273–14279.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., Bukau, B. (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* 419, 171–174.

Kubelka, J., Hofrichter, J., Eaton, W. A. (2004) The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14,** 76−88.

Kumar, S. Mohanty S.K.., Udgaonkar, J.B. (2007). Mechanism of Formation of Amyloid Protofibrils of Barstar from Soluble Oligomers: Evidence for Multiple Steps and Lateral Association Coupled to Conformational Conversion. *J. Mol. Biol.* 367, 1186–1204.

Kundu, B., Shukla, A., Chaba, R., Guptasarma, P. (2004) The excised heat-shock domain of alphaB crystallin is a folded, proteolytically susceptible trimer with significant surface hydrophobicity and a tendency to self-aggregate upon heating. *Protein Expres Purif.* 36:263–271.

Laganowsky, A., and Eisenberg, D. (2010) Non-3D domain swapped crystal structure of truncated zebrafish aA crystallin. *Protein Sci.* 19, 1978–1984.

Laganowsky, A., Benesch, J.L.P., Landau, M., Ding, L., Sawaya, M.R., Cascio, D., Huang, Q., Robinson, C.V., Horwitz, J., and Eisenberg, D. (2010) Crystal structures of truncated aA and aB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* 19, 1031–1043.

Laksanalamai, P., Maeder, D.L., Robb, F.T. (2001). Regulation And Mechanism Of Action Of The Small Heat Shock Protein From The Hyperthermophilic Archaeon *Pyrococcus furiosus*. *Journal of Bacteriology* 183(17): 5198–5202.

Langer, T., Lu, C., Echos, H., Flangan, J., Hayer, M.K., Hartl, F.U. (1992) Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* 356, 683–689.

Large, A.T., Lund, P.A. (2009) Archaeal chaperonins. *Front. Biosci.* 14: 1304–1324.

Lee, G.J., Roseman, A.M., Saibil, H.R., Vierling, E. (1997) A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* 16, 659–671.

Lee, K., Hwang, S., Lee, J. (2013) Neuronal Autophagy and Neurodevelopmental Disorders. *Exp Neurobiol.* 22(3):133-142.

Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44−45.

Levitt, M., Warshel, A. (1975) Computer simulation of protein folding. *Nature* **253**, 694–698.

Li, Z., Scheraga, H.A. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *PNAS.* **84**, 6611–6615.

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A. Ploegh, H., Matsudaria, P. (2007). *Molecular Cell Biology*. W. H. Freeman; 6th Edition.

Lopez-Maury, L., Marguerat, S., Bahler, J. (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Rev. Genet.* 9, 583–593.

Lu, Q., Han, J., Zhou, L., Coker, J.A., DasSarma, P., DasSarma, S., Xiang, H. (2008). Dissection of the regulatory mechanism of a heat-shock responsive promoter in Haloarchaea: a new paradigm for general transcription factor directed archaeal gene regulation. *Nucleic Acids Research*. 36: 3031–3042.

Lund, P. (2011) Insights into chaperonin function from studies on archaeal thermosomes. *Biochemical Society Transactions* 39: 94-98.

Maier, R., Eckert, B., Scholz, C., Lilie, H., Schmid, F.X. (2003) Interaction of Trigger factor with the ribosome. *J. Mol. Biol*. 326, 585–592.

Maiorov, V.N., Crippen, G.M. (1994). Significance of Root-Mean-Square Deviation in Comparing Three-Dimensional Structures of Globular Proteins. *J. Mol. Biol.* 235, 625-634

Matthews, B.W. (1995) Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* 46, 249–278.

Matthews, B.W. (1996) Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J.* 10, 35–41.

Mayer, M.P., Rudiger, S., Bukau, B. (2000) Molecular basis for interactions of the DnaK chaperone with substrates. *Biol. Chem.* 381, 877–885.

McDonald, E.T., Bortolus, M., Koteiche, H.A., Mchaourab, H.S. (2012) Sequence, structure, and dynamic determinants of Hsp27 (HspB1) equilibrium dissociation are encoded by the N-terminal domain. *Biochemistry* 51, 1257−1268.

Mchaourab, H.S., Godar, J.A., Stewart, P.L. (2009) Structure and Mechanism of Protein Stability Sensors: Chaperone Activity of Small Heat Shock Proteins. *Biochemistry.* 48, 3828–3837.

McHaourab, H.S., Lin, Y.L,. Spiller, B.W. (2012) Crystal structure of an activated variant of small heat shock protein Hsp16.5. *Biochemistry.* 26;51(25):5105-5112.

Mogk, A., Homuth, G., Scholz, C., L., Kim, Schmid, F.X., Schumann, W. (1997) The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *EMBO J.* **16:**4579–4590.

Morimoto, R.I. (2008) Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22,** 1427−1438.

Morley, J.F., Brignull, H.R., Weyers, J.J., Morimoto, R.I. (2002) The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caenorhabditis elegans*. *PNAS.* 99:10417–10422.

Morley, J.F., Morimoto, R.I. (2004) Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* 15:657−664.

Morrison, L.E., Whittaker, R.J., Klepper, R.E., Wawrousek, E.F., Glembotski, C.C. (2004) Roles for alphaB-crystallin and HSPB2 in protecting the myocardium from ischemia-reperfusion-induced damage in a KO mouse model. *Am J Physiol Heart Circ Physiol*. 286, H847−H855.

Muchowski, P.J., Wacker, J.L. (2005) Modulation of neurodegeneration by molecular chaperones. *Nat Rev Neurosci*. 6, 11-22.

Mymrikov, E.V., Seit-Nebi, A.S., Gusev N.B. (2011) Large potentials of small heat shock proteins. *Physiol Rev* 91: 1123−1159.

Narberhaus, F. (1999) Negative regulation of bacterial heat shock genes. *Mol. Microbiol*. **31:**1−8.

Narberhaus, F. (2002) α-Crystallin-Type Heat Shock Proteins: Socializing Minichaperones in the Context of a Multichaperone Network. *Microbiology and Molecular Biology Reviews*. 66, 64−93.

Narberhaus, F., Weiglhofer, W., Fischer, H.M., Hennecke, H. (1996) The *Bradyrhizobium japonicum rpoH*1 gene encoding a δ^{32}-like protein is part of a unique heat shock gene cluster together with *groESL*1 and three small heat shock genes. *J. Bacteriol*. **178:**5337−5346.

Natera, S.H., Guerreiro, A., Djordjevic, N.A. (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. *Mol. Plant-Microbe Interact*. **13:**995−1009.

Neckers, L. (2007) Heat shock protein 90: the cancer chaperone. *J. Biosci*. **32,** 517−530.

Nelson, D.L., & Cox, M.M., (2008). *Lehninger Principles of Biochemistry*. W. H. Freeman; 5th Edition.

Nocker, A., T., Hausherr, S., Balsiger, N.P., Krstulovic, Hennecke, H., Narberhaus, F. (2001) A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res.* **29:**4800–4807.

Ohlson,T., Wallner, B.,Elofsson, A. (2004) Profile-profile methods provide improved fold recognition: a study of different profile-profile alignment methods. *Proteins* **57**, 188–197.

Olzscha, H., Schermann, S.M.,Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., Vabulas, R.M. (2011) Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144:67–78.

Parry, D.A. (2005) Structural and functional implications of sequence repeats in fibrous proteins, *Adv Protein Chem*, **70**, 11–35.

Pasta, S.Y., Raman, B., Ramakrishna, T., Rao, C.M. (2004) The IXI/V motif in the C-terminal extension of alpha-crystallins: alternative interactions and oligomeric assemblies. *Mol. Vis*. 10, 655–662.

Pauling, L., Corey, R.B., (1951) Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *PNAS.* **37**(11): 729–740.

Pawlowski, M., Gajda, M.J., Matlak, R., Bujnicki, J.M. (2008). MetaMQAP: a meta-server for the quality assessment of protein models, *BMC Bioinformatics*, **9**, 403.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G., North, A.C.T., (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5A resolution, obtained by x-ray analysis. *Nature.*185, 416–422.

Peschek J, Braun, N., Franzmann, T.M., Georgalis, Y., Haslbeck, M., Weinkauf, S., Buchner, J. (2009) The eye lens chaperone alpha-crystallin forms defined globular assemblies. *PNAS.* 106:13272–13277.

Petesch, S.J., Lis, J.T. (2008) Rapid, transcription independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134**, 74–84.

Qian, J., Ren, X., Wang, X., Zhang, P., Jones, W.K., Molkentin, J.D., Fan, G.C., Kranias, E.G. (2009) Blockade of Hsp20 phosphorylation exacerbates cardiac ischemia/reperfusion injury by suppressed autophagy and increased cell death. *Circ Res*. 105, 1223–1231.

Quinlan, R.A., Zhang, Y., Lansbury, A., Williamson, I., Pohl, E., Sun, F. (2013) Changes in the quaternary structure and function of MjHSP16.5 attributable to deletion of the IXI motif and introduction of the substitution, R107G, in the α-crystallin domain. *Philos Trans R Soc Lond B Biol Sci*. 25;368(1617):20120327.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., (1963) Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**: 95–9.

Ray, P.S., Martin, J.L., Swanson, E.A., Otani, H., Dillmann, W.H., Das, D.K. (2001) Transgene overexpression of alphaB crystallin confers simultaneous protection against cardiomyocyte apoptosis and necrosis during myocardial ischemia and reperfusion. *FASEB J*. 15,393– 402.

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., Fraser, C.M. (2000) Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res*. **28:**1397–1406.

Reddy, V.S., Raghu, G., Reddy, S.S., Pasupulatim A.K., Suryanarayanam P., Reddy, G.B. (2013) Response of small heat shock proteins in diabetic rat retina. *Invest Ophthalmol Vis Sci*. 54(12), 7674-7682.

Rohl, C.A., Strauss, C.E., Misura, K.M., Baker, D. (2004) Protein structure prediction using Rosetta, *Methods Enzymol*, **383**, 66–93.

189

Rudiger, S., Buchberger, A., Bukau, B. (1997) Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol*. 4, 342–349.

Santos, H., da Costa, M.S. (2002) Compatible solutes of organisms that live in hot saline environments. *Environ. Microbiol*. **4:**501–509.

Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., Neher, S.B., Oakes, E.S.C., Siddiqui, S.M., Wah, D.A., Baker, T.A. (2004) Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* 119, 9–18.

Schiene-Fischer, C., Habazettl, J., Schmid, F.X., Fischer, G. (2002) The Hsp70 chaperone DnaK is a secondary amide peptide bond cis-trans isomerase. *Nat. Struct. Biol.* 9, 419–424.

Schubert, U., Anton, L.C., Gibbs, J., Orbyry, C.C., Yewdell, J.W., Bennink, J.R. (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* 404:770-774.

Schuler, J., Frank, J., Saenger, W., Georgalis, Y. (1999). Thermally Induced Aggregation of Human Transferrin Receptor Studied by Light-Scattering Techniques. *Biophysical Journal* 77: 1117–1125.

Seale, J.W., Martinez, J.L., Horowitz, P.M. (1995). Photoincorporation of 4,4'-Bis(1- anilino-8-naphthalenesulfonicA cid) into the Apical Domain of GroEL: Specific Information from a Nonspecific Probe. *Biochemistry 34,* 7443-7449.

Selkoe, D.J. (2003) Folding proteins in fatal ways. *Nature* 426, 900–904.

Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., Nudler, E. (2006) RNA-mediated response to heat shock in mammalian cells. *Nature* 440, 556–560.

Shashidharamurthy, R., Koteiche, H. A., Dong, J., McHaourab, H. S.(2005) Mechanism of chaperone function in small heat shock proteins: Dissociation of the HSP27 oligomer is required for recognition and binding of destabilized T4 lysozyme. *J. Biol. Chem*. 280,5281–5289.

Shi, J., Koteiche, H.A., McHaourab, H.S., Stewart, P. L. (2006) Cryoelectron microscopy and EPR analysis of engineered symmetric and polydisperse Hsp16.5 assemblies reveals determinants of polydispersity and substrate binding. *J. Biol. Chem*. 281, 40420–40428.

Shi, Y., Mosser, D.D., Morimoto, R.I. (1998) Molecular chaperones as HSF1-specific transcriptional repressors. *Genes Dev.* **12**, 654–666.

Shiau, A. K., Harris, S. F., Southworth, D. R. & Agard, D. A. (2006) Structural analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127,** 329−340.

Shigenobu, S., Watanabe, H., Hattori, Sakaki, M., Ishikawa, H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp*. *Nature* 407:81–86.

Shockley, K.R., Donald, E.W., Chabra, S.R., Conners, S.B., Montero, C.I., Kelly, R.M. (2003) Heat Shock Response by the Hyperthermophilic Archaeon *Pyrococcus furiosus. Applied And Environmental Microbiology*. 69, 2365–2371.

Simon, S., Michiel, M., Skouri-Panet, F., Lechaire, J.P., Vicart, P., Tardieu, A. (2007) Residue R120 is essential for the quaternary structure and functional integrity of human alphaB-crystallin. *Biochemistry.* 46(33), 9605-9614.

Soding , J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960.

Soppa, J. (1999) Transcription initiation in Archaea: facts, factors and future aspects. *Mol. Microbiol*. **31:**1295–1305.

Specht, S., Miller, S.B.M., Mogk, A., Bukau, B. (2011) Hsp42 is required for sequestration of protein aggregates into deposition sites in Saccharomyces cerevisiae. *J. Cell Biol.* 195:617–29.

Spriggs, K.A., Bushell, M., Willis, A.E., (2010) Translational Regulation of Gene Expression during Conditions of Cell Stress. *Molecular Cell*. 40, 228-237.

Srere, P.A., Brazil, H., Gonen, L. (1963). The citrate−condensing enzyme of pigeon, breast muscle and moth flight muscle. *Acta Chem Scand*. 17:129−134.

Stengel, F., Baldwin, A.J., Painter, A.J., Jaya, N., Basha, E., Kay, L.E., Vierling, E., Robinson, C.V., Benesch, J.L. (2010) Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *PNAS*. 107(5), 2007-2012.

Stotz, K., Bostanci, A., Griffiths, P.E. (2006) Tracking the Shift to 'Postgenomics. *Community Genetics*. 9, 190–196.

Stromer, T., Ehrnsperger, M., Gaestel, M., Buchner, J. (2003) Analysis of the interaction of small heat shock proteins with unfolding proteins. *J. Biol. Chem*. 278, 18015–18021.

Sui, X., Li, D., Qiu, H., Gaussin, V., Depre, C. (2009) Activation of the bone morphogenetic protein receptor by H11kinase/Hsp22 promotes cardiac cell growth and survival. *Circ Res*. 104, 887– 895.

Summa, C.M., Levitt, M. (2007) Near-native structure refinement using in vacuo energy minimization, *PNAS*, **104**, 3177–3182.

Taipale, M., Jarosz, D. F. & Lindquist, S. (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Rev. Mol. Cell Biol.***11,** 515−528.

Takeda, K., Hayashi, T., Abe, T., Hirano, Y., Hanazono, Y., Yohda, M., Miki, K. (2011) Dimer structure and conformational variability in the N-terminal region of an archaeal small heat shock protein, StHsp14.0. *J Struct Biol*. 174(1):92-9.

Thompson, D.K., Daniels, C.J. (1998) Heat shock inducibility of an archaeal TATA-like promoter is controlled by adjacent sequence elements. *Mol Microbiol* 27: 541-551.

Thompson, D.K., Palmer, J., Daniels, C.J. (1999) Expression and heat-responsive regulation of a TFIIB homologue from the archaeon *Haloferax volcanii. Molecular Microbiology*. 33(5), 1081-1092.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G (1997) The Clustal X Windows Interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic.Acids Res.*, **25**, 4876–4882.

Thompson, J.D., Plewniak, F., Thierry, J., Poch, O. (2000) Dbclustal: Rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res*, **28**, 2919–2926 (2000).

Tissieres, A., Mitchell, H.K., Tracy, U.M. (1974) Protein Synthesis in salivary glands of Drosophila melanongaster: Relation to chromosome puffs. *J. Mol. Biol* 85: 389-398.

Tyedmers, J., Mogk, A., Bukau, B. (2010) Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* 11:777–88.

Usui, K., Hatipoglu, O.F., Ishii, N., Yohda, M. (2004) Role of the N-terminal region of the crenarchaeal sHsp, StHsp14.0, in thermal-induced disassembly of the complex and molecular chaperone activity. *Biochem Biophys Res Commun*. 315(1), 113-118.

van Montfort, R.L., Basha, E., Friedrich, K.L., Slingsby, C., and Vierling, E. (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol*. 8, 1025–1030.

Vierke, G., Engelmann, A., Hebbeln, C., Thomm, M. (2003) A Novel Archaeal Transcriptional Regulator of Heat Shock Response. *The Journal Of Biological Chemistry.* 278, 18–26.

Vucetic, S., Xie, H., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., Uversky, V.N. (2007) Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding

sequence diversities correlated with long disordered regions, *J Proteome Res*, **6**, 1899–1916.

Wales, D.J., Scheraga, H.A. (1999) Global optimization of clusters, crystals, and biomolecules. *Science* **285**, 1368–1372.

Wallner, B., Elofsson, A. (2005) All are not equal: a benchmark of different homology modeling programs. *Protein Sci*. **14**, 1315–1327.

Walter, P., Ron, D. (2011) The unfolded protein response: from stress pathway to homeostatic regulation. *Science* 334:1081–1086.

Wandinger, S. K., Richter, K., Buchner, J. (2008) The Hsp90 chaperone machinery. *J. Biol. Chem.* **283,** 18473−18477.

Wang, J., Martin, E., Gonzales, V., Borchelt, D.R., Lee, M.K. (2007) Differential regulation of small heat shock proteins in transgenic mouse models of neurodegenerative diseases. *Neurobiol Aging.* 4, 586-597.

Weadick, C.J., Chang, B.S.W. (2009) Molecular Evolution of the βγ Lens Crystallin Superfamily: Evidence for a Retained Ancestral Function in βN Crystallins?. *Mol. Biol. Evol.* 26(5):1127–1142.

Wegrzyn, R.D., Deuerling, E. (2005) Molecular guardians for newborn proteins: ribosomeassociated chaperones and their role in protein folding. *Cell. Mol. Life Sci.* 62, 2727–2738.

Whitakerand, J.R., Granum, P.E. (1980). An Absolute Method for Protein Determination Based on Difference in Absorbance at 235 and 280 nm. *Analytical Biochemistry* 109: 156-159

White, H.E., Orlova, E.V., Chen, S., Wang, L., Ignatiou, A., Gowen, B., Stromer,T., Franzmann, T.M., Haslbeck, M., Buchner, J., Saibil, H.R. (2006) Multiple distinct assemblies reveal conformational flexibility in the small heat shock protein Hsp26. *Structure* 14, 1197–1204.

Wickner, S., Gottesman, S., Scowra, D., Hoskins, J., McKenney, K., Maurizi, M.R. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. *PNAS*. 91, 12218– 12222

Winkler, J., Tyedmers, J., Bukau, B., Mogk, A. (2012) Hsp70 targets Hsp100 chaperones to substrates for protein disaggregation and prion fragmentation. *J. Cell Biol*. 6;198, 387-404.

Xiang, Z., Steinbach, P.J., Jacobson, M.P., Friesner, R.A., Honig, B. (2007) Prediction of side-chain conformations on protein surfaces. *Proteins* **66**, 814–823.

Xie, H., Vucetic , S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., Uversky, V.N. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions, *J Proteome Res*, **6**, 1882–1898.

Xie, H., Vucetic, S., Iakoucheva, L.M. Oldfield, C.J., Dunker, A.K., Obradovic, Z., Uversky, V.N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins, *J Proteome Res*, **6**, 1917–1932.

Young, J.C., Barral, J.M., Hartl, F.U. (2003) More than folding: localized functions of cytosolic chaperones. *Trends Biochem. Sci*. 28, 541–547.

Yura, T., Kanemori, M., Morita M.T. (2000) The heat shock response: regulation and function, p. 3–18. *In* G. Storz and R. Hengge-Aronis (ed.), *Bacterial stress responses*. ASM Press, Washington, D.C.

Zhang, Y. (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* **69**, 108–117.

Zhu, X., Zhao, X, Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., Handrickson, W.A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 272, 1606–1614.

Zuber, U., Schumann, W. (1994) CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *Bacillus subtilis*. *J. Bacteriol.* **176:**1359–1363

APPENDIX A

BUFFERS AND SOLUTIONS

T.volcanium **Medium** (1000 ml) PH: 2.3 (Adjust with 98% H2SO4)

 (KH_2PO_4)…….…..…………3g

 $(MgSO_4)$…………............... 1g

$(CaCl_2.2H_2O)$……...………. 0.25g

 $((NH_4)_2SO_4)$...…………….. 0.2g

LB Agar for XL1 Cell Growth (1000ml) PH 7.0 (Adjust with 5N NaOH)

NaCl………………………..10g

Tryptone……………………10g

Yeast Extract (Bacto)……....5g

Agar………………………...20g

Double-Distilled H_2O………Complete to 1L

LB Agar Medium For Cell Plating (1000 ml) Adjust pH to 7.4 with 10 M NaOH

Tryptone medium.................18 g
Yeast Extract........................5 g
NaCl.....................................2 g
Agar.....................................15g
Distilled H_2O........................Adjust to 1l.

Stock YE Solution (5%)

Yeast Extract.......................... 2.5 g in 50ml H_2O

Stock Glucose Solution (25%)

Glucose..................................12.5g in 50 ml H_2O

NZY$^+$ Broth (1000ml) For Cell Growth after Transformation.

NZ Amine (Casein Hydrolysate)…10g

Yeast Extract (Bacto)……………..5g

NaCl…………………………………5g

$MgCl_2$ (1M)…………………………12.5 ml

Double-Distilled H_2O…………….Complete to 1L and Autoclave. Then add the followings after filter sterilization:

$MgSO_4$ (1M)…………………………2.5 ml

Glucose (2M)………………………...10 ml

SDS Gel Running Buffer (500ml) pH: 8.7.

Tris …………………….…………….1.51 g

Glycin……………………..………… 7.2 g

SDS……………………………………0.5 g

50x TAE Buffer (1000 ml)

Tris Base…………………………….242 g

Glacial Acetic Acid.……………......57.1 ml

EDTA(0.5M- pH: 8.0)…..……..…….100 ml

SDS-PAGE (12.5%) Gel

Chemicals	Seperating Gel	Stacking Gel
Acrylamide/ Bisacrylamide (30%)	2.5ml	330μl
dH2O	1.1ml	870μl
Tris (pH 8.9)	1.2ml (1.88 M)	400μl (0.625 M)
SDS (0.5%)	1.2ml	400μl
APS (10%)	30μl	10μl
TEMED	5μl	3μl

Agarose Gel Running buffer (1X)

50x TAE Buffe…………….……….6 ml

Distilled Water………..……….……300 ml

Agarose Gel (8%)

Agarose……….……….……….… 0.32 g

Running buffer..…………….......... 40 ml

Etidium Bromide……………........5 μl

Buffers for Protein Purification

Lysis buffer: (50ml) 10mM Imidazole (pH8.0 adjusted with 0.5M NaOH)

NaH_2PO_4………………………50 Mm

NaCl………………………….....300 mM

Imidazole……………………...10 mM

Wash buffer: (50ml) 30mM Imidazole (PH 8.0)

NaH_2PO_4………………………50 mM

NaCl…………………………...300 mM

Imidazole……………………...30 mM

Elution buffer: (50ml) 250mM Imidazole

NaH_2PO_4………………………50 mM

NaCl…………………………...300 mM

Imidazole……………………....250 mM

APPENDIX B

MARKERS AND LADDERS
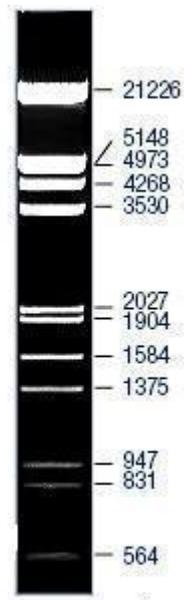
Figure B1: Lambda DNA/*EcoRI+HindIII* marker (Fermentas, Lithuania)
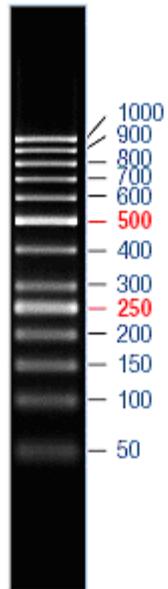
Figure B2: GeneRuler™ 50 bp DNA Ladder (50-1000 bp) (Fermentas, Lithuania)
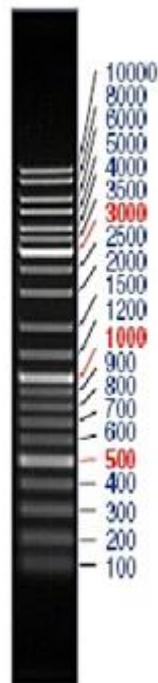
Figure B3: GeneRuler™ DNA Ladder Mix (10000-100) (Fermentas, Lithuania)
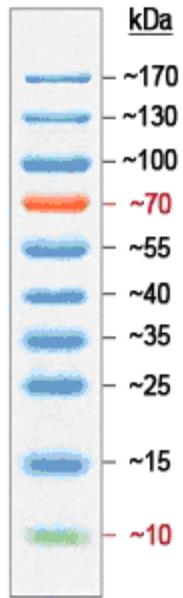
Figure B4: PageRuler™ Prestained Protein Ladder (Fermentas, Lithuania)

APPENDIX C

SEQUENCING RESULTS

Wild-Type sequence is shown at the beginning of every position. The wild-type and mutated sequences are underlined and bold.

R69 Position.

Wild-Type tpv-Hsp14.3 Gene Sequence

ATGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAG
GTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGTATTGGAAGCAGA
AATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGC
GGAG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCACGCGTTGACAAAGTGTATAAAGTAGTT
AAGCTGCCCGTAGAGATTGAGCAGCAGGACATTCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAA

The Sequence AGA is to be changed into ATG for R69M change.

Sequence 1

AGTGCTCATCATTGGAAAACGTTCTTCGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AG**ATG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
```

```
WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT      201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250
            |||..|||||||||||||||||||||||||||||||||||||||||||||
R69M    201 AG**ATG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT      351 TAAAGAACGTAGAAATAGAATAA 373
            |||||||||||||||||||||||
R69M    351 TAAAGAACGTAGAAATAGAATAA 373
```

Sequence 2
```
AGTGCTCATCATTGGAAAACGTTCTTCGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AG**ATG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA
```

Sequence Alignment
```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
```

206

```
WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250
            |||..|||||||||||||||||||||||||||||||||||||||||||||
R69M    201 AGATGAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT      351 TAAAGAACGTAGAAATAGAATAA 373
            |||||||||||||||||||||||
R69M    351 TAAAGAACGTAGAAATAGAATAA 373
```

The Sequence AGA is to be changed into GAG for R69E change.

Sequence 1

```
AGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACC
TGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTT
CGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATT
CAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGAT
CTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATT
GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGT
CTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAA
ATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCG
GAGGAGAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTT
AAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTT
AGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTTAATAATAATATAT
TAAAAAGA
```

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAA 249
            ||...|||||||||||||||||||||||||||||||||||||||||||||
R69M    201 AGGAGAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAA 249

WT      250 AGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATAT 299
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    250 AGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATAT 299
```

207

```
WT       300 CTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAAC 349
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     300 CTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAAC 349

WT       350 ATAAAGAACGTAGAAATAGAATAA 373
             ||||||||||||||||||||||||
R69M     350 ATAAAGAACGTAGAAATAGAATAA 373
```

Sequence 2

```
AGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACC
TGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTT
CGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATT
CAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGAT
CTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATT
GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGT
CTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAA
ATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCG
GAG**GAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTT
AAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTT
AGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTTAATAATAATATAT
TAAAAAGA
```

Sequence Alignment

```
WT         1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50

WT        51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT       101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT       151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT       201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAA 249
             ||**...**|||||||||||||||||||||||||||||||||||||||||||
R69M     201 AG**GAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAA 249

WT       250 AGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATAT 299
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     250 AGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATAT 299

WT       300 CTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAAC 349
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M     300 CTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAAC 349

WT       350 ATAAAGAACGTAGAAATAGAATAA 373
             ||||||||||||||||||||||||
R69M     350 ATAAAGAACGTAGAAATAGAATAA 373
```

208

The Sequence AGA is to be changed into AAG for R69K change.

Sequence 1

AGTGCTCATCATTGGAAAACGTTTTCGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AG**AAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment

```
WT         1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT        51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT       101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT       151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT       201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
             |||..|||||||||||||||||||||||||||||||||||||||||||||
R69K     201 AG**AAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250

WT       251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT       301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT       351 TAAAGAACGTAGAAATAGAATAA   373
             |||||||||||||||||||||||
R69K     351 TAAAGAACGTAGAAATAGAATAA   373
```

Sequence 2

AGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT
TTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGA
TTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGG

209

ATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAG
ATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGA
GGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCA
GAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGT
GCGGAG**AAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATAAAGTA
GTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACA
GTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATAT
ATATAAAAGA

Sequence Alignment

```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA    50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA    50


WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100


WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA   150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA   150


WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG   200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG   200


WT     201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA   250
           |||..|||||||||||||||||||||||||||||||||||||||||||||
R69K   201 AG**AAG**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA   250


WT     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC   300


WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA   350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA   350


WT     351 TAAAGAACGTAGAAATAGAATAA   373
           |||||||||||||||||||||||
R69K   351 TAAAGAACGTAGAAATAGAATAA   373
```

The intended site was changed into Asparigine codon instead of Lysine as the codon AAT
stands for Asn.

Sequence 3 (Asparigine Codon)
GACTGGGAGTACAACCAAGTCATTTGAGAATAGTATGCGGCGACCGAGTTGTTTGCCCGGCGTCAA
TACGGGATAATACCGGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACA
TTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCACCTCGAGAAATCATAAAAAATTT
ATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGA
ATTCATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCA
TGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTGTATACACCCATAAAGTTCTTTACGAATGAG
ATGATAAAAAACGTATCGAATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTAT
ATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAAGAAAAACATAAAGG
TCTCGGTAAATAAGAATGTACTCACTATAAGTGCGGAG**AAT**AAGAGAGAATACTCTACCGTATATA
TCGATCAGCGCGTTGACAAAGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACA

210

```
TATCTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAG
AAATAGAATAAAATCATTTTTAATAATAATATATATAAAAGA
```

Sequence Alignment

```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT     201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
           |||..|||||||||||||||||||||||||||||||||||||||||||||
R69K   201 AG**AAT**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250

WT     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R69K   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT     351 TAAAGAACGTAGAAATAGAATAA   373
           |||||||||||||||||||||||
R69K   351 TAAAGAACGTAGAAATAGAATAA   373
```

The R81 Position

The Sequence AGA is to be changed into AAG for R81K change.

Sequence 1
```
AGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT
TTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGA
TTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGG
ATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAG
ATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGA
GGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCA
GAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGT
GCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**AAG**GTTGACAAAGTGTATAAAGT
AGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTAC
AGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATA
TATATAAAAGA
```

211

Sequence Alignment

```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50

WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT     201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
           |||||||||||||||||||||||||||||||||||||||...|||||||||
R81K   201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGAAGGTTGACAAA  250

WT     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K   251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT     351 TAAAGAACGTAGAAATAGAATAA    373
           |||||||||||||||||||||||
R81K   351 TAAAGAACGTAGAAATAGAATAA    373
```

Sequence 2

```
GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**AAG**GTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGTA
```

```
WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250
            |||||||||||||||||||||||||||||||||||||||...||||||||
R81K    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGAAGGTTGACAAA 250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT      351 TAAAGAACGTAGAAATAGAATAA   373
            |||||||||||||||||||||||
R81K    351 TAAAGAACGTAGAAATAGAATAA   373
```

Sequence 3

```
CTTCAGCATCTTTTACTTTCACCAGCGTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAG
GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATT
AACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCACCTCGAGAAATCATAAAAAATTTAT
TTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAAT
TCATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCATG
CGAGCTCGGTACCCCGGGTCGACGAATTCAGATTGTATACACCCATAAAGTTCTTTACGAATGAGAT
GATAAAAAACGTATCGAATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT
CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTC
TCGGTAAATAAGAATGTACTCACTATAAGTGCGGAGAGAAAGAGAGAATACTCTACCGTATATATC
GATCAGAAGGTTGACAAAGTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATA
TCTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAA
ATAGAATAAAATCATTTTTTAATAATAATATATAAAAAGTATG
```

Sequence Alignment 3

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250
            |||||||||||||||||||||||||||||||||||||||...||||||||
R81K    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGAAGGTTGACAAA 250
```

213

```
WT       251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT       301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81K     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT       351 TAAAGAACGTAGAAATAGAATAA   373
             |||||||||||||||||||||||
R81K     351 TAAAGAACGTAGAAATAGAATAA   373
```

The Sequence AGA is to be changed into AAG for R81E change.

In this sequence, there is a short insertion shown in itallics and left as gaps in sequence alignment. For this reason, this colony was discarded.

Sequence 1 (Insertion)

GGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACA
TTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCACCTCGAGAAATCATAAAAAATTT
ATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGA
ATTCATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCA
TGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTGTATACACCCATAAAGTTCTTTACGAATGAG
ATGATAAAAAACGTATCGAATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTAT
ATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAAGAAAAACATAAAGG
TCTCGGTAAATAAGAATGTACTCACTATAAGTGCGGAGAGAAAGAGAGAATACTCTACCGTATATA
TCGATCAG*GAGGTTGACAAAGTGTATATATCGATCAG***GAG**GTTGACAAAGTGTATAAAGTAGTTAAGC
TGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAA
TGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATATA
AAAAGA

Sequence Alignment
```
WT         1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT        51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT       101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT       151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT       201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG------------  238
             |||||||||||||||||||||||||||||||||||||
R81E     201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGGAGGTTGACAAA  250

WT       239 ----------CG-------CGTTGACAAAGTGTATAAAGTAGTTAAGCTG  271
                       ||        .|||||||||||||||||||||||||||||||
R81E     251 GTGTATATATCGATCAGGAGGTTGACAAAGTGTATAAAGTAGTTAAGCTG  300
```

214

```
WT      272 CCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCAT  321
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    301 CCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCAT  350

WT      322 ACTTACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAAT  371
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    351 ACTTACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAAT  400

WT      372 AA  373
            ||
R81E    401 AA  402
```

Sequence 2

```
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCC
ACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCC
CTTTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATA
GATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGA
GGATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTC
AGATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAA
GAGGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAG
CAGAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAA
GTGCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**GAG**GTTGACAAAGTGTATAAA
GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTT
ACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAA
TATATATAAAG
```

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**CGC**GTTGACAAA  250
            ||||||||||||||||||||||||||||||||||||||...||||||||||
R81E    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**GAG**GTTGACAAA  250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT      351 TAAAGAACGTAGAAATAGAATAA  373
            |||||||||||||||||||||||
R81E    351 TAAAGAACGTAGAAATAGAATAA  373
```

215

Sequence 3
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT
TTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGA
TTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGG
ATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAG
ATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGA
GGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCA
GAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGT
GCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**<u>GAG</u>**GTTGACAAAGTGTATAAAGT
AGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTAC
AGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATA
TATAAAAAGA

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**CGC**GTTGACAAA  250
            ||||||||||||||||||||||||||||||||||||||...||||||||||
R81E    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**GAG**GTTGACAAA  250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81E    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT      351 TAAAGAACGTAGAAATAGAATAA   373
            |||||||||||||||||||||||
R81E    351 TAAAGAACGTAGAAATAGAATAA   373
```

The Sequence AGA is to be changed into ATG for R81M change.

Sequence 1
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT
TTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGA
TTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGG

216

ATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAG
ATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGA
GGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCA
GAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGT
GCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**ATG**GTTGACAAAGTGTATAAAGTA
GTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACA
GTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATAT
ATAAAAAGA

Sequence Alignment

```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT     201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**CGC**GTTGACAAA  250
           |||||||||||||||||||||||||||||||||||||...||||||||||
R81M   201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**ATG**GTTGACAAA  250

WT     251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M   251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT     351 TAAAGAACGTAGAAATAGAATAA   373
           |||||||||||||||||||||||
R81M   351 TAAAGAACGTAGAAATAGAATAA   373
```


Sequence 2
GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGT
GAGCAAAAACAGGAAGGCAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**ATG**GTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAAGA

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100


WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**CGC**GTTGACAAA  250
            |||||||||||||||||||||||||||||||||||||...||||||||||
R81M    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAG**ATG**GTTGACAAA  250

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M    251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT      351 TAAAGAACGTAGAAATAGAATAA    373
            |||||||||||||||||||||||
R81M    351 TAAAGAACGTAGAAATAGAATAA    373
```

Sequence 3

```
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGAT
CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCC
ACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCC
CTTTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATA
GATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGA
GGATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTC
AGATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAA
GAGGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAG
CAGAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAA
GTGCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGATGGTTGACAAAGTGTATAAAG
TAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTA
CAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAAT
ATATTAAAAGA
```

Sequence Alignment

```
WT        101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
              ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT        151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
              ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT        201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
              ||||||||||||||||||||||||||||||||||||||   ||||||||||
R81M      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGATGGTTGACAAA  250

WT        251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT        301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
              ||||||||||||||||||||||||||||||||||||||||||||||||||
R81M      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT        351 TAAAGAACGTAGAAATAGAATAA  373
              |||||||||||||||||||||||
R81M      351 TAAAGAACGTAGAAATAGAATAA  373
```

The QR(80-81) Position

The Sequence CAGCGC is to be changed into GAGCTC for QR(80-81)EL change.

Sequence 1
AGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATTA
AAAAGA

Sequence Alignment
```
WT            1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50
                ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA  50

WT           51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
                ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL  51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT          101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
                ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
```

```
WT          151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT          201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**CAGCGC**GTTGACAAA 250
                ||||||||||||||||||||||||||||||||||||.|||.|||||||||
QR(80-91)EL 201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAA 250

WT          251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300

WT          301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT          351 TAAAGAACGTAGAAATAGAATAA 373
                |||||||||||||||||||||||
QR(80-91)EL 351 TAAAGAACGTAGAAATAGAATAA 373
```

Sequence 2
```
AAGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG
ACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
TCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCA
ATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATCT
CACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTGT
ATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTCT
CATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAAT
GGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGGA
GAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAAGTGTATAAAGTAGTTAA
GCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTAG
AATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATTAA
AAGA
```

Sequence Alignment

```
WT            1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50

WT           51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL  51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT          101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT          151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT          201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**CAGCGC**GTTGACAAA 250
                ||||||||||||||||||||||||||||||||||||.|||.|||||||||
QR(80-91)EL 201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAA 250

WT          251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
                |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL 251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
```

220

```
WT            301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
                  |||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT            351 TAAAGAACGTAGAAATAGAATAA 373
                  |||||||||||||||||||||||
QR(80-91)EL   351 TAAAGAACGTAGAAATAGAATAA 373
```
Sequence 3
```
AGTGTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA
```

Sequence Alignment

```
WT              1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA 50

WT             51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT 100

WT            101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT            151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT            201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**CAGCGC**GTTGACAAA 250
                  |||||||||||||||||||||||||||||||||||.|||.||||||||||
QR(80-91)EL   201 AGAGAAAGAGAGAATACTCTACCGTATATATCGAT**GAGCTC**GTTGACAAA 250

WT            251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300

WT            301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
QR(80-91)EL   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350

WT            351 TAAAGAACGTAGAAATAGAATAA 373
                  |||||||||||||||||||||||
QR(80-91)EL   351 TAAAGAACGTAGAAATAGAATAA 373
```

221

The Position K87

The Sequence AAA is to be changed into ATA for K87I change.

Sequence 1

AGTGCTCATCATTGGAAAACGTTTTCGGGGGAAAACTTCAAGGATCTTACCGCTGTTGAGATCCAGT
TCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG
ACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
TCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCA
ATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATCT
CACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTGT
ATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTCT
CATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAAT
GGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGGA
GAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTAT**ATA**GTAGTTAA
GCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATAC**TTACAGTTAG**
AATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATTAA
AAAGA

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250

WT      251 GTGTAT**AAA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
            |||||||.||||||||||||||||||||||||||||||||||||||||||
K87I    251 GTGTAT**ATA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT      351 TAAAGAACGTAGAAATAGAATAA   373
            |||||||||||||||||||||||
K87I    351 TAAAGAACGTAGAAATAGAATAA   373
```

In this sequence, the R81I codon was actually K87R. This may be a labeling error or a mutation as well.

Sequence 2

AGTGCTCATCATTGGAAAACGTTCTTCGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG

TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTAT**<u>AGA</u>**GTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment

```
WT        1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I      1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT       51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I     51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT      201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250

WT      251 GTGTAT**AAA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
            |||||||.||||||||||||||||||||||||||||||||||||||||||
K87I    251 GTGTAT**AGA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
K87I    301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT      351 TAAAGAACGTAGAAATAGAATAA   373
            |||||||||||||||||||||||
K87I    351 TAAAGAACGTAGAAATAGAATAA   373
```

The Sequence AAA is to be changed into AGA for K87R change.

In this sequence, the K87R codon was actually K87I. This may be a labeling error, as backed by the previous sequence, or a mutation as well.

Sequence 1(It is Ile Codon)
AGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTACTTTCACCAGCGTTTCTGGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT

GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATCGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATATAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment

```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT  100

WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA  150

WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG  200

WT     201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R   201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA  250

WT     251 GTGTAT**AAA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300
           |||||||.||||||||||||||||||||||||||||||||||||||||||
K87R   251 GTGTAT**ATA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC  300

WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87R   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA  350

WT     351 TAAAGAACGTAGAAATAGAATAA   373
           |||||||||||||||||||||||
K87R   351 TAAAGAACGTAGAAATAGAATAA   373
```

This sequence was shorter because it was obtained in the first sequencing set, so there is no sequence alignment.

Sequence 2
ATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTAT**AGA**GTAGTTAAGCTGCCCGTAGA
GATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGNTAGAATGAAAACCAA
GAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTTAATAATAATATATATGAAAAGTATTGC
ATCTATTGCTATAAGTATGCTTGATGTGATTGGGTGAATCACGAATTCTGGATCCGATACGTAACGC
GTCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAAC
TCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTA
GCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAATCACTGGATATACCACCG
TTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTA
TAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTT
TTATCCGGNCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGA
AAGACNGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCATGAGCAAACTGAA
ACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTCCGGCAGTTTCTACAC

The Sequence AAA is to be changed into AGA for K87E change.

Sequence 1
AGTGCTCATCATTGGAAAACGTTTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTC
AATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATC
TCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAGATTG
TATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGAGGTC
TCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCAGAAA
TGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG
AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTATGAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment
```
WT       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E     1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50

WT      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E    51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100

WT     101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA   150
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E   101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA   150

WT     151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG   200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E   151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG   200

WT     201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA   250
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E   201 AGAGAAAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA   250

WT     251 GTGTAT**AAA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC   300
           ||||||| .|||||||||||||||||||||||||||||||||||||||||
K87E   251 GTGTAT**ATA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC   300

WT     301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA   350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
K87E   301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA   350

WT     351 TAAAGAACGTAGAAATAGAATAA   373
           |||||||||||||||||||||||
K87E   351 TAAAGAACGTAGAAATAGAATAA   373
```

The following two sequences were shorter because they were obtained in the first sequencing set, so there is no sequence alignment for them.

Sequence 2
AGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAAGTGTAT**GAA**GTAGTTAAGCTGCCC
GTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAA

ACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTTAATAATAATATATATGAAAAGT
ATTGCATCTATTGCTATAAGTATGCTTGATGTGATTGGGTGAATCACGAATTCTGGATCCGATACGT
AACGCGTCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTC
AGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCA
AGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATAC
CACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGT
ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCAC
AAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGC
AATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAA
ACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATT
CGCAAGATG

Sequence 3
TAAATAAGAATGTACTCACTATAAGTGCGGAGAGAAAGAGAGAATACTCTACCGTATATATCGATC
AGCGCGTTGACNAAGTGTAT**GAA**GTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTG
CTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACATAAAGAACGTAGAAATA
GAATAAAATCATTTTTTAATAATAATATATATGAAAAGTATTGCATCTATTGCTATAAGTATGCTTG
ATGTGATTGGGTGAATCACGAATTCTGGATCCGATACGTAACGCGTCTGCAGCCAAGCTTAATTAGC
TGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACG
CTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCT
AAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGT
AAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATA
TTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCT
TGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATG
GGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGT
GAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA
ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGT
TTCACCAGTTTTGATTAAACGTGGCC

This was one of the promising colonies obtained by Transformer Site Directed Mutagenesis Kit (Clontech). The sequence was changed but there is an internal deletion of four nucleotides.

Sequence 1. It has an internal 4-nt deletion
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
CAGTTCGATGTAACCCACTCGTGCACCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT
TTCGTCTTCACCTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGA
TTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGG
ATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACGAATTCAG
ATTGTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGAATACTGTGAAAGA
GGTCTCATCCTTTATATATCCACCAGTCACGTTATATCAAGATAGCTCTGATCTGGTATTGGAAGCA
GAAATGGCCGGGTTTGACAAGAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGT
GCGGAGATGAAGAGAGAATACTCTACCGTATATATCGATCAGTTGACAAAGTGTATAAAGTAGTTA
AGCTGCCCGTAGAGATTGAGCAGCAGGACATATCTGCTAAGTATAGTGAAGGCATACTTACAGTTA
GAATGAAAACCAAGAACATAAAGAACGTAGAAATAGAATAAAATCATTTTTAATAATAATATATAT
AAAAGA

Sequence Alignment
```
WT         1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M       1 GTATACACCCATAAAGTTCTTTACGAATGAGATGATAAAAAACGTATCGA   50


WT        51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M      51 ATACTGTGAAAGAGGTCTCATCCTTTATATATCCACCAGTCACGTTATAT   100
```
226

```
WT      101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    101 CAAGATAGCTCTGATCTGGTATTGGAAGCAGAAATGGCCGGGTTTGACAA 150

WT      151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    151 GAAAAACATAAAGGTCTCGGTAAATAAGAATGTACTCACTATAAGTGCGG 200

WT      201 AG**AGA**AAGAGAGAATACTCTACCGTATATATCGATCAGCGCGTTGACAAA 250
            |||..|||||||||||||||||||||||||||||    ||||||||||
R69M    201 AG**ATG**AAGAGAGAATACTCTACCGTATATATCGATCA----GTTGACAAA 246

WT      251 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    247 GTGTATAAAGTAGTTAAGCTGCCCGTAGAGATTGAGCAGCAGGACATATC 296

WT      301 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
R69M    297 TGCTAAGTATAGTGAAGGCATACTTACAGTTAGAATGAAAACCAAGAACA 346

WT      351 TAAAGAACGTAGAAATAGAATAA   373
            |||||||||||||||||||||||
R69M    347 TAAAGAACGTAGAAATAGAATAA   369
```

The othere sequences were out of range due to sequencing primer position and are not shown here.

APPENDIX D

DLS GRAPHS AND RAW DATA

The DLS graphs for all the experiments are shown in this appendix. After the graphs, the tables of the data for each experiment are shown.
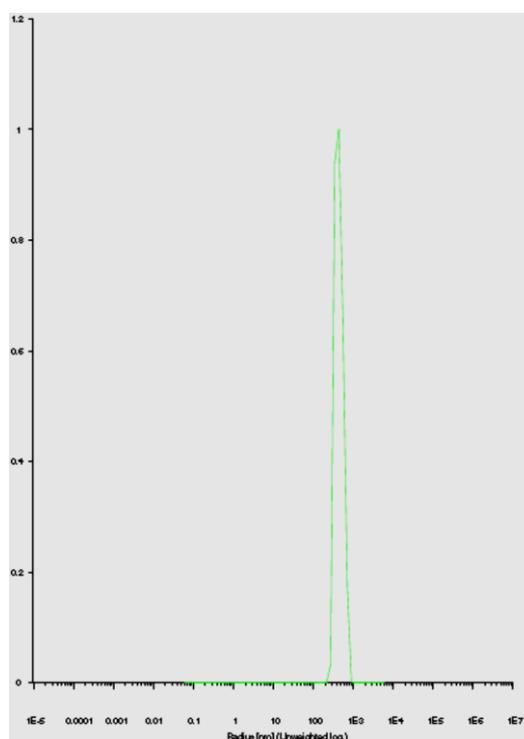
Figure D1: DLS graph for 30 second measurement at 13.6°C.
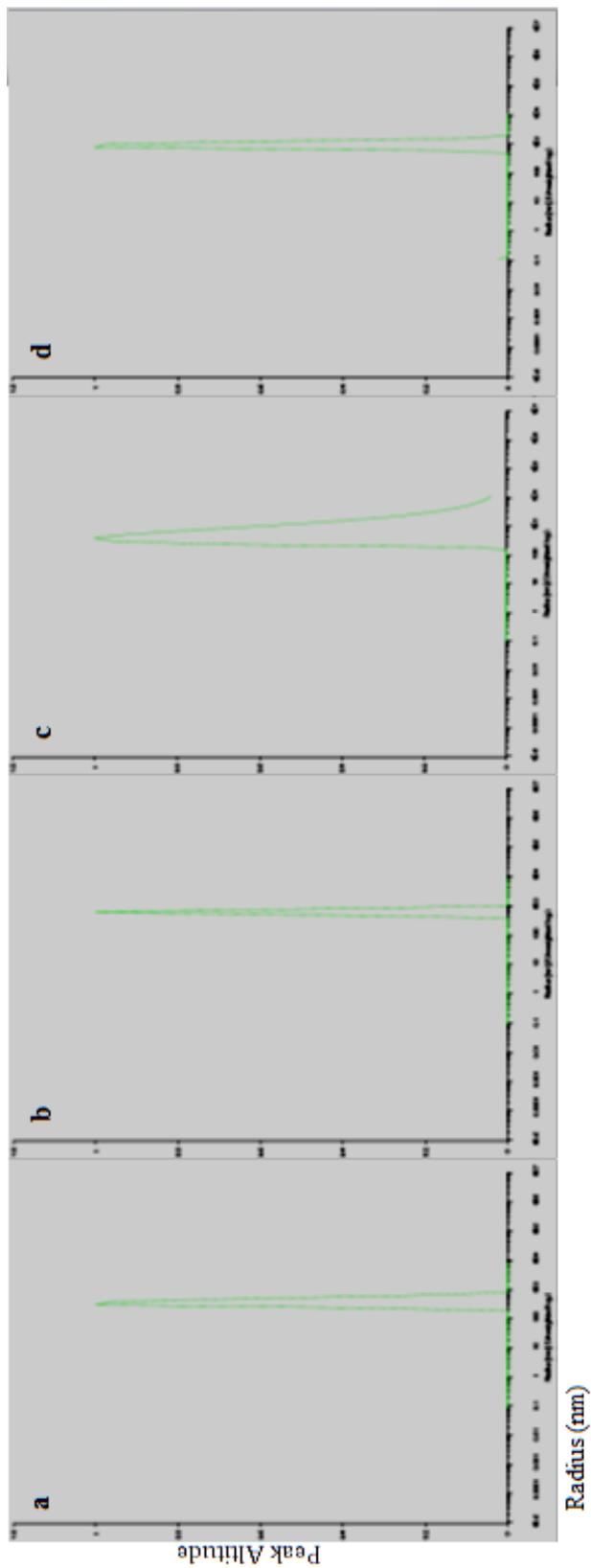
Figure D2: DLS graphs for 30 Seconds measurements. (a): First DLS measurement for 25.4°C. (b) Second DLS measurement for 25.4°C. (c) First DLS measurement for 35°C. (d) Second DLS measurement for 35 °C

230
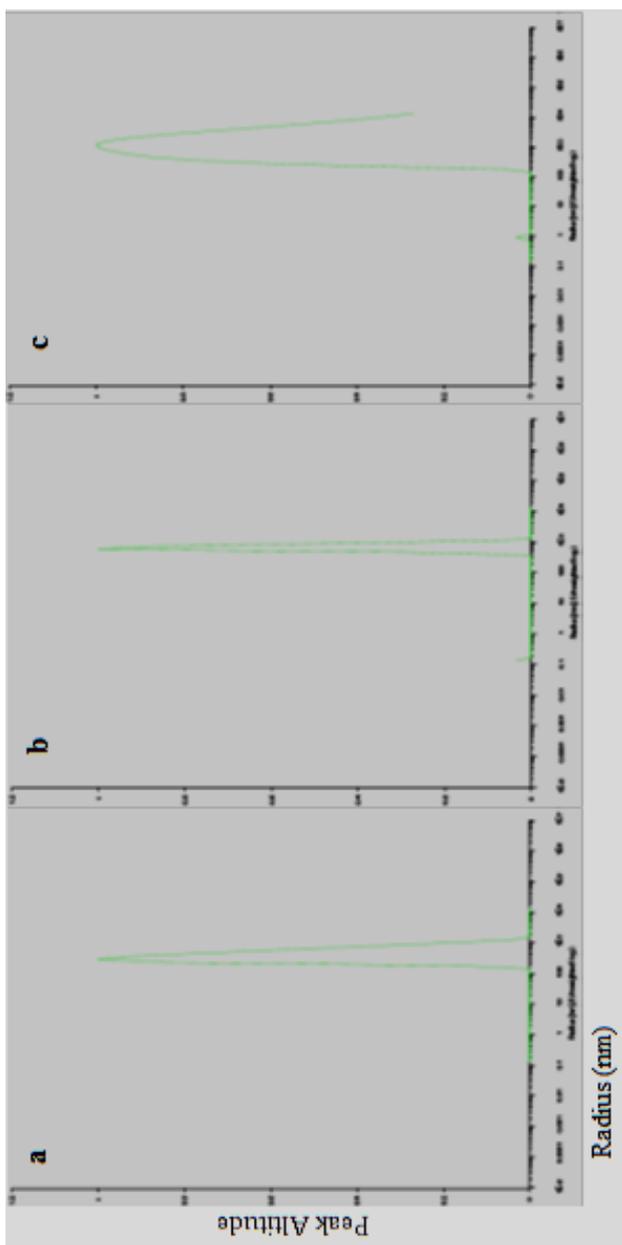
Figure D3: DLS graphs for 30 Seconds measurements. (a): First DLS measurement for 45°C. (b) Second DLS measurement for 45°C. (c) First DLS measurement for 60°C.

231
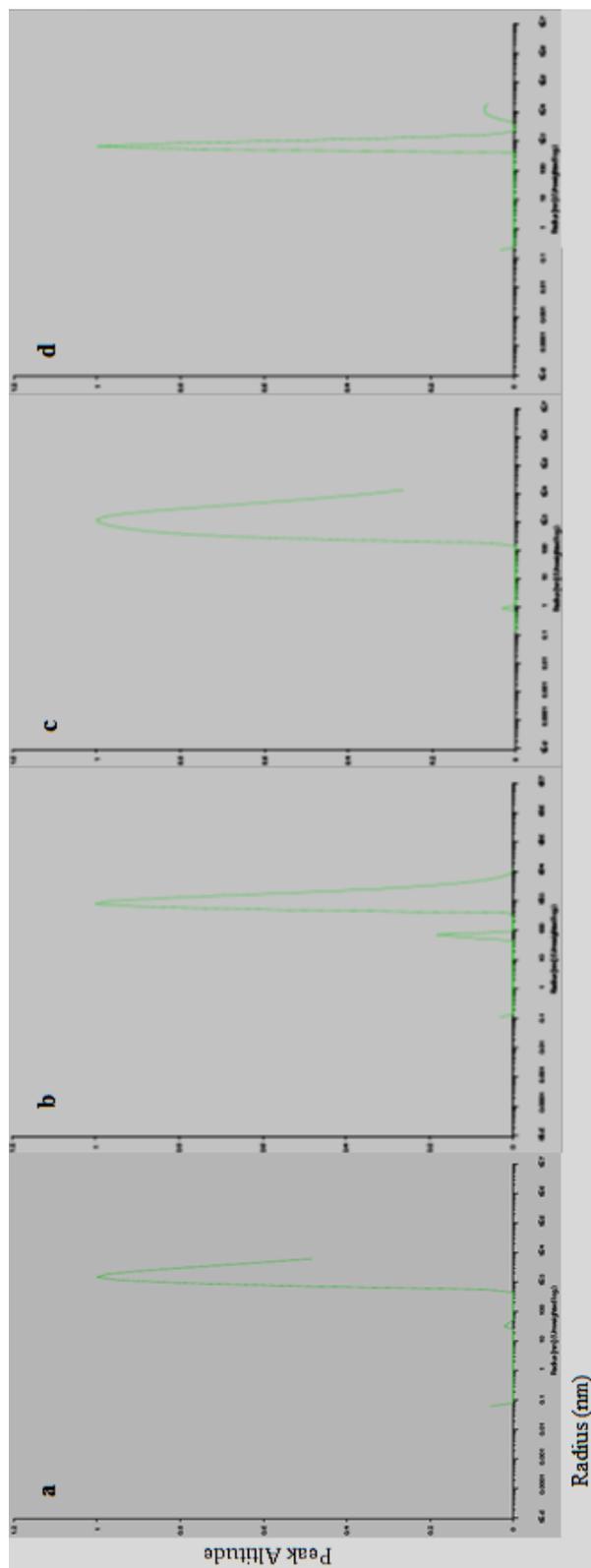
Figure D4: DLS graphs for 10 Seconds measurements. (a) DLS measurement at 13.6°C. (b) DLS measurement at 35°C. (c) DLS measurement for 44°C. (d) DLS measurement for 60 °C.

232
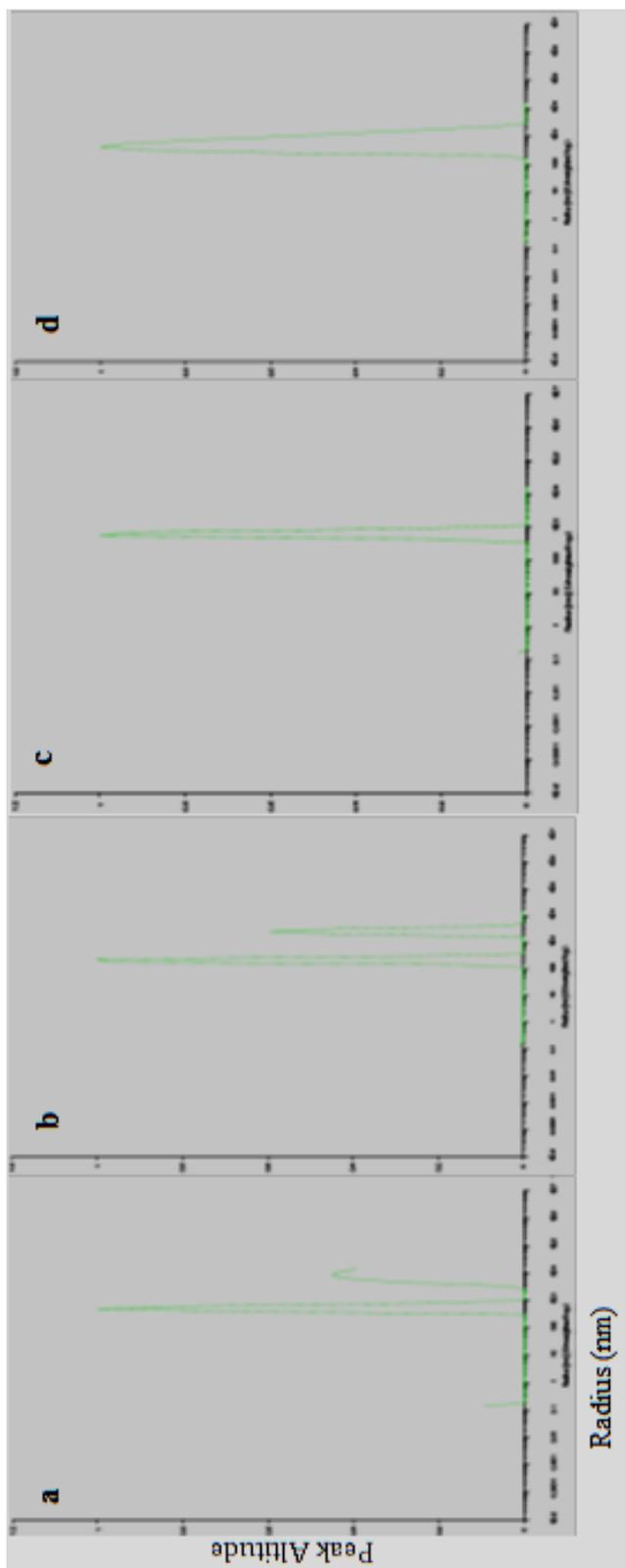
Figure D5: DLS graphs for Wild-Type Tpv-Hsp14.3 incubated with CS and GDH Enzymes. (a) 1mg total sHSPs incubated at 47°C for 10 minutes. (b) 1mg sHSP plus 15μl CS, incubated for 10 minutes at 47°C. (c) 0.85mg total sHSPs incubated at 53°C for 10 minutes. (b) 0.85mg sHSP plus 15μl GDH, incubated for 10 minutes at 53°C.

233
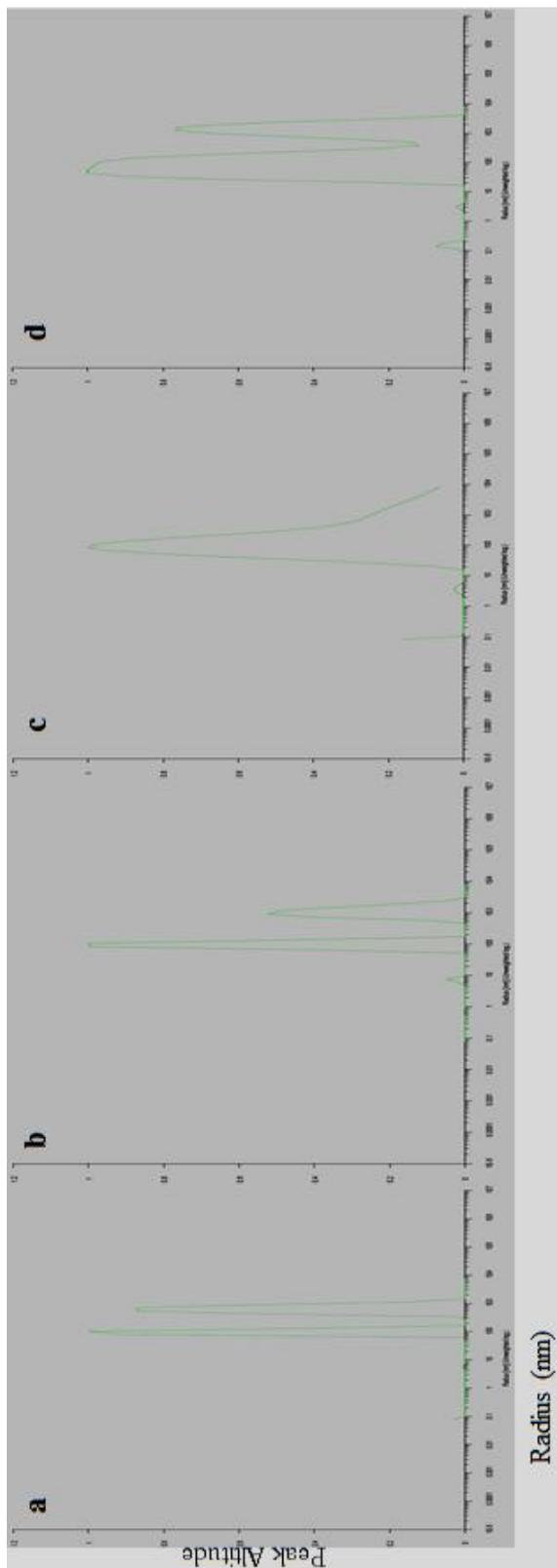
Figure D6: DLS Graphs for the Wild-type Tpv-Hsp14.3: (a) sHSP incubated at 20°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.

234
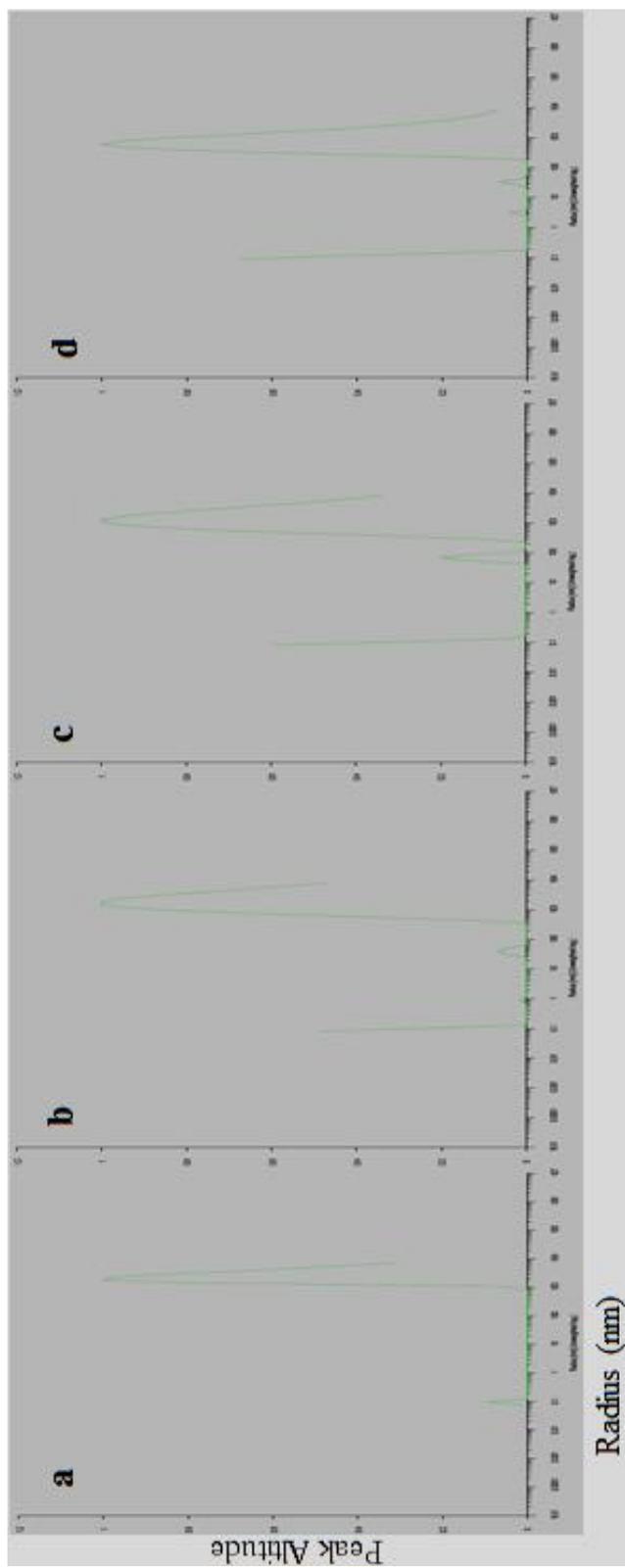
Figure D7: DLS Graphs for the R69K Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 20°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
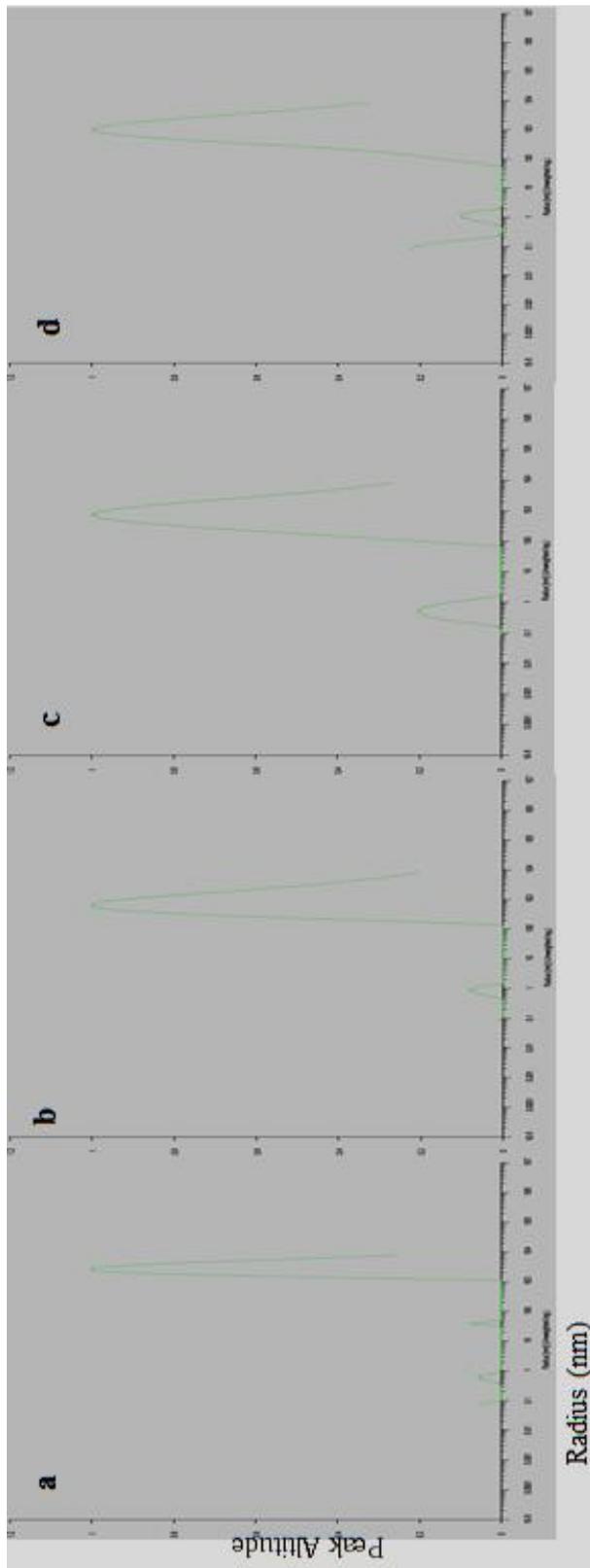
Figure D8: DLS Graphs for the R69E Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
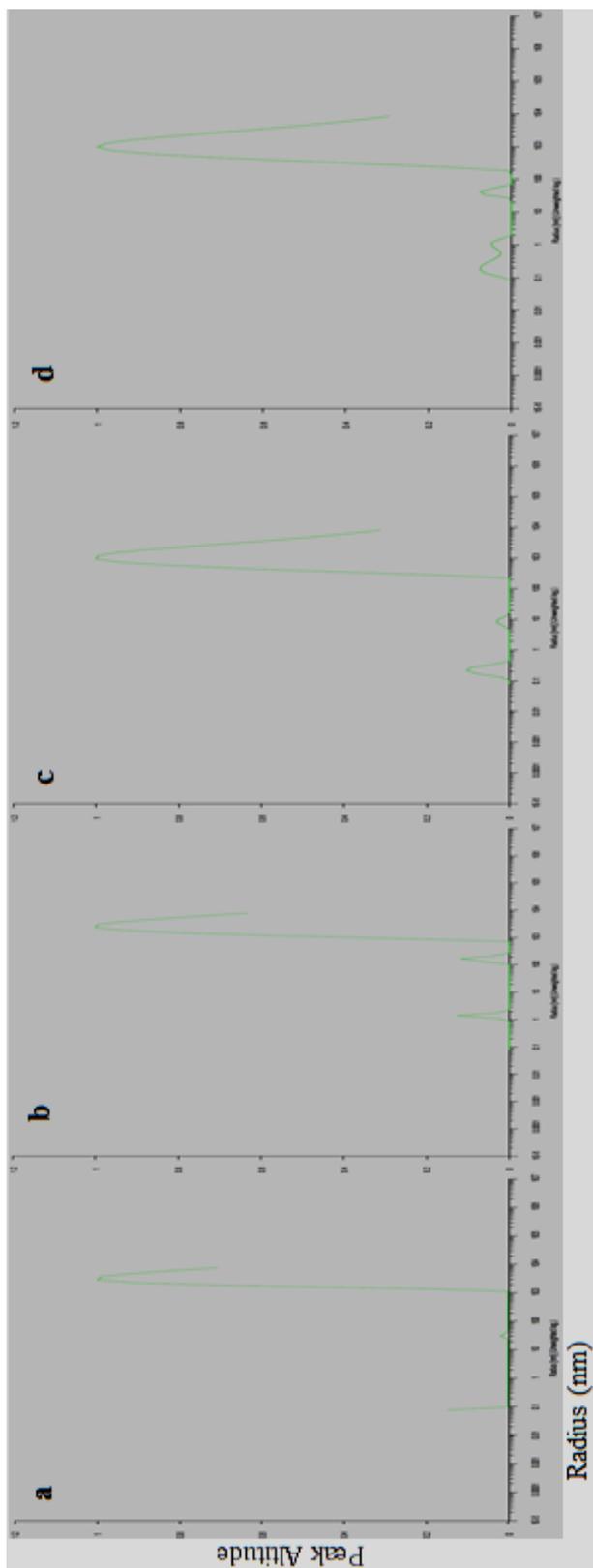
Figure D9: DLS Graphs for the R69M Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
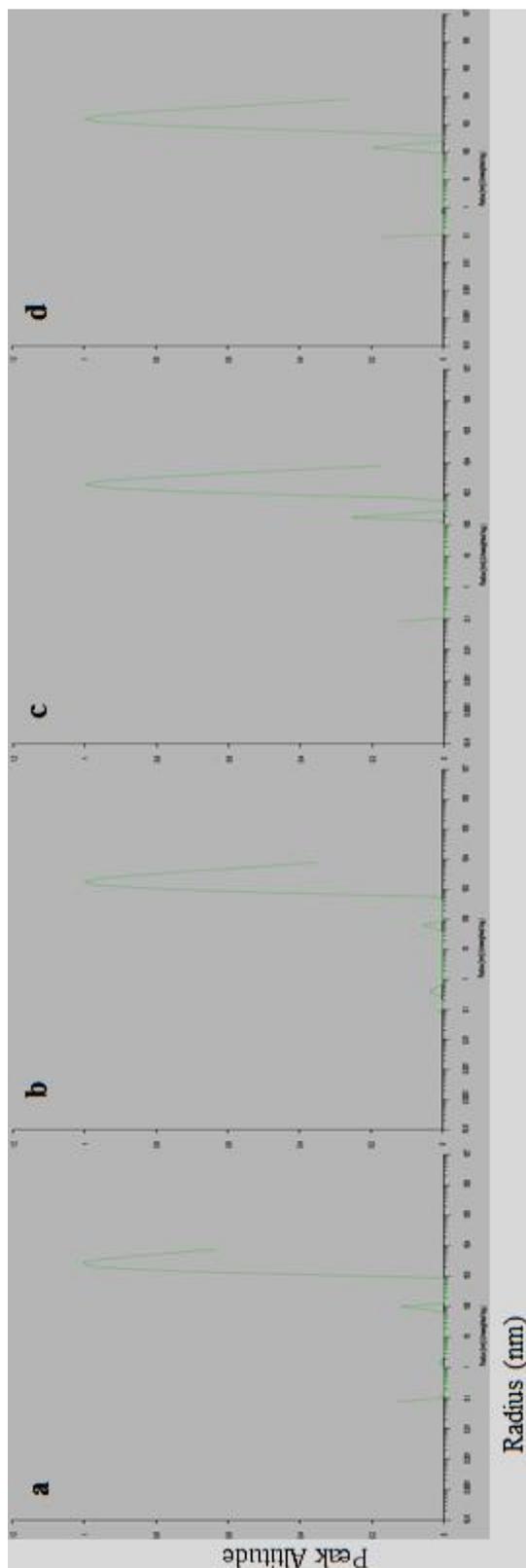
Figure D10: DLS Graphs for the R81K Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.

238
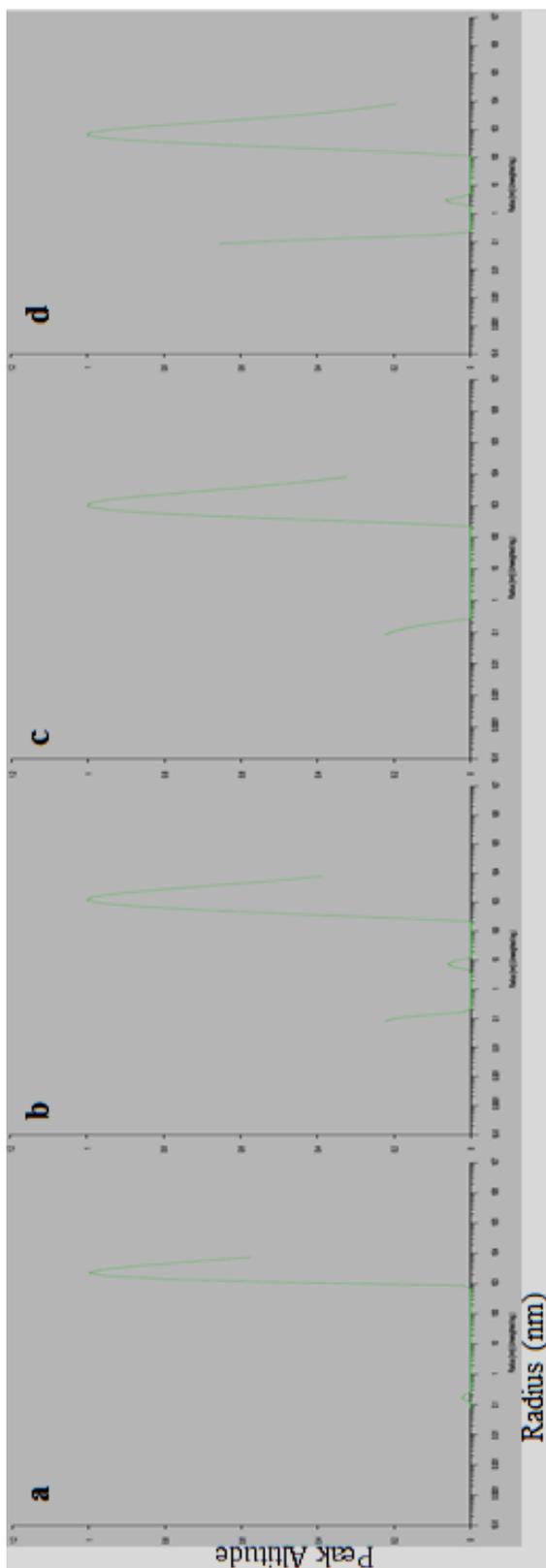
Figure D11: DLS Graphs for the R81E Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.

239
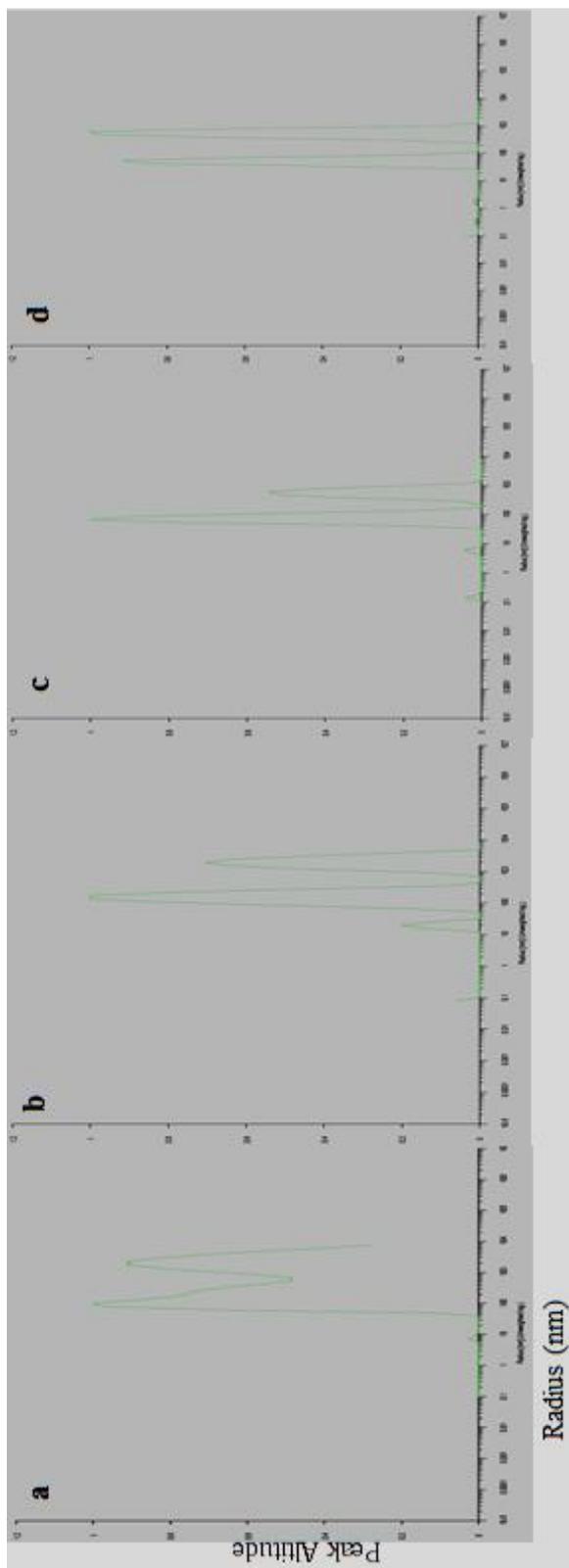
Figure D12: DLS Graphs for the R81M Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.

240
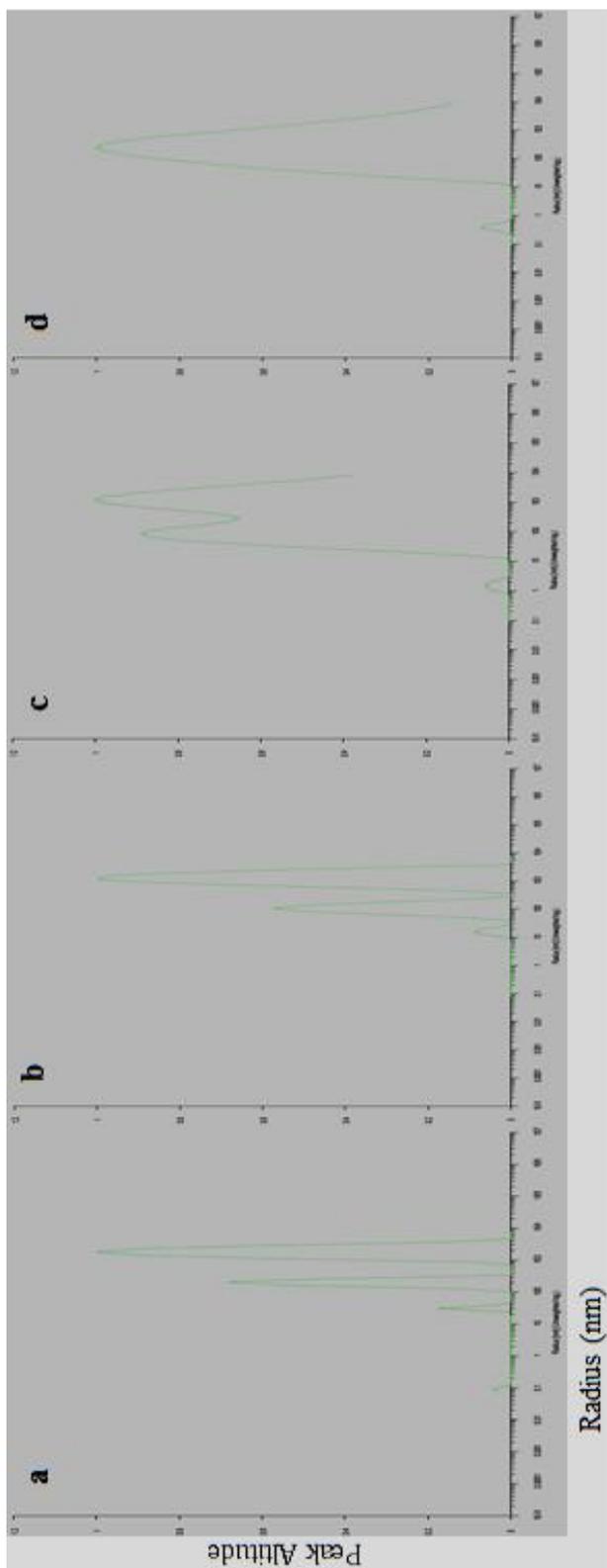
Figure D13: DLS Graphs for the QR(80-81)EL Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
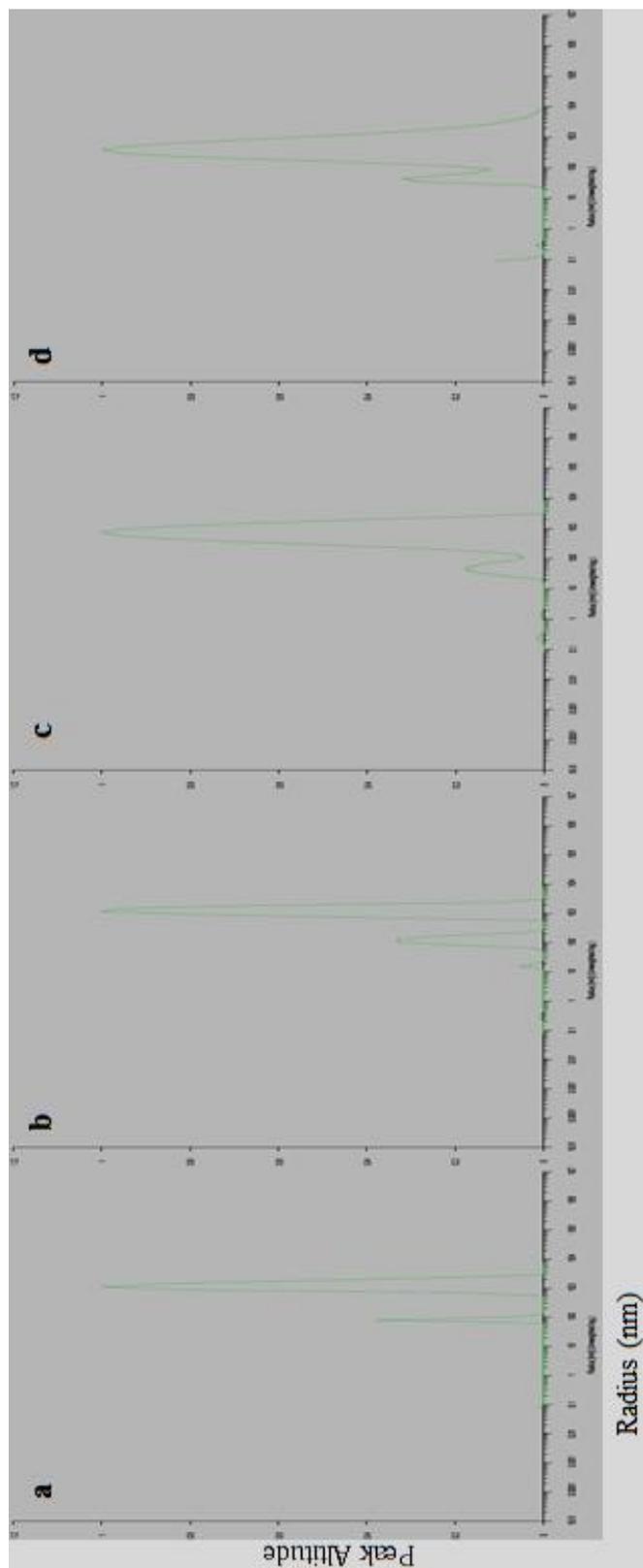
Figure D14: DLS Graphs for the K87R Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
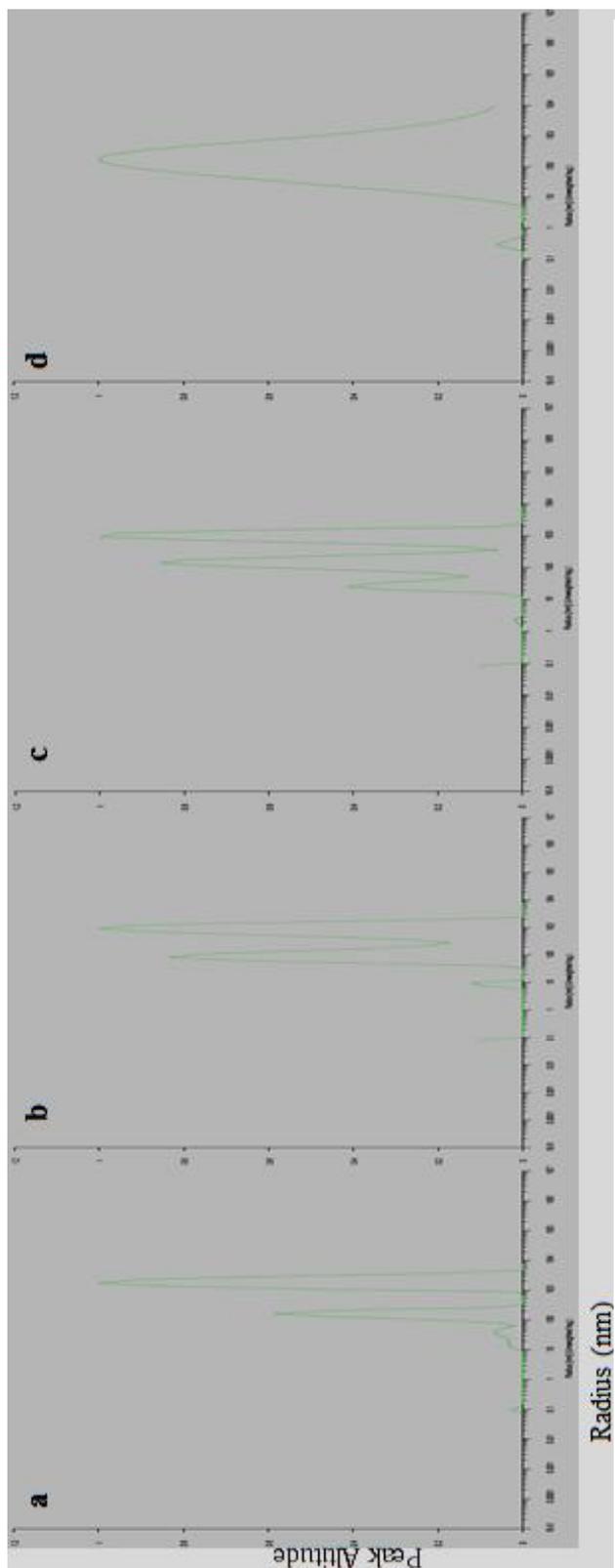
Figure D15: DLS Graphs for the K87E Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.
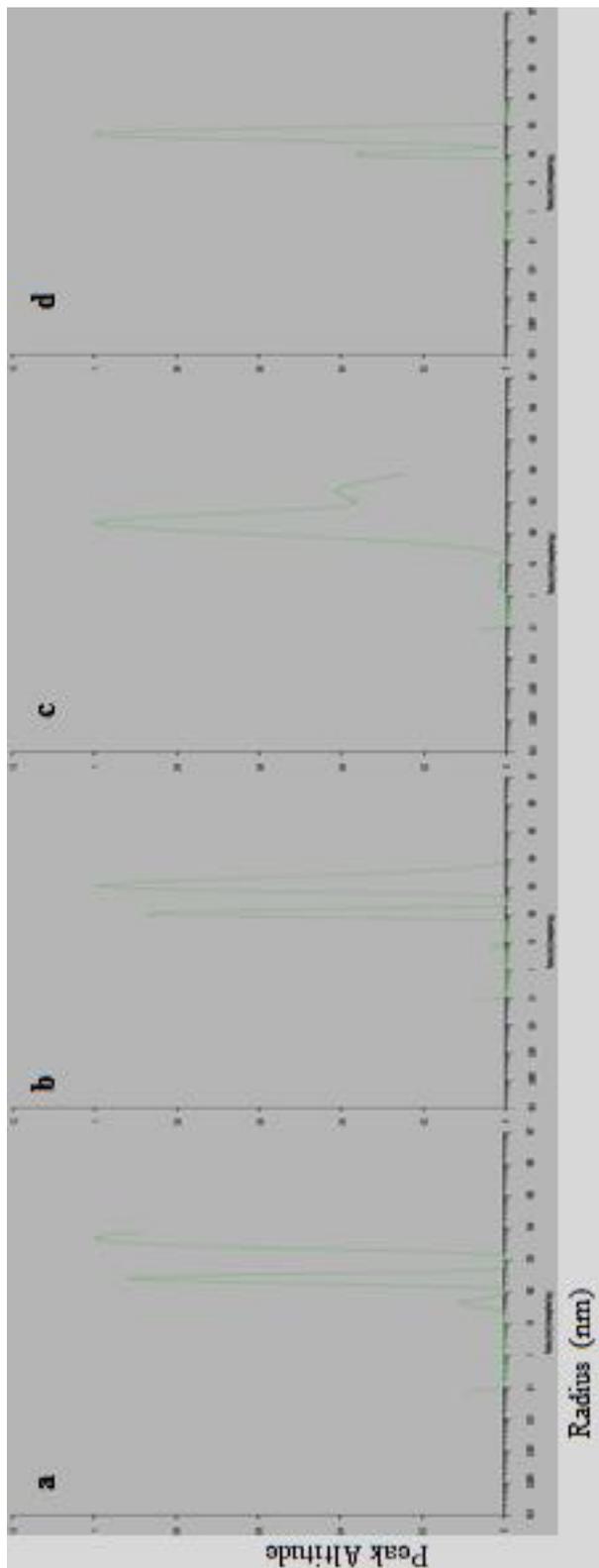
Figure D16: DLS Graphs for the K87I Mutant Variant Tpv-Hsp14.3: (a) sHSP incubated at 23°C for 10 minutes. (b) sHSP incubated for 10 minutes at 35°C. (c) sHSP incubated at 45°C for 10 minutes. (d) sHSP incubated for 10 minutes at 60°C.

Table D1: DLS Peak Values of Wild-Type sHSP for 30 Seconds Measurements.

Temperature (°C)	Peaks	Radius (nm)	Peak. Weight	% Peak Weight	Peak Width (nm)
	1	0	0	0	0
13.6°C	2	443.2	2.8	100	0.22
	Entire	443.2	2.8	100	0.22
25.4°C First Measure	1	0	0	0	0
	2	365.4	3.2	100	0.27
	Entire	365.4	3.2	100	0.27
25.4°C Second Measure	1	0	0	0	0
	2	619	2	100	0.16
	Entire	619	2	100	0.16
35°C First Measure	1	0	0	0	0
	2	692.7	7.8	100	0.55
	Entire	692.7	7.8	100	0.55
35°C Second Measure	1	0	0.02	0.85	0
	2	950.5	2.8	99.2	0.23
	Entire	880.1	2.86	100	0.86
44°C First Measure	1	0	0	0	0
	2	392.6	4.9	100	0.44
	Entire	392.6	4.9	100	0.44
44°C Second Measure	1	0	0.03	1.4	0
	2	676.3	2.458	98.6	0.21
	Entire	599.9	2.494	100	1
60°C	1	0	0.03	0.3	0
	2	1406	12.92	99.8	1.1
	Entire	1381	12.95	100	1.1

Table D2: DLS Peak Values of Wild-Type sHSP for 10 Seconds Measurements.

Temperature (°C)	Peaks	Radius (nm)	P eak Weight	% of Peak Weight	Peak Width (nm)
	1	0.06241	0.06	0.7	0.00
13.6°C	2	34.6	0.03	0.4	0.11
	3	1985	7.7	98.9	0.63
	Entire	1812	7.8	100.0	1.10
35.1°C	1	0.11	0.03	0.5	0.00
	2	65.59	0.3	4.6	0.12
	3	1121	6	94.9	0.57
	Entire	938.6	6.3	100.0	1.00
44°C	1	0	0	0.0	0.00
	2	0.92	0.03	0.3	0.00
	3	1406	12.9	99.8	1.10
	Entire	1381	12.9	100.0	1.00
60°C	1	0.19	0.03	0.9	0.00
	2	764.2	3.4	89.3	0.30
	3	10910	0.4	9.8	0.39
	Entire	922.3	3.8	100.0	1.16

Table D3: DLS Peak Values of Wild-Type sHSP with Substrates for 10 Seconds Measurements.

Protein Samples	Temperature	Peaks	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)
1mg sHSP	47°C	1	0.14	0.09	1.8	0.00
		2	518.9	2.6	51.5	0.21
		3	7999	2.36	46.6	0.40
		Entire	1597	5.1	100.0	1.89
1mg sHSP + 15µL CS	47°C	1	0	0	0.0	0.00
		2	0.1415	0.009	0.2	0.00
		3	218.2	2.4	61.6	0.19
		4	2660	1.5	38.2	0.20
		Entire	556.2	3.9	100.0	1.30
0.85mg sHSP	53°C	1	0	0	0.0	0.00
		2	0	0	0.0	0.00
		3	0.1588	0.01	0.8	0.00
		4	627.5	2.4	99.2	0.20
		Entire	588	2.5	100.0	0.76
0.85mg sHSP + 15µL GDH	53°C	1	0	0	0.0	0.00
		2	0	0	0.0	0.00
		3	0.3265	0.002	0.0	0.00
		4	587.3	6.2	100.0	0.55
		Entire	586.1	6.2	100.0	0.56

Table D4: DLS Peaks Values of Tpv-Hsp14.3 Variants.

Mut. Variant			Wild-Type				R69K				R69E			
Temp. (°C)	Peaks		Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)
20°C	1		0.075	0.03	0.6	0.00	0	0	0	0.00	0.076	0.05602	1.0	0.00
	2		5.66	0.01	0.1	0.00	0	0	0	0.00	0.59	0.125	2.1	0.18
	3		96.37	2.40	43.7	0.18	0.094	0.1034	1.8	0.07	38.87	0.0765	1.3	0.93
	4		630.2	3.05	55.6	0.27	2730	5.602	98.2	0.50	3014	5.59	95.6	0.46
	Entire		261.5	5.48	100.0	1.15	2266	5.705	100.0	1.46	2144	5.847	100.0	1.72
35°C	1		0.69	0.00	0.0	0.00	0.087	0.7171	7.1	0.11	0.08	0.01047	0.1	0.00
	2		7.12	0.07	1.3	0.10	0.89	0.01458	0.1	0.00	0.82	0.2353	2.2	0.23
	3		94.83	2.89	57.3	0.20	40.34	0.1572	1.6	0.18	944.3	10.51	97.7	0.94
	4		1071	2.09	41.3	0.34	2053	9.143	91.2	0.74	0	0	0.0	0.00
	Entire		249.5	5.04	100.0	1.30	929.3	10.03	100.0	2.73	802.1	10.76	100.0	1.42
45°C	1		0.083	0.17	1.5	0.05	0	0	0.0	0.00	0	0	0.0	0.00
	2		3.64	0.06	0.5	0.18	0.09	0.9581	8.6	0.14	0	0	0.0	0.00
	3		178	10.78	97.9	1.31	68.88	0.4617	4.1	0.19	0.51	1.263	9.2	0.50
	4		0	0.00	0.0	0.00	1555	9.736	87.2	0.83	862.6	12.49	90.8	1.10
	Entire		154.9	11.01	100.0	1.63	591.6	11.16	100.0	2.87	436	13.75	100.0	2.38
60°C	1		0.15	0.15	1.1	0.16	0.1	1.4	14.8	0.18	0.113	0.664	5.0	0.25
	2		2.94	0.03	0.2	0.10	3.12	0.0415	0.4	0.00	1.071	0.3533	2.7	0.29
	3		189.4	13.42	98.7	1.47	33.03	0.09851	1.0	0.13	1074	12.19	92.3	1.09
	4		0	0.00	0.0	0.00	875.6	7.938	83.8	0.79	0	0	0.0	0.00
	Entire		173.6	13.60	100.0	1.65	216.3	9.478	100.0	3.31	563.3	13.21	100.0	2.48

Table D4: DLS Peaks Values of Tpv-Hsp14.3 Variants (continuing).

Variants		R69M				R81K				R81E			
Temp. (°C)	Peaks	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)
23°C	1	0.076	0.1494	2.3	0.00	0.079	0.1476	2.0	0.07	0	0	0.0	0.00
	2	32.41	0.02702	0.4	0.10	1.468	0.01788	0.2	0.11	0	0	0.0	0.00
	3	3567	6.226	97.2	0.48	95.5	0.1569	2.1	0.10	0.17	0.03952	0.5	0.12
	4	0	0	0.0	0.00	3026	7.124	95.6	0.55	3842	7.176	99.6	0.56
	Entire	2721	6.403	100.0	1.71	2238	7.449	100.0	1.68	2695	7.207	100.0	0.91
35°C	1	1.411	0.1784	2.2	0.13	0.07945	0.01107	0.1	0.00	0	0	0.0	0.00
	2	165.4	0.22	2.7	0.16	0.41	0.07994	1.0	0.19	0.099	0.5645	5.1	0.21
	3	2920	7.667	95.1	0.60	60.84	0.08045	1.0	0.11	7.365	0.157	1.4	0.20
	4	0	0	0.0	0.00	2218	7.841	97.9	0.65	1528	10.28	93.5	0.86
	Entire	2280	8.066	100.0	1.34	1936	8.012	100.0	1.19	862.9	11	100.0	2.36
45°C	1	0.22	0.3275	3.2	0.26	0.08191	0.1222	1.7	0.00	0	0	0.0	0.00
	2	8.63	0.08018	0.8	0.22	188.1	0.3939	5.4	0.13	0	0	0.0	0.00
	3	1434	9.956	96.1	0.85	2373	6.794	92.9	0.58	0.12	0.7927	7.3	0.31
	4	0	0	0.0	0.00	0	0	0.0	0.00	1445	10.07	92.6	0.85
	Entire	1044	10.36	100.0	1.80	1744	7.31	100.0	1.53	727	10.87	100.0	2.59
60°C	1	0.35	0.5418	4.9	0.80	0.085	0.1687	1.9	0.00	0	0	0.0	0.00
	2	41.54	0.1824	1.7	0.19	0.74	0.005864	0.1	0.00	0.1052	1.525	12.4	0.22
	3	1353	10.31	93.5	0.89	151.4	0.4351	4.8	0.19	2.985	0.1394	1.1	0.17
	4	0	0	0.0	0.00	1975	8.382	93.2	0.72	942.4	10.68	86.5	0.97
	Entire	8512	11.03	100.0	2.03	1437	8.992	100.0	1.62	286.8	12.34	100.0	3.16

Table D4: DLS Peaks Values of Tpv-Hsp14.3 Variants (continuing).

Variants		81M				QR(80-81)EL				K87R			
Temp. (°C)	Peaks	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)
20°C	1	0	0	0	0.00	0.087	0.07141	1.2	0.12	0.076	0.01381	0.3	0.00
	2	0.0755	0.005622	0.0	0.00	31.61	0.2204	3.6	0.10	82.23	0.4481	11.0	0.09
	3	7.577	0.03168	0.2	0.01	197	1.778	28.7	0.21	1230	3.615	88.7	0.31
	4	567.7	14.91	99.7	1.44	1816	4.131	66.6	0.34	0	0	0.0	0.00
	Entire	560.6	14.95	100.0	1.46	741.5	6.201	100.0	1.56	884	4.077	100.0	1.05
35°C	1	0.07945	0.05874	0.7	0.00	0	0	0.0	0.00	0.2906	0.01445	0.3	0.12
	2	19.69	0.0449	0.5	0.18	16.32	0.2378	2.8	0.22	15.78	0.05767	1.1	0.06
	3	148.7	4.807	56.5	0.40	107.3	2.493	29.4	0.37	121.8	1.165	22.7	0.28
	4	1932	3.2	37.6	0.37	1264	5.754	67.8	0.48	1241	3.894	75.9	0.31
	Entire	332.6	8.515	100.0	1.60	542.1	8.485	100.0	1.34	681.2	5.132	100.0	1.17
45°C	1	0.1411	0.05278	0.9	0.11	0	0	0.0	0.00	0.23	0.02707	0.3	0.12
	2	5.67	0.06538	1.1	0.11	0	0	0.0	0.00	1.34	0.01527	0.2	0.15
	3	73.46	3.567	61.6	0.30	1.477	0.2015	1.1	0.28	527.5	8.372	99.5	1.00
	4	567.6	2.11	36.4	0.32	420.6	18.99	99.0	1.64	0	0	0.0	0.00
	Entire	141.9	5.795	100.0	1.27	396.4	19.19	100.0	1.73	509	8.414	100.0	1.13
60°C	1	0.0854	0.03314	0.5	0.00	0	0	0.0	0.00	0.09	0.144	1.5	0.10
	2	0.33	0.01393	0.2	0.12	0	0	0.0	0.00	0.29	0.01955	0.2	0.07
	3	52.14	2.957	44.6	0.26	0.39	0.1866	1.1	0.21	362.4	9.448	98.3	1.10
	4	535.1	3.608	54.4	0.29	287.5	16.12	98.9	1.44	0	0	0.0	0.00
	Entire	175.5	6.632	100.0	1.36	266.5	16.3	100.0	1.59	315.5	9.611	100.0	1.49

Table D4: DLS Peaks Values of Tpv-Hsp14.3 Variants (continuing).

Variants			K87E				K87I			
Temp. (°C)	Peaks		Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)	Radius (nm)	Peak Weight	% Peak Weight	Peak Width (nm)
20°C	1		0.09796	0.03038	0.5	0.07	0.076	0.08631	1.0	0.00
	2		120.2	2.131	33.9	0.67	44.09	0.2827	3.3	0.21
	3		1878	4.121	65.6	0.34	258.3	2.823	33.1	0.26
	4		0	0	0.0	0.00	4272	5.334	62.6	0.41
	Entire		704.9	6.282	100.0	1.52	1298	8.526	100.0	1.80
35°C	1		0.07948	0.01008	0.1	0.00	0.079	0.06977	0.9	0.00
	2		8.667	0.2168	2.3	0.12	7.63	0.03651	0.4	0.04
	3		327.3	9.21	96.7	1.15	116.9	2.493	30.7	0.23
	4		0	0	0.0	0.00	1431	5.532	68.0	0.51
	Entire		275.9	9.528	100.0	1.51	596.1	8.132	100.0	1.52
45°C	1		0.0819	0.09878	1.1	0.00	0.082	0.06865	0.6	0.00
	2		2.164	0.02642	0.3	0.11	4.44	0.1273	1.1	0.63
	3		254.9	9.275	98.7	1.33	441.1	11.79	98.4	1.38
	4		0	0	0.0	0.00	0	0	0.0	0.00
	Entire		231.2	9.4	100.0	1.58	399.9	11.98	100.0	1.58
60°C	1		0.3	0.1447	0.9	0.21	0	0	0.0	0.00
	2		1.16	0.002406	0.0	0.08	0	0	0.0	0.00
	3		185.5	15.59	99.0	1.45	0.085	0.007449	0.1	0.00
	4		0	0	0.0	0.00	406.7	5.353	99.9	0.68
	Entire		174.7	15.74	100.0	1.57	402	5.361	100.0	0.75

APPENDIX E

INFORMATION REGARDING THE PROTEINS USED IN MULTIPLE SEQUENCE ALIGNMENTS

Table E1: The 63 Best-Matching Proteins of Tpv-Hsp14.3 with Sequence Identity 30% and Higher Used in MSA.

No.	Name of the Organism	Accession Number	Protein Name
1	*Thermoplasma volcanium GSS1*	NP_111294.1	Hsp20-Related
2	*Thermoplasma acidophilum DSM 1728*	NP_394323.1	Hsp20
3	*Ferroplasma acidarmanus*	YP_008142169.1	Hsp 20
4	*Halobacterium sp. NRC-1*	NP_395730.1	Heat Shock Protease Protein
5	*Picrophilus torridus DSM 9790*	YP_023517.1	Hsp 20 Family sHSP
6	*halophilic archaeon DL31*	YP_004809151.1	Hsp20
7	*Anaeromyxobacter dehalogenans 2CP-C*	YP_466070.1	Hsp20
8	*Sulpholobus tokadii ST1653*	Q970D9	Hsp14.0
9	*crenarchaeote HF4000_ANIW141J13*	ABZ07832.1	Hsp20
10	*Nocardiopsis lucentensis*	WP_017599436.1	Hsp20
11	*Sulfolobus islandicus Y.N.15.51*	YP_002841880.1	Hsp20
12	*Nesterenkonia sp. F*	WP_010524295.1	Hsp20
13	*Candidatus Nitrosoarchaeum koreensis*	WP_007551556.1	Molecular Chaperone
14	*Candidatus Nitrosoarchaeum limnia*	WP_010189564.1	Molecular Chaperone
15	*Nocardiopsis dassonvillei DSM 43111*	YP_003681484.1	Hsp20
16	*Nocardiopsis synnemataformans*	WP_017566186.1	Hsp20
17	*Candidatus Nitrosopumilus salaria*	WP_008298667.1	Molecular Chaperone
18	*Chthoniobacter flavus*	WP_006981010.1	Hsp20
19	*Candidatus Nitrososphaera gargensis Ga9.2*	YP_006863216.1	Hsp21
20	*Candidatus Nitrosoarchaeum limnia*	WP_007402545.1	Molecular Chaperone

Table E1: The 63 Best-Matching Proteins of Tpv-Hsp14.3 with Sequence Identity 30% and Higher Used in MSA (Continuing).

21	*Geobacter lovleyi SZ*	YP_001953023.1	Hsp20
22	*Sulfolobus solfataricus P2*	NP_343935.1	Hsp20
23	*Nocardiopsis sp. CNS639*	WP_019607970.1	Hsp20
24	*Nocardiopsis alkaliphila*	WP_017604807.1	Hsp20
25	*Nocardiopsis halotolerans*	WP_017571483.1	Hsp20
26	*Nocardiopsis ganjiahuensis*	WP_017584932.1	Hsp20
27	*Petrotoga mobilis SJ95*	YP_001568567.1	Hsp20
28	*Hydrogenobacter thermophilus TK-6*	YP_003431999.1	Hsp20
29	*Sphaerochaeta pleomorpha str. Grapes*	YP_005063435.1	Molecular Chaperone
30	*Geobacter metallireducens GS-15*	YP_006721940.1	ATP-Independent Chaperone
31	*delta proteobacterium NaphS2*	YP_001900983.1	Hsp20
32	*Ralstonia pickettii*	WP_004628761.1	Molecular Chaperone
33	*Geobacter sulfurreducens PCA*	NP_951596.1	ATP-Independent Chaperone
34	*Candidatus Nitrosopumilus sp. AR2*	YP_006776550.1	Hsp20
35	*Microcoleus vaginatus*	WP_006634026.1	Molecular Chaperone
36	*Thermosipho africanus TCF52B*	YP_002335217.1	Hsp20
37	*uncultured crenarchaeote*	ABB88979.1	Hsp20
38	*Candidatus Nitrosopumilus koreensis AR1*	YP_006774607.1	Hsp20
39	*Candidatus Kuenenia stuttgartiensis*	CAJ73249.1	Similar To sHSPs
40	*Anabaena sp. 90*	YP_007000777.1	Hsp20
41	*Microcystis aeruginosa*	WP_002767035.1	Molecular Chaperone
42	*Microcoleus vaginatus*	WP_006633768.1	Molecular Chaperone
43	*Oscillatoria sp. PCC 6506*	WP_007354249.1	Molecular Chaperone
44	*Methanocella paludicola SANAE*	YP_003357413.1	Putative Shsps
45	*Cenarchaeum symbiosum A*	YP_876436.1	Shsps
46	*Nitrosopumilus maritimus SCM1*	YP_001583042.1	Hsp20
47	*Pedosphaera parvula*	WP_007418376.1	Hsp20
48	*Desulfohalobium retbaense DSM 5692*	YP_003197589.1	Hsp20
49	*Desulfurispirillum indicum S5*	YP_004112624.1	Hsp20
50	*Oscillatoria formosa*	WP_019488126.1	Molecular Chaperone
51	*delta proteobacterium NaphS2*	WP_006421140.1	Hsp20
52	*Methylobacterium sp. GXF4*	WP_007563476.1	Hsp20
53	*Desulfomonile tiedjei DSM 6799*	YP_006446048.1	Molecular Chaperone
54	*Candidatus Poribacteria sp.*	WP_009370651.1	Hsp20

254

Table E1: The 63 Best-Matching Proteins of Tpv-Hsp14.3 with Sequence Identity 30% and Higher Used in MSA (Continuing).

55	*Thermotoga petrophila RKU-1*	YP_001244140.1	Hsp20
56	*Thermotoga sp. EMP*	WP_008193529.1	Hsp20
57	*Thermotoga maritima MSB8*	NP_228185.1	Class I Hsp
58	*Microcystis aeruginosa*	WP_004160648.1	Hspa
59	*Metallosphaera yellowstonensis*	WP_009070347.1	Hsp20
60	*Sulfolobus acidocaldarius DSM 639*	YP_256268.1	Hsp20
61	*Thermoanaerobacterium xylanolyticum LX-11*	YP_004469872.1	Hsp20
62	*Thermodesulfovibrio yellowstonii DSM 11347*	YP_002248210.1	Hsp20
63	*Candidatus Chloracidobacterium thermophilum B*	YP_004861811.1	Molecular Chaperone

Table E2: The Archaeal Best-Matching Proteins of Tpv-Hsp14.3 with Sequence Identity 30% and Higher Used in MSA.

No.	Name of the Organism	Accession Number	Protein Name
1	*Thermoplasma volcanium GSS1*	NP_111294.1	Hsp20-Related
2	*Thermoplasma acidophilum DSM 1728*	NP_394323.1	Hsp20
3	*Ferroplasma acidarmanus*	YP_008142169.1	Hsp 20
4	*Halobacterium sp. NRC-1*	NP_395730.1	Heat Shock Protease Protein
5	*Picrophilus torridus DSM 9790*	YP_023517.1	Hsp 20 Family sHSP
6	*Sulpholobus tokadii ST1653*	Q970D9	Hsp14.0
7	*crenarchaeote HF4000_ANIW141J13*	ABZ07832.1	Hsp20
8	*Sulfolobus islandicus Y.N.15.51*	YP_002841880.1	Hsp20
9	*Candidatus Nitrosoarchaeum koreensis*	WP_007551556.1	Molecular Chaperone
10	*Candidatus Nitrosoarchaeum limnia*	WP_010189564.1	Molecular Chaperone
11	*Candidatus Nitrosopumilus salaria*	WP_008298667.1	Molecular Chaperone
12	*Candidatus Nitrososphaera gargensis Ga9.2*	YP_006863216.1	Hsp21
13	*Candidatus Nitrosoarchaeum limnia*	WP_007402545.1	Molecular Chaperone
14	*Sulfolobus solfataricus P2*	NP_343935.1	Hsp20
15	*Candidatus Nitrosopumilus sp. AR2*	YP_006776550.1	Hsp20
16	*uncultured crenarchaeote*	ABB88979.1	Hsp20
17	*Candidatus Nitrosopumilus koreensis AR1*	YP_006774607.1	Hsp20
18	*Cenarchaeum symbiosum A*	YP_876436.1	Shsps
19	*Nitrosopumilus maritimus SCM1*	YP_001583042.1	Hsp20
20	*Metallosphaera yellowstonensis*	WP_009070347.1	Hsp20
21	*Sulfolobus acidocaldarius DSM 639*	YP_256268.1	Hsp20