November 2014

Properties of Potential Substrates of a Cyanobacterial Small Heat Shock Protein

Yichen Zhang
University of Massachusetts Amherst

PROPERTIES OF POTENTIAL SUBSTRATES OF A CYANOBACTERIAL SMALL HEAT SHOCK PROTEIN

A Thesis Presented

by

YICHEN ZHANG

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE

September 2014

Molecular and Cellular Biology

PROPERTIES OF POTENTIAL SUBSTRATES OF A CYANOBACTERIAL SMALL HEAT SHOCK PROTEIN

A Thesis Presented

by

YICHEN ZHANG

Approved as to style and content by:

Elizabeth Vierling, Chair

Danny Schnell, Member

Peter Chien, Member

Barbara Osborne, Program Director
Molecular and Cellular Biology Program

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Elizabeth Vierling, for three years of patient, helpful guidance and support. I also thank all my lab members for any help to my research. Appreciation also goes to my program director Dr. Barbara Osborne and my program manager Sarah Czerwonka for helping with my Masters studies. I would also thank my committee members Dr. Danny Schnell and Dr. Peter Chien for sharing suggestions about my research.

I wish to appreciate my parents for supporting my study abroad and all their love. I also would like to thank all my friends who have helped me during my years in the US.

I want to thank the National Institutes of Health (NIH) for providing funds for this research and offering travel expenses during preparation of this manuscript.

ABSTRACT

PROPERTIES OF POTENTIAL SUBSTRATES OF A CYANOBACTERIAL SMALL HEAT SHOCK PROTEIN

SEPTEMBER 2014

YICHEN ZHANG, B.S., SUN YAT-SEN UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Elizabeth Vierling

Most proteins must fold into native three-dimensional structures to be functional. But, newly synthesized proteins are at high risk of misfolding and aggregating in the cell. Stress, disease or mutations can also cause protein aggregation. A cyanobacterial small heat shock protein, Hsp16.6, can act as a chaperone to prevent irreversible protein aggregation during heat stress. This thesis is focused on the properties of proteins that were associated with Hsp16.6 during heat stress, and which therefore may be "substrates" of Hsp16.6. Bioinformatics were used to determine if Hsp16.6 preferentially binds to proteins with certain properties, and biochemical studies were performed to investigate how the substrates actually behave with Hsp16.6 during heat stress. It was found that Hsp16.6 preferentially binds to proteins with higher molecular weight, higher acidity, higher percentage of charged residues (especially negatively charged residues), and a lower percentage of hydrophobic residues compared to all proteins encoded by the *Synechocystis* genome. Proteins bound to Hsp16.6 were also slightly enriched in VQL motifs. The potential substrate fructose bisphosphate aldolase class II (FBA) was expressed in *E.coli* and purified. FBA could be protected by Hsp16.6 from aggregation through forming a complex with Hsp16.6

during heat stress in vitro, consistent with it being a substrate of Hsp16.6. Another potential substrate, elongation factor G1 (EF-G1) was also expressed in *E.coli* and purified. EF-G1 did not form insoluble aggregates even at 47℃, but circular dichroism spectroscopy revealed the secondary structure has melted at this temperature, and the protein eluted earlier than unheated protein on size exclusion chromatography. Thus, EF-G1 appears heat sensitive, and may also be an in vivo substrate of Hsp16.6. Lastly, in vivo study studies were performed to determine the amount of FBA and EF-G1 in *Synechocystis* cells. Both proteins are abundant, with FBA levels (around 2% of total cell protein) being about twice that of EF-G1. Further in vivo experiments will be needed to confirm that FBA and EF-G1 are substrates of Hsp16.6.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Most proteins must fold into their native, three-dimensional structures to be functional (Tyedmers, Mogk et al. 2010). However, newly synthesized proteins are at a high risk of misfolding and aggregating in the cell (Hartl, Bracher et al. 2011). Stress, disease or mutations can also cause protein aggregation (Gidalevitz, Kikis et al. 2010, Basha, O'Neill et al. 2012). In the cell, such aggregated proteins can be rescued by a protein quality control network of chaperones and proteases (Tyedmers, Mogk et al. 2010, Basha, O'Neill et al. 2012).

Chaperones

Chaperones help proteins fold effectively and also facilitate refolding of misfolded proteins, preventing or reversing protein aggregation (Tyedmers, Mogk et al. 2010, Hartl, Bracher et al. 2011, Basha, O'Neill et al. 2012). Heat shock proteins (HSPs) were the first proteins defined as chaperones (Hartl, Bracher et al. 2011). Major HSP chaperones include HSP40, HSP60 (chaperonins), HSP70, HSP90, HSP100 and the small heat shock proteins (sHSPs), when classified according to their molecular weight (MW) (Hartl, Bracher et al. 2011). HSP70, HSP90 and HSP60 all have ATPase activity and generally recognize and bind the hydrophobic regions that are exposed by non-native proteins. In this way, HSP70, HSP90 and HSP60 can facilitate *de novo* protein folding and refolding with the involvement of ATP (Hartl, Bracher et

al. 2011). They may also coordinate with ATP-independent chaperones, such as the sHSPs, which are believed as “holdases” to make substrates available to the other chaperones (Hartl, Bracher et al. 2011).

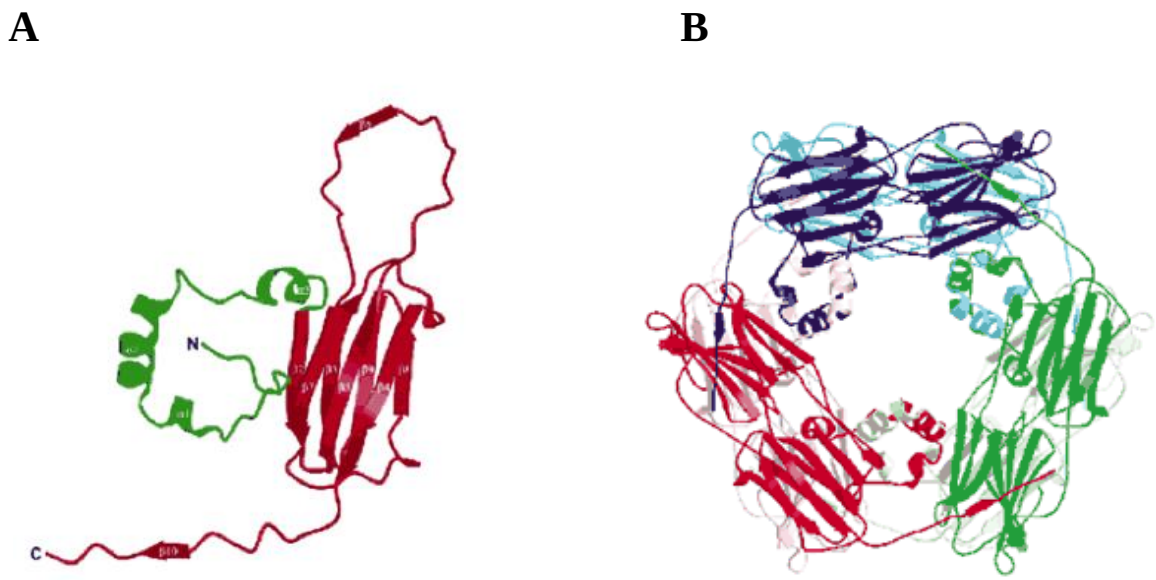
sHSPs

Unlike Hsp70, Hsp90 and chaperonins, sHSPs are ATP-independent chaperones (MacRae 2000, Haslbeck 2002, Giese and Vierling 2004). Monomers of sHSPs differ in size from ~12 to 42 kDa, but most of sHSPs exist in nature as oligomers that are 12 to >48 subunits (Basha, O'Neill et al. 2012). An alpha-crystallin (αC) domain (ACD) of ~90 amino acids, flanked by a highly variable N-terminal arm and a C-terminal extension, is the defining signature of sHSPs (Kriehuber, Rattei et al. 2010, Poulain, Gelly et al. 2010, Basha, O'Neill et al. 2012). The monomer structure of Hsp16.9 (1GME) as an example is shown as Figure 1A. The ACD consists of a β-sandwich, formed by seven or eight anti-parallel β-stands (Basha, O'Neill et al. 2012). The C-terminal extension is also divergent, except for a conserved I/V/L-X-I/V/L (IxI) motif involved in formation of sHSP oligomers (Basha, O'Neill et al. 2012). The structure of the oligomer of Hsp16.9 is shown in Figure 1B. The Hsp16.9 oligomer is a dodecamer, arranged as two stacked disks linked together by the C-terminal tails and N-terminal arms. Dimers are the building block of the oligomer and each disk consists of a trimer of dimers (van Montfort, Basha et al. 2001). However, not all sHSPs form homogeneous oligomers. Hsp16.6, a sHSP of *Synechocystis*, is a heterogeneous oligomer, comprising a distribution of oligomers from 12 to > 24 subunits.

Figure 1 Crystal Structure of Wheat Hsp16.9 (1GME) (van Montfort, Basha et al. 2001)

A Structure of an Hsp16.9 monomer. Hsp16.9 comprises a β-sandwich ACD (red), flanked by highly divergent N-terminal (green) and C-terminal extension (red).

B Structure of the Hsp16.9 dodecamer. Hsp16.9 is arranged as two interconnect, stacked disks. Each disk comprises three dimers. Each dimer is shown in a different color.



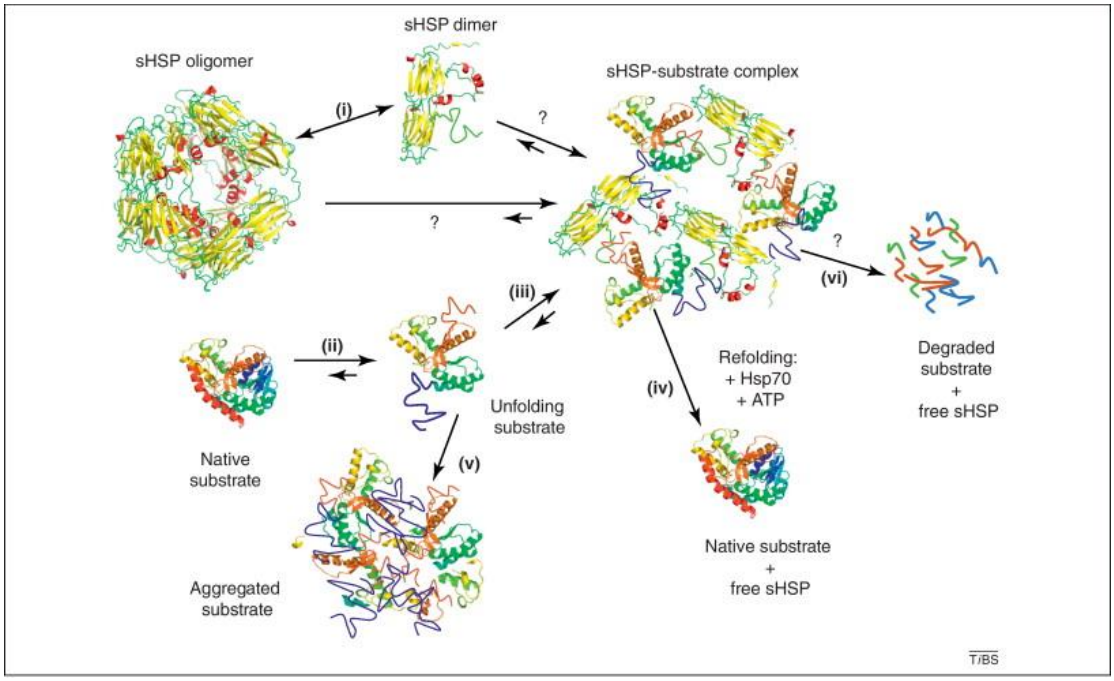
How sHSPs Functions have Become of Significant Interest to Medicine

Human sHSPs are linked to a variety of diseases. Defects of sHSPs may cause cataract (Graw 2009), muscle degeneration (Rajasekaran, Connell et al. 2007, Simon, Fontaine et al. 2007, Goldfarb, Olive et al. 2008, Tannous, Zhu et al. 2008, Willis, Schisler et al. 2009) and inherited neuropathies (Dierick, Irobi et al. 2005, Sun, Fontaine et al. 2010).

A current model for sHSP chaperone action (Figure 2), which has been developed almost entirely from in vitro studies, proposes that sHSP oligomers dissociate to dimers that form complexes with unfolding substrates (Giese and Vierling 2004, Gidalevitz, Kikis et al. 2010, Basha, O'Neill et al. 2012). Association with sHSPs limits protein aggregation and facilitates substrate delivery to ATP-dependent chaperones for refolding, and possibly to proteases for degradation (MacRae 2000, Haslbeck 2002, Basha, O'Neill et al. 2012). My research focuses on understanding the properties of potential substrates of a cyanobacterial sHSP, Hsp16.6 and provides a foundation for testing the current model of sHSP function in vivo.

Figure 2 A Proposed Model for the Chaperone Mechanism of sHSP/αC Proteins (Basha, O'Neill et al. 2012)

(i) Shows the oligomer of Hsp16.9 (1GME) dissociating into the dimer. sHSP/ αC oligomers are dynamic structures. (ii) Native substrates are denatured during heat stress. (iii) sHSP dimer, as a "holdase", forms a complex with unfolding substrates. (iv) sHSPs can coordinate with ATP-dependent HSP70 and help protein refolding. (v) Unfolding substrates are at a high risk of aggregating when not bound to sHSPs (vi) Substrates may also be delivered to proteases for degradation.



Synechocystis

The model organism used in my work is *Synechocystis sp.* PCC6803, a single-celled cyanobacterium. Cyanobacteria are oxygenic photosynthetic prokaryotes, a diverse group of organisms capable of living in a wide range of habitats where they are exposed to stresses such as high temperature, high light, high salinity, and lack of nutrients (Lee, Owen et al. 2000, Slabas, Suzuki et al. 2006). The whole genome of *Synechocystis*, 3725 genes in total, has been sequenced. *Synechocystis* can be transformed by homologous recombination (Kaneko, Sato et al. 1996) making it easy to perform gene deletions and replacements. A database containing all genomic information of *Synechocystis sp.* PCC6803 (Nakamura, Kaneko et al. 1998), CyanoBase, (http://genome.microbedb.jp/cyanobase/), has been valuable for my research. There is only a single sHSP in *Synechocystis*, Hsp16.6 (Giese and Vierling 2002). Hsp16.6 performs a protective role in the heat shock response and is required for the development of thermotolerance in *Synechocystis* (Lee, Prochaska et al. 1998, Lee, Owen et al. 2000). Deletion of Hsp16.6 makes cells heat sensitive, causing loss of viability (Giese and Vierling 2002).

Thesis Overview

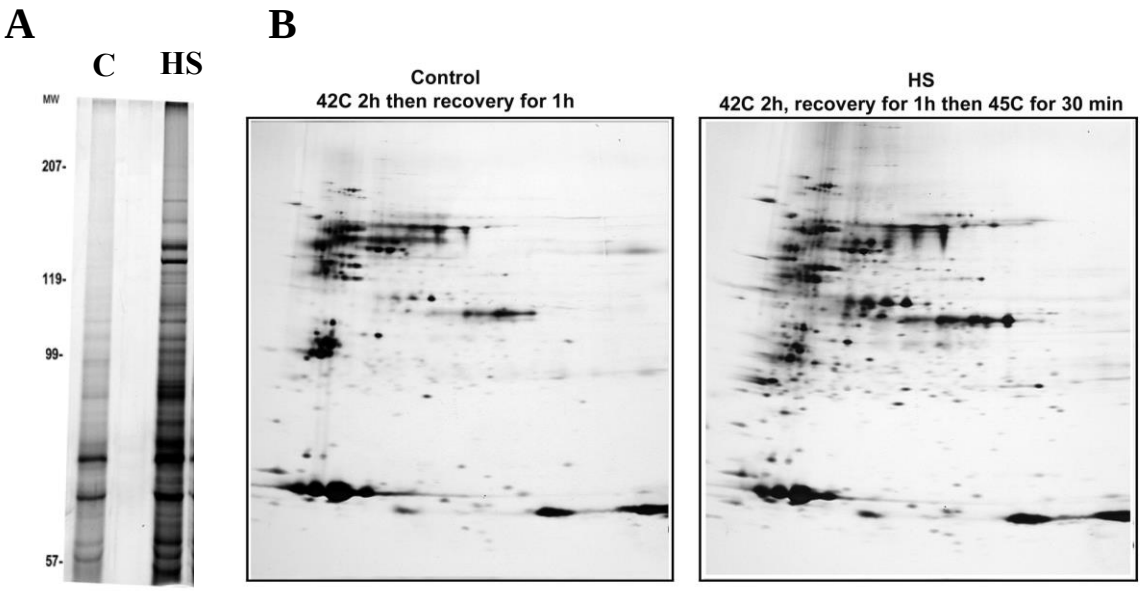
To identify Hsp16.6-associated proteins, which could represent substrates of this chaperone, a comparison of Hsp16.6 affinity pull downs from heat stressed wild-type and ΔHsp16.6 *Synechocystis* strains was carried out previously in the Vierling lab. Initial studies identified thirteen Hsp16.6-interacting proteins and one of them, serine esterase, proved to be heat sensitive in vitro and could be protected by Hsp16.6 (Basha, Lee et al. 2004). Additional experiments using the same approaches identified

a total of 84 potential Hsp16.6 substrates (Figure 3). This group of 84 proteins provided the starting point for my thesis research (Appendix I).

Chapter two uses bioinformatics and statistical analysis to determine the properties of proteins that were found associated with Hsp16.6 during heat stress. Chapter three goes on to test two of the putative substrates of Hsp16.6, FBA and EF-G1, for heat sensitivity and interaction with Hsp16.6, and determines the in vivo abundance of both proteins. The final chapter gives a conclusion of this thesis and provides a direction for future work.

Figure 3 1D and 2D-PAGE of Hsp16.6 Substrates

Hsp16.6 recovered in vivo. The *Synechocystis* cells carry a single copy of Hsp16.6 with a C-terminal Strep-tag that was integrated at the Hsp16.6 locus to replace the endogenous gene. After pre-treatment of a cell culture at 42 ℃, one half of the culture was left at 30 ℃ and the other half was heat stressed at 45 ℃ for 30min. Proteins associated with Hsp16.6 were then recovered by Strep-affinity chromatography and separated by 1D (A) and 2D-PAGE (B). Gels were stained with silver. The proteins present in the 45 ℃ treated sample, but absent in the control were isolated and identified by mass spectrometry. C: control; HS: heat shock.



CHAPTER 2

BIOINFORMATIC ANALYSIS OF HSP16.6-ASSOCIATED PROTEINS

Introduction

To understand which proteins are protected by sHSPs is critical, since it will provide an idea of how sHSPs protect cells from damage. An interesting question about Hsp16.6 is why it binds to certain proteins and potentially protects them from irreversible aggregation. One possibility is that Hsp16.6 binds and protects proteins with specific properties. A total of 84 Hsp16.6-associated proteins was identified by affinity isolation and mass spectrometry (unpublished), including 13 proteins from previous publication (Basha, Lee et al. 2004). For the following text, Hsp16.6-associated proteins will be referred to as Hsp16.6 "substrates" for convenience, even though they have not been shown directly to be Hsp16.6 "substrates".

To search for shared features of Hsp16.6 substrates, a primary characteristic to consider would be protein structure. However, the high-resolution structures of all the substrates are not currently available, and the actual structures adopted by substrates under stress conditions are not known. In the absence of these data, the formulas to calculate Molecular weight (MW), Isoelectric point (pI), the percentage of charged residues (including negatively charged residues and positively charged residues) and the percentage of hydrophobic residues can be used to determine if substrates show

commonality of these properties. Therefore, these properties will be discussed in this chapter. Furthermore, a IxI motif is conserved in the C-terminal region of the sHSPs and is invovled in sHSP oligomerization (Poulain, Gelly et al. 2010). The ACD of αB-crystallin could bind to a peptide with an IxI motif, indicating that the IxI motif of the C-terminal is a possible binding motif (Delbecq, Jehle et al. 2012). Additionally, from the crystal structure, a hydrophobic groove formed by β4-8 in the ACD is seen to be covered by the IxI motif from the C-terminus of another monomer, and this hydrophobic groove was suggested as one of the substrate binding sites on the sHSP (in addition to the N-terminal arm) (van Montfort, Basha et al. 2001). VQL is corresponding conserved motif in the C-terminal arm of Hsp16.6. It would be intriguing to see if Hsp16.6 substrates have relatively more VQL than others, driving them to bind to the Hsp16.6.

This chapter concludes that Hsp16.6 preferentially binds to proteins with higher MW, lower pI, higher percentage of charged residues (especially negatively charged) and lower percentage of hydrophobic residues compared to proteins encoded by the whole *Synechocystis* genome. Also, Hsp16.6 substrates have a slightly higher number of VQL motifs.

Methods

Bioinformatic and Statistical Analysis

All of the protein characteristics, MW, pI, percentage of charged residues (positive and negative residues), percentage of hydrophobic residues, and frequency of the tripeptide VQL for each protein were calculated by writing algorithms using Python (Python Software Foundation). All the scripts are included in Appendix II. The

Synechocystis proteome sequence data were downloaded from cyanobase (http://genome.microbedb.jp/cyanobase/). The format of the protein names and sequences was reorganized using Python programming, changing it from one line of the sequence name with the amino acid sequence in multiple lines, to one line of name with one line of sequence. The formula for calculating MW was the sum of the mean isotopic masses of amino acids in the protein and the mean isotopic mass of one water molecule, sourced from ExPASy (http://www.expasy.org). The formula for calculating pI was from Innovagen (http://www.innovagen.com). The net charge Z of a peptide at a specific pH was estimated by the following equation.

$$Z = \sum_i N_i \frac{10^{pKa_i}}{10^{pH} + 10^{pKa_i}} - \sum_j N_j \frac{10^{pH}}{10^{pH} + 10^{pKa_j}}$$

where N_i is the number, and pKa_i is the pKa values for the N-terminus and the side chains of arginine (Arg), lysine (Lys), and histidine (His). The symbol j stands for the C-terminus and the aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), tyrosine (Tyr) amino acids. The positive residues included Arg, His and Lys. The negative residues included Asp, Glu. The hydrophobic residues included Valine (Val), Isoleucine (Ile), Leucine (Leu), Methionine (Met), Phenylalanine (Phe), Tryptophan (Trp) and Cys. The pI calculated from Innovagen was slightly different from ExPAsy. For example, the pI of slr0244 was 4.93 from Innovagen, rather than 5.12 from ExPAsy. The formula was available from the Innovagen, but not ExPAsy, so the pI as derived from the Innovagen formula was used.

A side-by-side bar chart was prepared to compare the characteristics of the cutoff proteome (defined as explained in Results) and Hsp16.6 substrates by MATLAB (The Mathworks, Inc). All other statistical analysis was performed using Minitab16

(Minitab, Inc). A boxplot was used to compare the median of each characteristic. Unlike the effect that outliers have on the mean, the advantage of a boxplot is that it avoids the effect of outliers. Interval estimation was performed to gauge if there were any statistically significant differences between the mean of any two characteristics (Appendix III and IV).

Results

Hsp16.6 Prefers to Bind Substrate Proteins with Higher Molecular Weight (MW) and Higher Acidity

To investigate if Hsp16.6 substrates have any common characteristics, the MW and the pI of Hsp16.6 putative substrates were first compared to all predicted proteins of the *Synechocystis* genome, hereafter referred to as the *Synechocystis* proteome. Because substrate proteins were resolved by 2D-PAGE prior to mass spectrometry, the proteins that could be identified experimentally were limited to the MW range (10 to 200kDa) and the pI range (pH 4 to 9.5) of the gels. Therefore, before comparing substrate characteristics to the whole proteome, proteins outside these ranges of MW and pI were removed from the complete list of *Synechocystis* proteins, generating a new list of proteins, referred to as the "cutoff proteome". The cutoff proteome totally includes 2857 proteins. The MW and pI of the substrate proteins and proteins in the cutoff proteome protein were calculated using Python programming (see script in Appendix II) and analyzed using Minitab16. The shape of the distribution of MWs of the *Synechocystis* cutoff proteome and Hsp16.6 putative substrates appears very similar, but the overall distribution of Hsp16.6 substrates is shifted to the right compared to the cutoff proteome (Figure 4A). Boxplot analysis shows that the median MW of putative substrates (~51kDa) is higher than that of the cutoff proteome

(~32.7kDa) (Figure 4B). From the interval estimation (Table 1), the 99% confidence interval (CI) of the difference between the mean for the MW of the Hsp16.6 substrates and cutoff proteome is 9.8 to 28.8kDa and the p-value is less than 0.001, meaning that there is a statistically significant difference between the mean MW of Hsp16.6 substrates and the cutoff proteome. Therefore, Hsp16.6 appears more likely bind to substrates with a higher MW.

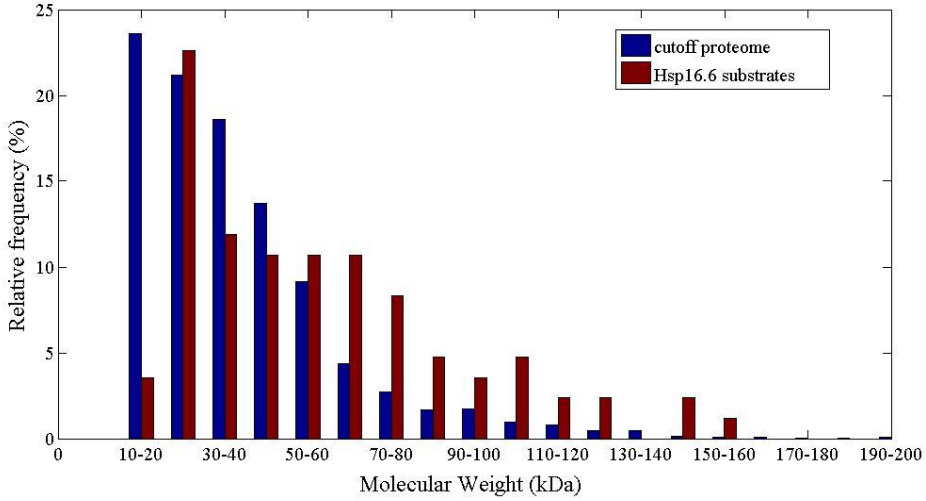
Unlike the distribution of the MW, the distribution of the pI of the cutoff proteome is bimodal. Based on the distribution (Figure 5A), it is hard to determine a difference between the pI of the cutoff proteome and Hsp16.6 substrates. According to the boxplot analysis (Figure 5B), the median pI of the cutoff proteome is 5.47, which is about 0.5 units lower than the median pI of Hsp16.6 substrates. Similarly, the 99% CI of the difference between the mean of the pI of the cutoff proteome and Hsp16.6 substrates is 0.46 to 1.00 units and the p-value is less than 0.001 (Table 1), suggesting that the mean pI of Hsp16.6 substrates is statistically significantly lower than that of the cutoff proteome mean. These results suggest that Hsp16.6 prefers to bind to more acidic proteins.

To further verify these results, a scatterplot of the MW vs pI was generated, comparing the cutoff proteome to the Hsp16.6 substrates (Figure 6). For the cutoff proteome, most proteins are found in the region with a MW of 20 to 75kDa and are spread from a pI 4 to 9.5. However, most Hsp16.6 substrates converge in an area with the MWs of 23kD to 125kD, and a pI of 4 to 6.5.

Figure 4 Comparison of the Molecular Weight (MW) Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the MW of the cutoff proteome and Hsp16.6 substrates. Blue bars: cutoff proteome; red bars: Hsp16.6 substrates. B Boxplot showing the difference of the median MW of the cutoff proteome and Hsp16.6 substrates. The values next to the box are the medians for each group.

A


cutoff proteome
Hsp16.6 substrates
Relative frequency (%)
25
20
15
10
5
0
0
10-20
30-40
50-60
70-80
90-100
110-120
130-140
150-160
170-180
190-200
Molecular Weight (kDa)

B

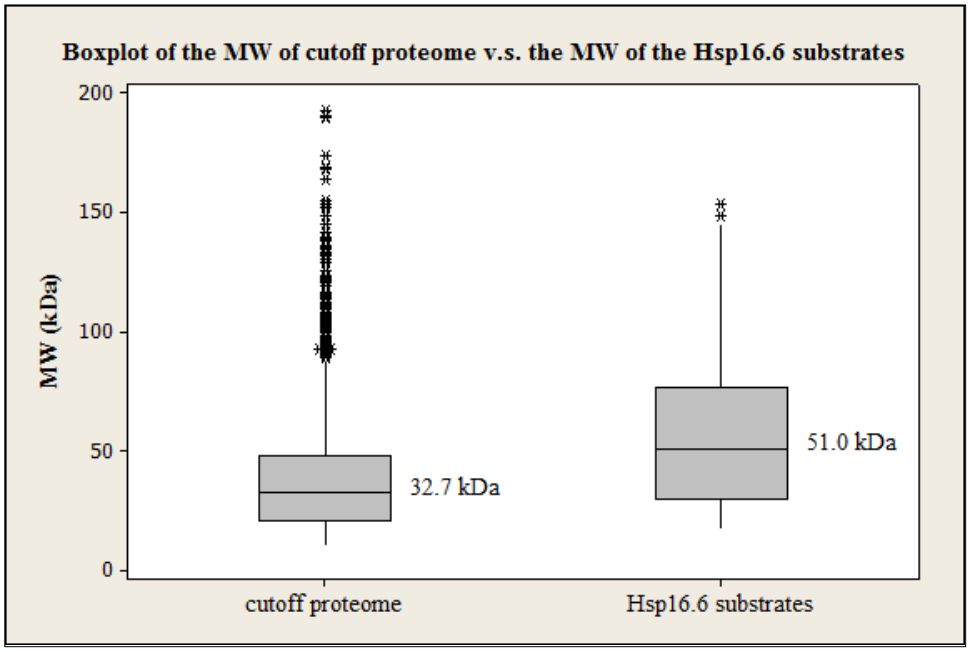

Boxplot of the MW of cutoff proteome v.s. the MW of the Hsp16.6 substrates
MW (kDa)
200
150
100
50
0
32.7 kDa
51.0 kDa
cutoff proteome
Hsp16.6 substrates

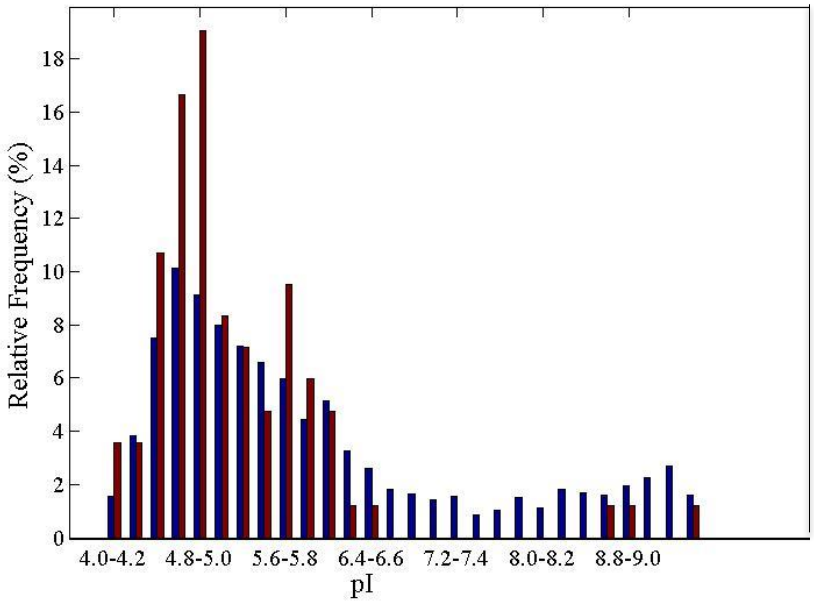
Table 1 Statistical Analysis of the Difference in Protein Characteristics Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

Protein characteristic	Sample	Mean	SD	Estimate for difference (cutoff proteome vs Hsp16.6 substrates)	99% Confidence interval (CI)	p-value	Relationship of substrates to cutoff proteome
Molecular weight (kDa)	Cutoff proteome	38.3	24.7	-19.3	-28.8 to -9.8	<0.001	substrates >cutoff proteome
	Substrates	57.7	32.7				
pI	Cutoff proteome	5.96	1.47	0.73	0.46 to 1.00	<0.001	substrates < cutoff proteome
	Substrates	5.24	0.92				
Charged residues (%)	Cutoff proteome	22.85	4.86	-2.24	-3.37 to -1.11	<0.001	substrates > proteome cutoff
	Substrates	25.09	3.85				
Negatively charged residues (%)	Cutoff proteome	11.49	2.92	-1.85	-2.59 to -1.11	<0.001	substrates > cutoff proteome
	Substrates	13.34	2.53				
Positively charged residues (%)	Cutoff proteome	11.36	2.64	-0.39	-0.98 to 0.20	0.085	Not significantly different at 0.01 significance level
	Substrates	11.76	2.01				
Hydrophobic residues (%)	Cutoff proteome	33.06	4.26	2.21	1.36 to 3.06	<0.001	substrates < cutoff proteome
	Substrates	30.85	2.86				

Figure 5 Comparison of the Isoelectric Point (pI) Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the pI of the cutoff proteome and Hsp16.6 substrates. Red bars: cutoff proteome; blue bars: Hsp16.6 substrates. B Boxplot showing the difference of the median pI of the cutoff proteome and Hsp16.6 substrates. The values next to the box are the medians for each group.

A


Relative Frequency (%)
0
2
4
6
8
10
12
14
16
18
4.0-4.2
4.8-5.0
5.6-5.8
6.4-6.6
7.2-7.4
8.0-8.2
8.8-9.0
pI

B

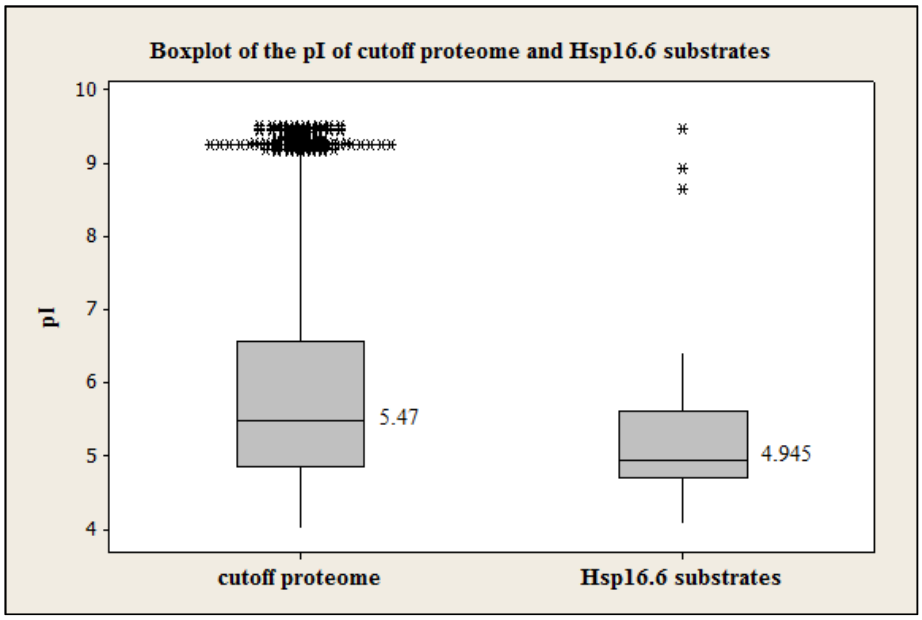
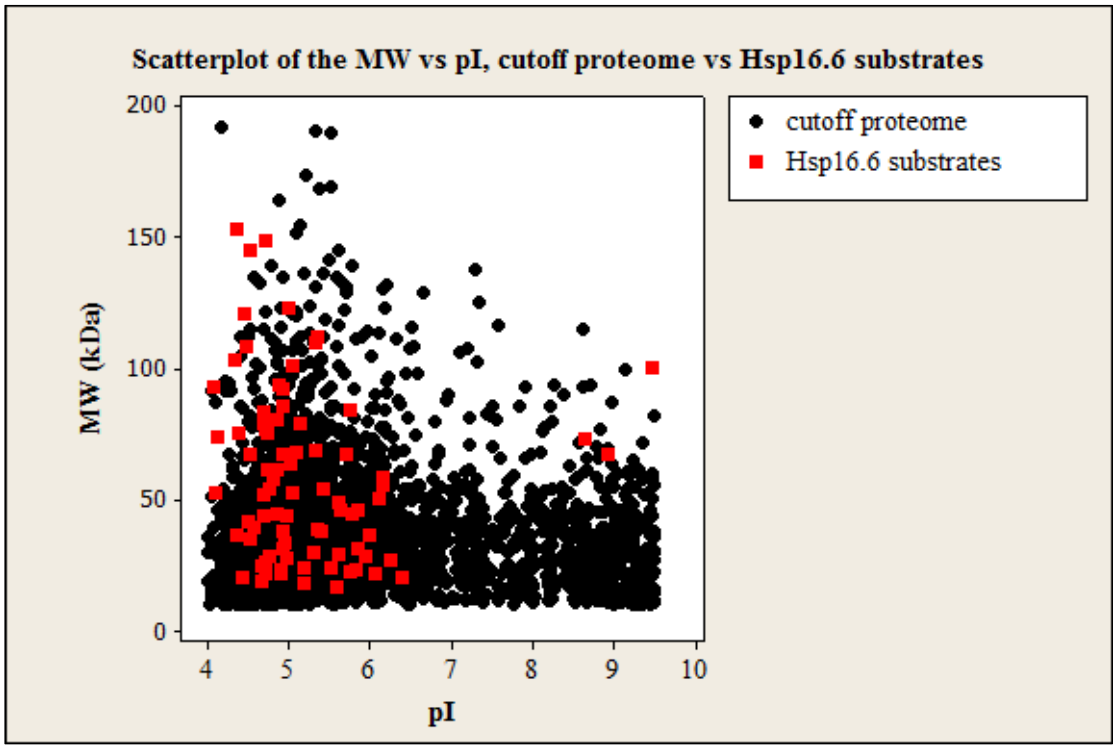

Boxplot of the pI of cutoff proteome and Hsp16.6 substrates
pI
10
9
8
7
6
5
4
5.47
4.945
cutoff proteome
Hsp16.6 substrates

Figure 6 Molecular Weight vs pI for the *Synechocystis* Cutoff Proteome vs Hsp16.6 Substrates

A scatterplot is profiled with the pI as the x-axis and MW as the y-axis. Black: cutoff proteome proteins; red dots: Hsp16.6 substrates.



Hsp16.6 Favors Binding Substrates with a Higher Percentage of Negatively Charged Residues and a Lower Percentage of Hydrophobic Residues

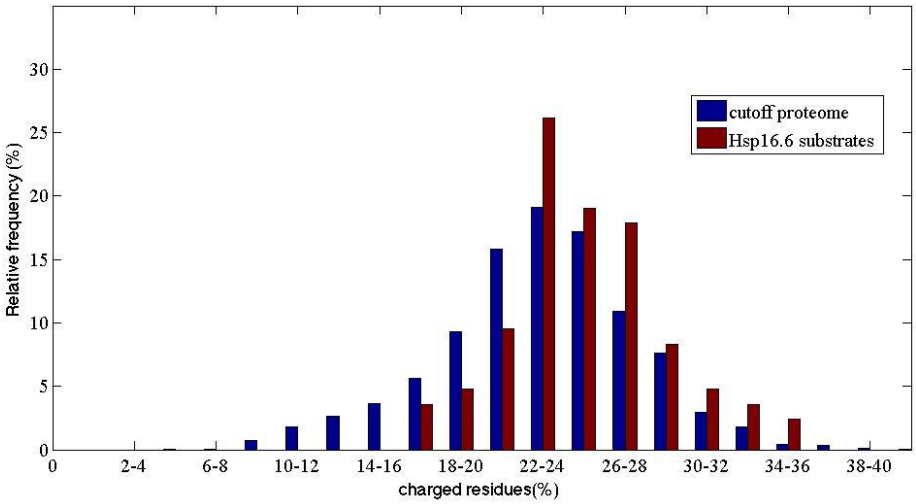
To understand other properties of proteins to which Hsp16.6 preferentially binds, the percentage of charged residues (Arg, His, Lys, Asp, Glu) and the percentage of hydrophobic residues (Val, Ile, Leu, Met, Phe, Trp, Cys) in each protein from the *Synechocystis* cutoff proteome and Hsp16.6 substrates were calculated using Python programming (Appendix II) and statistically analyzed with MATLAB and Minitab16, similar to MW and pI. Again, parameters from the cutoff proteome were compared to Hsp16.6 substrates. The percentage of charged residues from the cutoff proteome showed an approximately normal distribution (Figure 7A) and the distribution of Hsp16.6 substrates was shifted to the right of the cutoff proteome. From a boxplot analysis (Figure 7B), the median percentage of charged residues in Hsp16.6 substrates was 28.16%, which was larger than that from the cutoff proteome (27.35%). From interval estimation, the 99% CI for the difference between the percentage of charged residues from the cutoff proteome and Hsp16.6 putative substrates was -3.372% to -1.109% and the p-value is less than 0.001 (Table 1), suggesting that the mean percentage of charged residues in Hsp16.6 putative substrates was statistically significantly larger than the cutoff proteome.

To investigate whether positive residues (Arg, His, Lys) or negative residues (Asp, Glu) determine this difference, the same distribution, boxplot analysis and hypothesis testing were performed considering both types of residues. The results (Figure 8) showed that there was no statistically significant difference between the percentage of positive residues from the cutoff proteome and Hsp16.6 substrates, because the 99%

Figure 7 Comparison of the Percentage of Charged Residues Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the percentage of charged residues of the cutoff proteome and Hsp16.6 substrates. Red: cutoff proteome; blue: Hsp16.6 substrates. B Boxplot showing the difference of the median percentage of charged residues in the cutoff proteome and Hsp16.6 substrates. Values next to the box are the medians for each group.

A


Relative frequency (%)
charged residues(%)
cutoff proteome
Hsp16.6 substrates
0
5
10
15
20
25
30
0
2-4
6-8
10-12
14-16
18-20
22-24
26-28
30-32
34-36
38-40

B

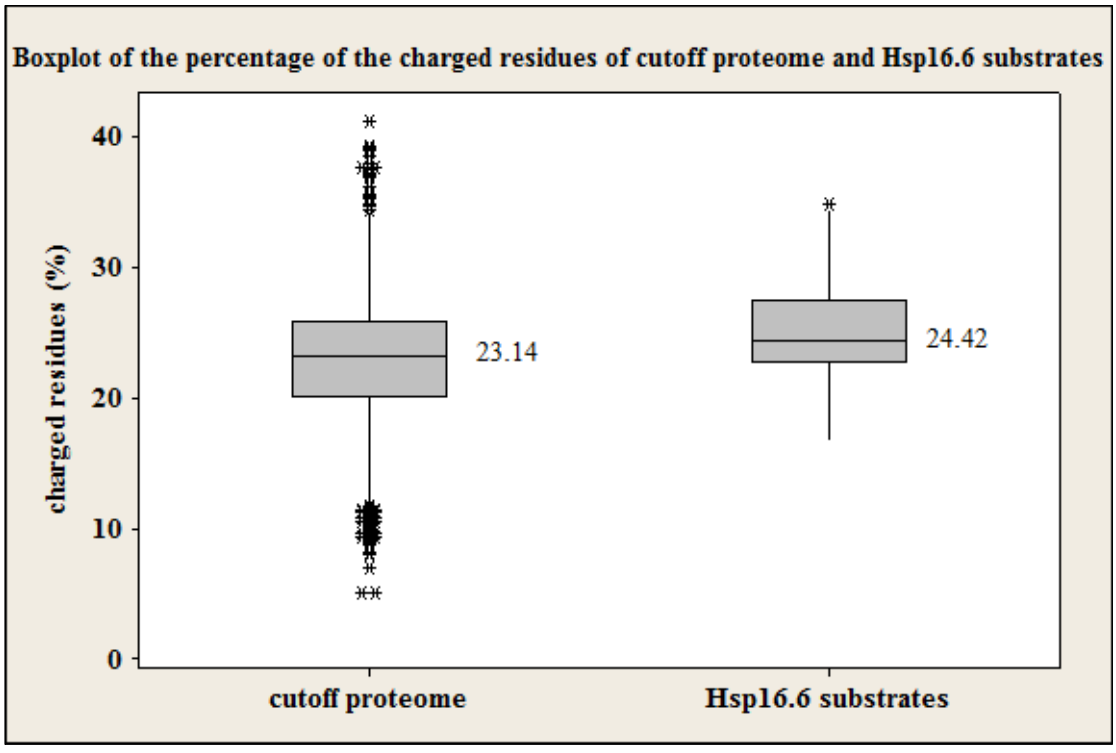
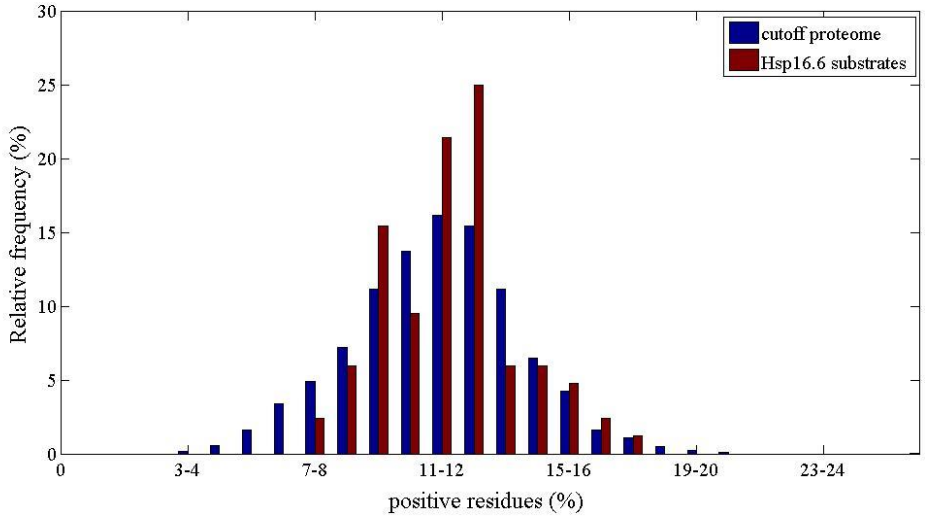

Boxplot of the percentage of the charged residues of cutoff proteome and Hsp16.6 substrates
charged residues (%)
0
10
20
30
40
23.14
24.42
cutoff proteome
Hsp16.6 substrates

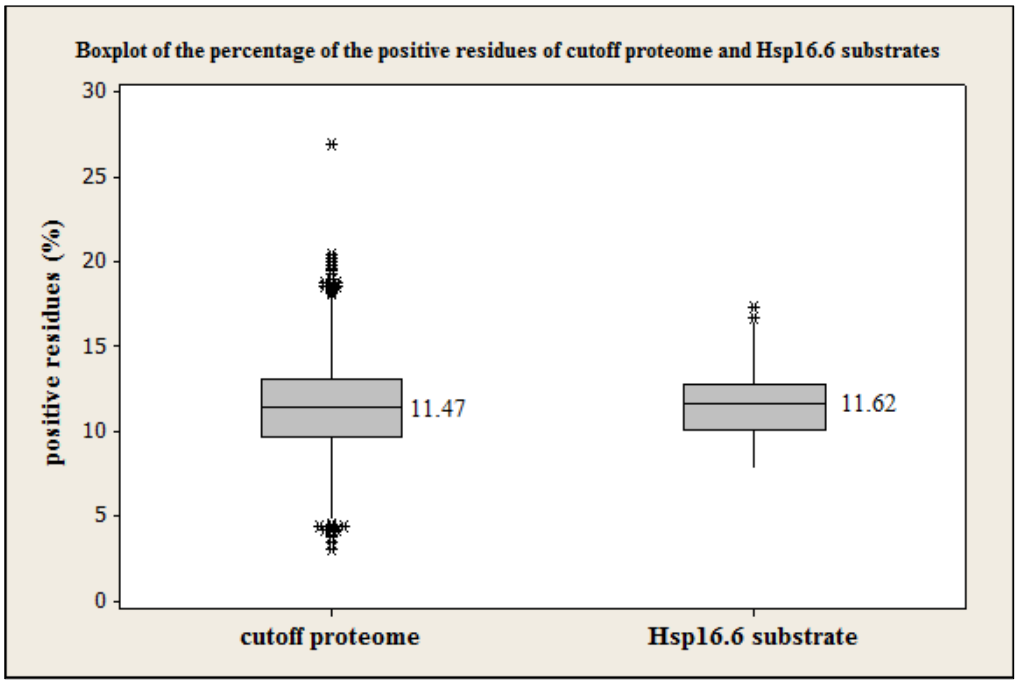
Figure 8 Comparison of the Percentage of Positively Charged Residues Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the percentage of positively charged residues of the cutoff proteome and Hsp16.6 substrates. Red: cutoff proteome; blue: Hsp16.6 substrates. B Boxplot showing the difference of the median percentage of positively charged residues of the cutoff proteome and Hsp16.6 substrates. Values next to the box are the medians for each group.

A


cutoff proteome
Hsp16.6 substrates
Relative frequency (%)
positive residues (%)
0
3-4
7-8
11-12
15-16
19-20
23-24
30
25
20
15
10
5
0

B


Boxplot of the percentage of the positive residues of cutoff proteome and Hsp16.6 substrates
positive residues (%)
30
25
20
15
10
5
0
11.47
11.62
cutoff proteome
Hsp16.6 substrate

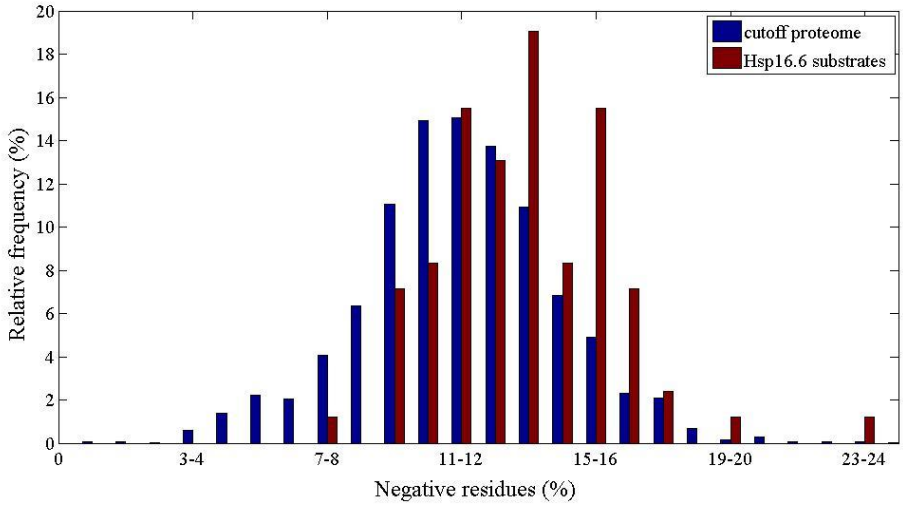
CI, -0.98% to 0.20% (Table 1), included zero and the p-value was 0.085. Nevertheless, from the interval estimation (Table 1), the 99% CI for the difference between the mean percentage of negative residues from the cutoff proteome and Hsp16.6 substrates was -2.59% to -1.11% and the p-value was less than 0.001, indicating that the mean of the percentage of negative residues in Hsp16.6 substrates was much higher than that from the cutoff proteome. Also, the median percentage of negative residues in Hsp16.6 substrates was 16.36%, which is greater than that of the cutoff proteome (Figure 9B). This result is consistent with the analysis showing that Hsp16.6 interacts with proteins that are more acidic (Figure 9). All of these data indicate that Hsp16.6 preferentially binds to putative substrates with a relatively higher percentage of charged residues, especially negative residues.

The next property examined was the enrichment of hydrophobic residues. From the bar chart in Fig 10A, the percentage of hydrophobic residues was almost normally distributed. The median percentage of hydrophobic residues of the cutoff proteome was 32.46%, which was greater than 31.01%, the median percentage of hydrophobic residues in Hsp16.6 substrates. Furthermore, from the interval estimation, the 99% CI for the difference between the mean percentage of hydrophobic residues in the cutoff proteome versus Hsp16.6 substrates was 1.362% to 3.055%. To clearly locate where the group of Hsp16.6 substrates fall among the proteins of the cutoff proteome, a scatterplot of the percentage of charged residues (y-axis) and the percentage of hydrophobic residues (x-axis) was generated (Figure 11). Hsp16.6 substrates (red) occupy the upper-middle part in the distribution relative to the cutoff proteome (black). All of this analysis suggests that Hsp16.6 favors binding putative substrates with a lower percentage of hydrophobic residues.

Figure 9 Comparison of the Percentage of Negatively Charged Residues Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the percentage of the negatively charged residues in the cutoff proteome and Hsp16.6 substrates. Red bars: cutoff proteome; blue: Hsp16.6 substrates. B Boxplot showing the difference of the median percentage of negatively charged residues in the cutoff proteome and the Hsp16.6 substrates. Values next to the box are the medians for each group.

A


cutoff proteome
Hsp16.6 substrates
Relative frequency (%)
Negative residues (%)
0
3-4
7-8
11-12
15-16
19-20
23-24

B

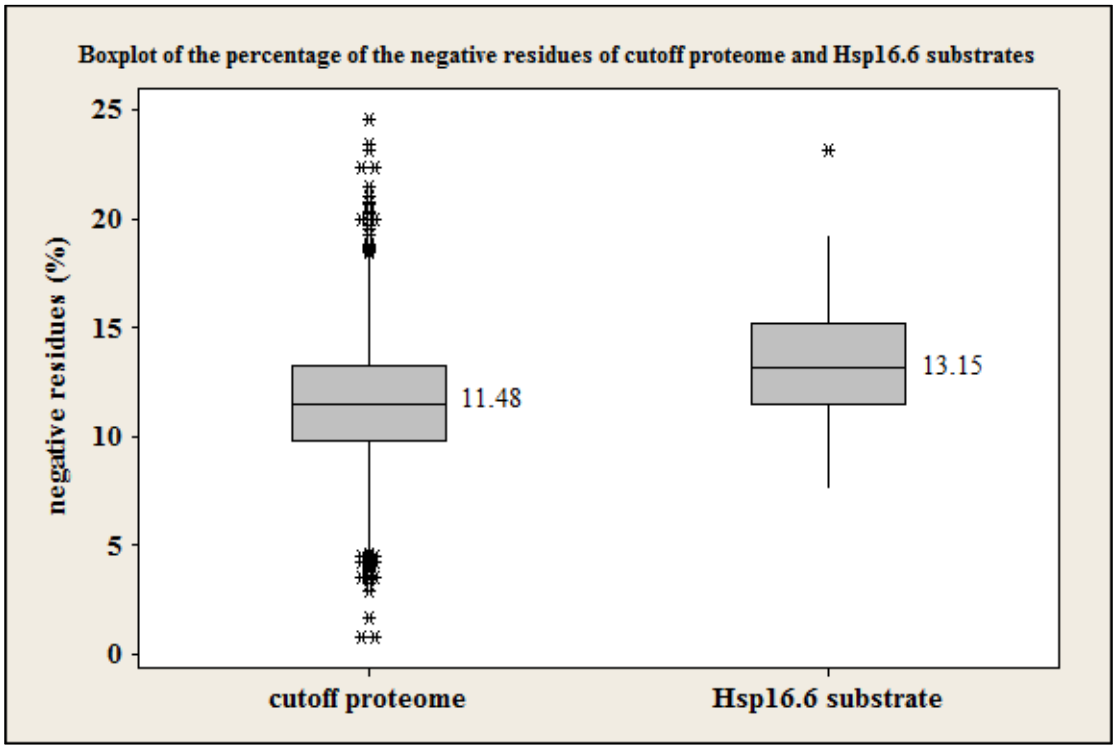
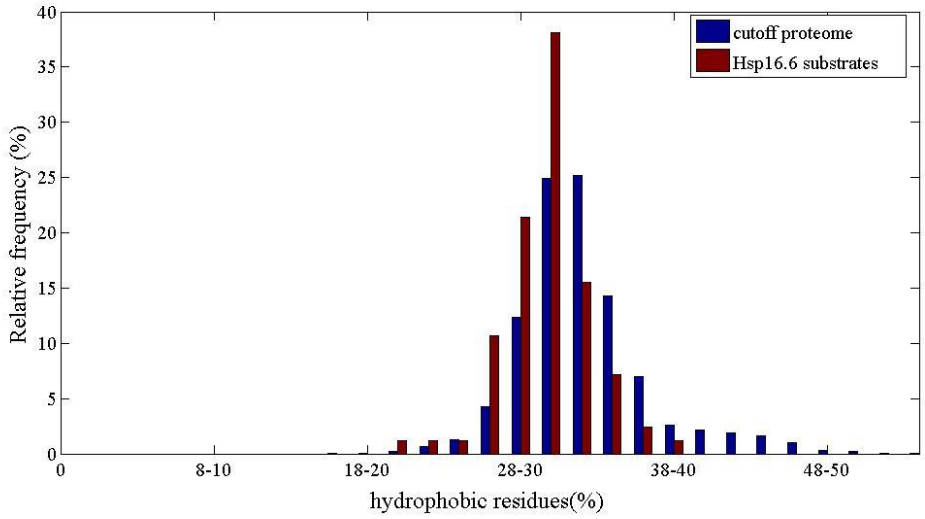

Boxplot of the percentage of the negative residues of cutoff proteome and Hsp16.6 substrates
negative residues (%)
11.48
13.15
cutoff proteome
Hsp16.6 substrate

Figure 10 Comparison of the Percentage of Hydrophobic Residues Between the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

A Distribution of the percentage of hydrophobic residues in the cutoff proteome and Hsp16.6 substrates. Red: cutoff proteome; blue: Hsp16.6 substrates. B Boxplot showing the difference of the median percentage of hydrophobic residues in the cutoff proteome and Hsp16.6 substrates. Values next to the box are the medians for each group.

A


cutoff proteome
Hsp16.6 substrates
Relative frequency (%)
0
5
10
15
20
25
30
35
40
0
8-10
18-20
28-30
38-40
48-50
hydrophobic residues(%)

B

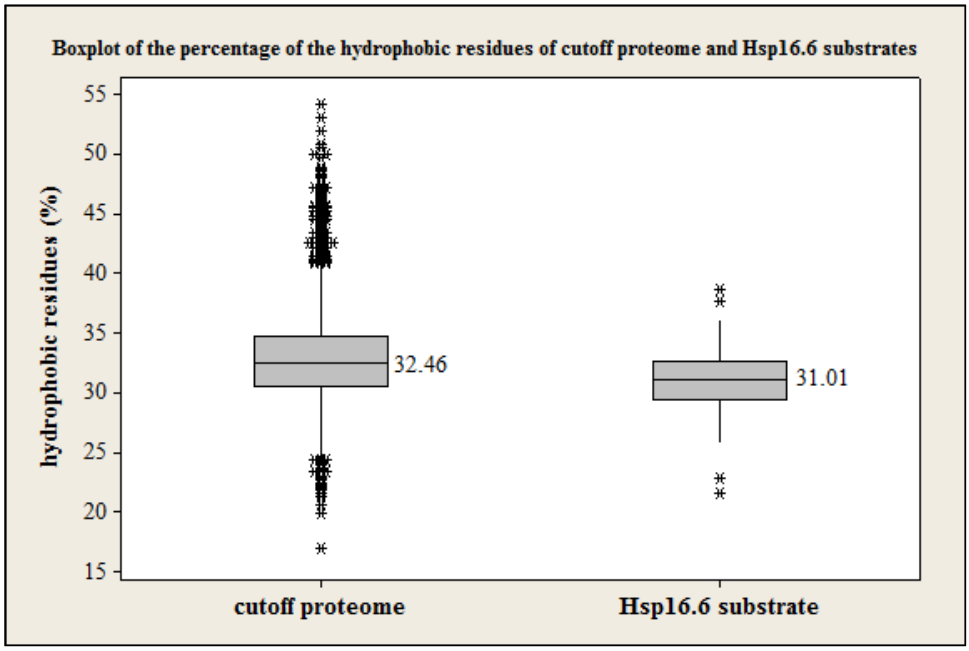
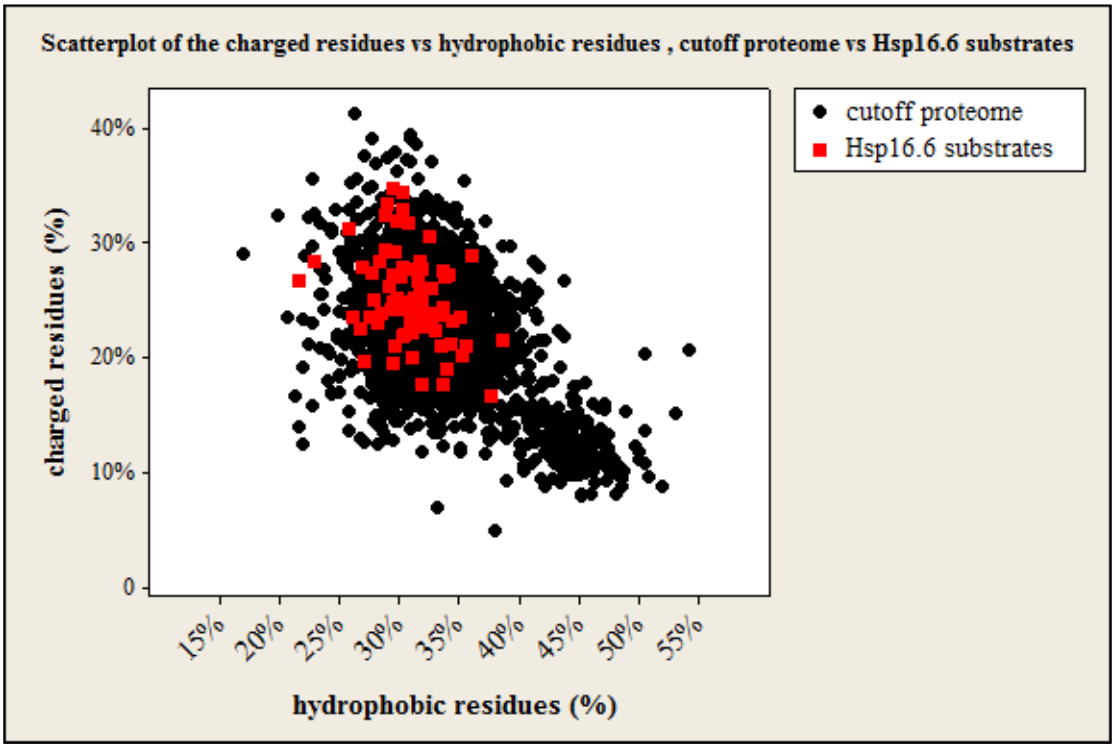

Boxplot of the percentage of the hydrophobic residues of cutoff proteome and Hsp16.6 substrates
hydrophobic residues (%)
15
20
25
30
35
40
45
50
55
32.46
31.01
cutoff proteome
Hsp16.6 substrate

Figure 11 Charged Residues vs Hydrophobic Residues for the *Synechocystis* Cutoff Proteome vs Hsp16.6 Substrates

A Scatterplot with the percentage of hydrophobic residues as the x-axis, and the percentage of charged residues as the y-axis. Black: cutoff proteome; red: Hsp16.6 substrates.



VQL Motif Analysis

To explore the role the VQL motif might play in the recognition of substrates by Hsp16.6, the number of times the amino acid motif VQL appeared in each of the Hsp16.6 substrates and proteins of the *Synechocystis* cutoff proteome were calculated using Python programming. To determine if Hsp16.6 substrates have overall more VQL motifs than proteins of the *Synechocystis* cutoff proteome, interval estimation was performed to check if Hsp16.6 substrates have a higher proportion of proteins with more than zero VQL motifs (Table 2). From the table, the 90% CI for the difference in the proportion of proteins which have more than zero VQL motifs was -0.1358 to -0.0008, with a p-value of 0.096. Even though the 90% CI and p-value did not support a strong difference, Hsp16.6 substrates still have a statistically somewhat higher proportion of proteins having more than zero VQL motifs compared to the *Synechocystis* cutoff proteome. However, all Hsp16.6 substrates have no more than one VQL motif while the cutoff proteome has 15 proteins containing more than one VQL motif (Table 2).

To get an idea of the frequency of occurrence of VQL motifs, the total length of the proteins in number of amino acids was divided by the number of VQL motifs (denoted as D(Hsp16.6 substrates), D(cutoff proteome)). D(Hsp16.6 substrates) was approximately equal to 3144 amino acids and D(cutoff proteome) was near 3304 amino acids, meaning that Hsp16.6 substrates have a slightly higher density of VQL motifs.

Table 2 Statistical Analysis of VQL Motifs in the *Synechocystis* Cutoff Proteome and Hsp16.6 Substrates

	Number proteins with more than zero VQL motifs	Total number of proteins	Proportion (Number proteins with more than zero VQL motifs/total number of proteins)	Number proteins which have more than one VQL motif	Total number of proteins	Proportion (Number proteins with more than one VQL motif/total number of proteins)
Cutoff proteome (1)	281	2857	0.0984	15	2857	0.005250
Hsp16.6 substrates (2)	14	84	0.1667	0	84	0
Estimate for the difference p(1)-p(2)	-0.0683			N/A		
90% CI for the difference	-0.1358 to -0.0008					
p-value	0.096					
Relationship	Hsp16.6 substrates have higher proportion of VQL motif than cutoff proteome					

The Synechocystis Cutoff Proteome vs the Total Proteome

To rule out the possibility that generating the cutoff proteome introduced a bias in the analysis leading to the difference in the protein properties observed between the cutoff proteome and Hsp16.6 substrates, all sequence properties of the total proteome were also calculated and compared to the cutoff proteome (Table 3). A p-value 0.05 was used as the default significance level in this analysis. Considering MW as an example, the 95% CI for the difference between the proteome and the cutoff proteome was ~-5.0kDa to ~-2.4kDa and the p-value was less than 0.001, indicating that the mean MW of the proteome was significantly less than the mean MW of the cutoff proteome. Together with the results of Hsp16.6 substrates and cutoff proteome, the relationship among these three was proteome < cutoff proteome < Hsp16.6 substrates.

Similarly, for the other properties, the cutoff proteome was always in the middle of the three values regardless of the difference between the proteome and the cutoff proteome. Thus, restricting the comparison of the substrates to the cutoff proteome only decreased the difference between the substrates and the total proteome, meaning that the "cutoff" did not create the statistically significantly difference between the cutoff proteome and Hsp16.6 substrates.

Discussion

The bioinformatic and statistical analysis suggest that Hsp16.6 has the propensity to bind to proteins with higher MW, lower pI, abundance of charged residues (especially negatively charged residues) and a low percentage of hydrophobic residues compared to the average protein encoded by the *Synechocystis* genome. It is not clear if proteins

Table 3 Statistical Analysis of the Difference Between Protein Characteristics of the *Synechocystis* Whole Proteome and the Cutoff Proteome

	Estimate for difference (proteome – cutoff proteome)	95% CI	p-value	Relationship
Molecular weight (kDa)	-3.7	-5.0 to -2.4	<0.001	Proteome < cutoff proteome < Hsp16.6 substrates
pI	0.57	0.49 to 0.66	<0.001	Hsp16.6 substrates < Cutoff proteome < proteome
charged residues (%)	0.14	-0.12 to 0.40	0.293	Proteome ≈ cutoff proteome < Hsp16.6 substrates
negatively charged residues (%)	-0.41	-0.59 to 0.23	<0.001	Proteome < cutoff proteome < Hsp16.6 substrates
positively charged residues (%)	0.50	0.35 to 0.66	<0.001	Proteome > cutoff proteome ≈ Hsp16.6 substrates
hydrophobic residues (%)	0.25	0.02 to 0.48	0.033	Proteome > cutoff proteome > Hsp16.6 substrates
Proportion of proteins which have VQL motif (more than zero)	-0.0145	-0.0286 to -0.0003	0.045	Proteome < cutoff proteome < Hsp16.6 substrates

with those features are more likely to be aggregated during heat stress or if Hsp16.6 binds proteins with these properties no matter if they aggregate during heat stress. It is possible that larger and more acidic proteins are more likely to aggregate during heat stress. Also, potentially it is changes in ionic interactions, not hydrophobic interactions that drive the aggregation of proteins during heat stress.

Notably, the statistical analysis of substrate characteristics in this thesis has limitations. The proteome used for the analysis was assumed to be the whole proteome, even though the proteome was derived from the annotated open reading frames of the sequenced genome, which may not have correctly identified all protein coding genes. However, the proteome in this thesis could be considered as a random sample of the actual proteome, which would add validity to the statistical analysis. Also, the interval estimation for statistically significant differences was based on the assumption that the 84 Hsp16.6 substrates were randomly selected from the actual Hsp16.6 substrates. Comparing the substrates to the cutoff proteome attempted to correct for the limitation of the experimental techniques used to identify substrates with regard to pI and MW, but another limitation of the experimental technique that could not be corrected for was protein abundance. The affinity and gel electrophoresis approach used for substrate identification was limited to identifying more abundant proteins in *Synechocystis* cells, and for this reason the substrates may not represent a completely random sample of substrates. Unfortunately, it is not possible to compare the isolated substrates to only the more abundant *Synechocystis* proteins, as information on protein abundance is not available for the whole proteome. Therefore, the relative properties of substrates reported here, could by biased due to the inclusion of low abundance proteins in the cutoff proteome, which were not sampled by the

experimental method used to recover substrates. The buffer used to lyse the *Synechocystis* cells for analysis of substrates, may also have precluded identification of any intrinsic membrane proteins with which Hsp16.6 might associate. Again, no attempt was made to remove intrinsic membrane proteins from the cutoff proteome; the extent to which this type of bias may contribute to the apparent lower hydrophobicity of the substrates compared to the proteome cannot be determined.

Previous experiments have examined the properties of proteins identified as interacting with chaperones as potential substrates, including sHSPs from other bacteria. GroEL in *E.coli* does not have a bias for binding proteins with a specific range of isoelectric points (pI), but most GroEL substrates are larger than 20kD (Houry, Frishman et al. 1999). In addition, GroEL substrates are less hydrophobic compared to GroEL independent proteins (Raineri, Ribeca et al. 2010, Azia, Unger et al. 2012). Both IbpB (a sHSP of *E.coli*) and Hsp20.2 (a sHSP of *Deinococcus radiodurans*) were proposed to be more likely to bind substrates with high MW, moderate acidity, abundant charged residues, but not hydrophobic residues (Fu, Chang et al. 2013). These conclusions are more or less consistent with my observations. However, these previous results have limitations. The 2D-gel used for obtaining GroEL substrates only identified proteins with MW 10-100kDa and pI 4-9. It makes less sense to compare the potential substrates with total soluble cytoplasmic proteins because there might be substrates out of the range sampled. IbpB and Hsp20.2 were concluded to prefer proteins with properties in specific ranges. Fu. *et al* calculated the p-value for each difference interval. For example, IbpB was reported to prefer proteins with 11-13% and 14-15% of positively charged residues at $p<0.05$. But, when the interval is changed, the p-value might change. So, I believe it is hard to

compare the two groups on an individual interval. The authors used 1D SDS-PAGE to separate proteins that were isolated with His-tagged IbpB in which benzoylphenylalanien (Bpa) replaced Phe16. Cells carrying the modified IbpB were heat stressed at 50℃ for 30 min and then in vivo photocrosslinking on ice, followed by lysis with urea denaturation. Recovered proteins were compared to a control of proteins recovered by the same methods without the photocrosslinking step. Mass spectrometry was carried out on all proteins using gel slices and only proteins not detected in the control were analyzed as substrates. Not surprisingly the spectrum of proteins detected was also found to be biased towards abundant *E.coli* proteins (supplemental data), a limitation of all current biochemical methods for sHSP substrate identification. In addition, it is possible that the presence of the His-tag or the Bpa substitution at Phe16 produced artifactual interactions with non-substrate proteins or biased the results to specific substrates binding that part of the sHSP. The severe heat stress conditions may also have an effect, as well as the position of the photocrosslinking residue. The identification of *Deinococcus radiodurans* substrates used immunoprecipitation, but of Hsp20.2 added to cell lysates that were then heated to drive interaction with the sHSP. How this would compare to in vivo conditions is not clear. Nonetheless, it is interesting that three different approaches with three different organisms led to some similar conclusions regarding potential sHSP substrates.

Moreover, the overlap among Hsp16.6, IbpB and Hsp20.2 substrates were also analyzed (Appendix V). Only seven of the 84 Hsp16.6 substrates have a homolog in the list of both 145 IbpB and 118 Hsp20.2 substrates. The function of most of these proteins is related to transcription. Remarkably, FBA is also found in this group.

Thirteen and seventeen Hsp16.6 substrates, respectively, share a homolog with IbpB substrates and Hsp20.2 substrates. These proteins span a wide range of functional categories.

From the VQL analysis, Hsp16.6 substrates have a slightly, but not very significantly, higher frequency of VQL motifs than the *Synechocystis* proteome. However, it is still the case that the majority of substrates identified have no VQL motifs (70 of 84). This observation suggests that the hydrophobic interaction between IxI motifs and β4-8 groove is just one of the factors that determines the substrates of Hsp16.6. But, intriguingly, when Valine was substituted by Alanine in the motif, the frequency of occurrence of AQL motifs for Hsp16.6 substrates was one per 1914 amino acids, which is lower than one per 1706 amino acids calculated for the cutoff proteome. Similarly, the frequency of GQL motifs in Hsp16.6 substrates was one per 2751 amino acids, which is also lower than one per 1706 amino acids calculated for the cutoff proteome. The same results apply to VNL, LQL, IQL and VQA motifs. This might indicate that VQL motifs are an important substrate interaction motif on some Hsp16.6 substrates. Interestingly, none of the Hsp16.6 substrates have more than one VQL motif. One of the possible reasons is that duplicate VQL motifs might negatively affect the possibility of substrates being recognized by Hsp16.6.

To analyze Hsp16.6 substrates beyond the sequence based analysis, a spreadsheet of the 84 Hsp16.6 substrates with different categories is shown in Appendix VI. Since the categories were obtained from different websites and no simple algorithm was available, it is less possible to make the same table for the entire *Synechocystis* proteome. According to the table, 13 out of 84 substrates have predicted membrane

regions (http://www.cbs.dtu.dk/services/TMHMM-2.0/), 5 out of 84 substrates have an available high-resolution structure in Protein Data Bank (PDB) (PDB access number is displayed), 20 out of 84 substrates were identified as soluble proteins in pH 4.5-5.5 ("yes" in the cells means the proteins were identified as soluble, and "no" means that the proteins were not identified as soluble, although this does not indicate that the proteins are not soluble) (Simon, Hall et al. 2002), and 21 out of 84 substrates were up-regulated during heat shock ("yes" in the cells means the proteins were identified as up-regulated during heat shock, and "no" means that the proteins were not identified as up-regulated, which does not necessarily mean that the proteins are not up-regulated during heat shock) (Slabas, Suzuki et al. 2006). Substrates, which bind to nucleotide phosphate, are also indicated in AppendixVI. According to the functional categories, substrates are basically related to physiological growth, amino acid synthesis, protein modification and degradation.

These bioinformatic and statistical analyses provide a general idea of what kind of proteins sHSPs may recognize and bind. The data may help people develop hypotheses and discover how sHSPs select substrates.

CHAPTER 3

BIOCHEMICAL ANALYSIS OF TWO PUTATIVE HSP16.6 SUBSTRATES

Introduction

From the previous analysis of the proteins of the *Synechocystis* proteome and Hsp16.6 substrates, two proteins were chosen as candidates to examine if their properties and behavior in vitro and in vivo are consistent with model for Hsp16.6 protection of substrates. To select candidate proteins for further testing, properties listed in Appendix VI were considered, as was the previous bioinformatics analysis.

Based on this analysis, two proteins were chosen as candidate substrates: fructose-1, 6-bisphosphate aldolase class II (FBA) (Cyanobase accession number: sll0018) and elongation factor-G 1 (EF-G1) (Cyanobase accession number: slr1463). FBA has a lower MW (38.9kDa), higher pI (5.46), 25.06% charged residues, 27.86% hydrophobic residues and no VQL motif. EF-G1 has a higher MW (76.7kDa), lower pI (4.71), 29.21% charged residues, 29.64% hydrophobic residues and one VQL motif.

FBA catalyzes a reversible reaction, cleaving fructose-1, 6-bisphosphate (FBP) into dihydroxyacetonephosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) (Gefflaut, Blonski et al. 1995, Nakahara, Yamamoto et al. 2003). FBA enzymes are divided into two groups, class-I FBA and class-II FBA, according to the organism in which they are found and their catalytic mechanism (Rutter 1964, Gefflaut, Blonski et

al. 1995). Class-I FBA forms a Schiff base with DHAP and FBP (Rutter 1964, Gefflaut, Blonski et al. 1995, Nakahara, Yamamoto et al. 2003). Class-II FBA requires divalent cations for activity (Rutter 1964, Gefflaut, Blonski et al. 1995, Nakahara, Yamamoto et al. 2003). Although the *Synechocystis* genome encodes both a Class-I and Class-II FBA, only the Class-II FBA was recovered in association with Hsp16.6; therefore the FBA discussed in the following text refers to the Class-II type, divalent cation-dependent FBA.

EF-G is a GTPase that promotes translocation of the peptidyl-tRNA during mRNA translation (Green 2000, Kojima, Motohashi et al. 2009). The EF-G1 corresponding to slr1463 will be discussed in this paper. Based on phylogenetic analysis, *Synechocystis* EF-G1 is more closely related to chloroplast EF-G, than are two other EF-G homologs (sll1098 and sll0830) of EF-G in *Synechocystis* (Kojima, Oshita et al. 2007, Kojima, Motohashi et al. 2009).

It was previously shown that citrate synthase (CS) and malate dehydrogenase (MDH) thermally aggregate at 45°C (Lee, Pokala et al. 1995, Lee, Roseman et al. 1997), and that Hsp16.6 can protect both of these proteins from irreversible aggregation in vitro (Basha, Lee et al. 2004). For the identification of Hsp16.6 substrates during heat stress in vivo, *Synechocystis*, which normally grows at 30 ℃, was first pretreated at 42 ℃ for 2 hrs to allow the accumulation of Hsp16.6. After 12 hrs of recovery at 30 ℃, cells were heated at 46 ℃ for 0-20 min prior to immunoprecipitation or affinity chromatography (Basha, Lee et al. 2004) (E. Basha unpublished). Based on this method of recovering Hsp16.6 substrates and the proposed mechanism of sHSP

function (Basha, O'Neill et al. 2012), I expect that a true substrate would be heat sensitive (at 46℃ or lower) and be protected by Hsp16.6 from irreversible aggregation by forming a complex with the substrate that is large and heterogeneous.

In vitro and in vivo properties of FBA and EF-G1 with Hsp16.6 during heat stress will be discussed in this chapter. FBA was expressed in *E.coli* and purified. FBA could be protected by Hsp16.6 from aggregation by forming a complex with Hsp16.6 during heat stress in vitro, consistent with it being a substrate of Hsp16.6. EF-G1 was also expressed in *E.coli* and purified. EF-G1 did not form insoluble aggregates even at 47℃, but circular dichroism spectroscopy revealed the secondary structure has melted at this temperature and the protein eluted earlier than unheated protein on size exclusion chromatography. Thus, EF-G1 appears heat sensitive, and may also be an in vivo substrate of Hsp16.6. At last, in vivo study studies were performed to determine the amount of FBA and EF-G1 in *Synechocystis* cells. Both proteins are abundant, with FBA levels (around 2% of total cell protein) being about twice that of EF-G1. Further in vivo experiments will be needed to confirm that FBA and EF-G1 are substrates of Hsp16.6.

Methods

Synechocystis Strains and Growth Conditions

The single-celled freshwater cyanobacterium *Synechocystis sp.* PCC 6803 was used as a source of DNA for cloning and for in vivo experiments. A wild-type strain in which a spectinomycin resistance gene had been inserted next to the Hsp16.6 locus (Cyanobase access number: sll1514) was created as described previously (Torok,

Goloubinoff et al. 2001, Giese and Vierling 2002). Cells were shaken in 10mM HEPES (pH 7.8)-buffered liquid BG-11 amended with 5mM glucose at 30°C under illumination at around 40μmol photons $m^{-2}s^{-1}$ under fluorescent lamps (Nakahara, Yamamoto et al. 2003, Basha, Lee et al. 2004). When cells were used to extract genomic DNA, the volume was 200ml (OD ~2.0 per ml at 730nm). When cells were prepared for measuring the FBA and EF-G1 amount, the volume was 5ml (OD ~2.0 at 730nm). Shaking speed was 170rpm.

Gel Electrophoresis

SDS-PAGE was performed following standard protocols. 15% acrylamide gels were used for assays with FBA, since FBA is 38.9kDa and Hsp16.6 is 16.6kDa. 10% acrylamide gels were used for assays of EF-G1 since EF-G1 is 76.7kDa.

Cloning and Purification of FBA and EF-G1

The plasmid containing *Synechocystis* FBA (sll0018) was obtained from a previous lab member. The backbone of the plasmid was pJC20 (Clos and Brandau 1994). Recombinant, untagged FBA was expressed in BL21 *E.coli* cells and purified following the general procedure published previously (Nakahara, Yamamoto et al. 2003). Basically, the supernatant prepared from lysed *E.coli* cells expressing FBA was brought to 30% saturation by adding solid ammonium sulfate, and the supernatant recovered after the precipitation was separated by butyl-Toyopearl chromatography (Tohsoh, Japan) with a gradient of 30% to 0% saturation of $(NH_4)_2SO_4$ in buffer (20mM potassium phosphate buffer pH7.5). The FBA concentration was determined using a calculated extinction coefficient (25820 $cm^{-1}M^{-}$

[1]). The mass of FBA was confirmed by MALDI (UMASS Mass spec. center). The results are available in Appendix VII.

To obtain recombinant EF-G1 it was first necessary to purify *Synechocystis* genomic DNA. 200ml of *Synechocystis* cells (OD ~ 2.0 at 730nm) were collected by centrifugation at 6000rpm for 10min. Total whole genomic DNA was extracted by treatment of the cell pellet following the procedures of the DNeasy Plant kit (QIAGEN). The coding region of EF-G1 (slr1463) was amplified from *Synechocystis* genomic DNA using a forward primer for polymerase chain reaction (PCR) with the sequence 5'-TCTGCCGGCGGCATGGAAAAAG-3' and the reverse primer 5'-AAGCTCGAGTTAAGCAGCGGCTTG-3'. Herculase II polymerase (Agilent) was used for the first round of PCR, and then the same primers, but Phusion polymerase (New England Biolabs, Inc), was used to re-amplify the isolated fragment to further purify the fragment and obtain additional material. After restriction digestion, the isolated fragment was inserted into the AgeI and XhoI restriction sites of pET23b-6His-SUMO (Wang, Sauer et al. 2007).

Purification of EF-G1 was performed according to procedures described in a previous paper (Malakhov, Mattern et al. 2004). Basically the 6His-SUMO tagged EF-G1 was expressed in BL21 *E.coli* cells and 1.0 liter of cells was harvested by centrifugation at 10,000 rpm for 10min. The cell pellet was resuspended in 25ml lysis buffer (50mM Tris pH8.0, 300mM NaCl, 10% Glycerol, 10mM imidazole, 1mM benzamidine, 5mM ε-Aminocaproic acid, 1mM PMSF) and sonicated using a 3s pulse/7s pause cycle for 10min. Then crude extract was spun down at 8000g for 15min. The supernatant of the crude extract was run through a 1.0ml Ni-NTA column (equilibrated with lysis buffer),

and the column washed in the same buffer. Then, 6His-SUMO-EF-G1 was eluted with 6 x 1ml elution buffer (50mM Tris pH8.0, 300mM NaCl, 10% glycerol, 300mM imidazole, 1mM benzamidine, 5mM ε-aminocaproic acid, 1mM PMSF). To remove the 6His-SUMO tag, a highly active SUMO specific cysteinyl protease with a His-tag (Ulp1-his) (Wang, Sauer et al. 2007) was purified by affinity chromatography using the same lysis buffer and elution buffer as for the His-SUMO-EF-G1. 5~10μl of ~8μM Ulp-His was incubated with the eluted 6His-SUMO-EF-G1 (6ml was diluted 10x to 60ml with lysis buffer) at either 4°C overnight or room temperature for 1 hr, and the mixture was passed through the 1.0ml Ni-NTA column again to remove cleaved 6His-tag, or any uncleaved 6His-SUMO-EF-G1 and to capture the Ulp1-His. Purified EF-G1 was recovered in the flow through fraction of the column and the concentration was determined using a calculated extinction coefficient (51520 $cm^{-1}M^{-1}$). In this way, the column was overloaded with the protein from 1.0 liter of cells. Future purifications should use less starting material or a larger nickel NTA column.

FBA was also successfully cloned into pET23b-6His-SUMO following the procedure as described for EF-G1. The primers for amplification from genomic DNA of *Synechocystis* were 5'AACCCGGGGGAATGGCTCTTGTACCAATG3' and 5'TTCTCGAGCTACACAGCAACGGAGGTG3'. The isolated fragments were inserted into XmaI and XhoI restriction sites of pET23b-6His-SUMO. 6His-SUMO-FBA was successfully expressed in BL21, but unfortunately it was not soluble, so the conventional purification method was used for all experiments reported here.

Generation of FBA and EF-G1 Antibodies

1.0mg of purified protein was submitted to Agrisera (Vännäs, Sweden) for preparation of polyclonal rabbit antiserum. The antibodies had high sensitivity, being able to detect as little as 10ng for anti-FBA antiserum (#274) and 5ng for anti-EF-G1 antiserum (#90).

Aggregation and Protection Assays with FBA, EF-G1 and Hsp16.6

The thermal sensitivity of FBA was tested by incubating 100µl of 0, 5 and 10µM FBA in reaction buffer (20mM HEPES, 100mM NaCl, pH6.5, 2mM DTT) at 40°C and 45°C for 2 hrs with shaking at 350rpm. After heat stress, the protein was centrifuged at maximum speed, 16,100 rcf, for 15min in a microcentrifuge at 4°C. 75µl of supernatant was boiled with 25µl 4X sample dye (240mM Tris pH8.0, 8% SDS, 0.038g/ml DTT, 0.6g/ml sucrose, 2.6mg/ml ε-aminocaproic acid, 0.8mg/ml benzamidine, 0.4mg/ml bromophenol blue) and saved for SDS-PAGE analysis. The pellet was washed with 475ul of reaction buffer, and the sample was centrifuged again at 16,100 rcf for 15min. 480µl of wash supernatant was discarded. The remaining 20µl of sample was boiled with 113µl 1X sample dye and used for SDS-PAGE analysis. Total proteins at the above concentrations without any treatment were used for comparison; 100µl of total protein was boiled with 33µl 4X sample dye and saved for SDS-PAGE.

The same basic procedure was used to test the thermal sensitivity of EF-G1, except a different reaction buffer and heat shock temperature were used. One of the reaction buffers tested for EF-G1 was 25mM sodium phosphate, 100mM NaCl, pH 7.5, 2mM

DTT, and another was 20mM HEPES, 100mM NaCl, pH 6.5, 2mM DTT. The heat shock temperature used for EF-G1 was 47°C for 2 hrs.

For the protection assay using FBA with Hsp16.6, 0, 5, 10 and 20μM Hsp16.6 were incubated with 10μM FBA at 45°C for 2 hrs with shaking at 350rpm. Subsequent sample preparation was the same as for analysis of FBA thermal sensitivity.

Circular Dichroism (CD) of FBA and EF-G1

10μM FBA and 5μM EF-G1, both in 10mM sodium phosphate pH7.5, were prepared for CD spectroscopy (Jasco J-715 Spectropolarimeter). A spectral scan was first performed with protein at room temperature (20 ℃). Afterwards, protein was heated at 95°C for approximately 5min, and the spectrum of the protein was reacquired. Then, protein was cooled back to room temperature, and scanned again. All the scans were taken four times at a scan rate of 50nm/min. Data pitch was 1nm. The cuvette path length was 1mm.

For the CD melting experiment, 10μM FBA and 5μM EF-G1 were prepared, but in 20mM HEPES pH 7.5 buffer. The spectrometer was programed to heat from 20°C to 95°C at the rate of 1°C/min and then cooled down from 95°C to 20°C at the same rate. The CD (mdeg) was monitored during the melt using 222nm for FBA and 215nm for EF-G1. During the EF-G1 experiment, the spectrometer stopped functioning, and it was not possible to acquire the 95 ℃ to 20 ℃ scan. Thus, only the data for FBA include the scan from 95°C to 20°C.

CD data for both FBA and EF-G1 were acquired only once, and need to be repeated.

Complex Formation of FBA with Hsp16.6

Size exclusion chromatography (SEC) was used to examine the physical association of FBA and Hsp16.6. 2.5, 5 and 10μM Hsp16.6 were separately incubated with 5μM FBA (total sample volume 150μl) at 40°C or 45°C, or at room temperature for 2 hrs. The reaction buffer was 20mM HEPES, 100mM NaCl, pH 6.5, 2mM DTT. After heating, the samples were centrifuged at 13,000rpm for 5min. Supernatant (100μl) was injected onto a G5000-SEC column (Tosoh Bioscience LLC) on a HPLC system (Waters) at room temperature. The running buffer for the chromatography was 20mM sodium phosphate, 100mM NaCl, pH 6.5. The column was pre-equilibrated with running buffer. The flow rate was at 0.6ml/min. 5μM FBA alone and 10μM Hsp16.6 alone (150μl) were prepared and analyzed by SEC using the same conditions as above.

In parallel, 100μl samples were prepared as described above for SDS-PAGE analysis. The supernatant and pellet from these samples were obtained by the procedures described for the protection assay of FBA.

Self-aggregation of EF-G1

SEC was also used to check the size of EF-G1 unheated and after heat stress (47℃). 120μl of 5μM EF-G1 in buffer (20mM HEPES, 100mM NaCl, pH 6.5, 2mM DTT) was prepared in duplicate. One sample was maintained at 4℃ and the other was incubated at 47℃ for 2 hrs. 100μl samples were respectively analyzed by SEC. Other conditions were same as for SEC of FBA.

In vivo Analysis of FBA and EF-G1 in *Synechocystis*

To estimate what percentage FBA and EF-G1 represent of the total proteins in *Synechocystis*, the total *Synechocystis* protein was analyzed first using a Coomassie stain protein assay. 1.0ml of *Synechocystis* cells at 0.5, 1.0, 2.0, 3.0, or 4.0 OD ml^{-1} (730nm) were collected by centrifugation at 6000rpm for 10min, then resuspended and boiled in 100µl 1X sample buffer. 0 (control), 0.25, 0.5, 1.0, 2.0, 3.0, or 4.0 µg BSA μl^{-1} in 1X sample buffer were also prepared as a standard. 1µl of each sample was spotted on a clean piece of Whatman 3M paper. After drying, the paper was stained for 30min in 0.2% Coomassie blue solution (0.1% Coomassie Blue in 10% acetic acid, 50% methanol and 40% ddH_2O) and destained with destaining solution (20% methanol, 10% acetic acid, 70% ddH_2O). Once the paper was fully destained, it was dried completely. Each protein spot was cut out and placed in tubes containing 1.0ml 2% SDS in H_2O overnight to elute Coomasie dye from the paper. Finally, the absorbance at 590nm was obtained for the eluted solution. Background control value (1X sample buffer alone) were substracted from OD values. A regression line could be made that fit $y=0.041x+0.0068$, where y is the OD at 590nm, and x is the concentration of BSA (µg/µl). The absorbance of the BSA was used to generate a standard curve. The total protein concentration in *Synechocystis* cells was estimated based on the BSA regression line.

To estimate the amount of FBA per µg total *Synechocystis* protein, each of the whole cell protein samples alone, and 10, 20, 50, 80, 100ng recombinant FBA were first separated on 15% SDS-PAGE. Similarly, 5, 10, 15, 20, 25ng recombinant EF-G1 were separated for comparison to the total protein samples on 10% SDS-PAGE. Proteins were transferred from SDS-PAGE to PVDF membrane (BioRad) at a

constant 30mA for 1 hr using a semi-dry blotter (BioRad). Blots were blocked with 5% milk in 1XTBST (50 mM Tris.HCl, pH 7.4, 150 mM NaCl, 0.1% Tween 20) at room temperature for 0.5 hr, incubated with a dilution of 1:1000 primary antibody (Agrisera) in 5% milk solution at room temperature for 1hr and a dilution of 1:10,000 ECL Rabbit IgG (HRP-linked whole Ab from donkey) (GE Healthcare Life Sciences) at room temperature for 0.5 hr both with agitation. In between the incubations with primary and secondary antibody and between the secondary antibody incubation and exposure, the blots were washed with 3X10ml 1XTBST for 7min each time. The blots were exposed in ECL substrates (Thermo). Data were acquired from the blots using a G-box (New England BioGroup, LLC) and quantified with imageJ (NIH).

Results

Heat Sensitivity of Fructose Bisphosphate Aldolase Class II (FBA)

Because the putative substrates were observed associated with Hsp16.6 only after *Synechocystis* cells had been heat stressed (Basha, Lee et al. 2004), it would be expected that these proteins are heat sensitive and would form a complex with Hsp16.6 when heated. To test these predictions, FBA was cloned, expressed in *E.coli* and purified as described in the Methods section.

The heat sensitivity of FBA was first tested by analysis of the formation of aggregates that sediment during centrifugation at 13,500rpm for 15min (see Methods), followed by separation of the resulting supernatant and pellet fractions by SDS-PAGE. When heated at 40°C (Figure 12A) for 2 hrs, both 5μM and 10μM of FBA began to form insoluble aggregates, which end up in the pellet fraction after centrifugation. However,

only less than 50% of FBA became insoluble at either concentration tested at this temperature. On the other hand, when heated at 45°C (Figure 12B), more than 50% of 5μM and 10μM FBA end up in the insoluble pellet. FBA in buffer (pH 7.5) and FBA in buffer without DTT were also tested for aggregation behavior at 47°C, but much less insoluble FBA was found in pellet. Therefore, further experiments were performed at 45°C, with buffer at pH 6.5 containing 2 mM DTT, which led to the greatest aggregation. Higher temperatures were avoided, as they are non-physiological.

In contrast to the results of the aggregation assays, the CD spectroscopy (Figure 13 A, B, C), suggests FBA retains secondary structure until heated to 80°C, although FBA did not recover native secondary structure when the temperature was reduced from 95°C to 20°C. The apparent difference in heat sensitivity of FBA in the aggregation assay and CD experiment might result in part from differences in the buffer conditions for the two experiments. The CD data were obtained at pH 7.5 in 20mM HEPES buffer, while the aggregation assay was performed at pH 6.5 in 20mM HEPES buffer.

Hsp16.6 Prevents Thermal Aggregation of FBA

As shown previously, Hsp16.6 protects serine esterase from forming insoluble aggregates when the molar ratio of Hsp16.6 to *Synechocystis* serine esterase is between 0.5 and 2 (Basha, Lee et al. 2004). To test if Hsp16.6 could also protect FBA from insolublization, different concentrations of Hsp16.6 were mixed and heated with FBA for 2 hrs at 45°C (Figure 14), holding the FBA concentration constant at 10μM.

When the molar ratio (monomer to monomer) of Hsp16.6 to FBA was 0.5, Hsp16.6 started to protect FBA from transitioning to the insoluble fraction. When the ratio was increased to 2.0, FBA was fully protected from aggregation by Hsp16.6. In a control experiment, when same concentrations of bovine serum albumin (BSA) were mixed and heated with FBA, most of the FBA ended up in the insoluble fraction.

Figure 12 Thermal Sensitivity of Fructose Bisphosphate Aldolase Class II (FBA)

The heat sensitivity of FBA at different concentrations and under different temperatures (A) 40 ℃ (B) 45 ℃. Heat aggregation was performed as described in the text. More than 50% of FBA became insoluble after a 45 ℃ heat stress. T: Total protein; S: soluble; P: protein aggregates (pellet); * is a minor contaminant protein.

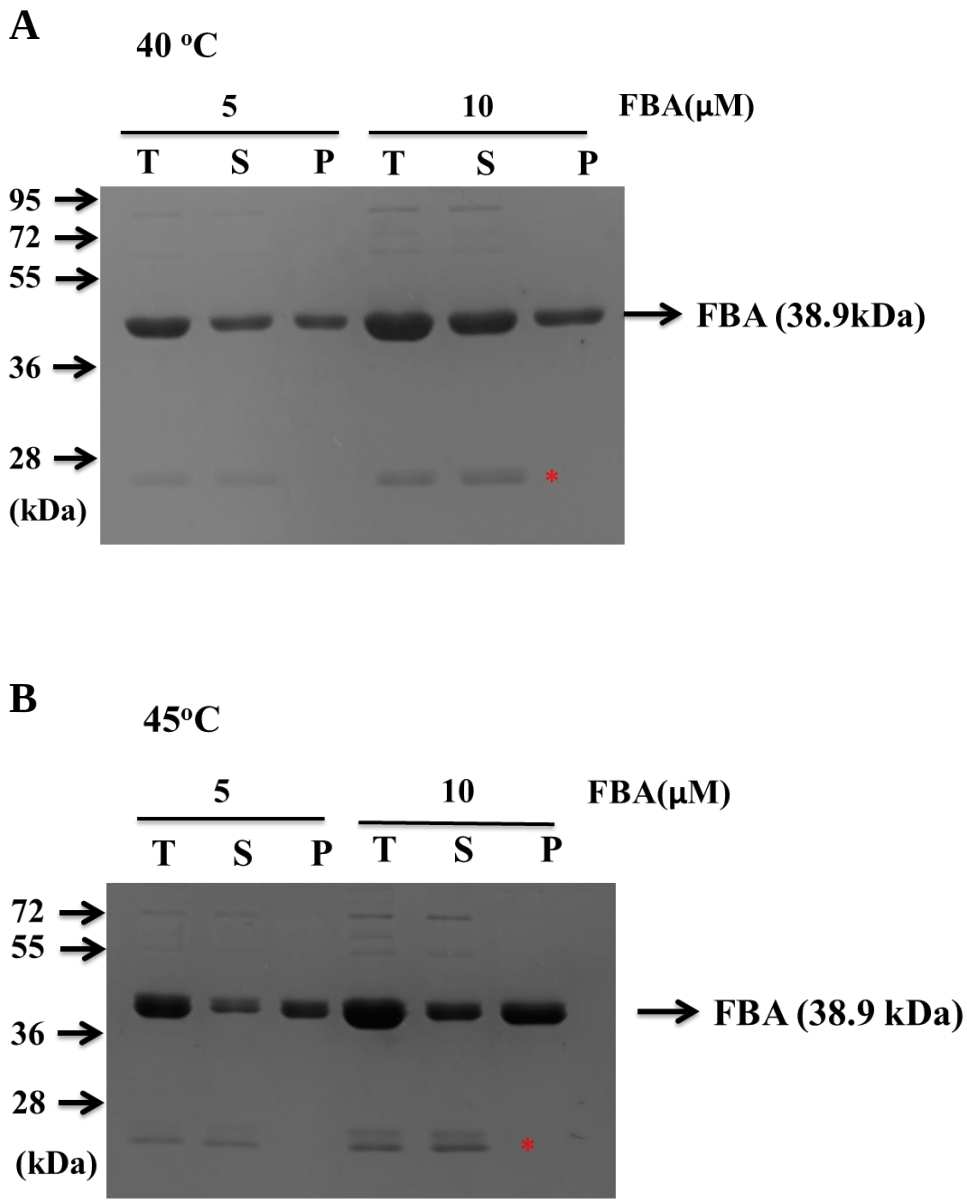
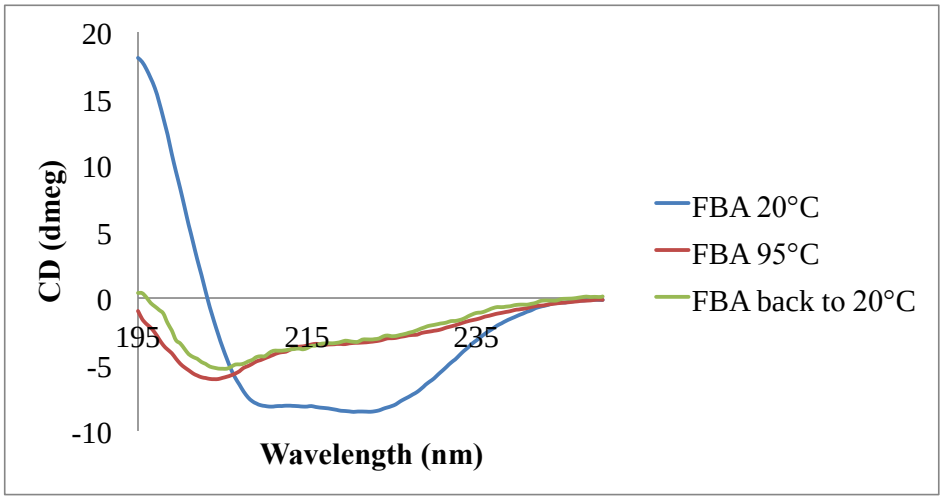


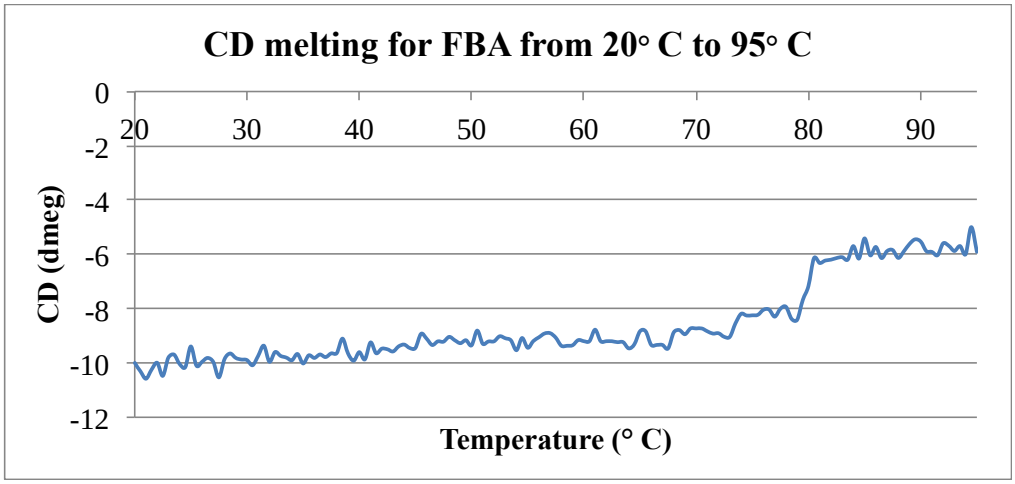
Figure 13 Secondary Structure of FBA

(A) CD spectroscopy showed that FBA unfolded during heating up to 95 ℃ and never refolded when the temperature was returned to 20 ℃. (B) When the temperature was slowly increased from 20 ℃ to 95 ℃, the FBA started to unfold at around 80 ℃. (C) FBA did not reform secondary structure when the temperature was decreased from 95 ℃ to 20 ℃.

A


20
15
10
5
0
-5
-10
CD (dmeg)
195
215
235
Wavelength (nm)
FBA 20°C
FBA 95°C
FBA back to 20°C

B


CD melting for FBA from 20° C to 95° C
0
-2
-4
-6
-8
-10
-12
20
30
40
50
60
70
80
90
CD (dmeg)
Temperature (° C)

C

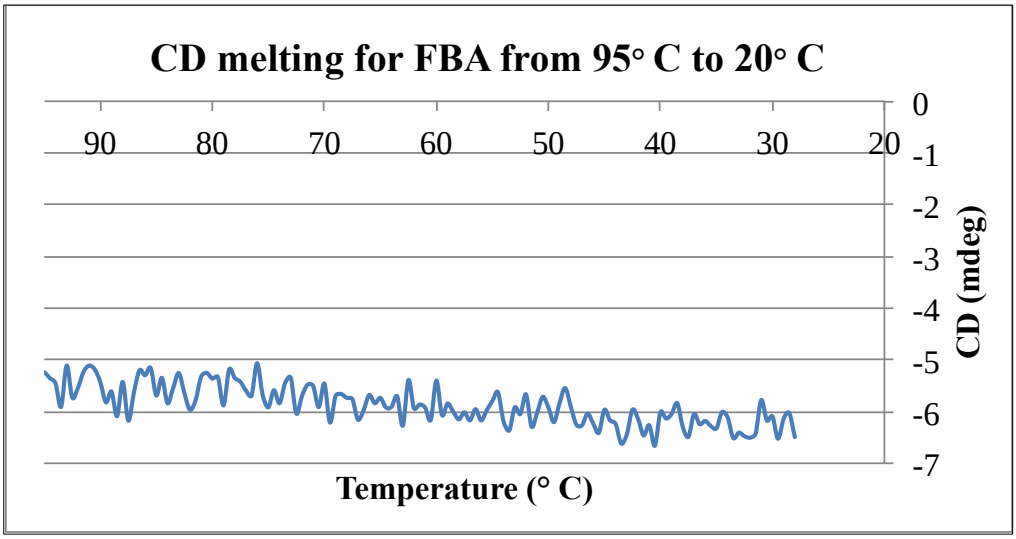
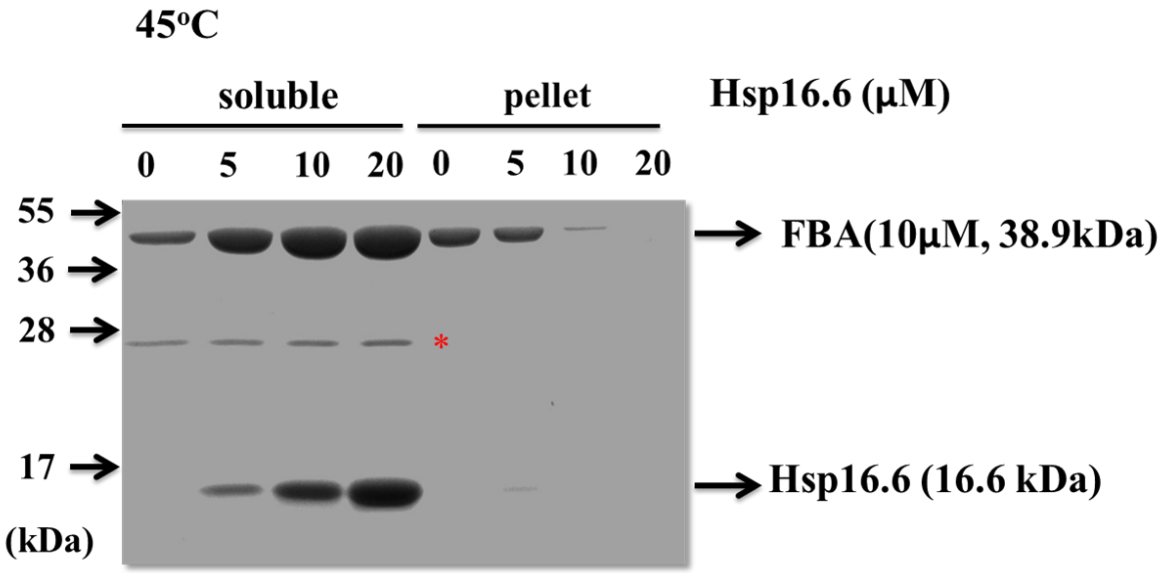

CD melting for FBA from 95° C to 20° C
90
80
70
60
50
40
30
20
0
-1
-2
-3
-4
-5
-6
-7
CD (mdeg)
Temperature (° C)

Figure 14 Hsp16.6 Prevents Thermal Aggregation of FBA

Different molar ratios of Hsp16.6 to FBA were incubated at 45℃ for 2 hrs as described in the text. Soluble and pellet fractions were separated by SDS-PAGE and the gel stained with Coomassie Blue. At a molar ratio of Hsp16.6:FBA of 1:1, Hsp16.6 already provides full protection to FBA. * indicates a minor contaminant in the FBA preparation.



Hsp16.6 Forms a Complex with FBA during Heat Stress

10μM Hsp16.6 alone and 5μM FBA alone were incubated at room temperature, 40°C or 45°C. The soluble fractions were separated by size-exclusion chromatography (SEC) (Figure 15 A and B) and both soluble and insoluble fractions were checked by SDS-PAGE. From SEC, the quantity of FBA clearly decreased with increasing temperature, while Hsp16.6 didn't change at all as temperature increased. From SDS-PAGE, lanes 4 and 9, Hsp16.6 remained in the soluble fraction even after incubation at both 40℃ and 45℃. On the other hand, more FBA aggregates turned up in the pellet fraction when heated to 45°C compared to 40°C. As predicted, FBA remained soluble at room temperature. Additionally, based on SEC FBA elutes around 158kDa, consistent with it being a native tetramer, since the monomer size of FBA is 38.9kDa.

Typically, sHSPs form large heterogeneous complexes with heat sensitive proteins (Basha, Lee et al. 2004). By SEC, Hsp16.6 has previously been shown to form a complex with serine esterase (Basha, Lee et al. 2004). To investigate Hsp16.6-FBA interaction, 2.5, 5 and 10μM Hsp16.6 were separately incubated with 5μM FBA at room temperature or at 40°C or 45°C for 2 hrs. The soluble and insoluble fractions were separated by SDS-PAGE, and in parallel the soluble fractions were separated by SEC using high-performance liquid chromatography (HPLC). From the chromatogram (Figure 15 C-E), when 2.5μM Hsp16.6 was incubated with 5μM FBA at different temperatures, the quantity of FBA eluting from the column decreased as incubation temperature increased, and no additional peaks eluting earlier appeared. From SDS-PAGE separation of the same samples in lanes 1 and 6 (Figure 15F), more aggregates appeared in the pellet fraction and less soluble protein in the supernatant as

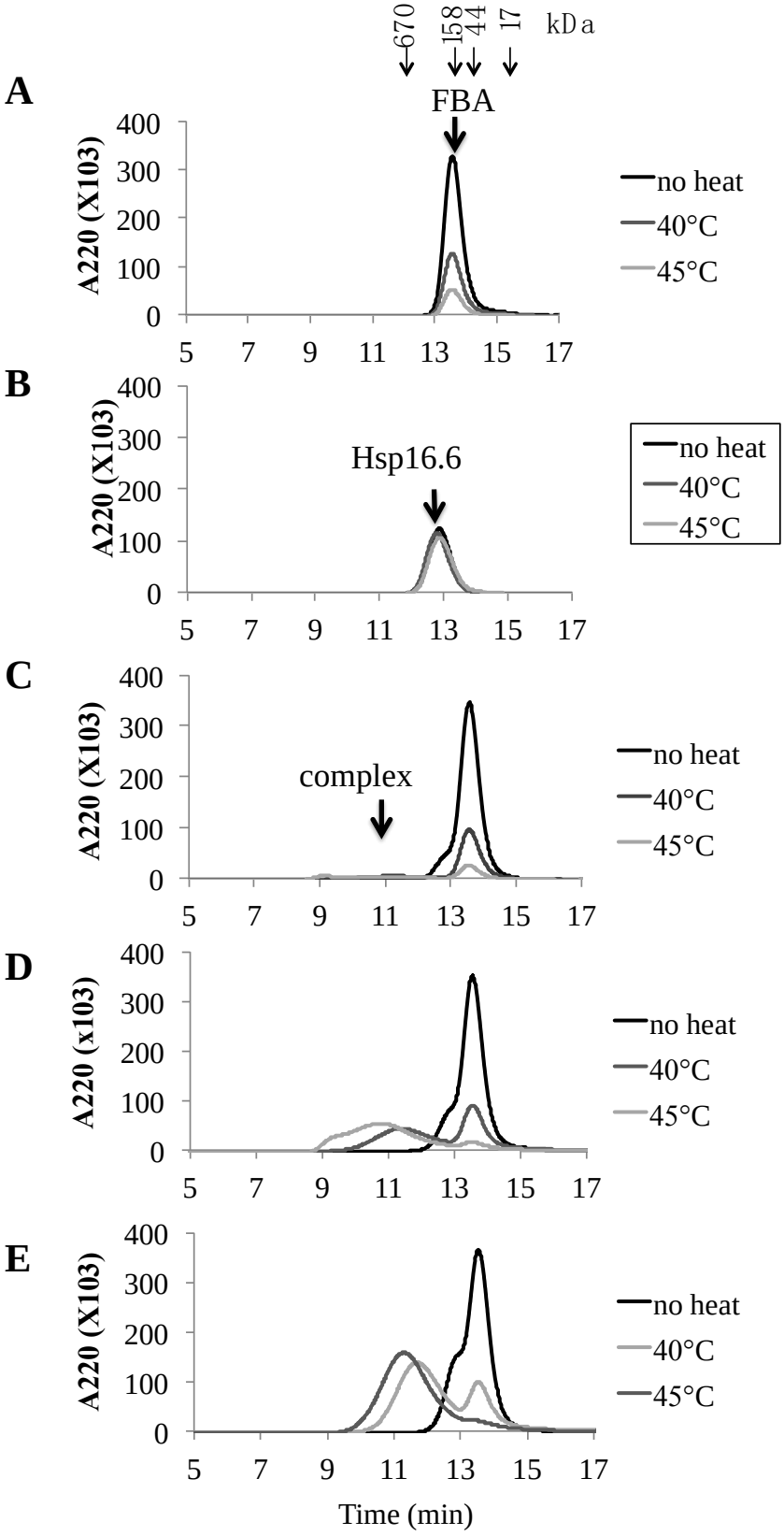
temperature increased. When Hsp16.6 was increased to 5μM and incubated with 5μM FBA, the amplitude of the FBA peak decreased further, but a new peak corresponding to Hsp16.6-FBA complex eluted at a larger size, and the Hsp16.6-FBA complex formed at 45°C was even larger than that formed at 40°C (Figure 15D). From lanes 2 and 7 in SDS-PAGE, greater than 70% of FBA stayed in the soluble fraction, but there was somewhat more insoluble FBA at 45°C than at 40°C. When Hsp16.6 was increased to 10μM with 5μM FBA, the peak of FBA still decreased, and a new peak of complex appeared. The complex formed at 45°C was also bigger than that formed at 40°C. However, at the 2:1 ratio of Hsp16.6 to FBA, both complex peaks eluted later than complexes formed at a 1:1 ratio of Hsp16.6 to FBA, indicating complexes formed at the 2:1 ratio are smaller. Again in lanes 3 and 8 from SDS-PAGE, more than 90% of FBA ended up in the soluble fraction.

Heat Sensitivity of Elongation Factor G1 (EF-G1)

As for FBA, the heat sensitivity of EF-G1 was analyzed by SDS-PAGE (Figure 16 A and B). EF-G1 did not form insoluble aggregates at various concentration even when heated at 47°C for 2 hrs either at pH 7.5 or pH 6.5. In contrast, when CD (Figure 17 A and B) was used to monitor the unfolding and refolding of EF-G1 as a function of temperature (Greenfield 2006), EF-G1 appeared to almost completely lose secondary structure above 47°C, suggesting the protein is heat sensitive, although the formation of large aggregates that could be separated by centrifugation were not observed.

Figure 15 Hsp16.6 Forms a Complex with FBA during Heat Stress

Different ratios of FBA plus Hsp16.6, or each alone were prepared either with heating at 40℃ or 45℃, or without heating. Each sample was separated by SEC. A FBA alone. B Hsp16.6 alone. C, D and E were prepared with 0.5, 1, 2 molar ratios, respectively of Hsp16.6 to FBA. The markers on the top are the standards size (kDa). F SDS PAGE of the soluble and insoluble fractions.


670
158
44
17
kDa
A
FBA
A220 (X103)
400
300
200
100
0
5 7 9 11 13 15 17
no heat
40°C
45°C
B
Hsp16.6
A220 (X103)
C
complex
A220 (X103)
D
A220 (x103)
E
A220 (X103)
Time (min)

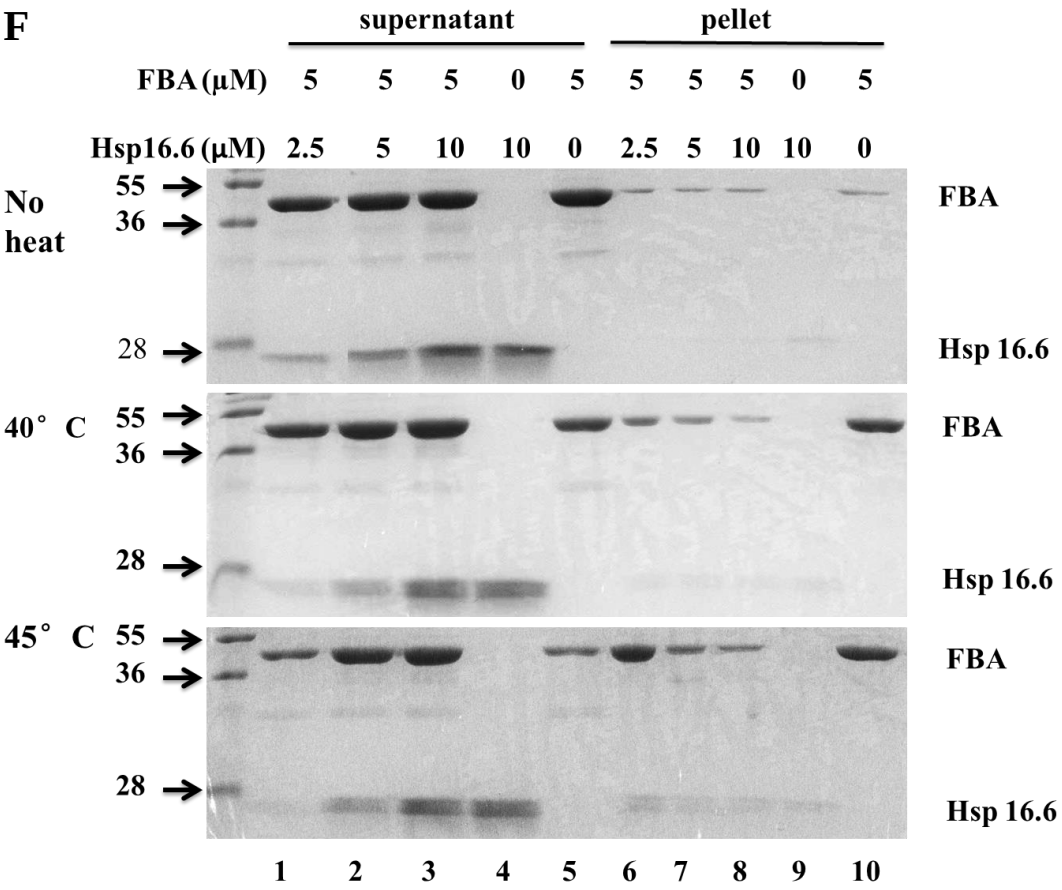
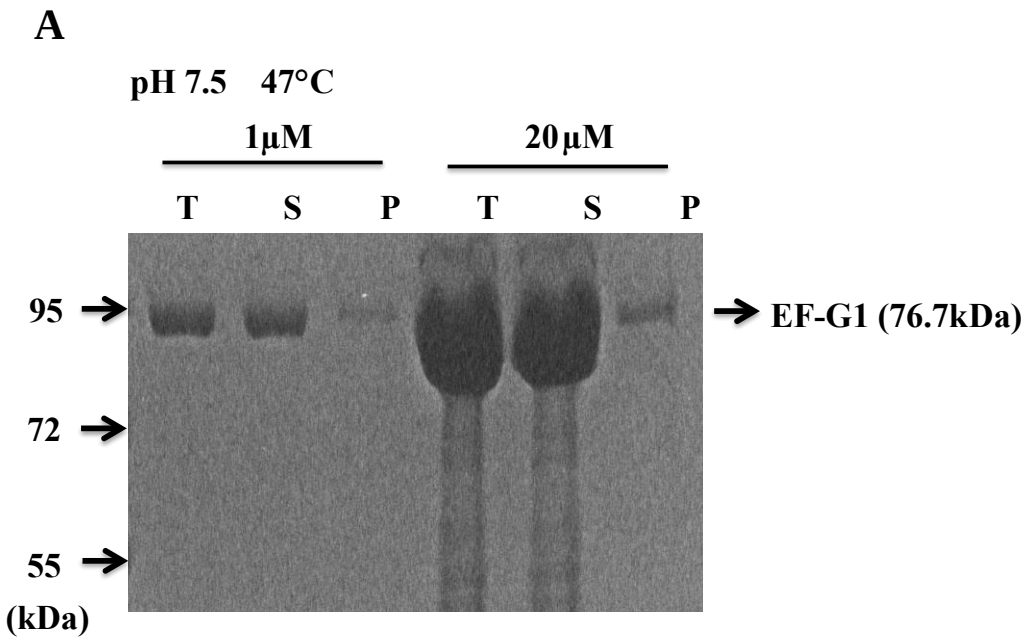

F
supernatant
pellet
FBA (μM) 5 5 5 0 5 5 5 5 0 5
Hsp16.6 (μM) 2.5 5 10 10 0 2.5 5 10 10 0
No heat
55
36
28
FBA
Hsp 16.6
40° C
55
36
28
FBA
Hsp 16.6
45° C
55
36
28
FBA
Hsp 16.6
1 2 3 4 5 6 7 8 9 10

Figure 16 Thermal Sensitivity of Elongation Factor G1 (EF-G1)

Different concentrations of EF-G1 were incubated at up to 47 ℃ for 2 hrs. All the EF-G1 remained in the soluble fraction whether heating was performed at pH 7.5 (A) or at pH 6.5 (B). T: total protein, S: soluble, P: aggregate protein (pellet)



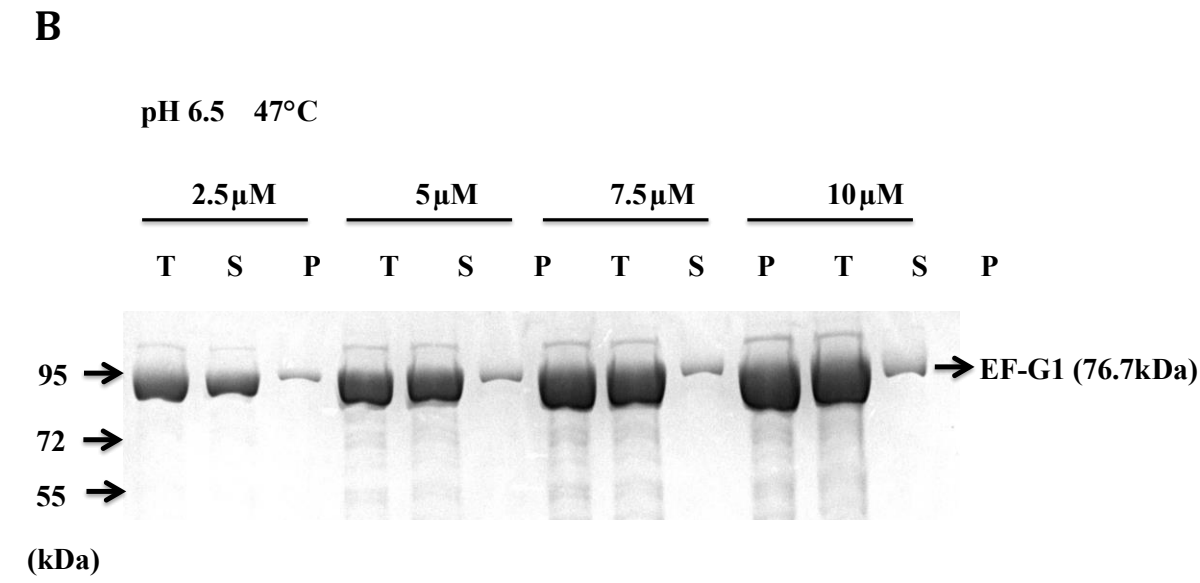
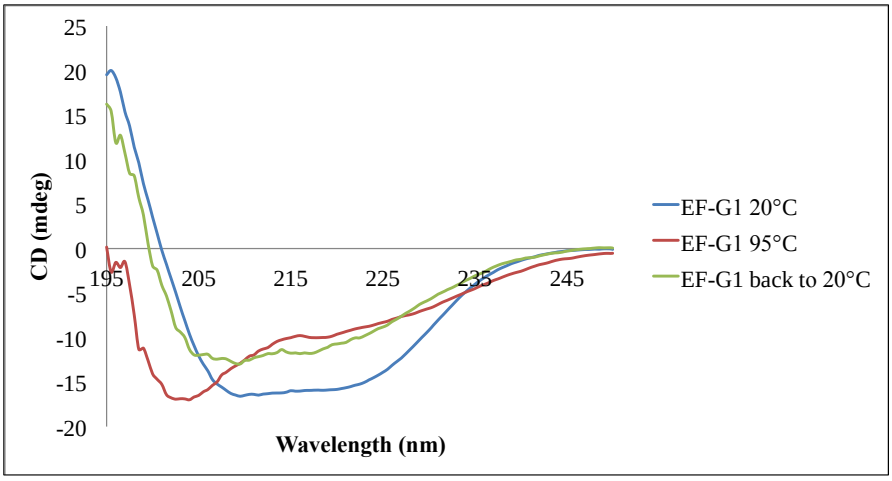


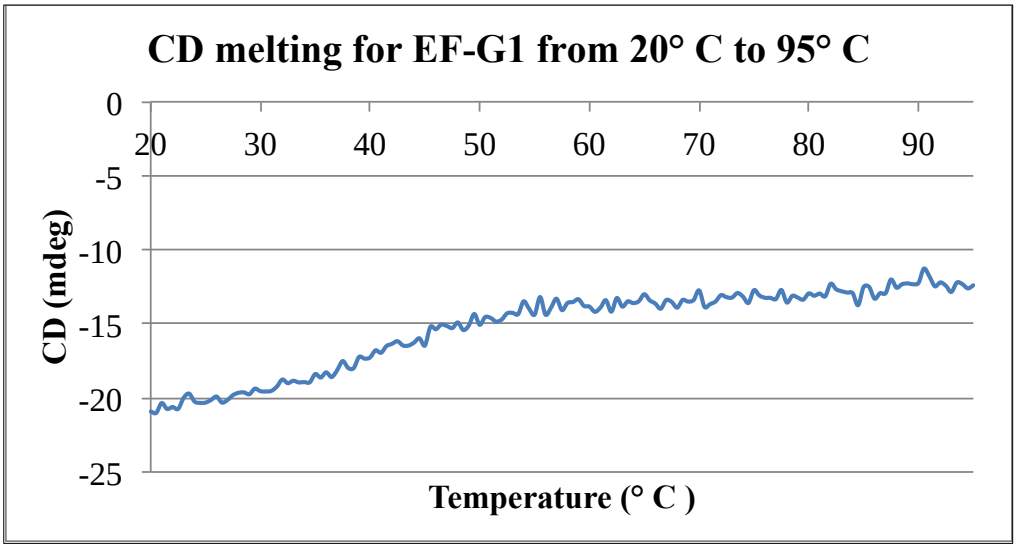
Figure 17 Secondary Structure of EF-G1

(A) CD spectroscopy indicates that EF-G1 unfolded when heated to 95 ℃ and partially refolded when cooling back to room temperature. (B) From the CD melt, EF-G1 started unfolding at around 40 ℃.

A



B



EF-G1 Forms Self-aggregates

To check how EF-G1 behaves during heat stress, SEC was performed. EF-G1 formed a self-aggregate as large as 670kDa during heat stress, while native EF-G1 was eluting at around 158kDa. Based on the elution of EF-G1 relative to the MW markers, EF-G1 appears to be dimeric, since the monomer of EF-G1 is 76.7kDa (Figure 18).

Analysis of FBA and EF-G1 in vivo

To explore how much FBA and EF-G1 are present in *Synechocystis* cells, western blotting with primary antibody against *Synechocystis* FBA and EF-G1 was performed, using purified recombinant FBA and EF-G1 as standards (Figure 19). On average, *Synechocystis* cells have 2.16±0.83% FBA of total protein and 0.73±0.16% EF-G1 of total protein. (outliers were excluded from this calculation). Therefore, both of these substrates represent abundant cellular proteins in *Synechocystis*.

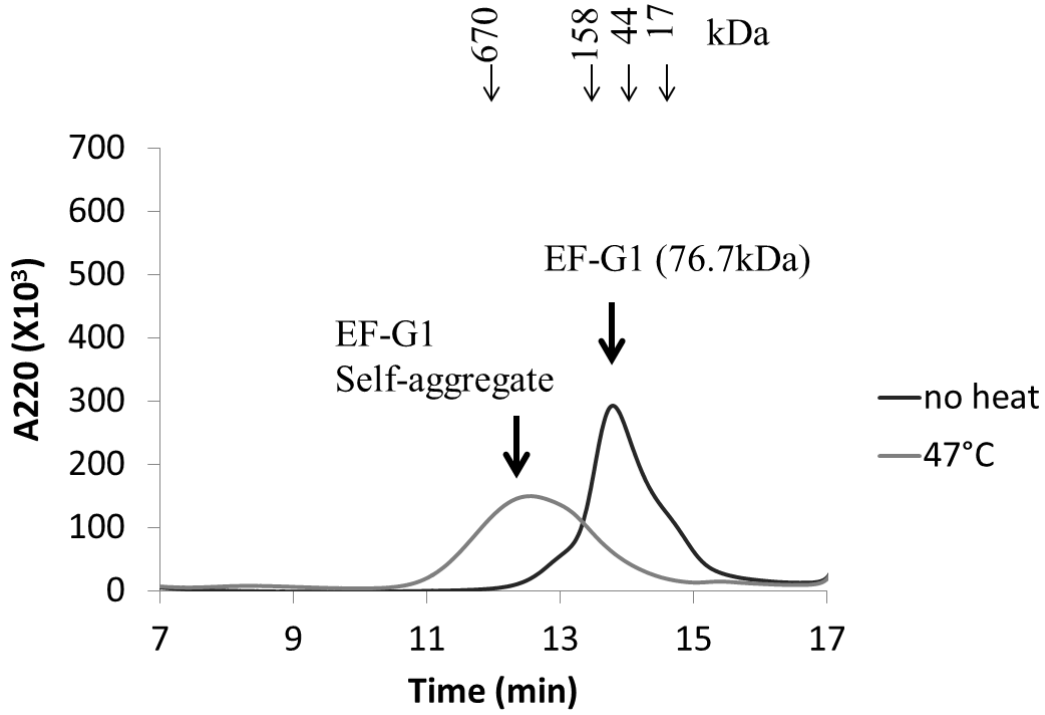
Discussion

FBA is a heat sensitive protein according to the thermal sensitivity assay and Hsp16.6 is able to protect FBA from aggregation. More interestingly, Hsp16.6 binds to FBA and forms a large complex during heating in vitro. These data support the conclusion that FBA is an actual substrate of Hsp16.6 in vivo.

Even though EF-G1 does not aggregate during heat stress in vitro under the conditions tried, it does not necessarily mean that EF-G1 does not unfold during heat stress. In fact, CD spectroscopy indicates that EF-G1 unfolds when heated to 40°C or

Figure 18 EF-G1 Forms Self-aggregates during Heat Stress

EF-G1 (no heat) eluted at 158kDa, while after heating, EF-G1 self-aggregates eluted at around 670kDa.



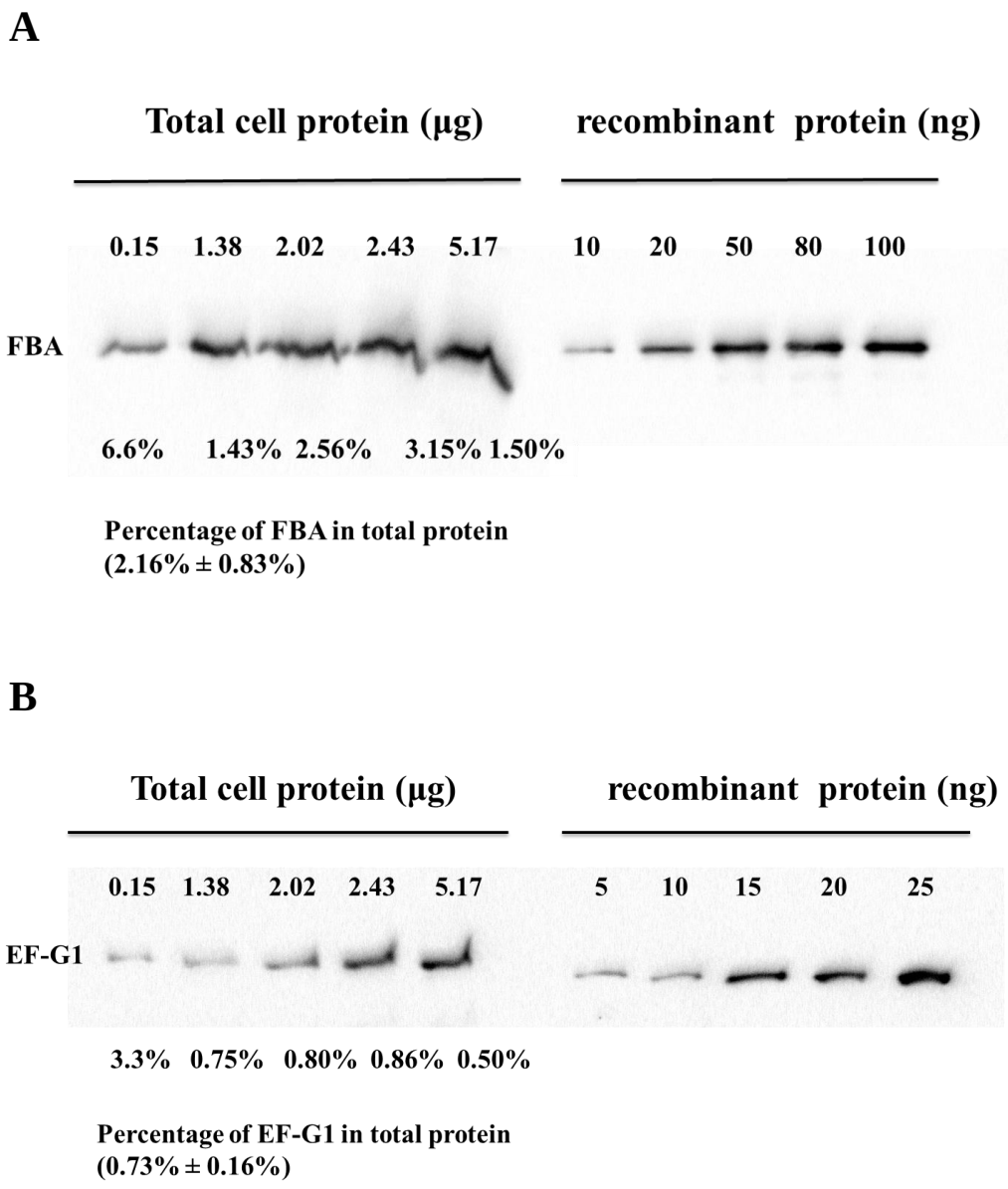
higher with losing the secondary structure. It is possible that the protein does not form large aggregates when it is unfolded, and therefore would not appear to be heat sensitive in the assay for insolubility.

Even though FBA aggregates in vitro, the thermal sensitivity of FBA in vivo is still not known. Also, we have no idea if Hsp16.6 would protect FBA from irreversible aggregation in vivo. Similarly, EF-G1 might aggregate in vivo even though it does not form large aggregates in vitro, perhaps in association with other proteins of the translation machinery. It remains to be determined if Hsp16.6 binds to EF-G1. SEC analysis of heated EF-G1 with Hsp16.6 showed no evidence of complex formation, although the apparent size of EF-G1 shifted with temperature. It was hard to determine if complexes formed using SEC analysis because of the shift of EF-G1 after heating, making it difficult to visualize a complex between EF-G1 and Hsp16.6, which might elute at a similar size range. Further studies will be required to understand the effect of heat on EF-G1 both in vitro and in vivo to understand any possible interactions with Hsp16.6.

From the in vivo estimate of the abundance of FBA and EF-G1, both proteins are readily detectable using polyclonal antibodies, making it possible to perform the in vivo protection assay in cells with or without Hsp16.6. The experiment in Figure 19 also provided the valuable information that while both proteins are major cellular components, EF-G1 is not as abundant as FBA. Since Hsp16.6 has been estimated to accumulate to 0.5% of the total cell protein during 42℃ heat stress for 2 hrs (Basha, Lee et al. 2004), Hsp16.6 might not be able to fully protect FBA which is about 2% of the total cell protein during the heat stress, indicating that part of FBA will end up in

Figure 19 Analysis of FBA and EF-G1 Levels in vivo

Western blotting was performed to detect what percentage of FBA (A) and EF-G1 (B) were present in the *Synechocystis* cell, with recombinant FBA and EF-G1 proteins as standards. The amount of the total protein of *Synechocystis* cell was estimated by Coomassie stain protein assay which was described in the Methods section.



the insoluble fraction if western blotting were performed to check FBA in the soluble and insoluble fraction of *Synechocystis* cells. We have less knowledge about how Hsp16.6 might protect EF-G1, therefore it is hard to estimate if Hsp16.6 could fully protect EF-G1 in vivo.

From these results, a new in vivo substrate was found and could be considered as a model substrate for continued work on sHSPs. The results with FBA fit well with the proposed model for the mechanism of sHSP function. The data also indicate that like EF-G1, other proteins might form soluble aggregates, in contrast to the traditional thought that aggregated proteins are insoluble.

CHAPTER 4

FUTURE DIRECTIONS

In future bioinformatics study, additional properties of Hsp16.6 substrates could be assessed. New algorithms could be written to obtain the aggregation propensity of Hsp16.6 substrates and the cutoff proteome as referred to in TANGO (Fernandez-Escamilla, Rousseau et al. 2004), and to classify the structure of proteins according to SCOP (http://scop.berkeley.edu/). Machine learning attempts could also be made to hypothesize whether a protein is a Hsp16.6 substrate or not. Additionally, our group also has information about Arabidopsis sHSP potential substrates that were identified in the same way in which Hsp16.6 substrates were acquired. The same bioinformatics analysis could be applied to these substrates.

In the future work, enzymatic assays could be performed to analyze how FBA and EF-G1 activity change during heat stress with and without Hsp16.6 in vitro. Will Hsp16.6 also help substrates recover activity during heat stress? Other experiments could be considered to check if Hsp16.6 could prevent EF-G1 from self-aggregating, with, for example, mass spectrometry. The behavior of FBA and EF-G1 could be determined in vivo using the antiserum from Agrisera. For example, FBA and EF-G1 could be checked to determine if they aggregate in different isogenic *Synechocystis* strains, with and without Hsp16.6, during heat stress. Also, Sally Chu from the Department of Microbiology has successfully purified recombinant glutamate ammonia ligase (cyanobase accession number: slr0288), another substrate. The same

experiments in this thesis could be performed with glutamate ammonia ligase to check whether or not it is a substrate of Hsp16.6. More candidate substrate proteins could also be chosen for study.

APPENDIX I

SYNECHOCYSTIS SUBSTRATE TABLE (1D AND 2D)

The first set of data is from bands excised from a 1-D gel. Therefore there are multiple proteins per band. The second set is from 2-D gel spots. Proteins associated with Hsp16.6 after heat stress were isolated essentially as described in Basha et. al. (2004) except that instead of immunoprecipitation, Hsp16.6 was recovered from cell extracts using a C-terminal Strep-tagged version of Hsp16.6 which had been replaced into the Hsp16.6 locus in *Synechocystis* cells as described (Basha, Lee et al. 2004). Data below are unpublished and were kindly provided by Dr. Eman Basha.

Band/ Spot	protein name	sequence	MW	pI	Acc #
2	RNA polymerase beta prime subunit [Synechocystis sp.	K.DVVGPDGEIIAK.R R.IAAVTDEVYVR.S K.VFDEPAAPSQGSQNEEGGR.Q K.GDNYQLVLR.R K.TGDIVQGLPR.I	144777	4.8	S77517
	methyl-accepting chemotaxis protein	K.IVAVISQIASR.T R.ALEDIIEVSNR.I	93.2	4.4	S74988 slr1044
	Contains similarity to gb|D90908 DNA mismatch repair protein MutS2 from Synechocystis sp.	K.TLGLLSLM@SKS.G (4X)	96.9	6.8	E96674
3	methyl-accepting chemotaxis protein	R.SSASFGTSGAR.S K.VRSEDELGALTQR.F R.SEDELGALTQR.F R.LLDDVEGASR.G R.EIVLQVK.N R.EIVLQVK.N R.EAEEVAHTSSLTAL	93.3	4.4	S74988 slr1044

		K.G R.TVGGILQIR.E K.IVAVISQIASR.T R.GFAIVADEVR.Q R.ALEDIIEVSNR.I			S76044
	CheA like protein	R.DLLTSVER.F R.TLLEGALLASR.S R.QVTTQLQEGMTK.S R.GNFSEEAPTIVR.S	120.5	4.8	slr0322
	methyl-accepting chemotaxis protein sll0041 - Synechocystis sp. (strain PCC 6803)	R.IPVAMITSR.G R.NESAQQAQILK.E K.ANLVAPINYK.G R.GFAVVADEVR.S	96.8 148.6	5.1 5	sll0041 S75000
	Mg-chelatase subunit; ChlH	K.GLQGNGYDVQDLP GSAK.E (2X) R.DTIVGSVYR.K (2X)	96.9	6.8	D90908
	Contains similarity to gb|D90908 DNA mismatch repair protein MutS2 from Synechocystis sp. CheA like protein [	K.TLGLLSLM@SKS.G (3X) K.IPVAMLTSR.G	153	4.6	S75938
4	methyl-accepting chemotaxis protein sll0041 - Synechocystis sp. (strain PCC 6803)	T.SIQAPTQSGGLSLR. N R.LPVPQTEQQVK.D V.PQTEQQVK.D K.AQTALALK.A R.HQQDLSLK.Q K.QAELLTELSR.A R.ANLSDIDEIQGVIQK .N K.ASLTVPLHR.D R.NESAQQAQILK.E K.ALGATIADPCFADS YVEK.Y K.ANLVAPINYK.G K.ANLVAPINYK.G R.SDLLAQQK.I R.QALDVAEALER.L K.SIQAVAENAAQAES	96.8	5.1	sll0041

	RNA polymerase beta subunit	AVQR.A K.SIQAVAENAAQAES AVQR.A R.ATQTVDQGEDAMN R.T R.ATQTVDQGEDAM @NR.T R.TVDGIVAIR.E V.DGIVAIR.E R.GFAVVADEVR.S R.GFAVVADEVR.S F.AVVADEVR.S K.NSSEASGVSATFK.E R.GTFIINGAER.V R.VIVNQIVR.S R.TYSASLIPNR.G R.SVGELLQNQIR.V R.ISALGPGGLTR.E K.LGPEEITR.E R.NLDEHGIIR.I R.STGPYSLVTQQPLG GK.A R.NEALNAIVK.G	123.4	5.3	P77965
5	HlyB family	R.TYLFVDTTNR.I R.VNELENIR.Q R.LADIVDTPQESER.D R.YYYLYQQQGAGGD DV.-	112.1	5.7	S75806
6	methyl-accepting chemotaxis-like protein [Synechocystis sp. PCC 6803]gi|7470738|pir|| S75285 methyl-accepting chemotaxis protein homolog sll1294 - Synechocystis sp. (strain PCC 6803)gi|1652276|dbj| BAA17199.1| ORF_ID:sll1294~methyl-accepting chemotaxis pr	K.YAAATDDLALDEE R.S R.GQSDNLAIIQAAR.L R.DIEYATLVGQDQR.I R.YTVTPVQDPQSK.K K.AQENPDM@PLVGR .T R.LLTDIEESSR.G R.LAESSLEISK.I K.IVGIISGISEK.T K.SQLVSQSLQSLAK.T	103.2	4.6	S75285 sll1294
7	delta-1-pyrroline-5-carboxylate	R.IYADAALVQPGQK. H	110	5.7	S75910

	dehydrogenase	R.TITGAIVSR.Q			
8	core-membrane linker ClpB protein phycobilisome LCM core-membrane linker polypeptide	R.AASADYFR.A R.ALELAFR.H R.GLGVEAQECR.N R.GPAVNNQVGNPSA VGEFPGSLGAK.V K.FGESSTQALIR.A K.SELFLK.L R.QEM@NQYFDIASK. Q K.EYSDAFGEDTVPYE R.Y K.EFYAPYPNTK.V K.EIQQYNQILASQGL K.A K.AIDLVDEAAAR.L K.EAVAAVSAAIR.R R.FVELGQVSAIR.T K.LSNNEINVK.E	100.5 101.4 100.3	9.3 5.3 9.3	Q02907 S76431 S76064
9	hypothetical protein slr0869 elongation factor EF-G [	R.SQQDTSIVETALGK. A R.VSVNGYETSNENY VR.I R.FGALTFTR.I K.GFDQSVVK.G	92.5 76.8	5.2 5	S74856 S76751
10	hypothetical protein sll1033 hypothetical protein sll0169	R.LYENFIDVGQR.Y R.VVDCQPLQPSVLK. V K.VLLSQQGDLLTR.L K.IYEVNQSNASSGSG R.M R.VEEISQPFTLGNQQ QK.G K.IDQIQVVNGPR.-	73.7 79.4	4.5 5	S74619 S76082
12	DNA ligase	R.NQLTENYQAQEK.M K.LNQELLQENTNLSD R.L R.NQQLSDQLSYVEQ NQAK.A	61.5	5.1	S75308 sll1583

		K.AVDEVLDQEEK.Q			
13	hypothetical protein 5-methylcytosine-specific restriction related enzyme [Bacillus cereus ATCC 10987]	R.EILLLAEQELSK.T K.INNTLSELQEQK.K K.INNTLSELQEQKK.I K.IIDKDITR.L K.RDSIEAEIK.N R.DSIEAEIK.N K.NLQAVQQNLESR.V K.ANPTLENLEIR.Q K.EISEQIQQGQVK.L R.NTYDALER.E K.NNVSELEQR.I K.QEISDLEDSAR.V R.ANVLTFGRPNELK.L K.YVVNYPGAEQDLQIR.R	84.2	6.2	BAD02128 slr6071
	DnaK protein	R.YGQFQWK.G K.IVDFLAGEFQK.A	67.6	4.9	C39025
	hybrid sensory kinase	K.EQSISITGASTLPDTEVDR.M K.NQADSLVYQAEK.Q K.ILIEYNESLQK.Q	94.1	5.2	S74654 sll1672
	exopolysaccharide export protein	R.NAEQQEVINPETSTEPK.N R.NAQEGTGLGLAITR.Q K.VLALTPGQPVYK.I R.VAELQAQM@LALQQQYK.F K.FFDPSQTAENLSSR.L	83.6	5	S74742
14	DnaK protein	K.QFAPEEISAQVLR.K K.IAGIEVLR.I R.IINEPTAASLAYGLDK.K K.IVDFLAGEFQK.A K.SALDEIVLVGGSTR.I R.IPAVQEVVK.K	67.6	4.9	C39025
	glutamate--ammonia ligase	K.EQSISITGASTLPDTEVDR.M K.NQADSLVYQAEK.Q	79.2	5.5	BAA18510.1

		K.LDPSVADAVATAM@R.D K.VLVQGEPDGSSFPNGGIR.D R.TSPFAFTGNR.F R.TTADALPVLK.E K.YIEDLFEK.T	75.2	4.7	S74372
	DnaK protein	K.TGVLTPVELESR.F K.IADLTNQMVGAVAK.L	84.2	6.1	BAD02128.1
	hypothetical protein	K.IADLTNQM@VGAVAK.L	78.9	5.3	slr6071 S75209
	DnaK protein	K.DAGTIAGLEVLR.I R.IINEPTAAALAYGLDK.Q	80.3	5.1	S76766
	oligopeptidase A	K.EISEQIQQGQVK.L K.YVVNYPGAEQDLQIR.R K.IAGLEVLR.I R.IINEPTAASLAYGLDQ.G K.AAEQEFADLQK.F			
16	dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase complex	K.IVSWTK.S K.VDLATIAGTGPHGR.I K.VDLATIAGTGPHGR.I K.PVTASIAAPSAPAPK.T R.VTSTPSVPVGQTVPLTTFQK.A K.ALVQNM@VAAM@AAPTFR.V R.VGYTITTDGLDQLYK.Q	44.9 67.1	6.2 8.7	S76485 S77114 sll1770
	BC1-like [Synechocystis sp. PCC 6803]gi|7450863|pir||S77114 ABC transporter sll1770 –	K.GVTMTALLAK.A R.PQVVANEEGLIGTK.R K.AQLHTGEDVVVK.V R.IVNSLVALGALK.E R.QAVQVGNSALGLPR.R	57.7 68.5	5.1 5.4	Q05972 S76378
	60kD chaperonin 1		67.7	6.2	S75115
	cell division protein;	K.DNTTIVAEGNEAAVK.S			

	FtsH	R.IAENAGQNGAVISE R.V	69.1	5.7	S75808
	acetohydroxy acid synthase	R.FLEYVDAGR.I R.ITSVDLYENGR.T R.TAIVQVSDPEVDR.T	67.3	5.2	S74970
	nodulation protein	R.IVTEAFHLASTGR.P R.APDVPIVGDVR.H			
	cell division protein; FtsH	K.DGVIVAAVQEER.F K.ALAELGDCK.T K.GVLLVGPPGTGK.T R.VRDLFEQAK.A			
17	trigger factor	K.GSDFEVTLEDGR.F R.LVAQTAMELER.M R.LVNFVESSLTESK.V	52.6	4.4	BAA10 868.1
	circadian rhythm protein	K.DSIILATGATGTGK. T R.AILFAYEESR.A	58.3	6.6	S76850
	dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase complex	R.VTSTPSVPVGQTVP LTTFQK.A (2X good peptide)	44.9	6.2	S76485
18	ATP synthase a subunit	K.TTGQIAQIPIGDAM @VGR.V R.VVDSLGRPIDGK.G R.LLESPAPGIIER.K K.STLVIYDDLSK.Q K.VTEFAQGLR.D K.YVEIINSSK.A K.YVEIINSSK.A K.ALTDEAETLLK.E	54	5.1	CAA41 135.1
	Zeta-carotene desaturase precursor	R.IGELDFR.F K.AFFTTSQLDTK.D K.IANSIALATSPIVR.G R.VTGLIINDGVETK.T	54.4	5.8	P74306

19	ferredoxin-NADP oxidoreductase	R.SGSTFITVPLK.R K.VLENYPLVR.E R.LYSIASTR.H K.SENILYKDDLEK.M	46.2	6.2	
	ferredoxin--nitrite reductase	K.LESCGLTSVQSGM@DNVR.N R.LADTYGSGEVR.L R.SVVSCTGAQFCK.F	55.6	6.6	S75713
	protochlorophillide reductase subunit; ChlN	K.LIGAPFPIGPDGTR.A R.YQAAELALLEK.T R.NSQLGELGWDK.L	52.5 73.1	5.3 8.6	JT0601 S74621
	carbon dioxide concentrating mechanism protein; CcmM	K.SAPVSSAGGSSAGGLTPEVIATVR.G R.LDNSVVTQVR.S			
20	hypothetical protein sll0245	K.STLFNALVANAK.A K.LAEISQSVK.V R.EVDAIVHVVR.C K.APQAAGVIHTDFER.G	39.3	4.8	S75088
	argininosuccinate synthetase	K.AIADTPDEPEYVDIGFEK.G R.LNEIAGNHGVGR.L R.DLESLTQTADVTHYK.N	44.5 44.1	5.1 5.1	S76929 S74347
	hypothetical protein slr0049	K.TAALDAFQVSDTVK.L R.IDEVEYQGQK.I	50.4	6.7	
	photosystem II chlorophyll a-binding protein psbC - Synechocystis sp. (strain PCC 6803)	R.LINLSGK.L R.SPSGEIIFGGETM@R.F			

1	light repressed protein	K.VDVHLSVAR.N K.HKAEVTVYANGTVIR.A K.AEVTVYANGTVIR.A R.APELPSEVLR.M R.APELPSEVLR.M R.NKDTDEINVIYIR.N R.NHGGYGVIQPHQA	21.9	6.5	S76493

		S.-			
2	elongation factor Tu	K.AVDDYIDTPER.E K.VGEEISIVGIK.D	43.7	5.3	S75862
3	glutathione S-transferase	R.KYPENSLLPHDPVQR.G K.YPENSLLPHDPVQR.G R.VAMVGALNQNPGLR.A K.IPGGNYLNIAQELK.G	29.8	5.7	S76871
4	dihydrolipoamide acetyltransferase component	R.VPTPNVSVVDLKIIAKK.A	44.9	6.2	S76485
5	fructose-1,6-bisphosphate aldolase	K.TQVDALAVAIGTSHGAYK.F R.YQQFWTAGNASK.I	39.0	5.7	S76332
6	The same as 5				
7	The same as 5	K.TQVDALAVAIGTSHGAYK.F R.KPTGEVLAISR.I R.KPTGEVLAISR.I			
9, 10, 12 14, 19 20, 21	glyceraldehyde-3-phosphate dehydrogenase	K.LDADISADENSITVNGK.T K.VLITAPGK.G K.GPNIGTYVVGVNAHEYK.H R.AAAVNIVPTSTGAAK.A R.VVDLAEIVAK.N	36.5	6.4	S54141
13	DNA polymerase III beta subunit	R.QSDLSSGLSLVSR.A R.KLEGAYPAYDQLIPR.Q	42.1	4.9	S74720
14	LysR transcriptional regulator	K.AQLTEAGHLLLNYGEK.I K.FISLDSQSTIR.K R.EVLPQFSTHPDALDPER.L	38.0	5.8	S75235
17	serine hydroxymethyltransferase	K.AVAFGEALKPEFK.V K.VGDQLLGEINITANK.N K.NTVPFDPESPFVTSGLR.L R.LLSPEDEGVKADC*LR.R	46.3	6.3	S75210
23	hypothetical protein slr0244	R.DYPEGELILAR.V R.VNPDLKPDLLPLSR	31.2	5.2	S74555

		.Q R.QEIEENPVLAPAIA K.A R.VNAPC*PVLLTR.K			
24	elongation factor TS	K.AETNFAEEVAAAA K.G	24.2	5.4	S75585
25	aspartate beta-semialdehyde dehydrogenese	K.GCDLVLASAGGSTS K.R K.AGAVMVDNSSAFR. M K.AGAVM@VDNSSAF R.M	30.6	5.8	sp|Q55 512
27	ATP-dependent Clp protease proteolytic	R.IVYLGM@PLFSSDE VK.Q R.ASLPHATIVLNQNR. T R.TGAQGQATDIQIR.A K.QTM@LEILSLNTGQ TQEK.L R.TFYLTPAQAK.E R.VLESPAELPKPM@A VI.-	24.9	5	sp|P744 66
28	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)				
29	hypothetical protein slr0552	K.ELDPTLVNEQFLK.F D.PTLVNEQFLK.F K.FSGIVSNEWELNQQ PVVSK.A K.AGSQLVFK.N K.NGLSIVAQPR.S K.NGLSIVAQPR.S L.SIVAQPR.S R.SLTFLEGMNDK.T R.SLTFLEGM@NDK.T K.TAEVVTVGK.V K.LPNAQYNGVVVTP K.C N.AQYNGVVVTPK.C K.CLIPLPDQNDGAR.K R.KFITGTLLASGAWQ DLGK.A L.ASGAWQDLGK.A K.APVQAAVEFTYLLE GCQFNLK.V K.VNQATLQIPDR.Q F.AGNFNYSLNNPNPQ	26.7	5	slr0552

		ER.V K.QYIEAWQSDLDTFR .G K.FLAEQQPQTVFG.-			
30	unknown protein	K.NGLSIVAQPR.S R.SLTFLEGM@NDK. T K.LPNAQYNGVVVTP K.C K.VNQATLQIPDR.Q	26.7	5	S76025
32	UDP-N- acetylglucosamine pyrophosphorylase	M.VAVAVLAAGK.G R.SDVEFVEQR.E C.LQDYQGDLLVLNG DAPLLR.S R.SETLENLLATHQR. H K.QLAAANDILQNR.I K.SVIGAQSNVAHLS Y.L R.VNVGAGTITANYD GVSK.H R.DVPAGSLAIAR.P	48.9	6.1	S76009
33	RecA gene product	K.ALNAALAQIER.S R.AEIEGEMGDTSVG SQAR.L R.AEIEGEM@GDTSV GSQAR.L K.IGVTYGSPEVTTG GNALK.F R.M@GCTIDLAEK.C K.GAWYSYNGENIA QGR.D K.YLEENPEIAATIDQ QVR.E	37.8	5.2	BAA18 857.1
33	fructose-1,6- bisphosphate aldolase	R.LAITAAFR.E R.YQQFWTAGNASK. I	39.0	5.7	Q55664
34	30S ribosomal protein S1	K.TLEMVVTGTNK.G K.TLEM@VVTGTNK. G K.GGVVGDVEGLR.G K.DNMDALVGQVLK. A	33.8	5.2	S75667

		K.DNM@DALVGQVLK.A K.AHILEANQDNNK.L K.IAAGNIYEGK.V R.ILETYPGELVEK.F K.FDEMMADAPNR.L			
35	chloroplast membrane-associated 30 kD protein	R.ANLNDLVSK.A K.VLEQAVIDM@QEDLVQLR.Q K.LALTNGEENLAR.E K.SLTDTAAAYQTQLAQQR.T K.ANAELQQTLGGLGTSSATSAFER.M	28.9	5	sll0617
37	uridine monophosphate kinase	R.VLTAIAM@QEVAEPYIR.R K.VMDSTAIALCK.D K.VM@DSTAIALCK.D	27.7	5.3	S76429
38	rehydrin	K.VIALSVDDVESHK.G K.VSDLYGMIHPN.A K.VSDLYGM@IHPN.A N.ALNNLTVR.S R.SVFIIDPAK.K R.SVFIIDPAK.K R.LTFTYPASTGR.N R.NFDEILR.V K.CVVVPSISTEDAK.V K.GVEEIKPYLR.L	23.6	5.3	S77532
39	As 27				
40	anthranilate synthase component II	R.NDQISLEEVK.S	17.9	5.9	S74362
41	ATP-dependent protease; ClpP	R.IVFLGQEVR.D R.IM@IHQPLGGAQGQATDIEIQAK.E K.SLEEITADTER.D K.EYGLIDQVINR.R	21.7	5	S75989
42	plastoquinol--plastocyanin reductase	M.TQISGSPDVPDLGR.R K.YLIPPSSGGSGGGVTAK.D K.VTEFLASHNAGDR.V	19.0	5.1	P26290
43	N-acetylglutamate kinase	M.SSTQDYIGEEAATR.V K.VGIEPQFK.D K.ELVNLINQAGGK.A K.DVGFVGEVSSVDAR.V	31.5	6.3	S77509

		K.LILLTDTR.G R.ELIGSGIVAGGMIPK.V R.ELIGSGIVAGGM@IPK.V			
44	glyceraldehyde-3-phosphate dehydrogenase				
45	Three different glyceraldehyde-3-phosphate dehydrogenase				
46	glyceraldehyde-3-phosphate dehydrogenase				
47	aspartate kinase	K.FGGTSVGTVER.I R.GGSDTSAVALAAALK.A K.AVDGVEYDADQAK.V R.SYPEADQEAEIIVEK.G K.IAIAGAGM@IGR.P R.GVALDQDQAQIAIR.H	63.5	5.4	S76764
48	elongation factor EF-G	K.ALQSLSEEDPTFR.V K.VEANVGAPQVAYR.E	75.4	5	sll1098
52	Dihydrolipoamide dehydrogenase	R.DIETYTGVFATK.I K.AGSPVEIELTDAK.T K.NLGLETVGVETDR.R R.GFIEVNDQM@QVIK.D	50.8	5.5	CAA88451.1
53	hypothetical protein sll1218	K.VLVIGATGETGK.R K.VLVIGATGETGKR.V K.AAIAGCTVVINAAGAR.P R.NLVDIAK.A K.VAEACVESLFSPSAK.N	23.5	6.2	sll1218
54	phosphoribosyl formylglycinamidine synthase	R.DIATVTAGLLDQPTR.F K.GYQSQQVITLPIAHGEGR.Y K.ALEDNEQILFR.Y	24.4	6	Q55843
55	SOS function	K.GVSVIGELK.G	22.7	6.1	S74809

	regulatory protein	K.GGELVEADAEEVEK.I R.SVTGEEEIEDGELVAASIK.G			
56	heme oxygenase	M.SVNLASQLR.E K.ISAAGQAYVDR.V R.YLGDLSGGQILK.K R.QAMNDLPIDQATAER.I K.M@FNELEGNLIK.A K.AIGIM@VFNSLTR.R	27.1	6.7	S74713
57	unknown protein	R.GNVICIQR.R R.TYLQTVSPLGK.V	28.27	6.4	slr1742
58	glyceraldehyde-3-phosphate dehydrogenase				

APPENDIX II

SCRIPTS FOR CALCULATING PROTEIN PROPERTIES

All scripts were written in Python language. The Integrated Development Environment used was "Eclipse" (Eclipse Foundation). The ".txt" files in the f = open() are reference, which are the protein name with the sequence. The ".txt" files in the f = open(, "w") are the new file created for writing proteins name along with the calculated properties.

A. Script for Calculating Molecular Weight

```
f = open('synecho aa list (after cutoff).txt')
lines = f.readlines()
f.close()

def molWeight(aminoacid):
    aaMW=[]
    MW=[]
    molecularWeight=0
    tempMW=0

aaList=['A','R','N','D','C','E','Q','G','H','I','L','K','M','F
','P','S','T','W','Y','V']

mwList=[71.0788,156.1875,114.1038,115.0886,103.1388,129.1155,1
28.1307,57.0519,137.1411,113.1594,113.1594,128.1741,131.1926,1
```

```
47.1766,97.1167,87.0782,101.1051,186.2132,163.1760,99.1326]
    for e in aminoacid:
        if e[0]==">":
            aaMW.append(e[:-1])
            molecularWeight=0
            tempMW=0
        else:
            for aa in e:
                if aa in aaList:
                    location=aaList.index(aa)
                    tempMW+=mwList[location]
            molecularWeight=tempMW+18
            aaMW.append(molecularWeight)
            MW.append(molecularWeight)
    return aaMW, MW
result1, result2=molWeight(lines)


f=open("synecho MW for cut off proteome.txt","w")
i=1
for res in result1:
    if i%2 == 0:
        f.write(str(res))
        f.write('\n')
    else:
        f.write(str(res))
        f.write("<       ")
    i+=1
f.close()
```

B. Script for Calculating Isoelectric Point

```
f = open('synecho aa list (after cutoff).txt')
proteome = f.readlines()
f.close()

f = open('synecho Hsp16.6 substrate aa list.txt')
substrates = f.readlines()
f.close()

aaNotation=['R','K','H','D','E','C','Y']
pKaReference=[12.4,10.5,6.00,3.86,4.25,8.33,10.0]

def pKa(sequence):
    pKa=[9.69,2.34]
    for e in sequence:
        if e in aaNotation:
            location=aaNotation.index(e)
            Rgroup=pKaReference[location]
            pKa.append(Rgroup)
    return pKa

pKa1=[9.69,2.34,12.4,10.5,6.00,3.86,4.25,8.33,10.0]
positiveList=[9.69,10.5,12.4,6.00]
negativeList=[2.34,3.86,4.25,8.33,10.0]
def pI(protein):
```

```
pKaList=pKa(protein)
positive=0
negative=0
pH=7
for e in pKa1:
    if e in positiveList:
        Ni1=pKaList.count(e)
        positive+=Ni1*(10**e)/(10**pH+10**e)
    if e in negativeList:
        Ni2=pKaList.count(e)
        negative+=Ni2*(10**pH)/(10**pH+10**e)
netCharge=positive-negative
tempZ=abs(netCharge)
positive=0
negative=0
if netCharge>0:
    pH+=.01
    while pH>=7 and pH<=10.5:
        for e in pKa1:
            if e in positiveList:
                Ni1=pKaList.count(e)
                positive+=Ni1*(10**e)/(10**pH+10**e)
            if e in negativeList:
                Ni2=pKaList.count(e)
                negative+=Ni2*(10**pH)/(10**pH+10**e)
        netCharge=positive-negative
        if abs(netCharge)<tempZ:
            tempZ=abs(netCharge)
```

```
                pI=pH
            pH+=.01
            positive=0
            negative=0
    if netCharge<0:
        pH-=.01
        while pH<=7 and pH>=3.5:
            for e in pKa1:
                if e in positiveList:
                    Ni1=pKaList.count(e)
                    positive+=Ni1*(10**e)/(10**pH+10**e)
                if e in negativeList:
                    Ni2=pKaList.count(e)
                    negative+=Ni2*(10**pH)/(10**pH+10**e)
            netCharge=positive-negative
            if abs(netCharge)<tempZ:
                tempZ=abs(netCharge)
                pI=pH
            pH-=.01
            positive=0
            negative=0

    return pI


def pIList(protein):
    pIList=[]
    for e in protein:
```

```
        if e[0]=='>':
            pIList.append(e[:-1])
        else:
            pIList.append(pI(e))
    return pIList

result=pIList(proteome)

f=open("synecho pI for cut off proteome.txt","w")
i=1
for res in result:
    if i%2 == 0:
        f.write(str(res))
        f.write('\n')
    else:
        f.write(str(res))
        f.write("<       ")
    i+=1
f.close()
```

C. Script for Calculating the Percentage of Charged, Negatively Charged, Positively Charged Residues in Each Protein

```
f = open('synecho Hsp16.6 substrate aa list.txt')
proteome = f.readlines()
f.close()
```

```
def PositiveResiduesPercentage(protein):


    R=0

    H=0

    K=0



    for e in protein:

            proteinLength=len(protein[:-1])

            for aa in e:

                if aa=='R':

                    R+=1

                if aa=='H':

                    H+=1

                if aa=='K':

                    K+=1

    RPercentage=float(R)/float(proteinLength)

    HPercentage=float(H)/float(proteinLength)

    KPercentage=float(K)/float(proteinLength)

    result=RPercentage+HPercentage+KPercentage

    positivePercentage=round(result, 4)

    return positivePercentage


def PosiResPerList(proteinList):

    positive=0 #the positive residue percentage for each

protein
```

```
    positiveList=[]#the final list of positive residue
percentage for multiple proteins
    for e in proteinList:
        if e[0]=='>':
            positiveList.append(e[:-1])
        else:
            positive=PositiveResiduesPercentage(e)
            positiveList.append(positive)
    return positiveList

result1=PosiResPerList(proteome)

f=open("synecho positive charge percentage for Hsp16.6
substrates.txt","w")
i=1
for res in result1:
    if i%2 == 0:
        f.write(str(res))
        f.write('\n')
    else:
        f.write(str(res))
        f.write("<       ")
    i+=1
f.close()

def NegativeResiduesPercentage(protein):
    D=0
    E=0
```

```
    proteinLength=0


    for e in protein:
        if e[0]!='>':
            proteinLength=len(protein[:-1])
            for aa in e:
                if aa=='D':
                    D+=1
                if aa=='E':
                    E+=1
    DPercentage=float(D)/float(proteinLength)
    EPercentage=float(E)/float(proteinLength)
    result=DPercentage+EPercentage
    negativePercentage=round(result, 4)
    return negativePercentage


def NegaResPerList(proteinList):
    negative=0 #the negative residue percentage for each
protein
    negativeList=[]#the final list of negative residue
percentage for multiple proteins
    for e in proteinList:
        if e[0]=='>':
            negativeList.append(e[:-1])
        else:
            negative=NegativeResiduesPercentage(e)
            negativeList.append(negative)
    return negativeList
```

```
result2=NegaResPerList(proteome)


f=open("synecho negative charge percentage for Hsp16.6
substrates.txt","w")
i=1
for res in result2:
    if i%2 == 0:
        f.write(str(res))
        f.write('\n')
    else:
        f.write(str(res))
        f.write("<       ")
    i+=1
f.close()


def chargedResPerList(proteinList):
    charged=0 #the negative residue percentage for each
protein
    chargedList=[]#the final list of negative residue
percentage for multiple proteins
    for p in proteinList:
        if p[0]=='>':
            chargedList.append(p[:-1])
        else:

charged=PositiveResiduesPercentage(p)+NegativeResiduesPercenta
ge(p)
```

```
            chargedList.append(charged)
    return chargedList
result3=chargedResPerList(proteome)
f=open("synecho charged residue percentage for Hsp16.6
substrates.txt","w")
i=1
for res in result3:
    if i%2 == 0:
        f.write(str(res))
        f.write('\n')
    else:
        f.write(str(res))
        f.write("<       ")
    i+=1
f.close()
```

D. Script for Calculating the Percentage of Hydrophobic Residues in Each Protein

```
f = open('synecho aa list (after cutoff).txt')
proteome = f.readlines()
f.close()

def HydrophobicResiduesPercentage(singleProtein):
    V=0
    I=0
    L=0
```

```
M=0
F=0
W=0
C=0
for e in singleProtein:
    proteinLength=len(singleProtein[:-1])
    for aa in e:
        if aa=='V':
            V+=1
        if aa=='I':
            I+=1
        if aa=='L':
            L+=1
        if aa=='M':
            M+=1
        if aa=='F':
            F+=1
        if aa=='W':
            W+=1
        if aa=='C':
            C+=1
VPercentage=float(V)/float(proteinLength)
IPercentage=float(I)/float(proteinLength)
LPercentage=float(L)/float(proteinLength)
MPercentage=float(M)/float(proteinLength)
FPercentage=float(F)/float(proteinLength)
WPercentage=float(W)/float(proteinLength)
CPercentage=float(C)/float(proteinLength)
```

```
result=VPercentage+IPercentage+LPercentage+MPercentage+FPercen
tage+WPercentage+CPercentage
    hydrophobicPercentage=round(result, 4)
    return hydrophobicPercentage

def HydroResPerList(proteinList):
    hydrophobic=0 #the positive residue percentage for each
protein
    hydrophobicList=[]#the final list of positive residue
percentage for multiple proteins
    for e in proteinList:
        if e[0]=='>':
            hydrophobicList.append(e[:-1])
        else:
            hydrophobic=HydrophobicResiduesPercentage(e)
            hydrophobicList.append(hydrophobic)
    return hydrophobicList

result1=HydroResPerList(proteome)




f=open("synecho hydrophobic percentage for cutoff
proteome.txt","w")
i=1
for res in result1:
    if i%2 == 0:
```

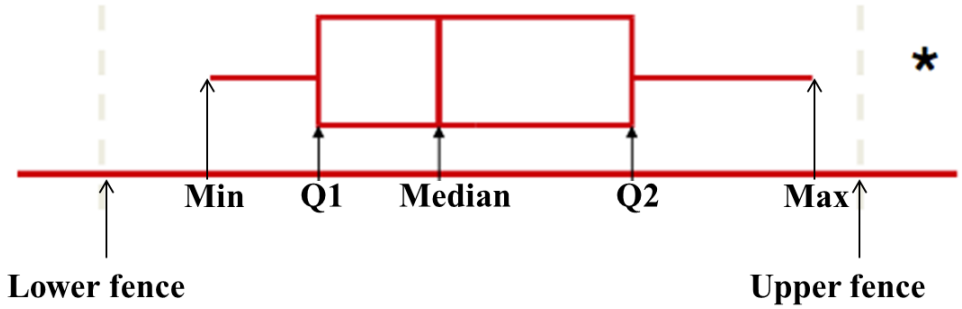
```
            f.write(str(res))
            f.write('\n')
        else:
            f.write(str(res))
            f.write("<        ")
        i+=1
    f.close()
```

APPENDIX III

INTRODUCTION OF BOXPLOT STATISTICS

(Introduction to Probability and Statistic 13th edition, by William Mendenhall, Robert J. Beaver, Barbara M. Beaver)

All boxplots shown in this thesis were generated by Minitab 16.



Min: minimum

Q1: larger than 25% and less than 75% of the ordered measurements

Median: the middle value of the ordered measurements

Q3: larger than 75% and less than 25% of the ordered measurements

Max: maximum

Lower fence: Q1-1.5(Q3-Q1)

Upper fence: Q3+1.5(Q3-Q1)

* outlier

APPENDIX IV

INTRODUCTION OF INTERVAL ESTIMATION

STATISTICS AND P-VALUE

(Introduction to Probability and Statistic 13th edition, by William Mendenhall, Robert J. Beaver, Barbara M. Beaver)

Interval estimation: two numbers are calculated to create an interval within which the parameter is expected to lie.

e.g.

Group 1

True mean: μ_1

Sample mean: $\overline{x_1}$

Sample standard deviation: s_1

Group 2

True mean: μ_2

Sample mean: $\overline{x_2}$

Sample standard deviation: s_2

Confidence interval for μ_1- μ_2

$$(\overline{x_1} - \overline{\overline{x_2}}) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

α is the significance level.

e.g. if α=0.01, it is 99% confidence.

Z value is referred to the normal distribution table.

p-value：

p-value<0.01, very strong difference

0.01<p-value<0.05, strong difference

0.05<p-value<0.1, have difference, but not strong

p-value>0.1, no difference

APPENDIX V

OVERLAPS AMONG HSP16.6, IBPB AND HSP20.2 SUBSTRATES

IbpB and Hsp20.2 substrates were referred to in previously published papers (Bepperling, Alte et al. 2012, Fu, Shi et al. 2013). "yes" means that IbpB or Hsp20.2 shares the homologous substrate with Hsp16.6.

1D		
Cyanobase accession number	IbpB substrates?	Hsp20.2 substrates?
sll1789	yes	yes
sll1787	yes	yes
slr1044		
sll0041		
sll1294		
slr1055		yes
slr0750		
slr0288		
slr0898		
slr0585		
sll0169		
slr0228		
slr1604		
slr0156		
slr0659	yes	
sll0170		
sll0058		
sll1932		
slr2076		
slr0322		
sll0043		
sll1672		
sll1561		
sll1178		
slr1463		yes
sll1841		yes
sll1031		
sll1180		

sll0533	yes	
slr0335		
slr2088		
sll1326	yes	
slr0940		
slr0869		
sll1033		
sll1583		
slr6071		
sll0923		
sll1770	yes	yes
slr0758		
sll0245		
slr1643		
slr0049		
sll0851		
2D		
Cyanobase accession number	IbpB substrates?	Hsp20.2 substrates?
sll1184		
slr1742		
slr1898		
slr0164		
slr0542	yes	
sll0998		
sll1626		
slr0657		
sll1099	yes	
sll1261		
sll1098		
sll1841		yes

sll1342		
slr0965		yes
sll0569	yes	
sll0947		yes
sll1545		
sll0018	yes	yes
sll1931		yes
slr0549		
sll0899		
slr1984	yes	yes
sll0144		
slr1198		
sll1316		
slr0520		
slr0244		
slr0552		
sll0617		
sll1218		
From previous paper		
Cyanobase accession number	IbpB substrates?	Hsp20.2 substrates?
slr1105		yes
slr1329		
slr1356	yes	yes
sll1818	yes	yes
sll1284		
sll0643		
sll1669		
slr2024		yes
slr1251		yes
slr0992		

APPENDIX VI

VARIOUS PROPERTIES OF THE 84 HSP16.6 ASSOCIATED PROTEINS

1D				
spot on gel	cyanobase number	uniprot number	Protein names	Gene names
1	sll1789	P73334	DNA-directed RNA polymerase subunit beta' (RNAP subunit beta') (EC 2.7.7.6)	rpoC2 sll1789
4	sll1787	P77965	DNA-directed RNA polymerase subunit beta (RNAP subunit beta) (EC 2.7.7.6)	rpoB sll1787
2, 3	slr1044	P73008	Methyl-accepting chemotaxis protein	mcpA slr1044
3,4	sll0041	Q55445	Putative methyl-accepting chemotaxis protein sll0041	sll0041
6	sll1294	P73173	PilJ protein	pilJ sll1294
3	slr1055	P73020	Mg-chelatase subunit; ChlH	chlH slr1055
19	slr0750	P28372	Light-independent protochlorophyllide reductase subunit N (DPOR subunit N) (LI-POR subunit N) (EC 1.18.-.-)	chlN slr0750
14	slr0288	P77970	Glutamate--ammonia ligase	glnN slr0288
19	slr0898	Q55366	Ferredoxin--nitrite reductase	nirA slr0898
20	slr0585	P77973	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)	argG slr0585
10	sll0169	H0PFL3	Putative uncharacterized protein sll0169	sll0169 SYNPCCP_2095
16	slr0228	Q55700	ATP-dependent zinc metalloprotease FtsH 2 (EC 3.4.24.-)	ftsH2 slr0228
16	slr1604	P72991	ATP-dependent zinc metalloprotease FtsH 3 (EC 3.4.24.-)	ftsH3 slr1604
8	slr0156	P74459	Chaperone protein ClpB 1	clpB1 slr0156
14	slr0659	P74571	Oligopeptidase A	prlC slr0659
13,14	sll0170	P22358	Chaperone protein dnaK2 (HSP70-2) (Heat shock 70 kDa protein 2) (Heat shock protein 70-2)	dnaK2 dnaK sll0170
14	sll0058	Q55154	Chaperone protein dnaK1 (HSP70-1) (Heat shock 70 kDa protein 1) (Heat shock protein 70-1)	dnaK1 sll0058
14	sll1932	P73098	Chaperone protein dnaK3 (HSP70-3) (Heat shock 70 kDa protein 3) (Heat shock protein 70-3)	dnaK3 sll1932
16	slr2076	Q05972	60 kDa chaperonin 1 (GroEL protein 1) (Protein Cpn60 1)	groL1 cpn60-1 groEL-1 groEL1 slr2076
3	slr0322	H0PFH5	CheA-like protein PilL/TaxAY3/Hik43	slr0322 SYNPCCP_2057
6	sll0043	H0PID4	CheA like protein Hik18	sll0043 SYNPCCP_2820
13	sll1672	H0PI99	Hybrid sensory kinase Hik12	sll1672 SYNPCCP_0230
7	sll1561	P74275	Delta-1-pyrroline-5-carboxylate dehydrogenase	putA sll1561
16	sll1178	P74178	Uncharacterized protein sll1178	sll1178
9	slr1463	P28371	Elongation factor G 1 (EF-G 1)	fusA fus slr1463
16,17	sll1841	P74510	Dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase complex	odhB sll1841
19	sll1031	P72758	Carbon dioxide concentrating mechanism protein; CcmM	ccmM sll1031
5	sll1180	P74176	HlyB family	hlyB sll1180

17	sll0533	Q55511	Trigger factor (TF) (EC 5.2.1.8) (PPIase)	tig sll0533
8*2	slr0335	Q55544	Phycobiliprotein ApcE (EC 4.-.-.-) (Phycobilisome LCM core-membrane linker polypeptide)	apcE slr0335
16	slr2088	P73913	Acetohydroxy acid synthase	ilvG slr2088
18	sll1326	P27179	ATP synthase subunit alpha (EC 3.6.3.14)	atpA sll1326
18	slr0940	P74306	Zeta-carotene desaturase (EC 1.3.5.6)	crtQ slr0940
9	slr0869	H0PM77	Putative uncharacterized protein slr0869	slr0869 SYNPCCP_1240
10	sll1033	H0PI64	Putative uncharacterized protein sll1033	sll1033 SYNPCCP_0195
12	sll1583	P73196	DNA ligase	lig sll1583
13,14	slr6071	Q6YRT3	Slr6012 protein (Slr6071 protein)	slr6012 slr6071
13	sll0923	P72877	Exopolysaccharide export protein	epsB sll0923
16	sll1770	P73627	Uncharacterized protein sll1770	sll1770
17	slr0758	P74646	Circadian clock protein kinase kaiC (EC 2.7.11.1)	kaiC slr0758
20	sll0245	P73886	Sll0245 protein probable GTP binding protein	sll0245
19	slr1643	Q55318	Ferredoxin--NADP reductase (FNR) (EC 1.18.1.2)	petH slr1643
20	slr0049	H0PG67	Putative uncharacterized protein slr0049	slr0049 SYNPCCP_2299
20	sll0851	F7UPA9	Photosystem II CP43 protein	psbC SYNGTS_1223
2D				
spot on gel	cyanobase number	uniprot number	Protein names	Gene names
29	sll1184	P72849	Heme oxygenase 1 (EC 1.14.99.3)	pbsA1 sll1184
30	slr1742	H0PM42	Putative uncharacterized protein slr1742	slr1742 SYNPCCP_1205
23	slr1898	P73326	Acetylglutamate kinase (EC 2.7.2.8) (N-acetyl-L-glutamate 5-phosphotransferase) (NAG kinase) (AGK)	argB slr1898
13	slr0164	P74466	Putative ATP-dependent Clp protease proteolytic subunit-like (Endopeptidase Clp-like)	clpR slr0164
21	slr0542	P54416	ATP-dependent Clp protease proteolytic subunit 1 (EC 3.4.21.92) (Endopeptidase Clp 1)	clpP1 slr0542
8	sll0998	P73123	Probable RuBisCO transcriptional regulator	rbcR sll0998
28	sll1626	P73722	SOS function regulatory protein	lexA sll1626
24	slr0657	P74569	Aspartokinase (EC 2.7.2.4)	lysC slr0657
2	sll1099	P74227	Elongation factor Tu (EF-Tu)	tuf tufA sll1099
11	sll1261	P74070	Elongation factor Ts (EF-Ts)	tsf sll1261
25	sll1098	P74228	Elongation factor G 2 (EF-G 2)	fusB fus sll1098
4	sll1841	P74510	Dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase complex	odhB sll1841

6	sll1342	P80505	Glyceraldehyde-3-phosphate dehydrogenase 2 (EC 1.2.1.59) (GAPDH 2)	gap2 sll1342
7	slr0965	P72856	DNA polymerase III subunit beta (EC 2.7.7.7)	dnaN slr0965
16	sll0569	P74737	Protein RecA (Recombinase A)	recA sll0569
1	sll0947	P74518	Light-repressed protein A homolog	lrtA sll0947
3	sll1545	P74665	Glutathione S-transferase	gst1 sll1545
5	sll0018	Q55664	Fructose-bisphosphate aldolase class 2 (FBP aldolase) (FBPA) (EC 4.1.2.13)	fbaA fda sll0018
9	sll1931	P77962	Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)	glyA sll1931
12	slr0549	Q55512	Aspartate-semialdehyde dehydrogenase (ASA dehydrogenase) (ASADH) (EC 1.2.1.11)	asd slr0549
15	sll0899	Q55504	Bifunctional protein GlmU [Includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23)	glmU sll0899
17	slr1984	P74142	30S ribosomal protein S1 homolog B	rps1b slr1984
19	sll0144	P74457	Uridylate kinase (UK) (EC 2.7.4.22) (Uridine monophosphate kinase) (UMP kinase) (UMPK)	pyrH sll0144
20	slr1198	H0PKZ1	Rehydrin	slr1198 SYNPCCP_0804
22	sll1316	P26290	Cytochrome b6-f complex iron-sulfur subunit 2 (EC 1.10.9.1)	petC2 sll1316
27	slr0520	Q55843	Phosphoribosylformylglycinamidine synthase 1 (EC 6.3.5.3)	purQ slr0520
10	slr0244	H0PI00	Putative uncharacterized protein slr0244	slr0244 SYNPCCP_0131
14	slr0552	H0PIM1	Putative uncharacterized protein slr0552	slr0552 SYNPCCP_2907
18	sll0617	Q55707	Uncharacterized protein sll0617	sll0617
26	sll1218	P74029	Ycf39 protein	ycf39 sll1218
From previous paper				
spot on gel	cyanobase number	uniprot number	Protein names	Gene names
	slr1105	P72749	GTP-binding protein TypA/BipA homolog	typA slr1105
	slr1329	P26527	ATP synthase subunit beta (EC 3.6.3.14) (ATP synthase F1 sector subunit beta) (F-ATPase subunit beta)	atpD atpB slr1329
	slr1356	P73530	30S ribosomal protein S1 homolog A	rps1A slr1356
	sll1818	P73297	DNA-directed RNA polymerase subunit alpha (RNAP subunit alpha) (EC 2.7.7.6)	rpoA sll1818
	sll1284	P73192	Serine esterase	sll1284
	sll0643	P72955	Urease accessory protein UreG	ureG sll0643
	sll1669	P72796	Shikimate kinase (SK) (EC 2.7.1.71)	aroK sll1669
	slr2024	H0PKN1	CheY subfamily protein Rre13	slr2024 SYNPCCP_0694
	slr1251	P73789	Peptidyl-prolyl cis-trans isomerase slr1251 (PPIase slr1251) (EC 5.2.1.8)	slr1251
	slr0992	P74516	Putative tRNA (cytidine(34)-2'-O)-methyltransferase (EC 2.1.1.207)	slr0992

1D						
Cyanobase number	MW (Da)	Length	pI (Expasy)	Extinction coefficient ($cm^{-1}M^{-1}$)	Membrane regions	structure available (PDB number)
sll1789	144770	1317	4.72	92920	no	no
sll1787	123358	1102	5.17	86390	no	no
slr1044	93206	869	4.28	57990	2	no
sll0041	108327	1000	4.69	55790	2	no
sll1294	103172	953	4.55	44640	2	no
slr1055	148547	1331	4.9	163330	no	no
slr0750	52471	469	5.23	59190	no	no
slr0288	79208	724	5.29	69160	no	no
slr0898	55923	502	6.17	49870	no	1GXI
slr0585	44483	400	5.05	54410	no	no
sll0169	79418	714	4.88	68590	1	no
slr0228	68494	627	5.29	37530	2	no
slr1604	67248	616	5.11	28710	2	no
slr0156	101385	898	5.23	51730	no	no
slr0659	80292	713	5.04	103890	no	no
sll0170	67612	636	4.72	12000	no	
sll0058	75174	692	4.6	27100	no	no
sll1932	86029	771	5.13	53840	no	no
slr2076	57650	541	5.01	14890	no	no
slr0322	120550	1095	4.64	49890	no	no
sll0043	153091	1402	4.56	99010	no	no
sll1672	94090	834	5.07	95190	8	no
sll1561	110022	990	5.46	109660	no	no
sll1178	69139	615	5.47	83600	no	no
slr1463	76748	695	4.9	51520	no	no
sll1841	44897	433	5.85	23020	no	no
sll1031	73119	687	8.83	71490	no	no
sll1180	112080	1011	5.51	96140	4	no

(Membrane regions were calculated from http://www.cbs.dtu.dk/services/TMHMM-2.0/, orange shading indicates that the protein has more than zero membrane region.)

sll0533	52607	471	4.31	20580	no	no
slr0335	100294	896	9.25	73750	no	3OSJ
slr2088	67727	621	5.75	60940	no	no
sll1326	53963	503	4.96	26410	no	no
slr0940	54368	489	5.55	63350	no	no
slr0869	92526	812	5.11	53540	no	no
sll1033	73706	668	4.31	65830	no	no
sll1583	61464	562	5.06	36250	no	no
slr6071	84201	730	5.87	50920	1	no
sll0923	83635	756	4.9	68820	1	no
sll1770	67125	585	8.78	91110	2	no
slr0758	58292	519	6.19	35350	no	1WWJ
sll0245	39314	363	4.77	18140	no	no
slr1643	46359.54	413	5.72	56410	no	no
slr0049	44058.15	398	4.88	59420	no	no
sll0851	50303.04	460	6.11			
2D						
Cyanobase number	MW (Da)	Length	pI (Expasy)	Extinction coefficient ($cm^{-1}M^{-1}$)	Membrane regions	structure available (PDB number)
sll1184	27050	240	6.24	19890	no	
slr1742	28273	256	5.98	27910	no	no
slr1898	31524	297	5.89	13850	no	no
slr0164	24880	225	4.86	20010	no	no
slr0542	21740	198	4.92	6640	has	no
sll0998	38014	345	5.54	16170	no	no
sll1626	22743	203	5.84	22190	no	no
slr0657	63530	600	5.19	17340	no	3L76
sll1099	43732	399	5.16	14200	no	no
sll1261	24230	218	5.37	9890	no	no
sll1098	75427	691	4.94	38150	no	no
sll1841	44897	433	5.85	23020	no	no

sll1342	36491	337	6	44980	no	2D2I
slr0965	42085	391	4.7	6520	no	no
sll0569	37804	354	5.13	16290	no	no
sll0947	21893	191	6.07	10240	no	no
sll1545	29763	271	5.43	23020	no	no
sll0018	38971	359	5.46	25820	no	no
sll1931	46258	427	5.88	20370	no	no
slr0549	36639	338	5.72	29760	no	no
sll0899	48920	456	5.68	33270	no	no
slr1984	33793	305	5.12	20460	no	no
sll0144	25592	260	5.14	7920	no	no
slr1198	23558	211	5.08	30800	no	no
sll1316	18996	180	4.87	29640	has	no
slr0520	24427	224	5.57	18260	no	no
slr0244	31202	284	5.12	16410	no	no
slr0552	26719	244	4.91	22430	no	no
sll0617	28904	267	4.95	6970	no	no
sll1218	23533	219	5.91	16860	no	no
From previous paper						
Cyanobase number	MW (Da)	Length	pI (Expasy)	Extinction coefficient ($cm^{-1}M^{-1}$)	Membrane regions	structure available (PDB number)
slr1105	66013.26	597	4.98	25640	no	no
slr1329	51733.02	483	4.89	22360	no	no
slr1356	36570.06	328	4.55	18970	no	no
sll1818	35003.77	314	4.72	23170	no	no
sll1284	22209.53	204	5.08	31630	no	no
sll0643	22012.55	206	5.09	12480	no	no
sll1669	20697.69	189	4.63	20940	no	no
slr2024	20232.81	180	6.34	28420	no	no
slr1251	18534.94	171	5.34	15610	no	no
slr0992	17033.44	153	5.64	39080	no	no

1D						
Cyanobase number	Oligomer state	No. of negatively charged residues	No. of positively charged residues	binding nucleotide phosphate	soluble protein or not	up-regulated after exposure to heat shock
sll1789		222	151		Yes	no
sll1787		171	132		no	no
slr1044		136	62		Yes	no
sll0041		137	82		no	no
sll1294		135	83		no	no
slr1055		181	127		no	no
slr0750		60	49		no	no
slr0288		93	71		no	no
slr0898	monomeric	66	61		no	no
slr0585		60	45		Yes	Yes
sll0169		95	65		no	no
slr0228		84	71	ATP	no	no
slr1604		81	67	ATP	no	no
slr0156		155	125		Yes	no
slr0659		95	65		no	Yes
sll0170		100	71		Yes	Yes
sll0058		101	61		no	no
sll1932		128	111		no	no
slr2076		84	66		Yes	Yes
slr0322		180	113		no	no
sll0043		204	105		no	no
sll1672		97	70		no	no
sll1561		127	103		no	no
sll1178		75	60		no	no
slr1463		112	78		Yes	Yes
sll1841		42	38		no	Yes
sll1031		63	70		no	Yes
sll1180		104	91		no	no

("soluble proteins or not" was referred to (Simon, Hall et al. 2002), "up-regulated after exposed to heat shock" was referred to (Slabas, Suzuki et al. 2006))

sll0533		109	55		Yes	no
slr0335		94	111		Yes	no
slr2088		66	50		no	no
sll1326		58	46		Yes	Yes
slr0940		57	49		no	no
slr0869		118	92		no	no
sll1033		204	105		no	no
sll1583		82	63		no	Yes
slr6071		122	114		no	no
sll0923		93	69		no	no
sll1770		70	75		no	no
slr0758	tetramer	68	63	ATP	no	no
sll0245		57	39		no	no
slr1643		57	49	FAD, NADP	no	no
slr0049		52	30		no	no
sll0851		35	29		no	no
2D						
Cyanobase number	Oligomer state	No. of negatively charged residues	No. of positively charged residues	binding nucleotide phosphate	soluble protein or not	up-regulated after exposure to heat shock
sll1184		30	28		no	no
slr1742		29	26		no	no
slr1898		34	30		no	no
slr0164		25	18		Yes	no
slr0542		27	19		no	no
sll0998		42	36		no	no
sll1626		31	29		no	Yes
slr0657	homodimer	70	53		Yes	Yes
sll1099		67	50	GTP	Yes	Yes
sll1261		39	34		Yes	Yes
sll1098		105	74	GTP	no	Yes
sll1841		42	38		no	no

sll1342	tetramer	38	34	NAD	no	Yes
slr0965		51	33		Yes	no
sll0569		47	39	ATP	no	no
sll0947		29	25		no	no
sll1545		32	26		no	no
sll0018		45	32		no	Yes
sll1931		49	40		no	no
slr0549		37	31	NADP	no	no
sll0899		49	40		no	no
slr1984		44	30		no	no
sll0144		29	24	ATP & UMP	no	no
slr1198		29	21		Yes	no
sll1316		19	12		no	no
slr0520		23	17		no	no
slr0244		27	26		no	no
slr0552		24	18		Yes	Yes
sll0617		42	34		no	no
sll1218		23	22		no	no
From previous paper						
Cyanobase number	Oligomer state	No. of negatively charged residues	No. of positively charged residues	binding nucleotide phosphate	soluble protein or not	up-regulated after exposure to heat shock
slr1105		92	67	GTP	no	
slr1329		62	46	ATP	Yes	Yes
slr1356		63	34		Yes	Yes
sll1818		49	32		Yes	no
sll1284		19	11		no	no
sll0643		26	21	GTP	no	no
sll1669		26	15	ATP	no	no
slr2024		25	24		no	Yes
slr1251		22	16		no	Yes
slr0992		14	10		no	no

1D		
Cyanobase number	function category	function in Synechocystis (uniprot)
sll1789	RNA synthesis, modification, and DNA transcription	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
sll1787	RNA synthesis, modification, and DNA transcription	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
slr1044	Chemotaxis	required for the biogenesis of thick pilli
sll0041	Chemotaxis	phytochrome-like photoreceptor protein for positive phototaxis; homologous to methyl-accepting chemotaxis protein
sll1294	Chemotaxis	homologous to methyl-accepting chemotaxis protein (MCP) methyl-accepting chemotaxis protein
slr1055	Cobalamin, heme, phycobilin and porphyrin	chlH magnesium-protoporphyrin methyltransferase
slr0750	Cobalamin, heme, phycobilin and porphyrin	Uses Mg-ATP and reduced ferredoxin to reduce ring D of protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide) This reaction is light-independent.
slr0288	Glutamate family / Nitrogen assimilation	
slr0898	Glutamate family / Nitrogen assimilation	nirA ferredoxin--nitrite reductase
slr0585	Glutamate family / Nitrogen assimilation	ATP + L-citrulline + L-aspartate = AMP + diphosphate + N(omega)-(L-arginino)succinate
sll0169	Cell division	
slr0228	Cell division	Acts as a processive, ATP-dependent zinc metallopeptidase for both cytoplasmic and membrane proteins. Plays a role in the quality control of integral membrane proteins Plays a role in the selective replacement of photosystem II (PSII) protein D1 in the PSII repair cycle following visible-light and UV-B induced damage. If damaged D1 is not removed then new D1 cannot be inserted to restore the PSII reaction center. Seems to also degrade damaged and/or unassembled PSII proteins D2 and PsbB (CP47). May recognize D1 via its first 20 amino acids, as deletion of these prevents the PSII repair cycle. Also seems to degrade cytoplasmic GGPS, glucosylglycerol-phosphate synthase
slr1604	cell division protein FtsH	Acts as a processive, ATP-dependent zinc metallopeptidase for both cytoplasmic and membrane proteins. Plays a role in the quality control of integral membrane proteins

slr0156	Degradation of proteins, peptides, and glycopeptides	Part of a stress-induced multi-chaperone system, it is involved in the recovery of the cell from heat-induced damage, in cooperation with DnaK, DnaJ and GrpE. Acts before DnaK, in the processing of protein aggregates. Protein binding stimulates the ATPase activity; ATP hydrolysis unfolds the denatured protein aggregates, which probably helps expose new hydrophobic binding sites on the surface of ClpB-bound aggregates, contributing to the solubilization and refolding of denatured protein aggregates by DnaK
slr0659	Degradation of proteins, peptides, and glycopeptides	
sll0170	Chaperones	act as chaperone
sll0058	Chaperones	Acts as a chaperone
sll1932	Chaperones	Acts as a chaperone
slr2076	Chaperones	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions
slr0322	Regulatory functions	two-component hybrid sensor and regulator
sll0043	Regulatory functions	positive phototaxis protein, homologous to chemotaxis protein CheA, two-component hybrid histidine kinase
sll1672	Regulatory functions	two-component hybrid histidine kinase
sll1561	Amino acids and amines	proline oxidase
sll1178	Amino acids and amines	probable carbamoyl transferase
slr1463	Protein modification and translation factors	Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome
sll1841	Pyruvate dehydrogenase	
sll1031	CO2 fixation	carbon dioxide concentrating mechanism protein CcmM, putative carboxysome structural protein
sll1180	Transport and binding proteins	toxin secretion ABC transporter ATP-binding protein
sll0533	Protein and peptide secretion	Involved in protein export. Acts as a chaperone by maintaining the newly synthesized protein in an open conformation. Functions as a peptidyl-prolyl cis-trans isomerase
slr0335	Phycobilisome	This protein is postulated to act both as terminal energy acceptor (by its phycobilin-like domains) and as a linker polypeptide (by its repeats and arms) that stabilizes the phycobilisome core architecture. Has intrinsic bilin lyase activity
slr2088	Branched chain family	Belongs to the TPP enzyme family
sll1326	ATP synthase	Produces ATP from ADP in the presence of a proton gradient across the membrane. The alpha chain is a regulatory subunit
slr0940	Carotenoid	Catalyzes the conversion of zeta-carotene to lycopene via the intermediary of neurosporene. It carries out two consecutive desaturations (introduction of double bonds) at positions C-7 and C-7

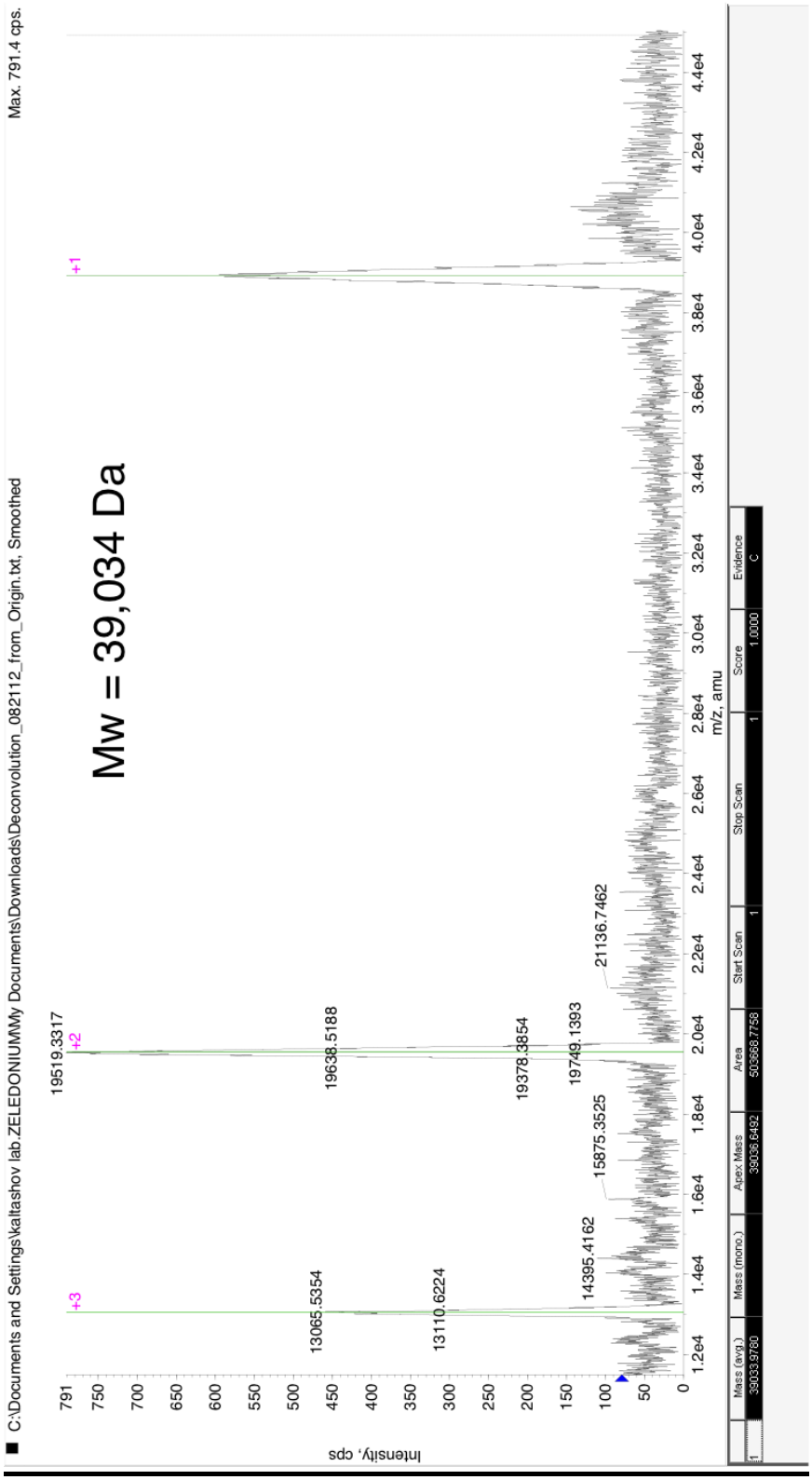
slr0869	Hypothetical	not found
sll1033	other	
sll1583	Unknown	ligase
slr6071	Unknown	
sll0923	Unknown	
sll1770	Hypothetical	Belongs to the protein kinase superfamily
slr0758	Other	Core component of the kaiABC clock protein complex, which constitutes the main circadian regulator in cyanobacteria. Binds to DNA. The kaiABC complex may act as a promoter-nonspecific transcription regulator that represses transcription, possibly by acting on the state of chromosome compaction
sll0245	Other	
slr1643	Photosynthesis and respiration	Soluble electron carriers
slr0049	Hypothetical	
sll0851	Photosynthesis and respiration	Photosystem II
2D		
Cyanobase number	function category	function in Synechocystis (uniprot)
sll1184	Cobalamin, heme, phycobilin and porphyrin	Catalyzes the opening of the heme ring with the release of iron. Key enzyme in the synthesis of the chromophoric part of the photosynthetic ante
slr1742	Cobalamin, heme, phycobilin and porphyrin	Not found
slr1898	Glutamate family / Nitrogen assimilation	Catalytic activity: ATP + N-acetyl-L-glutamate = ADP + N-acetyl-L-glutamate 5-phosphate.
slr0164	Degradation of proteins, peptides, and glycopeptides	Has lost the two conserved residues (Ser and His) proposed to be part of the active site. Therefore it could be inactive.
slr0542	Degradation of proteins, peptides, and glycopeptides	Cleaves peptides in various proteins in a process that requires ATP hydrolysis. Has a chymotrypsin-like activity. Plays a major role in the degradation of misfolded proteins
sll0998	Regulatory functions	transcription, DNA-dependent
sll1626	Regulatory functions	Catalytic activity: Hydrolysis of Ala-|-Gly bond in repressor lexA
slr0657	Amino acids and amines	Catalytic activity: ATP + L-aspartate = ADP + 4-phospho-L-aspartate

sll1099	Protein modification and translation factors	This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis
sll1261	Protein modification and translation factors	Associates with the EF-Tu.GDP complex and induces the exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-Tu.GTP complex up to the GTP hydrolysis stage on the ribosome
sll1098	Protein modification and translation factors	Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome
sll1841	Pyruvate dehydrogenase	Pyruvate dehydrogenase
sll1342	CO2 fixation	Glycolysis The chemical reactions and pathways resulting in the breakdown of a monosaccharide (generally glucose) into pyruvate, with the concomitant production of a small amount of ATP. Glycolysis begins with phosphorylation of a monosaccharide (generally glucose) on the sixth carbon by a hexokinase, and ends with the production of pyruvate. Pyruvate may be converted to ethanol, lactate, or other small molecules, or fed into the TCA cycle.
slr0965	DNA replication, restriction, modification, recombination, and repair	DNA polymerase III is a complex, multichain enzyme responsible for most of the replicative synthesis in bacteria. This DNA polymerase also exhibits 3' to 5' exonuclease activity. The alpha chain is the DNA polymerase
sll0569	DNA replication, restriction, modification, recombination, and repair	Can catalyze the hydrolysis of ATP in the presence of single-stranded DNA, the ATP-dependent uptake of single-stranded DNA by duplex DNA, and the ATP-dependent hybridization of homologous single-stranded DNAs. It interacts with LexA causing its activation and leading to its autocatalytic cleavage
sll0947	Adaptations and atypical conditions	Might modulate either transcription and/or translation
sll1545	Thioredoxin, glutaredoxin, and glutathione	transferase
sll0018	Glycolysis	Glycolysis Catalysis of the reaction: D-fructose 1,6-bisphosphate = glycerone phosphate + D-glyceraldehyde-3-phosphate.
sll1931	Serine family / Sulfur assimilation	Catalyzes the reversible interconversion of serine and glycine with tetrahydrofolate (THF) serving as the one-carbon carrier. This reaction serves as the major source of one-carbon groups required for the biosynthesis of purines, thymidylate, methionine, and other important biomolecules. Also exhibits THF-independent aldolase activity toward beta-hydroxyamino acids, producing glycine and aldehydes, via a retro-aldol mechanism
slr0549	Aspartate family	Catalyzes the NADPH-dependent formation of L-aspartate-semialdehyde (L-ASA) by the reductive dephosphorylation of L-aspartyl-4-phosphate
sll0899	Murein sacculus and peptidoglycan	Catalyzes the last two sequential reactions in the de novo biosynthetic pathway for UDP-GlcNAc. Responsible for the acetylation of Glc-N-1-P to give GlcNAc-1-P and for the uridyl transfer from UTP to GlcNAc-1-P which produces UDP-GlcNAc
slr1984	Ribosomal proteins: synthesis and modification	Binds mRNA.

sll0144	Pyrimidine ribonucleotide biosynthesis	Catalyzes the reversible phosphorylation of UMP to UDP
slr1198	Drug and analog sensitivity	Not found
sll1316	Cytochrome b6/f complex	Component of the cytochrome b6-f complex, which mediates electron transfer between photosystem II (PSII) and photosystem I (PSI), cyclic electron flow around PSI, and state transitions
slr0520	Purine ribonucleotide biosynthesis	Catalytic activity: ATP + N(2)-formyl-N(1)-(5-phospho-D-ribosyl)glycinamide + L-glutamine + H2O = ADP + phosphate + 2-(formamido)-N(1)-(5-phospho-D-ribosyl)acetamidine + L-glutamate
slr0244	Hypothetical	Not characterized
slr0552	Hypothetical	Not characterized
sll0617	Hypothetical	uncharacterized
sll1218	Hypothetical	nucleotide binding
From previous paper		
Cyanobase number	function category	function in Synechocystis (uniprot)
slr1105	Protein modification and translation factors	Not known; probably interacts with the ribosomes in a GTP dependent manner
slr1329	ATP synthase	Produces ATP from ADP in the presence of a proton gradient across the membrane. The catalytic sites are hosted primarily by the beta subunits
slr1356	Ribosomal proteins: synthesis and modification	Binds mRNA.
sll1818	RNA synthesis, modification, and DNA transcription	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
sll1284	Other	
sll0643	Other	Facilitates the functional incorporation of the urease nickel metallocenter. This process requires GTP hydrolysis, probably effectuated by UreG
sll1669	Aromatic amino acid family	Catalyzes the specific phosphorylation of the 3-hydroxyl group of shikimic acid using ATP as a cosubstrate
slr2024	Regulatory functions	
slr1251	Protein modification and translation factors	PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
slr0992	Aminoacyl tRNA synthetases and tRNA modification	Could methylate the ribose at the nucleotide 34 wobble position in tRNA

APPENDIX VII

MALDI ANALYSIS OF FBA



REFERENCES

Azia, A., R. Unger and A. Horovitz (2012). "What distinguishes GroEL substrates from other Escherichia coli proteins?" FEBS J **279**(4): 543-550.

Basha, E., G. J. Lee, L. A. Breci, A. C. Hausrath, N. R. Buan, K. C. Giese and E. Vierling (2004). "The identity of proteins associated with a small heat shock protein during heat stress in vivo indicates that these chaperones protect a wide range of cellular functions." J Biol Chem **279**(9): 7566-7575.

Basha, E., H. O'Neill and E. Vierling (2012). "Small heat shock proteins and alpha-crystallins: dynamic proteins with flexible functions." Trends Biochem Sci **37**(3): 106-117.

Bepperling, A., F. Alte, T. Kriehuber, N. Braun, S. Weinkauf, M. Groll, M. Haslbeck and J. Buchner (2012). "Alternative bacterial two-component small heat shock protein systems." Proc Natl Acad Sci U S A **109**(50): 20407-20412.

Clos, J. and S. Brandau (1994). "pJC20 and pJC40--two high-copy-number vectors for T7 RNA polymerase-dependent expression of recombinant genes in Escherichia coli." Protein Expr Purif **5**(2): 133-137.

Delbecq, S. P., S. Jehle and R. Klevit (2012). "Binding determinants of the small heat shock protein, alphaB-crystallin: recognition of the 'IxI' motif." EMBO J **31**(24): 4587-4594.

Dierick, I., J. Irobi, P. De Jonghe and V. Timmerman (2005). "Small heat shock proteins in inherited peripheral neuropathies." Ann Med **37**(6): 413-422.

Fernandez-Escamilla, A. M., F. Rousseau, J. Schymkowitz and L. Serrano (2004). "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins." Nat Biotechnol **22**(10): 1302-1306.

Fu, X., Z. Chang, X. Shi, D. Bu and C. Wang (2013). "Multilevel structural characteristics for the natural substrate proteins of bacterial small heat shock proteins." Protein Sci.

Fu, X., X. Shi, L. Yan, H. Zhang and Z. Chang (2013). "In vivo substrate diversity and preference of small heat shock protein IbpB as revealed by using a genetically incorporated photo-cross-linker." J Biol Chem **288**(44): 31646-31654.

Gefflaut, T., C. Blonski, J. Perie and M. Willson (1995). "Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects." Prog Biophys Mol Biol **63**(3): 301-340.

Gidalevitz, T., E. A. Kikis and R. I. Morimoto (2010). "A cellular perspective on conformational disease: the role of genetic background and proteostasis networks." Curr Opin Struct Biol **20**(1): 23-32.

Giese, K. C. and E. Vierling (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**(48): 46310-46318.

Giese, K. C. and E. Vierling (2004). "Mutants in a small heat shock protein that affect the oligomeric state. Analysis and allele-specific suppression." J Biol Chem **279**(31): 32674-32683.

Goldfarb, L. G., M. Olive, P. Vicart and H. H. Goebel (2008). "Intermediate filament diseases: desminopathy." Adv Exp Med Biol **642**: 131-164.

Graw, J. (2009). "Genetics of crystallins: cataract and beyond." Exp Eye Res **88**(2): 173-189.

Green, R. (2000). "Ribosomal translocation: EF-G turns the crank." Curr Biol **10**(10): R369-373.

Greenfield, N. J. (2006). "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions." Nat Protoc **1**(6): 2527-2535.

Hartl, F. U., A. Bracher and M. Hayer-Hartl (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.

Haslbeck, M. (2002). "sHsps and their role in the chaperone network." Cell Mol Life Sci **59**(10): 1649-1657.

Houry, W. A., D. Frishman, C. Eckerskorn, F. Lottspeich and F. U. Hartl (1999). "Identification of in vivo substrates of the chaperonin GroEL." Nature **402**(6758): 147-154.

Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata (1996). "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement)." DNA Res **3**(3): 185-209.

Kojima, K., K. Motohashi, T. Morota, M. Oshita, T. Hisabori, H. Hayashi and Y. Nishiyama (2009). "Regulation of translation by the redox state of elongation factor G in the cyanobacterium Synechocystis sp. PCC 6803." J Biol Chem **284**(28): 18685-18691.

Kojima, K., M. Oshita, Y. Nanjo, K. Kasai, Y. Tozawa, H. Hayashi and Y. Nishiyama (2007). "Oxidation of elongation factor G inhibits the synthesis of the D1 protein of photosystem II." Mol Microbiol **65**(4): 936-947.

Kriehuber, T., T. Rattei, T. Weinmaier, A. Bepperling, M. Haslbeck and J. Buchner (2010). "Independent evolution of the core domain and its flanking sequences in small heat shock proteins." FASEB J **24**(10): 3633-3642.

Lee, G. J., N. Pokala and E. Vierling (1995). "Structure and in vitro molecular chaperone activity of cytosolic small heat shock proteins from pea." J Biol Chem **270**(18): 10432-10438.

Lee, G. J., A. M. Roseman, H. R. Saibil and E. Vierling (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**(3): 659-671.

Lee, S., H. A. Owen, D. J. Prochaska and S. R. Barnum (2000). "HSP16.6 is involved in the development of thermotolerance and thylakoid stability in the unicellular cyanobacterium, Synechocystis sp. PCC 6803." Curr Microbiol **40**(4): 283-287.

Lee, S., D. J. Prochaska, F. Fang and S. R. Barnum (1998). "A 16.6-kilodalton protein in the Cyanobacterium synechocystis sp. PCC 6803 plays a role in the heat shock response." Curr Microbiol **37**(6): 403-407.

MacRae, T. H. (2000). "Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas." Cell Mol Life Sci **57**(6): 899-913.

Malakhov, M. P., M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks and T. R. Butt (2004). "SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins." J Struct Funct Genomics **5**(1-2): 75-86.

Nakahara, K., H. Yamamoto, C. Miyake and A. Yokota (2003). "Purification and characterization of class-I and class-II fructose-1,6-bisphosphate aldolases from the cyanobacterium Synechocystis sp. PCC 6803." Plant Cell Physiol **44**(3): 326-333.

Nakamura, Y., T. Kaneko, M. Hirosawa, N. Miyajima and S. Tabata (1998). "CyanoBase, a www database containing the complete nucleotide sequence of the genome of Synechocystis sp. strain PCC6803." Nucleic Acids Res **26**(1): 63-67.

Poulain, P., J. C. Gelly and D. Flatters (2010). "Detection and architecture of small heat shock protein monomers." PLoS One **5**(4): e9990.

Raineri, E., P. Ribeca, L. Serrano and T. Maier (2010). "A more precise characterization of chaperonin substrates." Bioinformatics **26**(14): 1685-1689.

Rajasekaran, N. S., P. Connell, E. S. Christians, L. J. Yan, R. P. Taylor, A. Orosz, X. Q. Zhang, T. J. Stevenson, R. M. Peshock, J. A. Leopold, W. H. Barry, J. Loscalzo, S. J. Odelberg and I. J. Benjamin (2007). "Human alpha B-crystallin mutation causes oxido-reductive stress and protein aggregation cardiomyopathy in mice." Cell **130**(3): 427-439.

Rutter, W. J. (1964). "Evolution of Aldolase." Fed Proc **23**: 1248-1257.

Simon, S., J. M. Fontaine, J. L. Martin, X. Sun, A. D. Hoppe, M. J. Welsh, R. Benndorf and P. Vicart (2007). "Myopathy-associated alphaB-crystallin mutants:

abnormal phosphorylation, intracellular location, and interactions with other small heat shock proteins." J Biol Chem **282**(47): 34276-34287.

Simon, W. J., J. J. Hall, I. Suzuki, N. Murata and A. R. Slabas (2002). "Proteomic study of the soluble proteins from the unicellular cyanobacterium Synechocystis sp. PCC6803 using automated matrix-assisted laser desorption/ionization-time of flight peptide mass fingerprinting." Proteomics **2**(12): 1735-1742.

Slabas, A. R., I. Suzuki, N. Murata, W. J. Simon and J. J. Hall (2006). "Proteomic analysis of the heat shock response in Synechocystis PCC6803 and a thermally tolerant knockout strain lacking the histidine kinase 34 gene." Proteomics **6**(3): 845-864.

Sun, X., J. M. Fontaine, A. D. Hoppe, S. Carra, C. DeGuzman, J. L. Martin, S. Simon, P. Vicart, M. J. Welsh, J. Landry and R. Benndorf (2010). "Abnormal interaction of motor neuropathy-associated mutant HspB8 (Hsp22) forms with the RNA helicase Ddx20 (gemin3)." Cell Stress Chaperones **15**(5): 567-582.

Tannous, P., H. Zhu, J. L. Johnstone, J. M. Shelton, N. S. Rajasekaran, I. J. Benjamin, L. Nguyen, R. D. Gerard, B. Levine, B. A. Rothermel and J. A. Hill (2008). "Autophagy is an adaptive response in desmin-related cardiomyopathy." Proc Natl Acad Sci U S A **105**(28): 9745-9750.

Torok, Z., P. Goloubinoff, I. Horvath, N. M. Tsvetkova, A. Glatz, G. Balogh, V. Varvasovszki, D. A. Los, E. Vierling, J. H. Crowe and L. Vigh (2001).

"Synechocystis HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding." Proc Natl Acad Sci U S A **98**(6): 3098-3103.

Tyedmers, J., A. Mogk and B. Bukau (2010). "Cellular strategies for controlling protein aggregation." Nat Rev Mol Cell Biol **11**(11): 777-788.

van Montfort, R. L., E. Basha, K. L. Friedrich, C. Slingsby and E. Vierling (2001). "Crystal structure and assembly of a eukaryotic small heat shock protein." Nat Struct Biol **8**(12): 1025-1030.

Wang, K. H., R. T. Sauer and T. A. Baker (2007). "ClpS modulates but is not essential for bacterial N-end rule degradation." Genes Dev **21**(4): 403-408.

Willis, M. S., J. C. Schisler, A. L. Portbury and C. Patterson (2009). "Build it up-Tear it down: protein quality control in the cardiac sarcomere." Cardiovasc Res **81**(3): 439-448.