Characterization of ClpB-DnaK bi-chaperonic system of *Mycobacterium tuberculosis* and identification of ClpB inhibitors through high-throughput screening

THESIS

SUBMITTED TO THE

JAWAHARLAL NEHRU UNIVERSITY

NEW DELHI, INDIA

FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

PRASHANT SINGH

CSIR-INSTITUTE OF MICROBIAL TECHNOLOGY (CSIR-IMTECH)

CHANDIGARH, INDIA

2020

सीएसआईआर – इमटैक
CSIR-IMTECH

सीएसआईआर – सूक्ष्मजीव प्रौद्योगिकी संस्थान
सैक्टर 39 – ए, चण्डीगढ़ – 160 036 (भारत)
CSIR-INSTITUTE OF MICROBIAL TECHNOLOGY
(A CONSTITUENT ESTABLISHMENT OF CSIR)
Sector 39-A, Chandigarh-160 036 (INDIA)

Certificate

The project work titled "**Characterization of ClpB-DnaK bi-chaperonic system of *Mycobacterium tuberculosis* and identification of ClpB inhibitors through high-throughput screening**", submitted by **Prashant Singh** for the degree of **Doctor of Philosophy**, has been carried out under the guidance of **Dr. Deepak Sharma** at CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh.

The research work is novel and has not been either partially or completely submitted for any other degree or diploma from any other institute or university.

Dr. Deepak Sharma
(Principal Investigator)

Prashant Singh

दूरभाष 0172-6665201 } Reception
0172-6665202 }
0172-6665..... (Direct)
फैक्स : 0091-172-2690632 (COA)
Fax : 0091-172-2690585 (Director)
0091-172-2690056 (Purchase)
एस. टी. डी. कोड : STD CODE : 172
Web : http://imtech.res.in

Dedicated

To my family

Acknowledgements

The journey of Ph.D was a special time of my life which taught me a lot of things. I not only gained scientific experience but also have grown as a person. I learned perseverance, patience and have become a more passionate and mature person. All this would not have been possible without the involvement and constant encouragement by numerous people including my well wishers, my friends, my colleagues and my family. I would like to take this opportunity to thank all those who made this thesis possible with an unforgettable experience for me.

First and foremost I would like to express my sincere gratitude to my PhD supervisor, ***Dr Deepak Sharma*** *for his vision and guidance. I was very fortunate to work under his supervision during early period of my scientific career, where i learnt the alphabet of science from him. I must acknowledge his extraordinary patience, optimism and confidence that I tried to learn from him. His suggestions expertise, freedom of thinking and constant encouragement was the key towards the meaningful completion of this work. Apart from being a Guide, he is a good human being who shared his life experiences with us which helped me in making few of the tough decision of my life.*

I am thankful to all the directors of IMTECH for providing us world-class infrastructure and MIL labs. I acknowledge former directors Dr. GirishSahni, Dr. R.S. Jolly, Dr. R.K. Sinha, Dr. Anil Koul and our present Director, Dr. SanjeevKhoisla for providing excellent research facilities and healthy working environment. I also express my gratitude towards Ph.D coordinators Dr. P.K. Chakraborti, Dr. S. Karthikeyan Subramanian, Dr.Charu Sharma and Ms.ShashiBatra, for conducting pre-Ph.D Course work smoothly and organising the wonderful seminars which helped us in broadening our knowledge. I owe special thanks to Mr.JanakiJi for his efforts in organizing my second research seminar.

I would like to thank my official collaborators Dr.Ramandeep Singh (THSTI) and Dr. Ashish Ganguly for helping in completion of my thesis. I had learned a lot from their experience and guidance which further improved my scientific aptitude.

I would like to thank my RAC committee members Dr Manoj Raje, Dr Ashish Ganguly and Dr Alka Rao for providing me valuable suggestions

during my RAC meetings which helped me to improve myresearch and completion of this thesis.

I find this opportunity to thank talented IMTECH staff and for helping me in exploring various domains related to my research interest. I would like to thank Dr. S. Kumaran, Dr.KrishanGopal for their valuable suggestions.

I would like to thank my seniors Dr. Navinder for helping in my biophysical experiments, Dr. Deepika for helping me throughout my refolding experiments, Dr. Arpit for teaching me the basics of research. Dr. Sonikaundal for AUC experiments, Shiv and Kanika for SAXS experiments, Dipesh for crystallisation trials, Adarsh sir for scientific discussion.

I would like to make special thanks for Priyanka (Poppy) for her scientific discussion, suggestion and troubleshooting for most of my experiments.Her stories are joyful and make everyone laugh in the Lab. I would also like to thank Anuradhika and Pradeep for their valuable suggestion and support during my tough times.

I also acknowledge my current and past labmates Monika, Mithun, Richa, Vibhuti,Isha,Rakesh, Jyoti, Debasree, swati mam Sonal, Abhishek, Krishna, Dr Nagesh, Dr Manish, and Dr. Mehrotra, surjit sir, Rajender sir for their help and support.I would like to thank my juniorsRahis, Madhuri, Happy, Saphy for their support.I would like to thank Dr. AnkurGautam for his scientific discussion and experience which helped me in troubleshooting of my experiments.Pradeep ji for most of technical and advisory suggestions

I would like to make special thanks to Dr. Harleen (THSTI) for helping in Highthroughput screening studies. I would like to acknowledge Aishwarya (IISER-Mohali) for helping me in my circular dichroism experiments and NEF assay. I would like to thank Ashish andAjit (ILS-bhubneshwar) for awaring me about biophysical conferences and workshop related to my research.

I would like to thank my dear friends Anil, Prabhatand Yash for their support and special birthday parties.I am always grateful for my long time friends Jaskaran and Parminder for their continuous support and help.

The financial assistance provided by DBT (Department of Biotechnology) throughout my research tenure is duly acknowledged. The cooperation and help provided by IMTECH such as Bioinformatic centre, Library, Instrumentation, Administration, Store and Purchase as well as electrical and engineering section is deeply acknowledged. I also thank Housekeeping and horticulture department for providing aesthetic, hygienic, and beautiful environment in the campus which make journey extremely pleasurable.

I am most grateful to the unconditional love and support provide by my wife ***Shelley****. I could not have gone so far without her support. I am thankful to her for being so patient and sometimes bearing my irritated behaviour. I am highly thankful to her for encouragement, confidence and faith in me during my difficult times. She has been a constant source of inspiration to me all these years.*

Words will never be enough to express my gratitude towards my family. All that I am today or aspire to be is because of ***my parent's*** *unconditional love, faith and support. The quality of life I am enjoying today is because of the endless scarifies they have made for me. They have always been a great strength throughout my life. I owe a lot to my mother and father and I dedicate this thesis to them. My brother* ***Sagar*** *is my partner in crime for everything. He took care of most of the things on my behalf during my busy PhD schedule. His much needed affection and good wishes supported me during my Ph.D. Motivation and support provide by my parents-in-law kept my spirit high.*

I am thankful to everyone who has been a part of my life shared my joy and sorrows. Life has been wonderful with all of you by my side. I extend my regards to all those who have supported me during my project and helped me weave all of the loose ends into a seamless whole. I wish to acknowledge everyone involved in this journey and beg for any omission, definitely not a deliberate one.

Prashant Singh

Table of Contents

List of Abbreviations

TB	Tuberculosis
PQC	Protein Quality Control
HSP	Heat Shock Protein
ClpB	Caseinolytic peptidase B
GrpE	*Gro-P* like protein E
NBD	Nucleotide Binding Domain
SBD	Substrate Binding Domain
CTD	Carboxy-Terminal Domain
NEF	Nucleotide Exchange Factor
ZNFR	Zinc Finger like Region
NMR	Nuclear Magnetic Resonance
HspBP1	HSPA (Hsp70) binding protein 1
Sse1	Stress Seventy subfamily E
BAG	BCL2-associated athanogene
SNL1	Suppressor of Nup116-C Lethal
[*PSI+*]	Prion form of sup35
HtpG	High temperature protein G
ROS	Reactive Oxygen Species
CD	Circular Dicroism
M.R.E	Mean residue elipticity
AUC	Analytical Ultracentrifugation technique
RPM	Rotations per minute
SPR	Surface Plasmon Resonance
SAXS	Small-angle X-ray scattering
TEV	Tobacco Etch Virus
His6X	Hexa-Histidine Tag
dNTP	Deoxy nucleotides tri phosphate
EtBr	Ethidium rromide
EDTA	Ethylenediaminetetraacetic acid
SDS	Sodium dodecyl sulfate
PAGE	Polyacrylamide gel electrophoresis

Rosette (DE3)	*E.coli* strain with rare codons for protein expression
LB	Luria broth
PCR	Polymerase chain reaction
NaCl	Sodium chloride
PMSF	Phenylmethylsulfonyl fluoride
PIC	Protease inhibitor cocktail
Tris	Trisaminomethane
Triton X-100	Octylphenol Ethoxylate
KCl	Pottasium Chloride
$MgCl_2$	Magnesium Chloride
PCR	Polymerase chain reaction
DTT	Dithiothreitol
LRB	Luciferase refolding buffer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Tris	dsdsdsdsd
DMSO	Dimethyl sulfoxide
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
AMP-PNP	Adenylyl imidodiphosphate
ATP-γS	Adenosine 5′-(3-thiotriphosphate) tetralithium salt
KD	Dissociation constant
RU	Response Unit
s	Sedimentation coefficient
D	Diffusion coefficient
mAU	Milli Absorbance Unit
Dmax	Maximum Dimension
Rg	Radius of Gyration
O.D	Optical Density
kDa	Kilo Dalton
nM, µM, mM, M	Nano molar, Micro molar, milli molar, Molar concentration
µg, mg, g	Microgram, Milligram, Gram
µl, ml, l/L	Microliter, Milliliter, Liter
Sec, Min	Second, minute

cm, mm, nm	Centimeter, millimeter, nanometer
nm, µm	Nonometer, Micrometer
µg/ml, mg/ml, mg/l, g/l	Microgram per milli liter, Milli gram per milli liter, Milli gram per liter
pH	Potential Hydrogenii
α	Alpha
β	Beta
γ	Gamma
µ	Micro
Δ	Delta
%	Percentage
º	Degree
=	Equal
Φ	Phi
π	Pie
θ	Theta
C	Celsius

List of Figures

List of Tables

List of Publications

1. **Prashant Singh**, Harleen Khurana, Shiv Pratap Singh Yadav, Kanika Dhiman, Padam Singh, Ashish, Ramandeep Singh and Deepak Sharma. Biochemical characterization of ClpB protein from Mycobacterium tuberculosis and identification of its small-molecule inhibitors. International Journal of Biological Macromolecules *165*, 375-387. https://doi.org/10.1016/j.ijbiomac.2020.09.131

2. **Prashant Singh**, Breezy Unik, Anuradhika Puri, Gandharva Nagpal, Balvinder Singh, Ankur Gautam, Deepak Sharma. HSPMdb: A computational repository of heat shock protein modulators, *Database*, Volume 2020, baaa03, https://doi.org/10.1093/database/baaa003.

3. Deepika Gaur, **Prashant Singh**, Jyoti Guleria, Arpit Gupta, Satinderdeep Kaur and Deepak Sharma. The Yeast Hsp70 Co-chaperone Ydj1 Regulates Functional Distinction of Ssa Hsp70s in the Hsp90 Chaperoning Pathway, *GENETICS*, 2020; https://doi.org/10.1534/genetics.120.303190.

4. Jasjeet Kaur, **Prashant Singh**, Deepak Sharma, Kusum Harjai and Sanjay Chibbar. A potent enzybiotic against methicillin-resistant *Staphylococcus aureus. Virus Genes* (2020). https://doi.org/10.1007/s11262-020-01762-4.

5. Arpit Gupta, Anuradhika Puri, **Prashant Singh**, Surbhi sonam, Richa Pandey and Deepak Sharma. The yeast stress inducible Ssa Hsp70 reduces α-synuclein toxicity by promoting its degradation through autophagy. 2018, *PLOS Genetics* 14(10): e1007751. https://doi.org/10.1371/journal.pgen.1007751.

List of Poster presentation

1. **HyThaBio 2020**, Hydrodynamic and thermodynamic analysis of biological macromolecules and their interactions, Institut de Biologie Structurale (IBS), EPN Campus, Grenoble, France.

Title: Biophysical characterization of *Mycobacterium tuberculosis* Hsp70 chaperone machinery, and identification of novel inhibitors against ClpB.

Introduction

Introduction

Tuberculosis or TB remains one of leading causes of the deaths in humans worldwide. WHO estimated that, one third of the total world's population is infected with *M. tuberculosis*. In 2018 alone, 1.2 million people died and 10 million people are affected with (WHO, Global Tuberculosis Report, 2019). TB, which most commonly affects the lungs and transmitted from person to person via droplets from the lungs and throat of people affected with the active respiratory disease(Delogu et al., 2013). Most people who are exposed to *M. tuberculosis* never develop symptoms because the bacteria survives in an inactive form in the body, and becomes active when immune system is compromised. Treatment comprises of antibiotic course for 6-9 months and could be fatal if left untreated (CDC, Center for Disease control and Prevention, U.S. Department of Health & Human Services, https://www.cdc.gov/tb/topic/treatment/default.htm). In case of drug resistant TB, combination of antibiotics such as isoniazid, rifampicin, ethambutol etc. are given which is provided for about 27 months(Leung et al., 2011). BCG vaccine is available for children but is less effective in case of adults and ineffective against XDR-TB(Davenne and McShane, 2016).

M. tuberculosis is the principal TB causative agent in humans. *M. tuberculosis* enters the host via inhalation of respiratory droplet nuclei through the respiratory tract. Inside the host lungs there is competition between *M. tuberculosis* and host system which resulted in four outcomes: 1: Host neutralizes the *M. tuberculosis* bacilli, 2: *M. tuberculosis* invade the host and results in its multiplication causing primary TB, 3: *M. tuberculosis* become dormant and does not have any effect 4: *M. tuberculosis* goes into latent phase upon activation that results in TB disease (Schluger and Rom, 1998).

M. tuberculosis genome is fully sequenced and consists of 4411529 bp encoding 4000 genes with a characteristic feature of very high G+C contents (≈ 60%) (Cole et al., 1998b). A large portion of *M. tuberculosis* protein encoding genes are dedicated for its virulence and pathogenesis majorly lipogenesis and lipolysis & glycine-rich proteins which provide antigenic variations (Bishai, 2000).The

characteristic features of the bacteria include its slow growth, dormancy, intracellular pathogenesis, genetic homogeneity and complex cell envelope. *M. tuberculosis* is successfully isolated in 1905 and it gains an extensive focus in biomedical research because of retention of virulence in animal models of TB (Cole et al., 1998a).

About 20.5 % of TB patients developed Multi-drug resistant TB (MDR-TB) due to resistance developed against 1st line of antibiotics like rifampicin, isoniazid and ethambutol (Zhang and Yew, 2015). Further 6.2 % of people develop Extensively drug resistant TB (XDR-TB) resistant to 2nd line of antibiotics such as amikacin, cycloserines and fluoroquinoles (Institute of, 2012). *M. tuberculosis* ability to continuously develop resistance makes the treatment regime very aggressive and expensive. Therefore, there is an urgent need for novel therapeutic approach to combat *M. tuberculosis*.

During host infection and combination of antibiotic drugs, *M. tuberculosis* undergoes several stresses like reactive oxygen and nitrogen species and various proteostasis stress which ultimately results in protein aggregation (Kumar et al., 2011a; Lupoli et al., 2018a). Molecular chaperones Hsp100, Hsp70, Hsp40 and NEFs help in the refolding the protein or mark them for degradation. Apart from their refolding function, these molecular chaperones are involved in various biological processes, such as disaggregation of protein aggregates, polypeptide transport across biological membranes, assembling multi-subunit proteins and proteolysis (Gupta et al., 2018; Reeg et al., 2016; Rohde et al., 2005; Truman et al., 2012).

Refolding chaperone machinery (Hsp70) in cooperation with disaggregase Hsp100 (ClpB) plays a crucial role in *M. tuberculosis* survival (Glover and Lindquist, 1998; Goloubinoff et al., 1999b). The inhibition of disaggregation machinery of *M. tuberculosis* results in growth defect during stress conditions and no persistence of the pathogen in mice model (Vaubourgeix et al., 2015). Further deletion of ClpB results in altered morphology, impaired biofilm formation and reduced infectivity of *M. tuberculosis ex-vivo* (Tripathi et al., 2020a).

Unlike *E. coli*, *M. tuberculosis*'s Hsp70 DnaK is predicted to be essential for growth (Fay and Glickman, 2014; Griffin et al., 2011; Sassetti et al., 2003). Refolding activity of DnaK is regulated by its co-chaperone Hsp40 (DnaJ) and Nucleotide exchange factor (GrpE). Presence of two DnaJ isoforms enhance the refolding capability of DnaK. Though each of the isoform are not essential for bacterial survival, , however one isoform at a time can be deleted but collectively they are indispensable for growth of *M. tuberculosis* (Lupoli et al., 2016b). GrpE exchanges the nucleotide to repeat DnaK refolding cycle and is essential for in-vitro growth of *M. tuberculosis*(Fay and Glickman, 2014). Therefore complete bichaperonic machinery is essential for *M. tuberculosis* survival. Also, the homolog of disaggregase ClpB and NEF GrpE in mammals are structurally far distant as compared to *M. tuberculosis* and functionally not involved in classical chaperoning mechanism thus, making them a potential target for combating TB. Based on the available information we have framed our following objectives.

Primary objectives for current study are:

1. Characterization and identification of small molecule inhibitors against ClpB protein from *Mycobacterium tuberculosis*.

2. Characterization of Hsp70 system and exploring nucleotide exchange activity of GrpE of *Mycobacterium tuberculosis.*

3. Developing a comprehensive database of modulators for Heat shock proteins (Hsps).

Review of Literature

Tuberculosis

There is a dread disease which so prepares its victim, as it were, for death; which so refines it of its grosser aspect, and throws around familiar looks, unearthly indications of the coming change dread disease, in which the struggle between soul and body is so gradual, quiet, and solemn, and the result so sure, that day by day, and grain by grain, the mortal part wastes and withers away, so that the spirit grows light and sanguine with its lightening load, and, feeling immortality at hand, deems it but a new term of mortal life; a disease in which death takes the glow and hue of life, and life the gaunt and grisly form of death; a disease which medicine never cured, wealth warded off, or poverty could boast exemption from; which sometimes moves in giant strides, and sometimes at tardy pace; but, slow or quick, is ever sure and certain.

—Charles Dickens, 1870, in Nicholas Nickleby, Wiendenfeld and Nicholson,
London, p. 243

Tuberculosis is one of the oldest diseases known to mankind that has co-evolved with humans for over thousand years (Hirsh et al., 2004). The earliest known molecular evidences of tuberculosis have been found in 17000 years old fossil of extinct bison and 9000 years old human remains from eastern Mediterranean (Hershkovitz et al., 2008; Rothschild et al., 2001). Tuberculosis accounted for about 25% deaths in Europe in 19th century. In London it was responsible for death of 1 per 1000 persons per year and was described as “the captain among these men of death” by John Bunyan, an English preacher and writer (TM, 2009). Dr. Richard Morton in 1689 established that pulmonary form of tuberculosis was related to tubercles. Robert Koch isolated and cultured *Mycobacterium tuberculosis* from crushed tubercles (Sakula, 1983). He identified *M. tuberculosis* as etiological agent for air-borne infectious disease tuberculosis(Sakula, 1983). Subsequently he was awarded Nobel Prize in Physiology or medicine for his discovery.

Mycobacterium tuberculosis

M. tuberculosis is a non-spore forming weak Gram-positive rod shaped bacteria (Figure 1.1) It does not produce capsule and has no flagellum (Hett and Rubin,

2008). It has a complex cellular envelope with relatively slow growth (Jackson, 2014). It is an intracellular macrophage pathogen that establishes in and infects the pulmonary system. It represents a dynamic spectrum varying from latent asymptomatic and non-transmissible infection to active life threatening disease. Latent TB refers to clinical disorder where *M. tuberculosis* infects the host but the immune system of host does not allow its replication and is devoid of any tissue damage (Gideon and Flynn, 2011; Kumar, 2016). The organism usually resides in the host and conditions into a dormancy state until its immune system is compromised. Although *M. tuberculosis* is primarily a pulmonary pathogen it can cause disease throughout the host body (Smith, 2003).

Figure 1.1 Scanning electron micrographs for *M. tuberculosis*. *M. tuberculosis* showing rod shaped morphology under Scanning electron microscope at 15549X magnification. Lower panel displays the *M. tuberculosis* stained pink colour through Ziehl-Neelson staining. Figures are adapted and generated by Center for Disease control (CDC), Division of Tuberculosis elimination, United States of America. (Sterling and Haas, 2006).

M. tuberculosis generation time in both infected animals and synthetic laboratory media is around 24 h (Cole et al., 1998a). In laboratory bacterial growth requires 3-4 weeks to form visual colonies that have dry and wrinkled surface (Bloom and Murray, 1992). The complete genome sequence for *M. tuberculosis* strain H37Rv is available enabling identification of its unique features for development of novel diagnostic tests, vaccines and biomarkers that differentiate it from other bacteria. *M. tuberculosis* is equipped with variety of genes involved in lipolysis and lipogenesis essentially required for survival inside host and synthesis of components of its cell envelope. The genome has high G+C content and repetitive DNA with many insertion sequences (Forrellad et al., 2013). Cellular envelope of *M. tuberculosis* is also unique as it comprises of N-glycolylmuramic acid. Its cell wall contains lipids consisting of uncommon long chain mycolic acids (Brennan and Nikaido, 1995). Cell wall of Mycobacteria retain carbolfuchsin with acidified organic solvents, thus they are called acid fast bacilli (Riello et al., 2016).

Global burden of TB

Despite the fact that first drug to treat TB was discovered 80 years ago, TB still affects approximately 1.2 million people annually. TB is responsible for more deaths than any other single infectious agent (WHO Global TB reports, 2013-2019). One in every three individuals is believed to be suffering from latent TB. Figure 1.2 shows India and China shares more than one-third burden of TB worldwide. In India, every three minutes two deaths occur due to TB (Sandhu, 2011). TB is a disease of poverty that is indistinguishably related to under nutrition and overcrowding. About 12% TB cases are HIV associated (Trinh et al., 2015). TB was declared a global public health emergency by WHO in 1993. Constant progress in tuberculosis diagnosis and treatment all over the world has been challenged by HIV epidemic and drug resistance.

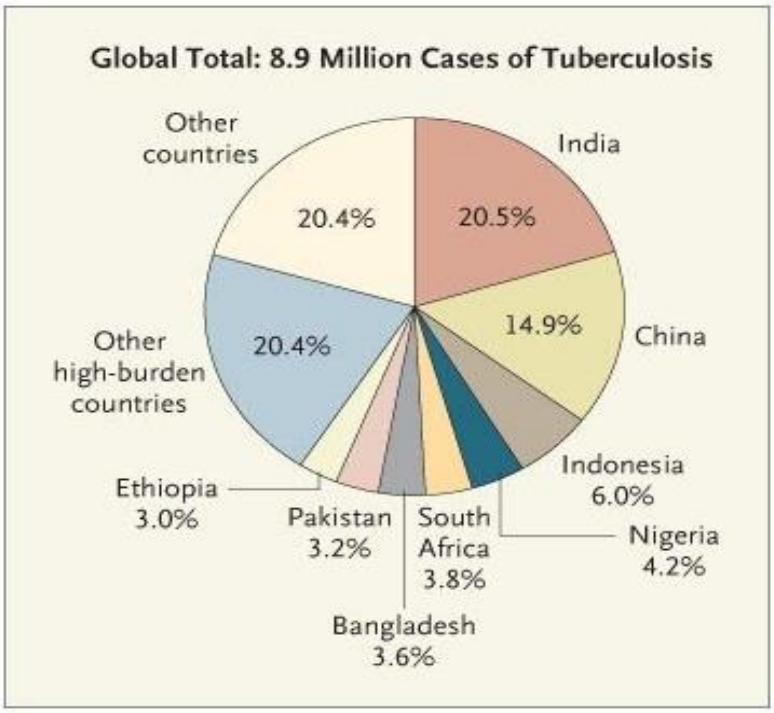

Figure 1.2 Global Burden of TB. India and china shares one-third of the global burden, while more than 60% of burden is shared by Asia continent. Figure is adapted from previous studies (Steinbrook, 2007).

Diagnosis

Koch also proposed use of tuberculin as a diagnostic tool to diagnose the disease (Fredericks and Relman, 1996). He established staining method for visualization of Bacillus. Paul Ehrlich, a German bacteriologist and doctor further improved this method that served as the basis for development of Ziehl-Nielson staining. Tuberculin skin test (TST) shown in Figure 1.3 is the only method to detect a recent or years past latent infection through delayed type hypersensitivity against antigens of *M. tuberculosis* (Glickman and Jacobs, 2001). This test is presently disseminated due to high false positives because reactivity of Bacille Calmette Guerin (BCG) to Purified Protein Derivative (PPD) and false negatives in immune compromised patients (Fätkenheuer et al., 1999). *In vitro* blood tests of cell mediated immune response quantify interferon-gamma secretion following RD1 encoded antigen stimulation. RD1 antigens are not encoded in genomes of non-tuberculous mycobacteria or BCG vaccination strains and therefore are more precise for *M. tuberculosis* than PPD. However, this test has poor predictive value (Arend et al., 2002).

For diagnosis of active TB, imaging techniques like chest X-rays, sputum smear microscopy, culture dependent and molecular methods are used. Chest radiography is popular tool owing to advent of digital radiology (Figure 1.3, Right panel). However, abnormal chest X-rays must be confirmed with microbiological tests as X-rays are not specific (Ryu, 2015). Microscopy of sputum smears remains widely used test in developing countries despite of its many limitations, molecular techniques through PCR offer more sensitive, specific and rapid detection and identification (Ahmed et al., 1998). Bacillary sensibility tests for anti-TB drugs are important to diagnose MDR-TB.

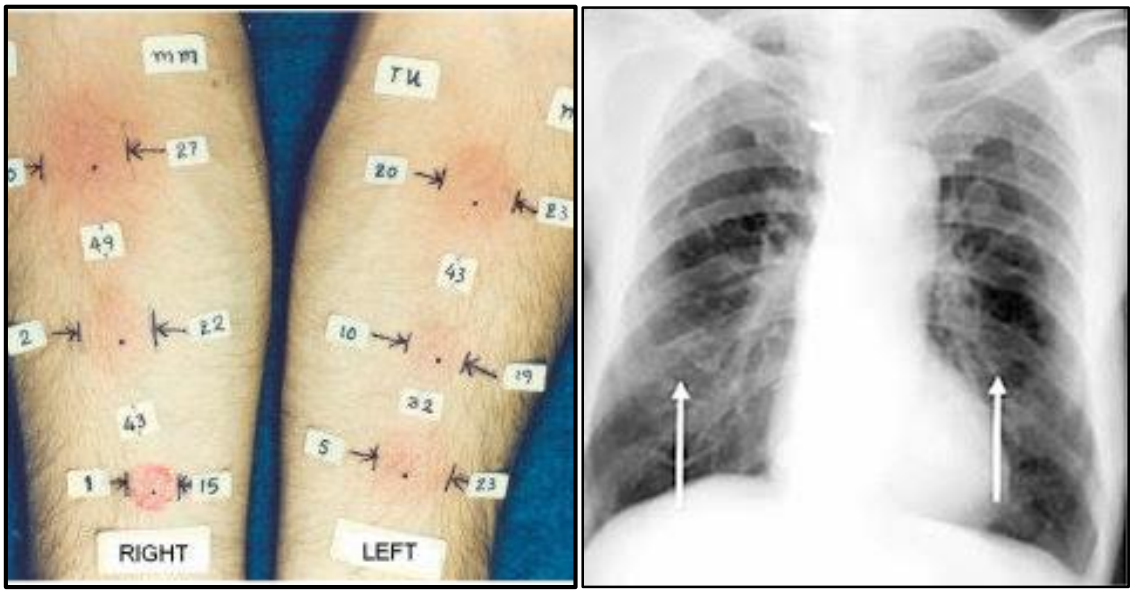

Figure 1.3 Diagnosis of TB. Left panel shows the Tuberculin skin test for a patient injected with 0.1 ml of tuberculin solution records blister formation after 48-72 hours. Size of blisters in mm determines the infectivity of *M. tuberculosis.* Right panel shows the chest radiograph of a patient highlights the infection of pulmonary TB. White arrows indicates the presence of bacteria residing on the surface. Figure is adapted from previous studies (Dimoliatis and Liaskos, 2008; Med and Al-Ubaidi, 2018).

Treatment

BCG vaccine is the only licensed vaccine at present for prevention for treatment of active TB. More than 90 % newborns are vaccinated each year with this vaccine around the globe(Hussey et al., 2007). BCG vaccine is effective in morbidity and mortality due to active TB in children below the age of 5 years. Conversely, its efficacy in adolescents and adults is uncertain (Hussey et al., 2007).

First line drugs like isoniazid, rifampicin, streptomycin, ethambutanol, pyrazinamide and fluoroquinolones are largely bactericidal and provide high efficacy with relatively low toxicity (Mitchison and Davies, 2008). The preferred existing regimen for treatment of active drug sensitive TB is therapy with first line drugs for a period of minimum six months. First two months include intensive phase of treatment where rifampicin, isoniazid, pyroinamide and ethambutol are used followed by isoniazid and rifampicin in continuation phase for four months. Ethionamide, para-aminosalicylic acid and cycloserine and others are second line bacteriostatic drugs offering lower efficacy and more toxicity (Pfyffer et al., 1999).

Despite high success rate of six-month regimen, many patients discontinue their treatments due to long duration and toxicity of drugs. Hence, supportive methods are crucial to confirm optimal adherence and escape treatment failure, relapse and development of drug resistance. Directly observed treatment short course (DOTS) is a worldwide practiced scheme for TB cure (Sisay et al., 2014). It relies on clinical methodology to treat patients and public health management systems including reporting and recording of TB cases. The health care workers counsel patients and ensure that each dose of medication is taken correctly.

Drug resistance

The phenomenon of drug resistance in TB was initially observed in 1948 during the first human trial for TB therapy suggesting the spontaneous adaptability of *M. tuberculosis* by modifying its genome (Daniels and Hill, 1952). Genetic mutations are responsible for emergence of drug resistant strains of *M. tuberculosis* against handful of presently existing first and second line drugs. Recurrent mismanagement of few available drugs has led to consecutive accumulation of mutations creating strains that are simultaneously resistant to all or multiple drugs (Fair and Tor, 2014). MDR-TB is caused by strain of *M. tuberculosis* that is resistant to at least rifampicin and isoniazid. XDR-TB is a form of MDR resistant to isoniazid, rifampicin, any fluoroquinolone and atleast one of three second-line injectable drugs amikacin, kanamycin or capreomycin (Kurz et al., 2016; Palomino and Martin, 2014).

Protein quality control

Protein quality control system (PQC) is essential for an organism to survive against various physiological as well as environmental stresses. *M. tuberculosis* overcomes the stress conditions through PQC system by modulating the proteastasis and proteolysis components shown in figure 1.4. Proteolysis system degrades the misfolded proteins through various proteases such as ClpP, ClpC, ClpX e.t.c (Kar et al., 2008; Schmitz and Sauer, 2014). While proteastasis network involves protein folding process assisted by chaperones. ClpB, DnaK GroEL-ES are the major players performing the energy dependant refolding of misfolded or aggregated protein to their native state. These chaperones are well characterized for their function in mycobacteria (Fay and Glickman, 2014; Lupoli et al., 2016b). Chaperones maintaining PQC emerges as a potential therapeutic targets for small molecule drug discovery against *M. tuberculosis* (Lupoli et al., 2018a).

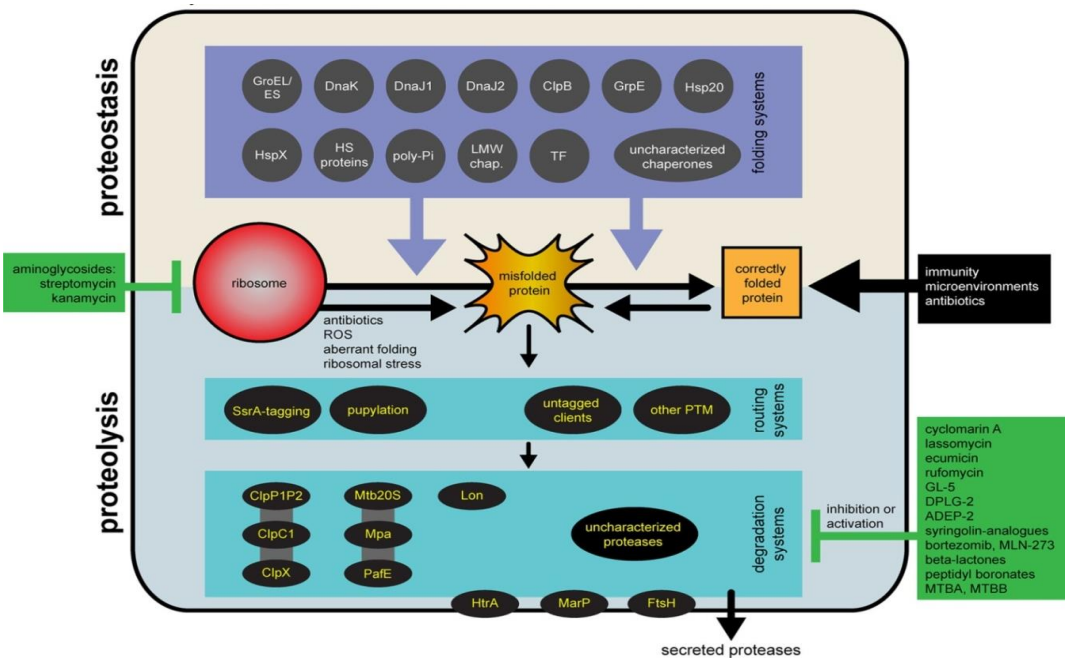

Figure 1.4. Mycobacterial protein quality control system. Mycobacteria possess two major system for maintaining protein structure and function via proteostasis (purple) or proteolysis (light blue). HS proteins are heat shock proteins, poly-Pi represents polyphosphate; LMW chaperones represents low molecular weight chaperones; sHsps are represented by HspX (Acr1/α-crystallin-1) and Hsp20 (Acr2/α-crystallin-22), TF represents Trigger Factor, PTM represents post-translational modification and ROS represents reactive oxygen species. Figure is adapted from previous study (Lupoli et al., 2018a).

Protein folding

Proteins are the biological molecules that are most versatile and important to all organisms. Proteins assist almost all the physiological activities such as muscle movement, food metabolism, sensory regulation, and protection against infection. Except for inherently unstructured protein, which lacks a stable tertiary structure (Fink, 2005), protein must be folded into a given three-dimensional conformation, called the native state to perform its function (Dill et al., 1995). The physical mechanism through which an unstructured polypeptide fits into a stable tertiary structure is known as protein folding (Anfinsen, 1973; Seckler and Jaenicke, 1992). In the 1970s, Anfinsen gave the famous hypothesis that the -native conformation is determined by amino acid sequence, relying on the studies on Ribonulcease A spontaneous folding(Anfinsen, 1973). The comprehensive studies in the following forty years allow us to obtain a much broader understanding of the protein folding theory. In 1968, Levinthal suggested the classic concept of folding mechanism, based on his observations that proteins fold by some guided process to achieve their native conformation (Levinthal, 1968). During research into protein folding, which supports this model, folding intermediates such as unstable transient states and molten globule state were reported (Brockwell et al., 2000; Creighton, 1997; Ohgushi and Wada, 1983).Nevertheless, this model is questioned by the discovery of parallel folding pathways, for example folding process for lysozyme and cytochrome c (Bieri et al., 1999; Brockwell et al., 2000; Tezcan et al., 2002).

Based on in vitro protein denaturation and refolding experiments (Figure 1.5), the new model which uses the concept of "folding funnel" and "energy landscape" was created (Dill et al., 1995; Veitshans et al., 1997; Wolynes et al., 1996). This explanation offers a mathematical description of the free energies of different molecular systems. The native states are situated at the bottom of the funnel that represents the lowest global energy, whereas the top side of the funnel represents the higher-energy ensemble of unfolded states or denatured states. Most protein folds on rough and rocky surfaces, through, which the polypeptide chain will traverse to the native states, likely through one or more inhabited intermediates (Dill and Chan, 1997; Gruebele, 2002).

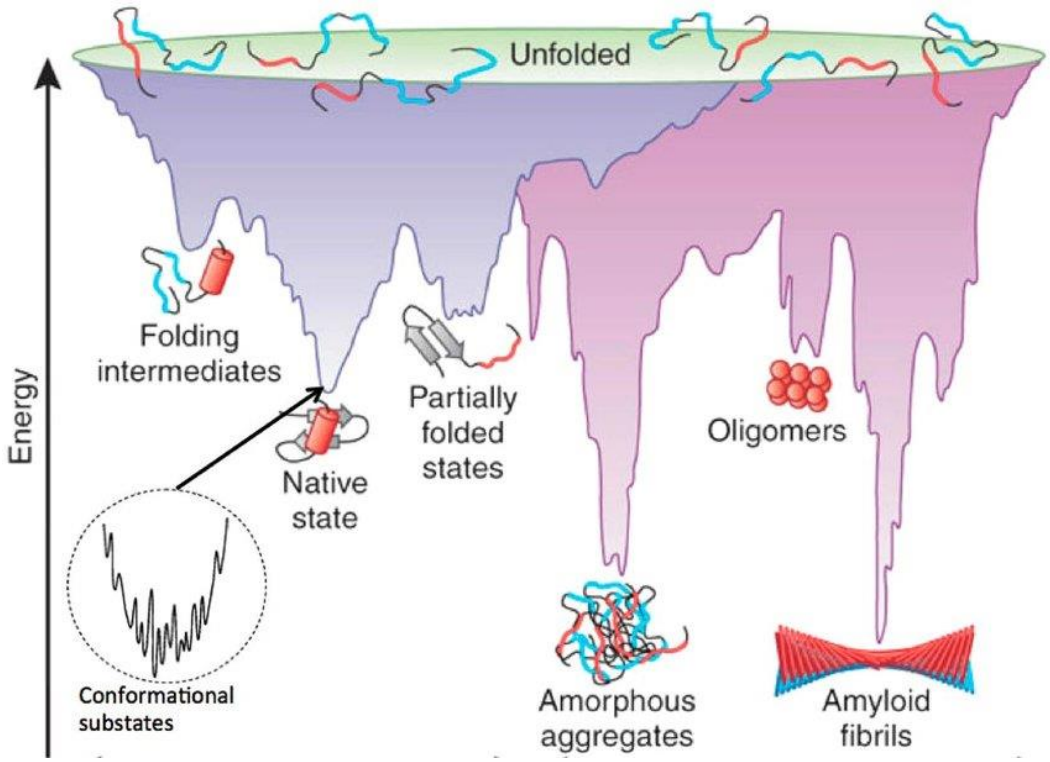

Figure 1.5 Energy landscape schematics of protein folding and aggregation. The native state at the bottom of the funnel is in a given state with minimal energy, while the unfolded state is characterized by a high-energy ensemble at the top of the funnel. The purple area displays the majority of configurations 'funneling' to native state and pink area represents the conformation approaching towards amorphous aggregates or amyloid fibrils. Figure is adapted from previous studies(Raskatov and Teplow, 2017).

Protein folding in the cell

The existing folding funnels cannot elucidate how well the polypeptide chains fold during physiological conditions, since they explain only the folding action of an independent single polypeptide chain at infinite dilution (Clark, 2004). The concentration of protein in cytoplasm is nearly 200 mg / ml in living cells (Lorimer, 1996). Protein-proteins interaction occurs immediately as the nascent polypeptide chain exits the ribosome channel (Braakman et al., 1991). Additionally, certain macromolecules including multiple types of RNA often lead to a more crowded system (Ellis, 2001). Interactions between multiple biomolecules and the crowding effects profoundly affect the polypeptide chain's folding nature(van den Berg et al., 1999). Aggregated and unfolded proteins

have been found to be correlated with neurodegenerative disorders (van den Berg et al., 1999; Winklhofer et al., 2008). Schematic for folding pathway is shown in figure 1.6.

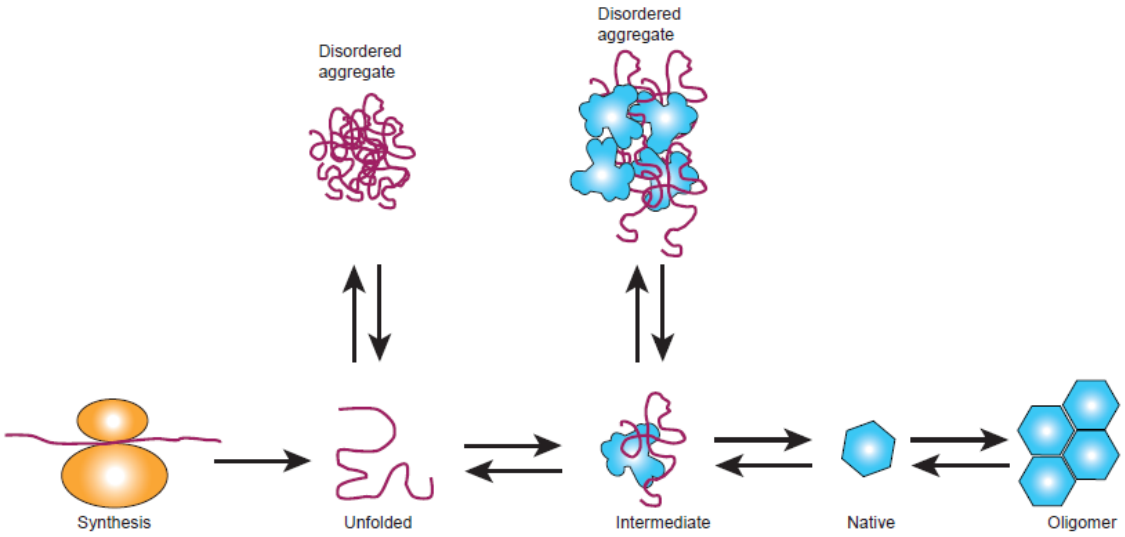

Figure 1.6 Folding pathway for proteins. Polypeptide (red line) chains are processed and released from the ribosome. They then fold into one or more intermediate states (irregular shape in blue) or directly to their natural conformations (blue hexagon). During the folding phase misfolded protein may form aggregates. Figure adapted from review article "Folding and Misfolding of Proteins" (Dobson, 2003).

Role of Heat shock proteins (Hsps) as molecular chaperones in virulence of *Mycobacterium tuberculosis*

Molecular Chaperones

In 1978 Lasky coined the word molecular chaperones. A nuclear protein called nucleoplasmine was discovered which facilitated nucleosome assembly but was not part of the nucleosome (Laskey et al., 1978). A molecular chaperone is defined as proteins that assist the folding or unfolding of other macromolecule non-covalently without being a part of its final structure (Ellis, 1993). Molecular chaperones are a broad family of proteins that are ubiquitous and highly conserved from prokaryotes to eukaryotes (Georgopoulos and Welch, 1993; Lindquist and Craig, 1988). Molecular chaperones interact with unfolded substrate proteins by binding to their hydrophobic patches and assist the protein

fold in a sequestered environment. Besides their fundamental role in de novo protein folding, chaperones are involved in various aspects of proteome assembly, like assistance in molecular assembly, regulation of cell cycle and protein trafficking and degradation (Bukau and Horwich, 1998; Gupta et al., 2018; Hartl and Hayer-Hartl, 2002; Reeg et al., 2016; Rohde et al., 2005; Truman et al., 2012). Most of these proteins were found in abundance during heat shock response in Drosophila and later in *E. coli,* therefore termed as heat shock proteins (Hsp).

Classification

Chaperones are classified into 5 major families on the basis of their molecular masses. Major families are Hsp100 (Clp), Hsp90, Hsp70 (DnaK), Hsp60 (GroEL), Hsp40, and sHsps (small Hsps).

Hsp100 Family

Hsp100 family members were first discovered in *E. coli* and are identified as ATPases associated with various cellular activities (AAA+ Super family), due to the nucleotide binding domains (NBD1 and NBD2) present in their primary sequences (Agarwal et al., 2001; Neuwald et al., 1999; Patel and Latterich, 1998; Schirmer et al., 1996). The prokaryotic equivalent is called ClpB. These chaperones exist as 80-100 kDa protomer oligomerizes into hexameric ring structures (Lee et al., 2003b), which act as a protein unfolding element when the aggregate proteins are transported to a protease proteolytic chamber.Hsp100 family is further divided into two subclasses shown in figure 1.7. The class I family includes ClpA, ClpB, ClpC, ClpE and ClpL which contain two nucleotide binding domain (NBD1 and NBD2) with ATPase activity. The class II family members contains only one nucleotide binding domain includes ClpX, ClpP and ClpY (Lindquist and Craig, 1988; Schirmer et al., 1996).The Hsp100 chaperones performs protein disaggregation or protein degradation in coordination with Clp (protease), and thus regulates the cellular homeostasis (Beuron et al., 1998; Goloubinoff et al., 1999a; Horwich et al., 2001; Lindquist and Craig, 1988; Weber-Ban et al., 1999).

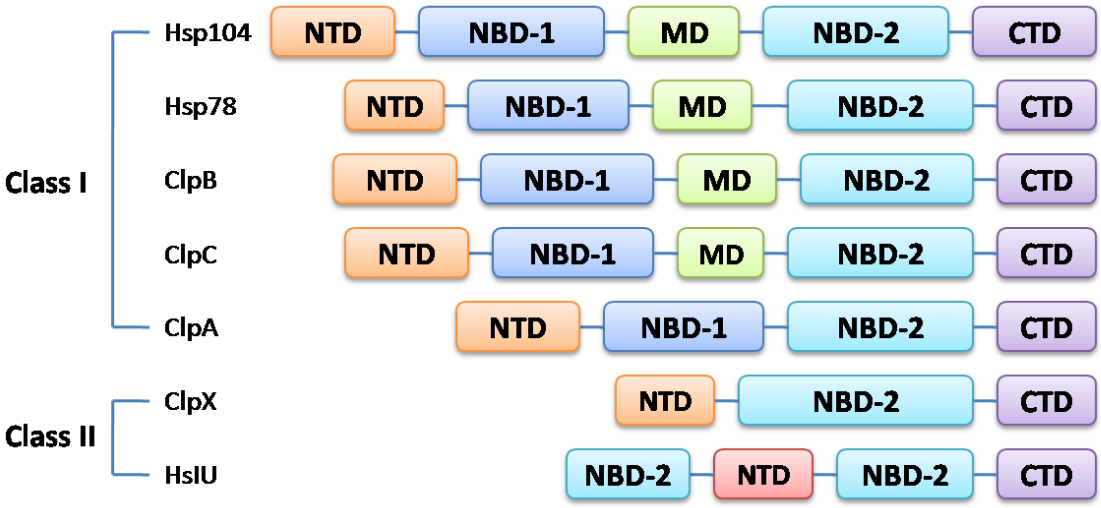

Figure 1.7 Hsp100 family are categorized into two classes based on their domain organization. Class I family members contains two NBD domain while Class II family contains only one.

ClpB

ClpB is a unique member of Hsp100 class from prokaryotes which did not interact with proteases but coordinates with Hsp70 and Hsp40 to direct aggregated substrates to native state for degradation or refolding (Doyle et al., 2007; Miot et al., 2011). *E. coli* ClpB has been extensively studied for its structural and functional properties (Barnett et al., 2000; Kedzierska-Mieszkowska et al., 2004; Lee et al., 2003a). In *E. coli*, ClpB is not essential during optimal conditions, however is required for survival of organism under stresses such as high temperature (Lourdault et al., 2011; Thomas and Baneyx, 1998). Structurally, *E. coli* ClpB forms nucleotide dependent hexamer where each protomer consists of an amino terminus domain followed by two ATP binding domains (NBD1 and NBD2) linked through a middle domain (Rizo et al., 2019).Analytical ultracentrifugation and gel filtration studies show the presence of ClpB in a dynamic equilibrium with its higher order oligomers (Akoev et al., 2004; del Castillo et al., 2011; Lin and Lucius, 2016; Zolkiewski et al., 1999). The amino terminus domain of ClpB is required for stable substrate interaction and the middle domain mediates its interaction with DnaK (Rosenzweig et al., 2015; Seyffer et al., 2012; Tanaka et al., 2004).Truncation studies suggest that amino terminal of ClpB is required for disaggregation activity (Barnett et al., 2000).

ClpB from *Mycobacterium tuberculosis*

M. tuberculosis genome harbours single copy of *clpB* gene encoded by Rv0384c (Cole et al., 1998a). ClpB forms 92.5 kDa monomeric protein which further oligomerizes into its functional hexameric form in the presence of ATP (Lupoli et al., 2016b; Tripathi et al., 2018).

ClpB is class I family member composed of 5 structural domains shown in figure 1.7. N-terminal domain (NTD) plays an important role for ClpB functioning. Previous studies shows that NTD truncated ClpB shows higher ATPase activity with similar disaggregation activity and is required for stable interaction with substrate (Tripathi et al., 2018).Nucleotide binding domain 1 and 2 (NBD-1 and NBD-2) are highly conserved and are responsible for ATPase activity of ClpB. Middle domain (MD) is the connecting domain for two NBDs, interacts with NBD region of DnaK regulating the disaggregation activity (Kedzierska-Mieszkowska et al., 2004; Yu et al., 2018a). C-terminal domain (CTD) has not been explored much for *M. tuberculosis* ClpB but studies from *E. coli* shows that CTD might play an important role in oligomerization or stabilization of hexameric assembly (Barnett et al., 2000).

Mechanism of disaggregation

Cryo-EM studies in 2018 captures the high resolution ring shaped structure of *M. tuberculosis* ClpB resolved at 3.6 Å in the presence of substrate κ-casein (Figure 1.8) (Yu et al., 2018a). NBD1 and NBD2 form a long helical region which binds 2 ATP molecules per monomer resulting in oligomerization of ClpB into a hexameric assembly. ATP bound state of ClpB show high affinity for substrate binding. The M-domain of ClpB forms a helical loop which protrudes outwards from NBD-1 and NBD-2. NBD region of DnaK makes inter helical contacts with M-domain loop resulting into translocation of substrate to the central pore of hexameric assembly. The hydrolysis of ATP drives the conformational changes in hexameric assembly, allowing substrate channel through the central pore exiting into a disaggregated form (Yu et al., 2018a). Disaggregated substrate further enters DnaK cycle for refolding back to native state (Doyle et al., 2007; Miot et al., 2011).

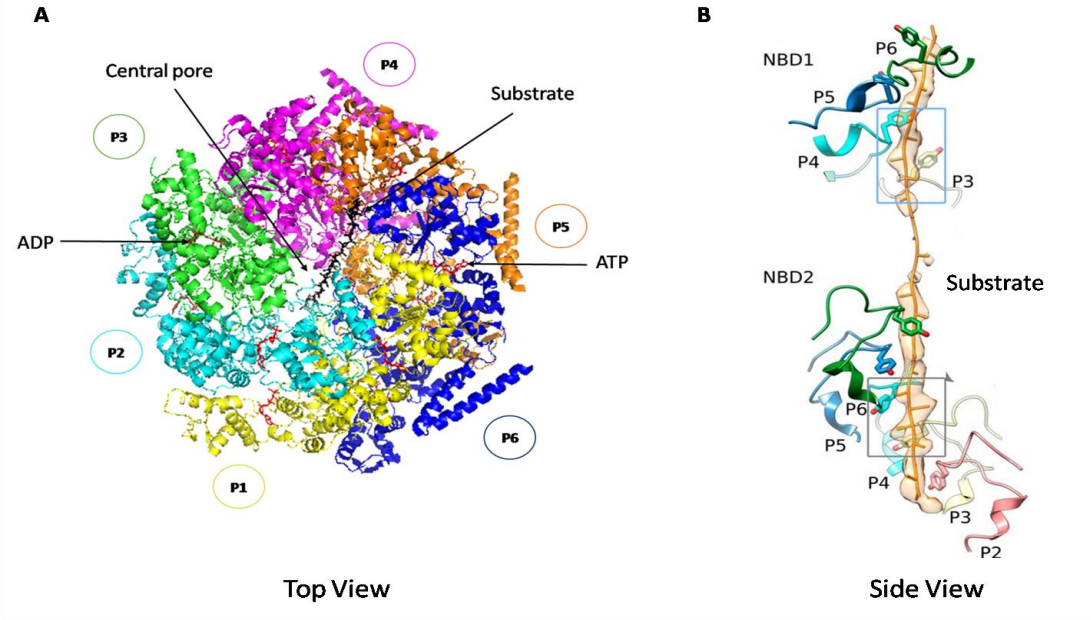

Figure 1.8 Disaggregation of substrate by *M. tuberculosis* ClpB. (A) P1-P6 are subunits of ClpB in nucleotide bound state forming a hexameric assembly. (B) Side view of the hexameric complex showing the disaggregation of substrate through NBD1 and NBD-2. Figure 1.8B is adapted from a previous study (Yu et al., 2018a).

ClpB disaggregates heat denatured Luciferase aggregates by collaborating with DnaK system in *M. tuberculosis* (Lupoli et al., 2016b). Hsp20 enhances the disaggregation activity of ClpB by interaction and stabilizing the aggregates or slows down the aggregation (Lupoli et al., 2016b). ClpB has been shown to be non-essential and dispensable under physiological conditions but it is required to overcome various stress conditions such as Heat, starvation, oxidative e.t.c (Tripathi et al., 2020b; Vaubourgeix et al., 2015). Previous studies show that ClpB is essential for the recovery of *M. tuberculosis* from stresses that promote protein aggregation (Vaubourgeix et al., 2015). Also, ClpB is required to establish infection in host tissues (Vaubourgeix et al., 2015). Recent studies show that the deletion of ClpB also alters the morphology and biofilm formation of *M. tuberculosis* (Tripathi et al., 2020b). ClpB is essential for not only active *M. tuberculosis* but also for its persistence in latent state (Tripathi et al., 2020b).

Hsp104

Yeast Hsp104 performs two major functions in vivo, thermotolerence and prion maintenance in yeast (Chernoff et al., 1995; Shorter and Lindquist, 2004; Shorter

and Lindquist, 2006). Thermolorerence activity requires Hsp104 to collaborate with Ssa1 and co-chaperone Ydj1 for disaggregation of amorphous and amyloid aggregates (Glover and Lindquist, 1998; Shorter and Lindquist, 2004; Shorter and Lindquist, 2006, 2008). Hsp104 plays a crucial role in disaggregation of large pre-formed fibrils into small fibrils which serves as seed for further prion formation and propagation. (Chernoff et al., 1995). Deletion of Hsp104 leads to destabilization of [*psi+*] prion and inhibition of Sup35 propagation in *S. cerevisiae* (Shorter and Lindquist, 2004). Domain characterization studies show that Hsp104 lacking the NTD (Hsp104(ΔN)) is able to dissolve the aggregates but unable to propagate prions(Sweeny et al., 2015b). Further, solution studies revealed a peristaltic pumping mechanism upon ATP hydrolysis, which drives unidirectional substrate translocation across the central pore and is significantly impared in Hsp104(ΔN) (Sweeny et al., 2015b).

Hsp90 Family

Hsp90 belongs to a conserved group of chaperones requiring ATP for their function and is one of the most abundant proteins in cells representing 1–2% of the total cell protein (Buchner, 1999; Frydman, 2001). The Hsp90 family can be divided into 5 different subfamilies according to the HUGO nomenclature committee: cytosolic HSP90A, endoplasmic reticulum localized HSP90B, chloroplast dependent HSP90C, mitochondrial TRAP (TNFR-related protein) and HtpG (high temperature protein G) in prokaryotes.(Chen et al., 2006). Homodimeric Hsp90 is important for all eukaryotic cells however bacterial counterpart (HtpG) is not necessary under stress-free conditions and has only a small growth defect at high temperatures (Honoré et al., 2017; Tanaka and Nakamoto, 1999). Two cytosolic Hsp90 isoforms sharing a 99 percent gene similarity are present in *S. cerevisiae* and *H. sapiens*, one of which is constitutively expressed (Hsc82, Hsp90β) and another is expressed under stress such as heat shock (Hsp82, Hsp90α) (Borkovich et al., 1989).

Owing to the broad variety of functions from morphological development and stress adaptation in Drosophila and Arabidopsis, the Hsp90 family is distinct from other molecular chaperone families, to the key role in assembling and

preserving 26S proteasome (Imai et al., 2003; Pratt et al., 2001; Queitsch et al., 2002; Rutherford and Lindquist, 1998). Hsp90 facilitate the late stage folding and maturing of proteins in coordination with Hsp70 and other co-chaperons. As most of its substrates are signal transduction proteins such as receptors of steroid hormones, signalling kinases, transcription factor, E3 Ubiquitin Ligase and protein-promoting tumors. Therefore, Hsp90 plays a key role in signal-transduction networks, cell-cycle control, protein degradation, and protein trafficking (Richter and Buchner, 2001; Young et al., 2001). The existence of extracellular Hsp90 in tumor cells is a constitutive phenomenon controlled by HIF-1α in around 40 per cent of the types of cancer cells (li et al., 2007), which in effect promotes motility of cells as an important occurrence in cancer and cell motility (Neckers et al., 2009).Therefore, Hsp90 is considered as a potential drug target not only for protein quality control but for various other cellular mechanisms.

Hsp70 Family

The most commonly known chaperone family of Hsp70 contains members with high sequence similarities across prokaryotes to eukaryotes and identical molecular mass. They serve as a platform for the folding, refolding and shuttling of protein aggregates together with trafficking of intracellular proteins (Gupta et al., 2018; Reeg et al., 2016; Rohde et al., 2005; Truman et al., 2012). This suggests that specific cellular roles have evolved from their multiple copies. Similarly, a large number of co-chaperons, Hsp40 (J-domain Protein) and Nucleotide exchange factor, GrpE regulates the Hsp70 activity (Szabo et al., 1994a). Multiple isoforms of Hsp70s are constitutively expressed, but others are expressed in response to different environmental conditions, including heat shock or growth conditions(Genevaux et al., 2007).

Structural organisation of Hsp70

Biochemical and structural analyses of *E. coli*, *T. thermophilus*, *S. cerevisiae* and bovine Hsp70 homologues have characteristic two domains of the Hsp70 protein family; extensively conserved 42-44 kDa N-terminal ATPase domain and C- terminal domain which has been divided into 15-18 kDa Substrate binding

domain and a less conserved 10 kDa lid domain (Young et al., 2004). A 10-12 residue linker which links these two domains mediates the contact between both the ATPase domain and the substrate binding domain (Figure 1.9).

Crystal structures of the Bovine Hsc70 ATPase domain has been solved in the presence of various nucleotides (ADP and ATP) which show that the ATPase domain consists of further two globular sub-domains I and II, each with its own smaller subdomains A and B, separated by a shallow cleft and bound by two helices forming a hydrophobic nucleotide binding pocket (Flaherty et al., 1990). Nucleotide bound Hsp70 complex with one $Mg^{2+}$and two K^{+} ions at the binding sites with the β and γ-phosphate binding loops, associates with the four subdomains (Meyer et al., 2003). The solution structure of the ATPase domain indicated that the leaning movements of the subdomains, which result in the opening and closing of the nucleotide binding cavity, may be regulated by the nucleotide, with the highest opening rate in nucleotide-free form and the lowest in the ATP bound form (Gassler et al., 2001; Zhang and Zuiderweg, 2004).

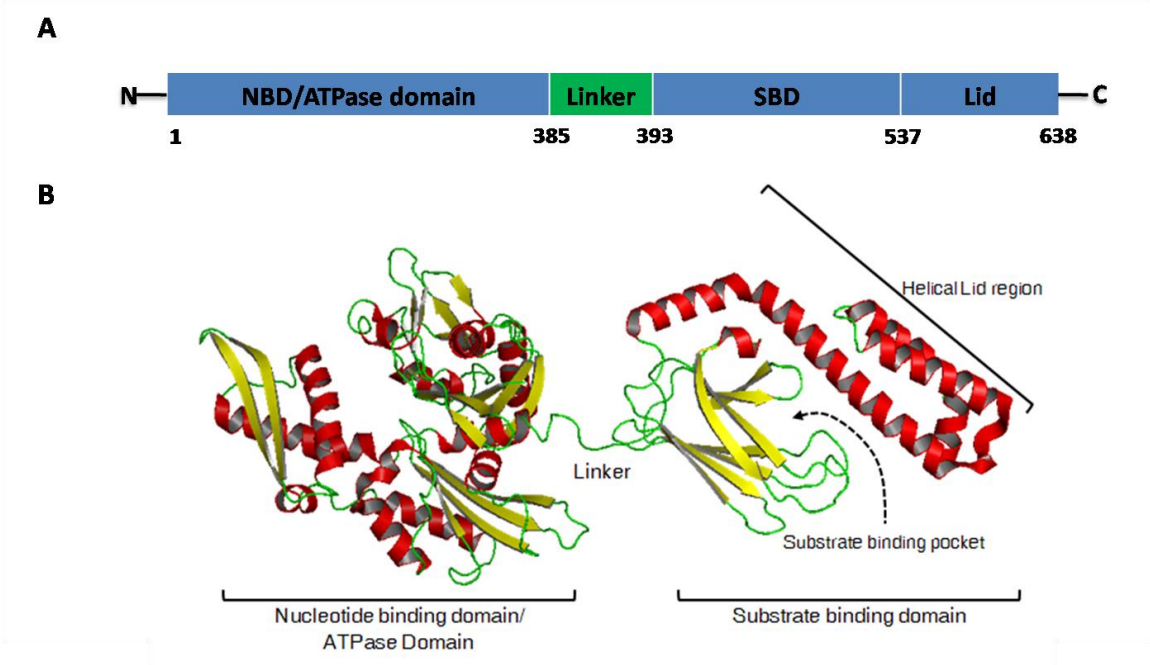

Figure 1.9 (A) Domain organisation of Hsp70 member DnaK from *E. coli* which specifies the domain boundaries of the Nucleotide binding domain (NDB), Linker, Substrate binding domain (SBD) and 10 kDa domain C-terminal. (B) Structure of DnaK describing the various domains along with their structure solved for *E. coli*. Figure is adapted from structure deposited to PDB:2KHO (Bertelsen et al., 2009).

Mechanism of refolding activity of Hsp70

Polypeptide substrate undergoes multiple cycles of binding and release in the Hsp70 assisted refolding, typically at a stoichiometry of one Hsp70 monomer binds to one substrate molecule (Buchberger et al., 1996; Szabo et al., 1994a). For efficient refolding, Hsp70 works in collaboration with its co-chaperones Hsp40 (DnaJ) and GrpE. Target sequences or misfolded polypeptides binding to Hsp70 are usually stretch of 7 to 13 hydrophobic amino acids residues (Schmid et al., 1994; Zhu et al., 1996). Hsp70 family eukaryotic members are shown to be associated with ribosomes and attaches to the nascent chains at the exit tunnel of the ribosome. The mechanism of action of the *E. coli* DnaK, is very well established by a significant number of studies at molecular level.

Hsp40, DnaJ has an affinity for substrate or misfolded proteins and binds them with its hydrophobic C-terminal domain (CTD). DnaJ has highly conserved 70 amino acid residues domain called the J-domain which has the ability to recognize DnaK in ATP bound form. DnaJ recruits the bounded misfolded substrate to DnaK which results in closing of lid. DnaJ significantly accelerates the ATPase activity of DnaK. While DnaK and DnaJ function separately in respect of peptide recognition, the cooperation between DnaK and DnaJ is important for substrate selectivity and accelerating DnaK's ATPase activity.

Prokaryotic Hsp70, DnaK exists in two different forms, (a) open and (b) closed depending on nucleotide bound state. DnaK exhibits low affinity with the substrate in ATP bound form and thus has high binding and release rates, between milliseconds to few seconds, while ADP bound form provides high affinity and low release and binding rates, from a few minutes to few hours. The ATP is hydrolysis locks the DnaK substrate binding pocket in a closed conformation. The ternary complex of DnaK-ADP-Substrate is stable until the other co-chaperone GrpE interacts. GrpE is the nucleotide exchange factor (NEF) which interacts with the NBD of DnaK in the presence of ADP (Harrison, 2003). This change in conformation leads to opening helical lid which frees the ADP and allows the re-binding of ATP. The ATP binding results into

the release of substrate protein in either a partially folded or fully folded state. Hsp70 cycle is described in the figure 1.10 mentioned below.

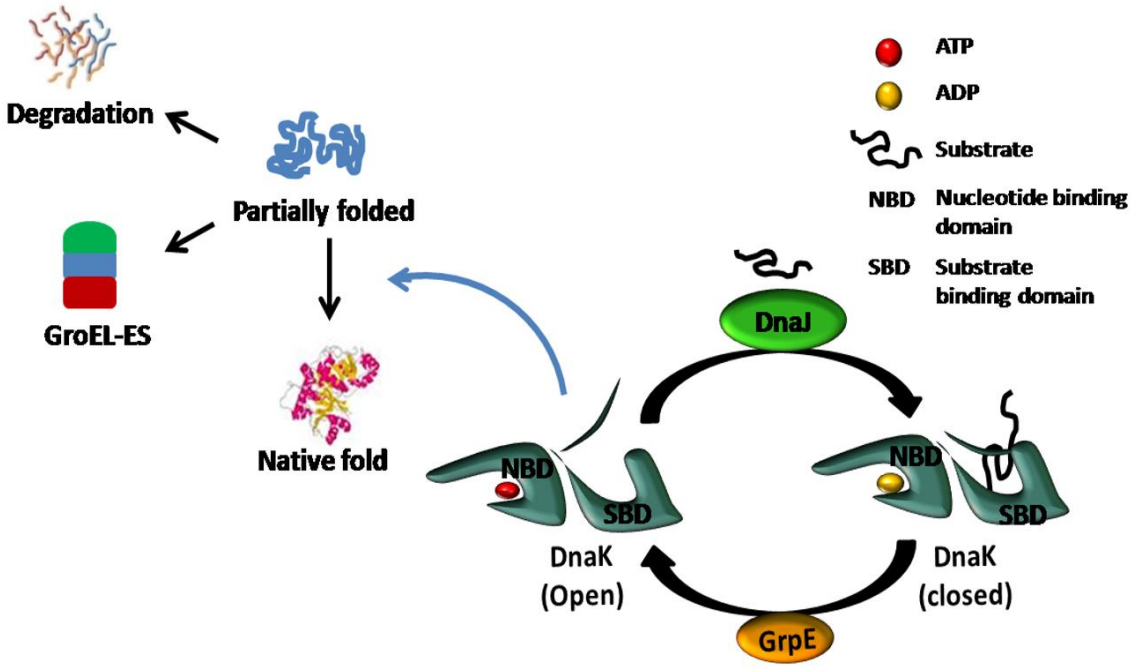

Figure 1.10 Schematic representation of Hsp70 refolding cycle. DnaJ recruits the substrate in the presence of ATP, while GrpE exchanges the ADP with ATP to complete the refolding cycle.

Hsp70 from *M.tuberculosis.*

The *M. tuberculosis* genome contains a single copy of the Rv0350 which encodes for dnaK gene (Cole et al., 1998). Figure 1.11 represents the DnaK operon constituting ORF for DnaK [hsp70, Rv0350] followed by ORF encoding for its co-chaperones GrpE (Rv0351), DnaJ1 (hsp40, Rv0352) and the repressor HspR (Rv0353). These genes with the size of 1878, 708, 1188 and 391 bp translates DnaK and co-chaperones with molecular mass of 66.8, 24.5, 41.3 and 14.1 kDa respectively. Although DnaJ2, an isoform of DnaJ1 is not a part of DnaK operon system but contributes significantly in refolding of substrate (Lupoli et al., 2016b).Previous studies suggests that DnaK operon is regulated by SigH (Rv3223c), a major stress response factor induced in response to oxidative stress, hypoxia and cell wall damage (Mehra et al., 2012; Raman et al., 2001). DnaK of *M. tuberculosis* are stated to be antigenic as they are secretary and are able to evoke immune response (Kim et al., 2018b).

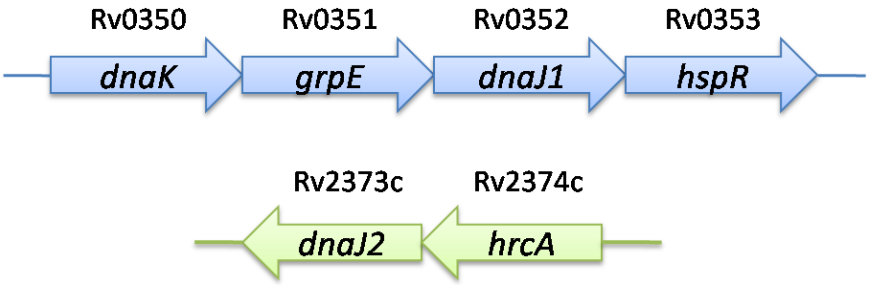

Figure 1.11 Schematics of *dnaK* operon system contains co-chaperones *grpE*, and *dnaJ1* along with the regulator *hspR*. A second gene for *dnaJ2* is present as a part of separate operon system with its regulator *hrcA* in *M. tuberculosis* genome.

The inverted repeats of HAIR elements in the upstream of DnaK operon with close association to HspR indicates additional regulation of DnaK operon system (Stewart et al., 2001; Stewart et al., 2002). In-vitro studies using denatured HspR describes the binding with HAIR element leads to the enhancement of HspR activity in the presence of Hsp70 and its co-chaperones (Das Gupta et al., 2008; Stewart et al., 2001). In the case of HspR knockout mutants, lack of survival also revealed that the indirect role of HspR in virulence could be performed by over-expression of DnaK, that may enhance the host's immune response (Stewart et al., 2003).

Hsp60 Family

The Hsp60 chaperone family is also known as chaperonin, differentiated by multimeric structure of broad ring assemblies assisting misfolded substrate proteins regulated by ATP-binding and release cycles. The misfolded protein or substrates are inserted in the refolding chamber for refolding. Chaperonins are further categorized into two sub-families, Group I and Group II chaperonins based on requirement of their co-chaperonins and localization in the cell.

Group I chaperonins need support from their co-chaperonin, Hsp10, which functions as a lid and is present in the cytoplasm of bacteria, mitochondria and eukaryotic chloroplasts. Group I consist of members found in the cytosol of prokaryotes, membrane of the eukaryotic organelles, mitochondria and chloroplasts involves co-chaperone, GroES or Cpn10. In addition, these

molecules are further characterized by the assembly of homo-tetradecameric ring surrounding two cavities for binding of substrate protein. Example are GroEL /GroES system of *E. coli* and various eubacteria (Bukau and Horwich, 1998; Horwich et al., 2001). Group II chaperonins contain a pre embedded lid translated along with their primary sequence which makes them independent of co-chaperonins for their function. They are usually present in the archea and eukaryotes cytosol (Ranson et al., 1998). Few other well studied members of this group are thermosomes and CCT chaperonins (Gutsche et al., 1999).

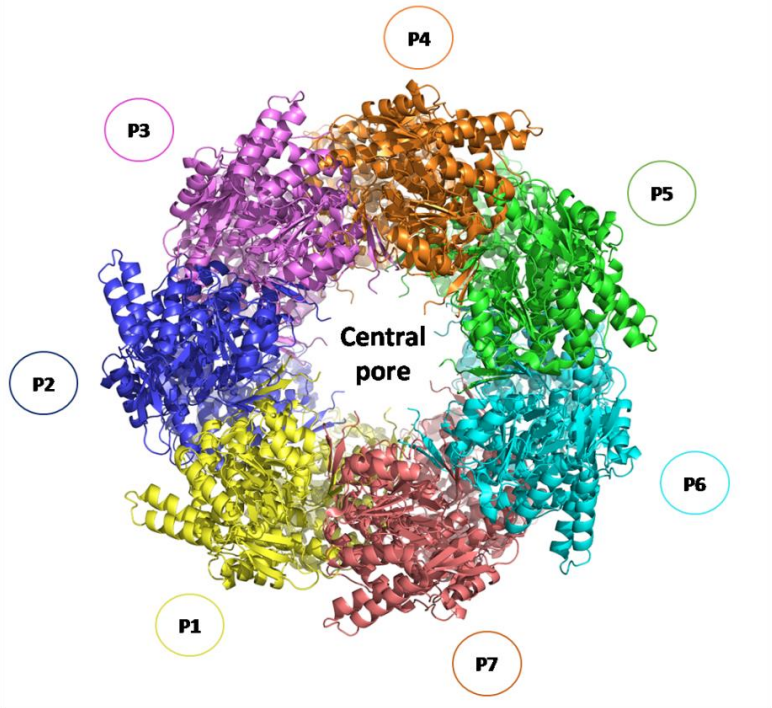

Figure 1.12 shows the crystal structure for Hsp60 resolved at 2Å through X-ray diffraction method from *E.coli* (PDB 1KP8). Structure shows assembly of heptameric ring with a central pore for substrate binding.

Mechanism of action of Hsp60

Most of the current understanding regarding chaperonin's mechanism of action emerged from biochemical, biophysical, and genetic experiments from prokaryotic homolog *E. coli* GroEL. GroEL activity exhibits a complex association with substrate protein based on conformational changes caused by binding and hydrolysis of ATP, along with GroES.(Hartl and Hayer-Hartl, 2002; Ueno et al., 2004; Weissman et al., 1995). Based on different size/cavity for substrate

binding and co-chaperonins, GroEL-ES mechanism can be divided into two pathways, *cis* and *trans*. Majority of the misfolded substrates follow the *cis* mechanism where GroES binds to the same side of the GroEL ring as of the substrate polypeptide and hence this mechanism is termed the *cis* mechanism. Although the holding of large substrate is difficult for central cavity, therefore the efficient folding of these substrates was proposed outside the *cis* opening. Unlike cis folding, it has been shown that only a part of larger substrate bounds to GroEL/ES machinery, indicating the nucleation phase, where remaining portion of the substrate folds independent of co-chaperonins (Chaudhuri et al., 2001). A representation of Hsp60 cycle has been shown in the figure 1.13 mentioned below.

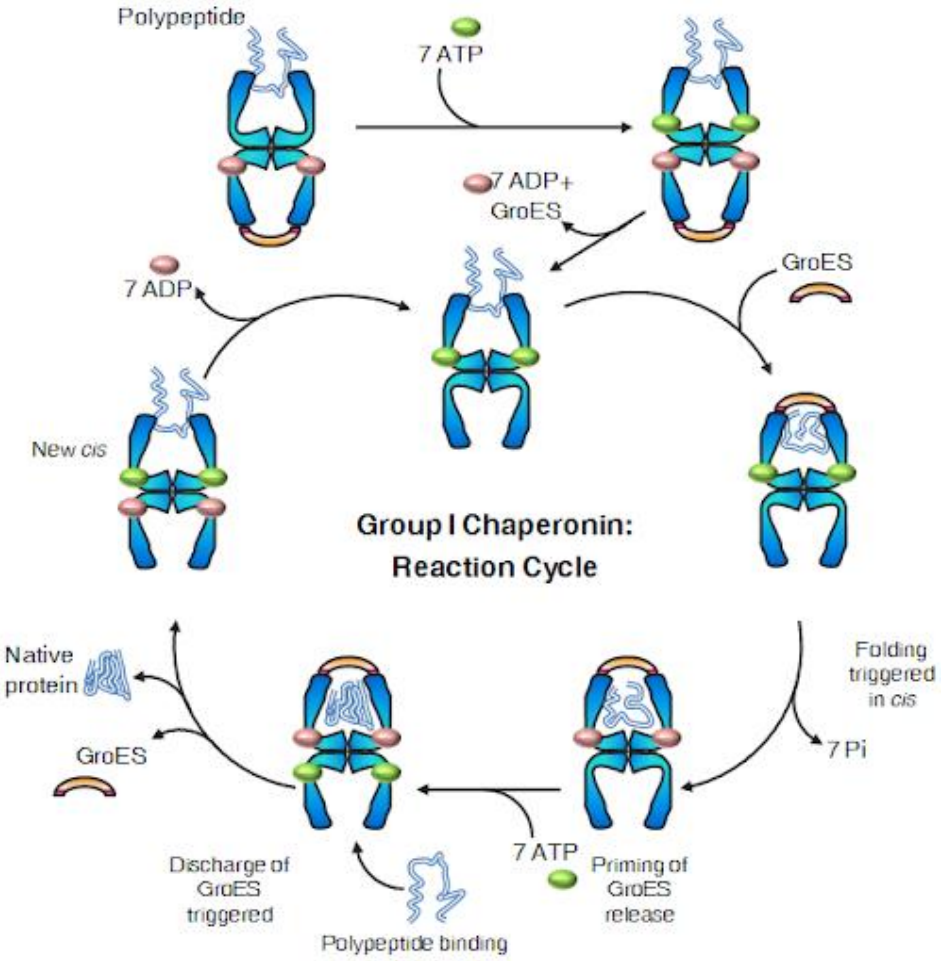

Figure 1.13 A schematic representation of Hsp60 refolding cycle. GroEL binding to polypeptide in the presence of ATP followed by binding of lid GroES results in refolding of polypeptide into is native form. The figure is adapted from previous study(R et al., 2012b).

GroEL from *Mycobacterium tuberculosis*

Genome annotation studies revealed that a few bacterial genomes possess multiple copies of *groEL* genes (Barreiro et al., 2005; Fischer et al., 1993; Karunakaran et al., 2003). GroEL system in *Mycobacterium tuberculosis* is a dimeric complex with ATPase activity that can partially refold proteins in vitro. The *M. tuberculosis* genome harbors two copies of *groEL* genes (*groEL*s). *GroEL1,* encoded by Rv3417c is arranged in an operon, with its co-chaperonin *groES*, encoded by Rv3418c, while the second copy, *groEL2* encoded by Rv0440 exists separately on the genome and is essential for *M. tuberculosis* survival (Cole et al., 1998a; Kong et al., 1993).

Structurally, apical domains of both GroEL1 and GroEL2 have almost identical 3-D structure however, GroEL2 behaves as house-keeping chaperone similar to *E. coli* GroEL. While GroEL1 works as a specialized chaperone with a specific substrate spectrum. Thus, considered to be major player in biofilm formation which probably confers extraordinary starvation survival and resistance to antibiotics (Ojha et al., 2005; Vinod Kumar et al., 2017). GroEL2 is an immune-dominant mycobacterial antigen and both GroEL1 and GroEL2 play role in cytokine stimulation and generating host immune response to *M. tuberculosis* infection (Kumar et al., 2015a).

However, the most prominent feature of *M. tuberculosis* GroELs is their oligomeric state, where they did not form the conventional tetradecameric arrangement in vitro, contrary to predictions, when purified from E. coli. Instead, the proteins were present as lower oligomers (dimers) independent of the presence or absence of factors such as GroES or ATP (Qamra and Mande, 2004; Qamra et al., 2004). In addition, GroES system displays weak ATPase activity for prevention of aggregation of denatured polypeptides (Illingworth et al., 2011).

Hsp40 family

For refolding function in vivo, Hsp70 requires the assistance of its co-chaperones, Hsp40 member DnaJ and GrpE. DnaJ is 41 kDaprotein which

accelerates the ATP-dependant substrate association of misfolded proteins to the substrate binding domain of DnaK. DnaJ can independently binds to misfolded protein or substrate and directs to DnaK with high binding affinity.

Hsp40 family is structurally organized into 4 different domains, J-domain, Glycine/Phenylalanine (G/F) rich domain, Zinc binding domain and C-terminal domain. J-domain binds directly to the nucleotide binding domain of Hsp70. Both Nuclear Magnetic Resonance (NMR) and X-ray crystallography have explored the structure of the J-domain (Jiang et al., 2005b; Pellecchia et al., 2000). The J-domain consists of four alpha-helices linked by a loop containing highly conserved Histidine-Proline-Aspartate residues, also referred as the HPD motif (Qian et al., 1996; Tsai and Douglas, 1996). J-domain is the characteristic feature of all the J proteins of Hsp40 members, which is essential for interaction with Hsp70. Once it associates with Hsp70 in the solution, J domain mobility is significantly decreased (Landry, 2003). (G/F) Domain is responsible for keeping the DnaK in high affinity state resulting in the maximal stimulation of DnaK's ATPase activity. Zinc binding domain is cysteine rich domain contains Zinc binding motif (Cys-X-X-Lys-X-Gly-X-Gly) responsible for binding 2 Zn^{2+} ions per DnaJ monomer (Banecki et al., 1996; Liberek et al., 1991; Szabo et al., 1996). The C-terminal domain is mainly involved in dimerization which helps DnaJ for interaction with DnaK(Liberek et al., 1991). Figure 1.14 represents the domain organization of Hsp40 class.

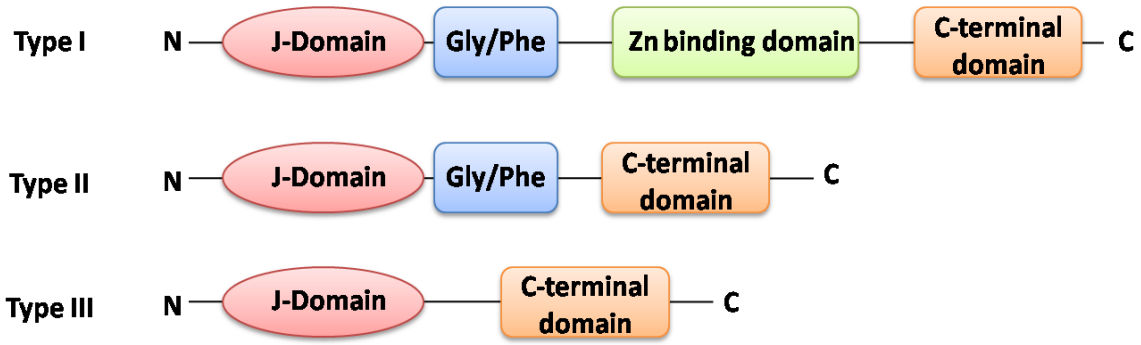

Figure 1.14 Structural domain organization of Hsp40 class. Type I is composed of 4 domains. J-domain consists of conserved HPD motif responsible for binding to Hsp70,

followed by glycine/phenylalanine rich region. Zinc binding domain consists of 2 zinc binding sites per monomer of DnaJ. C-terminal domain is involved in dimerization of Hsp40 and binding to polypeptides. Type II class has similar domain organization excepts Zinc binding domain, while type III has only J-domain and C-terminal domain for dimerization.

Based on presence and absence of these domains Hsp40 co-chaperones have been divided into 3 types. Figure 1.14 represents the division of Hsp40 chaperone family on the bases of various domains. *E. coli, M. tuberculosis, S. cerevisiase* DnaJ and Ydj1 are members of Type I class which contains all 4 domains. Type II class members Sis1, CbpA contains only 3 domains with a missing zinc binding domain. Type III class is less conserved and contains only J-domain and C-terminal domain. Type III proteins may require higher specificity for their Hsp70 partner and substrates and only recognize a restricted subset of substrates. Sec63, Zuol are the members of Type III class.

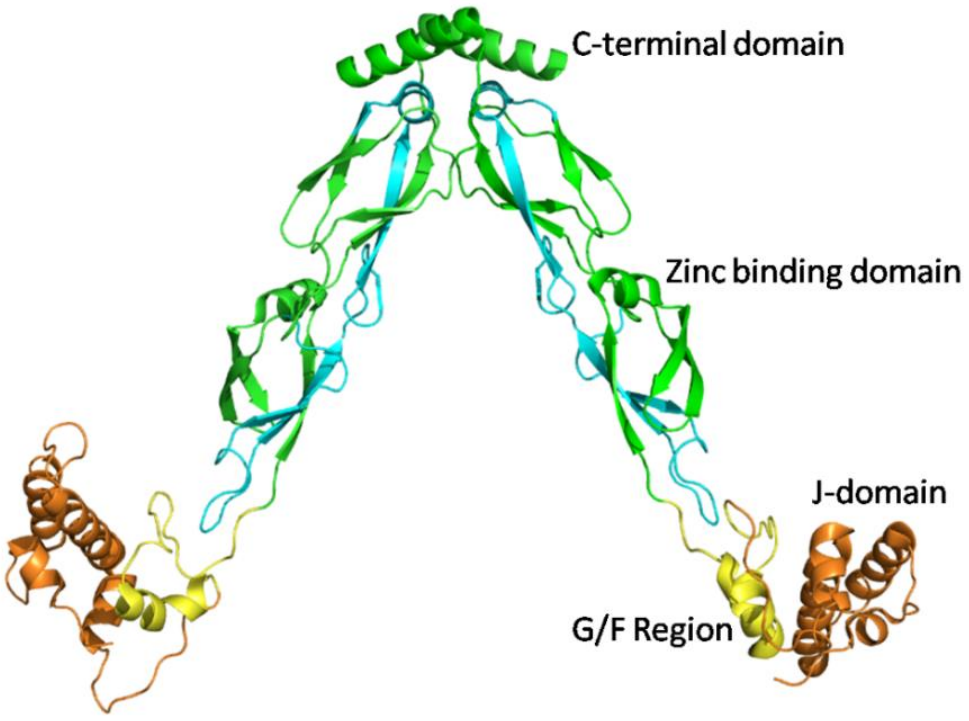

Figure 1.15 Represents the crystal structure for dimeric DnaJ from *T. hermophilus* (PDB: 4J80) resolved at 2.9 Å highlighting its various domains.

Hsp40 from *M. tuberculosis*

Mycobacterial genome harbours two copies of Hsp40, DnaJ1 is encoded by Rv0352 which translates into 41.4 kDa protein, while DnaJ2 is encode by Rv2373c and forms a 40.5 kDa protein (Cole et al., 1998a). Though DnaJ1 is a part of *DnaKJE* operon system but both the isoforms contribute towards refolding of substrate (Lupoli et al., 2016b). Previous studies shows that DnaJ1 or DnaJ2 can be dispensable individually but required collectively for Mycobacterial growth (Lupoli et al., 2016b). The refolding capabilities of DnaJ1 or DnaJ2 have been explored and but presence of two isoforms suggest broader functioning of J-proteins in the *M. tuberculosis* system.

Small Heat shock proteins (sHsps)

The small family of Heat shock proteins (sHsps) is the largest of the chaperone families containing Hsps with molecular masses ranging from 10-40 kDa and sharing a retained 90-residue C-terminal Alpha-crystalline Domain (Boston et al., 1996; Veinger et al., 1998; Vierling, 1997). The sHsps are ubiquitos, but primarily found in plants and are expressed in response to heat shock and other cellular stresses. While sHsps are unable to refold the non-native proteins on their own but they can bind to the unfolded substrates which stabilizes and prevents their aggregation (Ehrnsperger et al., 1997; Lee et al., 1997; Reddy et al., 2000). Further these substrate proteins are transferred to primary ATP powered chaperones such as Hsp100-Hsp70 bi-chaperonic system for subsequent refolding (Lee and Vierling, 2000; Mogk et al., 2003; Vierling, 1997).

Small heat shock proteins from *Mycobacterium tuberculosis*

M. tuberculosis contains two sHsps, Acr1 (α- crystalline related protein 1/ hsp16.3/ 16 kDa antigen), encoded by gene *hspX* (Rv2031c) and Acr2, encoded by *Hsp20* (Rv0251c) gene (Cole et al., 1998a).

Acr1, a stable protein prevents the aggregation of denaturing proteins and misfolding of nascent peptides under different stress conditions. It is synthesized

at a low level in logarithmic phase whereas; its synthesis markedly increases and becomes abundant under the transition from log to stationary phase. sHsp16.3 is a non-secretory protein that plays a major role in bacterial cell wall thickening. Its activity is affected in temperature range 25 to 37.5°C without any change in its native oligomeric structure (Mao et al., 2001). With further elevation in temperature such as during fever from infection, sHsp16.3 exposes its hydrophobic surfaces and dissociates into individual monomers which accelerates the chaperone like activity by binding to aggregation prone substrates. Thus, it is a potential important component facilitating the survival of *M. tuberculosis* as it accumulates during prolonged periods of macrophage infection (Yuan et al., 1998).

Acr2, a 18 kDa small heat shock protein has 30% amino acid sequence similarity to Acr1 of *M. tuberculosis*. Acr2 is significantly induced under starvation conditions, uptake by naive and activated macrophages and heat shock at 45°C (Mao et al., 2001). The stressed induced expression of Acr2 is regulated by the heat shock repressor protein HspR and by two component system called phoPR whereas SigE and SigH are alternative sigma factors that also down regulates the expression of acr2(Sun et al., 2004).Recently, Hsp20 has been shown to play an important role for disaggregation of substrate by *M. tuberculosis* ClpB. Hsp20 stabilizes the heat denatured luciferase aggregates and results in efficient disaggregation. (Lupoli et al., 2016b).

Nucleotide Exchange factors (NEF)

Nucleotide exchange factors (NEFs) are proteins which facilitate the release of ADP and allow rebinding of ATP resulting in the completion of Hsp70 refolding cycle. Figure 1.16 represents 4 different classes of NEFs GrpE, Hsp110, HspBP1 and Bag domain containing proteins (Mehdi, 2009). NEFs do not directly exchange one nucleotide with other but it is the high concentration of ATP in cells that occupies void created by ADP release (Brehmer et al., 2004; Gassler et al., 2001; Polier et al., 2008).

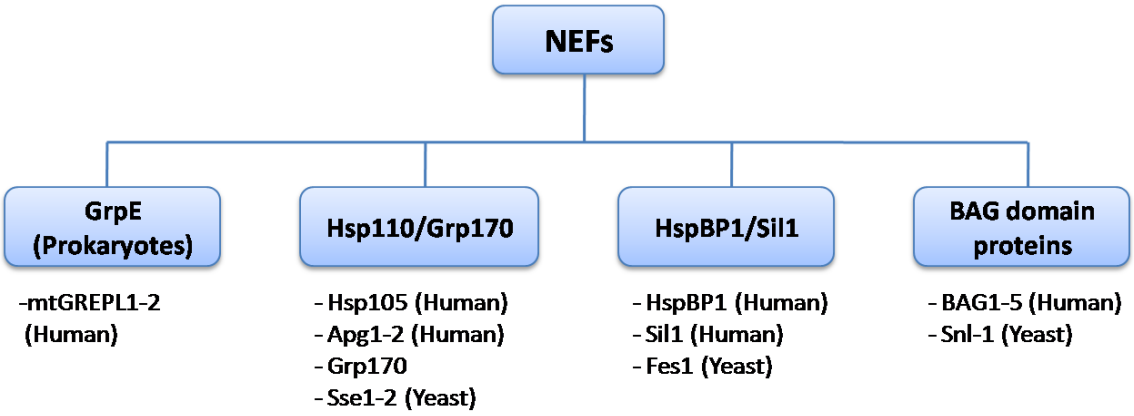

Figure 1.16 Classification of nucleotide exchange factors. Shown here is the schematic diagram for various nucleotide exchange factors present in prokaryotes and eukaryotes.

GrpE

GrpE (*Gro-P* like protein E) is a nucleotide exchange factor which plays an important role in regulation of Hsp70 protein refolding machinery. It is a heat-inducible protein which inhibits the aggregation of unfolded proteins in the cytoplasm under stress (Bracher and Verghese, 2015a; Harrison, 2003). GrpE was initially discovered for its role in λ-phage replication (Saito and Uchida, 1977). GrpE mutational studies shows the inhibition of λ-phage replication and significant decrease in replication *in-vitro,* and this defect was recovered by *in-vitro* overexpression of DnaK (Saito and Uchida, 1977). GrpE plays a major role in the replication of λ-phage by promoting bi-directional DNA unwinding through interacting with DnaK after the assembly of DnaB or other replication factors (Wyman et al., 1993).

GrpE performs the nucleotide exchange function by releasing the bound ADP from DnaK. DnaK binds ATP with low affinity in its open conformation and has a high exchange rate for proteins that are unfolded. When the substrate is transferred to the ATP bound form of DnaK by DnaJ, resulting into the hydrolysis of ATP to ADP. The DnaK forms a stable complex with ADP which requires the interaction of GrpE to bind DnaK and alters its conformation to free the bound ADP from DnaK's N-terminal ATPase domain (Harrison et al., 1997a).Transient kinetics shows a strong association of GrpE and DnaK with an affinity of 1nM (Harrison, 2003). Protein-protein interaction studies through Surface plasmon

resonance and Isothermal titration calorimetry also reveals the tight binding of DnaK and GrpE in 30-150 nM range (Chesnokova et al., 2003; Moro et al., 2007). The binding of GrpE to DnaK-ADP complex substantially decreases DnaK's affinity towards ADP 200-fold and accelerates the release of nucleotides by 5000-fold. This approach encourages DnaK's de novo refolding of misfolded substrate(Packschies et al., 1997).Biophysical studies from *T. thermophiles* GrpE show the down regulation of activity with increase in temperature suggesting a thermosensing activity(Bracher and Verghese, 2015a). The reversible unfolding of alpha-helices occurs at 35°C with Tm of 50°C, influences the structural stability of GrpE and prevents the binding of GrpE to DnaK. The DnaK's thermal regulation slows the rate of refolding and prevents the unfolded proteins from accumulating at elevated temperatures in the cytoplasm (Bhandari and Houry, 2015; Winter and Jakob, 2004)

Structure of GrpE

GrpE is a bipartite structure consists of three distinct domain, N-terminal unordered region, α-helices and small β-sheet domain at C-terminus. GrpE, either in solution or crystal, forms strong homodimer indicating asymmetrically interaction with only one ATPase domain of Hsp70 (Harrison et al., 1997a). Figure 1.17 shows the crystal structure for *E. coli* GrpE.

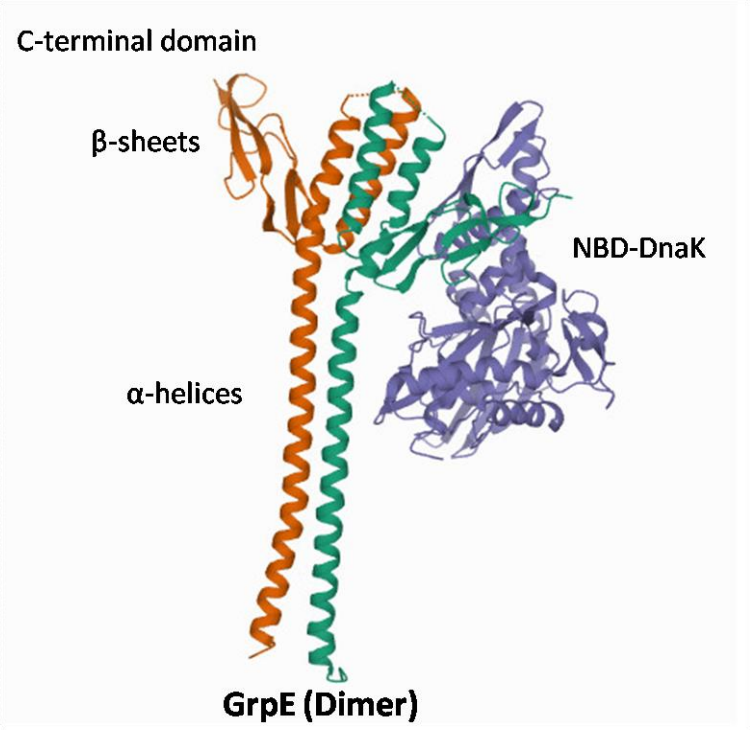

Figure 1.17: Crystal structure of dimeric GrpE in complex with Nucleotide binding domain of DnaK from *E. coli* resolved at 2.8 Å shows hetro3-mer symmetry. C-terminal region is directly involved in the asymmetrical binding with NBD region of DnaK.

The N-terminal domains contains 1-33 amino acids residues which have not been solved through the crystallization due to their unordered nature(Bracher and Verghese, 2015a).N-terminal domain is followed by four α-helices which forms together as bundles without showing any super helical twisting due to presence of hydrophobic residues (Bracher and Verghese, 2015a). Part of these helices are responsible to interacting Domain IIB of DnaK and also acts as thermosensors (Winter and Jakob, 2004). C-terminal is composed of two closely packed β-sheets which appears as arms from both the helices. The β-sheet adjacent to DnaK binds specifically with its ATP binding by incorporating itself into the cleft and inducing a conformational change in Domain IIB triggering the release of ADP (Blatch and Edkins, 2015).

GrpE from *M. tuberculosis*

M. tuberculosis genome contains single copy of *grpE* gene encoded by Rv0353 and is a part of DnaK/J/E operon system (Cole et al., 1998a; Lupoli et al.,

2016b). GrpE is 24.5 kDa monomer which is expected to form dimer in solution as shown before for its homolog from *E. coli* (Gelinas et al., 2002; Schönfeld et al., 1995). Previous studies show that GrpE plays a major role in substrate refolding and prevention of aggregation by mycobacterium DnaK system. GrpE enhances ATPase and luciferase refolding activity of bi-chaperonic system (Lupoli et al., 2016b; Tripathi et al., 2018). Recent studies on GrpE suggest that GrpE not only performs the refolding function but also involves in promoting naïve $CD4^+$/$CD8^+$T cell differentiation toward Th1-type T-cell immunity through interaction with dendritic cells (Kim et al., 2018a). GrpE induces maturation of dendritic cells by up-regulating the expression of cell surface molecules such as CD80, CD86, and MHC class I and II and production of several other pro-inflammatory cytokines TNF-α, IL-1β, IL-6, and IL-12p70 in dendritic cells (Kim et al., 2018a).

GrpE may contribute to the enhanced understanding of host-pathogen interactions as well as providing a rational basis for the discovery of new potential targets to develop an effective tuberculosis vaccine.

Hsp110 family

Hsp110 family member, Sse1 is the most prevalent NEF in yeast, accounting for around 10% of the overall Hsp70 in cytosol (Dragovic et al., 2006; Ghaemmaghami et al., 2003; Kulak et al., 2014). Its nearest counterpart in yeast is stress-inducible Sse2. Sse1 and Sse2 share 78% sequence similarity, although their resemblance to the Hsp70 Ssa family is 70%. Deletion of SSE1 results in a partial growth defect in yeast and also a constitutive activation of the heat shock response (Raviol et al., 2006). In addition, over-expression of Sse1 contributes to sluggish growth phenotypes indicating potential competition between the total ATP usable for Sse1 and Hsp70s (Liu et al., 1999). Sse1 is structurally categorized into 3 functional domains Nucleotide binding domain (NBD), Substrate binding domain-α (SBD-subject) & Substrate binding domain-β (SBD-β) but varies in the connecting region between SBD-α and SBD-β domain (Liu and Hendrickson, 2007). The Hsp110 class members functions

independent of the Hsp70s serve as a holdase to retain the substrate in a folding capable state (Easton et al., 2000).

HspBP1 Family

Fes1, HspBP1 family member stands for Factor Exchange for Ssa1 from *S. cerevisiae* (Kabani et al., 2002). There are two isoforms present for Fes1 in *S. cerevisiae*. Fes1-L is directed at the nucleus, whereas Fes1-S is found throughout the cytosol (Gowda et al., 2016). Deletion of Fes1 contributes to a constitutively heat shock response and results in slow growth of the phenotype at optimal condition and a very prominent growth defect under conditions that facilitate protein misfolding (Abrams et al., 2014). The human counterpart of Fes1 is HspBP1, which is distinguished by the presence of armadillo repeats. These repeats allow Fes1 to associate with sub-domain II B in the NBD of Hsp70, thereby helping to substitute ADP for ATP (Shomura et al., 2005). Fes1 and Sse1 also connect the Hsp70 chaperone cycle to degradation machines through ubiquitination of misfolded proteins or targeted substrate by interaction with E3 ligases (Gowda et al., 2013).

Bag domain family

Snl1 belongs to Bcl-2-associated athanogene (BAG) class of NEFs. Bag domain assists in various other cellular functions such as cell signalling pathways (Wang et al., 2013), refolding cycle (Takayama et al., 1997), degradation pathways (Kriegenburg et al., 2014) and transcription process (Frebel and Wiese, 2007). Snl1 is distinctive member of NEF family localized to ER membrane. Snl1 was initially described as a repressor of Nup116-C lethality, suggesting its involvement in nuclear pore biogenesis (Ho et al., 1998).It has been reported that the Bag domain of Snl1 cure [URE3] prion by modulation of Hsp70 in *S. cerevisiae* but not [PSI+] prion (Kumar et al., 2014).

Materials and Methods

2.1 Materials

All the materials required for this study was of analytical grade and purchased from sources through previously cited studies or mentioned below.

Table 2.1 List of material used in this study.

Sr. No	Materials	Source
1.	Enzymes for molecular cloning	Thermo Fermentas
2.	ATP, ADP, ATP-γS, AMP-PNP	Sigma
3.	Affinity purification resins	Thermo scientific/Qiagen
4.	Buffers and Salts	Sigma
5.	Dialysis tubing	Thermo scientific
6.	Centrifugal devices	Pall Biotech
7.	LB Media	BD Difco
8.	Antibiotics	Sigma
9.	Primers (Cloning and Sequencing)	Sigma

2.2 Methods for mycobacterial studies.

2.2.1. Construction of plasmids.

The gene encoding $ClpB_M$ was PCR amplified from *M. tuberculosis* H37Rv genomic DNA. The $ClpB_M$ was amplified with NdeI and HindIII restriction sites at 5' and 3' end respectively. The amplified PCR product was digested with NdeI and HindIII, and cloned into pET22b digested with the same enzymes to generate pET22b-$ClpB_M$. Similarly, $DnaK_M$ was cloned into pET29bHtv (Kumar et al., 2015b) using restriction sites BamHI and XhoI to construct plasmid pET29bHtv-$DnaK_M$. The plasmid pET29bHtv-$DnaK_M$ encodes from 5' to 3' direction Hexa-His tag, TEV protease recognition motif followed by $DnaK_M$. The gene encoding $DnaJ1_M$ was cloned into pE-SUMO plasmid (kind gift from Dr. Yogender Pal Khasa) using restriction sites BamHI and XhoI to generate plasmid pE-SUMO-$DnaJ1_M$. The plasmid pE-SUMO-$DnaJ1_M$encodes from 5' to 3' direction His_6-tag, SUMO tag, SUMO recognition motif followed by $DnaJ1_M$. The gene encoding $GrpE_M$ was sub-cloned at BamHI and XhoI restriction site in plasmid pET29bHtv to generate pET29Htv-$GrpE_M$. The plasmid encodes (from 5' to 3' direction) for His_6-tag, TEV recognition motif followed by $GrpE_M$.

Gene encoding *E. coli* ClpB was cloned into pET22b using restriction sites NdeI and XhoI. Plasmids encoding gene for *E. coli* DnaK, DnaJ and GrpE were kind gifts from Prof. Hideki Taguchi(Niwa et al., 2012).

Plasmids containing gene for expressing Tev and Ulp-1 protease were kind gifts from Dr. Krishan Gopal (Kaur et al., 2018) and Dr. Johannes Buchner.

All the clones were further confirmed by sequencing.

2.2.2. Protein Expression and purification.

All proteins were expressed in *E. coli* strain Rosetta (DE3) (Invitrogen).

2.2.2.1 Purification of *M. tuberculosis* Chaperones.

The cells harboring plasmid pET22b-$ClpB_M$ were grown in liquid growth media (LB), and $ClpB_M$ expression was induced at $O.D._{600nm}$ ~ 0.6 by 0.3mM IPTG. The cells were further grown at 37ºC for 3h before harvesting by centrifugation at 6000rpm for 15 minutes. The cells were resuspended in buffer A (50mMTris-Cl, 50mM NaCl, 1mM EDTA, 1mM DTT and 5% glycerol) and stored at -80 ºC overnight. The suspension was thawed at 30ºC and lysed using lysozyme (final concentration 6mg/ml) followed by sonication. The supernatant was collected and loaded onto a Q-sepharose FF anion exchange column (GE Healthcare). The column was first washed with buffer A, and elution was carried out using a linear gradient of 0-1M NaCl in buffer A. The eluted fractions were analyzed on SDS-PAGE, and the $ClpB_M$ was found in fractions eluted at 520mM NaCl. Fractions containing $ClpB_M$ were pooled and loaded onto a Superdex-200 PG column (GE Healthcare). The protein fractions were eluted with buffer A containing 100mM NaCl, 10mM $MgCl_2$ and 5% glycerol. The collected fractions were loaded onto SDS-PAGE, and those containing purified $ClpB_M$ were pooled, concentrated and stored at -80°C until used for various assays. The yield of purified $ClpB_M$ was about 10-12mg/L.

The His_6-tagged$DnaK_M$ and $GrpE_M$ were expressed in *E. coli,* similar to as described above for $ClpB_M$. The protein was purified from cellular lysate using Ni^{2+}-NTA (Qiagen)affinity purification. First, the cell debris and insoluble fraction

was separated from cellular lysate using centrifugation, and the supernatant was loaded onto Ni^{2+}-NTA column pre-equilibrated with in buffer B (25mM Hepes pH 7.4, 150mM NaCl, 20mM KCl and 20mM $MgCl_2$). The impurities were washed with buffer B containing 20mM imidazole. The proteins were further eluted using buffer B containing 300mM imidazole. The His_6- tag was later cleaved by overnight incubation of purified protein with His_6-TEV protease at a molar ration of 20:1 at 4°C. The His_6-TEV was removed by incubating the mixture with Ni^{2+}-NTA resin. Purified $DnaK_M$ and $GrpE_M$ were stored in -80ºC with a final yield of 10 mg/L and 6 mg/L.

His_6-SUMO tagged $DnaJ1_M$ was first purified using a method similar to as described above for $DnaK_M$ except that cobalt (Cat. no #89965, Thermo Scientific) based Talon metal affinity resin was used instead of Ni^{2+}-NTA. The His_6 and SUMO tag was cleaved by incubating purified protein overnight with Ulp-1 protease at 4°C (molar ratio $DnaJ1_M$/Ulp-1 :10/1) in Buffer C (25 mM Hepes pH 7.1, 400 mM NaCl, 10 mM $MgCl_2$ and 10% Glycerol). The His_6-Ulp-1 was removed using cobalt-based affinity purification. Whenever required the protein was further purified using superdex-200 (prep grade) column. The protein was stored at -80ºC until used for various assays.

2.2.2.2 Purification of*E.coli* Chaperones.

E. coli ClpB, was purified using similar method mentioned above for $ClpB_M$. The yield of *E. coli* ClpB was found to be about 5-6mg/L.

The *E. coli* His_6-tagged DnaK and GrpE were expressed in *E. coli,* similar to as described above for $ClpB_M$. The protein was purified from cellular lysate using Ni^{2+}-NTA affinity purification (Qiagen). First, the insoluble fraction was separated from cellular lysate using centrifugation at 13000 rpm for 45 minutes, and the supernatant was loaded onto Ni^{2+}-NTA column pre-equilibrated with buffer B. The impurities were washed with buffer B containing 40mM imidazole. The proteins were further eluted using buffer B containing 300mM imidazole. Purified DnaK and GrpE were stored in -80ºC with a final yield of 4.0 mg/L and 5.5 mg/L, respectively.

E. coli His_6-tagged DnaJ was purified using a method similar to as described above for mycobacterial $DnaJ1_M$with a final yield of 1.8 mg/L. The protein was stored at -80ºC until used for various assays.

2.2.2.3 Purification of *S. cerevisiae* chaperones.

Yeast Hsp104 was purified as described before (Sweeny et al., 2011) with some modifications. Briefly, cells harboring plasmid pROEX-Htv-Hsp104 were induced at $O.D._{600nm}$ ~ 0.6 by the addition of 1mM IPTG at 15ºC for 16 hours at 250rpm. Cells were harvested, and lysed similar to as described above for $ClpB_M$. The lysate was centrifuged at 12000rpm for 30 minutes, and the protein was partially purified using Ni^{2+}-NTA based affinity purification. The eluted fractions containing Hsp104 were pooled and subjected to anion exchange chromatography on resource Q column and eluted over a linear gradient of 0.050 - 1M NaCl in buffer A. The fractions containing purified protein were pooled and His_6-tag was cleaved by overnight incubation of purified protein with His_6-TEV protease at molar ratio of 20:1 (Protein: TEV protease) at 4°C in buffer D (20 mM Hepes pH 7.4, 150 mM KCl, 10 mM $MgCl_2$).The His_6-TEV was removed by incubating the mixture with Ni^{2+}-NTA resin. The protein purity was confirmed using SDS-PAGE and purified Hsp104 was stored at -80ºC with a final yield of 2 mg/L for further use.

2.2.2.4 Purification ofHis_6-Tag cleaving proteases.

Tev-protease and Ulp-1 protease, expression was induced at $O.D._{600nm}$ ~ 0.6 by 0.3mM IPTG. The cells were further grown at 18ºC for 16h before harvesting by centrifugation at 6000rpm for 15 minutes. Proteases was purified using similar method as described above for $DnaK_{M.}$ Proteins were stored in -80ºC with final yield of 8 mg/L and 4mg/L.

The protein purity for all the proteins was examined on SDS-PAGE and identity of all the chaperones were confirmed by intact mass through LC-MS (Agilent Technologies). Protein concentrations were measured by BCA method (Thermo Pierce).

2.2.3. Circular Dichroism studies.

The secondary structure was examined by monitoring Far-UV circular dichroism (CD) of proteins from 190nm to 250nm on a Jasco J-815 instrument. The spectra was recorded with 1.2µM $ClpB_{M,}$, 2.5µM $DnaK_M$, 5µM $DnaJ1_M$ and 5µM $GrpE_M$ in buffer containing 2.5mM Hepes (pH 7.4), 10mM NaCl,1mM $MgCl_2$, 0.5 % Glycerol. The wavelength scan was performed at a scan rate of 10nm/min using 1mm path length cuvette at 25°C. Each spectrum was averaged over three scans. For thermal denaturation, the protein was incubated with increasing temperature from 25°C to 95°C, and the change in CD signal was monitored at wavelength of 222nm. The temperature was increased at rate of 1°C/min. Similarly to examine the reversibility of thermal unfolding, temperature was decreased at rate of 1°C/min from 95°C to 25°C, and CD signal was monitored at same wavelength of 222nm. Raw CD signal (Φ_{obs}) was converted to Mean Residue Ellipticity (Φ_{MRE}) as follows:

$$\Phi_{MRE} = (100^{*}\Phi_{obs})/[d^{*}C^{*}(n - 1)]$$

Where Φ is the observed ellipticity (in degrees), C is protein concentration (molar), d is path length (in centimeters) and n is the total number of amino acids in the protein.

2.2.4. Size exclusion chromatography.

Size exclusion chromatography was employed to examine oligomerization of protein in its apo or nucleotide bound state. The analytical superdex-200 gel filtration column was first calibrated with globular proteins of known molecular weight. The proteins RNase A, Carbonic anhydrase, Ovalbumin, BSA, Conalbumin, Catalase and Blue dextran were used as standards for the estimation of oligomeric states of chaperones. About 50-150µg of desired standard protein or purified chaperone was loaded under native conditions onto analytical superdex-200 gel filtration column (column volume 24ml). The elution profile of proteins was monitored at both 220nm and 280nm. The apparent molecular weight of the chaperone was calculated using standard curve of elution volume versus log of molecular weight of standard globular proteins.

To examine the effect of nucleotides on $ClpB_M$ or $DnaK_M$ oligomerization, the proteins (21.5µM $ClpB_M$ and 10µM $DnaK_{M)}$ were pre-incubated with 5mM of nucleotides for 15minutes at 4°C before loading onto the gel-filtration column.

2.2.5. Analytical Ultracentrifugation Studies.

Sedimentation velocity experiments were performed using Beckman optima XL-I analytical ultracentrifuge. The samples were centrifuged at 40,000 rpm (160992 X g) at 20ºC using four place analytical titanium rotor (An-50 Ti) with double sector centre pieces of sample path length 1.2cm. Purified proteins were dialyzed against appropriate buffer, and buffer alone was used as reference. Radial absorbance scans were monitored with 3 minutes intervals using UV absorption detection system at wavelength of 280nm. The obtained data was further analyzed with 'SedFit analysis' program using continuous distribution(s) model based on Lamm equation.

$$\partial\chi/\partial t = 1/r.\ \partial/\partial r\ [rD.\ \partial\chi/\partial r - s\omega^2 r^2 \chi]$$

Where χ is the concentration, *D* is diffusion coefficient, *s* is the sedimentation coefficient of the particle, *t* is the time, *r* is the radius from the centre of rotation, and ω is the angular velocity of the rotor.

The partial specific volume (υ) of $ClpB_M$, buffer viscosity (η) and buffer density (ρ) were measured by Sednterp program. For $ClpB_M$, υ=0.740517, η=0.010399 poise, ρ=1.0072.For $DnaK_M$ and $GrpE_M$, υ= 0.735689 and υ= 0.718455 respectively, η=0.0104258 poise, ρ=1.00876 and for $DnaJ1_M$, υ=0.729971, η=1.4661 poise, ρ=1.05240. Standard sedimentation coefficients of samples at 20°C were represented as $S_{20,W}$.

To determine the effect of nucleotides, 5 mMof ATPor ADP was added to protein samples and incubated for 30 minutes at 4°C prior to sedimentation experiments.

To study the effect of different salt concentrations on $ClpB_M$ oligomerization, the $ClpB_M$ was dialyzed against buffers containing desired NaCl concentration (100mM, 200mM or 300mM). The dialyzed protein sample was used for sedimentation experiments with dialysate as reference buffer.

2.2.6. Native PAGE analysis for $DnaK_M$ and $GrpE_M$.

The proteins were added to loading buffer (62.5mM Tris-HCl, pH 6.8, 25% glycerol, 0.01% bromophenol blue) lacking any reducing or denaturaing agent, and incubated at room temperature for 10 minutes before loading onto10% Native-PAGE. The proteins were separated by electrophoresis at 150V for 5 hours at 4°C using running buffer composed of 25mM Tris-HCl, pH 8.3, and 192mM glycine.

2.2.7. Surface plasmon resonance studies.

The surface plasmon resonance studies with $DnaK_M$ and $GrpE_M$ were performed at 25°C on Biacore T-200 (GE Healthcare) with series S, NTA sensor chip. The purified His_6-tagged$GrpE_M$was immobilized to 100-120 Response Unit (RU) on NTA surface (Series S NTA sensor chip, GE). Varying concentrations of $DnaK_M$ pre-incubated with ADP (5 mM) was passed over $GrpE_M$ coated sensor surface, in a running Buffer B with a flow rate of 30 µL/min. The binding of $DnaK_M$ with immobilized $GrpE_M$ was monitored by increase in RU as $DnaK_M$ was flown over $GrpE_M$. Both association and dissociation were monitored for 120 seconds. An activated and further blocked flow cell without $GrpE_M$ was used as the negative control to correct non-specific increase in RU signal. Equilibrium dissociation constant was calculated by fitting the raw data to 1:1 model.

2.2.8. SAXS data acquisition and processing.

All SAXS experiments presented in this work were acquired at in-house SAXSpace instrument with a sealed tube source (Anton Paar, Graz, Austria). Data collection was done with source X-rays in line collimation and scattering was captured on 1D CMOS Mythen detector. For each sample of $ClpB_M$ in the presence or absence of nucleotide and matched buffer, about 70µl of sample/buffer was exposed for 1 hour at 20°C in a thermostat quartz capillary of 1mm diameter. Nucleotides were mixed with a protein: nucleotide molar ratio of 1:10 and incubated for 5 minutes before initiating sample loading for data collection. Latter process took about additional 3 minutes to load the sample and

activate the vacuum in the sample chamber prior to data collection. Thus, about 8 minutes elapsed from addition of nucleotides to protein and initiation of SAXS data collection for all datasets. For SAXS experiments of $ClpB_M$ and inhibitor drug samples, data on $ClpB_M$ was collected with either 1% DMSO or individual drugs. In later experiments, $ClpB_M$ was incubated with inhibitors for 10 minutes, followed by addition of ATP and incubated for further 5 minutes, and subsequently data was collected. In experiments with inhibitor drugs, the protein was incubated with 200μM drug for 30 minutes and data was acquired. Data processing included correction for beam position using SAXStreat program, followed by subtraction of matched buffer and desmearing of data to represent scattering from true point collimation using SAXSquant program. Finally, datasets were stored in ASCII format of three columns, Q, I(Q) and dI(Q) were Q is momentum transfer vector defined as $Q = 4\pi\sin(\theta/\lambda)$ having units in nm^{-1}.

Table 2.2: SAXS data collection parameters used for this study.

Data Collection Parameters	
Instrument	SAXSpace (Anton Paar)
Beam geometry	10 mm slit
Wavelength (nm)	0.15418
Primary beam calibration	SAXStreat Program
Data collection software	SAXSdrive Program
Temperature	293 K
Sample mounting	Quartz capillary
Detector type	Mythen
q range (nm^{-1})	0.10–5.00
Exposure time	60 and 30 minutes
Analysis & Representation	
Desmearing Software	SAXSquant
Data processing	SAS Data Analysis (ATSAS 2.7.4)
***Ab initio* analysis**	DAMMIF/DAMMIN
Validation and averaging	DAMAVER

2.2.9. SAXS Data Analysis and Shape restoration.

Estimation of size/shape parameters and restoration of scattering shape of $ClpB_M$ in the presence or absence of either nucleotides or inhibitors was performed using ATSAS 2.7.4 suite of programs. Inbuilt plugins for estimation of distance distribution profile of interatomic vectors was used to compute radius of gyration (R_g) and maximum linear dimension (D_{max}) of the protein molecules in solution. For shape restoration, synchronized plugins were employed to generate ten uniform density models using dummy residues (by DAMMIF program), their inertial axes alignment based superimposition (by SUPCOMB program), averaging (DAMAVER program) and re-optimization of final model to best represent the experimental data (DAMMIN program). Taking cue from previous publications that $ClpB_M$ forms assembly with P6 or hexameric symmetry (Sweeny et al., 2015a) initially two types of modeling runs were done for Apo $ClpB_M$, +ATP and +ADP datasets: 1) with P1 symmetry or no symmetry, and 2) with P6 symmetry. Consistently, it was found that normalized spatial disposition (NSD) value which represents shape profile variation between the computed ten models was similar between P1 and P6 symmetry. This suggested that the NSD values are due to the inherent error or noise in the dataset, not due to symmetry imposition, but χ^2 value, a quantitative assessment, between theoretical SAXS profile of the dummy residue modeled structure and converted experimental SAXS data (Q range: 0.012-0.1 $Å^{-1}$) was closer to unity for P6 symmetry vs. P1 or P3 symmetry. Based on this observation, we opted to consider interpretations of models done under P6 symmetry for all datasets. All graphs and images pertaining to SAXS datasets in this work were made using Origin5 and UCSF Chimera softwares.

2.2.10. Malachite green assay to measure $ClpB_M$ ATP hydrolysis activity.

For activity assays, 0.5µM of $ClpB_M$ was incubated with or without 5µM κ-casein (Sigma) in an assay buffer (50mM Tris-Cl, pH 7.4, 100mM NaCl, 10mM $MgCl_2$, 5mM DTT and 5 % Glycerol) at 37°C for 10 minutes. ATP (1mM) was added to solution containing $ClpB_M$, and reaction was incubated at 37°C for 30 minutes.

The aliquots were collected at regular intervals of time and 2N HCl was added to quench ATPase activity. The release of inorganic phosphate by ATP hydrolysis activity of $ClpB_M$ was monitored using malachite green assay (Sigma) by measuring absorbance at 620nm using M200PRO reader (Tecan). The amount of inorganic phosphate released in enzymatic reaction was calculated using a standard plot obtained from different known concentrations of phosphate. The enzymatic reactions had the desired controls such as buffer only, no enzyme and no substrate control.

2.2.11. Small molecule inhibitor libraries.

Two libraries were screened against $ClpB_M$ ATP hydrolysis activity. Small molecule library comprising of a subset of 160 active compounds that belonged to National Cancer Institute – Developmental Therapeutic Program (NCI-DTP).Spectrum collection containing 2320 FDA approved drugs was purchased from Micro source drug discovery system. The primary hits obtained were reordered from National Cancer Institute and Sigma for further assays.

2.2.12. High-throughput screen with Malachite green dye.

A high-throughput screening (HTS) was performed using small molecule libraries mentioned above. Purified $ClpB_M$, 0.25µM in buffer containing 50mM Tris (pH 7.4), 100mM NaCl, 10mM $MgCl_2$ and 5mM DTT was pre-incubated with small molecules (100µM) at 37ºC for 10 minutes. Reaction was initiated by adding 0.5mM ATP along with 5µM κ-casein and kept at 37ºC for 20 minutes. The release of inorganic phosphate was monitored in the presence and absence of drugs using malachite green assay as described above. The absorbance in the presence of drugs was normalized with that obtained from reaction containing DMSO without drugs as control, and the normalized absorbance was used to measure release of inorganic phosphate by ATPase activity of $ClpB_M$. Small molecules showing inhibition of ATP hydrolysis equal to or more than 70% were considered as potential inhibitors. In order to determine enzyme specificity, inhibition assays against *E. coli* ClpB and yeast Hsp104 with potential hits were also performed as described above.

2.2.13. Microscale thermophoresis studies.

Microscale thermophoresis (MST) experiments were performed by covalent labeling of 20 nM $ClpB_M$ with fluorescent label dye, NT-650 via NHS coupling. Labeled $ClpB_M$ was kept constant, while the concentration of ligands was varied between 50μM – 0.61nM. The assay was performed in buffer A supplemented with 0.05% Tween-20. After a short incubation the samples were loaded into Monolith™ NT.115 standard treated capillaries and data was recorded. For analysis, change in MST signal was expressed as the change in the normalized fluorescence (ΔF_{norm}), which is defined as F1/F0 where F1 is the fluorescence after thermodiffusion and F0 is initial fluorescence. The ΔF_{norm} obtained at each concentration of the ligand was plotted against the ligand concentration to yield a dose-response curve. Data were fitted to obtain binding constants.

2.2.14. K_m and Vmax plots.

The ATP hydrolysis activity of $ClpB_M$ (0.25μM) was monitored as a function of different ATP concentrations (50μM – 1000μM) using Malachite green based assay described above. The $ClpB_M$ in the presence and absence of hexachlorophene was incubated with desired concentration of ATP and κ-casein (5μM), and reaction was incubated at 37°C. The 30μl of aliquots were taken from reaction at every 3 minutes intervals for 15 minutes, and mixed with 30μl of malachite green dye at room temperature for 5 min. The ATP hydrolysis was monitored by recording increase in absorbance at 620nm. The initial velocity of reaction was obtained by a linear fit of the curve obtained as a function of time. The reaction velocities obtained over different concentrations of inhibitor and ATP, was further plotted as a function of ATP concentrations, and the curve was fit to rectangular hyperbola with non-linear regression analysis.

$$Y=a*x/(b+x)$$

Where Y is initial velocity, x is [ATP], a and b provides value for V_{max} and K_m respectively.

2.2.15. Luciferase refolding assay.

Luciferase refolding assay was performed as described before (Barnett et al., 2000) with few modifications. Briefly, firefly Luciferase (226 µM, Promega) was diluted 100 folds in denaturation buffer (25 mM Hepes (pH 7.6), 50 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM DTT and 7 M urea) containing 1 mM ATP for 30 minutes at 25ºC. Refolding was initiated by further dilution of denatured luciferase by 100 fold in refolding buffer (25 mM Hepes (pH 7.6), 50 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM DTT) containing 4.0µM ClpB, 1.0 µM DnaK, 0.15 µM DnaJ and 0.05 µM GrpE in the presence of 2.0 mM ATP at 25ºC for various time intervals. The 1µl of refolding solution was incubated with 60µl of luciferase assay reagent, and luminiscence was recoredusing multimode plate reader (Tecan). To determine the effect of identified small molecules on ClpB refolding activity, 100 µM of the compound or DMSO alone (as control) was added in refolding buffer. The luminiscence intensity obtained for refolded luciferase were normalized with thatof denatured luciferase.

2.2.16. Antimycobacterial assays.

THP-1 cells (human monocytic cell line) were cultured in RPMI medium supplemented with 10% heat inactivated-foetal bovine serum (HI-FBS). Before experiments, monocytes were differentiated into macrophages using 20 ng/ml phorbol 12-myristate 13-acetate (PMA). Intracellular killing experiment were performed as before (Arora et al., 2019). Briefly, THP-1 macrophages (0.2 million/well/ml) were infected with single-cell *M. tuberculosis* suspensions of 1:10 multiplicity of infection (MOI). Following 4 h infection, extracellular bacteria were removed by overlaying macrophages with RPMI containing 200 µg/ml of amikacin. The infected macrophages were washed twice with1xPBS twice and infected macrophages were overlaid with RPMI containing various drugs at a non-cytotoxic concentration. DMSO and INH (10 µM) were taken as negative and positive controls. Samples were collected by lysing macrophages with 1x PBS containing 0.1% Triton X-100. At designated time points, bacterial enumeration was by plating 10-fold serial dilution on middlebrook 7H11 plates at 37 ˚C for 3-4 weeks.

2.2.17. Stopped flow experiments.

Nucleotide exchange activity were performed as before (Packschies et al., 1997) with few modifications. 0.1µM MABA-ADP were mixed with 1 µM of $DnaK_M$ for stable complex formation at 4 ºC for 1 hour in Buffer B. Nucleotide exchange was performed by addition of $GrpE_M$ (2µM) containing 200 µM of ADP in buffer B. MABA-ADP-$DnaK_M$ complex were excited at 360 nm and emission were monitored at 435nm as a function of time. Readings were normalized to the maximum value of fluorescence of respective reaction and plotted as a function of time.

2.3 Methods for HSPMdb

2.3.1 Data collection

All articles available in PubMed were searched to collect and compile comprehensive information on Hsp modulators. To obtain research articles having information on Hsp modulators, systematic searches were performed using various keywords such as ‘heat shock protein modulators’, ‘heat shock protein inhibitors’ and ‘heat shock protein activators’. In addition, the name of individual chaperones such as Hsp100 modulators, Hsp70 modulators and ClpB modulators, was used as a query to search Pubmed articles. These searches ended in a total of 7005 research articles. Articles describing prediction methods, review articles and book chapters were excluded, and the rest of the complete research articles were manually screened by cautious reading for the relevant information of Hsp modulators. Only research papers providing information about experimentally validated Hsp modulators and their analogues were selected for further data curation. Thus, finally 176 research articles were shortlisted and data on Hsp modulators were manually curated. For modulators that have been examined in more than one study or tested against different Hsp types

2.3.2 Database architecture and web interface

HSPMdb is built using Linux–Apache–MySQL–PHP (LAMP), a package built on the Linux operating system. LAMP integrates Apache for the web server, and MySQL is a relational database management system. The PHP is scripting

language to bring the data fetched by MySQL on to the web pages for display. Additionally, the JAVA script was used to provide dynamic functionalities on web pages. Python scripts are implemented at the back-end to process data fetched from user queries. Data content HSPMdb provides comprehensive information on biological and chemical properties of small synthetic Hsp modulators. Biological information of these modulators has been compiled under two major categories: (i) enzymatic activity and (ii) cellular activity. Under the enzymatic activity field, detailed information of the type of enzymatic assay used, site of interaction of modulator, the effect of the modulator and *in vitro* enzymatic modulation activities (IC50, EC50, DC50, EC50, *Ki*, *K*d and percentage inhibition) have been provided. In the cellular activity field, complete information of the type of cell viability assays used, tested cell line and cellular activities (percentage cell growth inhibition, EC50 and GI50) has been compiled. Comprehensive information of targeted Hsps like their name, origin and localization has also been compiled. Additionally for each compound, the database provides information on 2D/3D structure, and its physical, elemental and topological properties. International Union of Pure and Applied Chemistry (IUPAC) names and Simplified Molecular Input Line Entry System (SMILES) of each modulator were extracted from literatures and further generated using OPSIN (Lowe et al., 2011). The physicochemical properties of all molecules were obtained using the PaDEL software (Yap, 2011) which calculates 2D/3D chemical descriptors from the SMILES of the compounds. The chemical structures of molecules are displayed in the database web pages using online 'SMILES to image' tool of RxnFinder (Hu et al., 2011).

2.3.3 Implementation of tools

A user-friendly web interface has been developed with various tools for convenience of data searching, browsing and analysis. The description of these tools is given below.

2.3.4 Search

Two searching options, 'Simple Search' and 'Advanced Search', have been designed for data searching. Simple search allows users to search for

modulators in HSPMdb using their desired keywords related to different fields such as the name of disease or compound or name of Hsp or PMID. A default six fields have been selected for display of the result of the query. Also, users can select various additional fields of their choice for display of results of their query. Advance Search allows users to search HSPMdb with complex queries e.g. more than one type of query at a time by selecting different conditions (e.g. AND &OR) between queries.

2.3.5 Browsing

To fetch information from HSPMdb, robust browsing pages have been developed. Users can browse on different fields such as enzymatic activity, cellular activity, Disease, Enzymatic and Cellular assays and Organism. In the case of the Enzymatic and Cellular activity field, users can fetch more information on Hsp modulators about its other properties such as IC50, EC50, DC50, EC50, *Ki*, *K*d and percentage inhibition. HSPMdb provides two options for the users. (i) All HSPs: from this browsing page, the user could get information about modulators against all different classes of Hsps for a desired property such as disease-specific or organism-specific. (ii) Individual Hsp: this browsing page will allow the user to fetch information about modulators on a particular Hsp such as Hsp70 or Hsp90.

2.3.6 Draw compound

The Draw compound is one of the very important tools which allow users to identify Hsp modulators having similarity to their query molecule based on the similarity index. The tool makes use of the JSME editor developed by Bruno Bienfait and Peter Ertl (Bienfait and Ertl, 2013). Users need either structure or SMILE of the query molecule to identify modulator(s) having a similar structure(s) in the database. At the back end, the tool compares the user-given SMILES with SMILES of all molecules in the HSPMdb database based on a method which performs fragmentation of SMILES strings into overlapping substrings of a defined size (four in case of this tool) called as LINGOs (Vidal et al., 2005). The similarity is calculated as the Tanimoto coefficient using the number of matching and non-matching LINGOs. The value of the Tanimoto

coefficient lies between 0 and 1 with value closer to 1 indicating higher similarity and the value closer to 0 as lower similarity. There are two ways to search for molecules having a structure similar to the query molecule: (i) users can draw the structure of query compound using tools available in the database and get the SMILE, or (ii) users can directly paste the SMILE of the query molecules. By clicking on the 'Compare with HSPMdb' button, users will get a list of Hsp modulators showing similarity with the query molecule along with the similarity index. Users can sort the entries by clicking (single and double) on similarity index. Thus, this tool will be very useful to scientific community for repurposing of existing drugs.

Chapter 3

Characterization and identification of small molecule inhibitors against ClpB protein from Mycobacterium tuberculosis.

3.1. Introduction

Despite several decades of extensive research, the biology of *Mycobacterium tuberculosis* (*M. tuberculosis*) remains poorly understood, hindering the development of effective diagnosis and therapeutics. Tuberculosis still poses a major health challenge and is the leading cause of death among infectious diseases (Lin and Flynn, 2010). Post-infection, *M. tuberculosis* could persist in its host for decades without developing any clinical symptoms. This latent or metabolically less active bacteria is reactivated as the host immune response gets compromised (Muñoz et al., 2015). In its host, *M. tuberculosis* encounters numerous stresses such as reactive oxygen species, reactive nitrogen species, low oxygen and nutritional stress, which are known to adversely affect stability and abundance of a number of cellular proteins, resulting in an imbalance in protein homeostasis (Kumar et al., 2011b; Lupoli et al., 2018b).To counter such imbalance, *M. tuberculosis* harbors an efficient protein quality control pathway consisting of DnaK and its co-chaperones to prevent aggregation of partially unfolded proteins, and a disaggregase ClpB that function to disassemble preformed protein aggregates. Therefore, better understanding of the *M. tuberculosis* chaperone machinery would provide insight into mechanism underlying its stress tolerance and aid ongoing efforts on design of novel therapeutic strategies.

Chaperones such as Hsp70 and Hsp100 family of proteins plays an essential role in the maintenance of protein homeostasis and thus, are critical for optimal cellular growth under stress conditions (Calloni et al., 2012; Saibil, 2013b; Vaubourgeix et al., 2015). Hsp70 primarily interacts with the exposed hydrophobic regions of unfolded proteins and prevent their aggregation. However, once these aggregates are formed, the activity oftheHsp100 family of proteins such as bacterial ClpB is required to resolubilize and reactivate aggregated proteins. The Hsp70 influences Hsp100 functions by stimulating its ATPase activity as well as recruiting it to larger aggregates (Glover and Lindquist, 1998; Goloubinoff et al., 1999b). Hsp70 is also known to aid the refolding of unfolded molecules rescued by Hsp100 from the preformed aggregates (Goloubinoff et al., 1999b). In addition to their role in protein folding,

these chaperones are also essential for the regulation of cell cycle, protein trafficking and degradation (Gupta et al., 2018; Reeg et al., 2016; Rohde et al., 2005; Truman et al., 2012).

Hsp100 family of chaperones is highly conserved among bacteria, yeast and plants however, no functional homologs have been reported in animals and humans. The Hsp100 family such as bacterial ClpB belongs to AAA^+ (ATPases associated with diverse cellular activities) superfamily of ATPase that uses energy driven by ATP hydrolysis to thread out protein molecules from aggregates (Yu et al., 2018b; Zolkiewski et al., 2012). The disaggregase activity of ClpB is essential for some pathogenic organisms such as *L. interrogans* to establish infection in the host (Krajewska et al., 2017; Vaubourgeix et al., 2015). In other organisms such as *E. coli*, ClpB is not essential during optimal conditions, however is required for survival of organism under stresses such as high temperature (Lourdault et al., 2011; Thomas and Baneyx, 1998). Structurally, *E. coli* ClpB forms nucleotide dependent hexamer where each protomer consists of an amino terminus domain followed by two ATP binding domains (NBD1 and NBD2) linked through a middle domain (Rizo et al., 2019). The amino terminus domain of ClpB is required for substrate interaction and the middle domain mediates its interaction with DnaK (Rosenzweig et al., 2015; Seyffer et al., 2012; Tanaka et al., 2004).Though *E. coli* ClpB has been extensively studied, the understanding about its homolog from *M. tuberculosis* is beginning to emerge.

As chaperones machinery play critical role in survival of various pathogens, various studies have focused towards identification of chaperone inhibitors for therapeutics (Cockburn et al., 2011; Davenport et al., 2014; Johnson et al., 2014; Singh et al., 2020).A similar approach targeting *E. coli* ClpB has shown promise in inhibition of the gram negative bacteria (Martin et al., 2013). *M. tuberculosis* ClpB ($ClpB_M$) is required to establish infection in host tissues (Vaubourgeix et al., 2015). Also, $ClpB_M$ is essential for the recovery of *M. tuberculosis* from stresses that promote protein aggregation (Vaubourgeix et al., 2015). The disaggregase $ClpB_M$ is essential for not only active *M. tuberculosis* but also for its persistence in latent state (Tripathi et al., 2020a). Considering the essential role of $ClpB_M$ in

pathogenesis of *M. tuberculosis*, strategies to inhibit $ClpB_M$ activity could lead to the development of novel therapeutics against the pathogen. Further such inhibitors would aid to our understanding of the mechanism of $ClpB_M$ action.

In order to expand our understanding of *M. tuberculosis* ClpB machinery here, we have performed a detailed biophysical characterization including understanding of its conformational changes through Small Angle X-ray Scattering (SAXS). We show that thermal unfolding of $ClpB_M$, at pH 7.4 is irreversible, we observed that $ClpB_M$ forms concentration, nucleotide and salt-dependent oligomers. Finally, high throughput screening led to the identification of novel small molecules that inhibited $ClpB_M$ ATPase activity, and growth of intracellular mycobacteria. One of the drug inhibited $ClpB_M$ activity in a competitive manner. Using SAXS data analysis and modeling, we further show that the identified drugs inhibit ATP driven structural rearrangements within $ClpB_M$. The study thus provides insight into biochemical aspects of the chaperone, and paves way for further design of inhibitors against *M. tuberculosis* chaperone machinery.

3.2. Results:

3.2.1 $ClpB_M$ thermal denaturation shows the presence of stable intermediates in its unfolding pathway.

To examine the folded nature of the purified chaperone, the secondary structure of $ClpB_M$ was studied under native conditions using Far-ultraviolet (UV) circular dichroism (CD). The Far-UV CD provides insight into the secondary structural content of proteins such as α-helix, β-sheet or random coil. $ClpB_M$ was purified to homogeneity and CD spectra were recorded in Far-UV range from 190nm to 250nm at pH 7.4 at varying temperature from 25°C - 95°C. Under native condition at 25°C, the $ClpB_M$ showed maximum negative ellipticity at 208nm and 222nm suggesting that the protein is predominantly α-helical which is in agreement with previous studies (Fig. 3.1A). CD studies suggest that at secondary structural level, the purified $ClpB_M$ is well folded and mostly similar to its respective counterparts in *E. coli* (Barnett et al., 2000).

We next determined the effect of increase in temperature on secondary structural content of $ClpB_M$ (Fig. 3.1B and 3.1C). The ellipticity of the protein at 222nm was monitored as the incubation temperature was raised from 25°C to 95°C. As shown in Figure 3.1B, the incubation of $ClpB_M$ with increasing temperature led to progressive decrease in ellipticity at 222nm indicating a gradual loss of its native structure. The thermal CD melt of $ClpB_M$ showed two transitions, the first transition was observed at apparent melting temperature (T_m) of about 52°C. There was no significant change in CD signal from 60-70°C with another gradual less cooperative transition at about 80°C. The first thermal transition showed a loss of about ~40 % of total ellipticity, whereas complete loss of residual structure was observed after a second high temperature thermal transition. Thus, $ClpB_M$ showed three state unfolding with an intermediate, which is stable upto 60°C. To examine the reversibility of $ClpB_M$ unfolding, the change in ellipticity at 222nm was further monitored as the temperature was lowered from 95°C to 25°C. As shown in Figure 3.1B, there was no decrease in ellipticity upon lowering the temperature suggesting that the unfolded protein could not regain its native structure on its own and thus $ClpB_M$ thermal unfolding is largely irreversible.

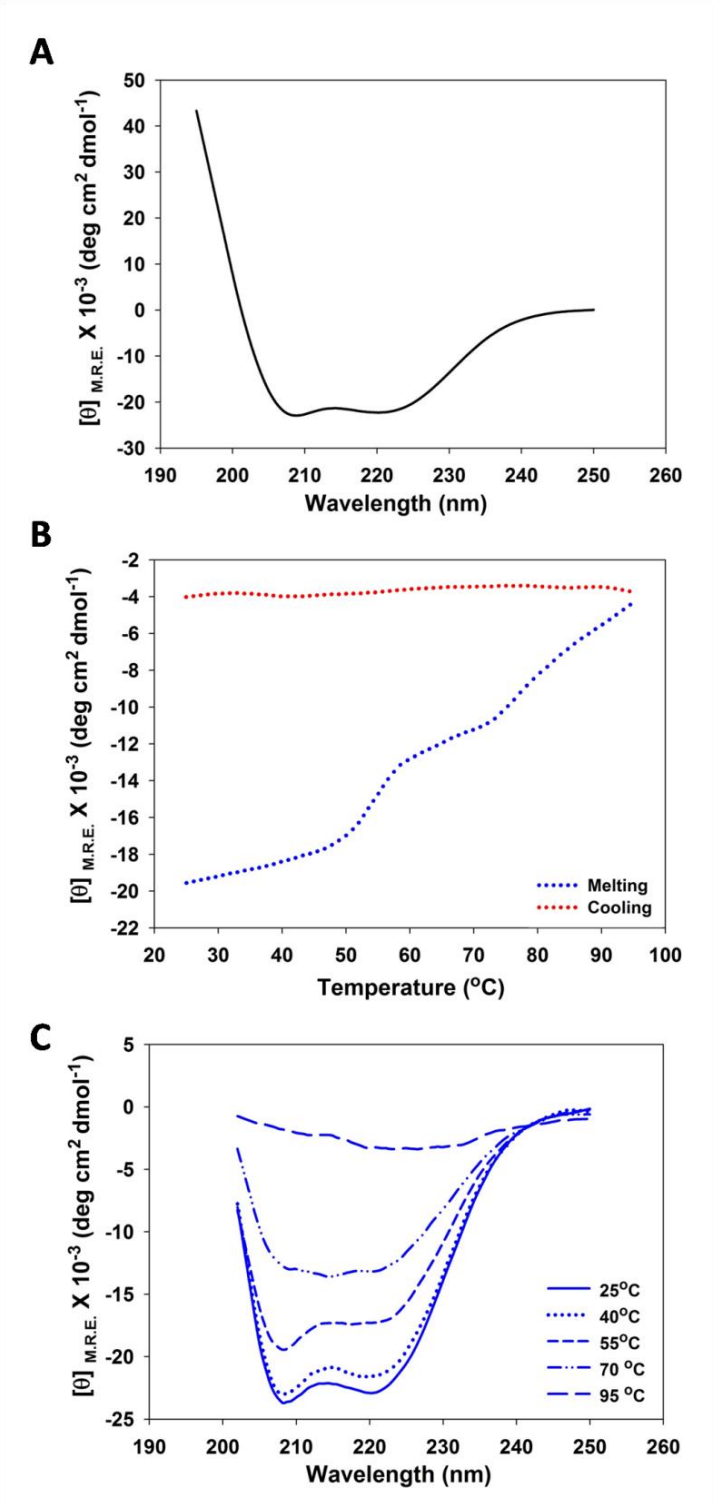

Figure 3.1. Secondary structural analysis of $ClpB_M$. (A) show Far-UV CD spectra of $ClpB_M$ recorded with 1mm path length cuvette at 25°C. The spectra shown here are the average of three independent scans.**(B)** The thermal stability of purified $ClpB_M$ was measured by monitoring changes in mean residue ellipticity of proteins at 222nm as a function of increase in temperature from 25°C-95°C at scan rate of 1°C/min. **(C)** The Far-UV CD spectra of chaperone $ClpB_M$, at various temperatures.

3.2.2 $ClpB_M$ oligomeric state is dependent upon its concentration, and solvent conditions.

Hsp100 class of proteins including ClpB utilizes energy from ATP hydrolysis for their disaggregation activity. The ATP or ADP binding also affects conformation as well as oligomeric state of Hsp100 proteins in solution (Akoev et al., 2004; del Castillo et al., 2011; Lin and Lucius, 2016; Zolkiewski et al., 1999).In order to examine the effect of nucleotides on oligomerization of $ClpB_M$, we performed analytical sedimentation ultracentrifugation (AUC) analysis as a function of its concentration, nucleotides and salts (Figure 3.2).

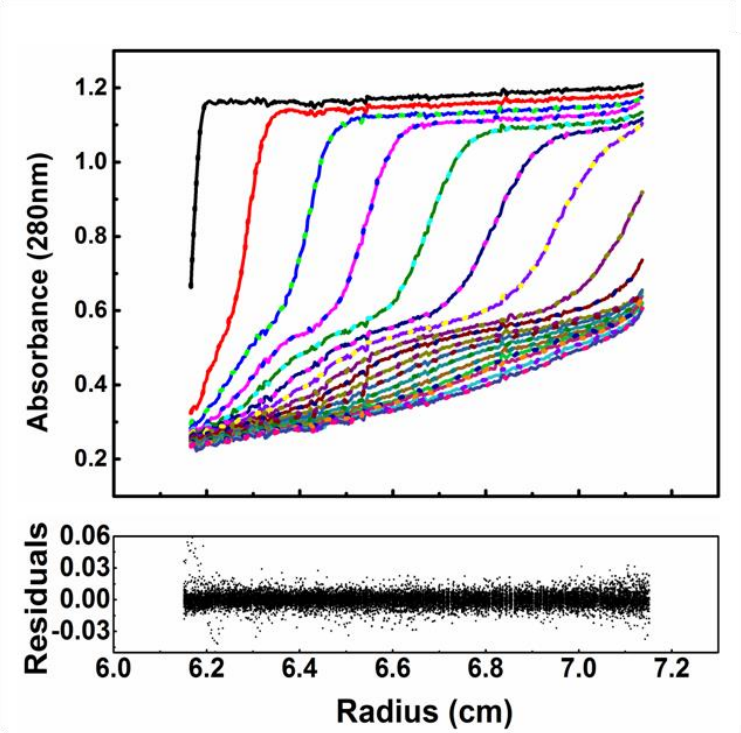

Figure 3.2 Absorbance scans show sedimentation of indicated chaperones as a function of radial distance over a period of time. Sedimentation velocity data for indicated proteins at 40,000 rpm and 20°C. Shown are radial absorption scans for $ClpB_M$.

We first examined the effect of changes in $ClpB_M$ protein concentration on its intermolecular association. The $ClpB_M$ concentration was increased from 0.75mg/ml to 2.25mg/ml and AUC absorption scans were monitored. As shown in Figure 3.3A, at concentration of 0.75mg/ml, $ClpB_M$ molecules appeared to

partition as two species with standard sedimentation coefficient $s_{20,w}$ of 5.30S and 11.36S. The species sedimenting at 5.30S (~53%) corresponds to the monomeric $ClpB_M$, and the higher sedimenting species (~38%) indicates the presence of self-associated higher order $ClpB_M$ oligomers. At a higher concentration of 1.5mg/ml, the first peak of lower sedimentation coefficient appeared at roughly similar position ($s_{20,w}$= 5.59S), however the second peak shifted from 11.36S to 13.37S. This shift in second peak to higher sedimentation coefficient suggests concentration dependent self-association of $ClpB_M$. This is in agreement with further shift of the second peak from $s_{20,w}$ = 13.37S to $s_{20,w}$ = 13.95S as $ClpB_M$ concentration was further increased to 2.25mg/ml (Table 1). The peak corresponding to 13.95S is consistent with weighted-average of oligomers (~53% of the overall population)with 4.5 units of $ClpB_M$. Taken together, the data suggest that in solution, Apo $ClpB_M$ exists as heterogeneous species, and an increase in protein concentration results is the self-association of $ClpB_M$ into higher order oligomers.

Next, we investigated the effect of different nucleotides (ADP or ATP) on oligomerization of $ClpB_M$ (1.5mg/ml). The pre-incubation of $ClpB_M$ with ADP led to an increase in standard sedimentation coefficient of higher order oligomeric form of $ClpB_M$ from 13.37S (for Apo form) to 13.74S (Fig. 3.3B and Table 3.1). As compared to its Apo form, the pre-incubation of ATP with $ClpB_M$ further increased the intensity as well as the standard sedimentation coefficient of the second peak to 14.68S. The observed increase in the intensity of the second peak suggests that in the presence of ATP, $ClpB_M$ majorly exists at a relatively higher oligomeric state than that observed in Apo or ADP state.

The presence of salt ions is also known to affect self-association in proteins. We next performed AUC study to examine the effect of increasing concentration of NaCl on $ClpB_M$ oligomerization (Figure 3.3C and Table 3.1). At 100mM NaCl, $ClpB_M$ sedimented as a broad peak at11.81S with another fraction of low molecular weight species sedimenting at 4.85S. The broadness of peak corresponding to sedimentation coefficient of 11.81S indicates the presence of oligomeric species of different association states in equilibrium with each other. The c(s) analysis of $ClpB_M$ scans in the presence of 200mM or 300mM NaCl

showed the presence of monomeric $ClpB_M$ as the predominant sedimenting species at 4.62S and 4.74S respectively. Also, at 200mMNaCl concentration, two small sedimenting species at 6.56S and 8.85S were observed. At 300mM NaCl, in addition to the monomeric species, a smaller peak with sedimentation coefficient $s_{20,w}$ of 7.60S was also observed, which could be due to the presence of monomeric and a higher order oligomeric $ClpB_M$ in equilibrium with each other. Overall, the data suggest that $ClpB_M$ exists in multiple oligomeric states in solution, which vary with salts and type of nucleotides.

Table 3.1. Sedimentation coefficient values of $ClpB_M$ obtained from analytical ultracentrifugation studies.

$ClpB_M$	$S_{20,w}$			Average Mol Wt (KDa) [Pred. units]			% Species		
	Peak 1	Peak 2	Peak 3	Peak 1	Peak 2	Peak 3	Peak 1	Peak 2	Peak 3
Concentration (mg/ml)									
0.75	5.30	11.36	-	99 [1.0]	305 [3.3]	-	51.29	38.01	-
1.5	5.59	13.37	-	103 [1.1]	381 [4.1]	-	42.28	20.73	-
2.25	5.05	13.95	-	94 [1.0]	430 [4.7]	-	28.64	52.49	-
Nucleotide (mM)									
APO	5.59	13.37	-	103 [1.1]	381 [4.1]	-	51.29	38.01	-
ADP	4.93	13.74	-	91[1.0]	424 [4.6]	-	25.87	59.39	-
ATP	5.60	14.68	-	102 [1.1]	433 [4.7]	-	25.38	64.40	-
NaCl (mM)									
100	4.85	-	11.81	93 [1.0]	-	341 [3.7]	23.53	-	70.71
200	4.62	6.56	8.85	91 [1.0]	152 [1.7]	291 [3.1]	23.63	22.73	25.21
300	4.74	7.60	-	89 [1.0]	120 [1.3]	-	46.86	36.02	-

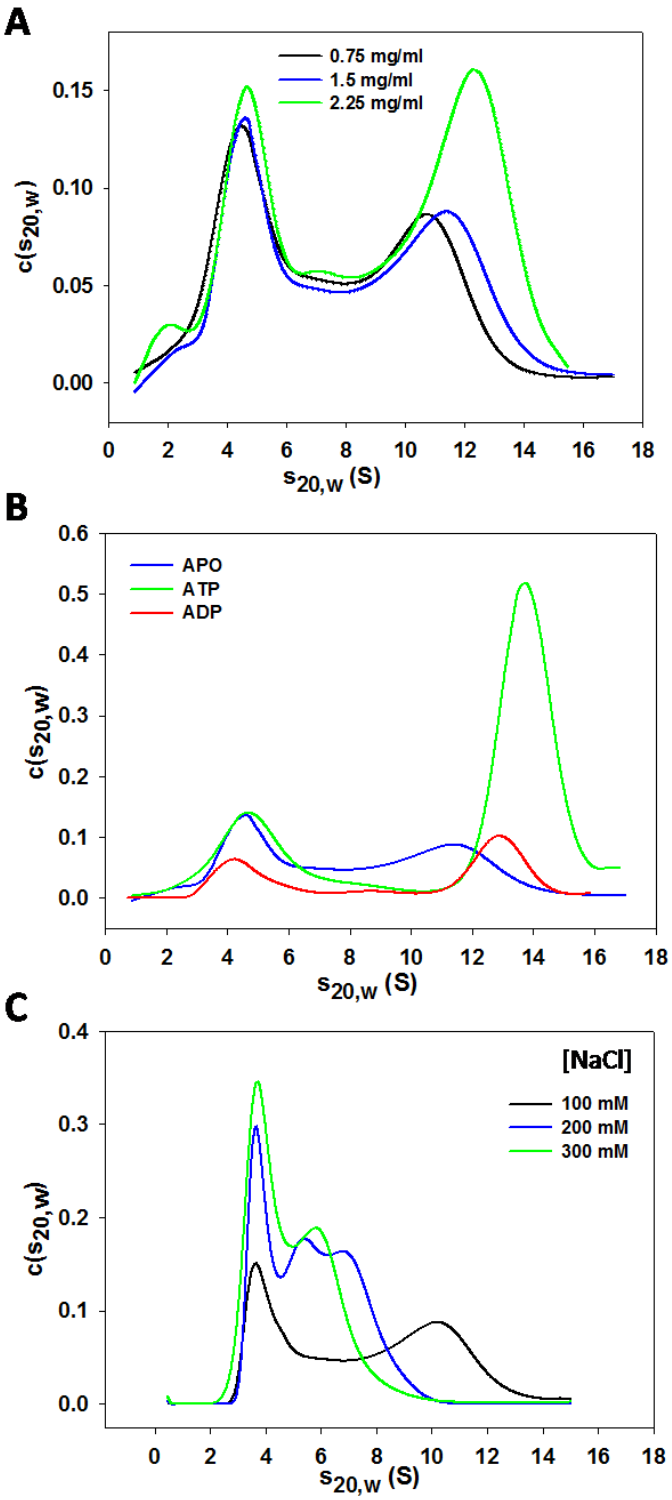

Figure 3.3 Sedimentation velocity studies of $ClpB_M$. Shown is sedimentation coefficient continuous distribution c(s) plots for $ClpB_M$, obtained from its sedimentation velocity data. **(A)** $ClpB_M$ sedimentation shows two major species corresponding to monomer or higher order oligomers. **(B)** $c(s_{20,w})$ distribution of $ClpB_M$ in the absence (blue) and presence of ATP (green) or ADP (red). **(C)** $ClpB_M$ was incubated with indicating concentration of NaCl for sedimentation studies.

3.2.3 Nucleotides affect oligomerization state of $ClpB_M$.

The sedimentation velocity studies suggest that self-assembly of $ClpB_M$ depends upon its concentration and the presence of nucleotides in solution. We further examined the association state of $ClpB_M$ using size exclusion chromatography. As shown in Figure 3.4A, in the absence of any nucleotides, apo $ClpB_M$ (2mg/ml) eluted at an approximate molecular weight of 175 kDa (based upon the column calibration with globular proteins of known molecular weight). The pre-incubation of ADP or ATP with the chaperone resulted in its elution at volume expected from relatively higher molecular weight species corresponding to an oligomer of trimer or tetrameric $ClpB_M$, respectively (Fig. 3.4A). The increase of protein concentration from 2 mg/ml to 6 mg/ml resulted in further decrease of elution volume and an asymmetrical elution peak with a sharp leading edge and relatively longer trailing edge which is suggestive of presence of multiple oligomers in equilibrium with each other. The weighted-average size of the protein at 6mg/ml corresponds to about pentameric and hexameric $ClpB_M$ in the absence and presence of ATP (Fig. 3.4B) respectively. Overall, these observations suggest that nucleotides induce significant alteration in solution shape of $ClpB_M$or initiate lower-order oligomerization, which is also in agreement with previously reported studies (Tripathi et al., 2018; Zolkiewski et al., 1999). Figure 3.4C show the linear standard graph for calculating molecular weight.

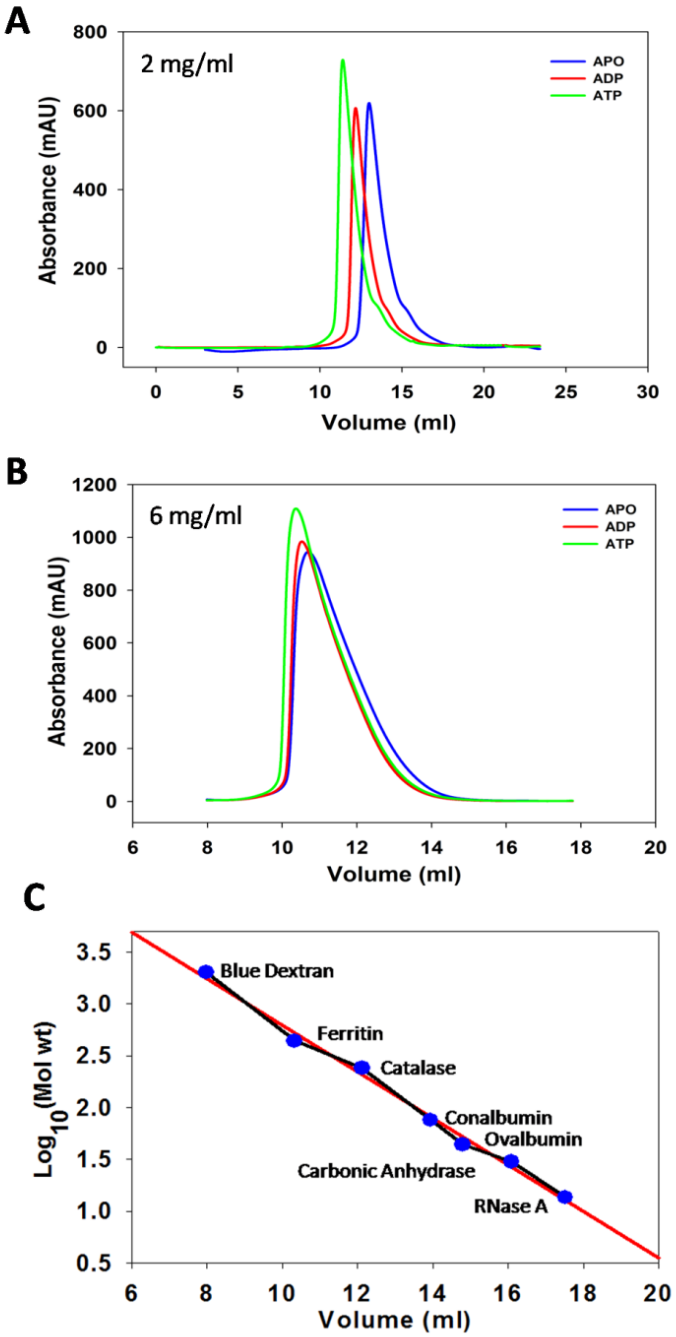

Figure 3.4 Nucleotide dependent changes in oligomerization of $ClpB_M$. (A) $ClpB_M$ elution profiles show shift in its elution volume towards higher oligomeric mass in the presence of nucleotides at a concentration of 2.0 mg/ml. **(B)** $ClpB_M$ elution profiles at 6.0 mg/ml. APO (blue), ADP (red) and ATP (green). **(C)** Calibration curve for Superdex 200 gel-filtration column. Proteins of known molecular weight, RNase (13.7 kDa), Carbonic anhydrase (29kDa), Ovalbumin (44 kDa), Conalbumin (75 kDa), Catalase (232 kDa) and Ferritin (440 kDa) in PBS were loaded onto the column. Blue dextran (2000 kDa) was also used as one of the standard marker. The elution volume of the proteins was plotted against Log of molecular weight to obtain a standard calibration curve. The curve was then used for estimating the molecular weight of chaperone $ClpB_M$.

3.2.4 ATP, and not its non-hydrolyzable analogs induce large scale conformational rearrangements in $ClpB_M$.

The above AUC and size exclusion studies suggest that $ClpB_M$ is more prone to oligomerization upon incubation with nucleotides which are its native ligands required for activity of the protein. We thus further performed structural characterization of $ClpB_M$ using SAXS data analysis in the presence and absence of nucleotides. SAXS studies have been extensively used to gain valuable information about the oligomeric states as well as quaternary structures of the proteins (Borges et al., 2016; Brosey and Tainer, 2019; Korasick and Tanner, 2018) SAXS Intensity, I(Q) profiles of $ClpB_M$ protein (6.1mg/ml) in the presence or absence of nucleotides were determined. Double logarithmic plots ($Log_{10}I(Q)$ vs. $Log_{10}Q$) show the expected first order exponential decay (Figure 3.5A), and none of the datasets reflected large order aggregation or interparticulate effect at the concentration of protein studied. The linear zones of Guinier analyses of the datasets presuming a globular scattering profile are shown in Figure2.4B. We observed the linear fit of the Apo and ATP-$ClpB_M$ is much shorter and steeper compared to other nucleotide bound state of $ClpB_M$ (Figure 3.5B). This suggested that the radius of gyration (R_g) value of the protein shape is higher for Apo and ATP state compared to other nucleotides. Indirect Fourier transformation of the datasets provided probability distribution of the interatomic vectors present in the SAXS data of the protein sample i.e. P(r) as a function of r (Fig.3.5C). This analysis in correlation with Guinier analysis showed that the D_{max} of the protein molecules was about 22nm and 29nm for $ClpB_M$ in Apo and ATP state, respectively. In comparison, the maximum dimension (D_{max}) values were 15-17nm when $ClpB_M$ was bound to ADP, ATPγ-S or AMP-PNP. Also, the P(r) profiles indicated the shape of $ClpB_M$ molecules to be extended in Apo state, bilobal in ATP state and single domain in presence of slow/non-hydrolyzable nucleotides ADP, ATPγ-S and AMP-PNP. Additionally, the computed R_g values from P(r) analysis were 8.14±0.07 nm, 8.29±0.12nm, 5.69±0.04 nm, 5.56±0.02nm and 5.78±0.03 nm for Apo, ATP, ADP, ATPγ-S and AMP-PNP state, respectively. In summary, SAXS data analysis indicated that there appears a large scale shape/size rearrangement, which depends on its nucleotide bound state.

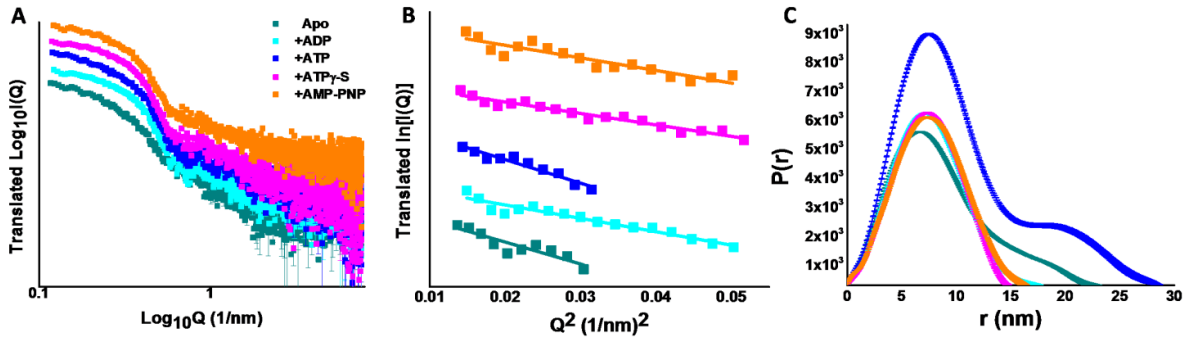

Figure 3.5 SAXS profiles of $ClpB_M$ in the presence of different nucleotides. (A) SAXS datasets of $ClpB_M$ protein in apo and nucleotide bound states are presented in double Log_{10} mode. Individual datasets and colour code are mentioned in the panels. Y-axis have been translated for clarity. **(B)** The Guinier analysis regions of the datasets have been shown with linear fit. **(C)** The estimated P(r) plot represents the maximum dimensions of $ClpB_M$ datasets. Colour of the plots have been retained with the same used for SAXS data representation and further shape restoration analysis.

3.2.5 Shape restoration of $ClpB_M$ in the presence and absence of nucleotides.

As mentioned in methods, shape restoration of $ClpB_M$ in the presence and absence of nucleotides using respective SAXS data profile and dummy residues were performed considering P6 symmetry. They have been presented in molecular map format and orthogonal views are presented in Figure 3.6 to collectively perceive the results on shape profile of $ClpB_M$ in the presence of different nucleotides. The Apo state showed an extended shape that lacked and adopted a clear and predominant P6 symmetry in the ATP state. Though shape profiles are dramatically different under these two states, the D_{max} and R_g values were similar since the length dimension of Apo state is similar to the cross-sectional length of the hexameric association. Furthermore, shape analysis seen in Figure 3.6 showed that the ATP state of $ClpB_M$ is hollow from center which is in agreement with similar observation reported before for its homolog (Lee et al., 2007; Sweeny et al., 2015a). In sharp contrast, the shape of $ClpB_M$ molecules bound to ADP, ATPγ-S and AMP-PNP showed a smaller compact dimension with P6 symmetry. These results indicated that ATP induces hexa-symmetric association and/or dramatic reorganization in the protein molecules, and similar ability is absent in other slow or non-hydrolyzable nucleotides.

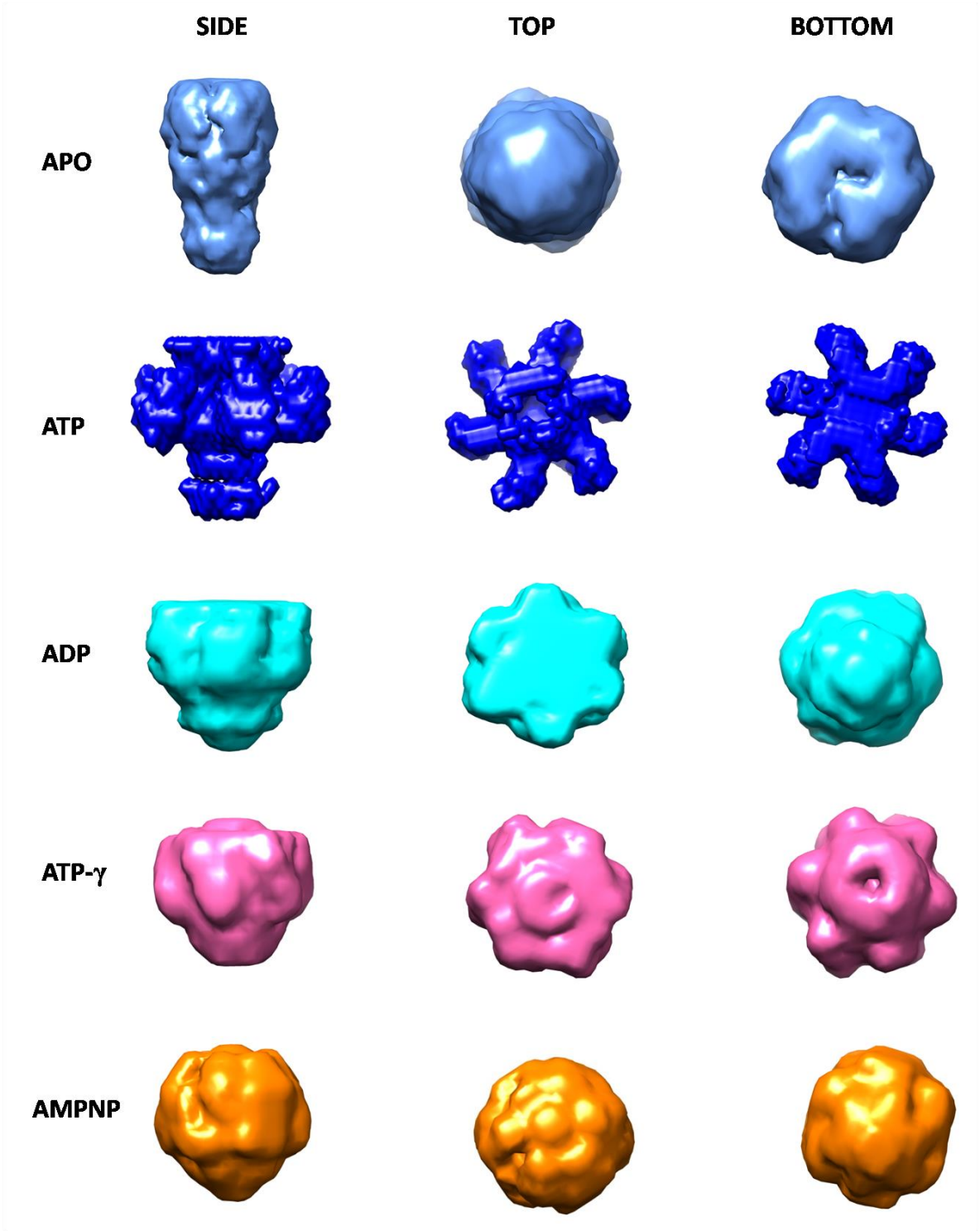

Figure 3.6 ATP and not non-hydrolyzable analogs lead to conformational changes in $ClpB_M$. SAXS data based shape models of $ClpB_M$ in the absence and presence of various nucleotides have been presented. Orthogonal views have been presented alongside the shape models. As seen the presence of ATP and not other non-hydrolyzable analogs causes large scale conformational changes in $ClpB_M$.

3.2.6. High throughput screening identifies small molecule inhibitors of $ClpB_M$ ATP hydrolysis activity.

$ClpB_M$ is an attractive target for development of anti-mycobacterial agents as there is no functional homolog of $ClpB_M$ in mammals and its inactivation enhances *M. tuberculosis* sensitiveness towards other antibiotics (Lupoli et al., 2018b; Vaubourgeix et al., 2015). Therefore, identification of novel $ClpB_M$ inhibitors might pave the way for the development of novel anti *M. tuberculosis* therapeutics. Here, we optimized a malachite green (MG) based high throughput screening assay to identify inhibitors of ATPase activity associated with $ClpB_M$ (Figure 3.8A). We screened two different libraries of small compounds collectively consisting of 2480 molecules, as described in Materials and Methods. As shown in Figure 3.7 the release of inorganic phosphate due to $ClpB_M$ ATPase activity was observed in a time dependent manner. Similar assay when carried out in the presence of libraries of small molecules resulted in identification of both enhancers and inhibitors of $ClpB_M$ ATPase activity (Figure 3.8B). The 13 small molecules that inhibited more than approximately 70% of $ClpB_M$ ATPase activity were considered as preliminary hits. However, in our repeat assay from these 13 compounds, only 6 showed significant reproducible inhibition of the $ClpB_M$ activity in the range of ~70 to~90% (Figure 3.8C). Among these 6 compounds, except Tannic acid and Juglone, remaining compounds were freshly purchased to reconfirm the activity. Tannic acid is of relatively high molecular weight (~1.78kDa), and Juglone could not be obtained from available commercial sources. Among these freshly purchased compounds, only 3 showed significant inhibition of $ClpB_M$ ATPase activity in the range of 70% to 90% (Figure3.8D).

We next examined the specificity of identified inhibitors against other $ClpB_M$ homolog. As mammals do not encode a functional homolog of $ClpB_M$, we evaluated the inhibitory effect of the above 3 compounds against ATPase activity associated with *E. coli* ClpB and yeast Hsp104. Hsp104 is a member of the Hsp100 family and is required for growth of *S. cerevisiae* under stress conditions. Similarly, *E. coli* ClpB is required for thermotolerance. As shown in Figure 3.8D, we observed that at 100µM, all three compounds were able to inhibitHsp104 activity by >80%. For its prokaryotic homologue *E. coli* ClpB, NSC71948 shows the maximum inhibition of 95%, while hexachlorophene and NSC34931 showed 64% and 21% inhibition,

respectively. These observations suggest that the NSC71948 and hexachlorophene possess broad specificity and inhibit the enzymatic activity of ClpB protein from mycobacteria, *E. coli* and yeast.NSC34931 was found to be more specific against $ClpB_M$ and Hsp104. The chemical structures of identified three potential small molecules inhibitors are shown in Figure 3.8E.

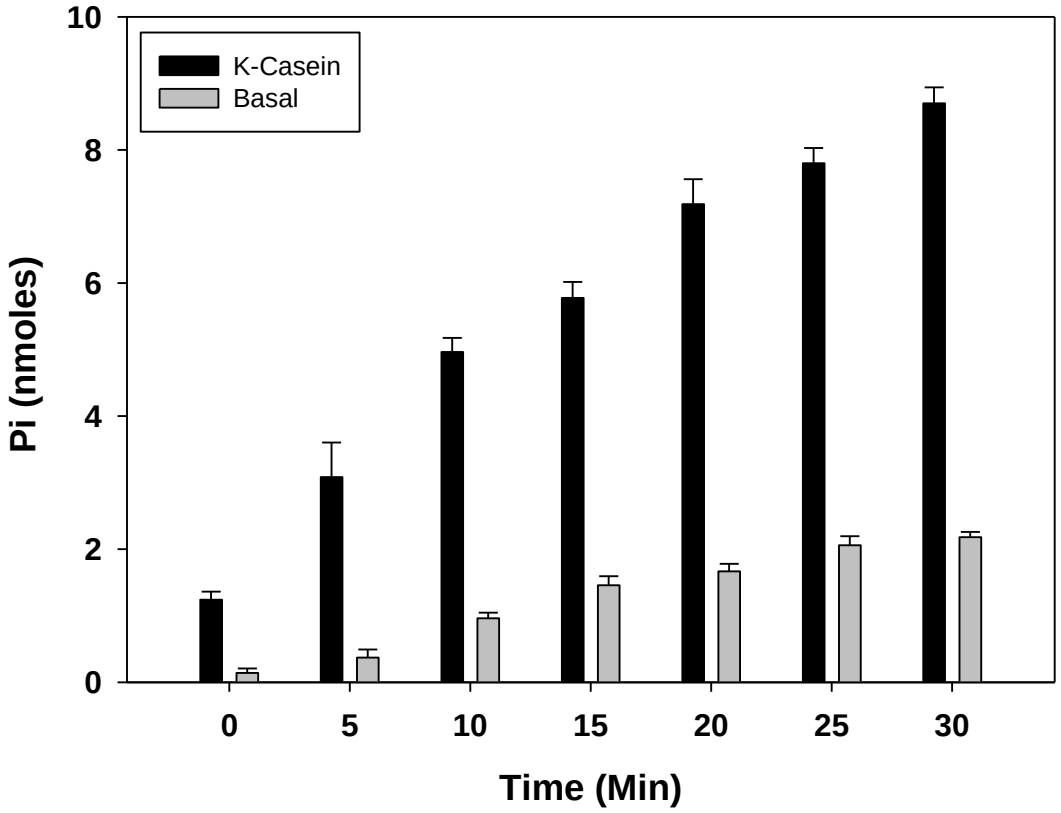

Figure 3.7. ATP hydrolysis activity for $ClpB_M$. Inorganic phosphate released upon ATP hydrolysis by $ClpB_M$ in the presence and absence (basal) of its substrate κ-casein was measured using malachite green assay by monitoring absorbance at 620nm of complex formed by malachite green oxalate and phosphomolybdate. The phosphate released was calculated using a standard plot obtained from plot of different known concentration of absorbance versus inorganic phosphate at 620nm.

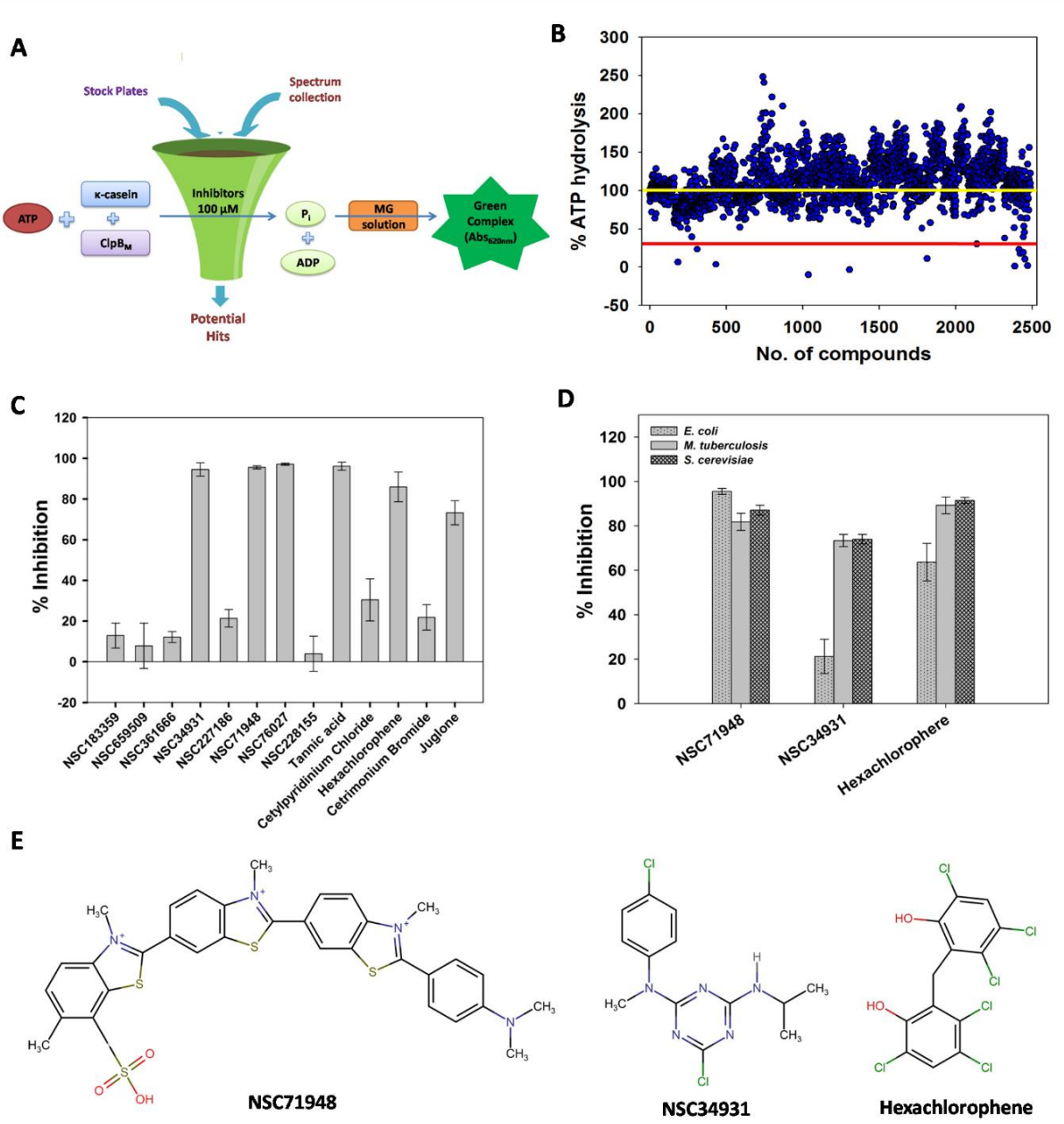

Figure 3.8. High-throughput screening identifies inhibitors of $ClpB_M$ ATP hydrolysis activity. (A) A schematic representation of the high-throughput screening system to identify potential inhibitors of $ClpB_M$ ATP hydrolysis using malachite green (MG) assay. The assay was carried out as described in Materials and Methods. **(B)** The obtained preliminary hits showing inhibition of $ClpB_M$ ATP hydrolysis activity. The red line shows desired threshold of 70% inhibition. The yellow line corresponds to $ClpB_M$ activity in the absence any drug. **(C)**Inhibition of $ClpB_M$ ATPase activity, using 13 preliminary hits obtained from our high throughput screening. **(D)** Cross inhibition activity of $ClpB_M$ inhibitors against *E. coli* ClpB and *S. cerevisiae* Hsp104. **(E)** Chemical structures of the best three identified inhibitors showing reproducible inhibition of $ClpB_M$ ATP hydrolysis.

3.2.7 *E. coli* chaperone system shows complete inhibition of luciferase refolding in the presence of drugs.

To further examine whether the identified small molecules inhibit chaperoning activity of ClpB, we examined the ability of *E. coli* ClpB to refold chemically denatured luciferase in the presence and absence of identified inhibitors as described in Materials and Methods. As shown in Figure 3.9A, no significant increase in luminescence was observed when only ClpB or DnaK was incubated with the denatured luciferase suggesting that these chaperones alone are unable to support luciferase refolding. The fraction of refolded luciferase increased with time, when chaperonic system consisting of DnaK and it co-chaperones DnaJ and GrpE was added with denatured luciferase in the refolding buffer. As expected, the addition of ClpB to DnaK chaperonic system not only decreased the lag-time for luciferase refolding but also had further increase in luminescence by around 25 fold in 60 minutes of incubation, suggesting a well-coordinated chaperoning activity of ClpB (Figure 3.9A). To examine the effect of drugs on refolding activity of ClpB chaperoning system, the denatured luciferase was incubated with the ClpB and the other chaperones along with 100μM of the identified inhibitor. We noticed reduced luciferase refolding in the presence of NSC71948, NSC34931 or hexachlorophene thereby suggesting complete inhibition of ClpB refolding activity (Figure 3.9B).

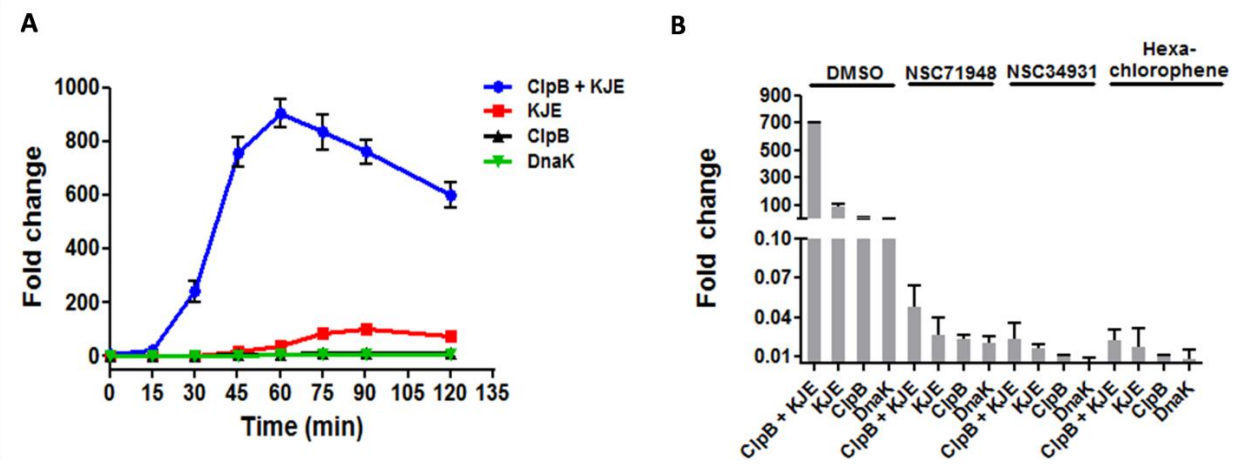

Figure 3.9 *E. coli* chaperone system shows complete inhibition of luciferase refolding in the presence of identified drugs. (A) The time course for refolding of chemically denatured luciferase by *E. coli* ClpB in collaboration with Hsp70 system (DnaK/DnaJ/GrpE). 22.6nM of luciferase was denatured in the presence of 7M urea and refolded in the presence

of 4.0 µM ClpB, 1.0 µM DnaK, 0.15 µM DnaJ and 0.05 µM GrpE. The KJE plot (red)represents luciferase refolding in the presence of DnaK+DnaJ+GrpE. **(B)** The luciferase refolding was measured in the presence and absence of drugs at 100 µM. The refolding reaction was performed with indicated combination of chaperones in the presence of either indicated compound or DMSO (as control). The luminiscence was recored after 60 minutes of addition of dentaured luciferase to the refolding reaction, and normalized with respect to that obtained from denatured luciferase.Error bars represents standard error of mean (S.E) from 2 different biological replicates.

3.2.8. Determination of IC_{50} value and affinity of lead compounds with $ClpB_M$.

The inhibitory activity of all three lead compounds on $ClpB_M$ was further measured as a function of their concentrations. As shown in Figure 3.10A, the compounds, NSC71948, NSC34931and hexachlorophene, showed IC_{50} value of 72.4µM, 60µM, and 40.5µM, respectively against $ClpB_M$ ATP hydrolysis activity *in vitro*.

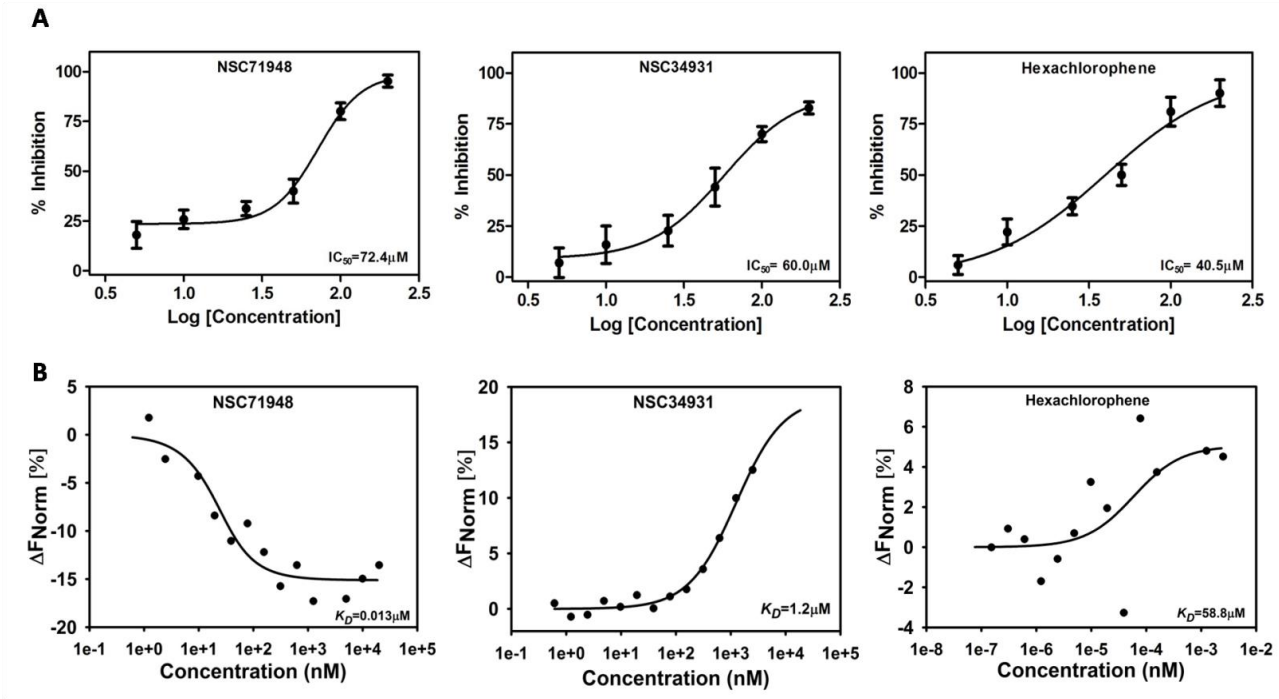

Figure 3.10. IC_{50} and binding affinity of inhibitors for $ClpB_M$. (A) Potential inhibitors were subjected to 6 point dose response concentrations from 200µM to 5µM with 0.25µM$ClpB_M$in the presence of κ-casein. IC_{50} values were calculated through a non-linear regression curve fitting by software Prism 5.0. Results are average of triplicate values and error bar represents standard error of mean (S.E). **(B)** Binding curves obtained by microscale thermophoresis for binding of indicated compounds with $ClpB_M$. For interaction, different concentrations of inhibitors ranging from 50µM to 0.61nM were incubated with a constant amount of fluorophore labeled $ClpB_M$ (20nM) and binding was measured as described in Materials and Methods.

We further examined the interaction of identified lead compounds with purified $ClpB_M$ using microscale thermophoresis (MST). The interaction was examined using a constant concentration of fluorophore labeled $ClpB_M$ with varying concentration of identified inhibitors. The MST analysis showed that NSC71948, NSC34931 and hexachlorophene interact with $ClpB_M$ with K_D of 0.013µM, 1.2µM and 58.8µM respectively (Figure 3.10B). Interestingly, binding of NSC71948 to $ClpB_M$ leads to decrease in normalized fluorescence, which is indicative of increased diffusion of bound complex as compared to free $ClpB_M$, thereby suggesting the drug induced significant conformational changes in the protein upon complex formation.

3.2.9. Hexachlorophene acts as a competitive inhibitor against $ClpB_M$ ATPase activity.

Among the three compounds, hexachlorophene was found to be relatively more potent than others in our ATPase inhibition and intracellular killing experiment. Next, we explored the mechanism underlying its inhibition of $ClpB_M$ ATPase activity using steady-state enzyme kinetics. Initially, the kinetics of $ClpB_M$ activity was determined at varying concentrations of ATP (Fig. 3.11). $ClpB_M$ displayed K_m of 408.48±12.37 µM with V_{max} of 8.79±0.01 µM/min for ATP as its substrate (Fig. 3.11).

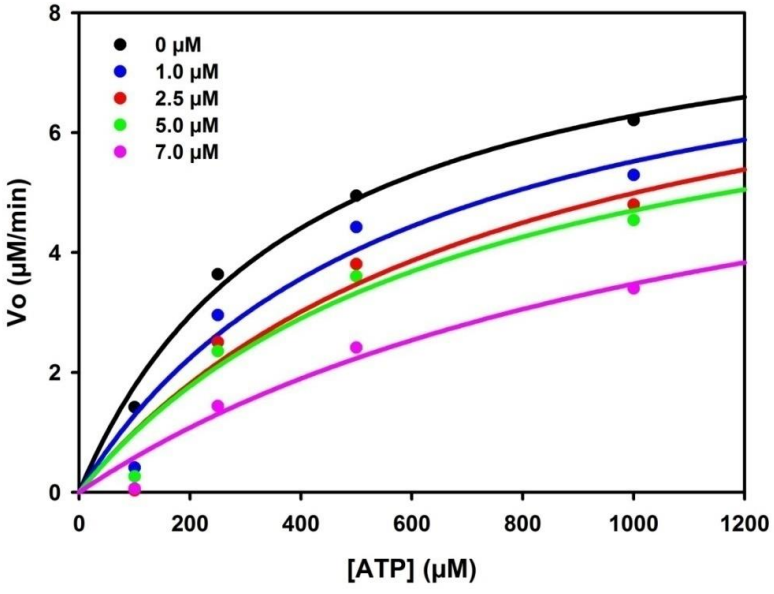

Figure 3.11. Hexachlorophene shows competitive mode of inhibition. The enzyme kinetic studies was carried out for ATP hydrolysis by $ClpB_M$ as a function of [ATP] (ranged from 100 µM to 1000 µM) in the presence and absence of hexachlorophene. Shown is the plot of initial velocity (V_o) versus indicated concentrations of ATP in the presence of 0 µM, 1.0 µM, 2.5 µM, 5.0 µM and 7.0 µM of Hexachlorophene.

Similar kinetic study in the presence of 1 µM hexachlorophene shows partial inhibition of ATPase activity with V_{max} around 8.6 ± 0.15 µM/min with increase in K_m to 572.21 ± 6.20 µM. A further increase of K_m was observed with increasing concentration of hexachlorophene (Table 3.2). At 7µM of the drug, K_m increased by about 3 fold to 1230.19 ± 23.86 µM in comparison to protein only control. The results suggest that the K_m and not V_{max} is dependent on the inhibitor concentrations indicating that hexachlorophene inhibits $ClpB_M$ ATPase activity in a competitive manner.

Table 3.2. Calculated V_{max} and K_m for $ClpB_M$ ATP hydrolysis.

Hexachlorophene (µM)	V_{max} (µM/min)	Km (µM)
0	8.79± 0.01	408.48± 12.37
1.0	8.55± 0.15	572.21± 6.20
2.5	8.91± 0.01	809.11± 27.57
5.0	8.43± 0.40	711.15± 4.11
7.0	7.47± 0.37	1230.19± 23.86

3.2.10. The identified drugs inhibits ATP induced conformational changes in $ClpB_M$.

Since we observed functional inhibition of $ClpB_M$ ATPase activity in the presence of the drugs, SAXS experiments were repeated in presence of the relatively more potent inhibitory molecules NSC71948 and hexachlorophene. As mentioned above, $ClpB_M$ was pre-incubated with drugs and subsequently ATP was added. Double logarithmic plots and Guinier analysis of the SAXS datasets with drugs showed that the $ClpB_M$ molecules were more like intermediates between hydrolyzable and non-hydrolyzable nucleotide bound state (Figure 3.12A and 3.12B). It appeared that Apo protein adopted shorter dimensions in the presence of small molecule inhibitors suggesting that addition of drugs did not allow the ATP mediated native-like reorganization of the protein (Figure 3.12C). Finally, shape restoration also showed P6 symmetry in the molecules, but they were not elongated like Apo or ATP state, but more like non-hydrolyzable nucleotide bound state (Figure 3.12D). Overall, SAXS data analysis of $ClpB_M$ with drug molecules indicated that these molecules

"capture" $ClpB_M$ molecules in a state of "limbo" which rendered them non-functional. These results suggest that dramatic shape rearrangement is a requirement for functionality of $ClpB_M$.

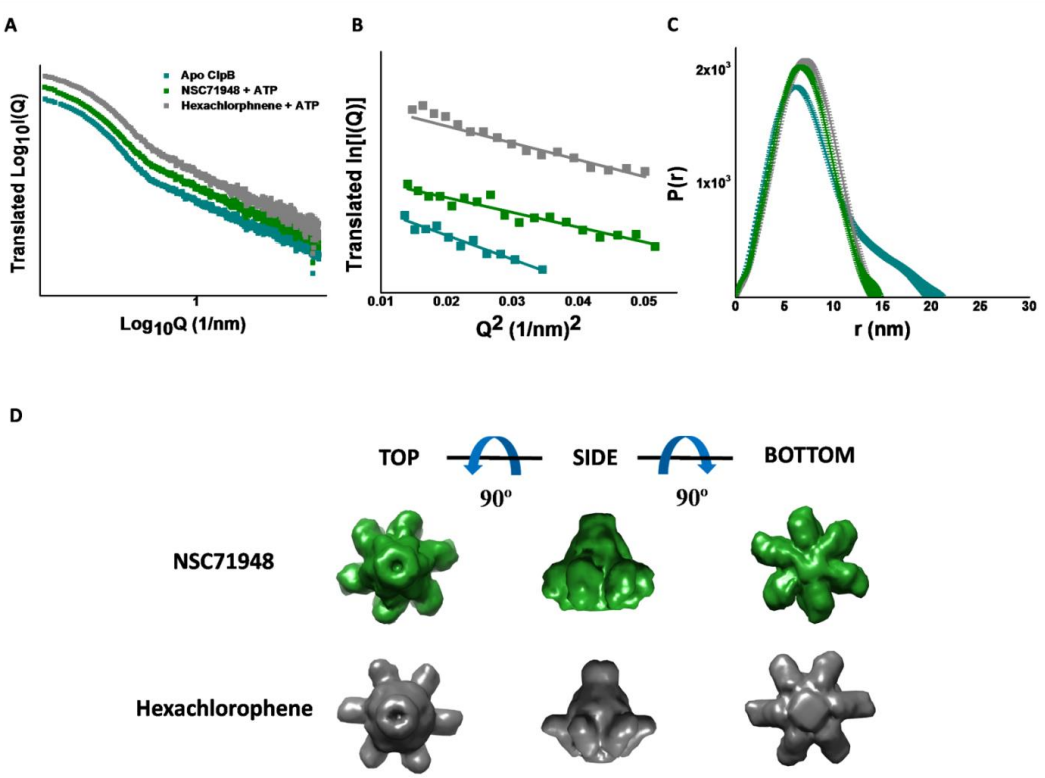

Figure 3.12 The lead identified drugs inhibit ATP induced conformational changes in $ClpB_M$. (A) SAXS datasets of $ClpB_M$ protein in the presence of drugs in double Log_{10} mode. Individual datasets and colour code are mentioned in the panels. Y-axis have been translated for clarity. **(B)** Guinier analysis shows the linear fit of the $ClpB_M$ datasets. **(C)** P(r) plots of the dataset represent the maximum dimensions in the presence of drugs. **(D)** SAXS profiles of $ClpB_M$ in the presence of inhibitors show drug induced structural changes in the chaperone in solution.

3.2.11 Anti-Mycobacterial assays.

We next determined the effect of identified lead compounds on *M. tuberculosis* and performed MIC_{99} determination assay. In our MIC_{99} determination assays, we observed that identified primary hits were inactive even at 100 µM concentration (the highest concentration in our assay) against *M. tuberculosis* in liquid cultures. These observations are in concordance with previous studies which have shown that $ClpB_M$ is dispensable for *M. tuberculosis* growth in vitro (Tripathi et al., 2020a; Vaubourgeix et al., 2015). However the mutant strain displayed a growth defect of 100.0-fold in bacterial numbers in comparison to wild type and complemented strain infected

animals. Therefore, we next determined the activity of the identified primary hits against intracellular mycobacteria in THP-1 macrophages. We observed that NSC 34931, NSC 71948, hexachlorophene and Isoniazid were non-cytotoxic to cells at 50 µM, 50 µM, 25 µM and 50 µM, respectively, till 96 hrs of incubation.

For intracellular killing experiments, macrophages represent physiological conditions that mimic disease and take into consideration the favourable contribution of host cells in the process of eradicating *M. tuberculosis*. Therefore, we checked the activity of the non-cytotoxic concentration of NSC 34931, NSC 71948, and hexachlorophene against intracellular *M. tuberculosis*inTHP-1 macrophages. As shown in Figure 3.13, maximum growth inhibition was seen in the case of hexachlorophene in comparison to untreated cells. We observed that exposure to hexachlorophene decreased the number of intracellular *M. tuberculosis* by ~8.0 fold in comparison to DMSO control. As expected, maximum growth inhibition was seen in THP-1 macrophages exposed to Isoniazid. The efficacies of compounds used in this assay against intracellular *M. tuberculosis* was Isoniazid>>Hexachlorophene> NSC 71948 > NSC 34931 respectively.

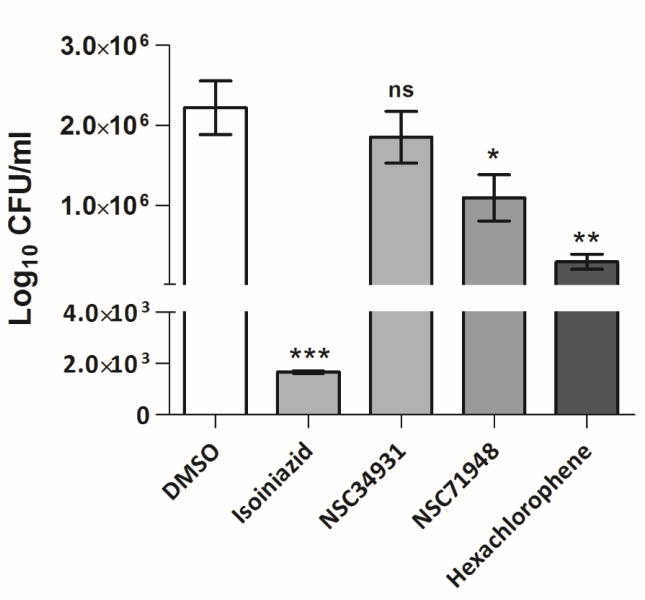

Figure 3.13 Intracellular activity of $ClpB_M$ ATPase Inhibitors. THP-1 macrophages were infected with *M. tuberculosis* H37Rv at a multiplicity of infection of 1:10 followed by treatment with various small molecules for 96 h. At designated time points, macrophages were lysed with 1x PBS containing 0.1% Triton X-100. The bacterial loads were determined by plating 100 µl of 10-fold serial dilutions on MB7H11 plates. Isoniazid was used as positive control. Each experiment was performed in triplicates and data shown in this panel is mean ± S.E. obtained of

Log_{10}CFU/ml obtained from two independent experiments. Statistically significant differences were obtained for the indicated groups (paired two-tailed t test, *P < 0.05, **P < 0.01, and ***P < 0.001).

3.3 Discussion

TB remains a major health concern primarily due to the emergence of various drug-resistant strains. How *M. tuberculosis* survives a prolonged dormant phase before its reactivation, as the host immune system is compromised, remains an unsolved aspect of *M. tuberculosis* biology. Various studies are now focused towards understanding of mechanism of stress tolerance for *M. tuberculosis* survival in host environment (Saito et al., 2017; Shastri et al., 2018; Zhai et al., 2019). Among various cellular factors, various chaperones including disaggregase $ClpB_M$ have emerged as key players that provide adaptation to various stresses encountered by the pathogen. The present study further broadens our understanding of $ClpB_M$, the primary disaggregase of *M. tuberculosis*, and provides small molecule inhibitors that inhibit $ClpB_M$ activity. The $ClpB_M$ inhibitors could provide an effective therapeutics against both symptomatic as well as latent tuberculosis.

CD spectroscopy studies showed that the purified $ClpB_M$ is properly folded into its native state. We observed that $ClpB_M$ thermally unfolds in a biphasic manner populating three states. This is in contrast to a previously reported study where authors showed $ClpB_M$ unfolds in a cooperative two-state manner when thermally heated to 90°C (Tripathi et al., 2018). This variation could be attributed to the use of His_6-tagged versus tagless $ClpB_M$ in previous versus present study respectively, and different purification protocol such as use of ammonium sulphate based precipitation, Ni-NTA based purification and use of 100mM KCl (than 200mM NaCl in present study) in previous study. As shown by AUC studies, based upon the concentration of protein and salts, $ClpB_M$ exists in different oligomeric states. A shift in equilibrium between different oligomeric states of $ClpB_M$ might affect its unfolding pathway. The biphasic transition observed in the present study could be due to the dissociation of a higher oligomeric state to relatively more stabilized smaller oligomer that subsequently unfolds as temperature is further increased.

The size exclusion chromatography studies show that the presence of nucleotides significantly affects $ClpB_M$ oligomerization. The relatively higher order oligomers were observed when $ClpB_M$ was incubated with ATP in comparison to ADP. AUC studies further confirmed these findings and we conclusively show that $ClpB_M$

oligomerization is sensitive to its concentration, nucleotides and salts in buffer. These observations are in concordance with previous reports for $ClpB_M$ and its other homologs (Krajewska et al., 2017; Lin and Lucius, 2015; Lupoli et al., 2016a; Tripathi et al., 2018; Zhang et al., 2013). The physiological significance of different oligomeric forms of $ClpB_M$ remains to be studied however it's possible that the dynamic oligomerization could be a mechanism to autoregulate $ClpB_M$ activity.

In order to identify small molecule inhibitors, we screened 2480 compounds with diverse structures from two different libraries, small molecule library from National Cancer Institute and spectrum collection. We observed that three compounds (NSC 71948, NSC 34931 and hexachlorophene) were consistent in their ability to inhibit the ATP hydrolysis. It is interesting to notice that hexachlorophene with the weakest affinity to $ClpB_M$ is the strongest inhibitor of its ATP hydrolysis activity. In agreement with its higher efficacy against the ATPase activity, the hexachlorophene was found be more effective than other two compounds in reducing the growth of *M. tuberculosis* in infected macrophages. This might be attributed to either the intracellular concentration of the identified compounds or their stability in experimental conditions.

The enzyme kinetics studies showed that the strongest inhibitor hexachlorophene inhibits $ClpB_M$ in a competitive mode of inhibition indicative of a direct competition with ATP. The designed chemical scaffold would be useful in the design of new derivatives with relatively stronger binding to $ClpB_M$ as well as increased intracellular activity. Since ClpB is highly conserved among different species, the identified compounds showed cross-inhibition with homologs from *E. coli* and yeast *S. cerevisiae*. Surprisingly, the NSC34931 showed better inhibition of $ClpB_M$ and Hsp104 than *E. coli* ClpB suggesting that though highly conserved, small structural differences exist between ClpB homologs from *M. tuberculosis* and *E. coli*. The understanding of such differences among ClpB from different organisms would be useful for specific targeting of the pathogenic organisms.

The shape model from the SAXS data profile showed that apo $ClpB_M$ undergoes a shape rearrangement from extended to more globular shape with P6 symmetry. Interestingly $ClpB_M$ with non-hydrolyzable ATP analogs could not attain a similar structure to that observed in the presence of ATP. The observations suggest that in addition to ATP binding, its hydrolysis is also required to drive $ClpB_M$ structure to more open form. We also observed that the identified lead compounds also inhibited

ATP driven $ClpB_M$ structural rearrangement. The $ClpB_M$ structure obtained in the presence of small molecules was more similar to that observed in the presence of non-hydrolyzable ATP. Therefore, we speculate that the identified small molecules drugs might interfere with the ATP hydrolysis of the $ClpB_M$.

Extensive research is underway to enhance understanding of *M. tuberculosis* biology, and to explore new therapeutic approaches against it. Previous approaches are primarily directed towards inhibition of DNA/RNA synthesis or energy metabolism as these pathways are critical for the bacterial survival. More recently, several studies have revealed the critical role of *M. tuberculosis* chaperone machinery in its stress tolerance and survival broadening our understanding of *M. tuberculosis* and opening new avenues for the design of novel therapeutic approaches. The biochemical characterization and high throughput screening assay performed in the present study paves the way for design of more potent small molecule inhibitors against ClpB protein.

Chapter 4

Characterization of Hsp70 system and exploring nucleotide exchange activity of GrpE of Mycobacterium tuberculosis.

4.1 Introduction

Tuberculosis commonly known as TB is a highly infectious disease caused by *Mycobacterium tuberculosis (M. tuberculosis)*. TB remains a global health problem and is the primary cause of deaths in humans among communicable diseases. WHO estimated that, one third of the total world's population approximately 2-2.5 billion people, are infected with *M. tuberculosis*. In 2018 alone, 1.2 million people died and 10 million people are affected with (WHO, Global Tuberculosis Report, 2019).The pathology of *M. tuberculosis* is still limited and insufficient, hindering the advancement of successful diagnosis and therapeutics, even after many decades of its discovery. *M. tuberculosis* can survive in its host without showing any symptoms for years. This dormant or metabolically less functioning mycobacteria is reactivated when the host immune system is weakened (Muñoz et al., 2015).During host infection and combination of antibiotic drugs, *M. tuberculosis* undergoes several stresses like reactive oxygen and nitrogen species and various proteostasis stress which ultimately results in protein aggregation (Kumar et al., 2011a; Lupoli et al., 2018a). *M. tuberculosis* deploys its chaperones system to overcome such stresses by refolding the misfolded proteins back to their native form. Therefore, a deeper understanding of the *M. tuberculosis* chaperone machinery will allow development of novel therapeutic strategies. Hsp70 chaperone family plays a crucial role in the maintenance of proteastasis network and thus, are critical for optimal growth under stress conditions (Calloni et al., 2012; Saibil, 2013b). Hsp70 effectively binds the exposed hydrophobic regions of misfolded proteins and further prevent their aggregation. Hsp70 collaborates with Hsp100 to perform efficient refolding the disaggregated substrate processed by Hsp100 (Goloubinoff et al., 1999b). These chaperones are also important for the regulation of cell cycle, protein trafficking and degradation, as well as their role in protein folding (Gupta et al., 2018; Reeg et al., 2016; Rohde et al., 2005; Truman et al., 2012).

The ubiquitously present Hsp70 is highly conserved across species from prokaryotes to humans. Though eukaryotes possess multiple cytosolic Hsp70 members, prokaryotes such as *E. coli* and *M. tuberculosis* harbors only a single Hsp70 isoform, DnaK. The activity of Hsp70is modulated by various co-chaperones such as Hsp40 (DnaJ1 and DnaJ2 in *M. tuberculosis*) and nucleotide exchange factors (NEF, GrpE

in *M. tuberculosis*) (Pierpaoli et al., 1998; Szabo et al., 1994b).Structurally, Hsp70 consists of an amino-terminus nucleotide-binding domain (NBD) and a carboxy-terminus substrate-binding domain (CTD) (Jiang et al., 2005a; Kityk et al., 2012). The binding of ATP at NBD leads to the opening of substrate binding pocket for substrate to interact (Qi et al., 2013). The Hsp40 stimulates Hsp70 ATPase activity and induces further conformational changes in Hsp70 that closes the C-terminal lid over the substrate-binding pocket trapping partially unfolded substrates (Fan et al., 2003; Lupoli et al., 2016a). Further, NEFs facilitate substrate release and initiation of the new Hsp70 reaction cycle by promoting exchange of ADP with ATP (Bracher and Verghese, 2015b; Harrison et al., 1997b; Lupoli et al., 2016a; Mally and Witt, 2001).
The DnaK/ClpB bichaperonic system of *M. tuberculosis* has been shown to interact with protein aggregates *in vivo*. *M. tuberculosis* encodes for DnaK and is part of an operon that also encodes for other cofactors such as DnaJ1 and GrpE. *M. tuberculosis* DnaK ($DnaK_M$) and GrpE ($GrpE_M$) are essential for its viability (Fay and Glickman, 2014). The other isoform of DnaJ, DnaJ2 ($DnaJ2_M$), is highly homologous (sequence identity ~40%) to DnaJ1 ($DnaJ1_M$). In *M. smegmatis,* the two DnaJ isoforms individually are non-essential however, their double deletion is synthetically lethal for the cellular growth (Lupoli et al., 2016a).

In order to expand our understanding of *M. tuberculosis* Hsp70 chaperone machinery, we have performed a detailed biophysical characterization of $DnaK_M$ and its co-chaperones ($DnaJ1_M$, $GrpE_M$). We show that thermal unfolding of $DnaK_M$ or $DnaJ1_M$ at pH 7.4 is irreversible, whereas $GrpE_M$ unfolds in a reversible two-state manner. Analytical ultracentrifugation studies reveal the monomeric form of $DnaK_M$, while the other proteins of chaperone machinery, $DnaJ1_M$ and $GrpE_M$ are primarily dimeric in solution. $DnaK_M$ exhibits a strong interaction with $GrpE_M$ with a binding of 47.6 nM. The release of MABA-ADP from $DnaK_M$.Maba-ADP complex confirms the nucleotide exchange activity of $GrpE_M$. The study thus provides insight into biophysical and biochemical aspects of *M. tuberculosis* Hsp70 refolding machinery, and paves way for further design of inhibitors against *M. tuberculosis* chaperone machinery. Thus, our study provides insight into biophysical and biochemical aspects of *M. tuberculosis* Hsp70 refolding machinery making it a potential target for combating TB.

4.2 Results

4.2.1 Secondary structure characterization of $DnaK_M$ and its co-chaperone $DnaJ1_M$ and $GrpE_M$ of *M. tuberculosis.*

To examine the folded nature of purified chaperones, the secondary structure of $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ was studied under native conditions using Far-ultraviolet (UV) circular dichroism (CD). The Far-UV CD provides insight into the secondary structural content of proteins such as α-helix, β-sheet or random coil. The proteins were purified to homogeneity and CD spectra were recorded in Far-UV range from 190nm to 250nm at pH 7.4, 25°C.

The CD studies show that all proteins are well folded. The CD spectra for $DnaK_M$ and $DnaJ1_M$ show features of α-helix however with relatively lower ellipticity than as expected from a completely α-helical protein, suggesting that these proteins consist of mixture of α-helical and β-sheet regions (Fig. 4.1 A and 4.1 B). Far-UV CD studies on purified $GrpE_M$ reveals signature of α-helices as indicated by local minima of CD signal at 208 and 222nm (Fig. 4.1C). The ratio of CD signal at 222nm and 208nm of 0.8 suggests lower inter-helical contacts in $GrpE_M$. Taken together, above studies suggest that at secondary structural level of $DnaK_M$ and its co-chaperones $DnaJ1_M$, $GrpE_M$ are mostly similar to their respective counterparts in *E. coli*.

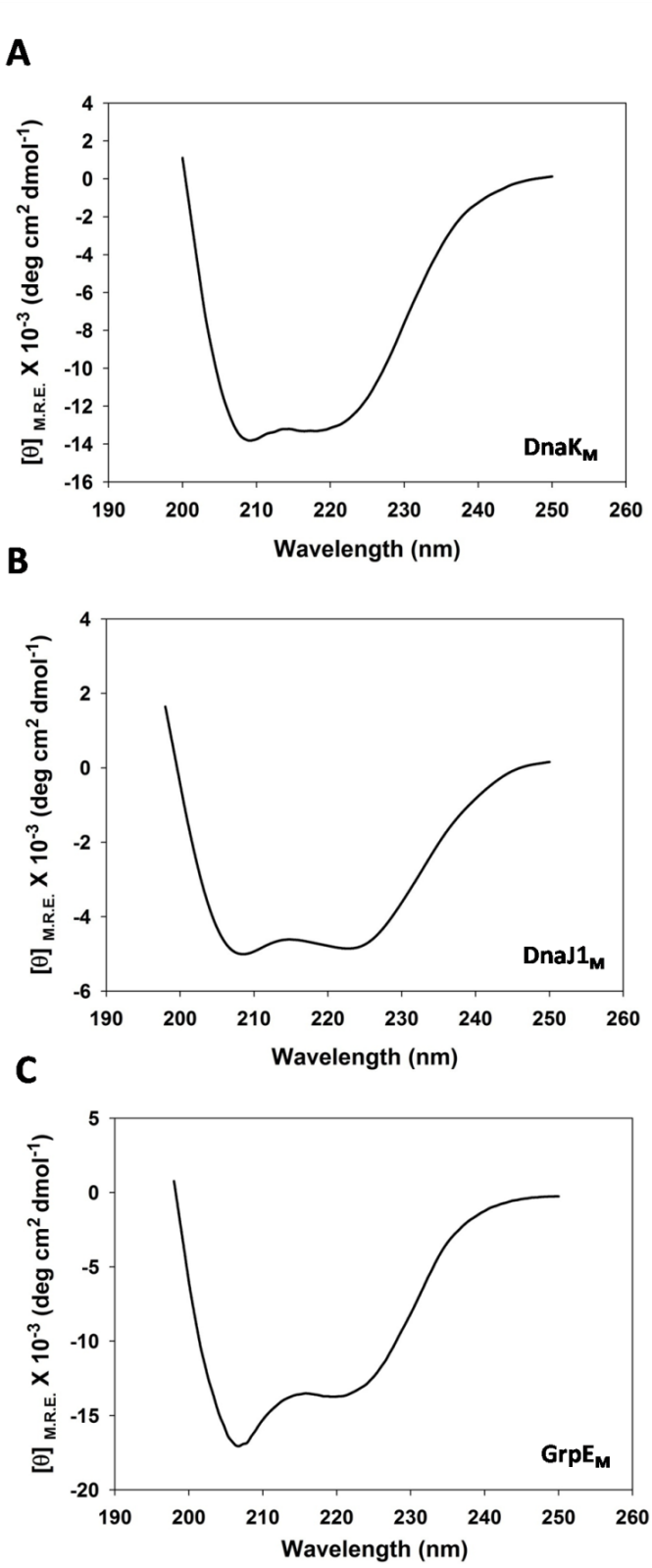

Figure 4.1 Secondary structural analysis of components of bichaperonic system. Panel **(A), (B) and (C)** show Far-UV CD spectra of $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ respectively. The spectra were recorded with purified proteins in 1mm path length cuvette at 25°C. The spectrum shown here is the average of three independent scans.

4.2.2 Thermal stability of $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ shows stable intermediates and renaturation ability of chaperones.

We next examined the effect of increase in temperature on secondary structural content of the components of mycobacterial chaperone machinery. The ellipticity of proteins at 222nm was monitored as the incubation temperature was raised from 25°C to 95°C.

As shown in Figures 4.2A and 4.2B, no significant changes in the ellipticity of $DnaK_M$ was noticed as temperature was increased upto 40°C. We observed two thermal transitions in the temperature range between 40-95ºC. Among the two transitions, the lower and higher temperature transition showed an apparent midpoint of about 45°C and 60°C, respectively. The $DnaK_M$ thermal transition was largely reversible as seen by an increase in negative ellipticity upon lowering the temperature from 95°C to 25°C (Fig. 4.2B). Unlike $DnaK_M$, $DnaJ1_M$ showed two-state unfolding with no secondary structural changes until 50°C, and a single thermal transition with a midpoint of about 55°C (Fig. 4.2C and 4.2D). We also observed that $DnaJ1_M$ thermal unfolding is irreversible when heated to a temperature of 95°C.

As shown in Figures 4.2E and 4.2F, $GrpE_M$ loses α-helical content significantly as the temperature is increased to 95°C. The $GrpE_M$, which is primarily α-helical, shows two state thermal unfolding with apparent midpoint of transition at 48°C. We further observed that $GrpE_M$ unfolds reversibly when heated to 95°C.

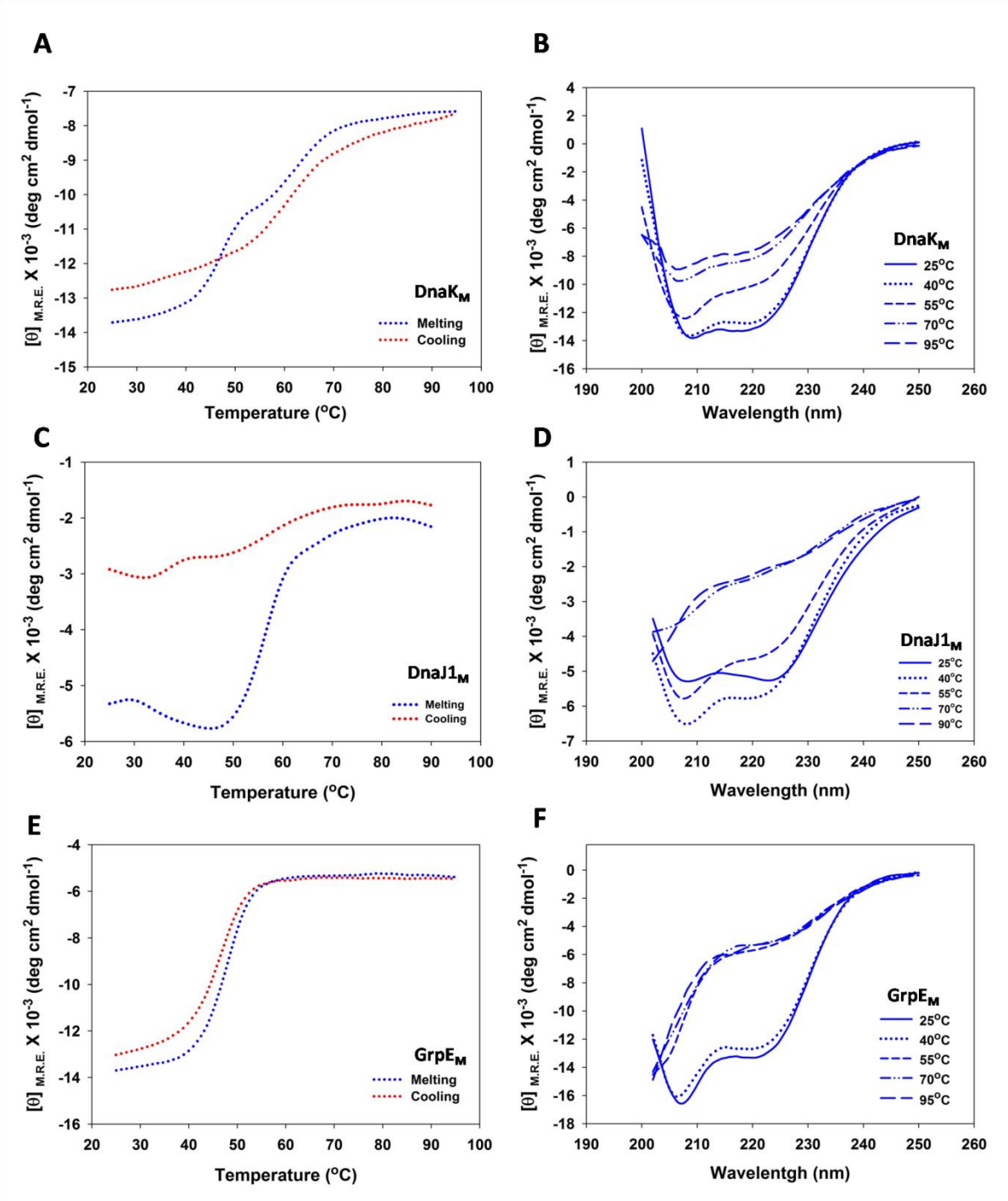

Figure 4.2. Thermal unfolding of $DnaK_M$, $DnaJ1_M$ and $GrpE_M$. The thermal stability of purified chaperones was measured by monitoring changes in mean residue ellipticity of proteins at 222nm as a function of increase in temperature from 25°C-95°C at scan rate of 1°C/min. **(A), (C) and (E)** show thermal denaturation curves (blue) as $ClpB_M$, $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ respectively were incubated at increasing temperature. Red curves show changes in ellipticity as incubation temperature was decreased from 95ºC to 25ºC. Panel **(B), (D) and (F)** show changes in Far-UV CD spectra of chaperone $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ respectively at various temperatures.

4.2.3 Gel-filtration analysis of DnaK$_M$, its co-chaperones.

DnaK$_M$ requires ATP to perform its refolding function (Lupoli et al., 2016b). In order to examine the effect of nucleotides on oligomerization of DnaK$_M$, we performed size exclusion chromatography analysis in the presence and absence of ATP or ADP. As shown in Figure 4.3A, in the absence of any nucleotide, DnaK$_M$ eluted at a volume corresponding to the molecular weight of 142kDa. Further, as shown in Figure 4.3A, we observed that pre-incubation with either ADP or ATP did not promoter self-association of DnaK$_M$.

Previous studies have shown that DnaJ, as well as GrpE of *E. coli* primarily exists as homodimer in solution (Li et al., 2009; Mehl et al., 2003; Schönfeld et al., 2007). The size exclusion chromatographic studies showed that DnaJ1$_M$eluted at a molecular weight of 189kDa, significantly higher than expected from the dimeric protein (~80kDa) which could either be due to self-association or its non-globular structure (Figure. 4.3B). The elution profile of GrpE$_M$ showed that the protein eluted corresponding to a molecular weight of 258kDa (Figure 4.3C).The sharpness and nearly symmetrical nature of the peak indicated that GrpE$_M$ primarily exists in a single oligomeric state and does not form aggregate in solution (Figure 4.3C).

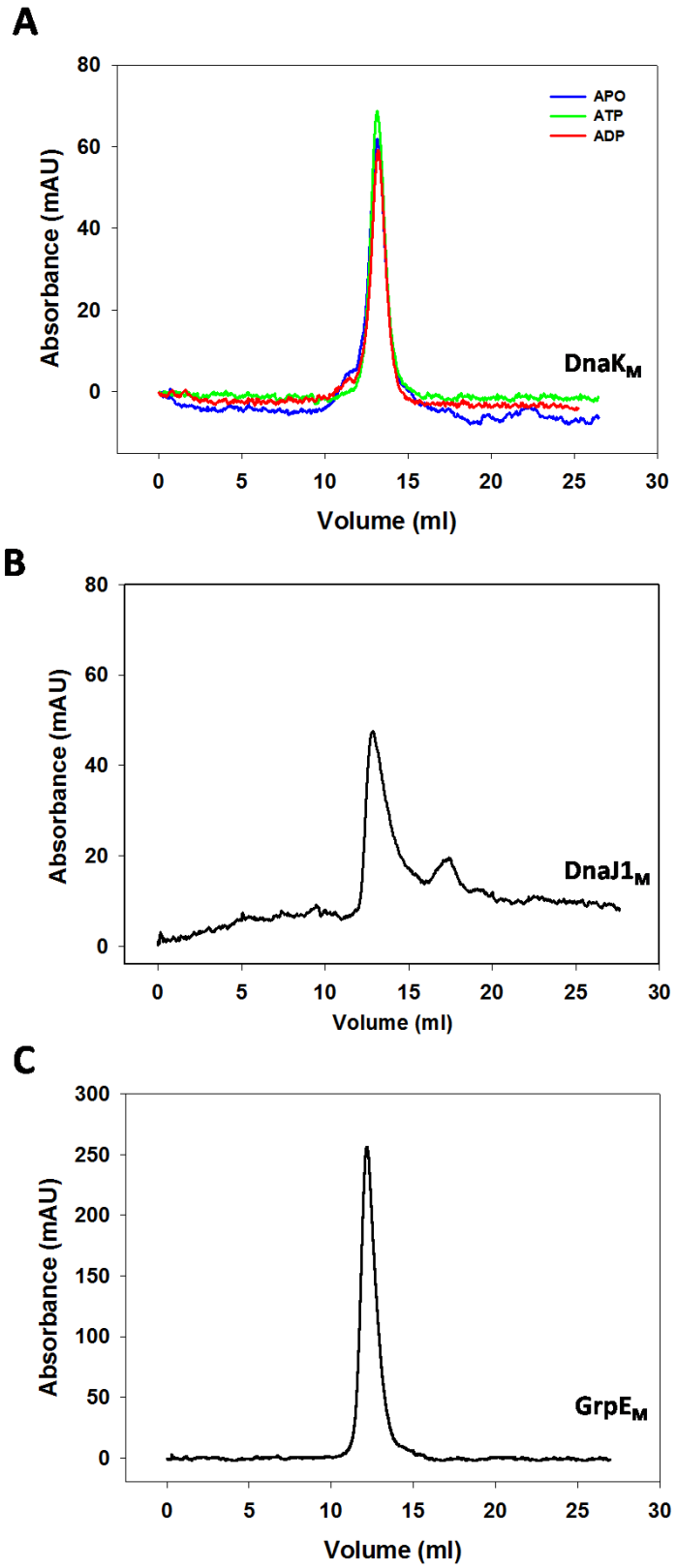

Figure 4.3 Gel filtration profiles for DnaK$_M$, DnaJ1$_M$ and GrpE$_M$. (A) DnaK$_M$ elution profiles showing that pre-incubation of nucleotides have no effect on its oligomerization. **(B)** and **(C)** shows chromatograms for DnaJ1$_M$ and GrpE$_M$ respectively.

4.2.4 The oligomeric status of $DnaK_M$ and its co-chaperones through analytical ultracentrifugation studies.

The apparent molecular weight as estimated from above gel-filtration studies for $DnaK_M$ and its co-chaperones is significantly higher than as expected based upon their respective homologs from other organisms such as *E. coli*. The observed disagreement between the observed and expected molecular weight could be attributed to their non-globular structure, self-association or interaction with the column matrix. We thus carried out AUC studies to examine oligomerization state of these chaperones.

In order to determine the effect of the $DnaK_M$ concentration on its oligomerization, we performed sedimentation velocity experiments at three different protein concentrations, 0.61mg/ml, 1.22mg/ml and 1.84mg/ml. Figure 4.4A show the radial absorbance plot for the sedimentation of $DnaK_M$. The c(s) analysis of absorbance scans at these concentrations revealed the presence of a single dominant species (~65% of total species) with $s_{20,w}$ of 4.17S, which corresponds to the molar mass of monomeric form of $DnaK_M$ (71kDa, Fig. 4.4D). Further, at 1.22 and 1.84mg/ml of protein concentration, we observed a small fraction of additional species sedimenting as broad peak of $s_{20,w}$ values of 6.25S and 7.33S respectively. The presence of these additional peaks suggests the accumulation of higher order oligomers with 2-3 units of $DnaK_M$ at relatively higher concentrations of protein (Table 4.1).

Next, we examined the association state of Hsp40 $DnaJ1_M$. The purified $DnaJ1_M$ was subjected to sedimentation velocity AUC experiments, at three different concentrations of 0.66mg/ml, 1.31mg/ml and 1.98mg/ml. The radial scans and c(s) analysis of absorbance scans showed that at lower concentrations of 0.66mg/ml and 1.31mg/ml, the protein primarily sedimented as single major species at sedimentation coefficient of $s_{20,w}$ = 4.71S which is consistent with molecular weight of dimeric $DnaJ1_M$ (~80kDa) (Figure 4.4B and 4.4E). We noticed that at relatively higher concentration of 1.98mg/ml, in addition to the dimeric $DnaJ1_M$, higher order oligomeric species (collectively 27% of total species) of sedimentation coefficient of 7.53S and 15.31S were also formed which correspond to 4and 10-11 units ofmonomeric $DnaJ1_M$ respectively (Table 4.1). The presence of these peaks

indicated thatDnaJ1$_M$tends to multimerize into larger size oligomersat higher concentrations.

The oligomerization state of GrpE$_M$ was also examined using AUC. The sedimentation velocity AUC experiments were performed at 2.5mg/ml, 5.0mg/ml and 7.5mg/ml of purified GrpE$_M$ (Fig. 4.4F).The c(s) plots show that at all concentrations examined, GrpE$_M$ sedimented predominantly as a single species of $s_{20,w}$ = 2.41S, which corresponds to the molecular weight of dimeric GrpE$_M$.

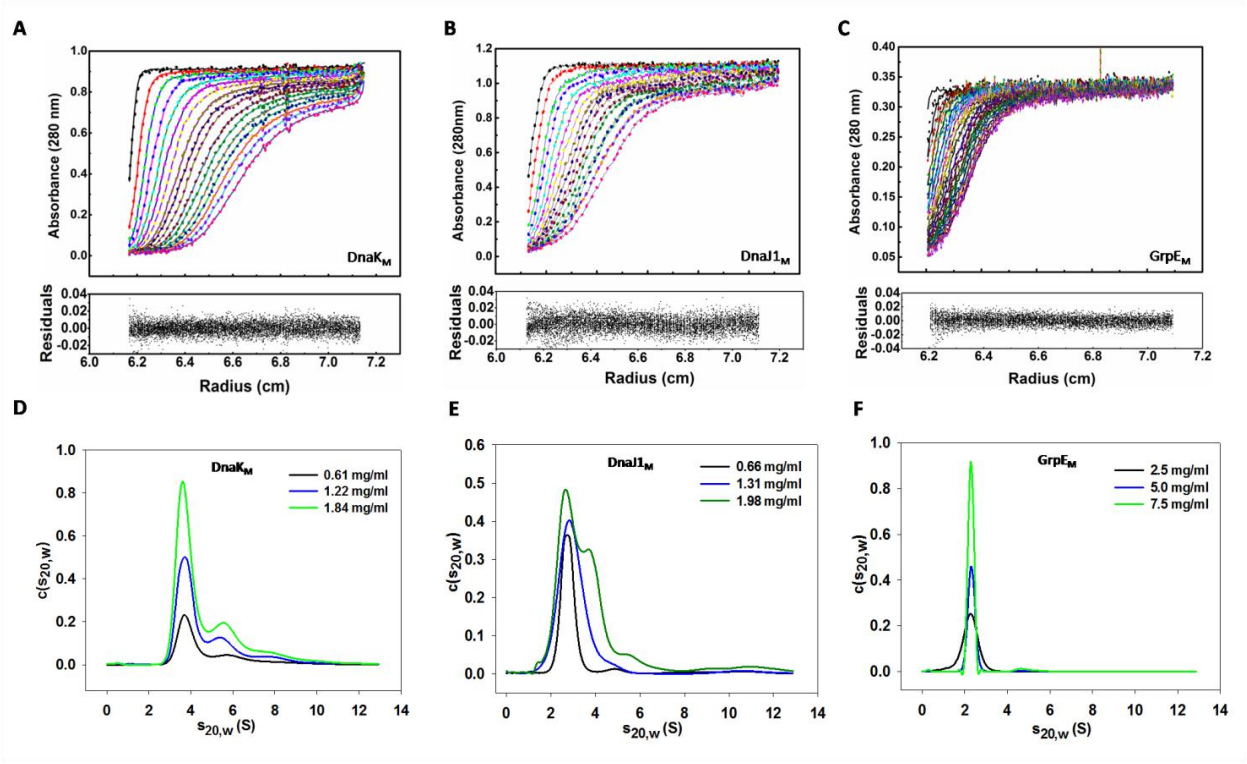

Figure 4.4 Sedimentation velocity studies of DnaK$_M$, DnaJ1$_M$ and GrpE$_M$. Sedimentation velocity data for indicated proteins at 40,000 rpm and 20°C. Shown are the radial absorbance sedimentation scans as a function of radial distance over a period of time and sedimentation coefficient continuous distribution c(s) plots for proteins, obtained from their respective sedimentation velocity data of indicated chaperones. **(A)** 1.84mg/ml DnaK$_M$. **(B)** 1.98mg/ml DnaJ1$_M$ and **(C)** 2.5mg/ml GrpE$_M$. The scans were recorded for every 3 minutes. For clarity, every 3rd scan is shown. **(D)** c($s_{20,w}$) distribution curves of DnaK$_M$ at different concentrations show that the protein predominantly exists as dimer with tendency to oligomerize at higher concentrations. **(E)** Sedimentation plots of DnaJ1$_M$ show that the protein sediments primarily as dimer with tendency to form higher order oligomers at higher concentration. **(F)** GrpE$_M$ sediments as single dimeric species.

Table 4.1 Sedimentation coefficient values of $DnaK_M$, $DnaJ1_M$ and $GrpE_M$ obtained from analytical ultracentrifugation studies.

	$S_{20,w}$				Mol. Wt (KDa) [Pred. units]				% species			
Protein	**Peak 1**	**Peak 2**	**Peak 3**	**Peak 4**	**Peak 1**	**Peak 2**	**Peak 3**	**Peak 4**	**Peak1**	**Peak2**	**Peak3**	**Peak4**
DnaK (mg/ml)												
0.61	4.17	7.71	-	-	71 [1.0]	178 [2.6]	-	-	64.48	30.12	-	-
1.22	4.09	6.25	9.77	-	70 [1.0]	133 [1.9]	261 [3.8]	-	68.53	21.86	9.22	
1.84	4.07	7.33	-	-	70 [1.0]	171 [2.5]	-	-	60.01	31.16	-	-
DnaJ1 (mg/ml)												
0.66	4.71	8.43	14.05	18.53	75 [0.9]	180 [4.3]	385 [9.2]	586 [13.9]	66.28	3.34	2.52	3.37
1.31	5.10	-	-	18.77	83 [1.0]	-	-	590 [14.0]	94.48	3.62	-	-
1.98	4.75	7.53	15.31	19.34	76 [0.9]	153 [3.6]	444 [10.6]	631 [15.0]	42.69	25.47	1.57	3.51
GrpE (mg/ml)												
2.5	2.48	-	-	-	52 [2.0]	-	-	-	93.70	-	-	-
5.0	2.41	-	-	-	50 [2.0]	-	-	-	82.61	-	-	-
7.5	2.38	-	-	-	49 [2.0]	-	-	-	77.15	-	-	-

4.2.5 DnaK$_M$ forms a stable complex with GrpE$_M$.

DnaK functions in coordination with DnaJ and GrpE (Schröder et al., 1993; Szabo et al., 1994b). Though DnaK$_M$ interaction with DnaJ1$_M$ is well characterized, not much is known about DnaK$_M$-GrpE$_M$ complex. The association between DnaK$_M$ and GrpE$_M$ was first studied using the gel-filtration column that separates proteins based upon their hydrodynamic radii. For gel filtration experiments, DnaK$_M$, GrpE$_M$ or solution containing both proteins at molar ration of 1:1 was loaded onto Superdex-200 gel-filtration column, and elution profile was monitored at 220nm. As shown in Figure 4.5A, when loaded alone, DnaK$_M$ and GrpE$_M$ eluted at a volume of 13.4ml and 12.2ml, respectively. The identity of the proteins in elution fractions was further confirmed on 10% SDS-PAGE (Figure 4.5B). However, when a solution containing pre-incubated DnaK$_M$ and GrpE$_M$ was loaded, two overlapping peaks at elution volume of 11.4ml (fraction 1) and 12.7ml (fraction 2) were observed (Figure 4.5A).The eluted fractions were further subjected to 10% SDS-PAGE and we observed that the fraction 1 consisted of both proteins, DnaK$_M$ and GrpE$_M$ (Figure 4.5B).The fraction 2showed the presence of 70kDa protein on SDS-PAGE consistent with free DnaK$_M$ protein. The co-elution of DnaK$_M$ and GrpE$_M$ in fraction 1 on the size exclusion column suggests the formation of a stable complex between these proteins.

The complex formation between DnaK$_M$ and GrpE$_M$ was further analyzed under non-reducing conditions on Native-PAGE. The DnaK$_M$, GrpE$_M$ or mixture of both proteins at a molar ratio1:1 was incubated at 4°C for 4 hours before loading onto Native-PAGE. In concordance with our AUC studies, the presence of monomeric as well as oligomeric form of DnaK$_M$ was observed (Figure 4.5C). The GrpE$_M$ protein showed the presence of predominantly one species. As expected, the sample corresponding to a mixture of DnaK$_M$ and GrpE$_M$ showed the presence of one additional protein band, absent in lanes loaded with DnaK$_M$ or GrpE$_M$ alone, indicative of the formation of a stable complex between the two proteins. As expected, this additional band disappeared upon boiling of protein samples at 99^{o}C before loading onto Native-PAGE. Overall, these results suggest that under native conditions, GrpE$_M$ forms a stable complex with DnaK$_M$.

We further quantified the binding affinity between $DnaK_M$ and $GrpE_M$ using Surface Plasmon Resonance (SPR). For SPR experiments, His_6-$DnaK_M$ was immobilized on Ni-NTA sensor surface and further incubated with ADP in running buffer. Varying concentrations of $GrpE_M$ was flown over the ligand bound surface. As shown in Figure 4.5D, we observed an increase in the response unit with increasing concentration of $GrpE_M$. The fit of the data suggests that complex forms with on-rate (k_a) of 8.8 x $10^3 M^{-1}s^{-1}$, dissociates with off-rate (k_d) of 4.2 x $10^{-4}s^{-1}$, and dissociation constant (K_D) of 47.6nM.

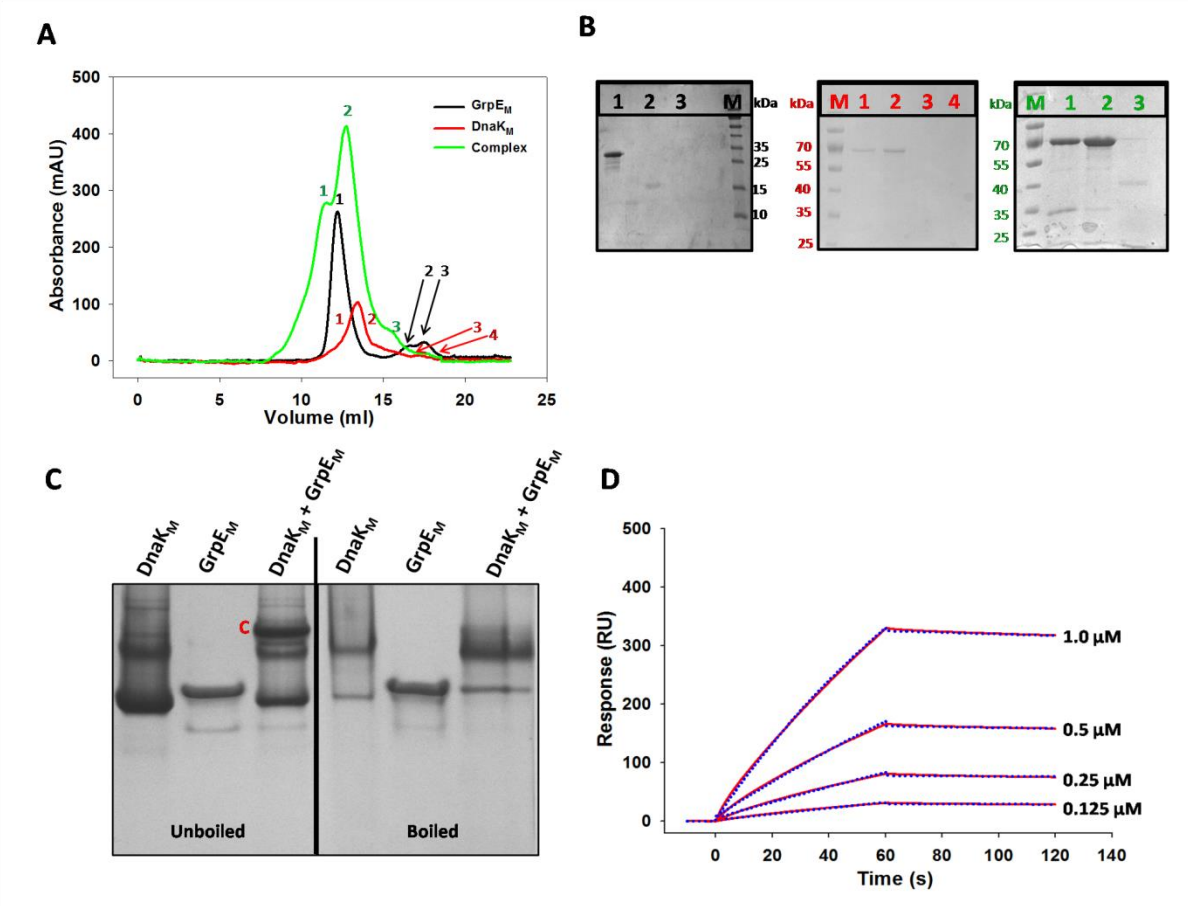

Figure 4.5 $DnaK_M$ forms stable complex with $GrpE_M$. (A) Size-exclusion chromatography profiles of $DnaK_M$ (red), $GrpE_M$ (Black) and sample containing both proteins at molar ratio of 1:1 (Green). **(B)**The fractions eluted as separate peaks shown in Panel A were collected and fractionated onto 10% SDS PAGE. **(C)** Fractionation of sample containing $DnaK_M$, $GrpE_M$ and their complex on native PAGE. The sample containing both proteins shows a high molecular weight species (labelled as 'C') (Left Panel) which disappears upon boiling of samples (Right Panel) suggesting the formation of stable complex under native conditions between $DnaK_M$ and $GrpE_M$. **(D)** Surface plasmon resonance studies to monitor interaction between $DnaK_M$ and $GrpE_M$. Indicated concentrations of $GrpE_M$ were flown over the $DnaK_M$ bound sensor surface, and binding was examined by monitoring increase in response unit.

4.2.6 $GrpE_M$ is a Nucleotide exchange factor for $DnaK_M$.

The essential regulatory role of the nucleotide exchange factor is evaluated by more than just how often it destabilises the association of its target protein with nucleotide, but also by the acceleration of the exchange of nucleotide. We therefore measured the change the fold change of nucleotide exchange through stopped flow experiment (Figure 4.6). A complex of $DnaK_M$ and MABA-ADP was formed by mixing 1.0 µM $DnaK_M$ and 0.1 µM MABA-ADP in syringe A of the stopped flow instrument. This complex was then rapidly mixed with 200 µM ADP in the presence or absence of 2 µM $GrpE_M$. The data was represented the dependence of the displacement of MABA-ADP by ADP as a function of time in the presence and absence of $GrpE_M$ was 2.5 folds. This shows the nucleotide exchange activity of $GrpE_M$ for $DnaK_M$.

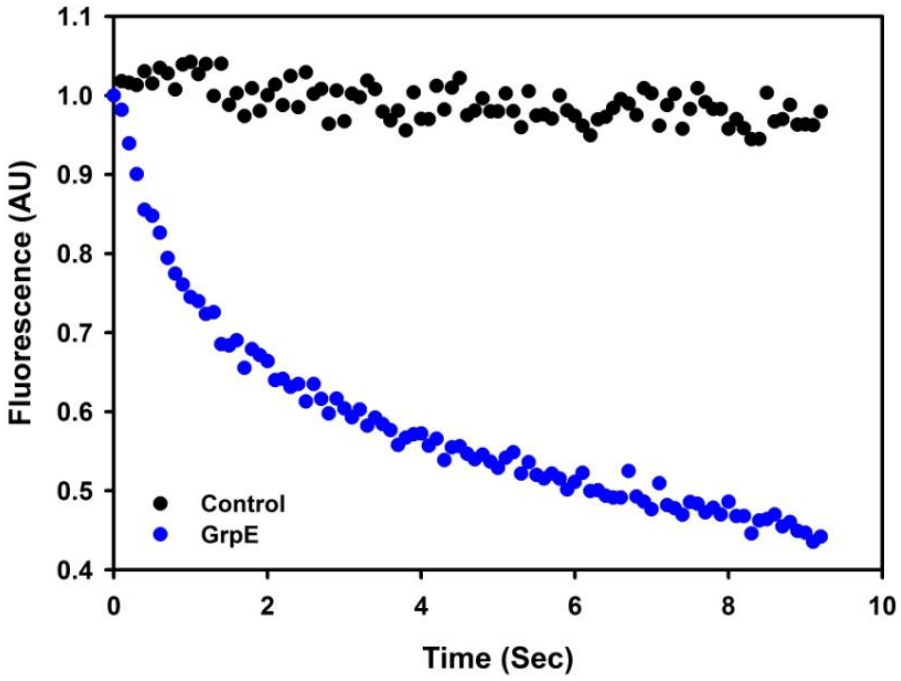

Figure 4.6 $GrpE_M$performs nucleotide exchange activity for $DnaK_M$. Time course shows the displacement of MABA-ADP from the complex $DnaK_M$.MABA-ADP upon addition of $GrpE_M$ (Blue). Signal for MABA-ADP exchange with ADP was recorded as decrease in fluorescence upon dissociation from $DnaK_M$.MABA-ADP complex. Complex was excited at 360nm and emission was monitored at 435nm.

4.3. Discussion

Tuberculosis remains one of the leading cause of deaths worldwide. *M. tuberculosis* evolving into a drug resistant form is the major reason for its survival. Numerous studies are already directed towards deciphering the stress survival mechanism of *M. tuberculosis* during the infection (Saito et al., 2017; Shastri et al., 2018; Zhai et al., 2019). Chaperones are identified as important players among different cellular components, ensuring resistance to different stresses experienced by the pathogen. The present study further expands our understanding of *M. tuberculosis* Hsp70 chaperone system, as a potential target.

CD spectroscopy studies showed that the purified *M. tuberculosis* Hsp70 chaperones are well folded into their native state and secondary structure was similar to that reported for the *E.coli* homologs. $DnaK_M$ showed three state denaturation with first transition being more cooperative than the second. The biphasic curve obtained during denaturation suggests the presence of at least one intermediate upon thermal unfolding of the protein. The observed three state unfolding is similar to previously reported thermal and Gdn.HCl mediated unfolding of its homolog from *E. coli* (Grimshaw et al., 2001; Palleros et al., 1992; Palleros, 1993). For *E.coli* DnaK, it is proposed that the intermediate formed during its unfolding possesses properties of a molten globule state with disrupted tertiary structure however similar secondary structure as of the native protein (Palleros et al., 1992; Palleros, 1993). Interestingly, intermediate formed during thermal unfolding of $DnaK_M$ has similar fraction of folded secondary structure (80%) as reported during denaturant mediated unfolding pathway of *E. coli* DnaK (85%). Considering the similarity of secondary structure between the two chaperones, it is thus possible that intermediate observed during $DnaK_M$ unfolding in the present study is also molten globule in nature. $DnaJ1_M$ unfolds in a two state unfolding pathway with irreversible renaturation. $GrpE_M$ thermal unfolding showed a characteristic of simple two-state unfolding. The loss of ellipticity upon heating was almost completely restored as the temperature was lowered to 25°C suggesting $GrpE_M$ unfolding is reversible. The temperature dependent sigmoidal unfolding suggests that dimeric $GrpE_M$ undergoes cooperative unfolding to an unfolded monomer without formation of a stable intermediate. The two state unfolding of $GrpE_M$ is in contrast to the thermal unfolding

of its homolog from *E. coli* which unfolds in a three state manner and shows the presence of a stable intermediate in its unfolding pathway (Gelinas et al., 2002; Grimshaw et al., 2001). The difference in the unfolding pathway of GrpE homologs from *M. tuberculosis* and *E. coli* could be because of their lower sequence identity (21%) as well as differences in interactions within secondary structural elements.

The size exclusion chromatography studies show that the incubation of $DnaK_M$ in the presence of nucleotides no affects on its oligomerization. $DnaJ1_M$ and $GrpE_M$ eluting nearly at it 4 and 10 times of their monomeric molecular mass signifies the self association and non-globular shape of the respective chaperones.

Sedimentation velocity studies further confirmed these findings for the oligomerization statuts of Hsp70 chaperones. The sedimentation velocity studies show that both $DnaK_M$ and its co-chaperone $DnaJ1_M$ primarily remain in their functionally active monomeric and dimeric form, respectively with tendency to further oligomerize at higher concentrations. These observations are in concordance with previous studies showing that in the absence of ATP, *E. Coli* DnaK has tendency to form higher order oligomers (Benaroudj et al., 1996; Schönfeld et al., 1995). While GrpE pre-dominantly sediments as dimeric species in agreement with previous results shown for *E. coli* (Schönfeld et al., 1995).

$DnaK_M$ functions with co chaperones $DnaJ1_M$ and $GrpE_M$ for its refolding function. Interaction of $DnaK_M$ with $DnaJ_M$ (both isoforms DnaJ1 and DnaJ2) is well characterized for *M. tuberculosis* (Lupoli et al., 2016b).The other half of refolding cycle is performed by interaction of GrpE with DnaK resulting in substrate release and rebinding of ATP for multiple round of cycles. Therefore, Interaction of $DnaK_M$ with $GrpE_M$ was confirmed by gel filtration chromatography where appearance of new Peak eluting at higher molecular weight than $DnaK_M$ and $GrpE_M$ alone shows the complex $DnaK_M$-$GrpE_M$ formation. Further, the complex was also confirmed by Native Page analysis where the band corresponding to complex migrating at higher size as compared to individual proteins. The complex band disappears upon boiling and migrates to the size $DnaK_M$ and $GrpE_M$ alone. The binding affinity between $DnaK_M$ and $GrpE_M$ was quantitated in vitro using Surface Plasmon Resonance (SPR). The SPR is a widely used method to obtain quantitative parameters such as on-rates, off-rates and dissociation constants. In SPR, one of the binding partner is

immobilized on a sensor surface, and other is flown over the surface, and binding is monitored by increase in resonance unit. For SPR experiments, His-$GrpE_M$ was immobilized using Ni-NTA sensor surface, and $DnaK_M$ in the presence or absence of nucleotides was run over the surface

Stopped flow studies show that the $GrpE_M$ performs the nucleotide exchange activity for $DnaK_M$. The release of MABA-ADP from $DnaK_M$.MABA-ADP complex upon the rapid mixing of $GrpE_M$ and ADP results in as the loss in fluorescence depicting the exchange of MABA-ADP with ADP.

Substantial research is still needed to expand understanding of *M. tuberculosis* pathology and the discovery of alternative clinical methods toward tuberculosis. Previous strategies developed were targeting the essential biology such as interference with cell metabolism or transcription machinery of the bacteria. Recently, several studies have revealed the significant role of *M. tuberculosis* chaperone machinery for overcoming various stresses ensuring its survival. The biophysical and biochemical characterization performed in current study broadens the understanding of *M. tuberculosis* Hsp70 chaperone machinery and designing of potential inhibitors for combating tuberculosis.

Chapter 5

Developing a comprehensive database of modulators for Heat shock proteins (Hsps).

5.1 Introduction

Cellular proteins are exposed to various kinds of stresses such as changes in temperature, pH and metal ion concentrations, which induces protein misfolding and aggregation (Hartl, 1996). The intracellular accumulation of protein aggregates adversely affects cell viability and is the underlying basis of various human diseases. Thus, each cell has an evolved set of proteins known as heat shock proteins, which are components of the cellular quality control system to prevent protein aggregation. The heat shock proteins interact with exposed hydrophobic patches of aggregation-prone proteins and thereby protect cells from the deleterious effects of protein aggregates (Mayer and Bukau, 2005; Saibil, 2013a). These ubiquitously present proteins in different organisms are highly conserved across different species from bacteria to humans. In addition to their role in preventing protein aggregation, Hsps are also involved in protein synthesis, protein trafficking, assembly of multi-protein complexes and protein degradation (Craig et al., 1993; Fernández-Fernández and Valpuesta, 2018). Based upon the approximate molecular weight, Hsps are categorized into different families such as Hsp100, Hsp90, Hsp70, Hsp60 or Hsp40 family. Hsp100 family of proteins is AAA+ (ATPase associated with diverse cellular activities) superfamily of ATPase that either in coordination with Hsp70 facilitates disaggregation or with a protease rings promotes protein degradation (Mogk et al., 2015). Hsp90 functions to promote refolding of various growth hormone receptors, kinases, transcription factors and many viral proteins (Schopf et al., 2017). Hsp70 functions in coordination with Hsp40s to bind to partially unfolded substrates and promote their folding (Mayer and Bukau, 2005). In addition to stimulating the ATPase activity of Hsp70, Hsp40 also facilitates substrate transfer to the substrate binding domain ofHsp70s. Hsp70 also binds to the number of other cellular factors that play a crucial role in regulating substrate fate, e.g. interaction with ubiquitin ligase CHIP at C-terminus of Hsp70 promotes substrate degradation. Hsp60 proteins are known to perform variety to functions such as maintenance of mitochondrial protein homeostasis, cellular signalling, and its inactivation is associated with multiple disorders such as in neurodegenerative diseases (Bukau and Horwich, 1998; Cheng et al., 1989). Many of the Hsp families possess highly homologous multiple members which perform both redundant as well as non-redundant functions (Sharma and Masison, 2008). Many previous studies have been focused on understanding the mechanism of Hsps action in various biological pathways. To comprehend such

enormous data from different studies, few databases have been designed that provide comprehensive understanding of the functions and roles of these chaperones. HSPIR provides information on sequence, structure, localization and biological roles of Hsps (R et al., 2012a). A comprehensive information of chaperone interaction could be accessed through Protein Homeostasis Database (Ramakrishnan et al., 2020). Similarly, sHSPDb (Jaspard and Hunault, 2016) and CrAgDb (Rani et al., 2017) decipher information about small heat shock proteins and archaeal chaperones respectively. As various human diseases are related to protein misfolding disorders such as neurodegenerative disorders (Leak, 2014; Sharma and Masison, 2009), and various forms of cancer (Chatterjee and Burns, 2017), Hsps have been extensively studied as potential therapeutic targets against these diseases (Kampinga and Bergink, 2016; Shirota et al., 2015), and over the last two decades, considerable efforts have been made towards developing modulators of Hsps activities. Many of these modulators are currently being evaluated for their efficacy in different phases of clinical trials (Hendriks and Dingemans, 2017; Jhaveri et al., 2012; Kim et al., 2009). However, the information about these modulators from different studies is largely scattered, and no common platform of Hsp modulators with their activities and physicochemical properties has been established until now. Such platform would enable a better understanding of various scaffolds used for targeting different Hsps and thus facilitate rational drug discovery approaches. In this study, we have made a systematic attempt to collect and compile comprehensive information of experimentally validated modulators (activators and inhibitors) of five major Hsps (Hsp100, Hsp90, Hsp70, Hsp60 and Hsp40) from published literature. The user interface developed in the database also enables users to find the similarity between their compounds of interest with any of the modulators deposited in the database. We anticipate that HSPMdb will be very useful for the scientific community working in the areas.

5.2 Results and Discussion

HSPMDb, is a freely accessible repository containing modulators for Heat shock proteins. Figure 5.2.1 shows the user interface for HSPMdb web page. The current version of HSPMdb catalogues 10,223 entries of Hsp modulators among which 10 159 entries are of Hsp inhibitors while 59 entries are of Hsp activators. The above information was manually extracted from 176 research articles. Figure 5.2.2 provides the architecture of HSPMdb where information of modulators against five different

Hsp classes with maximum entries of Hsp60 (4052 entries) followed by Hsp90 (3448 entries), Hsp70 (1488 entries), Hsp100 (1183 entries) and Hsp40 (52 entries) has been categorised.

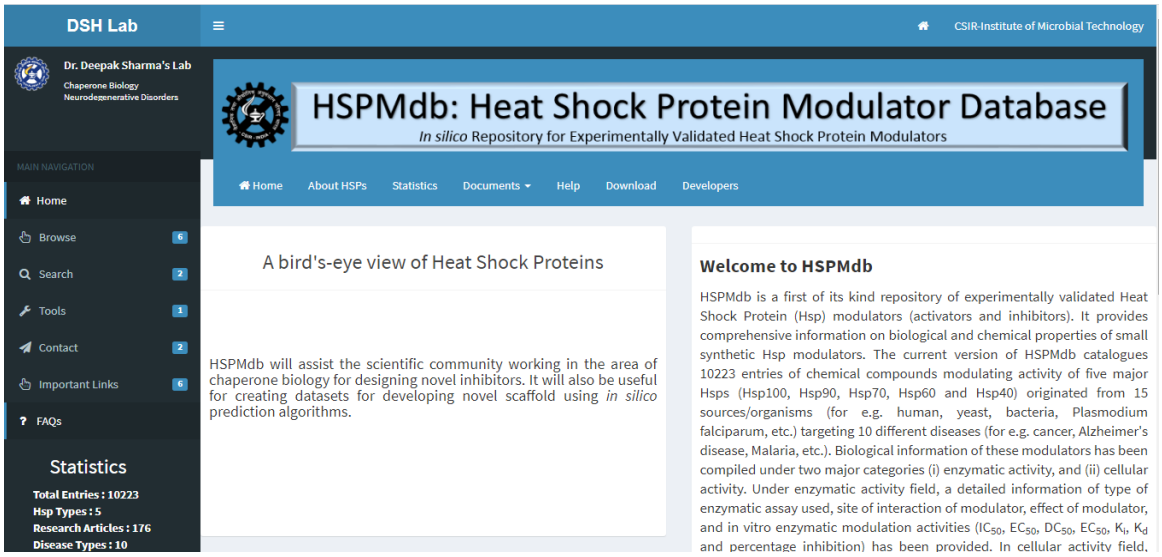

Figure 5.2.1 Interface of HSPMdb web page.

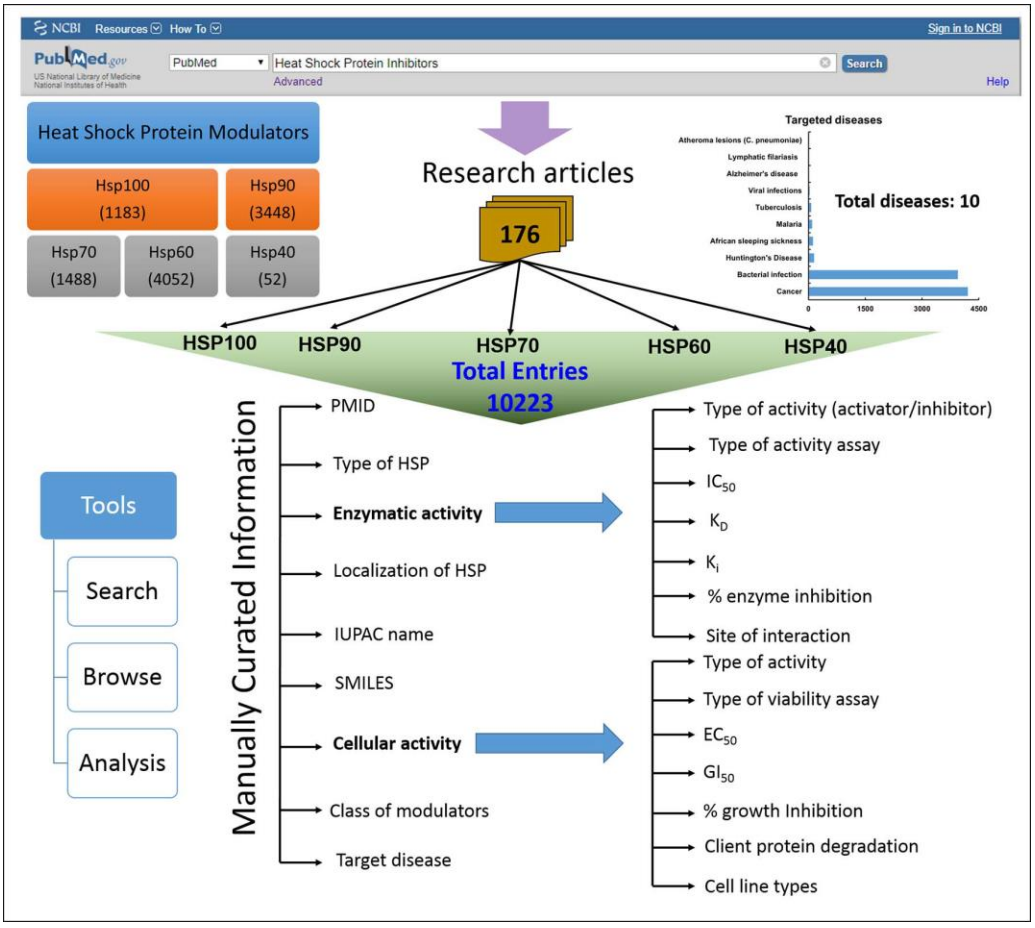

Figure 5.2.2 Architecture of HSPMdb.

These Hsps belong to different organisms (e.g. human, yeast, bacteria, *Plasmodium falciparum*) and currently information about Hsps from a total of 15 organisms has been compiled (Fig 5.2.3). A Schematic representation of HSPMdb browsing page has been shown in Figure 5.2.4).

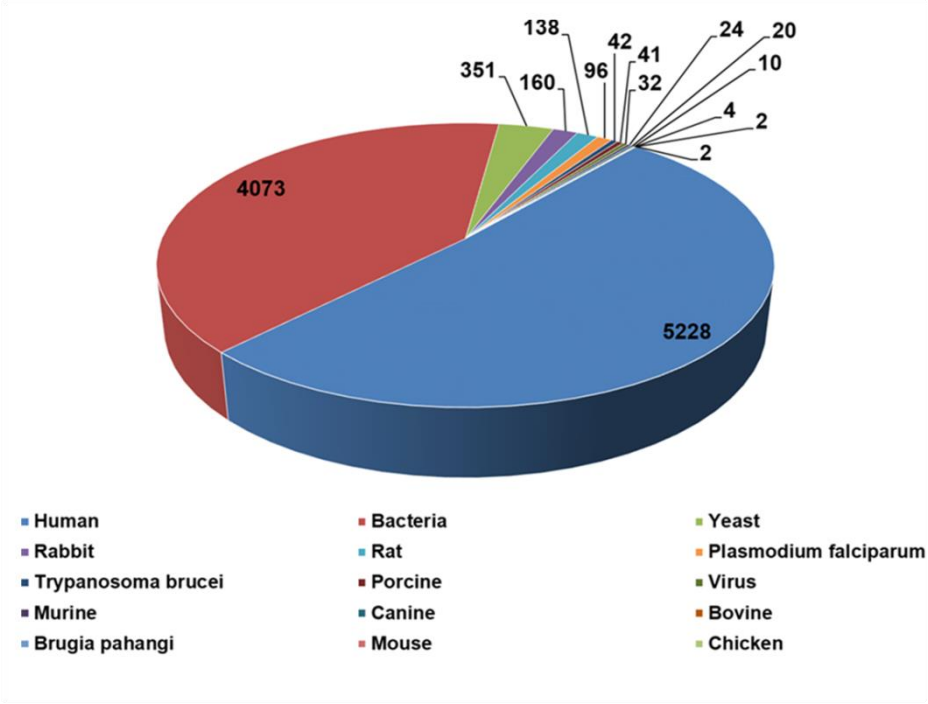

Figure 5.2.3 Total number of entries in HSPMdb based upon origin of Hsps.

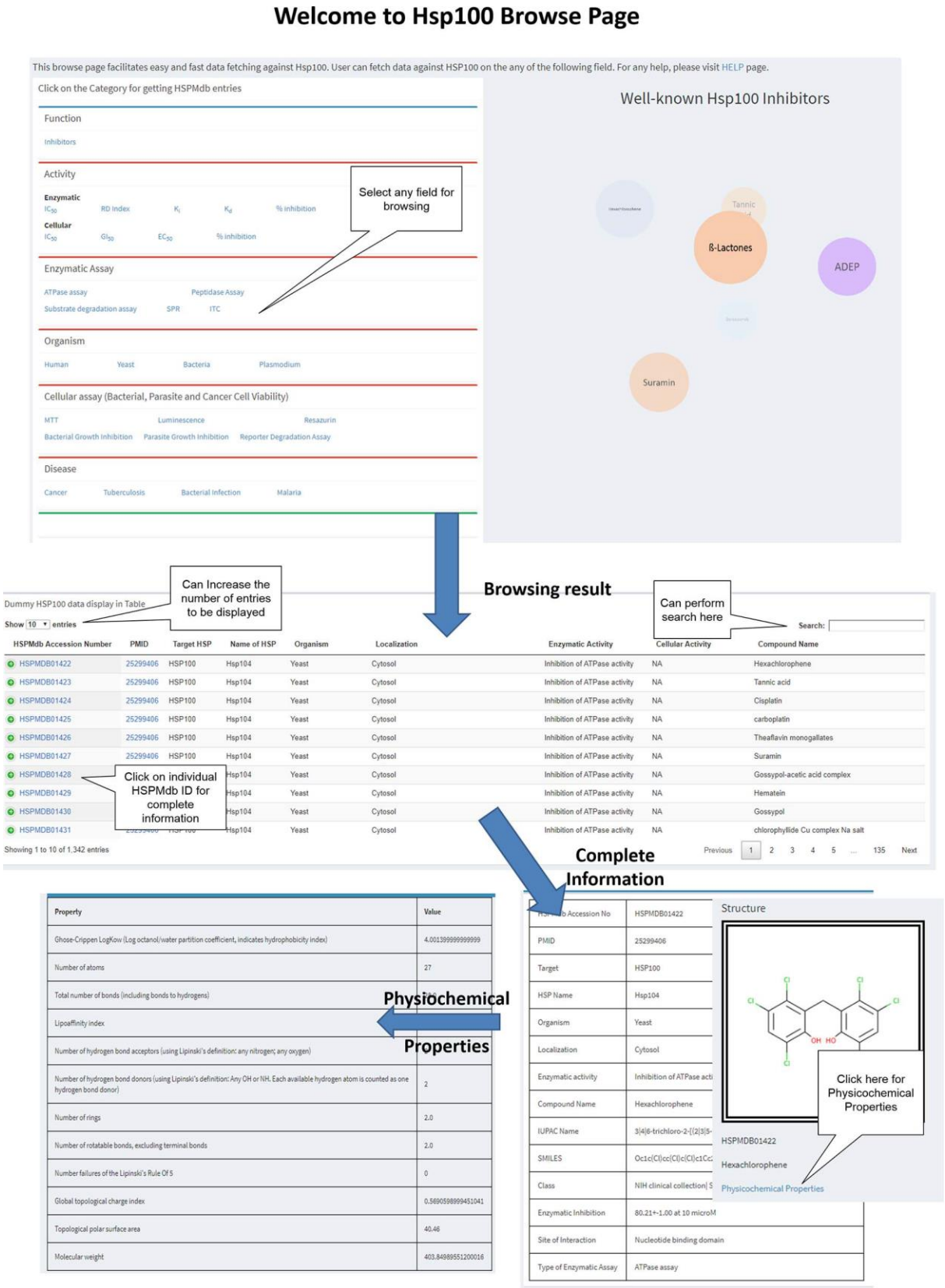

Figure 5.2.4 Schematic representation of browsing page of HSPMdb.

Maximum entries are compiled on modulators against human Hsps (5228 entries) which is as expected from the multiple roles played by these proteins in diverse cellular processes. The presence of a large number of modulators against human Hsps further suggests that targeting these Hsps is one of the major ongoing therapeutic intervention strategies. The second and third most common modulators were found against bacterial Hsps (4073 entries) and yeast Hsps (351 entries) suggesting studies targeting bacterial diseases are more than yeast diseases. Hsps have been extensively explored as therapeutic targets against various human diseases (Kampinga and Bergink, 2016; Nahleh et al., 2012; Shirota et al., 2015) primarily associated due to protein misfolding or aggregation. In addition, various pathogens also require their own chaperoning machinery for survival under stressful conditions encountered in the host (Pesce et al., 2010; Ramdhave et al., 2013). In the current version of HSPMdb, information about Hsps' modulators against 10 different diseases has been compiled (Figure 5.2.5).

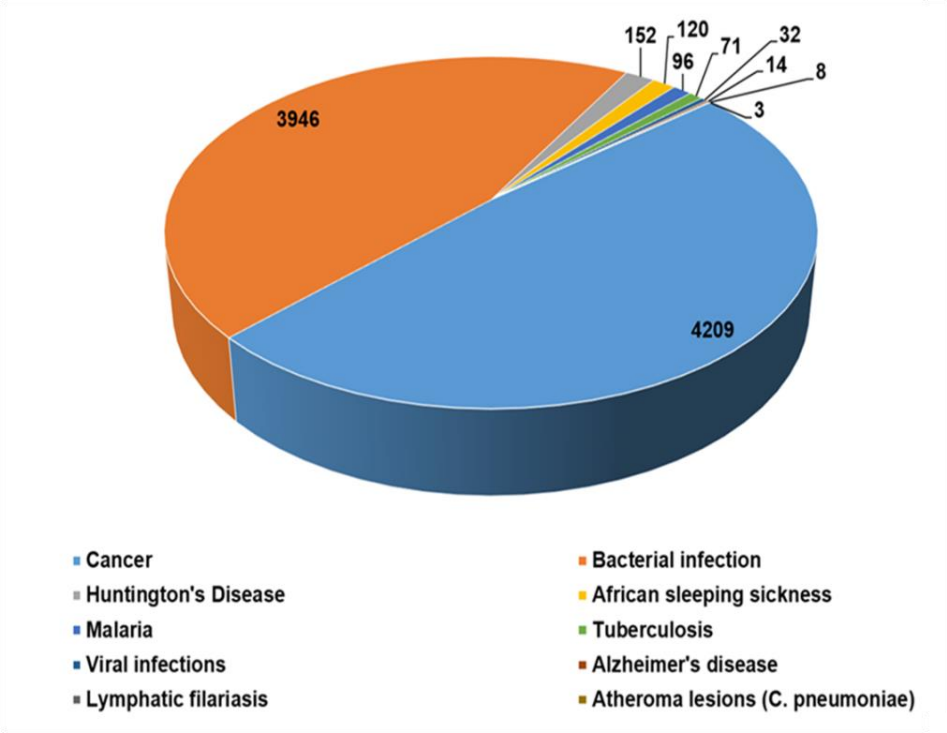

Figure 5.2.5 Total number of entries in HSPMdb based upon targeted diseases.

We found that most of these modulators are against cancer (4209 entries), followed by bacterial infections (3946entries). The total number of entries for Hsps' modulators against cancer, the most widely targeted disease, is 113,3053, 994 and

49 for Hsp100, Hsp90, Hsp70 and Hsp60,respectively, suggesting Hsp90 is the major target in cancer therapeutics. The studies mined for the design of current database reports activities of Hsps modulators either in the form of enzymatic activity and/or as cellular activity. The reported enzymatic assays are primarily *in vitro* with purified Hsps and provide information such as IC50, *Ki* and *K*d. The cellular-based activity assays are predominantly to examine the effect of modulator on activity of Hsps in a cell-based assay such as measurement of cell-based luminescence or cell growth using MTT (3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide)/ Alamar assay. Therefore, experimental data on both activities of Hsp modulators have been collected and reported in the current study. Almost equal entries of modulators for enzymatic (5244) and cellular-based activity assay (4985) have been observed. For enzymatic based activity, we have collected and reported all information about the modulators such as IC50, EC50, DC50, *Ki*, *K*d and percentage inhibition obtained from various functional assays. In total, information has been compiled from 26 different types of enzymatic assays. Our study shows that the substrate refolding assay is the most widely used assay followed by ATPase assay to examine the effect of molecules on Hsps enzymatic activity. Similarly, in the case of cellular activity, different cellular viability assays like MTT, Alamar blue and resazurin-based assays have been reported in the literature and, thus, we have collected data on such 15 different types of reported cellular assays. The database reports information from 140 different cell lines used for cell viability assay. The total number of entries of modulators found using cellular viability assay was observed to be 4985. For bacterial growth inhibition assay, 21 different bacterial species have been used resulting in 1594 entries of modulators against various Hsps. For some of the modulators (geldanamycin,MKT-077, MAL3- 101, 17-AAG, JG-98), multiple entries have been made as those were examined in multiple studies or tested against different Hsp types or validated by multiple functional/cellular assays. Hsps are multi-domain proteins, and interaction with other co-chaperones influences their activity. The modulation of Hsps' activity by various small molecules could be due to their interaction with different regions of the chaperone such as with substrate binding or nucleotide-binding pocket. In addition, many modulators obtained from previous studies have been reported to modulate the activity of Hsps by binding at the interface of the co-chaperone binding site. To enrich users with such information, we have collected and compiled information of binding site of these modulators on

their respective Hsps. We found that most of the modulators bind to the N-terminal domain (5222 entries) while a few (77 entries) were found to interact with the C-terminal domain of Hsps. The dominance of modulators binding to the N-terminal of Hsps suggests that the function of this domain is more sensitive to alteration by the small molecule binders. Hsp modulators compiled in HSPMdb belong to diverse classes or scaffolds. We observed that in the case of Hsp70 and Hsp90, most of the previous studies had explored the effect of different analogues of already existing modulators (such as of geldanamycin, resorcinol, radicicol, VER155008, YM-08, JG-98 and Apoptozole). For the Hsp100 and Hsp60 family of proteins, studies have primarily reported screening of various available commercial libraries of diverse compounds to identify molecules with modulatory activities. The present database thus provides comprehensive information of different classes/scaffolds of Hsp modulators from a large set of available studies in PubMed (Figure 5.2.6). The comprehensive information provided in the present study will facilitate the development of novel inhibitors or activators against various Hsps.

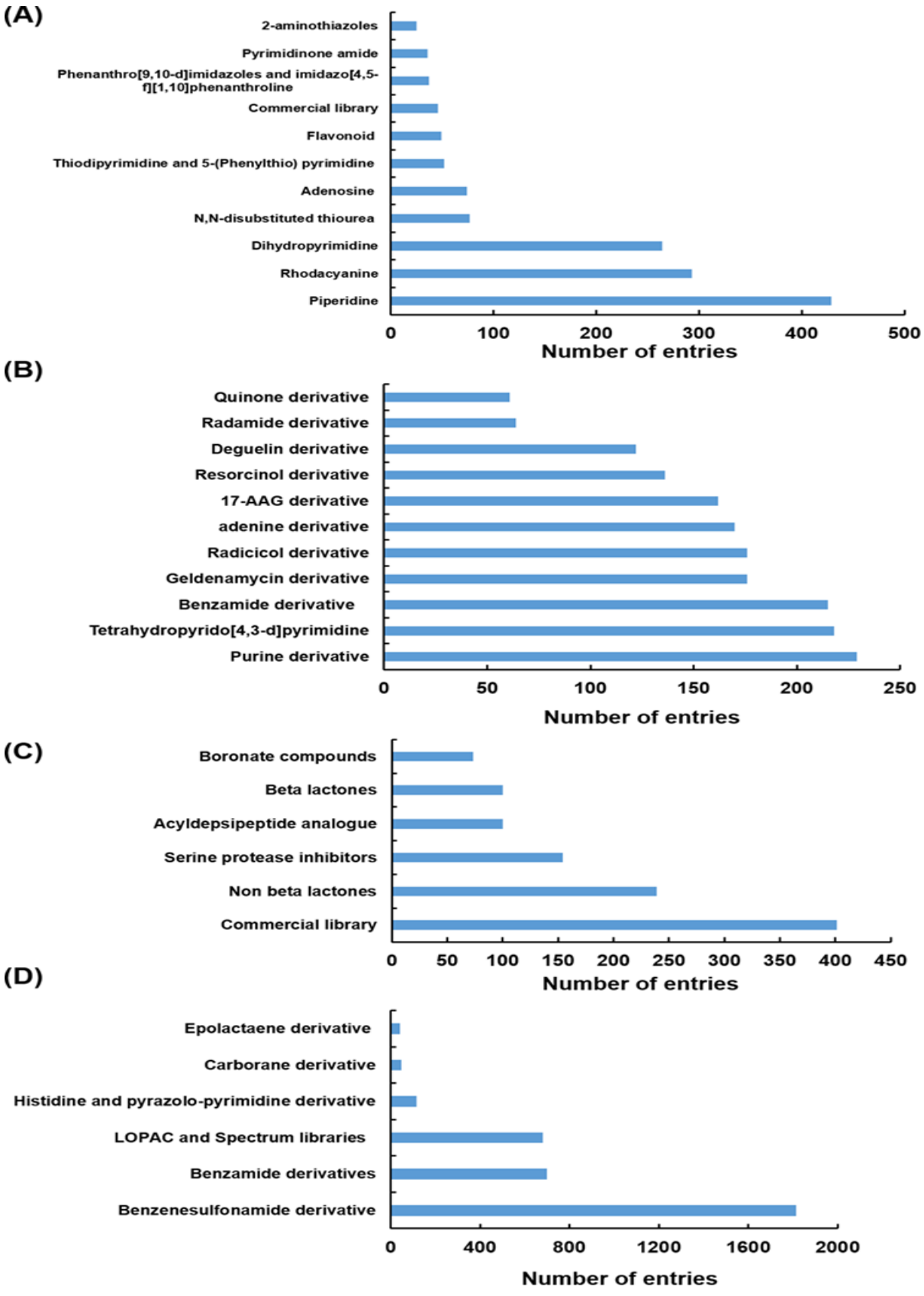

Figure 5.2.6 Different scaffolds/classes of modulators targeting Hsp70 (A), Hsp90 (B), Hsp100 (C) and Hsp60 (D).

5.3 Summary and future perspectives

HSPMdb will be very useful for a broader scientific community working in the area of chaperone biology and protein misfolding diseases in many ways: (i) the researcher can gather information of all Hsp modulators on a single platform which was not available until now, (ii) users could search HSPMdb for their newly designed molecules to examine whether similar scaffold or identical Hsp modulators have already been reported against any Hsps from 15 different organisms and (iii) HSPMdb provides a novel dataset of a large number of compounds targeting different Hsps which would be useful for developing novel algorithms for the prediction of Hsp modulators. This is the first version of HSPMdb where information of Hsp modulators has been compiled from available research articles. The only limitation is that information of Hsp modulators from patents have not been incorporated in this version of HSPMdb. Currently, we have compiled information of five major Hsps (Hsp100, Hsp90, Hsp70, Hsp60 andHsp40), and similar information from smallHsps needs to be provided. The information of small Hsp modulators as well as data from patents will be incorporated in the subsequent updated version of HSPMdb.

Availability

HSPMdb is available at http://bioinfo.imtech.res.in/bvs/hspmdb/index.php

5.4 HSPMdb field description

ID	All the Hsp modulators have been assigned a unique id number which is constant throughout the database.	HSPMDB0001
PMID	It is a PubMed identification number for further reference.	23961953
TARGET HSP	It provides the Class of Targeting Hsp	Hsp100
NAME OF HSP	It represents the isoform/member of respective Hsp class.	ClpB
ORIGIN	It represents the organism to which Hsp belongs.	Bacterial
LOCALIZATION	This field gives information about the sub-cellular localization of Hsp inside the cell.	Cytosol
LIBRARY	It represents the class/source of compound	Myria Screen Diversity Collection
NAME OF COMPOUND	This field gives information about the name of compound used for study	MKT-077
IUPAC NAMES	It provides the IUPAC nomenclature of the compound	(3Z)-3-[3-(4-Chlorophenyl)-4-oxo-2-thioxo-1,3-thiazolidin-5- ylidene]-5-nitro-1,3-dihydro-2H-indol-2-one
SMILES	It provides the template for generating structure of the compound	[O][N+](=O)c1ccc2NC(=O)\C(=C3/SC(=S)N(C3=O)c4ccc(Cl)cc4)c2c1
ENZYMATIC ACTIVITY	This field provides information for type of Hsp enzymatic activity	Inhibition of chaperone activity
TYPE OF ENZYMATIC ASSAY	This field contains the information about type of assay used for inhibition of enzymatic assay	Luciferase aggregate reactivation assay
ENZYME INHIBITION	It gives the percent inhibition of enzymatic activity.	90%
K_d	It provides the affinity constant value of	10 µM

	the binding assays.	
K_i	It provides the inhibition constant value of the assays.	2 µM
RD-INDEX	It provides the relative degradation value with respect to control reaction.	0.9
IC_{50}	It represents the concentration of an inhibitor where the response (or binding) is reduced by half	10 µM
CELLULAR ACTIVITY	This field provides the information for type of cellular activity.	Antibacterial activity
TYPE OF CELLULAR ASSAY	It gives the type of assay used for cellular inhibition	ATPase inhibition assay
CELL LINES /BACTERIA	Represents the cell lines used for in-vivo assays or organism for inhibition assays	HeLa/ S. aureus
% CELL GROWTH INHIBITION	It gives the percent inhibition of growth based cellular assay	50%
MIC/MBC	MIC represents is the lowest concentration of the drug, which prevents visible growth of bacterium	10 µM or 10 µg/ml
EC_{50}	It represents the concentration of a drug that gives half-maximal response.	5 µM
GI_{50}	It represents the concentration for 50% of maximal inhibition of cell proliferation.	2 µM
SITE OF INTERACTION	It tells about the site of binding/interaction of Drug with target Hsp	Nucleotide binding domain (NBD)
DISEASE	Represents the disease or condition involved with the target Hsp	Tuberculosis
HSP CLIENT PROTEIN ASSAY	It provides the information for Hsp client protein degradation assay.	Her-2 degradation assay

5.5 Assays descriptions for HSPMdb

1. ATPase assay:

ATPase assays are quantitative assays to measure the amount of ATP released from a biochemical reaction in terms of inorganic phosphate or formation of ADP. Released inorganic phosphate of ADP can be detected through various methods such as:

Absorbance based: Released inorganic phosphate can be detected calorimetrically by binding to malachite green solution which turns from orange to green in the presence of inorganic phosphate. The amount of inorganic phosphate is directly proportional to intensity of green colour measured by absorbance in 620-640 nm range.

Radioactivity based: Labelled ATP (P^{32}-γ) is used as substrate for a biochemical reaction. Hydrolysis of ATP (P^{32}-γ) yields inorganic labelled (P^{32}-γ) with which can be measured directly can be measured by scintillation counter for radioactivity counts.

Fluorescence based: This method measures fluorescence emitted from a biochemical reaction upon binding of Inorganic phosphate of ADP to fluorogenic substrate.

2. Substate refolding assays:

Refolding assays are the functional assays to calculate the activity of a chaperone protein. Firefly Luciferase, Malate dehydrogenase, Lactate dehydrogenase Glycerol-6-phosphate dehydrogenase e.t.c. are used as modelled substrates for studying denaturation and refolding. Thermal denaturing of Firefly Luciferase results in unfolded or misfolded intermediates which mimics as denatured substrate produced inside the cell upon to heat stress, while the chemically denaturation results in mimicking as nascent unfolded intermediate substrate. Firefly Luciferase refolding required principle chaperone (HSP100, Hsp70, Hsp60) along with their co-chaperones (Hsp40,NEF), ATP and presence of Mg^{2+} ions.

Luciferase in its native form emits luminescence which decreases upon heat or chemical denaturation. Chaperones will bind to denatured or unfolded Luciferase

and performs the refolding cycle in the presence of ATP. Refolding activity can be measures directly as increasing in luminescence over a period of time.

3. Surface Plasmon resonance (SPR):

SPR is a conventional technique to check the protein-protein interactions along with kinetics and equilibrium parameters quantitatively in real time manner. Interaction is measured by immobilizing the desired protein on sensor chip surface by various coupling methods to a certain response units followed by injecting the analyte over the surface of immobilized protein at constant flow rate. Binding of proteins will results in increasing in mass due to binding thereby changes the refractive index at the surface of sensor chip. These changes are measured as change in resonance angle ($\Delta\Theta$). The SPR signal can be quantified as resonance units (RU) and represented as change in resonance angle with respect to critical angle. The final output of interaction is depicted as a sensongram showing association and dissociation phase. The affinity constant (KD), K on rates and K off rates can be calculated according to various binding models.

4. Isothermal titration calorimetry (ITC):

ITC is Biophysical technique use to measure biomolecular interactions and thermodynamical parameters in solution by measuring change in heat of the reaction. Analyte will be injected through syringe into micromacorimeter cell containing protein. Heat of binding will be measured by cells in adiabatic chamber and represented as change in heat of binding per unit time. Binding affinity (Ka), stoichiometry (n) can be calculated from isotherm curves. Thermodynamic parameters such as Enthalpy change (ΔH), Gibbs free energy changes (ΔG) and Entropy changes (ΔS) can be determined.

5. Fluorescence polarization assay (FP):

This technique is used for measuring protein-protein interactions and enzyme activities in real time. Assay works on the principle that plane polarized light strikes the protein sample in the solution, exciting the fluorophore bounded to protein sample. The excited fluorophore results in emitting fluorescence back to the detector. The continuous movement of fluorescing sample in solution results light in

polarized form. When there is interaction between two proteins/biomolecules, increase in mass leads to slow movement of the complex in solution results in increase the polarization time resulting in increased fluorescence on detector. The fluorescence emission is detected through detector and plotted as a function of time.

Summary

Summary

Tuberculosis (TB) remains one of leading causes of the deaths in humans worldwide. Approximately one third of the world's population is affected out of which 10% of population will eventually develop active TB. *Mycobacterium tuberculosis* (*M. tuberculosis*), the causative agent of TB, is an extremely successful pathogen with multifactorial ability to control the host immune response. *M. tuberculosis* survives mainly due to emergence of drug resistance and stress tolerance mechanism. How *M. tuberculosis* survives a prolonged dormant phase before its reactivation, as the host immune system is compromised, remains an unsolved aspect of *M. tuberculosis* biology. Various studies are now focused towards understanding of mechanism of stress tolerance for *M. tuberculosis* survival in host environment during the infection. Studies show that among various cellular factors, chaperones play a key role in overcoming the unfavourable or stressed conditions. Chaperone maintains the protein quality control system by modulating the balance between refolding and degradation of misfolded substrates and ensures the pathogen's survival.

$ClpB_M$, a major disaggregase, protects the mycobacterium from various stresses encountered in the host environment. $ClpB_M$ works in collaboration with $DnaK_M$ system to ensure efficient protein refolding in the mycobacterium. In the present study we have performed a detailed biophysical, biochemical, and structural characterization of *M. tuberculosis* $ClpB_M$ machinery and provide insights for targeting the disaggregase $ClpB_M$ by a high throughput screening for identifying small molecule inhibitors. Circular dichroism spectroscopy results showed the dominant helical nature of $ClpB_M$ in its native state. Thermal denaturation studies reveal that $ClpB_M$ unfolds in a biphasic manner populating three states with a stable intermediate. The size exclusion chromatography studies show that the presence of nucleotides significantly affects $ClpB_M$ oligomerization. Analytical ultracentrifugation sedimentation velocity studies reveal that $ClpB_M$ oligomerization varies with its concentration and is dependent upon its surrounding milieu such as salt concentration and nucleotides. Further, using high throughput malachite green assay, we screened small molecule commercial libraries and identified 3 potential novel inhibitors of ClpB ATPase activity with 70-90% inhibition. The enzyme kinetics revealed that the lead molecule inhibits $ClpB_M$ activity in a competitive manner. These drugs were also able to inhibit ATPase activity associated with *E. coli* ClpB

and yeast Hsp104. The identified drugs inhibited the growth of intracellular bacteria in macrophages. Small angle X-ray scattering based modelling shows that ATP, and not its non-hydrolyzable analogs induce large scale conformational rearrangements in $ClpB_M$. surprisingly, the identified small molecules inhibited these ATP inducible conformational changes, suggesting that nucleotide induced shape changes are crucial for $ClpB_M$ activity.

We also biophysically characterized the *M. tuberculosis* Dnak system which collaborates with disaggregase $ClpB_M$ for efficient refolding of substrate. $DnaK_M$ works along with its co-chaperone, $DnaJ1_M$ and nucleotide exchange factor $GrpE_M$. Circular dichroism spectroscopy studies showed the native folded state of $DnaK_M$ and $DnaJ1_M$ as mixture of α-helices and β-sheets, while $GrpE_M$ is exclusively α-helix.

Thermal denaturation studies show that thermal unfolding of $DnaK_M$ or $DnaJ1_M$ at pH 7.4 is irreversible, whereas $GrpE_M$ unfolds in a reversible two-state manner. Gel filtration analysis showed the insignificant effect of nucleotides (ATP/ADP) on $DnaK_M$ oligomerization. Further, Analytical ultracentrifugation sedimentation velocity studies reveal the monomeric form of $DnaK_M$, while the other proteins of chaperone machinery, $DnaJ1_M$ and $GrpE_M$ are primarily dimeric in solution. $DnaK_M$ interaction with $GrpE_M$ was confirmed by gel filtration chromatography by appearance of additional Peak eluting at higher molecular weight than $DnaK_M$ and $GrpE_M$ alone. Further, the $DnaK_M$-$GrpE_M$ complex was also confirmed onto Native Page where the band corresponding to complex migrating at higher size as compared to individual proteins were identified. The affinity of interaction assessed by Surface plasmon resonance was found to be 47.6 nM. The release of MABA-ADP from $DnaK_M$.MABA-ADP complex by $GrpE_M$ with a fluorescence change of 2.5 folds shows the nucleotide exchange activity of $GrpE_M$. The study thus provides insight into biophysical and biochemical aspects of *M. tuberculosis* Hsp70 refolding machinery, and allow us to design inhibitors against essential activity of $GrpE_M$.

In another study, we found that Heat shock proteins (Hsps) are among highly conserved proteins across all domains of life. These proteins are also involved in a wide range of cellular functions such as protein refolding, protein trafficking and cellular signalling. A large number of potential Hsp modulators are under clinical trials against various human diseases. As the number of modulators targeting Hsps

is growing, there is a need to develop a comprehensive knowledge repository of these findings which is largely scattered. Therefore, we developed a web-accessible database, HSPMdb, which is a first of its kind manually curated repository of experimentally validated Hsp modulators (activators and inhibitors). The data was collected from 176 research articles and current version of HSPMdb holds 10,223 entries of compounds that are known to modulate activities of five major Hsps (Hsp100, Hsp90, Hsp70, Hsp60 and Hsp40) originated from 15 different organisms (i.e. human, yeast, bacteria, virus, mouse, rat, bovine, porcine, canine, chicken, *Trypanosoma brucei* and *Plasmodium falciparum*). HSPMdb provides comprehensive information on biological activities as well as the chemical properties of Hsp modulators. The biological activities of modulators are presented as enzymatic activity and cellular activity. Under the enzymatic activity field, parameters such as IC50, EC50, DC50, *Ki* and *K*D have been provided. In the cellular activity field, complete information on cellular activities (percentage cell growth inhibition, EC50 and GI50), type of cell viability assays and cell line used has been provided. One of the important features of HSPMdb is that it allows users to screen whether or not their compound of interest has any similarity with the previously known Hsp modulators. HSPMdb is a valuable resource for the broader scientific community working in the area of chaperone biology and protein misfolding diseases.

Overall, this comprehensive study broadens our understanding of *M. tuberculosis* bi-chaperone machinery and provides the basis for designing more potent inhibitors for combating TB.

Appendix-I

Cloning and purification of chaperones

Appendix-I

Construction and confirmation of the plasmids encoding ClpB, DnaK, DnaJ1 and GrpE used in this study is described briefly in section A1.1 and A1.2. Similarly, detailed purification methods are described in section A1.3.

A1.1 Cloning of *M. tuberculosis* ClpB.

The gene encoding ClpB gene was cloned at NdeI and HindIII restriction sites of plasmid pET22b. Figure A1 shows confirmation of ClpB gene cloned in pET22b vector.

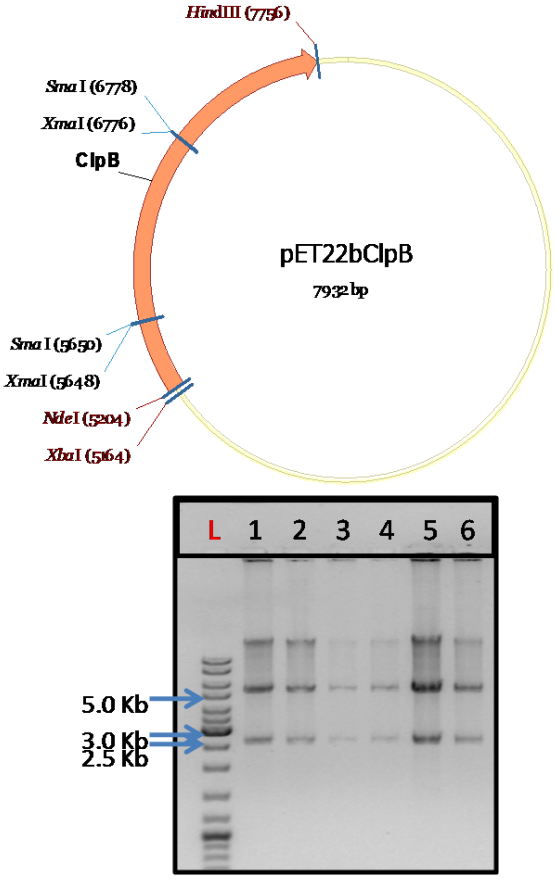

Figure A1. Shows the vector map of plasmid pET22bClpB. Restriction digestion of plasmid pET22bClpB with enzyme XbaI and HindIII. The insert size of 2.6 kb corresponds to gene encoding ClpB.

A1.2 Cloning of *M. tuberculosis* Hsp70 components DnaK, DnaJ1 and GrpE.

All the genes were amplified from Mycobacterium tuberculosis H37Rv genomic DNA. The gene encoding DnaK was cloned at BamHI and XhoI restriction site of plasmid pET29bHtv. DnaJ1 was cloned in pESUMO vector at restriction sites BamHI and

XhoI. GrpE was cloned at BamHI and XhoI restriction site of plasmid pET29bHtv by Overlap PCR. Figure A2 shows the clone confirmation of all the genes. All the clones were screened by different sets of restriction enzyme and further confirmed by sequencing.

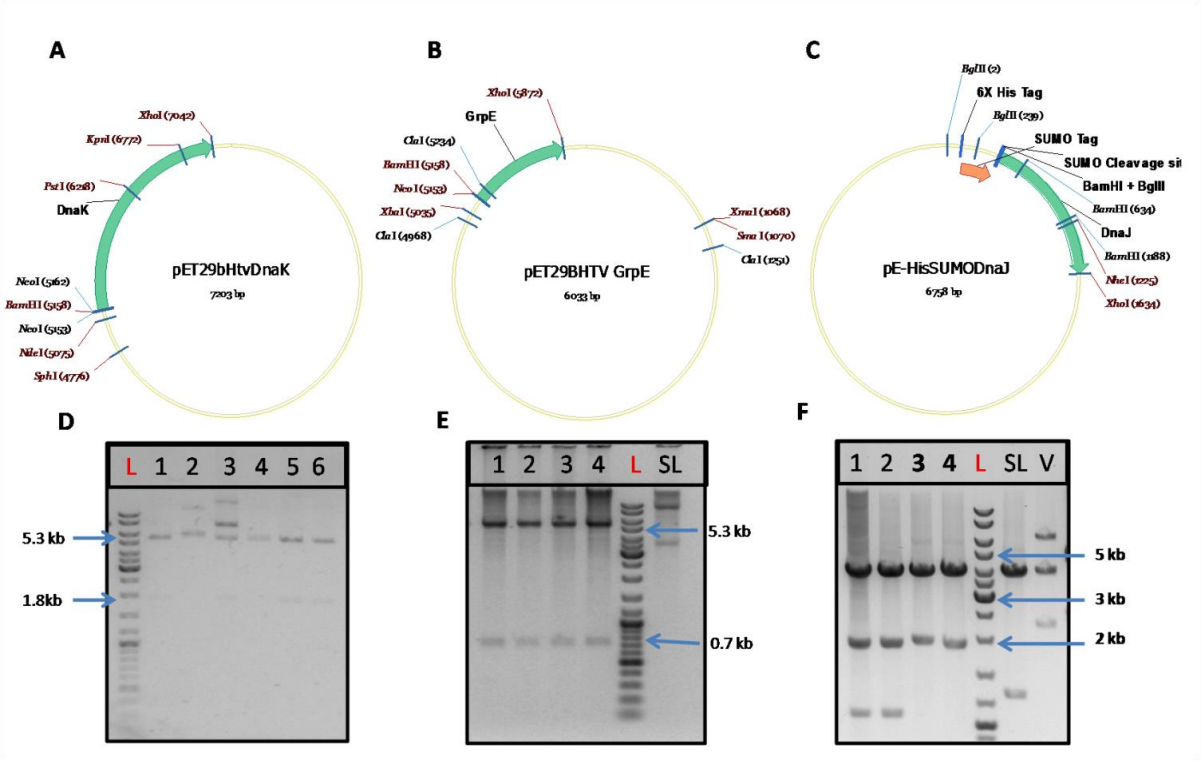

Figure A2. Cloning of *M. tuberculosis* DnaK, GrpE and DnaJ1. (A) and **(B)** shows the vector map of plasmid pET29Htv containing genes encoding for DnaK or GrpE. **(C)** shows the vector map of plasmid pESUMODnaJ1. **(D)** Restriction digestion of plasmid pET29HtvDnaK with enzyme BamHI and XhoI. The insert size of 1.8 Kb corresponds to gene *dnaK*. **(E)** Restriction digestion of plasmid pET29HtvGrpE with enzyme BamHI and XhoI, releasing an insert size of 0.7Kb corresponds to gene GrpE. **(F)** Restriction digestion of plasmid pEHisSUMODnaJ1 with enzyme NheI and ClaI gives insert size of 1.9 Kb confirms the cloning of gene DnaJ1.

A1.3 Protein expression and purification of chaperones.

All the chaperones were purified as mentioned in materials and methods. The protein purity for all the proteins were examined on SDS PAGE. Protein concentrations was measured by BCA kit (Thermo Pierce) and Absorbance at 280nm. Figure A3.1 shows the purification of *M. tuberculosis* ClpB. Figure A3.2 shows purification of DnaK. Figure A3.3 shows purification of GrpE. Figure A3.4 shows purification of DnaJ1.

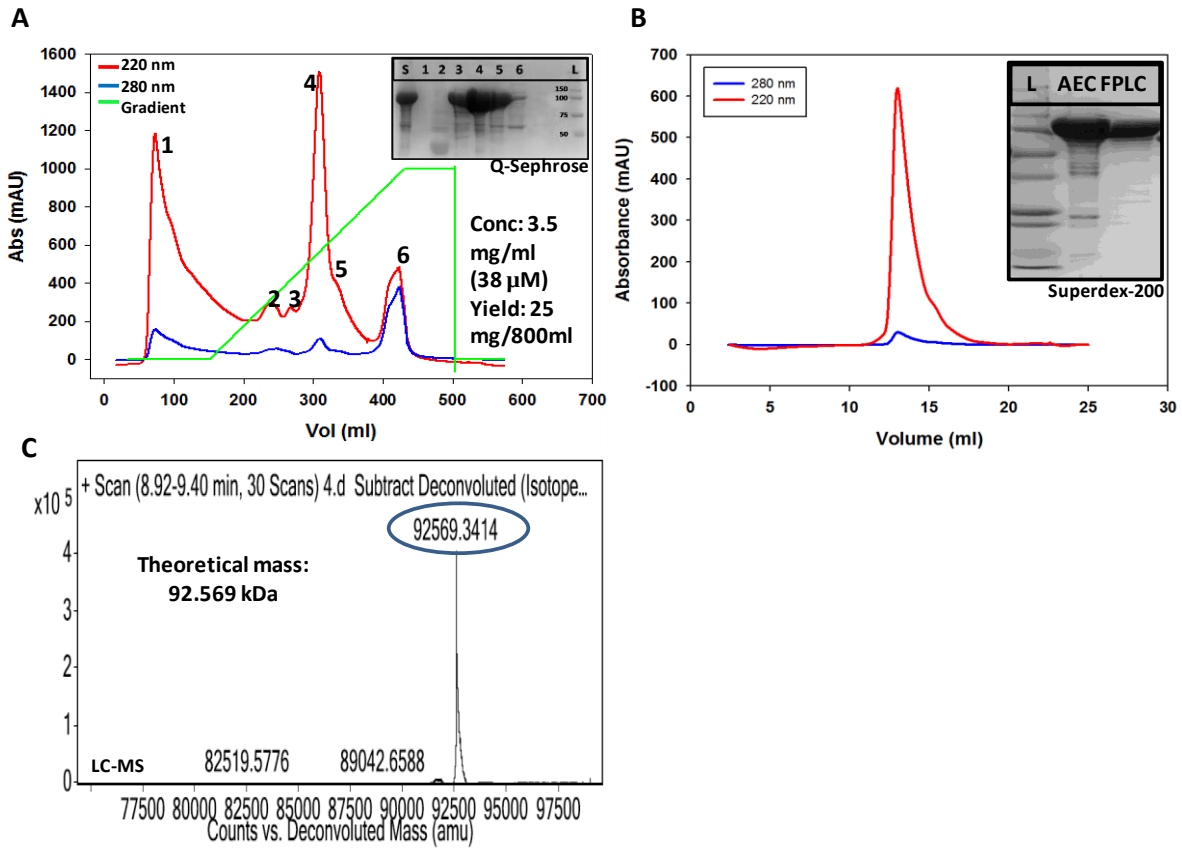

Figure A3.1 Purification of ClpB. (A) shows the chromatogram for anion exchange chromatography, inset shows the peaks containing ClpB fraction. **(B)** Shows the chromatogram for gel filtration chromatography, inset shows the fraction containing purified ClpB after gel filtration chromatography. **(C)** LC-MS of purified protein confirms the identity of ClpB by observed molecular mass of 92.569 kDa.

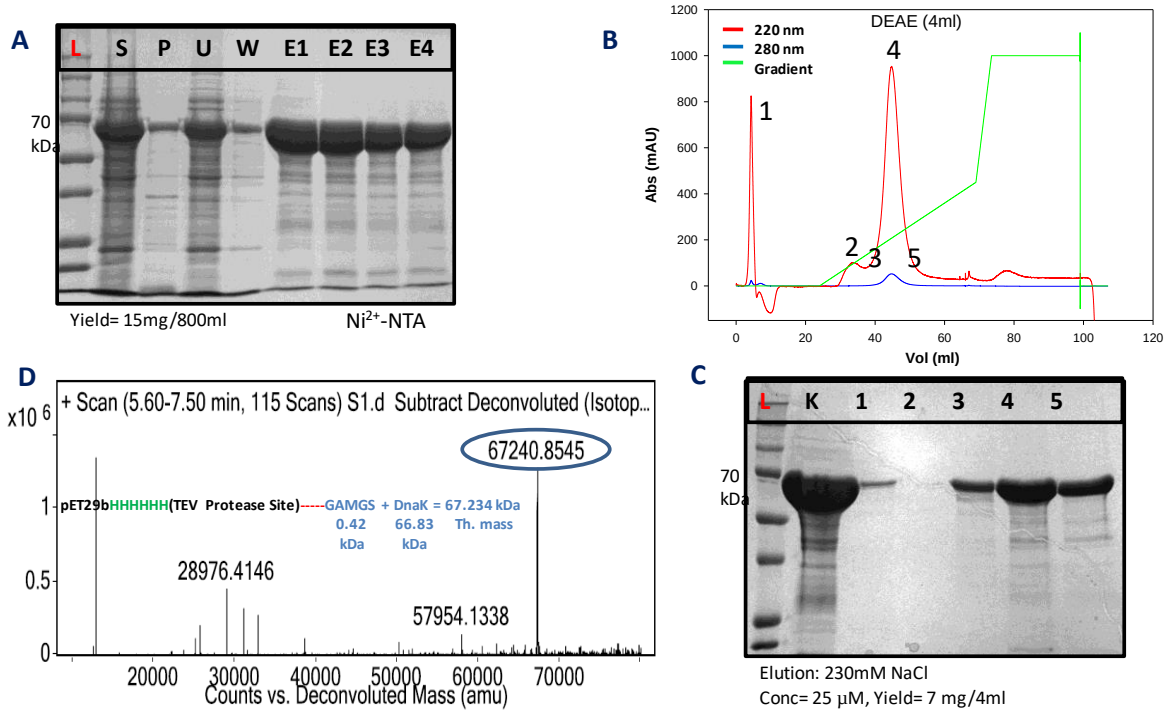

Figure A3.2. Purification of DnaK. (A) Shows the Affinity chromatography of DnaK purified using Ni^{2+} -NTA resin. Loaded fractions are Ladder (L), Supernatant (S), Pellet (P), Unbound (U), Wash (W) and Elutions (E1-E4). **(B)** Elution profile on DEAE column. Elution fractions from Ni^{2+}-NTA were loaded on DEAE column and DnaK was eluted at 230mM NaCl. **(C)** Peaks from DEAE anion exchange colum were collected and confirmed on 10 % SDS-PAGE. **(D)** LC-MS of purified protein. Identity of DnaK was confirmed by observed molecular mass of 67.240 kDa in LC-MS with a theoretical molecular mass of 67.234 kDa.

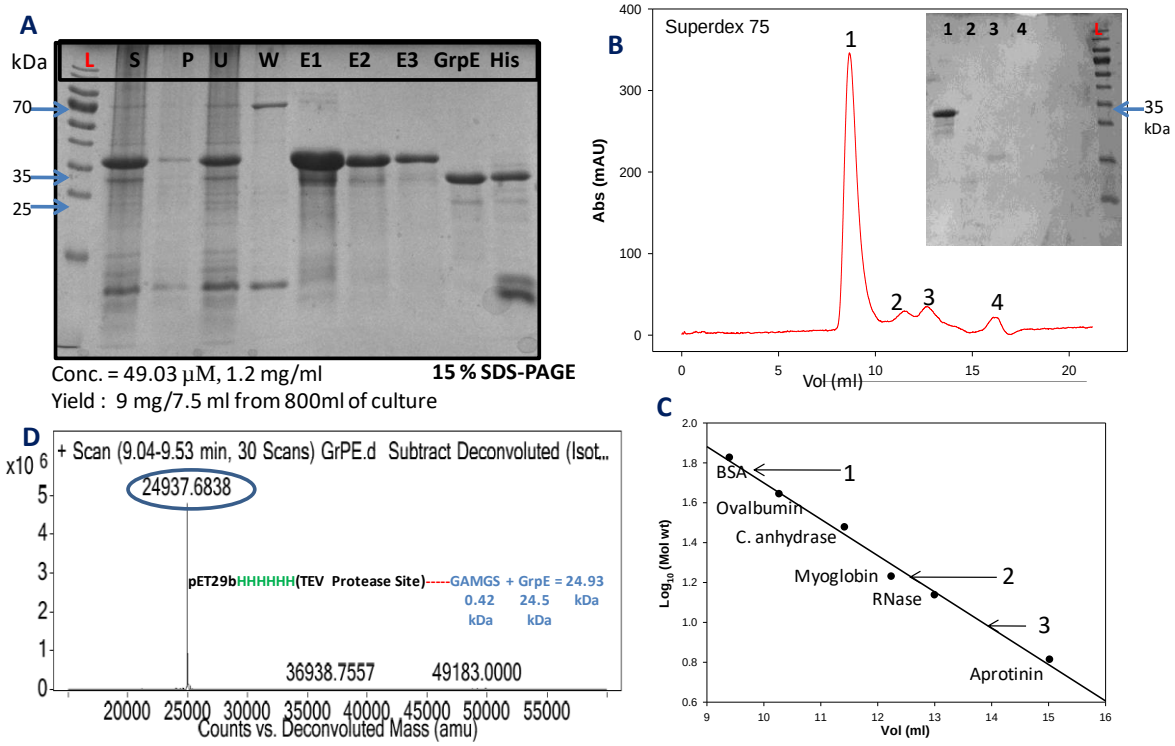

Figure A3.3 Purification of GrpE. (A) shows the affinity chromatography of GrpE. The protein was purified using Ni^{2+}-NTA resin. Loaded fractions are Ladder (L), Supernatant (S), Pellet (P), Unbound (U), Wash (W) and Elutions (E1-E3). His tag was removed with TEV protease digestion yielding a pure protein marked as GrpE. **(B)** Gel filtration chromatogram for GrpE. Elution fractions fromNi^{2+}-NTA were loaded on Superdex 75 gel filtration column **(C)** Protein standards for Superdex 75 plotted against Log_{10} Molecular weight Vs Elution Volume. Aprotinin (6.5 kDa), Rnase (13.7 kDa), Myoglobin (17 kDa), Carbonic anhydrase (30 kDa), Ovalbumin (44 kDa), BSA (66 kDa) were used as standards. **(D)** LC-MS of purified protein. Identity of GrpE was confirmed by observed molecular mass of 24.937 kDa in LC-MS with a theoretical molecular mass of 24.935 kDa.

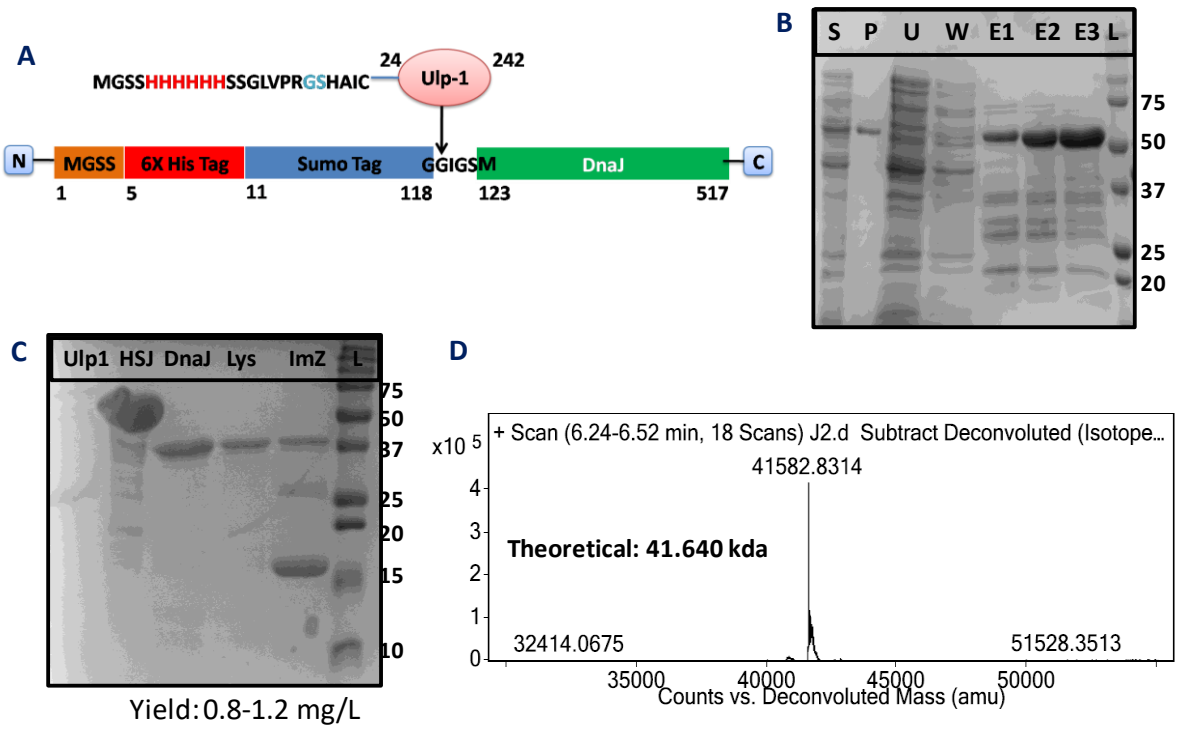

Figure A3.4. Purification of DnaJ1. (A) Schematic diagram showing DnaJ construct in pESUMO vector along with His6-SUMO cleaving site. **(B)** Affinity chromatography for DnaJ using Co^{2+}-NTA resin. Loaded fractions are Supernatant (S), Pellet (P), Unbound (U), Wash (W) and Elutions (E1-E3) and (L) Ladder. **(C)** Ulp-I protease reaction shows His6-SUMO tag cleaved from DnaJ. Ulp-1 represents the lane containing 29 kDa Ulp-1 protease. HSJ represents His6X-Sumo tagged DnaJ, 55 kDa followed by lane representing tag cleaved DnaJ with molecular mass of 42 kDa. Imz represents lane containing cleaved His6X sumo tag from DnaJ and Ulp-1 in the reaction. **(D)** LC-MS of purified protein. Identity of DnaJ was confirmed by observed molecular mass of 41.58 kDa in LC-MS with a theoretical molecular mass of 41.64 kDa.

Appendix-II

Oligonucleotide primers and plasmid vectors

Appendix-II

Biophysical, biochemical and structural studies on ClpB, DnaK, DnaJ1 and GrpE from *M. tuberculosis* described in the previous chapters were assisted by the molecular biology techniques in generating vectors for the complementation studies and for the purification of proteins. Generation and confirmation of the mentioned clones involved several oligonucleotide primers and appropriate vectors. The sequences of oligonucleotide primers used for cloning and sequencing are listed in the section A2.1 Similarly, Plasmid vectors used for cloning and the plasmids vectors generated in this study are listed in the section A2.2

A2.1 Oligonucleotide Primers Used for Cloning in this Study.

Primer	**Sequence (5'-3' amplification)**
T7_Forward	TAATACGACTCACTATAGGG
T7_Reverse	TAGTTATTGCTCAGCGGTGG
	***M. tuberculosis* chaperones**
ClpB_Fwd_BamHI	GTAGTAGGATCCATGGACTCGTTTAACCCGACG
ClpB_Rev_HindIII	GTCGTTAAGCTTTCAGCCCAGGATCAGCGAG
ClpB_FragA_Fwd	GTGCTGATCGGTGAGCCCG
ClpB_Seq_Fwd_597	GAACAACCCGGTGCTGATCG
ClpB_Seq_Rev_700	CGCGCAAGCTCTCCGGCAC
ClpB_Seq_Fwd_1378	CGTATCCCACATAGCCGGGC
ClpB_Seq_Fwd_1889	GATGGTCCGCATCGACATG
Dnak_Fwd_BamH1	GTAGTAGGATCCATGGCTCGTGCGGTCGG
DnaK_Rev_Xho1	GTCGTTCTCGAGTCACTTGGCCTCCCGGC
DnaK_Seq_R465	CTCCTTCTCGCCCTTGTC

DnaK_Seq_ F603	GACGACTGGGACCAGCGG
DnaK_Seq_F1293	GACCTGACCGGCATCCCG
DnaJ_FBAM_Sumo	GTAGTAGGATCC ATG GCT CAA AGG GAA TGG GTC
DnaJ_FBAM_Sumo	GTAGTAGGATCC ATG GCT CAA AGG GAA TGG GTC
DnaJ_Seq_R449	CAACTCGGTCTCGGTCTCC
DnaJ_Seq_F596	GGGTCGGGCGTGATCAACC
GrpE_Fwd_BamHI	GTAGTAGGATCCATGACGGACGGAAATCAAAAGC
GrpE_Rev_XhoI	GTCGTTCTCGAGTTAACTGCCCGACGGTTCTG
GrpE_Ovr_Fwd	GTGTACTGGACGATCTGGAGCGGGCGCGCAAGCACG
GrpE_Ovr_Rev	CGTGCTTGCGCGCCCGCTCCAGATCGTCCAGTACAC

E. coli chaperones

EC_ClpB_F_NdeI	GTAGTACATATGCGTCTGGATCGTCTTACTAAT
EC_ClpB_R_XhoI	GTAGTACTCGAGTTACTGGACGGCGACAATCCGGTC
EC_clpB_Seq_F513	GCCGAACAGGGCAAACTCGAT
EC_clpB_Seq_F1150	CGTCAGCTGCCGGATAAAG
EC_clpB_Seq_F1650	CATGATGGAAAGCGAGCGCGA
EC_ClpB_Seq_R2193	GCTCTTTCATGTGCGCATAATC
EC_DnaK_F_BamHI	GTAGTAGGATCCATGGGTAAAATAATTGGTATCG
EC_DnaK_R_XhoI	GTAGTACTCGAGTTATTTTTTGTCTTTGACTTCTTC
EC_DnaK_F900	ACTCGTGCGAAACTGGAA
E_DnaK_R1018	CGAGGATAACGTCGTCGA

EC_DnaJ_F_BamHI	GTAGTAGGATCCATGGCTAAGCAAGATTATTAC
EC_DnaJ_R_XhoI	GTCGTTCTCGAGTTAGCGGGTCAGGTC
EC_DnaJ_F_614	GTGTTGAGCGCAGCAAAA
EC_GrpE_F_BglII	GTAGTAAGATCTATGAGTAGTAAAGAACAGAA
EC_GrpE_R_XhoI	GTAGTACTCGAGTTAAGCTTTTGCTTTCGCTA

EC: refers to E. coli

A2.2 Plasmid vectors used in this study.

Name	Description	Source
pET20b	T7 based expression vector for expression in *E. coli Rossetta DE3*	Novagen Inc., USA
pET29bHtv	T7 based expression vector for expression in *E. coli Rossetta DE3*	(Kumar et al., 2014)
pET21a	T7 based expression vector for expression in *E. coli Rossetta DE3*	Novagen Inc., USA

Bibliography

Bibliography

Abrams, J.L., Verghese, J., Gibney, P.A., and Morano, K.A. (2014). Hierarchical functional specificity of cytosolic heat shock protein 70 (Hsp70) nucleotide exchange factors in yeast. J Biol Chem *289*, 13155-13167.

Agarwal, M., Katiyar-Agarwal, S., Sahi, C., Gallie, D.R., and Grover, A. (2001). Arabidopsis thaliana Hsp100 proteins: kith and kin. Cell Stress Chaperones *6*, 219-224.

Ahmed, N., Mohanty, A.K., Mukhopadhyay, U., Batish, V.K., and Grover, S. (1998). PCR-based rapid detection of Mycobacterium tuberculosis in blood from immunocompetent patients with pulmonary tuberculosis. J Clin Microbiol *36*, 3094-3095.

Akoev, V., Gogol, E.P., Barnett, M.E., and Zolkiewski, M. (2004). Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB. Protein science : a publication of the Protein Society *13*, 567-574.

Anfinsen, C.B. (1973). Principles that Govern the Folding of Protein Chains. Science (New York, NY) *181*, 223.

Arend, S.M., van Meijgaarden, K.E., de Boer, K., de Palou, E.C., van Soolingen, D., Ottenhoff, T.H., and van Dissel, J.T. (2002). Tuberculin skin testing and in vitro T cell responses to ESAT-6 and culture filtrate protein 10 after infection with Mycobacterium marinum or M. kansasii. The Journal of infectious diseases *186*, 1797-1807.

Arora, G., Gagandeep, Behura, A., Gosain, T.P., Shaliwal, R.P., Kidwai, S., Singh, P., Kandi, S.K., Dhiman, R., Rawat, D.S.*, et al.* (2019). NSC 18725, a Pyrazole Derivative Inhibits Growth of Intracellular Mycobacterium tuberculosis by Induction of Autophagy. Front Microbiol *10*, 3051.

Banecki, B., Liberek, K., Wall, D., Wawrzynów, A., Georgopoulos, C., Bertoli, E., Tanfani, F., and Zylicz, M. (1996). Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. J Biol Chem *271*, 14840-14848.

Barnett, M.E., Zolkiewska, A., and Zolkiewski, M. (2000). Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains. The Journal of biological chemistry *275*, 37565-37571.

Barreiro, C., González-Lavado, E., Brand, S., Tauch, A., and Martín, J.F. (2005). Heat shock proteome analysis of wild-type Corynebacterium glutamicum ATCC 13032 and a spontaneous mutant lacking GroEL1, a dispensable chaperone. Journal of bacteriology *187*, 884-889.

Benaroudj, N., Triniolles, F., and Ladjimi, M.M. (1996). Effect of nucleotides, peptides, and unfolded proteins on the self-association of the molecular chaperone HSC70. J Biol Chem *271*, 18471-18476.

Bertelsen, E.B., Chang, L., Gestwicki, J.E., and Zuiderweg, E.R. (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proceedings of the National Academy of Sciences of the United States of America *106*, 8471-8476.

Beuron, F., Maurizi, M.R., Belnap, D.M., Kocsis, E., Booy, F.P., Kessel, M., and Steven, A.C. (1998). At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. Journal of structural biology *123*, 248-259.

Bhandari, V., and Houry, W. (2015). Substrate Interaction Networks of the Escherichia coli Chaperones: Trigger Factor, DnaK and GroEL. In, pp. 271-294.

Bienfait, B., and Ertl, P. (2013). JSME: a free molecule editor in JavaScript. Journal of Cheminformatics *5*, 24.

Bieri, O., Wildegger, G., Bachmann, A., Wagner, C., and Kiefhaber, T. (1999). A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. Biochemistry *38*, 12460-12470.

Bishai, W. (2000). Lipid lunch for persistent pathogen. Nature *406*, 683-685.

Blatch, G., and Edkins, A. (2015). The Networking of Chaperones by Co-chaperones: Control of Cellular Protein Homeostasis, Vol 78.

Bloom, B.R., and Murray, C.J. (1992). Tuberculosis: commentary on a reemergent killer. Science (New York, NY) *257*, 1055-1064.

Borges, J.C., Seraphim, T.V., Dores-Silva, P.R., and Barbosa, L.R.S. (2016). A review of multi-domain and flexible molecular chaperones studies by small-angle X-ray scattering. Biophysical reviews *8*, 107-120.

Borkovich, K.A., Farrelly, F.W., Finkelstein, D.B., Taulien, J., and Lindquist, S. (1989). hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Molecular and Cellular Biology *9*, 3919.

Boston, R.S., Viitanen, P.V., and Vierling, E. (1996). Molecular chaperones and protein folding in plants. Plant Molecular Biology *32*, 191-222.

Braakman, I., Hoover-Litty, H., Wagner, K.R., and Helenius, A. (1991). Folding of influenza hemagglutinin in the endoplasmic reticulum. The Journal of cell biology *114*, 401-411.

Bracher, A., and Verghese, J. (2015a). The nucleotide exchange factors of Hsp70 molecular chaperones. Front Mol Biosci *2*, 10-10.

Bracher, A., and Verghese, J. (2015b). The nucleotide exchange factors of Hsp70 molecular chaperones. Frontiers in Molecular Biosciences *2*.

Brehmer, D., Gässler, C., Rist, W., Mayer, M.P., and Bukau, B. (2004). Influence of GrpE on DnaK-substrate interactions. J Biol Chem *279*, 27957-27964.

Brennan, P.J., and Nikaido, H. (1995). THE ENVELOPE OF MYCOBACTERIA. Annual Review of Biochemistry *64*, 29-63.

Brockwell, D.J., Smith, D.A., and Radford, S.E. (2000). Protein folding mechanisms: new methods and emerging ideas. Current opinion in structural biology *10*, 16-25.

Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Curr Opin Struct Biol *58*, 197-213.

Buchberger, A., Schröder, H., Hesterkamp, T., Schönfeld, H.J., and Bukau, B. (1996). Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. Journal of molecular biology *261*, 328-333.

Buchner, J. (1999). Hsp90 & Co. - a holding for folding. Trends in biochemical sciences *24*, 136-141.

Bukau, B., and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell *92*, 351-366.

Calloni, G., Chen, T., Schermann, Sonya M., Chang, H.-c., Genevaux, P., Agostini, F., Tartaglia, Gian G., Hayer-Hartl, M., and Hartl, F.U. (2012). DnaK Functions as a Central Hub in the E. coli Chaperone Network. Cell Reports *1*, 251-264.

CDC (Center for Disease control and Prevention, U.S. Department of Health & Human Services, https://www.cdc.gov/tb/topic/treatment/default.htm). In Center for Disease control and Prevention.

Chatterjee, S., and Burns, T.F. (2017). Targeting Heat Shock Proteins in Cancer: A Promising Therapeutic Approach. Int J Mol Sci *18*, 1978.

Chaudhuri, T.K., Farr, G.W., Fenton, W.A., Rospert, S., and Horwich, A.L. (2001). GroEL/GroES-mediated folding of a protein too large to be encapsulated. Cell *107*, 235-246.

Chen, B., Zhong, D., and Monteiro, A. (2006). Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. BMC Genomics *7*, 156.

Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L., and Horwich, A.L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature *337*, 620-625.

Chernoff, Y.O., Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G., and Liebman, S.W. (1995). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science (New York, NY) *268*, 880-884.

Chesnokova, L.S., Slepenkov, S.V., Protasevich, I.I., Sehorn, M.G., Brouillette, C.G., and Witt, S.N. (2003). Deletion of DnaK's Lid Strengthens Binding to the Nucleotide Exchange Factor, GrpE: A Kinetic and Thermodynamic Analysis. Biochemistry *42*, 9028-9040.

Clark, P.L. (2004). Protein folding in the cell: reshaping the folding funnel. Trends in biochemical sciences *29*, 527-534.

Cockburn, I.L., Pesce, E.R., Pryzborski, J.M., Davies-Coleman, M.T., Clark, P.G., Keyzers, R.A., Stephens, L.L., and Blatch, G.L. (2011). Screening for small molecule modulators of Hsp70 chaperone activity using protein aggregation suppression assays: inhibition of the plasmodial chaperone PfHsp70-1. Biological chemistry *392*, 431-438.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd*, et al.* (1998a). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature *393*, 537-544.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E.*, et al.* (1998b). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature *393*, 537-544.

Craig, E.A., Gambill, B.D., and Nelson, R.J. (1993). Heat shock proteins: molecular chaperones of protein biogenesis. Microbiological reviews *57*, 402-414.

Creighton, T.E. (1997). How important is the molten globule for correct protein folding? Trends in biochemical sciences *22*, 6-10.

Daniels, M., and Hill, A.B. (1952). Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council trials. British medical journal *1*, 1162-1168.

Das Gupta, T., Bandyopadhyay, B., and Das Gupta, S.K. (2008). Modulation of DNA-binding activity of Mycobacterium tuberculosis HspR by chaperones. Microbiology (Reading, England) *154*, 484-490.

Davenne, T., and McShane, H. (2016). Why don't we have an effective tuberculosis vaccine yet? Expert Rev Vaccines *15*, 1009-1013.

Davenport, J., Balch, M., Galam, L., Girgis, A., Hall, J., Blagg, B.S., and Matts, R.L. (2014). High-throughput screen of natural product libraries for hsp90 inhibitors. Biology *3*, 101-138.

del Castillo, U., Alfonso, C., Acebrón, S.P., Martos, A., Moro, F., Rivas, G., and Muga, A. (2011). A Quantitative Analysis of the Effect of Nucleotides and the

M Domain on the Association Equilibrium of ClpB. Biochemistry *50*, 1991-2003.

Delogu, G., Sali, M., and Fadda, G. (2013). The biology of mycobacterium tuberculosis infection. Mediterr J Hematol Infect Dis *5*, e2013070-e2013070.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., and Chan, H.S. (1995). Principles of protein folding--a perspective from simple exact models. Protein science : a publication of the Protein Society *4*, 561-602.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. Nature structural biology *4*, 10-19.

Dimoliatis, I.D., and Liaskos, C.A. (2008). Six Mantoux tuberculin skin tests with 1, 2, 5, 10, 20, and 50 units in a healthy male without side-effects - is skin reaction a linear function of tuberculin dose? Cases J *1*, 115-115.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884-890.

Doyle, S.M., Hoskins, J.R., and Wickner, S. (2007). Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system. Proceedings of the National Academy of Sciences of the United States of America *104*, 11138-11144.

Dragovic, Z., Broadley, S.A., Shomura, Y., Bracher, A., and Hartl, F.U. (2006). Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s. EMBO J *25*, 2519-2528.

Easton, D.P., Kaneko, Y., and Subjeck, J.R. (2000). The hsp110 and Grp1 70 stress proteins: newly recognized relatives of the Hsp70s. Cell Stress Chaperones *5*, 276-290.

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

Ellis, R.J. (1993). The general concept of molecular chaperones. Philosophical transactions of the Royal Society of London Series B, Biological sciences *339*, 257-261.

Ellis, R.J. (2001). Macromolecular crowding: obvious but underappreciated. Trends in biochemical sciences *26*, 597-604.

Fair, R.J., and Tor, Y. (2014). Antibiotics and bacterial resistance in the 21st century. Perspect Medicin Chem *6*, 25-64.

Fan, C.-Y., Lee, S., and Cyr, D.M. (2003). Mechanisms for regulation of Hsp70 function by Hsp40. Cell stress & chaperones *8*, 309-316.

Fätkenheuer, G., Taelman, H., Lepage, P., Schwenk, A., and Wenzel, R. (1999). The return of tuberculosis. Diagn Microbiol Infect Dis *34*, 139-146.

Fay, A., and Glickman, M.S. (2014). An Essential Nonredundant Role for Mycobacterial DnaK in Native Protein Folding. PLOS Genetics *10*, e1004516.

Fernández-Fernández, M.R., and Valpuesta, J.M. (2018). Hsp70 chaperone: a master player in protein homeostasis. F1000Res *7*, F1000 Faculty Rev-1497.

Fink, A.L. (2005). Natively unfolded proteins. Current opinion in structural biology *15*, 35-41.

Fischer, H.M., Babst, M., Kaspar, T., Acuña, G., Arigoni, F., and Hennecke, H. (1993). One member of a gro-ESL-like chaperonin multigene family in Bradyrhizobium japonicum is co-regulated with symbiotic nitrogen fixation genes. EMBO J *12*, 2901-2912.

Flaherty, K.M., DeLuca-Flaherty, C., and McKay, D.B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature *346*, 623-628.

Forrellad, M.A., Klepp, L.I., Gioffré, A., Sabio y García, J., Morbidoni, H.R., de la Paz Santangelo, M., Cataldi, A.A., and Bigi, F. (2013). Virulence factors of the Mycobacterium tuberculosis complex. Virulence *4*, 3-66.

Frebel, K., and Wiese, S. (2007). Signalling molecules essential for neuronal survival and differentiation. Biochemical Society transactions *34*, 1287-1290.

Fredericks, D.N., and Relman, D.A. (1996). Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clinical Microbiology Reviews *9*, 18.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem *70*, 603-647.

Gassler, C.S., Wiederkehr, T., Brehmer, D., Bukau, B., and Mayer, M.P. (2001). Bag-1M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration-dependent as positive and negative cofactor. J Biol Chem *276*, 32538-32544.

Gelinas, A.D., Langsetmo, K., Toth, J., Bethoney, K.A., Stafford, W.F., and Harrison, C.J. (2002). A Structure-based Interpretation of E.coli GrpE Thermodynamic Properties. Journal of molecular biology *323*, 131-142.

Genevaux, P., Georgopoulos, C., and Kelley, W.L. (2007). The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions. Molecular microbiology *66*, 840-857.

Georgopoulos, C., and Welch, W.J. (1993). Role of the major heat shock proteins as molecular chaperones. Annual review of cell biology *9*, 601-634.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Gideon, H.P., and Flynn, J.L. (2011). Latent tuberculosis: what the host "sees"? Immunologic research *50*, 202-212.

Glickman, M.S., and Jacobs, W.R., Jr. (2001). Microbial pathogenesis of Mycobacterium tuberculosis: dawn of a discipline. Cell *104*, 477-485.

Glover, J.R., and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell *94*, 73-82.

Goloubinoff, P., Mogk, A., Ben-Zvi, A., Tomoyasu, T., and Bukau, B. (1999a). Sequential Mechanism of Solubilization and Refolding of Stable Protein Aggregates by a Bichaperone Network. Proceedings of the National Academy of Sciences of the United States of America *96*, 13732-13737.

Goloubinoff, P., Mogk, A., Zvi, A.P., Tomoyasu, T., and Bukau, B. (1999b). Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proceedings of the National Academy of Sciences of the United States of America *96*, 13732-13737.

Gowda, N.K., Kaimal, J.M., Masser, A.E., Kang, W., Friedländer, M.R., and Andréasson, C. (2016). Cytosolic splice isoform of Hsp70 nucleotide exchange factor Fes1 is required for the degradation of misfolded proteins in yeast. Molecular biology of the cell *27*, 1210-1219.

Gowda, N.K.C., Kandasamy, G., Froehlich, M.S., Dohmen, R.J., and Andréasson, C. (2013). Hsp70 nucleotide exchange factor Fes1 is essential for ubiquitin-dependent degradation of misfolded cytosolic proteins. Proceedings of the National Academy of Sciences *110*, 5975.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog *7*, e1002251.

Grimshaw, J.P.A., Jelesarov, I., Schönfeld, H.-J., and Christen, P. (2001). Reversible Thermal Transition in GrpE, the Nucleotide Exchange Factor of the DnaK Heat-Shock System. Journal of Biological Chemistry *276*, 6098-6104.

Gruebele, M. (2002). Protein folding: the free energy surface. Current opinion in structural biology *12*, 161-168.

Gupta, A., Puri, A., Singh, P., Sonam, S., Pandey, R., and Sharma, D. (2018). The yeast stress inducible Ssa Hsp70 reduces α-synuclein toxicity by promoting its degradation through autophagy. PLOS Genetics *14*, e1007751.

Gutsche, I., Essen, L.O., and Baumeister, W. (1999). Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. Journal of molecular biology *293*, 295-312.

Harrison, C. (2003). GrpE, a nucleotide exchange factor for DnaK. Cell Stress Chaperones *8*, 218-224.

Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F., and Kuriyan, J. (1997a). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science (New York, NY) *276*, 431-435.

Harrison, C.J., Hayer-Hartl, M., Liberto, M.D., Hartl, F.-U., and Kuriyan, J. (1997b). Crystal Structure of the Nucleotide Exchange Factor GrpE Bound to the ATPase Domain of the Molecular Chaperone DnaK. Science *276*, 431-435.

Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. Nature *381*, 571-579.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. Science (New York, NY) *295*, 1852.

Hendriks, L.E.L., and Dingemans, A.C. (2017). Heat shock protein antagonists in early stage clinical trials for NSCLC. Expert opinion on investigational drugs *26*, 541-550.

Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E.*, et al.* (2008). Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. PloS one *3*, e3426.

Hett, E.C., and Rubin, E.J. (2008). Bacterial Growth and Cell Division: a Mycobacterial Perspective. Microbiology and Molecular Biology Reviews *72*, 126.

Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., and Small, P.M. (2004). Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proceedings of the National Academy of Sciences of the United States of America *101*, 4871-4876.

Ho, A.K., Raczniak, G.A., Ives, E.B., and Wente, S.R. (1998). The integral membrane protein snl1p is genetically linked to yeast nuclear pore complex function. Molecular biology of the cell *9*, 355-373.

Honoré, F.A., Méjean, V., and Genest, O. (2017). Hsp90 Is Essential under Heat Stress in the Bacterium Shewanella oneidensis. Cell Rep *19*, 680-687.

Horwich, A.L., Fenton, W.A., and Rapoport, T.A. (2001). Protein folding taking shape. Workshop on molecular chaperones. EMBO Rep *2*, 1068-1073.

Hu, Q.N., Deng, Z., Hu, H., Cao, D.S., and Liang, Y.Z. (2011). RxnFinder: biochemical reaction search engines using molecular structures, molecular fragments and reaction similarity. Bioinformatics (Oxford, England) *27*, 2465-2467.

Hussey, G., Hawkridge, T., and Hanekom, W. (2007). Childhood tuberculosis: old and new vaccines. Paediatric respiratory reviews *8*, 148-154.

Illingworth, M., Ramsey, A., Zheng, Z., and Chen, L. (2011). Stimulating the substrate folding activity of a single ring GroEL variant by modulating the cochaperonin GroES. J Biol Chem *286*, 30401-30408.

Imai, J., Maruya, M., Yashiroda, H., Yahara, I., and Tanaka, K. (2003). The molecular chaperone Hsp90 plays a role in the assembly and maintenance of the 26S proteasome. EMBO J *22*, 3557-3567.

Institute of, M. (2012). The National Academies Collection: Reports funded by National Institutes of Health. In Facing the Reality of Drug-Resistant Tuberculosis in India: Challenges and Potential Solutions: Summary of a Joint Workshop by the Institute of Medicine, the Indian National Science Academy, and the Indian Council of Medical Research (Washington (DC): National Academies Press (US) Copyright © 2012, National Academy of Sciences.).

Jackson, M. (2014). The mycobacterial cell envelope-lipids. Cold Spring Harb Perspect Med *4*, a021105.

Jaspard, E., and Hunault, G. (2016). sHSPdb: a database for the analysis of small Heat Shock Proteins. BMC Plant Biology *16*, 135.

Jhaveri, K., Taldone, T., Modi, S., and Chiosis, G. (2012). Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. Biochimica et biophysica acta *1823*, 742-755.

Jiang, J., Prasad, K., Lafer, E.M., and Sousa, R. (2005a). Structural Basis of Interdomain Communication in the Hsc70 Chaperone. Molecular Cell *20*, 513-524.

Jiang, J., Prasad, K., Lafer, E.M., and Sousa, R. (2005b). Structural basis of interdomain communication in the Hsc70 chaperone. Molecular cell *20*, 513-524.

Johnson, S.M., Sharif, O., Mak, P.A., Wang, H.T., Engels, I.H., Brinker, A., Schultz, P.G., Horwich, A.L., and Chapman, E. (2014). A biochemical screen for GroEL/GroES inhibitors. Bioorganic & medicinal chemistry letters *24*, 786-789.

Kabani, M., Beckerich, J.-M., and Brodsky, J.L. (2002). Nucleotide Exchange Factor for the Yeast Hsp70 Molecular Chaperone Ssa1p. Molecular and Cellular Biology *22*, 4677.

Kampinga, H.H., and Bergink, S. (2016). Heat shock proteins as potential targets for protective strategies in neurodegeneration. The Lancet Neurology *15*, 748-759.

Kar, N.P., Sikriwal, D., Rath, P., Choudhary, R.K., and Batra, J.K. (2008). Mycobacterium tuberculosis ClpC1: characterization and role of the N-terminal domain in its function. The FEBS journal *275*, 6149-6158.

Karunakaran, K.P., Noguchi, Y., Read, T.D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C.C., and Brunham, R.C. (2003). Molecular analysis of the multiple GroEL proteins of Chlamydiae. Journal of bacteriology *185*, 1958-1966.

Kaur, G., Kapoor, S., and Thakur, K.G. (2018). Bacillus subtilis HelD, an RNA Polymerase Interacting Helicase, Forms Amyloid-Like Fibrils. Front Microbiol *9*, 1934-1934.

Kedzierska-Mieszkowska, S., Akoyev, V., Barnett, M., and Zolkiewski, M. (2004). Structure and Function of the Middle Domain of ClpB from Escherichia coli. Biochemistry *42*, 14242-14248.

Kim, W.S., Jung, I.D., Kim, J.-S., Kim, H.M., Kwon, K.W., Park, Y.-M., and Shin, S.J. (2018a). Mycobacterium tuberculosis GrpE, A Heat-Shock Stress Responsive Chaperone, Promotes Th1-Biased T Cell Immune Response via TLR4-Mediated Activation of Dendritic Cells. Frontiers in Cellular and Infection Microbiology *8*.

Kim, W.S., Kim, J.S., Kim, H.M., Kwon, K.W., Eum, S.Y., and Shin, S.J. (2018b). Comparison of immunogenicity and vaccine efficacy between heat-shock proteins, HSP70 and GrpE, in the DnaK operon of Mycobacterium tuberculosis. Sci Rep *8*, 14411.

Kim, Y.S., Alarcon, S.V., Lee, S., Lee, M.J., Giaccone, G., Neckers, L., and Trepel, J.B. (2009). Update on Hsp90 inhibitors in clinical trial. Current topics in medicinal chemistry *9*, 1479-1492.

Kityk, R., Kopp, J., Sinning, I., and Mayer, Matthias P. (2012). Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. Molecular cell *48*, 863-874.

Kong, T.H., Coates, A.R., Butcher, P.D., Hickman, C.J., and Shinnick, T.M. (1993). Mycobacterium tuberculosis expresses two chaperonin-60 homologs. Proceedings of the National Academy of Sciences of the United States of America *90*, 2608-2612.

Korasick, D.A., and Tanner, J.J. (2018). Determination of protein oligomeric structure from small-angle X-ray scattering. Protein science : a publication of the Protein Society *27*, 814-824.

Krajewska, J., Modrak-Wójcik, A., Arent, Z.J., Więckowski, D., Zolkiewski, M., Bzowska, A., and Kędzierska-Mieszkowska, S. (2017). Characterization of the

molecular chaperone ClpB from the pathogenic spirochaete Leptospira interrogans. PLOS ONE *12*, e0181118.

Kriegenburg, F., Jakopec, V., Poulsen, E.G., Nielsen, S.V., Roguev, A., Krogan, N., Gordon, C., Fleig, U., and Hartmann-Petersen, R. (2014). A Chaperone-Assisted Degradation Pathway Targets Kinetochore Proteins to Ensure Genome Stability. PLOS Genetics *10*, e1004140.

Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nature methods *11*, 319-324.

Kumar, A., Farhana, A., Guidry, L., Saini, V., Hondalus, M., and Steyn, A.J. (2011a). Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert reviews in molecular medicine *13*, e39.

Kumar, A., Farhana, A., Guidry, L., Saini, V., Hondalus, M., and Steyn, A.J.C. (2011b). Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert reviews in molecular medicine *13*, e39-e39.

Kumar, C.M.S., Mande, S.C., and Mahajan, G. (2015a). Multiple chaperonins in bacteria--novel functions and non-canonical behaviors. Cell Stress Chaperones *20*, 555-574.

Kumar, N., Gaur, D., Gupta, A., Puri, A., and Sharma, D. (2015b). Hsp90-Associated Immunophilin Homolog Cpr7 Is Required for the Mitotic Stability of [URE3] Prion in Saccharomyces cerevisiae. PLOS Genetics *11*, e1005567.

Kumar, N., Gaur, D., Masison, D.C., and Sharma, D. (2014). The BAG homology domain of Snl1 cures yeast prion [URE3] through regulation of Hsp70 chaperones. G3 (Bethesda) *4*, 461-470.

Kumar, P. (2016). Adult pulmonary tuberculosis as a pathological manifestation of hyperactive antimycobacterial immune response. Clin Transl Med *5*, 38-38.

Kurz, S.G., Furin, J.J., and Bark, C.M. (2016). Drug-Resistant Tuberculosis: Challenges and Progress. Infect Dis Clin North Am *30*, 509-522.

Landry, S.J. (2003). Structure and energetics of an allele-specific genetic interaction between dnaJ and dnaK: correlation of nuclear magnetic resonance chemical shift perturbations in the J-domain of Hsp40/DnaJ with binding affinity for the ATPase domain of Hsp70/DnaK. Biochemistry *42*, 4926-4936.

Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature *275*, 416-420.

Leak, R.K. (2014). Heat shock proteins in neurodegenerative disorders and aging. Journal of cell communication and signaling *8*, 293-310.

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

Lee, G.J., and Vierling, E. (2000). A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant physiology *122*, 189-198.

Lee, S., Choi, J.M., and Tsai, F.T. (2007). Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB. Molecular cell *25*, 261-271.

Lee, S., Sowa, M.E., Watanabe, Y.-h., Sigler, P.B., Chiu, W., Yoshida, M., and Tsai, F.T.F. (2003a). The Structure of ClpB: A Molecular Chaperone that Rescues Proteins from an Aggregated State. Cell *115*, 229-240.

Lee, S., Sowa, M.E., Watanabe, Y.H., Sigler, P.B., Chiu, W., Yoshida, M., and Tsai, F.T. (2003b). The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell *115*, 229-240.

Leung, E.C.C., Yew, W.W., Leung, C.C., Leung, W.M., and Tam, C.M. (2011). Shorter treatment duration for selected patients with multidrug-resistant tuberculosis. European Respiratory Journal *38*, 227.

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

Li, J., Qian, X., and Sha, B. (2009). Heat shock protein 40: structural studies and their functional implications. Protein and peptide letters *16*, 606-612.

li, W., Li, Y., Guan, S., Cheng, C.-F., Bright, A., Chinn, C., Chen, M., and Woodley, D. (2007). Li W, Li Y, Guan S, Fan J, Cheng CF, Bright AM, Chinn C, Chen M, Woodley DTExtracellular heat shock protein-90alpha: linking hypoxia to skin cell motility and wound healing. Eur Mol Biol Organ J 26: 1221-1233. EMBO J *26*, 1221-1233.

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C., and Zylicz, M. (1991). Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. Proceedings of the National Academy of Sciences *88*, 2874.

Lin, J., and Lucius, A.L. (2015). Examination of the dynamic assembly equilibrium for E. coli ClpB. Proteins: Structure, Function, and Bioinformatics *83*, 2008-2024.

Lin, J., and Lucius, A.L. (2016). Examination of ClpB Quaternary Structure and Linkage to Nucleotide Binding. Biochemistry *55*, 1758-1771.

Lin, P.L., and Flynn, J.L. (2010). Understanding latent tuberculosis: a moving target. Journal of immunology (Baltimore, Md : 1950) *185*, 15-22.

Lindquist, S., and Craig, E.A. (1988). THE HEAT-SHOCK PROTEINS. Annual Review of Genetics *22*, 631-677.

Liu, Q., and Hendrickson, W.A. (2007). Insights into Hsp70 chaperone activity from a crystal structure of the yeast Hsp110 Sse1. Cell *131*, 106-120.

Liu, X.D., Morano, K.A., and Thiele, D.J. (1999). The yeast Hsp110 family member, Sse1, is an Hsp90 cochaperone. J Biol Chem *274*, 26654-26660.

Lorimer, G.H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. The FASEB Journal *10*, 5-9.

Lourdault, K., Cerqueira, G.M., Wunder, E.A., Jr., and Picardeau, M. (2011). Inactivation of clpB in the pathogen Leptospira interrogans reduces virulence and resistance to stress conditions. Infect Immun *79*, 3711-3717.

Lowe, D.M., Corbett, P.T., Murray-Rust, P., and Glen, R.C. (2011). Chemical Name to Structure: OPSIN, an Open Source Solution. Journal of Chemical Information and Modeling *51*, 739-753.

Lupoli, T.J., Fay, A., Adura, C., Glickman, M.S., and Nathan, C.F. (2016a). Reconstitution of a Mycobacterium tuberculosis proteostasis network highlights essential cofactor interactions with chaperone DnaK. Proceedings of the National Academy of Sciences *113*, E7947-E7956.

Lupoli, T.J., Fay, A., Adura, C., Glickman, M.S., and Nathan, C.F. (2016b). Reconstitution of a Mycobacterium tuberculosis proteostasis network highlights essential cofactor interactions with chaperone DnaK. Proceedings of the National Academy of Sciences *113*, E7947.

Lupoli, T.J., Vaubourgeix, J., Burns-Huang, K., and Gold, B. (2018a). Targeting the Proteostasis Network for Mycobacterial Drug Discovery. *4*, 478-498.

Lupoli, T.J., Vaubourgeix, J., Burns-Huang, K., and Gold, B. (2018b). Targeting the Proteostasis Network for Mycobacterial Drug Discovery. ACS Infectious Diseases *4*, 478-498.

Mally, A., and Witt, S.N. (2001). GrpE accelerates peptide binding and release from the high affinity state of DnaK. Nature Structural Biology *8*, 254-257.

Mao, Q., Ke, D., Feng, X., and Chang, Z. (2001). Preheat treatment for Mycobacterium tuberculosis Hsp16.3: correlation between a structural phase change at 60 degrees C and a dramatic increase in chaperone-like activity. Biochemical and biophysical research communications *284*, 942-947.

Martin, I., Underhaug, J., Celaya, G., Moro, F., Teigen, K., Martinez, A., and Muga, A. (2013). Screening and Evaluation of Small Organic Molecules as ClpB Inhibitors and Potential Antimicrobials. Journal of Medicinal Chemistry *56*, 7177-7189.

Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci *62*, 670-684.

Med, A., and Al-Ubaidi, B. (2018). The Radiological Diagnosis of Pulmonary Tuberculosis (TB) in Primary Care. *4*.

Mehdi, K. (2009). Structural and Functional Diversity Among Eukaryotic Hsp70 Nucleotide Exchange Factors. Protein & Peptide Letters *16*, 623-630.

Mehl, A.F., Heskett, L.D., Jain, S.S., and Demeler, B. (2003). Insights into dimerization and four-helix bundle formation found by dissection of the dimer interface of the GrpE protein from Escherichia coli. Protein Science *12*, 1205-1215.

Mehra, S., Golden, N.A., Stuckey, K., Didier, P.J., Doyle, L.A., Russell-Lodrigue, K.E., Sugimoto, C., Hasegawa, A., Sivasubramani, S.K., Roy, C.J.*, et al.* (2012). The Mycobacterium tuberculosis stress response factor SigH is required for bacterial burden as well as immunopathology in primate lungs. The Journal of infectious diseases *205*, 1203-1213.

Meyer, A.S., Gillespie, J.R., Walther, D., Millet, I.S., Doniach, S., and Frydman, J. (2003). Closing the Folding Chamber of the Eukaryotic Chaperonin Requires the Transition State of ATP Hydrolysis. Cell *113*, 369-381.

Miot, M., Reidy, M., Doyle, S.M., Hoskins, J.R., Johnston, D.M., Genest, O., Vitery, M.-C., Masison, D.C., and Wickner, S. (2011). Species-specific collaboration of heat shock proteins (Hsp) 70 and 100 in thermotolerance and protein disaggregation. Proceedings of the National Academy of Sciences of the United States of America *108*, 6915-6920.

Mitchison, D.A., and Davies, G.R. (2008). Assessment of the Efficacy of New Anti-Tuberculosis Drugs. Open Infect Dis J *2*, 59-76.

Mogk, A., Kummer, E., and Bukau, B. (2015). Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. Front Mol Biosci *2*, 22-22.

Mogk, A., Schlieker, C., Friedrich, K.L., Schönfeld, H.J., Vierling, E., and Bukau, B. (2003). Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. J Biol Chem *278*, 31033-31042.

Moro, F., Taneva, S.G., Velázquez-Campoy, A., and Muga, A. (2007). GrpE N-terminal Domain Contributes to the Interaction with DnaK and Modulates the Dynamics of the Chaperone Substrate Binding Domain. Journal of molecular biology *374*, 1054-1064.

Muñoz, L., Stagg, H.R., and Abubakar, I. (2015). Diagnosis and Management of Latent Tuberculosis Infection. Cold Spring Harbor perspectives in medicine *5*, a017830.

Nahleh, Z., Tfayli, A., Najm, A., El Sayed, A., and Nahle, Z. (2012). Heat shock proteins in cancer: targeting the 'chaperones'. Future medicinal chemistry *4*, 927-935.

Neckers, L., Mollapour, M., and Tsutsumi, S. (2009). The complex dance of the molecular chaperone Hsp90. Trends in biochemical sciences *34*, 223-226.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome research *9*, 27-43.

Niwa, T., Kanamori, T., Ueda, T., and Taguchi, H. (2012). Global analysis of chaperone effects using a reconstituted cell-free translation system. Proceedings of the National Academy of Sciences *109*, 8937.

Ohgushi, M., and Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. FEBS Letters *164*, 21-24.

Ojha, A., Anand, M., Bhatt, A., Kremer, L., Jacobs, W.R., and Hatfull, G.F. (2005). GroEL1: A Dedicated Chaperone Involved in Mycolic Acid Biosynthesis during Biofilm Formation in Mycobacteria. Cell *123*, 861-873.

Packschies, L., Theyssen, H., Buchberger, A., Bukau, B., Goody, R.S., and Reinstein, J. (1997). GrpE Accelerates Nucleotide Exchange of the Molecular Chaperone DnaK with an Associative Displacement Mechanism. Biochemistry *36*, 3417-3422.

Palleros, D.R., Reid, K.L., McCarty, J.S., Walker, G.C., and Fink, A.L. (1992). DnaK, hsp73, and their molten globules. Two different ways heat shock proteins respond to heat. Journal of Biological Chemistry *267*, 5279-5285.

Palleros, D.R.S., L; Reid, K L; Flink, A L (1993). Three-state denaturation of DnaK induced by guanidine hydrochloride. Evidence for an expandable intermediate. Biochemistry *32*, 4314-4321.

Palomino, J.C., and Martin, A. (2014). Drug Resistance Mechanisms in Mycobacterium tuberculosis. Antibiotics (Basel) *3*, 317-340.

Patel, S., and Latterich, M. (1998). The AAA team: related ATPases with diverse functions. Trends in cell biology *8*, 65-71.

Pellecchia, M., Montgomery, D.L., Stevens, S.Y., Vander Kooi, C.W., Feng, H.P., Gierasch, L.M., and Zuiderweg, E.R. (2000). Structural insights into substrate binding by the molecular chaperone DnaK. Nature structural biology *7*, 298-303.

Pesce, E.R., Cockburn, I.L., Goble, J.L., Stephens, L.L., and Blatch, G.L. (2010). Malaria heat shock proteins: drug targets that chaperone other drug targets. Infectious disorders drug targets *10*, 147-157.

Pfyffer, G.E., Bonato, D.A., Ebrahimzadeh, A., Gross, W., Hotaling, J., Kornblum, J., Laszlo, A., Roberts, G., Salfinger, M., Wittwer, F.*, et al.* (1999). Multicenter laboratory validation of susceptibility testing of Mycobacterium tuberculosis against classical second-line and newer antimicrobial drugs by using the radiometric BACTEC 460 technique and the proportion method with solid media. J Clin Microbiol *37*, 3179-3186.

Pierpaoli, E.V., Sandmeier, E., Schönfeld, H.-J., and Christen, P. (1998). Control of the DnaK Chaperone Cycle by Substoichiometric Concentrations of the Co-chaperones DnaJ and GrpE. Journal of Biological Chemistry *273*, 6643-6649.

Polier, S., Dragovic, Z., Hartl, F.U., and Bracher, A. (2008). Structural basis for the cooperation of Hsp70 and Hsp110 chaperones in protein folding. Cell *133*, 1068-1079.

Pratt, W.B., Krishna, P., and Olsen, L.J. (2001). Hsp90-binding immunophilins in plants: the protein movers. Trends in plant science *6*, 54-58.

Qamra, R., and Mande, S.C. (2004). Crystal structure of the 65-kilodalton heat shock protein, chaperonin 60.2, of Mycobacterium tuberculosis. Journal of bacteriology *186*, 8105-8113.

Qamra, R., Srinivas, V., and Mande, S.C. (2004). Mycobacterium tuberculosis GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. Journal of molecular biology *342*, 605-617.

Qi, R., Sarbeng, E.B., Liu, Q., Le, K.Q., Xu, X., Xu, H., Yang, J., Wong, J.L., Vorvis, C., Hendrickson, W.A.*, et al.* (2013). Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. Nature structural & molecular biology *20*, 900-907.

Qian, Y.Q., Patel, D., Hartl, F.U., and McColl, D.J. (1996). Nuclear magnetic resonance solution structure of the human Hsp40 (HDJ-1) J-domain. Journal of molecular biology *260*, 224-235.

Queitsch, C., Sangster, T.A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature *417*, 618-624.

R, R.K., N, S.N., S, P.A., Sinha, D., Veedin Rajan, V.B., Esthaki, V.K., and D'Silva, P. (2012a). HSPIR: a manually annotated heat shock protein information resource. Bioinformatics (Oxford, England) *28*, 2853-2855.

R, R.K., N. S, N., S. P, A., Sinha, D., Veedin Rajan, V.B., Esthaki, V.K., and D'Silva, P. (2012b). HSPIR: a manually annotated heat shock protein information resource. Bioinformatics *28*, 2853-2855.

Ramakrishnan, R., Houben, B., Kreft, Ł., Botzki, A., Schymkowitz, J., and Rousseau, F. (2020). Protein Homeostasis Database: protein quality control in E.coli. Bioinformatics (Oxford, England) *36*, 948-949.

Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W.R., Jr., and Husson, R.N. (2001). The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in Mycobacterium tuberculosis. Journal of bacteriology *183*, 6119-6125.

Ramdhave, A.S., Patel, D., Ramya, I., Nandave, M., and Kharkar, P.S. (2013). Targeting heat shock protein 90 for malaria. Mini reviews in medicinal chemistry *13*, 1903-1920.

Rani, S., Srivastava, A., Kumar, M., Goel, M., and Lund, P. (2017). CrAgDb-a database of annotated chaperone repertoire in archaeal genomes. FEMS Microbiol Lett *364*.

Ranson, N.A., White, H.E., and Saibil, H.R. (1998). Chaperonins. Biochem J *333 (Pt 2)*, 233-242.

Raskatov, J.A., and Teplow, D.B. (2017). Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins. Scientific Reports *7*, 12433.

Raviol, H., Sadlish, H., Rodriguez, F., Mayer, M.P., and Bukau, B. (2006). Chaperone network in the yeast cytosol: Hsp110 is revealed as an Hsp70 nucleotide exchange factor. EMBO J *25*, 2510-2518.

Reddy, G.B., Das, K.P., Petrash, J.M., and Surewicz, W.K. (2000). Temperature-dependent chaperone activity and structural properties of human alphaA- and alphaB-crystallins. J Biol Chem *275*, 4565-4570.

Reeg, S., Jung, T., Castro, J.P., Davies, K.J.A., Henze, A., and Grune, T. (2016). The molecular chaperone Hsp70 promotes the proteolytic removal of oxidatively damaged proteins by the proteasome. Free radical biology & medicine *99*, 153-166.

Richter, K., and Buchner, J. (2001). Hsp90: chaperoning signal transduction. Journal of cellular physiology *188*, 281-290.

Riello, F.N., Brígido, R.T.S., Araújo, S., Moreira, T.A., Goulart, L.R., and Goulart, I.M.B. (2016). Diagnosis of mycobacterial infections based on acid-fast bacilli test and bacterial growth time and implications on treatment and disease outcome. BMC Infectious Diseases *16*, 142.

Rizo, A.N., Lin, J., Gates, S.N., Tse, E., Bart, S.M., Castellano, L.M., DiMaio, F., Shorter, J., and Southworth, D.R. (2019). Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. Nature communications *10*, 2393-2393.

Rohde, M., Daugaard, M., Jensen, M.H., Helin, K., Nylandsted, J., and Jäättelä, M. (2005). Members of the heat-shock protein 70 family promote cancer cell growth by distinct mechanisms. Genes & Development *19*, 570-582.

Rosenzweig, R., Farber, P., Velyvis, A., Rennella, E., Latham, M.P., and Kay, L.E. (2015). ClpB N-terminal domain plays a regulatory role in protein disaggregation. Proc Natl Acad Sci U S A *112*, E6872-E6881.

Rothschild, B.M., Martin, L.D., Lev, G., Bercovier, H., Bar-Gal, G.K., Greenblatt, C., Donoghue, H., Spigelman, M., and Brittain, D. (2001). Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *33*, 305-311.

Rutherford, S., and Lindquist, S. (1998). HSP90 as a capacitor for morphological evolution. Nature *396*, 336-342.

Ryu, Y.J. (2015). Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. Tuberc Respir Dis (Seoul) *78*, 64-71.

Saibil, H. (2013a). Chaperone machines for protein folding, unfolding and disaggregation. Nature reviews Molecular cell biology *14*, 630-642.

Saibil, H. (2013b). Chaperone machines for protein folding, unfolding and disaggregation. Nature Reviews Molecular Cell Biology *14*, 630.

Saito, H., and Uchida, H. (1977). Initiation of the DNA replication of bacteriophage lambda in Escherichia coli K12. Journal of molecular biology *113*, 1-25.

Saito, K., Warrier, T., Somersan-Karakaya, S., Kaminski, L., Mi, J., Jiang, X., Park, S., Shigyo, K., Gold, B., Roberts, J.*, et al.* (2017). Rifamycin action on RNA polymerase in antibiotic-tolerant Mycobacterium tuberculosis results in differentially detectable populations. Proc Natl Acad Sci U S A *114*, E4832-E4840.

Sakula, A. (1983). Robert koch: centenary of the discovery of the tubercle bacillus, 1882. Can Vet J *24*, 127-131.

Sandhu, G.K. (2011). Tuberculosis: current situation, challenges and overview of its control programs in India. J Glob Infect Dis *3*, 143-150.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol *48*, 77-84.

Schirmer, E.C., Glover, J.R., Singer, M.A., and Lindquist, S. (1996). HSP100/Clp proteins: a common mechanism explains diverse functions. Trends in biochemical sciences *21*, 289-296.

Schluger, N.W., and Rom, W.N. (1998). The host immune response to tuberculosis. Am J Respir Crit Care Med *157*, 679-691.

Schmid, D., Baici, A., Gehring, H., and Christen, P. (1994). Kinetics of molecular chaperone action. Science (New York, NY) *263*, 971-973.

Schmitz, K.R., and Sauer, R.T. (2014). Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. Molecular microbiology *93*, 617-628.

Schönfeld, H.-J., Schmidt, D., Schröder, H., and Bukau, B. (1995). The DnaK Chaperone System of Escherichia coli: Quaternary Structures and Interactions of the DnaK and GrpE Components. Journal of Biological Chemistry *270*, 2183-2189.

Schönfeld, H.J., Schmidt, D., and Zulauf, M. (2007). Investigation of the molecular chaperone DnaJ by analytical ultracentrifugation. In, pp. 7-10.

Schopf, F.H., Biebl, M.M., and Buchner, J. (2017). The HSP90 chaperone machinery. Nature reviews Molecular cell biology *18*, 345-360.

Schröder, H., Langer, T., Hartl, F.U., and Bukau, B. (1993). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. The EMBO journal *12*, 4137-4144.

Seckler, R., and Jaenicke, R. (1992). Protein folding and protein refolding. The FASEB Journal *6*, 2545-2552.

Seyffer, F., Kummer, E., Oguchi, Y., Winkler, J., Kumar, M., Zahn, R., Sourjik, V., Bukau, B., and Mogk, A. (2012). Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces. Nature structural & molecular biology *19*.

Sharma, D., and Masison, D.C. (2008). Functionally Redundant Isoforms of a Yeast Hsp70 Chaperone Subfamily Have Different Antiprion Effects. Genetics *179*, 1301.

Sharma, D., and Masison, D.C. (2009). Hsp70 structure, function, regulation and influence on yeast prions. Protein and peptide letters *16*, 571-581.

Shastri, M.D., Shukla, S.D., Chong, W.C., Dua, K., Peterson, G.M., Patel, R.P., Hansbro, P.M., Eri, R., and O'Toole, R.F. (2018). Role of Oxidative Stress in the Pathology and Management of Human Tuberculosis. Oxid Med Cell Longev *2018*, 7695364-7695364.

Shirota, T., Ojima, H., Hiraoka, N., Shimada, K., Rokutan, H., Arai, Y., Kanai, Y., Miyagawa, S., and Shibata, T. (2015). Heat Shock Protein 90 Is a Potential Therapeutic Target in Cholangiocarcinoma. Molecular cancer therapeutics *14*, 1985-1993.

Shomura, Y., Dragovic, Z., Chang, H.C., Tzvetkov, N., Young, J.C., Brodsky, J.L., Guerriero, V., Hartl, F.U., and Bracher, A. (2005). Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. Molecular cell *17*, 367-379.

Shorter, J., and Lindquist, S. (2004). Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers. Science (New York, NY) *304*, 1793-1797.

Shorter, J., and Lindquist, S. (2006). Destruction or potentiation of different prions catalyzed by similar Hsp104 remodeling activities. Molecular cell *23*, 425-438.

Shorter, J., and Lindquist, S. (2008). Hsp104, Hsp70 and Hsp40 interplay regulates formation, growth and elimination of Sup35 prions. EMBO J *27*, 2712-2724.

Singh, P., Unik, B., Puri, A., Nagpal, G., Singh, B., Gautam, A., and Sharma, D. (2020). HSPMdb: a computational repository of heat shock protein modulators. Database *2020*.

Sisay, S., Mengistu, B., Erku, W., and Woldeyohannes, D. (2014). Directly Observed Treatment Short-course (DOTS) for tuberculosis control program in Gambella Regional State, Ethiopia: ten years experience. BMC Research Notes *7*, 44.

Smith, I. (2003). Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. Clinical Microbiology Reviews *16*, 463.

Steinbrook, R. (2007). Tuberculosis and HIV in India. The New England journal of medicine *356*, 1198-1199.

Sterling, T.R., and Haas, D.W. (2006). Transmission of Mycobacterium tuberculosis from health care workers. The New England journal of medicine *355*, 118-121.

Stewart, G.R., Robertson, B.D., and Young, D.B. (2003). Tuberculosis: a problem with persistence. Nature reviews Microbiology *1*, 97-105.

Stewart, G.R., Snewin, V.A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I.N., and Young, D.B. (2001). Overexpression of heat-shock proteins reduces survival of Mycobacterium tuberculosis in the chronic phase of infection. Nature medicine *7*, 732-737.

Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Butcher, P.D., and Young, D.B. (2002). The heat shock response of Mycobacterium tuberculosis: linking gene expression, immunology and pathogenesis. Comp Funct Genomics *3*, 348-351.

Sun, R., Converse, P.J., Ko, C., Tyagi, S., Morrison, N.E., and Bishai, W.R. (2004). Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. Molecular microbiology *52*, 25-38.

Sweeny, E.A., DeSantis, M.E., and Shorter, J. (2011). Purification of hsp104, a protein disaggregase. Journal of visualized experiments : JoVE, 3190.

Sweeny, Elizabeth A., Jackrel, Meredith E., Go, Michelle S., Sochor, Matthew A., Razzo, Beatrice M., DeSantis, Morgan E., Gupta, K., and Shorter, J. (2015a). The Hsp104 N-Terminal Domain Enables Disaggregase Plasticity and Potentiation. Molecular cell *57*, 836-849.

Sweeny, E.A., Jackrel, M.E., Go, M.S., Sochor, M.A., Razzo, B.M., DeSantis, M.E., Gupta, K., and Shorter, J. (2015b). The Hsp104 N-terminal domain enables disaggregase plasticity and potentiation. Molecular cell *57*, 836-849.

Szabo, A., Korszun, R., Hartl, F.U., and Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. EMBO J *15*, 408-417.

Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B., and Hartl, F.U. (1994a). The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. Proceedings of the National Academy of Sciences of the United States of America *91*, 10345-10349.

Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B., and U Hartl, F. (1994b). The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. Proceedings of the National Academy of Sciences of the United States of America *91*, 10345-10349.

Takayama, S., Bimston, D.N., Matsuzawa, S., Freeman, B.C., Aime-Sempe, C., Xie, Z., Morimoto, R.I., and Reed, J.C. (1997). BAG-1 modulates the chaperone activity of Hsp70/Hsc70. EMBO J *16*, 4887-4896.

Tanaka, N., and Nakamoto, H. (1999). HtpG is essential for the thermal stress management in cyanobacteria. FEBS Letters *458*, 117-123.

Tanaka, N., Tani, Y., Hattori, H., Tada, T., and Kunugi, S. (2004). Interaction of the N-terminal domain of Escherichia coli heat-shock protein ClpB and protein aggregates during chaperone activity. Protein Sci *13*, 3214-3221.

Tezcan, F.A., Findley, W.M., Crane, B.R., Ross, S.A., Lyubovitsky, J.G., Gray, H.B., and Winkler, J.R. (2002). Using deeply trapped intermediates to map the cytochrome c folding landscape. Proceedings of the National Academy of Sciences of the United States of America *99*, 8626-8630.

Thomas, J.G., and Baneyx, F. (1998). Roles of the Escherichia coli Small Heat Shock Proteins IbpA and IbpB in Thermal Stress Management: Comparison with ClpA, ClpB, and HtpG In Vivo. Journal of Bacteriology *180*, 5165.

TM, D. (2009). The history of tuberculosis: past, present and challenges for the future. Tuberculosis: a comprehensive clinical reference *Philadelphia, PA: Saunders*, 1–7.

Trinh, Q.M., Nguyen, H.L., Nguyen, V.N., Nguyen, T.V., Sintchenko, V., and Marais, B.J. (2015). Tuberculosis and HIV co-infection-focus on the Asia-Pacific

region. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases *32*, 170-178.

Tripathi, P., Parijat, P., Patel, V.K., and Batra, J.K. (2018). The amino-terminal domain of Mycobacterium tuberculosis ClpB protein plays a crucial role in its substrate disaggregation activity. FEBS open bio *8*, 1669-1690.

Tripathi, P., Singh, L.K., Kumari, S., Hakiem, O.R., and Batra, J.K. (2020a). ClpB is an essential stress regulator of Mycobacterium tuberculosis and endows survival advantage to dormant bacilli. International journal of medical microbiology : IJMM *310*, 151402.

Tripathi, P., Singh, L.K., Kumari, S., Hakiem, O.R., and Batra, J.K. (2020b). ClpB is an essential stress regulator of Mycobacterium tuberculosis and endows survival advantage to dormant bacilli. International Journal of Medical Microbiology *310*, 151402.

Truman, A.W., Kristjansdottir, K., Wolfgeher, D., Hasin, N., Polier, S., Zhang, H., Perrett, S., Prodromou, C., Jones, G.W., and Kron, S.J. (2012). CDK-dependent Hsp70 Phosphorylation controls G1 cyclin abundance and cell-cycle progression. Cell *151*, 1308-1318.

Tsai, J., and Douglas, M. (1996). A Conserved HPD Sequence of the J-domain Is Necessary for YDJ1 Stimulation of Hsp70 ATPase Activity at a Site Distinct from Substrate Binding. J Biol Chem *271*, 9347-9354.

Ueno, T., Taguchi, H., Tadakuma, H., Yoshida, M., and Funatsu, T. (2004). GroEL mediates protein folding with a two successive timer mechanism. Molecular cell *14*, 423-434.

van den Berg, B., Ellis, R.J., and Dobson, C.M. (1999). Effects of macromolecular crowding on protein folding and aggregation. EMBO J *18*, 6927-6933.

Vaubourgeix, J., Lin, G., Dhar, N., Chenouard, N., Jiang, X., Botella, H., Lupoli, T., Mariani, O., Yang, G., Ouerfelli, O*., et al.* (2015). Stressed Mycobacteria Use the Chaperone ClpB to Sequester Irreversibly Oxidized Proteins Asymmetrically Within and Between Cells. Cell Host & Microbe *17*, 178-190.

Veinger, L., Diamant, S., Buchner, J., and Goloubinoff, P. (1998). The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J Biol Chem *273*, 11032-11037.

Veitshans, T., Klimov, D., and Thirumalai, D. (1997). Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Folding & design *2*, 1-22.

Vidal, D., Thormann, M., and Pons, M. (2005). LINGO, an Efficient Holographic Text Based Method To Calculate Biophysical Properties and Intermolecular Similarities. Journal of Chemical Information and Modeling *45*, 386-393.

Vierling, E. (1997). The small heat shock proteins in plants are members of an ancient family of heat induced proteins. Acta Physiologiae Plantarum *19*, 539-547.

Vinod Kumar, K., Lall, C., Vimal Raj, R., Vedhagiri, K., Kartick, C., Surya, P., Natarajaseenivasan, K., and Vijayachari, P. (2017). Overexpression of heat shock GroEL stress protein in leptospiral biofilm. Microbial Pathogenesis *102*, 8-11.

Wang, A.M., Miyata, Y., Klinedinst, S., Peng, H.M., Chua, J.P., Komiyama, T., Li, X., Morishima, Y., Merry, D.E., Pratt, W.B.*, et al.* (2013). Activation of Hsp70 reduces neurotoxicity by promoting polyglutamine protein degradation. Nature chemical biology *9*, 112-118.

Weber-Ban, E.U., Reid, B.G., Miranker, A.D., and Horwich, A.L. (1999). Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. Nature *401*, 90-93.

Weissman, J.S., Hohl, C.M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H.R., Fenton, W.A., and Horwich, A.L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. Cell *83*, 577-587.

WHO (Global Tuberculosis Report, 2019).

Winklhofer, K.F., Tatzelt, J., and Haass, C. (2008). The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. EMBO J *27*, 336-349.

Winter, J., and Jakob, U. (2004). Beyond transcription--new mechanisms for the regulation of molecular chaperones. Critical reviews in biochemistry and molecular biology *39*, 297-317.

Wolynes, P.G., Luthey-Schulten, Z., and Onuchic, J.N. (1996). Fast-folding eriments and the topography of protein folding energy landscapes. Chemistry & Biology *3*, 425-432.

Wyman, C., Vasilikiotis, C., Ang, D., Georgopoulos, C., and Echols, H. (1993). Function of the GrpE heat shock protein in bidirectional unwinding and replication from the origin of phage ?? J Biol Chem *268*, 25192-25196.

Yap, C.W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. Journal of computational chemistry *32*, 1466-1474.

Young, J.C., Agashe, V.R., Siegers, K., and Hartl, F.U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. Nature reviews Molecular cell biology *5*, 781-791.

Young, J.C., Moarefi, I., and Hartl, F.U. (2001). Hsp90: a specialized but essential protein-folding tool. The Journal of cell biology *154*, 267-273.

Yu, H., Lupoli, T.J., Kovach, A., Meng, X., Zhao, G., Nathan, C.F., and Li, H. (2018a). ATP hydrolysis-coupled peptide translocation mechanism of Mycobacterium tuberculosis ClpB. Proceedings of the National Academy of Sciences *115*, E9560.

Yu, H., Lupoli, T.J., Kovach, A., Meng, X., Zhao, G., Nathan, C.F., and Li, H. (2018b). ATP hydrolysis-coupled peptide translocation mechanism of Mycobacterium tuberculosis ClpB. Proceedings of the National Academy of Sciences *115*, E9560-E9569.

Yuan, Y., Crane, D.D., Simpson, R.M., Zhu, Y.Q., Hickey, M.J., Sherman, D.R., and Barry, C.E., 3rd (1998). The 16-kDa alpha-crystallin (Acr) protein of Mycobacterium tuberculosis is required for growth in macrophages. Proceedings of the National Academy of Sciences of the United States of America *95*, 9578-9583.

Zhai, W., Wu, F., Zhang, Y., Fu, Y., and Liu, Z. (2019). The Immune Escape Mechanisms of Mycobacterium Tuberculosis. Int J Mol Sci *20*, 340.

Zhang, T., Kedzierska-Mieszkowska, S., Liu, H., Cheng, C., Ganta, R.R., and Zolkiewski, M. (2013). Aggregate-Reactivation Activity of the Molecular Chaperone ClpB from Ehrlichia chaffeensis. PLOS ONE *8*, e62454.

Zhang, Y., and Yew, W.W. (2015). Mechanisms of drug resistance in Mycobacterium tuberculosis: update 2015. Int J Tuberc Lung Dis *19*, 1276-1289.

Zhang, Y., and Zuiderweg, E.R.P. (2004). The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. Proceedings of the National Academy of Sciences of the United States of America *101*, 10272.

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

Zolkiewski, M., Kessel, M., Ginsburg, A., and Maurizi, M.R. (1999). Nucleotide-dependent oligomerization of C1pB from Escherichia coli. Protein Science *8*, 1899-1903.

Zolkiewski, M., Zhang, T., and Nagy, M. (2012). Aggregate reactivation mediated by the Hsp100 chaperones. Archives of biochemistry and biophysics *520*, 1-6.

Publications

International Journal of Biological Macromolecules 165 (2020) 375–387

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

Biochemical characterization of ClpB protein from *Mycobacterium tuberculosis* and identification of its small-molecule inhibitors

Prashant Singh [a], Harleen Khurana [b], Shiv Pratap Yadav [a], Kanika Dhiman [a], Padam Singh [b], Ashish [a], Ramandeep Singh [b], Deepak Sharma [a,*]

[a] Council of Scientific and Industrial Research-Institute of Microbial Technology, India
[b] Tuberculosis Research Laboratory, Translational Health Science and Technology Institute, India

ARTICLE INFO

Article history:
Received 9 July 2020
Received in revised form 25 August 2020
Accepted 18 September 2020
Available online 25 September 2020

Keywords:
ClpB
Chaperones
Hsp100
M. tuberculosis

ABSTRACT

Tuberculosis, caused by pathogenic *M. tuberculosis*, remains a global health concern among various infectious diseases. Studies show that ClpB, a major disaggregase, protects the pathogen from various stresses encountered in the host environment. In the present study we have performed a detailed biophysical characterization of *M. tuberculosis* ClpB followed by a high throughput screening to identify small molecule inhibitors. The sedimentation velocity studies reveal that ClpB oligomerization varies with its concentration and presence of nucleotides. Further, using high throughput malachite green-based screening assay, we identified potential novel inhibitors of ClpB ATPase activity. The enzyme kinetics revealed that the lead molecule inhibits ClpB activity in a competitive manner. These drugs were also able to inhibit ATPase activity associated with *E. coli* ClpB and yeast Hsp104. The identified drugs inhibited the growth of intracellular bacteria in macrophages. Small angle X-ray scattering based modeling shows that ATP, and not its non-hydrolyzable analogs induce large scale conformational rearrangements in ClpB. Remarkably, the identified small molecules inhibited these ATP inducible conformational changes, suggesting that nucleotide induced shape changes are crucial for ClpB activity. The study broadens our understanding of *M. tuberculosis* chaperone machinery and provides the basis for designing more potent inhibitors against ClpB chaperone.

1. Introduction

Despite several decades of extensive research, the biology of *Mycobacterium tuberculosis* (*M. tuberculosis*) remains poorly understood, hindering the development of effective diagnosis and therapeutics. Tuberculosis still poses a major health challenge and is the leading cause of death among infectious diseases [1]. Post-infection, *M. tuberculosis* could persist in its host for decades without developing any clinical symptoms. This latent or metabolically less active bacteria is reactivated as the host immune response gets compromised [2]. In its host, *M. tuberculosis* encounters numerous stresses such as reactive oxygen species, reactive nitrogen species, low oxygen and nutritional stress, which are known to adversely affect stability and abundance of a number of cellular proteins, resulting in an imbalance in protein homeostasis [3,4]. To counter such imbalance, *M. tuberculosis* harbors an efficient protein quality control pathway consisting of DnaK and its co-chaperones to prevent aggregation of partially unfolded proteins, and a disaggregase ClpB that function to disassemble preformed protein aggregates. Therefore, better understanding of the *M. tuberculosis* chaperone machinery would provide insight into mechanism underlying its stress tolerance and aid ongoing efforts on design of novel therapeutic strategies.

Chaperones such as Hsp70 and Hsp100 family of proteins play an essential role in the maintenance of protein homeostasis and thus, are critical for optimal cellular growth under stress conditions [5–7]. Hsp70 primarily interacts with the exposed hydrophobic regions of unfolded proteins and prevents their aggregation. However, once these aggregates are formed, the activity of the Hsp100 family of proteins such as bacterial ClpB is required to resolubilize and reactivate aggregated proteins. The Hsp70 influences Hsp100 functions by stimulating its ATPase activity as well as recruiting it to larger aggregates [8,9]. Hsp70 is also known to aid the refolding of unfolded molecules rescued by Hsp100 from the preformed aggregates [9]. In addition to their role in protein folding, these chaperones are also essential for the regulation of cell cycle, protein trafficking and degradation [10–13].

Hsp100 family of chaperones is highly conserved among bacteria, yeast and plants however, no functional homologs have been reported in animals and humans. The Hsp100 family such as bacterial ClpB belongs to AAA^{+} (ATPases associated with diverse cellular activities) superfamily of ATPases that use energy driven by ATP hydrolysis to

* Corresponding author at: CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India.
E-mail address: deepaks@imtech.res.in (D. Sharma).

https://doi.org/10.1016/j.ijbiomac.2020.09.131
0141-8130/

DATABASE
The Journal of Biological Databases and Curation

Database, 2020, 1–9
doi: 10.1093/database/baaa003
Original article

Original article

HSPMdb: a computational repository of heat shock protein modulators

Prashant Singh, Breezy Unik, Anuradhika Puri, Gandharva Nagpal, Balvinder Singh, Ankur Gautam and Deepak Sharma*

Council of Scientific and Industrial Research-Institute of Microbial Technology, Sector 39A, Chandigarh-160036, India

*Corresponding author: CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India. Tel: 91-172-6665478; Email: deepaks@imtech.res.in. Correspondence may also be addressed to Ankur Gautam. Tel: +91-172-6665478; Email: ankurgautam@imtech.res.in

Citation details: Singh,P., Unik,B., Puri,A. *et al.* HSPMdb: a computational repository of heat shock protein modulators. *Database* (2020) Vol. 2020: article ID baaa003; doi:10.1093/database/baaa003

Received 16 November 2019; Revised 23 December 2019; Accepted 8 January 2020

Abstract

Heat shock proteins (Hsp) are among highly conserved proteins across all domains of life. Though originally discovered as a cellular response to stress, these proteins are also involved in a wide range of cellular functions such as protein refolding, protein trafficking and cellular signalling. A large number of potential Hsp modulators are under clinical trials against various human diseases. As the number of modulators targeting Hsps is growing, there is a need to develop a comprehensive knowledge repository of these findings which is largely scattered. We have thus developed a web-accessible database, HSPMdb, which is a first of its kind manually curated repository of experimentally validated Hsp modulators (activators and inhibitors). The data was collected from 176 research articles and current version of HSPMdb holds 10 223 entries of compounds that are known to modulate activities of five major Hsps (Hsp100, Hsp90, Hsp70, Hsp60 and Hsp40) originated from 15 different organisms (i.e. human, yeast, bacteria, virus, mouse, rat, bovine, porcine, canine, chicken, *Trypanosoma brucei* and *Plasmodium falciparum*). HSPMdb provides comprehensive information on biological activities as well as the chemical properties of Hsp modulators. The biological activities of modulators are presented as enzymatic activity and cellular activity. Under the enzymatic activity field, parameters such as IC_{50}, EC_{50}, DC_{50}, K_i and K_D have been provided. In the cellular activity field, complete information on cellular activities (percentage cell growth inhibition, EC_{50} and GI_{50}), type of cell viability assays and cell line used has been provided. One of the important features of HSPMdb is that it allows users to screen whether or not their compound of interest has any similarity with the previously known Hsp modulators. We anticipate that HSPMdb would become a valuable resource for the broader scientific community working in the area of chaperone biology and protein misfolding diseases. HSPMdb is freely accessible at http://bioinfo.imtech.res.in/bvs/hspmdb/index.php

PLOS GENETICS

RESEARCH ARTICLE

The yeast stress inducible Ssa Hsp70 reduces α-synuclein toxicity by promoting its degradation through autophagy

Arpit Gupta, Anuradhika Puri, Prashant Singh, Surabhi Sonam, Richa Pandey, Deepak Sharma*

G. N. Ramachandran Protein Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology, Chandigarh, India

* deepaks@imtech.res.in

OPEN ACCESS

Citation: Gupta A, Puri A, Singh P, Sonam S, Pandey R, Sharma D (2018) The yeast stress inducible Ssa Hsp70 reduces α-synuclei n toxicity by promoting its degradation through autophagy. PLoS Genet 14(10): e1007751. https://doi.org/10.1371/journal.pgen.1007751

Editor: Tricia R. Serio, University of Massachusetts Amherst, UNITED STATES

Received: March 6, 2018

Accepted: October 9, 2018

Published: October 30, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: One of the grants is from the Council of Scientific & Industrial Research, and grant number is: OLP0090. The other source of funds is from the Department of Biotechnology, Government of India and the fellowship number is BT/RLF/RE-entry/33/2010. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The mechanism underlying the role of Hsp70s in toxicity associated with intracellular accumulation of toxic protein inclusions is under intense investigation. In current study, we examined the roles of all different isoforms of yeast cytosolic Ssa Hsp70 on α-synuclein mediated cellular toxicity. The study showed that yeast cells expressing stress-inducible Ssa3 or Ssa4 as sole Ssa Hsp70 isoforms, reduced α-synuclein toxicity better than those expressing a constitutive counterpart. The protective effect of stress-inducible Ssa Hsp70s was not α-syn specific, but more general to other inclusion forming proteins such as polyQ. We show that the protective effect is not by induction of a general stress response in Ssa3 cells rather by promoting α-synuclein degradation through autophagy. The present study revealed that effect of Hsp70s was isoform dependent, and that autophagy protects Ssa3 cells from the deleterious effects of toxic protein inclusions.

Author summary

The accumulation of protein inclusions is a common feature in many neurodegenerative diseases. The Hsp70 chaperones has emerged as potent suppressor of toxicity caused by protein inclusions such as those formed of α-synuclein and polyQ. The underlying mechanism of Hsp70 mediated effect remains unclear, and it is believed that anti-aggregation activity of Hsp70 reduces protein inclusion mediated toxicity. In the present study, we examined the role of yeast cytosolic Ssa Hsp70 family of chaperones on α-synuclein mediated toxicity. We show that stress inducible Ssa Hsp70s are better than their constitutive counterparts to protect cells from both α-synuclein and polyQ toxicity. Our study reveals that the role of stress inducible Hsp70 in autophagy mediated degradation of α-synuclein protect cells from α-synuclein mediated toxicity.

Genetics: Early Online, published on April 16, 2020 as 10.1534/genetics.120.303190

The yeast Hsp70 co-chaperone Ydj1 regulates functional distinction of Ssa Hsp70s in the Hsp90 chaperoning pathway

Deepika Gaur[1], Prashant Singh[1], Jyoti Guleria[1], Arpit Gupta[2], Satinderdeep Kaur[1], Deepak Sharma[1*]

[1]Council of Scientific and Industrial Research-Institute of Microbial Technology, India, [2]Council of Scientific and Industrial Research-Institute of Microbial Technology, India, Currently in Department of Chemistry, M/C 147-75, California Institute of Technology, 1200 E. California Blvd. Pasadena, CA 91125

***Corresponding author:**E-mail: deepaks@imtech.res.in, CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India.
Phone: 91-172-6665478

Abstract:

Hsp90 assists in the folding of diverse sets of client proteins including kinases and growth hormone receptors. Hsp70 plays a major role in many Hsp90 functions by interacting and modulating conformation of its substrates before being transferred to Hsp90s for final maturation. Each eukaryote contains multiple members of the Hsp70 family. However, the role of different Hsp70 isoforms in Hsp90 chaperoning actions remains unknown. Using v-Src as an Hsp90 substrate, we examine the role of each of the four yeast cytosolic Ssa Hsp70s in regulating Hsp90 functions. We show that the strain expressing stress-inducible Ssa3 or Ssa4 and the not constitutively expressed Ssa1 or Ssa2 as the sole Ssa Hsp70 isoform reduces v-Src mediated growth defects. The study shows that although different Hsp70 isoforms interact similarly with Hsp90s, v-Src maturation is less efficient in strains expressing Ssa4 as the sole Hsp70. We further show that the functional distinction between Ssa2 and Ssa4 is regulated by its C-terminal domain. Further studies reveal that Ydj1, which is known to assist substrate transfer to Hsp70s, interacts relatively weakly with Ssa4 compared to Ssa2, which could be the basis for poor maturation of the Hsp90 client in cells expressing stress-inducible Ssa4 as the sole Ssa Hsp70. The study thus reveals a novel role of Ydj1 in determining the functional distinction among Hsp70 isoforms with respect to the Hsp90 chaperoning action.

Virus Genes (2020) 56:480–497
https://doi.org/10.1007/s11262-020-01762-4

ORIGINAL PAPER

A potent enzybiotic against methicillin-resistant *Staphylococcus aureus*

Jasjeet Kaur[1] • Prashant Singh[2] • Deepak Sharma[2] • Kusum Harjai[1] • Sanjay Chhibber[1]

Received: 6 November 2019 / Accepted: 24 April 2020 / Published online: 4 May 2020

Abstract

Staphylococcus aureus is one of the most dreadful infectious agents, responsible for high mortality and morbidity in both humans and animals. The increased prevalence of multidrug-resistant (MDR) *Staphylococcus aureus* strains has limited the number of available treatment options, which calls for the development of alternative and effective modalities against MDR *S. aureus*. Endolysins are bacteriophage-derived antibacterials, which attack essential conserved elements of peptidoglycan that are vital for bacterial survival, making them promising alternatives or complements to existing antibiotics for tackling such infections. For developing endolysin lysin-methicillin-resistant-5 (LysMR-5) as an effective antimicrobial agent, we evaluated its physical and chemical characteristics, and its intrinsic antibacterial activity against staphylococcal strains, including methicillin-resistant *Staphylococcus aureus* (MRSA). In this study, we cloned, expressed, and purified LysMR-5 from *S. aureus* phage MR-5. In silico analysis revealed that LysMR-5 harbors two catalytic and one cell wall-binding domain. Biochemical characterization and LC–MS analysis showed that both catalytic domains were active and had no dependence on divalent ions for their action, Zn^{2+} exerted a negative effect. The optimal lytic activity of the endolysin was at 37 °C/pH 7.0 and in the presence of ≥ 300 mM concentration of NaCl. Circular dichroism (CD) demonstrated a loss in secondary structure with an increase in temperature confirming the thermosensitive nature of endolysin. Antibacterial assays revealed that LysMR-5 was active against diverse clinical isolates of staphylococci. It showed high lytic efficacy against *S. aureus* ATCC 43300, as an endolysin concentration as low as 15 µg/ml was sufficient to achieve maximum lytic activity within 30 min and it was further confirmed by scanning electron microscopy. Our results indicate that rapid and strong bactericidal activity of LysMR-5 makes it a valuable candidate for eradicating multidrug-resistant *S. aureus*.

Keywords Endolysin · Enzybiotics · Staphylococcal infections · Bacteriophage · Alternative therapies

Introduction

Staphylococcus aureus has long been considered as a major nosocomial pathogen, associated with various infections ranging from skin and soft tissue to other insidious infections like meningitis, osteomyelitis, and pneumonia [1, 2]. Further, rapid development of resistance by *S. aureus* strains against broad spectrum of antibiotics has posed additional limits to treat these infections. Besides controlling misuse of antibiotics, development of novel antimicrobials that are refractory to resistance generation is essential to eradicate MDR pathogens [3–6]. In recent years maximum attention has been directed towards bacteriophage endolysins, which are being investigated as a valuable alternative antimicrobial strategy for containing multidrug-resistant Gram-positive bacteria [7–10]. Endolysins are produced by phages in a bid to escape from their host at the end of phage lytic cycle, by targeting specific bonds in peptidoglycan, resulting in osmotic lysis of the cell and release of phage progeny [11–13]. These hydrolases can rapidly kill targeted Gram-positive bacteria even when applied exogenously, since they are able to make direct contact with the cell wall carbohydrates and peptidoglycan layers, which are inaccessible

Edited by Andrew Millard.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11262-020-01762-4) contains supplementary material, which is available to authorized users.

✉ Sanjay Chhibber
sanjaychhibber8@gmail.com

[1] Department of Microbiology, Panjab University, Chandigarh, India

[2] CSIR-Institute of Microbial Technology, Chandigarh, India