IDENTIFICATION OF AMINO ACIDS INVOLVED IN THE MOONLIGHTING FUNCTIONS OF HTPB, THE *LEGIONELLA PNEUMOPHILA* CHAPERONIN

by

Karla Valenzuela

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2015

To my husband Francisco

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Chaperonin 60s (Cpn60s) and their cognate co-chaperonin 10s (Cpn10s) are highly conserved housekeeping proteins that provide favorable conditions for the correct folding of other proteins. The Cpn60/10 of *Escherichia coli* (known as GroEL/GroES) has been widely studied. Based on these studies it is known that this molecular protein folding machinery is constituted by two GroEL heptameric rings each forming a folding chamber, and a GroES heptameric cap that keeps unfolded proteins inside the chamber. However, bacterial Cpn60s have also evolved additional functions independent of protein folding. The biochemical term recently adopted to describe multifunctional proteins that have a primary well-known function is "moonlighting", which colloquially refers to having a secondary job (usually at night) in addition to one's main day job. Some known moonlighting functions of Cpn60s are cell-signaling, proteolytic and toxigenic activity. Unlike GroEL, the Cpn60 of the intracellular bacterial pathogen *L. pneumophila* (HtpB) reaches the cytoplasm of infected host cells and has been implicated in host cell invasion, microfilament reorganization, mitochondria recruitment and cell-signaling. I hypothesized that HtpB must interact with a cytoplasmic protein in the host cell to exert its unique moonlighting functions, and that functional gains are due to substitutions in key amino acid positions. Therefore, a yeast two hybrid (Y2H) screening using a human cDNA library was performed to find an interaction partner specific for HtpB. In so doing, the human homolog of protein ECM29 (hECM29), which does not interact with GroEL, was found. In addition, putative key amino acids involved in HtpB moonlighting functions were predicted using the ET bioinformatics method. Mutational analysis of the predicted moonlighting-related amino acids led to the identification of 4 residues involved in the HtpB-hECM29 interaction, namely K298, N507, H473 and K474. Since hECM29 couples the 26S proteasome to molecular motors, endocytic vesicles and the endoplasmic reticulum, I propose that exploitation of the HtpB-hECM29 interaction is a previously undescribed strategy used by *L. pneumophila* to alter protein degradation and vesicular trafficking in the host cell. Additionally, I suggest that although HtpB is a conserved essential protein, it has substitution-prone amino acid positions that have accumulated mutations resulting in the acquisition of novel functions that support the intracellular lifestyle of *Legionella pneumophila*.

LIST OF ABBREVIATIONS USED

°	degree
%	percent
~	approximately
16S RNA	16S ribosomal RNA
APS	ammonium persulfate
ATP	adenosine triphosphate
BCYE	buffered charcoal yeast extract
bp	base pair
BSA	bovine serum albumin
BYE	buffered yeast extract
C	Celsius or cytosine
cDNA	complementary DNA
Cpn60	chaperonin 60
ddH_20	double distilled water
DDO	double drop out
DDO/X/A	double drop out supplemented with aureobasidin A and X-alpha-Gal
DMSO	dimethylsufoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
ER	endoplasmic reticulum
EtBr	ethidium bromide
g	centrifugal force
h	hour
hECM29	Human homolog of the yeast protein ECM29
HtpA	high-temperature protein A
HtpB	high-temperature protein B
IL	interleukin
kDa	kilo Dalton
kV	Kilovolts

LCV	*Legionella*-containing vacuole
LiAc	Lithium acetate
mRNA	Messenger ribonucleic acid
min	minute
NEB	New England Biolabs
nm	nanometre
OD	optical density
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	Polyethylene glycol
PNP	p-nitrophenyl-α-d-galactoside
QDO	quadruple drop out
QDO/X/A	quadruple drop out supplemented with aureobasidin A and X-alpha-Gal
SD	yeast minimal media
SDS	sodium dodecyl sulfate
sec	second
SOC	super optimal broth supplemented with glucose
ssDNA	single-stranded DNA
TAE	Tris-acetate-EDTA buffer
TE	Tris-EDTA buffer
TEMED	N,N,N´,N´-tetramethylethylenediamine
Y2H	yeast two hybrid
YPD	Yeast Peptone Dextrose media
α	alpha
β	beta
γ	gamma
δ	delta
ε	epsilon

ACKNOWLEDGEMENTS

There are many people I would like to acknowledge for their personal and academic advice during these 3 years.

First I would like to express my sincere gratitude to my supervisor Dr. Rafael Garduno. Thanks for your continuous support and patience. I know sometimes I can be difficult and stubborn, but we managed to be very good friends. I am very grateful because being away from home can be difficult sometimes, but you and your wife Elizabeth were always there to support me, give advice and cheer me up with wonderful music. I see you like family**.**

I would like to also thank the members of my committee, Dr. Christine Barnes and Dr. John Archibald for providing me with very helpful advice and feedback. Special mention to Dr. John Rohde who helped me with the yeast experiments and was always willing to answer my questions, thanks for letting me use your black board to draw my crazy ideas! Furthermore, I like to thank Dr. Andrew Roger for taking the time to read my thesis and serve as external examiner.

I also want to thank Dr. Gabriel Moreno-Hagelsieb for letting me spend some time in his lab at Wilfrid Laurier University, during the summers of 2013 and 2014, learning bioinformatics. Thanks for your patience, all the laughing and the many coffees we shared talking about evolution.

I would like to thank "my favorite master student in the Garduno lab", Peter Robertson for the endless jokes, readily available ice cream during difficult "nothing is working" times, and for constant support. The Garduno lab was just Peter and I for most of the time, but at some point Wolim Lee showed up to do his honours project, and became a wonderful friend as well. Thanks Lee for being the third musketeer. I was also very lucky because later on, after Lee joined the lab, my best Chilean friend, Cristian Oliver ("el Viejo") came to do an internship at the Garduno lab. We both share a deep interest in science, and became friends long time ago in Chile. Thanks Viejo, for helping me in every aspect of my life and being always there, even when you are thousands of kilometres away.

I couldn't forget to thank my dear friend Don for every "Friday at four", the endless caipirinias and, most importantly, for his invaluable help proofreading this thesis. Also thanks to Susan, for all the help during these years, and for taking time out of your very busy days to chat with me.

Lastly, but not least I want to thank my beloved husband Francisco, for being there for me all the time. For being supportive and understanding when I stayed late at work and for always believing in me.

CHAPTER 1 INTRODUCTION

1.1 *LEGIONELLA PNEUMOPHILA*, AN ACCIDENTAL PATHOGEN

1.1.1 History and biology

Legionella pneumophila is a Gram-negative, aerobic, pleomorphic, flagellated bacterium [1] ubiquitously found within freshwater environments [2]. However, *L. pneumophila* is not strictly a free-living bacterium; rather, it is a facultative intracellular parasite of protozoan species (e.g., amoebae),[3] or else persists in extracellular environments by forming biofilms.[4,5]. *L. pneumophila* was first identified as an opportunistic human pathogen after an outbreak of acute pneumonia, so called Legionnaires' disease, that occurred at an American Legion convention in Philadelphia, PA in 1976 [6]. The bacterium is also the causal agent of a mild febrile flu-like illness called Pontiac fever [7]. It is important to note that *L. pneumophila* is only accidentally a human pathogen; the infection usually occurs by inhalation of contaminated water droplets generated by human-made aquatic systems such as cooling towers, evaporative condensers and spas [8-10]. Person-to-person transmission has not been reported. Typically, patients susceptible to infection are immunocompromised or have one or more risk factors, such as chronic lung disease, diabetes or smoking [11]. Although *L. pneumophila* is an opportunistic pathogen, it is still a significant cause of nosocomial and community-acquired pneumonia. Indeed, the number of cases reported to the Centers for Disease Control and Prevention in the United States increased from ~1000 in 2000 to ~4000 in 2011, with an average mortality rate of 12% [12].

1.1.2 Infection of host cells

L. pneumophila successfully replicates within its natural protozoan hosts and is also able to colonize human alveolar macrophages. In nature, the bacterium is ingested by amoebae, escapes digestion, replicates to high numbers and then returns to the environment after killing the host [13]. Similarly, when *Legionella* is phagocytosed by macrophages, it escapes the endocytic/lysosomal pathway and recruits mitochondria, ribosomes and endoplasmic reticulum (ER) from the host cell to transform its phagosome into a specialized vacuole called the *Legionella*-containing vacuole or LCV (more details

in section 1.1.3). The pathogen replicates to high numbers within the LCV and kills its host cell by inducing pore formation and membrane lysis [14].

Virulence of *L. pneumophila* mainly depends on the presence of a functional Dot/Icm (for defect in organelle trafficking/intracellular multiplication) type IV secretion system (T4SS). Indeed, *dot/icm* mutants are non-virulent and fail to evade fusion with lysosomes, establish the LCV and multiply intracellularly [15,16]. Genetic analysis of such mutants led to the identification of 24 different genes that comprise the Dot/Icm complex [17,18]. Although the crystal structure of this secretory system has not been reported, DotC, DotD, and DotH have been found to be essential for assembly of the Dot/Icm ring-shaped core complex in both the inner and outer membranes[19]. This multiprotein complex delivers an exceptionally large number of *L. pneumophila* effector proteins (nearly 300) into the host cell cytosol, modulating many signalling and metabolic pathways [20-22]. Interestingly, considerable redundancy has been observed among effectors that target similar host processes; that is, mutations in a single secreted effector rarely lead to detectable phenotypes [23,24].

1.1.3 Lysosomal escape and establishment of a replication niche

A common attribute of intracellular pathogens is their ability to evade the endocytic pathway and scape of fusion of the lysosome to the phagosome. Normally, following phagocytosis, the bacteria contained in phagosomes are digested. In order for this to happen, phagosomes must undergo a series of maturation stages, first fusing with early and late endosomes, and finally becoming phagolysosomes following fusion with lysosomes. During this process, phagosomes acquire different protein markers in their limiting membranes, and their contents are gradually acidified leading to bacterial degradation [25]. The small GTPase Rab5 [26] and early endosome antigen 1 (EEA1) [27] are markers of the early phagosome; as maturation continues, association with the V-ATPase induces progressive acidification caused by proton pumping. In the late phagosome, Rab5 and EEA1 are replaced by Rab7 [28] and the lysosomal-associated membrane proteins (LAMPs) [29]. The process culminates with the formation of the phagolysosome, which is rich in V-ATPases, cathepsins and hydrolases [30]. Details of the phagocytic pathway are reviewed by Kinchen and Ravichandran [31].

L. pneumophila, subverts the endocytic/lysosomal pathway at very early stages. *Legionella*-containing phagosomes lack maturation markers such as Rab5 [32], LAMP-2 and cathepsin D [33]. Therefore, this pathogen impairs fusion of the early phagosome with endosomes and lysosomes, thus avoiding phagosome acidification during infection [34,35]. Shortly after entry, *L. pneumophila* remodels the phagosome in order to establish the LCV [36]. First, early secretory vesicles that transit between the ER and the Golgi are intercepted and their content is incorporated into the LCV creating an ER-like organelle [36]. This process is mediated by Dot/Icm effectors such as SidM and LidA that promote binding of Rab1, a regulator of ER traffic, to the LCV [37]. Similarly, the effector RalF recruits the host trafficking small GTPase Arf1 to the LCV [38]. The vesicle-trafficking protein SEC22b is also targeted to the LCV facilitating the transport and fusion of ER-derived vesicles with the LCV [39]. Second, mitochondria tightly associate with the LCV [40,41]. The advantage of this association is still unclear and no specific Dot/Icm effectors have as yet been associated with this phenomenon. However, since *dot/icm* mutants fail to recruit mitochondria [15], it is thought that the factor(s) involved could be Dot/Icm secretion substrates. The last step before *L. pneumophila* replication commences, is the recruitment of ribosomes and rough ER to the LCV[41,42]. Although the mechanism of ribosome recruitment is unknown, it is a Dot/Icm-dependent process [15,41].

1.2 THE CHAPERONIN 60 FAMILY

Molecular chaperones are proteins that assist the folding of other proteins [43]. A large percentage of these proteins are termed heat shock proteins (Hsp), as they are upregulated under sudden temperature increase or other stress conditions in which the concentrations of denatured (aggregated) proteins increase [43]. These chaperones are usually assigned names on the basis of their molecular mass (e.g., Hsp10, Hsp60, Hsp70 and Hsp90). Among them, Hsp60 chaperones are the most widely distributed chaperones as they are present in all kingdoms of life [43] except for some members of the genus *Mycoplasma* [44]. The Hsp60 chaperones are members of a larger family of highly conserved proteins called the chaperonin family, which is divided in two main groups [45]. Group I chaperonins are found in eubacteria, and eukaryotic organelles originated from endosymbiotic microbes

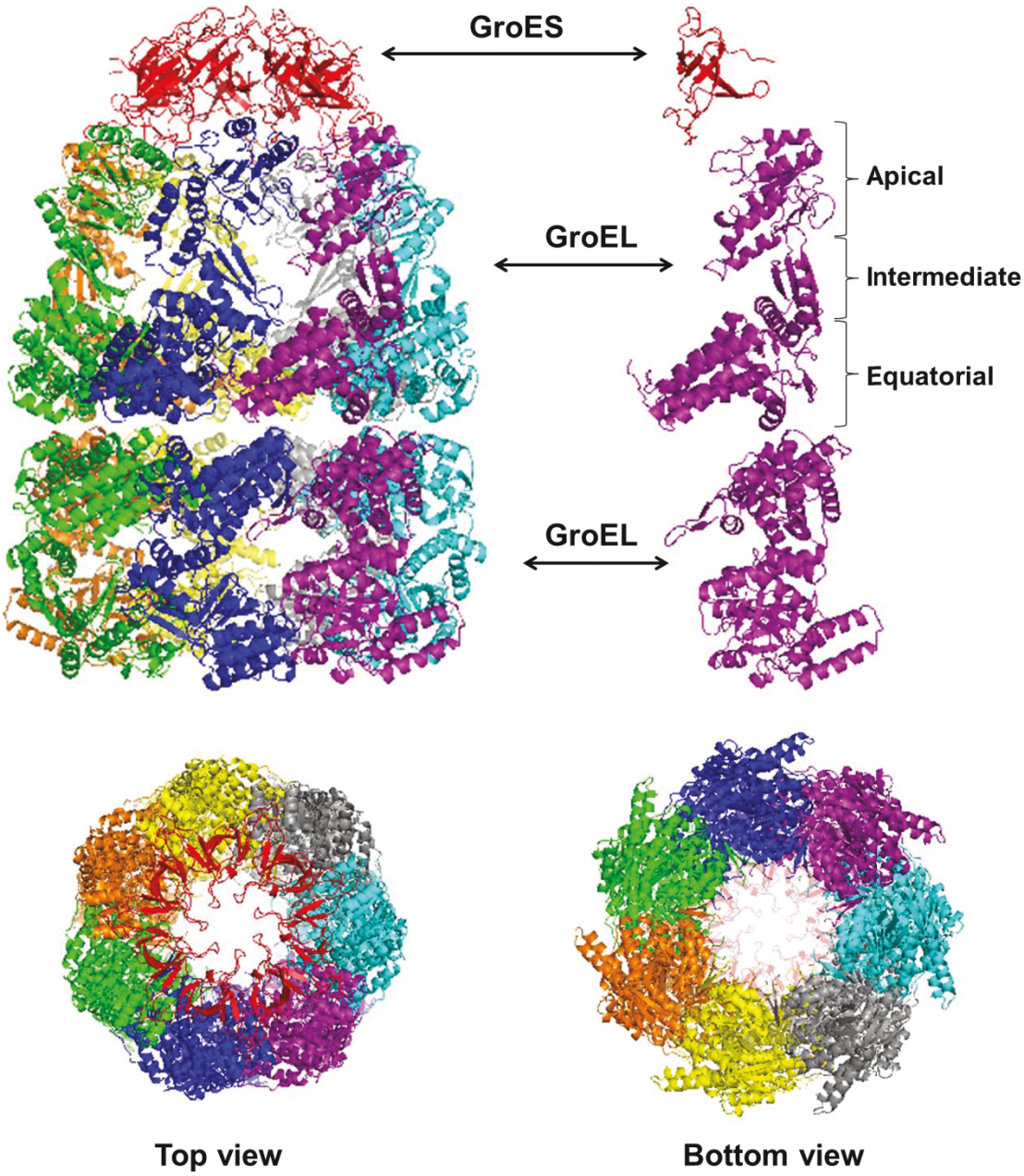
like mitochondria, chloroplasts, and hydrogenosomes. Group I chaperonins are also known as chaperonin 60s (Cpn60s) and are built of two rings each composed by seven identical subunits [46]. Group II chaperonins occupy the eukaryotic and archaea cytosol and their structure is more complex. Thermosomes, which are found in archaea, are formed by two octa- or nonameric rings built by one, two or three different chaperonin subunits. The most complex member of this group is the eukaryotic chaperonin, which is known as chaperonin containing TCP-1 (CCT) or TRiC and is formed by two rings composed of 8 different subunits [45].

This thesis will focus only on bacterial Cpn60s, which are very well known for their essential protein-folding function. The protein-folding Cpn60 nanomachine consists of 14 identical Cpn60 (60 kDa) subunits that oligomerize, forming two stacked rings (to 800–1000 kDa) [47]. Each ring defines a molecular chamber, inside which favorable conditions are provided for the folding of nascent polypeptides or the re-folding of denatured proteins (called substrates), in an ATP-mediated process (see section 1.2.1 below). To properly function, Cpn60s require the help of small 10-kDa heat shock proteins, or Hsp10 chaperones, also named co-chaperonin 10 (Cpn10s) that forms heptameric oligomers [48,49]. Group I chaperonins and their cognate co-chaperonins have been termed differently, depending on which organism they are from. For the purpose of this work, the chaperonin/co-chaperonin sets will be referred as Hsp60/Hsp10 in mitochondria, GroEL/GroES in *Escherichia coli*, HtpB/HtpA in *L. pneumophila,* and generically Cpn60/Cpn10 in any other bacteria.

1.2.1 The GroEL/GroES folding nanomachine

The GroEL/GroES protein-folding machinery is the best characterized group I chaperonin/co-chaperonin complex. The GroEL crystal structure has been resolved at different resolutions using X-ray crystallography, and the mechanism of how this multimer assembles and how it performs under physiological conditions has been elucidated to a great extent [50-52]. The GroEL monomer consists of three domains: apical, intermediate and equatorial (Fig. 1) [47]. The apical domain interacts with unfolded substrates and GroES [53] as explained below, whereas the equatorial domain houses an ATP binding pocket [54]. These two domains are connected by the intermediate domain that

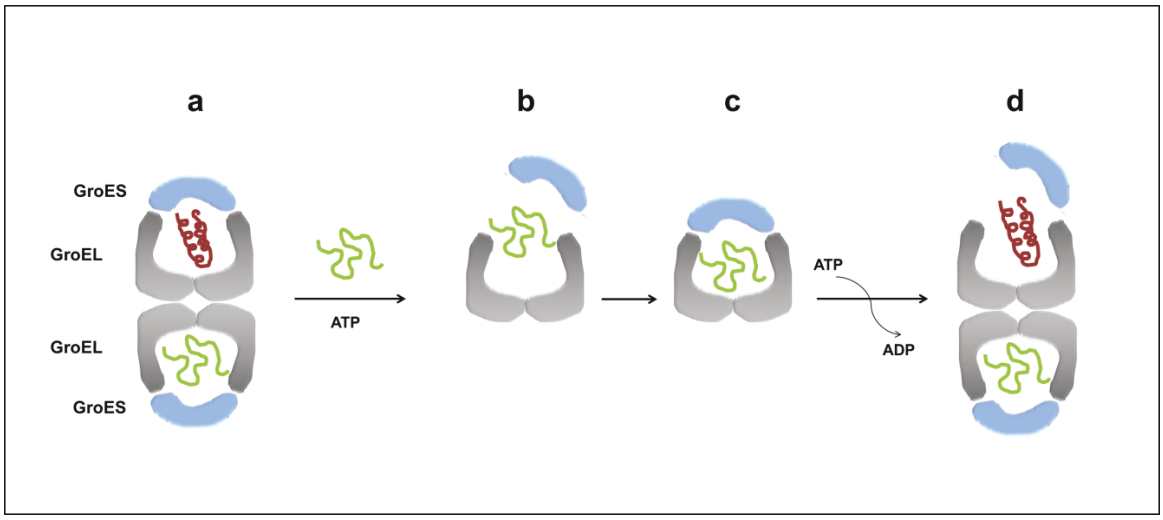
Figure 1. Structure of the asymmetric GroEL/GroES bullet complex. Diagram showing the crystal structure of the chaperonin complex GroEL(14)/GroES(7) represented as ribbons. Top (GroEL ring where the GroES ring is bound) and bottom (GroES free ring) views of the complex are shown at the bottom of the figure. Individual monomers of GroEL (purple) and GroES (red) are separately shown on the right side of the figure. GroEL domains are labeled on the top individual monomer structure (purple). Complete structure (ID: 1AON) was retrieved from the protein data bank (PDB) and colored using the protein viewer PyMol. The GroEL folding chamber is formed by fourteen identical GroEL monomers arranged in a symmetric fashion as two back-to-back seven-member rings (for clarity of viewing, the various colors indicate individual GroEL monomers). Additionally, seven identical monomers of GroES (red) interact with the chamber forming a domed structure. The ring that interacts with GroES (upper ring in the figure) undergoes conformational changes after ATP binding that lead to elevation of the apical domains increasing the volume of the central cavity.


GroES
GroEL
GroEL
Apical
Intermediate
Equatorial
Top view
Bottom view

allows for major conformational changes of the apical domain induced by the binding and hydrolysis of ATP [55] to the equatorial domain. The equatorial domains also experience subtle structural changes during the folding process [55]. The GroEL oligomer consists of two rings stacked back to back, each containing seven GroEL subunits, forming a barrel-like structure called apo-GroEL [47]. During the protein-folding process, the GroES ring covers one end of apo-GroEL (i.e., attaches to one of the two GroEL rings) forming an asymmetric configuration referred to as the bullet complex (Fig. 1) [53]. When two GroES rings simultaneously bind to both ends of apo-GroEL, the resulting symmetric configuration is referred to as the football-shaped complex [56,57].

The sequence of events that lead to the proper folding of protein substrates under physiological conditions is currently being re-examined. Previously, it had been thought that only one GroEL ring was used at any given time (i.e., only one end of apo-GroEL would be engaged by substrate and the GroES ring) and that alternations between apo-GroEL and the bullet complex shaped the folding reaction cycle [46]. However, recent reports indicate that the football-shaped complex is the protein folding functional form [57,58]. Although the football-shaped structure was described long ago by Llorca et al. [59], who showed by electron microscopy that a symmetric conformation of GroEL/GroES can contain substrate in both GroEL rings, only now it has received attention. In fact, the crystal structure of the football shaped complex was just recently solved by Fei et al. and Koike-Takeshita et al. [56,58]. They have proposed a model in which the two rings of GroEL perform folding in parallel, rather than alternately. Figure 2 shows a simplified version of the model proposed by Fei et al. in which both GroEL rings are occupied by substrate virtually all the time (Fig. 2 a and d). However, another study showed that equivalent amounts of football-shaped and bullet-shaped complexes can be found under folding conditions suggesting that these two structures alternate throughout the protein folding cycle [60]. Current knowledge of the folding cycle can be summarized as follows: the non-native polypeptide, with exposed hydrophobic surfaces, binds to apical domains in the GroEL ring [61] and then binding of ATP [62] to the equatorial domains of that ring induces displacement of the substrate into the hydrophilic chamber [55]. Following binding of GroES, the substrate attempts to fold. ATP hydrolysis induces release of GroES and the now properly-folded substrate [63](Fig. 2). This simultaneously happens in both rings

Figure 2. The GroEL/GroES mediated protein-folding process. Diagram showing the folding of protein substrates by GroEL/ES according to the "football symmetric complex" model. The apo-GroEL 14-mer (in grey) interacts with two GroES heptamers (in blue) and two protein substrates (green= unfolded, brown= folded) at the same time (a). For simplicity, the folding of only one peptide inside one GroEL ring is shown in (b) and (c). The GroEL ring opens and the apical domain of each monomer interacts with the unfolded substrate (b). Binding of ATP followed by closure of the chamber by the co-chaperonin GroES triggers productive folding of the substrate inside the now encapsulated and hydrophilic cavity (c). ATP hydrolysis allows release of GroES and the native folded protein from the ring (d). Adapted from Fei et al. 2014 [58].


a
b
c
d
GroES
GroEL
GroEL
GroES
ATP
ATP
ADP

(football complex), although intermediate stages where only one ring is occupied by substrate and GroES (bullet complex) are also part of the cycle [57]. It has been shown that this cycle occurs more rapidly than previously believed. Thus, a substrate undergoes various rounds of folding before the functional native stage is reached [57,58].

1.3 CPN60s ARE MOONLIGHTING PROTEINS

1.3.1 Introduction to moonlighting

The term "moonlighting" is colloquially defined as the practice of having a second job, usually at night, in addition to the regular employment (day job). This term was first applied in the context of protein multi-functionality by Campbell and Scanes who reported that somatostatin and growth hormone releasing hormone (GH-RH), two well characterized endocrine peptides, also exhibited immunological activity [64]. Later, Constance J. Jeffery defined a moonlighting protein as one having one or more independent biological functions, in addition to its canonical function [65]. The term moonlighting excludes proteins whose multi-functionality is due to gene fusions, multiple splice variants, proteolytic fragmentation or post-translational modification. Proteins that have multiple cellular roles as a manifestation of the same function in different locations, or utilize different substrates maintaining a single activity, are also not considered moonlighting proteins [65]. Over 200 proteins have been experimentally verified to be moonlighting proteins, enough to justify the creation of a database named MoonProt [66], which contains information about the sequences, structures and functions of moonlighting proteins.

Insights into the mechanisms by which a protein can perform multiple functions have in some cases been revealed using X-ray crystallography. Evidence shows that such proteins can have different binding sites for different substrates. For example, I-AniI, a mitochondrial DNA endonuclease that is also involved in RNA splicing, binds DNA and mRNA through two different molecular surfaces [67]. Another mechanism for moonlighting is provided by modification of active site characteristics through conformational changes that expose a different set of amino acids; this strategy is used by the *Sulfolobus tokodaii*

fructose-1,6-bisphosphate aldolase/phosphatase to catalyse two different reactions in gluconeogenesis [68].

How moonlighting proteins acquire additional functions is less well understood. One hypothesis is that moonlighting proteins represent a "transitional stage" between the original function and a novel evolving function. Whether the first or the second function remains, depends on environmental parameters [69]. However, this hypothesis implies that at some point a moonlighting protein will become a single-function protein again. Nevertheless, if the canonical function of a given moonlighting protein is essential, it cannot be lost, as is the case of moonlighting members of the Cpn60 family. These proteins are highly conserved and its protein-folding function is essential in most cases [70]. A possibility to solve this conflict is gene duplication. Thus, one copy of the gene can evolve a new function, while the other retains the canonical function [71]. This seems to be the case in the bacterial Cpn60 family, where paralogy is common [72].

Mario Fares has proposed a model for the evolution of moonlighting [69]. He hypothesises that a moonlighting protein was the last common ancestor of all variants for a given protein family and that the moonlighting ancestor encoded all the functions observed in the population. He also suggests that "moonlighting declined as species diversified and very likely, such decline may have had an important role in speciation" [69]. Therefore, if moonlighting functions have emerged from a common ancestor, the same function will be observed among phylogenetically related moonlighting proteins. There is evidence supporting this hypothesis. For example, studies in the enolase family, formed by glycolytic enzymes conserved among prokaryotes and eukaryotes that are also involved in bacterial pathogenesis and cancer-cell invasion [73,74], showed that moonlighting functions of this protein are conserved, even among phylogenetically distinct eukaryotic organisms [75,76]. However, more research is required in order to more fully understand the origins of protein moonlighting.

1.3.2 Moonlighting functions of bacterial Cpn60s

Moonlighting proteins are often identified by accident while studying a certain effect that turns out to be associated with a previously characterized protein. That is the case for the

first reported moonlighting function attributed to a Cpn60. Ensgraber and Loos were analysing a protein responsible for the binding of *Salmonella typhimurium* to intestinal mucus, and eventually determined that it belonged to the Cpn60 family [77]. Since then, many other moonlighting Cpn60s that act as adhesins have been identified and a large variety of additional functions have been described [78]. Special attention has been captured by bacterial chaperonin moonlighting functions, as some of them are related with virulence and pathogenesis [73].

Moonlighting Cpn60s of bacterial pathogens, are often exported to the bacterial cell surface, and have also been reported to be secreted proteins. For example, in *Mycobacterium tuberculosis,* Cpn60.2 is found exposed on the bacterial cell surface where it acts as a ligand that recognizes CD43 on the cell surface of macrophages facilitating bacterial binding/uptake [79]. Similarly, the Cpn60 of *Helicobacter pylori* functions as a cell surface adhesin [80], iron-binding siderophore, and it is also secreted [81].

Secreted and surface-exposed chaperonins are also able to act as signalling molecules in eukaryotic cells, which constitute an emerging role for chaperonins in bacterial pathogenesis [73]. For instance, the two chaperonins of *M. tuberculosis* (Cpn60.1 and Cpn60.2) stimulate human monocytes to synthesise proinflammatory cytokines [82] and a Cpn60.1 mutant fails to produce granulomatous inflammation in mice [83]. This suggests that *M. tuberculosis* chaperonins contribute to maintenance of a selective inflammatory stage beneficial for bacterial persistence in the lung [84]. Similarly, the secreted fraction of the Cpn60 of *Actinobacillus actinomycetemcomitans*, causative agent of periodontal disease, induces epithelial cell death through activation of p38 MAP kinases, and tumor necrosis factor alpha (TNF-α) upregulation [85,86].

Some bacterial chaperonins have even more surprising moonlighting functions; e.g. the Cpn60 of a symbiotic strain of *Enterobacter aerogenes* that lives in the saliva of antlions (larvae stage of *Myrmeleon bore* that feed on other insects), is also a potent insect neurotoxin [87]. Interestingly, only eleven amino acids are different between GroEL and the Cpn60 of symbiotic *E. aerogenes*, and mutational analysis showed that only four of these eleven amino acids (V100, N101, D338 and A471) are involved in the toxigenic activity of this protein. Indeed, mutations of these four residues in GroEL (which is non-toxic)

turned it into a potent insect neurotoxin [87]. A similar example is provided by the Cpn60.2 of *Mycobacterium leprae*, which also functions as a protease. It was reported that the cytosol from macrophages transfected with cDNA of the *M. leprae* Cpn60.2 displayed higher proteolytic activity compared to cytosol from control macrophages. Later on, the proteolytic activity of purified Cpn60 was confirmed and mutational analyses identified the amino acids T375, K409, and S502 as essential for proteolysis [88].

1.4 HTPB, A MOONLIGHTING CHAPERONIN

The chaperonin of *L. pneumophila*, here referred to as high temperature protein B (HtpB), was initially cloned and expressed by Hoffman et al. who reported that this chaperonin forms part of the *htpAB* operon and was over expressed upon heat shock [89]. The same study showed that HtpB cannot complement a GroEL deficient *E. coli* strain thus suggesting that HtpB is not functionally equivalent to GroEL [89]. However, an honours student in the Garduno laboratory recently demonstrated that HtpB is able to re-fold denatured malate dehydrogenase substrate, thus it performs as a chaperonin (Wolim Lee, Investigating the protein-folding function of high temperature protein B (HtpB), the putative chaperonin of *Legionella pneumophila,* honours thesis, Dalhousie University, April 2014). Additionally, as summarized in a recent review[90] HtpB has a variety of moonlighting functions that will be briefly presented here.

Since its initial characterization HtpB has been an intriguing protein. Early evidence suggested that HtpB was located in the membrane of *L. pneumophila* [91,92], periplasm and also in OMVs [93] derived from this pathogen. The ectopic localization of HtpB was first related to virulence by Hoffman et al. (1990) who found by immunofluorescence that HtpB localized to the outer-membrane of the Svir virulent strain of *L. pneumophila* but not on the salt-tolerant avirulent derivative strain [94]. Additionally, HtpB upregulation is not only induced by heat shock but is also differentially increased in virulent and non-virulent strains upon bacterial contact with L929 cells [95] suggesting a role during interaction of *L. pneumophila* with the host cell. In fact, is has been demonstrated that HtpB is involved in adhesion to the host cell. Indeed, Hsp60-specific antibodies inhibit

the adherence and invasiveness of *L. pneumophila* [96], and beads coated with HtpB are efficiently taken up by HeLa and CHO cells [96,97]. Therefore, over-expression under infection conditions would directly benefit uptake of *L. pneumophila* by the host.

Although it is not clear by which mechanism HtpB is translocated to the outer-membrane, there is evidence indicating that such a mechanism is not present in *E. coli* since recombinant HtpB expressed in this host accumulates in the cytoplasm [91]. Similar observations are reported in *dot/icm L. pneumophila* mutants where HtpB accumulates in the periplasm as shown by immuno-electron microscopy [98] suggesting that HtpB translocation from the periplasm to the outer-membrane is depending on a functional Dot/Icm secretion system.

Surprisingly, HtpB is profusely released into the *Legionella*-containing vacuole as shown by immuno-gold labelling [91]. Translocation experiments using HtpB fused to the CyaA (adenylate cyclase from *Bordetella pertussis*) domain that triggers increased levels of cAMP upon contact with host cell cytoplasm's calmodulin demonstrated that HtpB contacts the cytosol of *Legionella*-infected CHO and U937-derived human macrophages cells inducing high levels of cAMP [99]. HtpB has been immunolocalized free in the cytoplasm and also associated to the cytoplasmic side of phagosome membranes similar to Dot/Icm effectors [95,100].

Since HtpB is an essential protein, determination of the role of HtpB in the pathogenesis of *L. pneumophila* has been difficult to achieve. Experiments using latex microbeads coated with recombinant or purified HtpB, but not GroEL-coated beads, have shown that HtpB directly induces uptake by HeLa cells [96]. It has also been reported that beads coated with HtpB, but not GroEL-coated beads, induce alterations in the actin cytoskeleton of CHO cells similar to the cytoskeletal rearrangements observed in *L. pneumophila*-infected cells [97]. The same study also revealed that phagosomes containing HtpB-coated beads, but not GroEL-coated beads, closely associate with mitochondria. In fact, mitochondria are still attached to phagosomes purified from infected cells, suggesting direct interaction of HtpB with a mitochondrial protein [97,100]. As mentioned in section 1.1.3, mitochondria interact with the LCV and HtpB appears to be the only protein that has been described to contribute to this association. Recently, the interaction of HtpB

with mitochondrial Hsp10 was reported and the author suggests that this interaction could be relevant to the association of mitochondria with the LCV [101]. In addition, Fernandez-Moreira et al. reported that OMVs of *L. penumophila*, but not *E. coli* K-12 vesicles, attached to beads inhibit phagosome-lysosome fusion in human macrophages [102]. Similarly, HtpB coated beads delay phagosome maturation [97], suggesting that HtpB present in OMVs [93] could account for the results observed by Fernandez-Moreira [102].

1.5 RATIONALE AND OBJECTIVES

Bacterial chaperonin 60s are highly conserved housekeeping proteins, essential in most cases, which actively participate in the folding of other proteins [70]. However, bacterial chaperonins have also evolved functions independent of protein folding; these are defined as moonlighting functions [65]. Studies on the amino acids involved in some of these moonlighting functions have revealed that small changes in Cpn60 amino acid sequence can introduce novel biological activities (see section 1.3.2). HtpB has intriguing virulence-related moonlighting activities as described in section 1.4 and these biological functions are not shared by GroEL. Unlike other bacteria, *L. pneumophila* has just one copy of the chaperonin 60 gene and it has been demonstrated to be essential [103]. This imposes selective pressure to preserve HtpB's protein-folding function, but somehow it has been able to acquire additional functions that seem to support the intracellular lifestyle of *L. pneumophila*.

Based on this information I formulated the following hypotheses.

Hypothesis #1. HtpB must interact with a cytoplasmic protein in the host cell to exert some of its unique functions.

Hypothesis #2. Substitution in key amino acids, which occurred in HtpB, but not in GroEL, resulted in the acquisition of moonlighting functions by HtpB.

Accordingly, this work has been aimed at testing the above hypotheses through two specific experimental objectives:

(i) find, using a yeast two-hybrid screening, an HtpB interaction partner in human cells that does not interact with GroEL and can thus be used as an experimental moonlighting model for HtpB.

(ii) identify potential functionally-important amino acids in HtpB using the evolutionary trace method and evaluate them by systematic mutational analysis.

Additionally, I considered that phylogenetic analyses of HtpB are required to understand the multifunctional nature of this protein. Therefore, the phylogenetic relationships within the bacterial chaperonin 60 family and between moonlighting chaperonins (including HtpB) were assessed.

CHAPTER 2 MATERIALS AND METHODS

2.1 MICROBIAL STRAINS AND GROWTH CONDITIONS

2.1.1 *Legionella pneumophila*

The *L. pneumophila* strain JR32 used in this study was a restriction-deficient, streptomycin-resistant derivative of the clinical isolate strain Philadelphia I [104]. JR32 was routinely grown on (N-[2-acetamido]-2-aminoethane-sulfonic acid) – buffered charcoal yeast extract (BCYE, appendix B) agar at 37 °C in a humid incubator for 3 days. Liquid cultures were inoculated in (N-[2-acetamido]-2-aminoethane-sulfonic acid) – buffered yeast extract (BYE, appendix B) broth and grown overnight at 37 °C with shaking (200 rpm). Frozen stocks were maintained at -80 °C in freezing medium (appendix B).

2.1.2 *Escherichia coli*

The *E. coli* strains used in this study were DH5α (*F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1*) and JM109 *(endA1 recA1 gyrA96 thi hsdR17 (rk-, mk+) relA1 supE44 Δ(lac-proAB) [F′ traD36 proAB laqI^{q}ZΔM15]*) [105]. Both strains were grown on Lysogeny broth (LB, appendix B) or LB solidified with 2% agar at 37 °C. Ampicillin (100 µg/mL) or kanamycin (50 µg/mL) was added to the medium for plasmid selection. DH5α was used as the host strain for all transformations unless stated otherwise and JM109 genomic DNA was used as template to amplify the GroEL gene. Frozen stocks were maintained at -80 °C in freezing medium.

2.1.3 *Saccharomyces cerevisiae*

The yeast strain used in this study was Y2Hgold (*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1$_{UAS}$–Gal1$_{TATA}$–His3, GAL2$_{UAS}$–Gal2$_{TATA}$–Ade2 URA3::MEL1$_{UAS}$–Mel1$_{TATA}$ AUR1-C MEL1*. Clontech, Cat No. 630498) which is designed for use with the Clontech's Matchmaker™ Gold Yeast Two-Hybrid System (Clontech, Cat. No. 630489). This strain contains four distinct reporter genes (*ADE2, HIS3, MEL1* and *AUR1-C*) that are only expressed in the presence of GAL4-based protein interactions. This strain was used to perform the yeast two-hybrid (Y2H) screening and to evaluate the

mutated versions of HtpB. Yeast was grown in YPD agar (appendix B) plates or in the appropriate selection media (e.g., SD/-Leu, see appendix B) for 3 days at 30°C. Working stock plates were kept at 4 °C for up to a month. Frozen stocks were maintained long-term at -80 °C in freezing medium (appendix B) or in the appropriate selection media (appendix B) plus 25% glycerol.

2.2 GENERAL MOLECULAR BIOLOGY TECHNIQUES

2.2.1 Polymerase chain reaction (PCR)

PCR amplifications were routinely carried out using Taq DNA Polymerase (New England Biolabs Ltd. [NEB], Cat. No. M0267) or Platinum® Pfx DNA Polymerase (Invitrogen, Cat. No. 11708) following manufacturer's indications. All oligonucleotides used as PCR primers were synthetized by Integrated DNA Technologies, Inc. (IDT) and their sequences are specified in table 1. PCR primers were re-suspended in nuclease-free water (Invitrogen, Cat. No. 10977) to a concentration of 100 µM (frozen stock) and then diluted to the working solution concentration of 10 µM. A typical 25 µl PCR reaction consisted of 18.9 µl nuclease-free water, 2.5 µl 10X ThermoPol Buffer, 1 µl dNTP mix (NEB, Cat. No. N0446S), 1 µl of the working solution of each primer, 1 µl template DNA and 0.125 µl *Taq* DNA Polymerase. Higher volume reactions (50 to 100 µl) were performed for DNA purification. The following thermo-cycling conditions were routinely used (unless otherwise specified): initial denaturation at 94 °C for 5 min, 30 cycles of [94 °C for 15 sec, 55 to 62 °C (depending on primers Tm) for 30 sec and 68°C for 1 minute per kb] and final extension at 68 °C for 10 min, resulting PCR amplification products (amplicons) were stored at -20 °C. For colony PCR, single colonies of *E. coli* transformants from agar plates were suspended in 50 µl of sterile ddH_2O on 0.2 mL tubes (Eppendorf) and heated at 95°C for 5 min in the thermocycler (Biometra T1), then centrifuged at 16000 x *g* (MIKRO 20, Hettich) for 1 min and 1 µl supernatant was used in PCR reactions.

2.2.2 Agarose gel electrophoresis

Electrophoresis of DNA was performed in 1 or 2% agarose gels depending on the size of the bands analyzed. To prepare a 1% gel, 0.5 g of agarose were dissolved in 50 mL of 1X TAE buffer (appendix B); 1µg/mL (5µl) of ethidium bromide (EtBr) was added to the melted agarose, mixed-in, and poured into a gel trail with the comb in place. Samples were mixed with 6X loading buffer (NEB, Cat. No. B7024S), loaded into the gel wells, and then subjected to electrophoresis at 100 Volts for 20 to 30 min. A VersaDoc™ MP 5000 System (Bio Rad Laboratories Inc.) was used for DNA visualisation and imaging. Images of gels were saved as JPEG files.

2.2.3 Chromosomal DNA isolation

Three mL of an overnight culture of JM109 and JR32 were centrifuged at 4000 x *g* (UNIVERSAL 32R, Hettich) for 6 min, the pellets were resuspended in 600 µl of 1X TE 0.5% SDS buffer (appendix B) and transferred to an 1.5 mL tube. Then, 1.5 µl of RNase (30 mg/mL stock solution, Sigma, Cat. No. R6513) and 3 µl of proteinase K (20 mg/mL stock solution, Sigma, Cat. No. 9001632) and incubated for 45 min at 50 °C. After incubation, 600 µl of phenol:chloroform (1:1) were added and gently mixed by inversion until the phases were completely mixed. Following centrifugation at 16000 x *g* for 10 min the upper aqueous phase was transferred to a new tube and mixed by inversion with 600 µl of phenol:chloroform (1:1). This process was repeated twice or until the white protein layer disappeared. To remove phenol, an equal volume of chloroform was added to the aqueous layer and mixed well by vortexing the tubes. The tubes were then centrifuged at 16000 x *g* for 5 min and the aqueous layer was recovered on a new tube that was centrifuged again at maximum speed for 15 min at 4 °C. The aqueous layer was recovered and transferred to a new 1.5 mL tube and 50 µl of 5M NaCl and 1 mL of 95% ethanol were added to ~500 µl of the aqueous fraction. The tubes were gently mixed by inversion and then left overnight at -20 °C. The following day, the tubes were centrifuged at 16000 x *g* for 10 min and the pellet (genomic DNA) was rinsed with 100 µl of 70% ethanol. The DNA pellet was then air-dried and resuspended in 100 µl of nuclease-free water. Two µl of chromosomal DNA were routinely used as template in a 50 µl PCR reaction.

2.2.4 Non-chromosomal DNA purification

Plasmid purification (Qiagen, Cat. No. 27104), purification of DNA from agarose gels and cleaning of PCR amplicons (Qiagen, Cat. No. 28704) were carried out using commercial kits from Qiagen as described by the manufacturer using the centrifuge protocol.

2.2.5 Restriction endonuclease digestion

Restriction endonuclease reactions were performed following manufacturer's (NEB) indications. Briefly, digestions were set up in 25 µl for analytical reactions or 50 µl for excision of DNA fragment from agarose gels. Site-directed mutations introduced on *htpB* and cDNA inserts from the yeast two hybrid (Y2H) screening were analyzed by digestion with *Fat*I (NEB, Cat. No. R0650) or *Pst*I (NEB, Cat. No. R0140) and *Alu*I (NEB, Cat. No. R0137), respectively. A typical analytical reaction would have 2-5 µl DNA (approximately 1 µg of DNA), 2.5 µl of the corresponding 10X NEBuffer, 1 µl enzyme and nuclease-free water to final volume of 25 µl. After incubation for 3 hours at 37°C or 55°C (*Fat*I), reactions were analysed by electrophoresis. *Eco*RI (NEB, Cat. No. R0101), *Bam*HI (NEB, Cat. No. R3136) and *Sal*I (NEB, Cat. No. R3138) were used for cohesive-end cloning; double digestions were set as follows: 10 µl of DNA (approximately 3 µg of DNA), 5 µl of 10X NEBuffer 4, 2 µl of each enzyme and nuclease-free water to final volume of 50 µl. Following incubation for 3 hours at 37°C, the reactions were mixed with loading buffer and separated by electrophoresis and the desired fragment was purified using the Qiagen gel purification kit. In the case of digested PCR products, the enzymes were directly removed using the Qiagen PCR cleanup kit.

2.2.6 DNA ligation

Ligation reactions were carried out using T4 DNA ligase (NEB, Cat. No. M0202) following manufacturer's indications. Reactions were set up on ice as follows: 2 µl 10X T4 DNA Ligase Buffer, 200 ng of DNA (volume was adjusted to obtain a molar ratio of 1:3 vector to insert), 1 µl ligase and nuclease-free water to 20 µl. Following overnight incubation at 16°C, the ligation mixture was heat inactivated at 65°C for 10 minutes, chilled on ice and 2 µl were transformed into *E. coli*.

2.2.7 Electrocompetent *E. coli* cells

One single colony of DH5α from a fresh plate culture was inoculated in 10 mL of LB (appendix B) and grown overnight at 37 °C with shaking (this was the starter culture). The whole starter culture was then used to inoculate in 1 L of LB and incubated until the culture reached log phase growth (OD_{620} 0.4-0.6). The culture was chilled on ice for 20 min, and then transferred to pre-chilled 50 mL Falcon™ centrifuge tubes. Cells were recovered by centrifugation at 1000 x *g* (UNIVERSAL 32R) for 20 min at 4 °C and washed twice in ice-cold ddH_2O. Cells were then resuspended in 100 mL of an ice-cold solution of 10% glycerol in ddH_2O, pooled in two 50 mL Falcon™ tubes and harvested again by centrifugation at 1000 x *g* for 20 min at 4 °C. Cells were carefully resuspended in 1 mL of ice-cold 10% glycerol and 70 µl aliquots were stored at -80 °C into sterile 1.5 mL Eppendorf tubes.

2.2.8 Transformation of *E. coli* cells by electroporation

Electrocompetent DH5α cells, electroporation cuvette (2 mm gap) and DNA were placed on ice. Cells were mixed with 2 µl of DNA in a 1.5 mL sterile Eppendorf tube, transferred to the pre-chilled cuvette and electroporated on a MicroPulser™ apparatus (Bio-Rad) at 2.5 kV for 5 milliseconds. Pre-warmed at 37 °C SOC (appendix B) medium (500 µl) was immediately added to the cuvette, gently mixed and transferred to a tube. Cells were incubated at 37 °C for 1 hour with shaking and then plated onto pre-warmed at 37 °C LB selective plates. Colonies grown on the selective agar plates were considered transformants, and were confirmed by colony PCR analysis.

2.2.9 Cloning of *htpB* into pBS and subcloning into pGBKT7

The entire *htpB* gene was PCR amplified (50µl reaction volume) from JR32 genomic DNA using primers *Eco*RI-htpB_F/BamHI-htpB_R. The PCR product of ~1600 bp (predicted size = 1659 bp) was cleaned using the Qiagen cleanup kit. The *htpB* amplicon and plasmid pBS (Agilent Technologies, Inc.) were digested with *Eco*RI and *Bam*HI as described above (section 2.2.5). Following cleaning of the digested DNA (Qiagen cleanup kit), amplicon and plasmid were ligated as described above (section 2.2.6). The resulting construct, pBS:*htpB,* was transformed into *E. coli* DH5α by electroporation, and

transformants were selected on LB plates with ampicillin (100 µg/mL). Transformants were confirmed by colony PCR analysis using the primers EcoRI-htpB_F and BamHI-htpB_R. PBS:*htpB* was purified from one of the positive transformants, subjected to digestion using *Eco*RI/*Bam*HI to drop *htpB* that was then extracted from the gel and ligated into *Eco*RI/*Bam*HI restriction sites of pGBKT7 (Clontech, Cat. No. 630489) to generate pGBK:*htpB*. The pGBK:*htpB* construct contains *htpB* cloned in translational frame with the *GAL4* DNA-Binding Domain and the c-Myc epitope tag of the bait plasmid pGBKT7. This plasmid contains a *TRP1* nutritional marker that allows for selection in yeast (growth in medium lacking tryptophan) and also carries the Kan^r gene for selection in *E. coli* (growth in medium containing kanamycin). Plasmid pGBK:*htpB* was then electroporated into DH5α and transformant selection was performed on LB plates with kanamycin (50 µg/mL). Transformants were confirmed by colony PCR using EcoRI-htpB_F and BamHI-htpB_R primers. Plasmid pGBK:*htpB* was purified and verified by sequencing with primer sets EcoRI-htpB_F/BamHI-htpB_R, HtpB419_F/HtpB1200_R and T7 (provided by Genome Quebec).

2.2.10 Cloning of *groEL* into pBS and subcloning into pGBKT7

The *groEL* gene was also cloned into plasmid pBS and then subcloned into plasmid pGBKT7 using a strategy similar to that used to clone *htpB*. Briefly, g*roEL* was PCR amplified (50µl reaction volume) from JM109 chromosomal DNA using primers EcoRI-groEL_F/SalI-GroEL_R. The ~1600 bp (predicted size = 1653 bp) amplicon was then cleaned and subjected to endonuclease digestion with *Eco*RI and *Sal*I. After DNA cleaning, *groEL* was ligated into pBS to generate the construct pBS:*groEL*, which was then electroporated into DH5α. Transformant selection was performed on LB plates with ampicillin (100 µg/mL). A positive transformant was selected for plasmid purification and then pBS:*groEL* was subjected to digestion with *Eco*RI and *Sal*I. The resulting digestion product was ligated into pGBKT7 to generate pGBK:*groEL* that was then transformed into DH5α. PGBK:*groEL* was verified by sequencing using primers EcoRI-groEL_F/SalI-groEL_R, GroEL461_F/GroEL1154_R and T7.

2.3 PROTEIN TECHNIQUES

2.3.1 Protein samples preparation

E. coli and *L. pneumophila*: Overnight 2 mL cultures of *E. coli* or *L. pneumophila* were centrifuged at 4000 x *g* (Universal 32R) for 6 min at 4 °C, supernatants were discarded. The cells were resuspended in 500 µl of ddH_20, transferred to a conical 2 mL screw cap tube (BIO PLAS Inc., Ca. No. 4216R) containing 0.4 g of 0.1 mm silica beads (BioSpec, Cat. No. 11079101z) and beaten using a Mini-BeadBeater (3110BX, BioSpec) at 4800 oscillations/min for 1 min at 4 °C. Following centrifugation at 16000 x *g* for 1 min at 4 °C, the supernatants containing soluble proteins were transferred to a clean 1.5 mL centrifuge tube and frozen at -80 °C.

S. cerevisiae: Yeast cultures were grown overnight in 50 mL of appropriate selection medium, then chilled on ice for 20 min and centrifuged at 1000 x *g* for 5 min at 4 °C. Cells were resuspended in ice-cold ddH_2O and centrifuged again. Pellets were then resuspended in 1mL of ice-cold yeast lysis buffer (appendix B) and transferred to conical 2 mL screw cap tubes (BIO PLAS Inc., Ca. No. 4216R) containing 400 µl of 0.5 mm glass beads (BioSpec, Cat. No. 11079105) and beaten using the Mini-BeadBeater at 4800 oscillations/min for 3 min (in 3 x 1 min rounds chilling on ice in between) at 4 °C. Tubes were centrifuged at maximum speed for 5 min, and the supernatants containing soluble proteins were transferred to pre-chilled 1.5 mL Eppendorf tubes and stored at -80°C.

2.3.2 Protein quantification

Determination of protein concentration was carried out using the BCA Protein Assay Kit (Thermo scientific, Cat. No. 23225) as indicated by the manufacturer. Briefly, samples were diluted 1 in 10 with ddH_2O and 25 µl of each BSA standard or sample were pipetted into individual wells of a 96-well microplate. Then, 200 µl of the working reagent were added to each well and the microplate was incubated at 37 °C for 30 minutes. Absorbance at 562 nm was measured on a microplate reader (Benchmark Plus, Bio-Rad). Protein concentration of each unknown sample was calculated using the BSA standard curve as a reference.

2.3.3 SDS-PAGE and Western blotting

Polyacrylamide gels were cast using a 5% stacking gel and a 12% resolving gel (appendix B). Protein samples were mixed 4:1 with 5X SDS-PAGE loading buffer (appendix B) and heated at 95 °C for 3 min in a thermoblock (Standard Heatblock, VWR). The gel was placed into the electrophoretic chamber (Mini-PROTEAN® Multi-Casting Chamber, Bio Rad) filled with 1X running buffer (appendix B), 15 µl of each sample and 7 µl of protein ladder (NEB, Cat. No. P7712S) were loaded into separate wells of the gel. Electrophoresis was first run at 60 V for 20 min, and then at 100-150 V for 2 h. After electrophoresis, the gel was equilibrated in 1X transfer buffer (appendix B) and then the proteins were transferred to an activated PVDF membrane (EMD Millipore, Cat. No. IPVH00010) during 90 min at 90 V. The membrane was then stained for 5 min with Ponceau S (Allied Chemical, Cat. No.628, see appendix B for preparation) to visualize the bands. After imaging, the Ponceau S staining was removed by washing with 1X phosphate-buffered saline (PBS, appendix B) and then the membrane was incubated for 3 h with blocking solution (appendix B). After blocking, the membrane was rinsed with 1X TTBS buffer (appendix B) and then incubated overnight at 4 °C with the corresponding primary antibody diluted in 1X TTBS 0.01% BSA as follows: polyclonal anti-HtpB (supplied as hyperimmune rabbit serum cross-reacts with Cpn60 and HtpB) [97] 1:500, monoclonal anti-c-Myc (Clontech, Cat. No. 631206) 1:500, polyclonal anti-HA-Tag (Clontech, Cat. No. 631207) 1:500. The membrane was then washed 3 times for 10 min with 1X TTBS and incubated for 1 h with the corresponding conjugated (Alkaline Phosphatase) secondary antibody diluted 1:5000 in 1X TTBS 0.01% BSA: goat anti-rabbit IgG (Cedarlane, Cat. No. CLCC42008) or goat anti-mouse IgG (Cedarlane, Cat. No. CLCC30008). The membrane was then washed 3 times for 10 min with 1X TTBS, 2 times for 10 min with 1X TBS (appendix B) and 2 times for 5 minutes with AP buffer (appendix B). The PVDF membrane (Milipore) was then incubated with developing solution (appendix B) until the bands appeared, rinsed with ddH_2O and then allowed to dry.

2.4 YEAST TECHNIQUES AND YEAST TWO HYBRID (Y2H) SCREENING

2.4.1 Yeast transformation

The lithium acetate (LiAc)-mediated method was used to transform *S. cerevisiae* (see Appendix B for buffers and media recipes). To prepare competent yeast cells, several colonies of strain Y2Hgold from a fresh YPD plate culture were inoculated in 30 mL of YPDA medium and incubated at 30 °C with shaking at 200 rpm overnight. This culture was then inoculated into 300 mL of fresh YPD and incubated until the culture reached log-phase (OD_{600} between 0.4 and 0.6). Then, the culture was split into six 50 mL Falcon™ tubes and cells were harvested by centrifugation at 1000 x *g* for 5 min at room temperature, the supernatant was discarded and the cells were resuspended in sterile ddH_20 and pooled in one 50 mL Falcon™ tube. Cells were centrifuged again and then resuspended in 1.5 mL of freshly prepared, sterile 1X TE/1X LiAc solution, this competent cells should be used the same day they are prepared for optimal performance. In clean 1.5 mL Eppendorf tubes, 0.1 µg of plasmid DNA (0.1 µg of each plasmid were added when co-transformations were performed) and 0.1 mg of carrier DNA (Clontech, Cat No. 630440) were mixed by vortexing with 100 µl of competent yeast cells. Then, 0.6 mL of sterile PEG/LiAc solution was added to each tube and vortexed at high speed for 10 sec, this mixture was incubated overnight at room temperature. Next day, 70 µl of Dimethyl sulfoxide (DMSO) (Sigma, Cat. No. 472301) were added to each tube and mixed gently by inversion. The cells where then heat shocked for 30 min at 42 °C in the thermoblock, followed by a 2 min incubation on ice. Finally, the cells were pelleted at high speed for 5 sec, pellets were resuspended in 200 µl of sterile 1X TE buffer and plated on the appropriate SD agar (e. g. SD/-Leu) plate that will select for the desired transformants. Plates were incubated for three days at 30 °C and then stored at 4 °C for up to one month.

2.4.2 Plasmid purification from yeast

This protocol was adapted to disrupt yeast cells mechanically and the resulting lysate was subjected to plasmid DNA extraction using the QIAprep Spin Miniprep Kit (Qiagen, Cat No. 27104). Yeast cells carrying the desired plasmid were grown overnight on 2 mL of

the appropriate SD medium at 30°C. Then 200 µl of 0.5 mm glass beads (BioSpec, Cat. No. 11079105) were mixed with the yeast culture in a 2mL conical screw cap tube (BIO PLAS Inc., Ca. No. 4216R). The mixture was centrifuged for 2 minutes at 1000 x *g*, supernatant was discarded and the pellet was resuspended in 250 µl of buffer P1 (Miniprep kit). Then, 250 µl of buffer P2 (Miniprep kit) were added and the mixture was beaten using the Mini-BeadBeater at 4800 oscillations/min for 3 min at 4 °C. 350 µl of buffer N3 (Miniprep kit) were added and mixed by inversion. The mixture containing total soluble DNA was centrifuged for 10 minutes at maximum speed. The supernatant was applied to a QIAprep spin column and the rest of the protocol was performed following manufacturer indications. Plasmid DNA was eluted in 30 µl of nuclease-free water.

2.4.3 Y2H screening using yeast mating

To identify an interaction partner for HtpB, a Y2H screening was performed using a human cDNA library. The library (prey) contains human cDNAs fused to the Gal4 transcriptional activation domain and the bait plasmid (pGBK:*htpB*) carries HtpB fused to the Gal4 DNA binding domain. The Y2HGold strain (Clontech's Matchmaker™ Gold Yeast Two-Hybrid System) used in the screening has four integrated reporter genes: *AUR1-C* (resistance to aureobasidin A), *HIS3* (biosynthesis of histidine), *ADE2* (biosynthesis of adenine), and *MEL1* (encodes alpha-galactosidase that turns colonies blue in the appropriate medium). There are three distinct Gal4-responsive promoters controlling expression of HIS3, ADE2, and MEL1/AUR1-C reporter genes: G1, G2, and M1, respectively. *AUR1-C* and *MEL1* share the M1 promoter. These reporter genes are expressed in response to protein-protein interactions that bring the GAL4 transcriptional activation and DNA binding domains into close proximity allowing yeast growth in selective media.

The yeast two hybrid screening was performed as described in the Clontech user manual (Protocol No. PT4084-1). Briefly, 50 mL of SD/–Trp (appendix B) broth were inoculated with 5 colonies of Y2Hgold strain carrying the bait construct (pGBK:*htpB*) and incubated overnight with shaking (250 rpm) at 30 °C until the OD_{600} reached 0.8. Cells were pelleted (1000 x g for 5 min), resuspended in 5 mL SD/–Trp and counted using a

hemacytometer (Bright-Line™, Sigma, Cat. No. Z359629). This concentrated culture ($9.2x10^8$ cells) was mixed with 1 mL of the library strain (Mate & Plate™ Library - Universal Human, Clontech, Cat. No. 630481) plus 45 mL of 2X YPDA (with 50 µg/mL kanamycin) (appendix B) in a sterile 2 L flask and incubated at 30 °C for 24 h, slowly shaking (40 rpm) to favor the mating process. The number of viable cells in the library was determined by plating 100 µl of serial dilutions in SD/–Leu medium (appendix B). After 24 h of mating, the presence of zygotes in the culture was confirmed by observing a sample under the phase contrast microscope. Then, the cells were pelleted (1,000 x g for 10 min) and resuspended in 10 mL of 0.5X YPDA broth (appendix B). To calculate the number of clones screened, 100 µl of serial dilutions from the mated cell suspension (1:10, 1:100, 1:1000 and 1:10000) were spread on 10 cm diameter agar plates of SD/–Trp, SD/–Leu and DDO (appendix B). The remaining 9.8 mL of the cell suspension was plated on forty nine 15 cm diameter plates (200 µl per plate) of double drop-out medium supplemented with 200 ng/mL of aureobasidin A (Clontech, Cat. No. 630466) and 0.04 ng/mL of X-alpha-Gal (Clontech, Cat. No. 630463) (DDO/X/A, appendix B) and incubated at 30 °C for 5 days. The total number of screened clones (diploids) was calculated by counting the colonies from the DDO plates and multiplying for the total volume of cell plated. The mating efficiency was calculated by dividing the number of screened diploids (in CFU/mL) by the viability of the library (CFU/mL grown in the SD/–Leu plates) and multiplying by 100. All colonies (blue and white) grown in the 15 cm DDO/X/A plates were re-streaked as small patches on fresh 10 cm DDO/X/A plates, incubated at 30 °C for 3 days and then just the blue patches were re-streaked on quadruple drop-out media lacking Leu, Trp, His and adenine, containing 0.04 ng/mL of X-alpha-Gal and 200 ng/mL of aureobasidin A (QDO/X/A, appendix B) to select positive interactions. The colonies that grew on the higher stringency medium (QDO/X/A) were frozen on DDO broth plus 25% glycerol and kept at -80 °C.

2.4.4 Duplicates elimination and prey insert identification by sequencing

To identify potential duplicate inserts, the DNA inserts were amplified from the prey plasmids rescued from the Y2H screening using the MMAD primer set (Table 1), then subjected to digestion with *Alu*I for 3 hours at 37 °C and analyzed by electrophoresis on a

2% TAE Agarose/EtBr gel (appendix B). The plasmids that carried inserts that seemed to be unique were sequenced using the MMAD set (Table 1) and T7 primers (provided by Genome Quebec).

2.4.5 Library plasmids isolation and genuine interactions confirmation

All colonies that grew on QDO/X/A were further analyzed to verify that the interactions were genuine. Library plasmids were purified from yeast cells grown on QDO broth (appendix B) using the protocol described in section 2.4.2 and transformed in electrocompetent DH5α. The library plasmid has the Amp^r cassete that allows *E. coli* transformants selection in LB plates containing ampicillin. *E. coli* transformants were subjected to colony PCR with primer set MMAD (Table 1) to confirm presence of the plasmid. Positive transformants were grown in 5 mL of LB broth plus ampicillin (100 µg/mL) and used to purify the plasmid as described above (section 2.2.4). To confirm genuine interactions, the following pairs of vectors were co-transformed back into *S. cerevisiae* strain Y2HGold:

- pGBK:*htpB* + Candidate prey
- pGBK:*groEL* + Candidate prey
- Empty pGBKT7 + Candidate prey
- pGBK:*htpB* + Empty pGADT7 (Clontech, Cat. No. 630442)
- pGBK:*groEL* + Empty pGADT7
- Empty pGBKT7 + Empty pGADT7

The co-transformed yeast cells were then grown overnight in 2 mL cultures of DDO broth. Then, the cultures were serially diluted (1:10, 1:100, 1:1000), and 10 µl of each dilution as well as 10 µl of each undiluted culture were plated on DDO/X/A and QDO/X/A media and incubated for 5 days at 30°C.

2.4.6 Co-immunoprecipitation (Co-IP) in yeast

Total yeast protein extracts were prepared as follows: Y2HGold yeast cells carrying both pGBK:*htpB* and pGAD:*hECM29* were grown overnight. Cells were recovered by centrifugation (1000 x *g* for 5 min) and washed twice (by centrifugation) with ice cold

Table 1. Oligonucleotides used in this study

Name	SEQUENCE 5' → 3'	Rest.sit.
EcoRI-htpB_F	CCG**GAATTC**ATGATAATGGCTAAAGAATTA CG	*Eco*RI
BamHI-htpB_R	ATA**GGATCC**TTACATCATTCCGCCCATG	*Bam*HI
HtpB419_F	AAGACAGCAAAGCCATTG	
HtpB1200_R	AGCATCTTCAACACGAGC	
EcoRI-groEL_F	CGG**GAATTC**ATGGCAGCTAAAGACG	*Eco*RI
SalI-groEL_R	AGTC**GTCGAC**TTACATCATGCCGCCCA	*Sal*I
GroEL461_F	CCGACGAAACCGTAGGTAAA	
GroEL1154_R	TAGCAGCACCCACTTTGATAA	
MMAD_F	CTATTCGATGATGAAGATACCCCACCAAACCC	
MMAD_R	GTGAACTTGCGGGGTTTTTCAGTATCTACGAT	
MMBD_F	TCATCGGAAGAGAGTAGTAAC	
MMBD_R	CCTAAGAGTCACTTTAAAATTTGTATAC	
M68A_F	TGAGTTTGAGCATCGTTTCGCGAACATGGGCGCTCAAATG	
M68A_R	CATTTGAGCGCCCATGTTCGCGAAACGATGCTCAAACTCA	
M212A_F	TTTATCAACAACCAGCAAAACGCGAGCTGTGAACTTGAGCATCC	
M212A_R	GGATGCTCAAGTTCACAGCTCGCGTTTTGCTGGTTGTTGATAAA	
S236A_F	CAGTATTCGTGAAATGTTGGCCGTATTGGAAGGTGTTGC	
S236A_R	GCAACACCTTCCAATACGGCCAACATTTCACGAATACTG	
K298A_F	AGCGATGTTGCAAGACATTGCTATTTTGACTGCGGGTCAAGTTATTTC T	
K298A_R	AGAAATAACTTGACCCGCAGTCAAAATAGCAATGTCTTGCAACATCG CT	
N507A_F	CAAAGTAACCCGTATGGCTCTGCAAGCTGCAGCTTCTGTA	
N507A_R	TACAGAAGCTGCAGCTTGCAGAGCCATACGGGTTACTTTG	
M68E_R	CATTTGAGCGCCCATGTTCTCGAAACGATGCTCAAACTCA	
M212G	GGATGCTCAAGTTCACAGCTCCCGTTTTGCTGGTTGTTGATAAA	
S236P	GCAACACCTTCCAATACGGGCAACATTTCACGAATACTG	
K298G_R	AGAAATAACTTGACCCCCAGTCAAAATAGCAATGTCTTGCAACATCG CT	
N507Y_R	CTACAGAAGCTGCATATTGCAGAGCCATACGGG	

E472A_ F	TAGTAAACAAGGTAGCTGCGCACAAAGACAACTACGG	
E472A_ R	CCGTAGTTGTCTTTGTGCGCAGCTACCTTGTTTACTA	
H473A_F	GTAGTAAACAAGGTAGCTGAGGCCAAAGACAACTACGGTTTCAA	
H473A_R	TTGAAACCGTAGTTGTCTTTGGCCTCAGCTACCTTGTTTACTAC	
K474A_F	TAAACAAGGTAGCTGAGCACGCAGACAACTACGGTTTCAACG	
K474A_R	CGTTGAAACCGTAGTTGTCTGCGTGCTCAGCTACCTTGTTTA	
D475A_F	GGTAGCTGAGCACAAAGCCAACTACGGTTTCAACG	
D475A_R	CGTTGAAACCGTAGTTGGCTTTGTGCTCAGCTACC	
1415-17-18-20-21-24F	CTTCTGTTGTAGTAAACAAGGTAGCTGCGGCCGC-AGCCAACTACGGTTTCAACGCTGCAACTGG	
1415-17-18-20-21-24R	CCAGTTGCAGCGTTGAAACCGTAGTTGGCTGCGGCC-GCAGCTACCTTGTTTACTACAACAGAAG	
1411-12-13-15_F	ATGAAGCTTCTGTTGTAGTAAACAAGGTAAAAGGGC-ACAAAGACAACTACGGTTTCAAC	
1411-12-13-15_R	GTTGAAACCGTAGTTGTCTTTGTGCCCTTTTACCTT-GTTTACTACAACAGAAGCTTCAT	
1417-18-20-22-24_F	CTGTTGTAGTAAACAAGGTAAAAGGGGGCGATGGC-AACTACGGTTTCAACGCTGCAACTG	
1417-18-20-22-24_R	CAGTTGCAGCGTTGAAACCGTAGTTGCCATCGCCCC-CTTTTACCTTGTTTACTACAACAG	
E67M_R	CACCATCTGCGCACCCATATTCATGAACTT-GTCTTCCAGTTCGAT	
G211M_R	TTCCAGTTCTACTGCCATAGTTTCCGGCTT-GTTGATGAAGTAAGG	
P235S_R	AGCTTCCAGAACCGACAGCATTTCGCGGAT	
G297K_R	GATCTCTTCAGAGATCACGGTTTTGCCAGTC-AGGGTTGCGATATC	
Y506N_R	CACAGAAGCTGCGTTCTGCAGAGCAGAACG	
GroEL470-474_F	CTGTTGTTGCTAACACCGTTGCAGCCGCCGC-CGCCAACTACGGTTACAACGCAG	
GroEL470-474_R	CTGCGTTGTAACCGTAGTTGGCGGCGGCGGCTGCA-ACGGTGTTAGCAACAACAG	

Bold nucleotides indicate restriction sites corresponding to the enzymes under the column Rest. sit.
Underlined nucleotides indicate places were mutations were introduced

water. The cell pellet was then resuspended in 1mL of ice-cold lysis buffer (appendix B) and 1 µl of protease inhibitor mix (Sigma, Cat. No. P8340) was added. Then, the mixture was transferred to a conical 2 mL screw cap tube (BIO PLAS Inc., Ca. No. 4216R) containing 400 µl of 0.5 mm glass beads (BioSpec, Cat. No. 11079105) and beaten using the Mini-BeadBeater at 4800 oscillations/min for 3 min (in 3 x 1 min rounds chilling on ice in between) at 4 °C. Following centrifugation at 16000 x *g* for 5 min at 4 °C, the supernatant containing soluble proteins was transferred to a pre-chilled 1.5 Eppendorf tube and stored at -80 °C. The vectors pGBKT7 and pGADT7 provided HtpB with a c-Myc epitope tag and hECM29 with a HA tag, respectively. I subsequently performed Co-IP analyses using antibodies that recognize these tags as follows: 1 mg of total yeast protein extract was incubated with 1 µg of anti-HA (Clontech, Cat. No. 631207) or with 1 µg of anti cMyc (Clontech, Cat. No. 631206) for 2 hours at 4°C. Then 30 µl of Protein A/G PLUS-Agarose (Santa Cruz Biotech, Cat. No. sc-2003) were added and the mixture was rotated gently for 2 hours at 4°C using a rotator (Tube Rotator 400110Q, Labquake™). The agarose beads were then recovered by spin down and washed three times by spinning down and washing with 500 µl of lysis buffer w/o protease inhibitors (the first supernatant containing unbound soluble proteins was saved). The sedimented proteins (bound to the agarose beads) and supernatants were separately mixed with 5X sample buffer and boiled for 5 min to generate SDS-PAGE samples, and 12 µl of these samples were then loaded onto a SDS-PAGE gel and separated by electrophoresis. Proteins were then electro-transferred to a PVDF membrane and immunostained with anti-cMyc or anti-HA diluted 1:500 in TTBS 0.01% BSA, and the secondary antibodies goat anti-mouse IgG and goat anti-rabbit IgG, respectively.

2.5 AMINO ACID PREDICTION

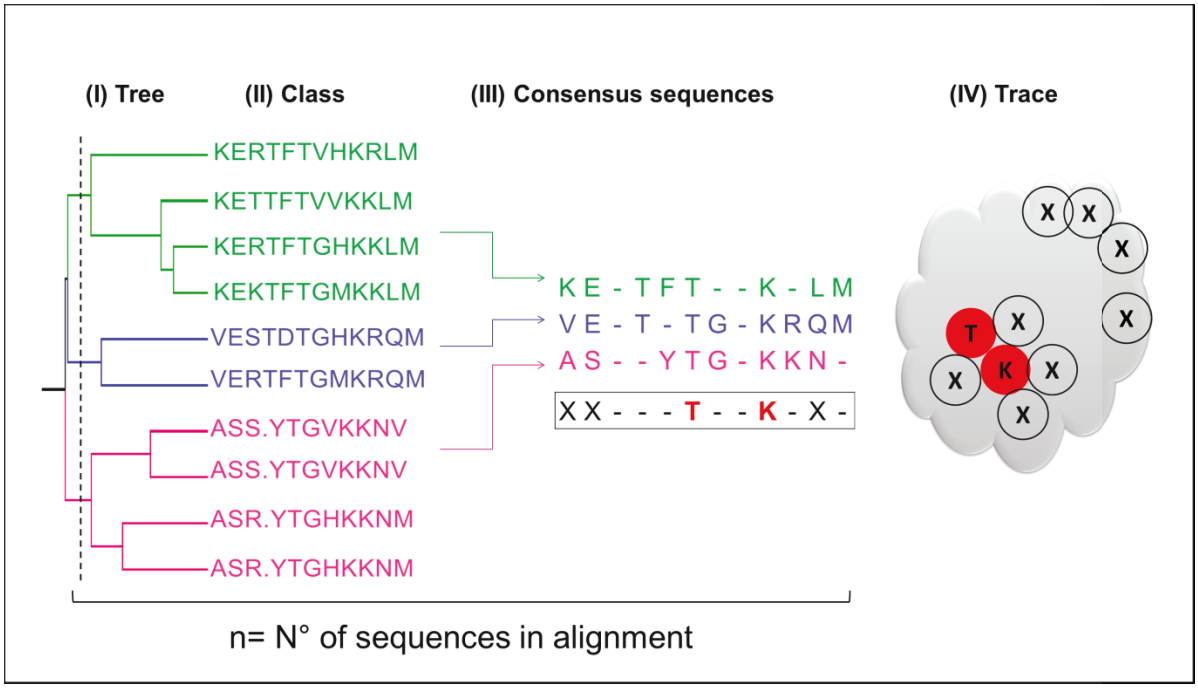
2.5.1 Evolutionary trace (ET) calculation and amino acids selection

To identify orthologs of HtpB (GI: 52840925), a BLAST (Basic Local Alignment Search Tool) search was performed against the NCBI's (National Center for Biotechnology Information) non-redundant protein sequence database, using the blastp under default

parameters. Orthologs with an E-value less than 1 x 10^{-6} were retrieved and then aligned using ClustalOmega [106]. Identical sequences were eliminated using the Usearch's derep_fulllength command [107]. Mitochondrial Hsp60s, group II chaperonins (Archaea) and sequences with less than 50% the length of HtpB were manually eliminated from the multiple sequence alignment (MSA). The alignment was then used as input into the ET code available at http://mammoth.bcm.tmc.edu/ as follows: the **Universal Evolutionary Trace tool** was selected. In step 1, **UniProt accession number** was indicated as the type of input provided; thus, the HtpB UniProt accession number (Q5ZXP3) was specified. Then, under **advanced options**, the **Real-Valued Trace (rvET)** option was selected and the HtpB alignment file in GCG format was uploaded under **Custom Sequence** Input. The Query name in the input file was then introduced: Q5ZXP3. Finally, the ET for each amino acid position in the alignment was calculated by the online tool (click in **run trace** button). Briefly, the ET tool builds a pairwise sequence similarity matrix using the sequences in the alignment and then the UPGMA method [108] is applied to generate an evolutionary tree where the sequences are separated in groups according to the branching. A consensus sequence is established for each branch group and then, the residue trace ranks are assigned based on the minimum number of branches into which the evolutionary tree must be partitioned for that residue to be invariant within each branch group [109]. Finally, the invariance within the individual branches is also introduced to the calculation to obtain the real-value ET (rvET) (Fig. 3). The rvET ranks (hereafter designated simply as ET ranks) were then mapped onto the HtpB 3D structure, which was independently predicted as per section 2.5.2 (see below). The ET rank is a relative ranking of evolutionary importance for each amino acid position in the MSA. Low ET ranks (the lowest value being 1.0) indicate sequence conservation and therefore, an implied functional importance.

To narrow down my search for amino acid residues that are more likely involved in moonlighting functions, I took advantage of the knowledge that most of the protein-folding-independent functions of HtpB cannot be performed by GroEL [97]. Therefore, I used the BLOSUM 62 matrix [110] as a secondary tool to specifically compare the amino acid substitutions between HtpB and GroEL, scoring each substitution in a numerical

Figure 3. The evolutionary trace (ET) method. Diagram illustrating the steps of the ET method. All of the sequences (under step II) in a protein family are aligned and a tree based on amino acid similarity is constructed to illustrate the relationships between individual family members. Dots in sequences represent gaps in the alignment (I). The tree is then delineated into groups termed classes; in this example 3 classes (green, blue and pink) were formed as indicated by the intersects with the vertical dashed line (II). A consensus sequence for each class is created (letters are conserved positions and dashes are non-conserved positions) and these are then compared to form the ET sequence (sequence inside the box). Residue positions that are invariant within each class, but that vary among them are called class-specific or trace residues (labeled X in the ET sequence, colored black) and those that are conserved in all sequences are denoted by amino acid single-letter code in the ET sequence and colored red. The number of classes into which the tree has to be divided for a residue to become class-specific is called the ET rank of that residue (e. g. class-specific positions in this example have rank 3). Steps I to III are repeated n times until all amino acids are ranked (III). Finally, trace residues are mapped onto the three-dimensional structure of a family member (grey). Residues conserved in all sequences of the alignment are shown as red circles and trace amino acids (ET rank = 3) are shown as no filled circles (IV). Residues with low ET ranks are considered to be evolutionary important.


(I) Tree
(II) Class
(III) Consensus sequences
(IV) Trace
KERTFTVHKRLM
KETTFTVVKKLM
KERTFTGHKKLM
KEKTFTGMKKLM
VESTDTGHKRQM
VERTFTGMKRQM
ASS.YTGVKKNV
ASS.YTGVKKNV
ASR.YTGHKKNM
ASR.YTGHKKNM
K E - T F T - - K - L M
V E - T - T G - K R Q M
A S - - Y T G - K K N -
X X - - - T - - K - X -
X
T
K
n= N° of sequences in alignment

scale from -3 to +9. A negative BLOSUM 62 score is given to substitutions with a low probability to occur. Therefore, a low ET rank and a negative BLOSUM 62 score, were used as the main criteria to select amino acids in evolutionary important positions with less likely substitutions, which would be expected to be involved in functional diversity (e.g., putative HtpB moonlighting functions). Ultimately, these selected amino acids were experimentally tested by mutagenesis and their impact in the HtpB-hECM29 interaction (see section 2.7 and 2.9 below).

2.5.2 HtpB 3D structure prediction and ET ranks mapping

The HtpB 3D structure was predicted using the ModWeb server available online at https://modbase.compbio.ucsf.edu/modweb/. The HtpB protein sequence was submitted and the slow (Seq-Prf, PSI-Blast) fold assignment method was selected to calculate the sequence-structure alignment. Three models were calculated based on the best hits found in the protein data bank (PDB): the Cpn60 of *E. coli* (1kp8A), Cpn60 of *Thermus thermophilus* (4V4O) and the apical domain of the Cpn60-1 of *Mycobacterium tuberculosis* (3m6cA). The HtpB model based on the *E. coli* chaperonin was selected as it had the highest sequence identity (76%), and was subsequently used to map the ET ranks. HtpB structure visualization and mapping of the ET ranks was performed using PyMol and the PyETV plugin downloaded from:

http://mammoth.bcm.tmc.edu/traceview/HelpDocs/PyETVHelp/pyInstructions.html.

Additionally, analysis of the secondary structure of the HtpB amino acids 471-475 was performed using the online software VADAR (Volume, Area, Dihedral Angle Reporter) following the developers instructions [111].

2.6 *HTPB* AND *GROEL* SITE-DIRECTED AND MULTISITE-DIRECTED MUTATIONS

All mutations were introduced into pBS:*htpB* or pBS:*groEL* and then the mutated genes (*htpB* or *groEL*) were excised from pBS and sub-cloned into pGBKT7 as described above (sections 2.2.9 and 2.2.10). All base changes were verified by sequencing using primer

pairs MMBD-set, EcoRI-htpB_F/BamHI-htpB_R, HtpB419_F/HtpB1200_R, EcoRI-groEL_F/SalI-groEL_R, GroEL461_F/GroEL1154_R and the T7 primer (provided by Genome Quebec).

2.6.1 Single amino acid replacements (site-directed mutations)

The QuikChange® Site-Directed Mutagenesis Kit (Agilent Technologies, Cat. No. 200518) was used to generate all the desired single amino acid replacements in HtpB and GroEL. Briefly, pBS:*htpB* or pBS:*groEL* was subjected to PCR using complementary primers that carry the desired mutation(s) (1 to 7 nucleotides). Oligonucleotides used to introduce the mutations (Table 1) were designed using the QuikChange® Primer Design Program available at http://www.stratagene.com/qcprimerdesign. Reaction was performed following manufacturer's instructions. The plasmid containing the desired mutations was then transformed into XL1-Blue super-competent cells following manufacturer's indications. Mutated constructs were recovered from XL1-Blue *E. coli* and the mutated *htpB* or *groEL* were excised from pBS and sub-cloned into pGBKT7 as described above (sections 2.2.9 and 2.2.10).

2.6.2 Multiple amino acid replacements (multi-site directed mutations)

The QuikChange Multi Site Directed Mutagenesis Kit (Agilent Technologies, Cat. No. 200515) was used to simultaneously introduce mutations to different sites in pBS:*htpB* and pBS:*groEL*. Oligonucleotides containing the desired mutations were designed using the QuikChange® Primer Design Program to bind the same strand of the template DNA. Two, three or five primers were used to mutate 2, 3 or 5 amino acids at once, respectively. The PCR reaction was set exactly as indicated in the manufacturer's instructions. The resulting mutated single-stranded circular DNA (ssDNA) was then transformed into *E. coli* XL10-Gold ultracompetent cells, where the mutant ssDNA was converted into double-stranded circular DNA plasmids (constructs) *in vivo*. Mutated constructs were then recovered from XL10-Gold *E. coli* and the mutated *htpB* or *groEL* were excised from pBS and sub-cloned into pGBKT7 as described above in sections 2.2.9 and 2.2.10.

The amino acids substituted in HtpB are shown in table 2. Single residue substitutions in HtpB (M68A, M212A, S236A, K298A, N507A, E472A, H473A, K474A and D475A) were done using Site-Directed Mutagenesis Kit as above, changing only one or two nucleotides as shown in Table 1. Multiple amino acid substitutions in the HtpB aa471-475 cluster (mutant HtpB 472-475/A), were also introduced using the site-directed kit as the nucleotides changed were close enough to be covered by a single primer. To make the mutant HtpB KGGDG, two primer sets had to be used in two separate reactions, the first pair (1411-12-13-15_F/1411-12-13-15_R) changed the first two residues (A471K,E472G) and the second pair (1417-18-20-22-24_F/1417-18-20-22-24_R) changed the other three amino acids (H472G, K474D, D475G). Similarly, mutant HtpB MMSKN/A was generated using primers M68A_R, M212A_R, S236A_R, K298A_R and N507A_R and mutant HtpB EGPGY with primers M68E_R, M212G_R, S236P_R, K298G_R and N507Y_R as above.

The amino acids substituted in GroEL are indicated in Table 3. GroEL-AEHKD was generated using primers GroEL470-474_F/GroEL470-474_R and site-directed mutagenesis method. Mutant GroEL-MMSKN was generated using primers E67M_R, G211M_R, P235S_R, G297K_R and Y506N_R. Then, GroEL-MMSKN was used as template to introduce the additional substitutions in positions 470-474 using primers GroEL470-474_F/GroEL470-474_R to generate the GroEL-Multi mutant.

2.7 HTPB-hECM29 INTERACTION EVALUATION BY Y2H

2.7.1 Evaluation of the interaction by plate and broth assays

Plate assay: direct interaction between WT or mutated versions of HtpB or GroEL and hECM29 was investigated by co-transformation of the respective Y2H plasmids in the *S. cerevisiae* strain Y2Hgold, followed by selection of transformants on DDO solid medium after incubation at 30 °C for 3 days. A single colony of each transformant was subsequently transferred to 2 ml cultures of DDO liquid medium and incubated overnight at 30 °C. Serial dilutions (1:10, 1:100 and 1:1000) of these cultures were then spotted (one 10 µl drop per dilution) on selective QDO/X/A plates. The interaction between

Table 2. Summary of the amino acids substituted in HtpB

	Amino acid position									
HtpB variant	**68**	**212**	**236**	**298**	**507**	**471**	**472**	**473**	**474**	**475**
WT-HtpB	**M**	**M**	**S**	**K**	**N**	**A**	**E**	**H**	**K**	**D**
M68A	A									
M212A		A								
S236A			A							
K298A				A						
N507A					A					
MMSKN/A	A	A	A	A	A					
EGPGY	E	G	P	G	Y					
E472A							A			
H473A								A		
K474A									A	
D475A										A
472-475/A							A	A	A	A
KGGDG						K	G	G	D	G

Amino acids that were substituted are shown in white background cells, where the letter indicates the substituting amino acid.
Wild type (WT-HtpB) amino acids and their corresponding positions are given in boldface at the top of the table.
The amino acid positions that were not mutated in the different variants are indicated by blank cells with grey background.

Table 3. Summary of the amino acids substituted in GroEL

	Amino acid position									
GroEL variant	**67**	**211**	**235**	**297**	**506**	**470**	**471**	**472**	**473**	**474**
WT-GroEL	**E**	**G**	**P**	**G**	**Y**	**K**	**G**	**G**	**D**	**G**
GroEL-MMSKN	M	M	S	K	N					
GroEL-AEHKD						A	E	H	K	D
GroEL-Multi	M	M	S	K	N	A	E	H	K	D

Amino acids that were substituted are shown in white background cells, where the letter indicates the substituting amino acid.
Wild type (WT-GroEL) amino acids and their corresponding positions are given in boldface at the top of the table.
The amino acid positions that were not mutated in the different variants are indicated by blank cells with grey background.

hECM29 and the chaperonin variants was qualitatively evaluated as positive or negative, by observing yeast growth (as blue colonies) after 5 days of incubation at 30 °C.

Broth assay: 2 ml of QDO broth were inoculated in triplicated with enough of an overnight DDO culture of yeast carrying the same plasmid combinations used in the plate assay to achieve a starting OD600 of 0.8 units. The OD600 of these 2 ml QDO cultures was measured after 6 days of incubation at 30 °C with shaking to quantitatively evaluate the HtpB-hECM29 interaction.

2.7.2 Alpha-galactosidase assay

Alpha-galactosidase activity was evaluated using a colorimetric assay following the recommendations found in the Yeast Protocols Handbook (Clontech, PT3024-1). To compare the strengths of the interaction between hECM29 and the HtpB variants, Y2Ggold carrying different plasmid combinations was grown at 30°C overnight in 2 mL of DDO medium. Then, subcultures were started adding 200 µl of the overnight DDO culture to 2 mL of fresh QDO medium and incubating for 6 days at 30 °C with shaking. The QDO cultures were then diluted 1 in 10 with water and OD_{600} was recorded. Then, 1 mL of the culture was centrifuged at 16000 x *g* for 1 min, and 16 µl of supernatant were loaded in triplicate into a 96-well plate containing 48 µl of PNP assay buffer (Appendix B). Microplates were incubated at 30 °C for 3 h covered with a lid and sealed with parafilm to prevent evaporation. Reactions were terminated by addition of 136 µl of 1 M Sodium carbonate and absorbance at 410 nm of each sample was measured using the Benchmark Plus microplate reader. Alpha-galactosidase activity was calculated as follows:

Milli units/(ml x cell) = A_{410} x Vf x 1,000 / [(ε x b) x t x Vi x OD_{600}]

A_{410} = Absorbance at 410

t = elapsed time (in min) of incubation

Vf = final volume of assay (200 µl)

Vi = volume of culture medium supernatant added (16 µl)

OD_{600} = optical density of culture

(ε x b) = p-nitrophenol molar absorptivity at 410 nm x the light path (cm) =

= 20.3 ml/µmol (determined by Clontech for 200 µl assay in 96-well plates) x 0.52 cm

2.8 PHYLOGENETIC TREE CONSTRUCTION

All available chaperonin 60 amino acid sequences (orthologs and paralogs) were gathered from the Pfam database (accession number: PF00118) and unique sequences were then selected using the Usearch's derep_fulllength command to create a database that I named Bact-Cpn60. Then, Cpn60 sequences from organisms of the phylum Proteobacteria were retrieved from the Bact-Cpn60 database to generate a dataset that I named All_ProteoBact-Cpn60, which included 1419 Cpn60 amino acid sequences. To decrease redundancy in the data set (i.e., sequences of several strains belonging to one species), only one organism per cluster of similar genomes was selected using the clustering tool available at:
http://microbiome.wlu.ca/research/redundancy/redundancy.cgi.
Construction of complete genome clusters was based on a genome similarity threshold of 0.9 (90% similarity). After correcting for redundancy only 710 Cpn60 amino acid sequences remained in the dataset. Then, the Cpn60 sequences of these non-redundant organisms were sorted, based on the bacterium's number of Cpn60s in decreasing order. Following this organization, the chaperonins were clustered based on 85% identity using Usearch (cluster_smallmem command, -usersort and –id 0.85 arguments). One representative sequence per cluster was chosen to create the Proteobacteria dataset reducing the number of sequences to 231 Cpn60s, corresponding to 180 bacterial species including *L. pneumophila's* HtpB. The clustering was biased, so that the sequences belonging to a bacterium that has multiple chaperonins tend to be selected as the representative sequences. Additionally, the Cpn60 sequence of *Thermus sp*. (GI: 504327362) was included as an outgroup reference. These 231 sequences were aligned using ClustalOmega, and then masked with Zorro [112], eliminating parts of the alignment with less than 0.4 confidence scores.

A second smaller data set was created, which only included amino acid sequences of Cpn60s that have been reported to have moonlighting functions, and their paralogs (if any). I named this dataset MoonlightingCpn60 and was composed of 48 chaperonins (including paralogs) from 35 bacterial species. The sequences in the MoonlightingCpn60 dataset were then aligned and masked as described for the Proteobacteria data set.

ProtTest3 [113] was used to select the best-fit amino acid substitution model for both chaperonin 60 alignments (one for the proteobacteria dataset and one for the MoonlightingCpn60 dataset) from the following models: JTT, LG, DCMut, MtREV, MtMam, MtArt, Dayhoff, WAG, RtREV, CpREV, Blosum62, VT, HIVb and HIVw plus three distribution parameters (+I, +G and +F). The LG+I+G substitution model [114] was the best fitting model I selected for both datasets as it had the smallest Akaike Information Criterion (AIC). The LG+I+G substitution model was thus used to estimate the maximum likelihood phylogeny using PhyML [115], and branch support was verified using 100 bootstrap replicates. Phylogenetic trees were visualized using FigTree [116].

CHAPTER 3 RESULTS

3.1 HTPB, BUT NOT GROEL, INTERACTS WITH THE HUMAN HOMOLOG OF ECM29 (hECM29)

3.1.1 Y2H bait plasmids were successfully constructed and correctly expressed their encoded bait proteins

The creation of Y2H bait plasmids began by cloning *htpB* and *groEL* into the pBS plasmid to create constructs pBS:*htpB* and pBS:*groEL* (Fig. 4). The correctness of the constructs and the fidelity whit which these chaperonin genes were PCR amplified was confirmed by DNA sequencing (Fig. 5). Evidence that *htpB* and *groEL* were sub-cloned in translational frame with the GAL4 DNA binding domain (BD) of the bait plasmid pGBKT7, to create the bait Y2H constructs pGBK:*htpB* and pGBK:*groEL*, is shown in Fig. 6. Genes cloned in the pGBKT7 vector are expressed both in yeast and *E. coli*. In yeast, bait proteins are expressed from the *ADH1* promoter generating fusion proteins composed of a GAL4 DNA binding domain followed by a c-Myc epitope tag in the N-terminus of HtpB or GroEL (Fig. 6A). Expression in *E. coli* is from the T7 promoter and the resulting fusion proteins have only the c-Myc tag in the N-terminus (Fig. 6). Expression of c-Myc tagged HtpB and GroEL in *E. coli* protein extracts was evaluated by western blot using anti-c-Myc and anti-Hsp60 antibodies. Anti-c-Myc antibody detected a ~60 KDa protein in the protein extracts of *E. coli* carrying pGBK:*htpB* and pGBK:gro*EL* but not in the strain carrying empty plasmid (pGBKT7); whereas, polyclonal anti-HtpB detected a ~60 KDa protein in all three extracts (Fig. 6B), the band detected in the empty vector lane (pGBKT7) corresponds to GroEL and is the result of cross reactivity of the antibody used (Fig. 6B). These results confirm the expression of tagged HtpB and GroEL in *E. coli,* and the integrity of the fusion proteins that were subsequently used in the Y2H experiments.

3.1.2 Y2H screening identified 29 potential HtpB interactions

The Y2HGold strain (Clontech's Matchmaker^{TM} Gold Yeast Two-Hybrid System) used in the screening has four integrated reporter genes: *AUR1-C* (resistance to aureobasidin A), *HIS3* (biosynthesis of histidine), *ADE2* (biosynthesis of adenine), and *MEL1* (encodes

Figure 4. Cloning of *htpB* and *groEL* in pBS. A: Photograph of a 1% agarose gel showing PCR products of ~1600 bp corresponding to *htpB* and *groEL* amplified from genomic DNA from *L. pneumophila strain* JR32 and *E. coli* strain DH5α, respectively. **B:** Photograph of a 1% agarose gel showing digestion products after digestion of pBS:*htpB* and pBS:*groEL* with *Eco*RI/*Bam*HI and with *Eco*RI/*Sal*I respectively. Two fragments are observed in the second and third lane of the gel, the largest corresponds to linearized pBS (~3000 bp) and the smallest (~1600 bp) is *htpB* (second lane) or *groEL* (third lane). For **A** and **B**, DNA sizes are given in base pairs under the bp column. First lane on the left of both gels contains a 1kbp ladder. **C:** Diagrams displaying pBS:*htpB* (left) and *pBS:groEL* (right) plasmids maps. Genes are represented as labelled rectangular boxes. Restriction sites used to clone *htpB* and *groEL* are shown by arrow heads. The vectors carry the ampicillin resistance gene (Amp^r). Sizes of both plasmids are indicated under the name of the construct in Kilobase pairs (Kbp).

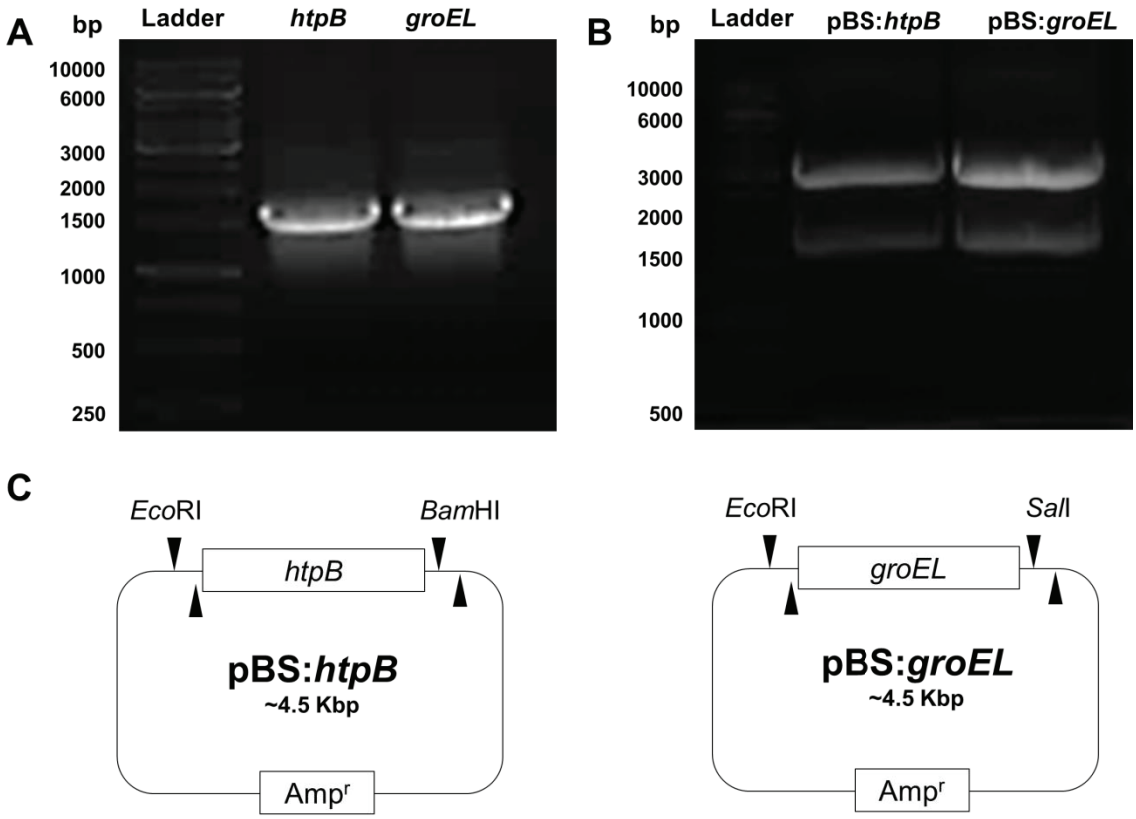

A
bp
Ladder
htpB
groEL
10000
6000
3000
2000
1500
1000
500
250
B
bp
Ladder
pBS:htpB
pBS:groEL
10000
6000
3000
2000
1500
1000
500
C
EcoRI
BamHI
htpB
pBS:htpB
~4.5 Kbp
Ampr
EcoRI
SalI
groEL
pBS:groEL
~4.5 Kbp
Ampr

Figure 5. Sequencing of HtpB and GroEL genes. Schematic representations of *htpB* (**A**) and *groEL* (**B**) sequence confirmation. pBS:*htpB* (**A**) and pBS:*groEL* (**B**) constructs were sequenced using various primers. The Sequencher software was used to assemble the readings obtained from each primer against the *htpB* (GI:AE017354.1:740047-741699) and *groEL* (GI:45686197) sequences retrieved from the NCBI database. Top green arrow is the reference sequence used for assembly; gene name and length (bp) are indicated above arrow. The following solid green arrows are forward readings whereas dashed red arrows are reverse readings. Primer names and length (bp) of the contigs are indicated above each arrow. Bottom bar represents coverage of the reference sequence by the contigs as indicated by legend.

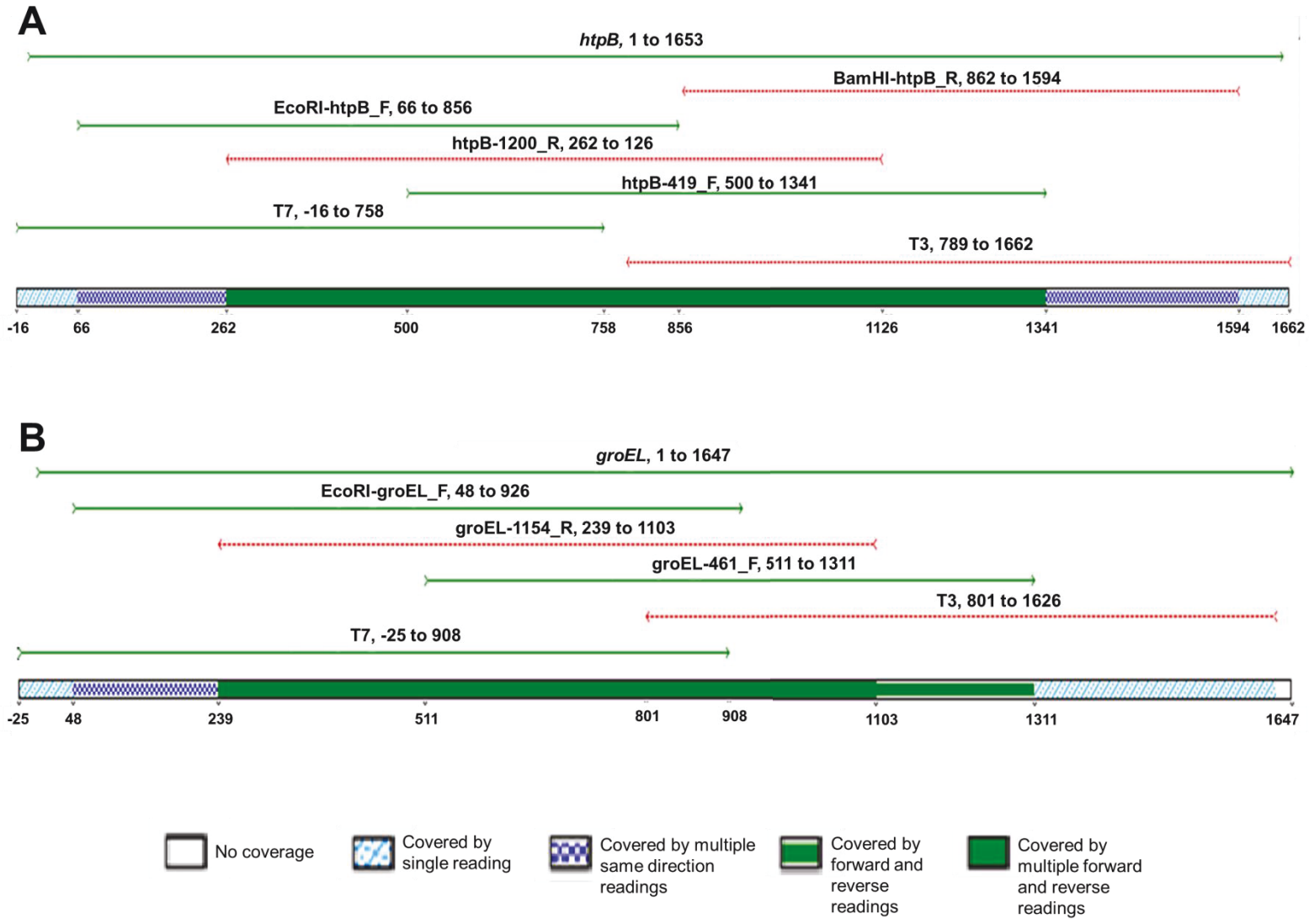
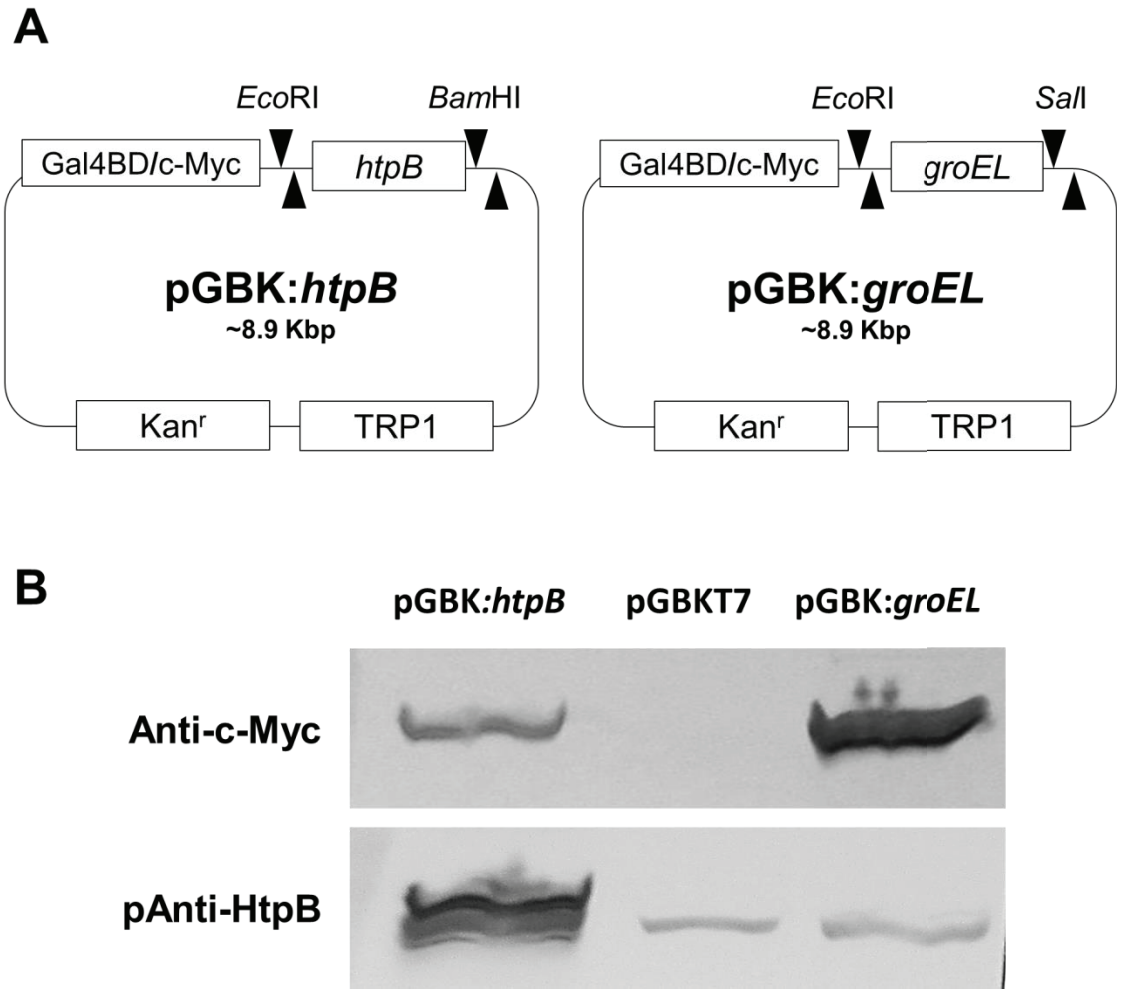

A
htpB, 1 to 1653
BamHI-htpB_R, 862 to 1594
EcoRI-htpB_F, 66 to 856
htpB-1200_R, 262 to 126
htpB-419_F, 500 to 1341
T7, -16 to 758
T3, 789 to 1662
-16
66
262
500
758
856
1126
1341
1594
1662
B
groEL, 1 to 1647
EcoRI-groEL_F, 48 to 926
groEL-1154_R, 239 to 1103
groEL-461_F, 511 to 1311
T3, 801 to 1626
T7, -25 to 908
-25
48
239
511
801
908
1103
1311
1647
No coverage
Covered by single reading
Covered by multiple same direction readings
Covered by forward and reverse readings
Covered by multiple forward and reverse readings

Figure 6. C-Myc tagged HtpB and GroEL are expressed in *E. coli*. A: Diagrams displaying pGBK:*htpB* (left) and pGBK:*groEL* (right) plasmid organization. *htpB* and *groEL* were excised from pBS:*htpB* and pBS:*groEL* respectively and subcloned into pGBKT7. The resultant constructs will express proteins fused to the Gal4 binding domain (Gal4BD) and c-Myc tagg on the N-terminus. Restriction sites used to subclone *htpB* and *groEL* are shown by arrow heads. Genes are represented as labelled boxes. The vectors carry the kanamycin resistance gene (Kan^r) for selection in *E. coli* and the TRP1 nutritional marker for selection in yeast. Sizes of both plasmids are indicated under the name of the construct in Kilobase pairs (Kbp). **B:** Western blots showing immunodetection of HtpB and GroEL in total protein from *E.coli* carrying the constructs pGBK:*htpB*, pGBKT7 (empty plasmid) and pGBK:*groEL* as indicated in each lane of the membrane. The top panel was probed with anti-c-Myc and the bottom panel with polyclonal anti-HtpB (pAnti-HtpB) antibodies.


A
EcoRI
BamHI
Gal4BD/c-Myc
htpB
pGBK:htpB
~8.9 Kbp
Kanr
TRP1
EcoRI
SalI
Gal4BD/c-Myc
groEL
pGBK:groEL
~8.9 Kbp
Kanr
TRP1
B
pGBK:htpB
pGBKT7
pGBK:groEL
Anti-c-Myc
pAnti-HtpB

alpha-galactosidase that turns colonies blue in medium containing X-alpha-Gal). These reporter genes are expressed in response to protein-protein interactions that bring the GAL4 transcriptional activation domain (in the library) and DNA binding domain (in the bait) into close proximity. Yeast carrying the plasmid pGBK:*htpB* (bait) was mated with yeast carrying the human library and then plated on double drop-out DDO/X/A (-Leu, -Trp, supplemented with aureobasidin A and X-alpha-gal) selective medium to identify putative interacting clones.

The number of clones screened was determined to be 3.4×10^6 and the mating efficiency of the screening was calculated as 17.9%. This indicates that more than 1 million diploids were screened ensuring detection of genuine interactions. Also, titration of the library indicated that the viability was 4.1×10^7 CFU/mL ensuring the quality of the library. All colonies (302 in total) that grew on DDO/X/A (moderate stringency selective medium) were picked up and re-streaked on fresh DDO/X/A selection medium , and then only blue colonies (196 in total), that indicate activation of *MEL1* reporter gene, were patched onto high stringency medium QDO/X/A (detects the activation of the four reporter genes) to identify potential interactions. Only 29 colonies, out of the 196 colonies patched, grew in QDO/X/A. Figure 7 summarizes the screening steps and the number and morphology of the colonies isolated. Size of the colonies was recorded as small colonies could indicate false interactions due to decay of the selection antibiotic (aureobasidin A) after 5 days of incubation.

3.1.3 Eight proteins were identified by sequencing of the isolated plasmid inserts

Prior to sequencing the isolated plasmids, duplicated inserts were identified by digesting with *Alu*I the PCR amplified prey library inserts. Figures 8 and 9 show that inserts of plasmids 13, 14, 17 and 34 had similar digestion profiles, as well as plasmids 32 and 35. Therefore, out of these six plasmids, only plasmids 34 and 35 were sequenced. The remaining 23 plasmids were all sequenced because no other evident similarities in their digestion profiles were found. From the sequencing results, only eight proteins were identified after searching in the NCBI database, of which the majority were represented by multiple clones. The amplified prey library inserts corresponding to proteins EIF1B, ECM29 and MRVI1 were the most abundant inserts found (Table 4).

3.1.4 HtpB interacts with the C-terminus of hECM29

To confirm the interactions identified under section 3.1.3, one library plasmid (prey) per protein identified by sequencing was selected (e.g., plasmid 20 for hECM29) and transformed back into *S. cerevisiae* strain Y2HGold along with the bait constructs for HtpB or GroEL, or the empty plasmid pGBKT7. An insert was classified as a genuine positive when both bait and prey were required to activate the reporter genes allowing growth of blue colonies on QDO/X/A medium, whereas a false positive clone would activate the expression of the reporters in the absence of bait (i.e., in the presence of empty pGBKT7). Six of the eight plasmids tested were false positives (Table 4), and I speculated that they encoded proteins with high affinity for random nucleotide sequences, explaining the non-specific activation of the reporter genes. For example, the most abundant insert, the *Homo sapiens* eukaryotic translation initiation factor 1B (EIF1B), would naturally bind non-specifically to nucleotides (Table 4). However, two genuine interactions were found several times: the inserts corresponding to the *Homo sapiens* KIAA0368 also known as ECM29 homolog (hECM29 from now on) and the *Homo sapiens* TOX high mobility group box family member 4 or TOX-4 (Table 4). Figure 10 shows the confirmation of the interaction of pasmids 20 (hECM29) and 15 (TOX-1) and an example of a false positive interaction (EIF1B plasmid 16). Growth was observed as blue colonies in both the double drop-out (DDO/X/A) and the high stringency medium (QDO/X/A), indicating a potential interaction between HtpB and EIF1B, hECM29 or TOX-1. However, when the empty plasmid (pGBKT7) was tested, there was no growth observed except in the EIF1B-pGBKT7 combination, showing that hECM29 and TOX-1 are true interactor proteins, but EIF1B is a false positive. Viability of the cultures was tested by plating 10 µl of the highest dilution (10^{-3}) in DDO medium that select only for the presence of bait and prey plasmids. All plasmids combinations tested showed similar growth (data no shown). Importantly, GroEL did not interact with hECM29, but it did with TOX-1 (Fig. 10). Therefore, hECM29 plasmid N° 20 was chosen to do all

Figure 7. Yeast two hybrid (Y2H) screening results. Table (top) and diagram (bottom) showing a summary of the Y2H screening performed. The table indicates the number and morphology of the colonies initially isolated in the DDO/X/A big plates. Colony size refers to diameter “>/=2mm“ means greater than or equal to 2mm and “<2mm” is less than 2mm. Clones that were able to grow in DDO/X/A plates (big plate on the left) were re-streaked on fresh DDO/X/A selective plates and then only blue colonies were patched on the high stringency QDO/X/A medium. Circular figures are plates and blue and white dots represent yeast colonies. The selective media used is showed under each plate. Number of colonies isolated each time is indicated on top of the red arrows. Number of putative interacting clones is indicated.

	Big colonies (>/=2mm)	Small colonies (<2mm)	Total
DDO/X/A 150mm plates * 50	222	80	302

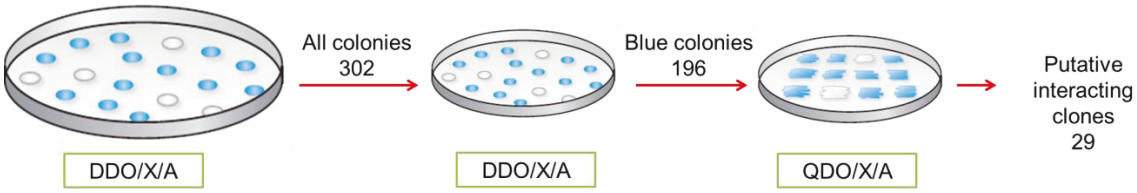


Figure 8. Plasmids 13, 14, 17 and 32 are duplicates. Pictures of 2% agarose gels where *Alu*I digested DNA inserts were separated. **A:** Digested and nondigested (indicated at the bottom of the gels) PCR amplified inserts from plasmids 13, 14, 16, 24, 32, 35 (labelled at the top of the gels) are shown. **B:** Digested and nondigested (indicated at the bottom of the gel) inserts from plasmids 30, 31, 34, 17, 18 and 19 (labelled at the top of the gel) are shown. Inserts that displayed identical digestion profiles (duplicates), as marked by the black horizontal square brackets, were not sequenced (red numbers). For **A** and **B**, DNA sizes are given in base pairs in the column labelled bp. First lane on the left of the gels contains a 100 bp ladder.

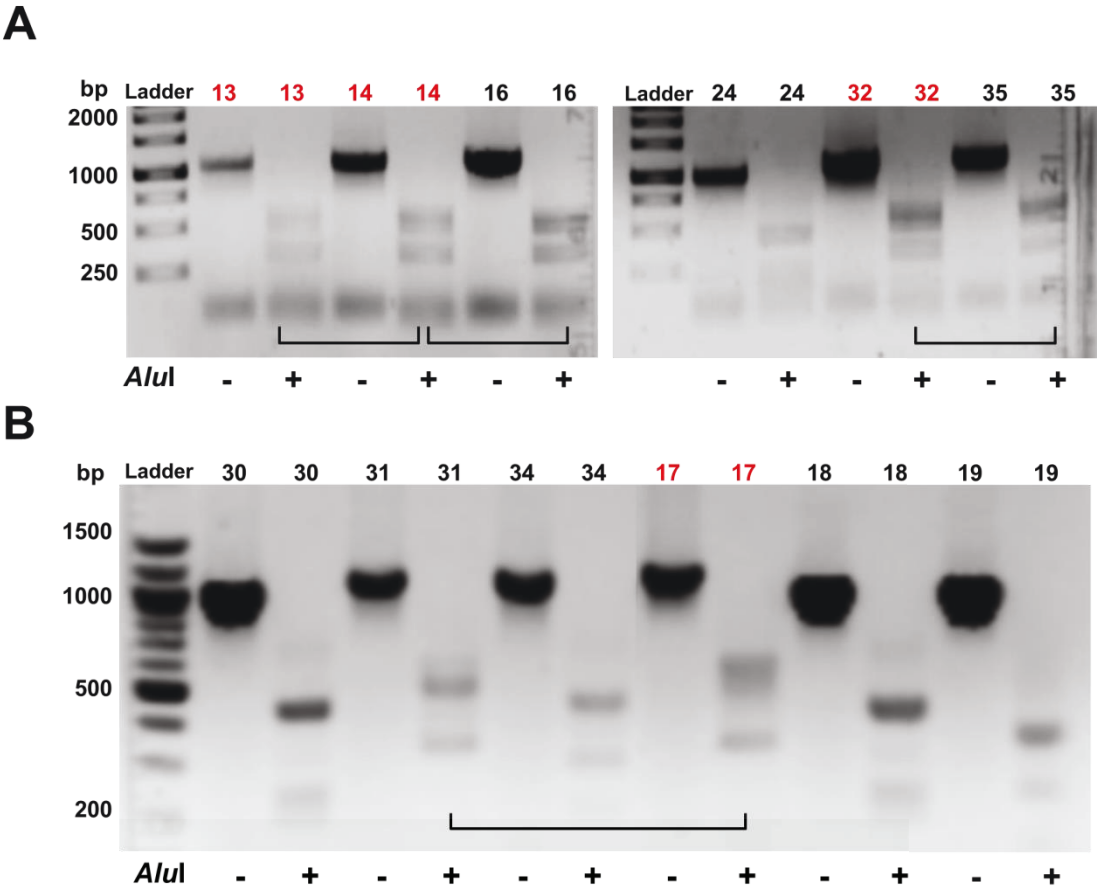

A
bp
Ladder
13
13
14
14
16
16
2000
1000
500
250
AluI
-
+
-
+
-
+
Ladder
24
24
32
32
35
35
-
+
-
+
-
+
B
bp
Ladder
30
30
31
31
34
34
17
17
18
18
19
19
1500
1000
500
200
AluI
-
+
-
+
-
+
-
+
-
+
-
+

Figure 9. Plasmids 14, 16 and 17 are duplicates. Pictures of a 2% agarose gel where *Alu*I digested DNA inserts were separated. Digested inserts from plasmids 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27 (labelled at the top of the gel) are shown. Inserts that displayed identical digestion profiles (duplicates), as marked by the black horizontal square brackets, were not sequenced (red numbers). DNA sizes are given in base pairs in the column labelled bp. First lane on the left of the gel contains a 100 bp ladder.

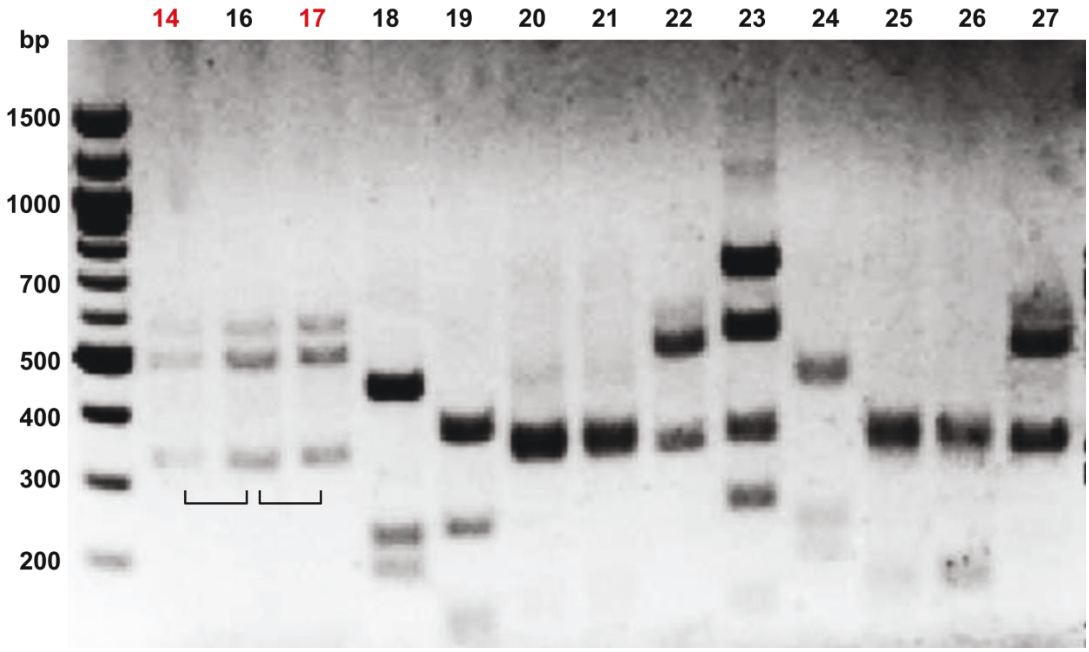
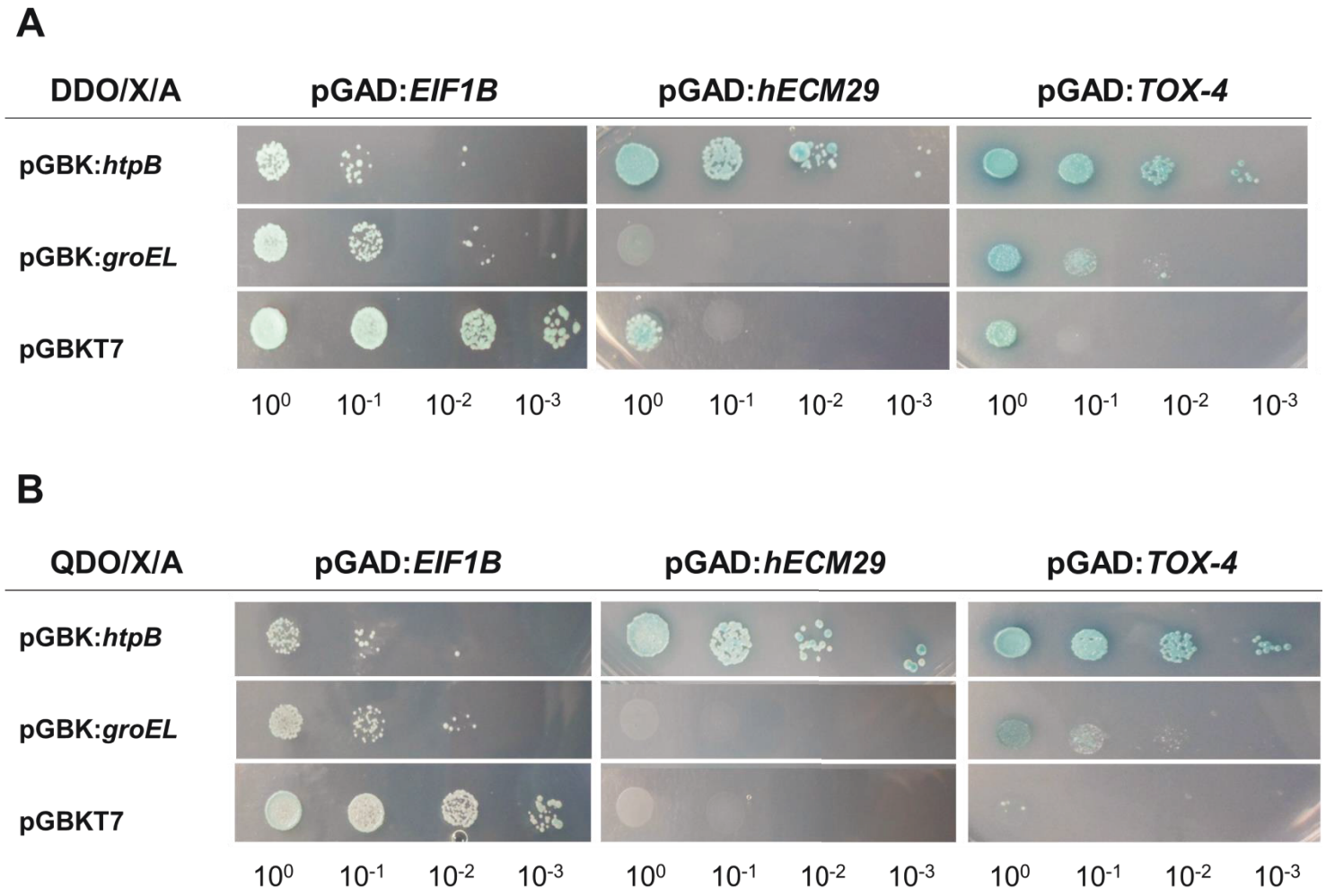

14
16
17
18
19
20
21
22
23
24
25
26
27
bp
1500
1000
700
500
400
300
200

Table 4. Summary of Y2H screening results

Plasmid name	Inserts**	Identified gene [GI]	Protein function	Interaction confirmation^
19	1	Homo sapiens ATPase, Na+/K+ transporting, beta 1 polypeptide. [49574487]	Hydrolysis of ATP coupled with the exchange of Na+ and K+ ions across the plasma membrane.	Negative
5	1	Homo sapiens SET domain containing 2 (SETD2). [378404895]	Histone-lysine N-methyltransferase activity.	False Positive
1, 6	2	Homo sapiens splicing factor, arginine/serine-rich 5. [34189763]	Contains 2 RNA recognition motif domains.	False Positive
3, 4	2	Homo sapiens hypothetical protein LOC284412. [28175098]	None.	Negative
15, 23	2	Homo sapiens TOX high mobility group box family member 4. [15489161]	Control of chromatin structure and cell cycle progression during the transition from mitosis into interphase.	Positive
18, 24, 30, 35, (32)*	5	Homo sapiens murine retrovirus integration site 1 homolog (MRVI1). [332634571]	It is a substrate of cGMP-dependent kinase-1 (PKG1) that can function as a regulator of IP3-induced calcium release.	False Positive
20, 21, 25, 26, 27, 33	6	Homo sapiens KIAA0368 or ECM29 homolog. [122937210]	Binds to the 26 S proteasome, motor proteins, endosome, endoplasmic reticulum and centrosome.	Positive
12, 16, 22, 28, 29, 31, 34, (13, 14, 17)*	10	Homo sapiens eukaryotic translation initiation factor 1B (EIF1B). [13937791]	Poly(A) RNA binding, translation initiation factor activity.	False Positive

* Duplicated plasmids identified by *Alu*I digestion
** Number of inserts that corresponded to the same gene
[] Gene identification number
^ Confirmation by Y2H

Figure 10. HtpB interacts with hECM29 and TOX-4. Plasmids isolated from the library screening (pGAD:*EIF1B*, pGAD:*hECM29* and pGAD:*TOX-4*) were co-transformed with the bait plasmids pGBK:*htpB*, pGBK:*groEL* or the empty plasmid (pGBKT7) into *S. cerevisiae*. Transformants were grown in DDO broth overnight and then 10µl drops of undiluted (10^0) and serially diluted (10^{-1}, 10^{-2}, 10^{-3}) culture were spotted onto DDO/X/A plates (**A**) and QDO/X/A plates (**B**). Pictures were taken after 5 days of incubation at 30°C.


A
DDO/X/A
pGAD:EIF1B
pGAD:hECM29
pGAD:TOX-4
pGBK:htpB
pGBK:groEL
pGBKT7
10^0 10^-1 10^-2 10^-3
B
QDO/X/A
pGAD:EIF1B
pGAD:hECM29
pGAD:TOX-4
pGBK:htpB
pGBK:groEL
pGBKT7
10^0 10^-1 10^-2 10^-3

subsequent experiments to evaluate the amino acids involved in the HtpB-hECM29 interaction (see section 3.2.3 below).

Interaction of HtpB and hECM29 was also confirmed by co-immunoprecipitation experiments. hECM29 was co-precipitated when HtpB was pulled down with the anti-c-Myc antibody and HtpB was co-precipitated when hECM29 was pulled down with the anti-HA tag antibody (Fig. 11). These results confirm the direct physical interaction between hECM29 and HtpB expressed in yeast (Fig. 11). Additionally, the DNA inserts corresponding to hECM29 (plasmids 20, 21, 25, 26, 27 and 33) were aligned against the mRNA sequence of hECM29 (GI: 122937210) showing that all the isolated plasmids corresponded to the C-terminus of hECM29 (Fig. 12).

3.2 HTPB AMINO ACIDS K298, N507, H473 AND K474 ARE REQUIRED FOR THE INTERACTION WITH hECM29

3.2.1 Substitutions in amino acids involved in oligomerization were found between GroEL and HtpB

A literature review was performed to find GroEL amino acids that have an identified role in folding and/or multimerization. GroEL has been the subject of intensive studies; therefore, amino acids involved in ATP binding (23), polypeptide recognition (16), intra-ring (39) and inter-ring interaction (14) have been identified [47,50,53,55,56,117-121]. Based on the alignment between GroEL and HtpB, homologous positions of the amino acids described were identified in HtpB (Table 5) and mapped into HtpB's predicted 3D structure (Fig. 13). These amino acids are important for correct functioning or oligomerization of GroEL; however, substitutions were found in amino acids involved in the interaction between same ring monomers (e.g., R246K), inter-ring contacts (e.g., D435E) and surprisingly in the highly conserved ATP-binding pocket (M32L).

Figure 11. HtpB interacts with hECM29 in *S. cerevisiae*. Explanatory diagrams (left) and photos of western blots (right) of co-immunoprecipitation (Co-IP) assays performed using protein from *S. cerevisiae* expressing c-Myc and HA tagged versions of HtpB and hECM29, respectively. **A:** The diagram shows the HtpB-hECM29 interaction complex captured by an anti-HA antibody immobilized on a protein A/G agarose bead (Prot. A/G). The western blot, probed with anti-c-Myc, shows a ~60 kDa band corresponding to HtpB detected in the anti-HA immunoprecipitate (IP-HA) but not in the supernatant (Unbound) fraction of the Co-IP. **B:** The diagram shows the HtpB-hECM29 interaction complex captured by an anti-c-Myc antibody immobilized on a protein A/G agarose bead (Prot. A/G). The western blot, probed with anti-HA, shows a ~80 kDa band corresponding to hECM29 detected in the anti-c-Myc immunoprecipitate (IP-c-Myc) but not in the supernatant (Unbound) fraction of the Co-IP. Molecular weight in kilo Daltons (kDa) is shown on the left side of both western blots and the first lane contains a protein ladder (11–245 kDa).

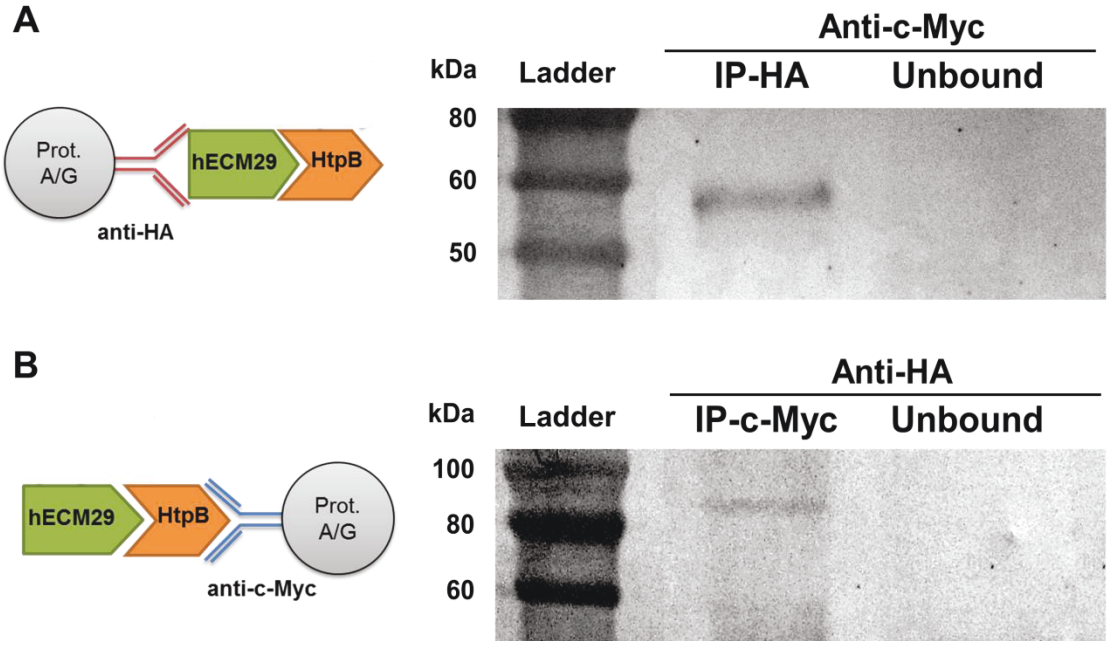

A
Prot.
A/G
hECM29
HtpB
anti-HA
Anti-c-Myc
kDa
Ladder
IP-HA
Unbound
80
60
50
B
hECM29
HtpB
Prot.
A/G
anti-c-Myc
Anti-HA
kDa
Ladder
IP-c-Myc
Unbound
100
80
60

Figure 12. HtpB interacts with the C-terminus of hECM29. Graphical overview of blastn alignment of the hECM29 gene (*KIAA0368*, GI:122937210) and DNA inserts identified in the Y2H screening. Inserts were PCR amplified from the plasmids isolated from the Y2H screening and sequenced using primer T7. Red bar on top represents the mRNA sequence of *KIAA0368.* Numbers under the red bar are bases. Shorter red bars under *KIAA0368* are the DNA insert sequences. Plasmid names are shown at the left side of each bar.

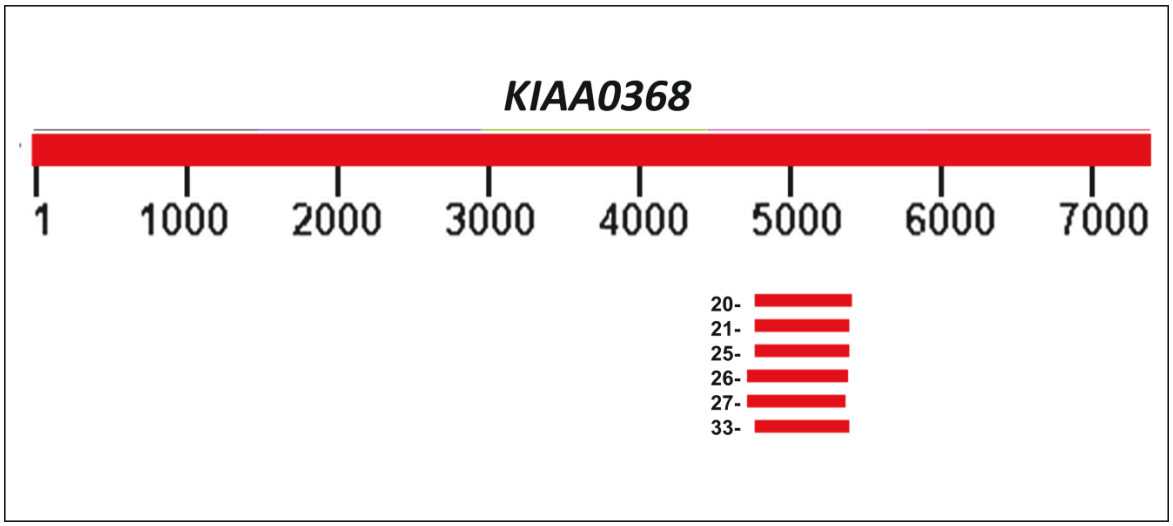
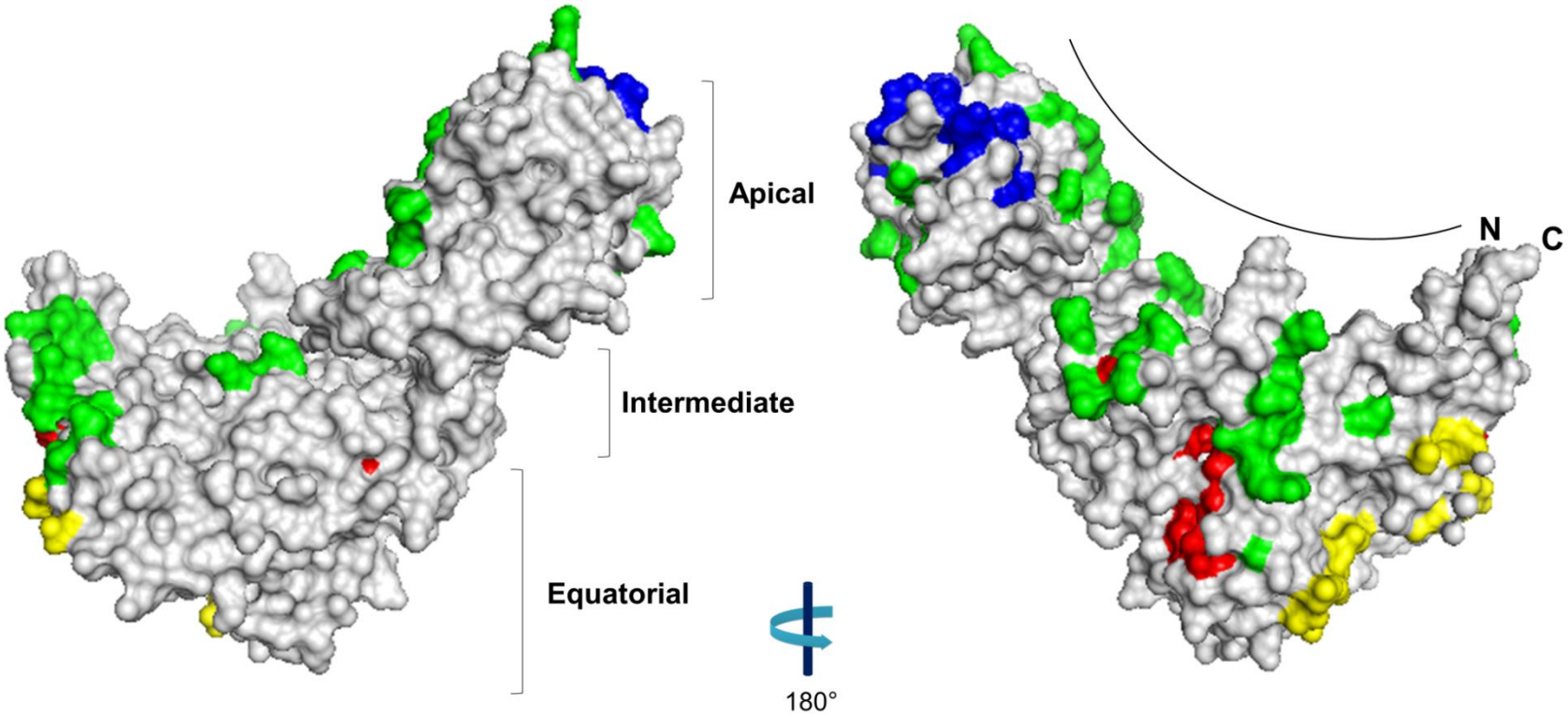

KIAA0368
1
1000
2000
3000
4000
5000
6000
7000
20-
21-
25-
26-
27-
33-

Table 5. HtpB amino acids that have a role in protein folding.

Intra-ring contacts	Nucleotide binding pocket	Polypeptide recognition	Inter-rings contacts
L7(47.14) [V]	R14(6.33)	Y200(10.29)	D12(133.91)
A23(125.52) [V]	T31(1)	S202(6.66)	L15(113.12) [V]
R37(9.63)	M32(16.1) [L]	Y204(1.65)	K106(36)
N38(9.8)	G33(1)	F205(14.17)	A109(48.43)
V39(5.59)	P34(1)	R232(32.97)	A110(24.42)
V40(44.13)	K52(6.27)	L235(31.35)	G111(2.21)
L41(39.88)	D53(1)	L238(3.42)	M112(49.9)
E42(62.78)[D]	G54(1)	E239(9.22)	D435(146.41) [E]
K81(21.07)	D88(1)	A242(46.79)	R446(79.93)
D84(9.44)	T92(1.64)	L260(7.37)	R453(48.94)
N113(22.45)	I151(29.03)	T262(14.39)	E462(32.73)
M115(68.31)	S152(4.14)	V264(19.47)	S464(50.97)
N182(43.17) [T]	A153(23.45)	V265(32.63)	V465(42.35)
L184(66.68)	A384(8.56)	N266(1.49)	N468(101.41)
R198(10.2)	D399(1.72)	R269(10.8)	
N208(47.69) [K]	A407(3.87)	I271(54.8)	
E217(117.15)	G416(1)		
K227(3.43)	I455(8.15)		
R232(32.97)	N480(41.39)		
R246(79.05) [K]	A481(13.62)		
E253(14.91)	A482(88.43)		
E256(41.75)	I494(30.94)		
E258(12.87)	D496(1.4)		
K273(74.67)			
F282(7.64)			
D284(16.16)			
R285(3.19)			
R286(2)			
Y361(9.78)			
A385(35.43)			
E387(4.19)			
M390(22.23)			
A459(37.54) [C]			
T517(4.29)			
E519(25.31)			
C520(50.52)			
M521(64.93)			
V522(68.71)			
A523(69.18) [T]			

ET rank is shown between parentheses.
Square brackets indicate GroEL (*E. coli*) amino acids only if different from HtpB.
Amino acids are in single letter code follow by their position in the HtpB sequence.

Figure 13. Amino acids involved in folding and multimerization of HtpB. Diagram showing the HtpB predicted structure as two surface representations generated after 180 degree rotation of the molecule. Amino acids that have a known function in GroEL were identified and mapped on the HtpB structure. Blue: polypeptide recognition, red: nucleotide binding pocket, green: intra-ring contacts and yellow: inter-rings contacts. HtpB apical, intermediate and equatorial domains are shown. The arc on the left structure indicates the area of HtpB that faces the inside of the chamber. **N**: N-terminus, **C**: C-terminus.


Apical
Intermediate
Equatorial
180°
N
C

3.2.2 Amino acids M68, M212, S236, K298, N507 and peptide 471-475 are possibly involved in HtpB moonlighting functions

Protein sequences of HtpB orthologs were obtained from the NCBI database; duplicates, fragments and Archaea and Eukarya chaperonins were eliminated from the data set used to run the ET analysis. The remaining 1373 Cpn60 sequences provided the ET rankings depicted in appendix A. Rankings obtained from the ET analysis ranged from 1.00 to 234.92. Approximately 300 amino acids (first two columns of the histogram in Fig. 14) have ranks under 50, indicating high conservation of the HtpB sequence along its molecular evolution. The ET ranks were mapped onto the 3D structural model of HtpB and the ranks of superficial amino acids are shown in figure 15A. Twenty seven amino acids were conserved in all sequences in the MSA and therefore had an ET rank of 1.00. Out of these 27 amino acids, those located on the surface of HtpB are shown in red in Figure 15A. Other evolutionary important amino acids with low ET ranks (orange, yellow and green) were mainly distributed along the molecular surface of HtpB that faces the internal cavity of the folding chamber, the apical domain (which interacts with the unfolded substrate) and the ATP-binding pocket (Figs. 15A compared to Fig. 13 and Table 5). Interestingly, the amino acids involved in recognition of the un-folded protein substrate and nucleotide binding had high ET ranks indicating high evolutionary importance and conservation of these amino acids among the sequences analysed (Table 5). Some of the amino acids that participate in ATP recognition were also fully conserved in all sequences (ET rank = 1). Less conserved amino acids had higher ET ranks and were mostly located in the equatorial domain and in areas of HtpB that do not have contact with any other monomer (Fig. 15A compared to Fig. 13). In addition, it was surprising that some of the amino acids involved in the interactions between rings had high ET ranks (Table 5).

To find the amino acids most likely involved in HtpB moonlighting functions, I focused on the HtpB versus GroEL alignment. HtpB and GroEL are 76% identical, meaning that only 137 out of 550 residues are different between them. The ET ranks of these 137 substitutions are mapped in Fig. 15B. Most of the substitutions have high ET ranks (blue and purple in figure 15B) but some amino acids have lower ET ranks (yellow, green and

Fig 14. Distribution of the evolutionary trace (ET) ranks. Histogram showing the distribution of the ET ranks of HtpB. The x axis gives the ET rank values in intervals of 25 arbitrary units and Y axis gives the number of amino acids showing that particular ET rank interval.

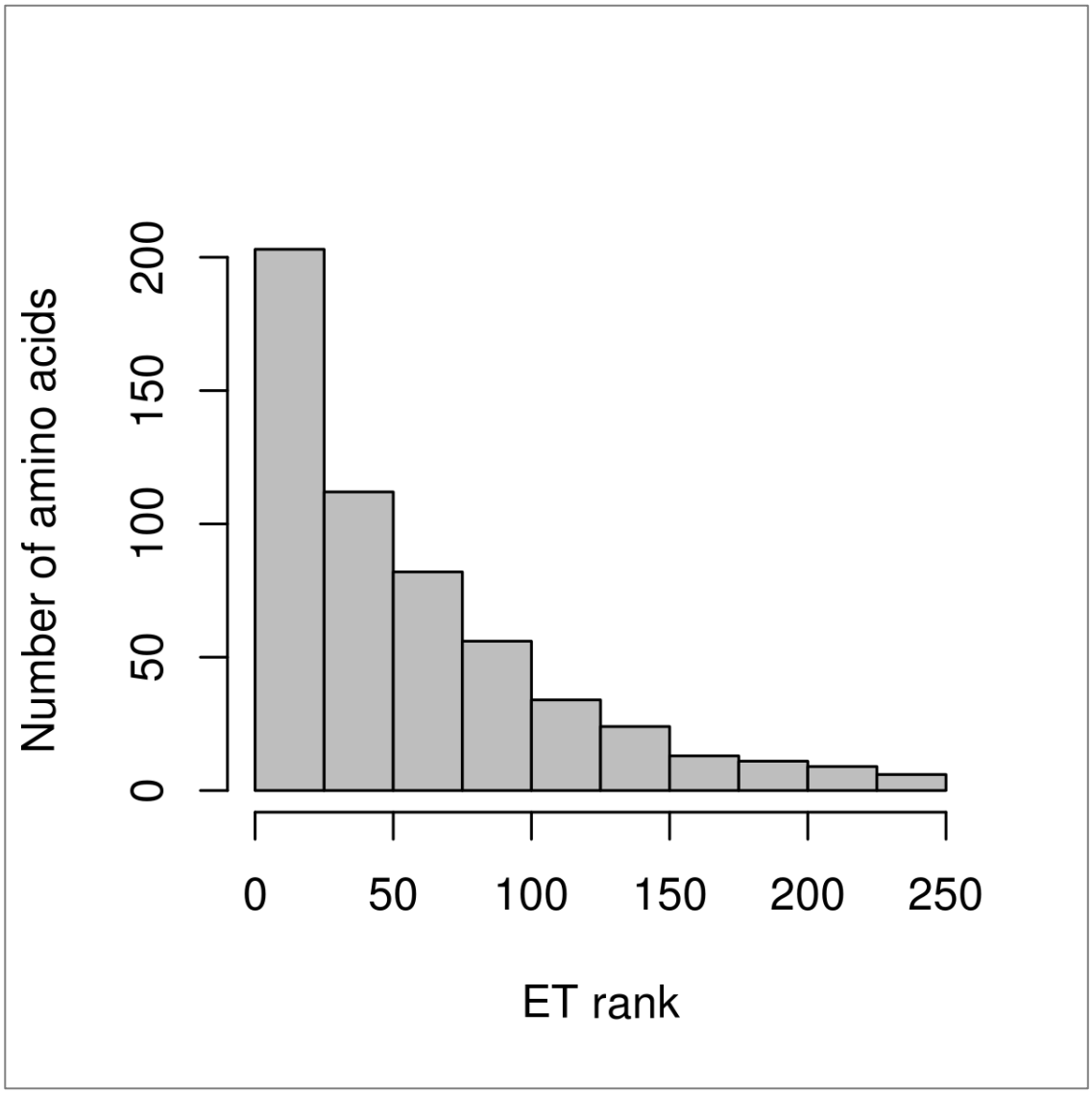
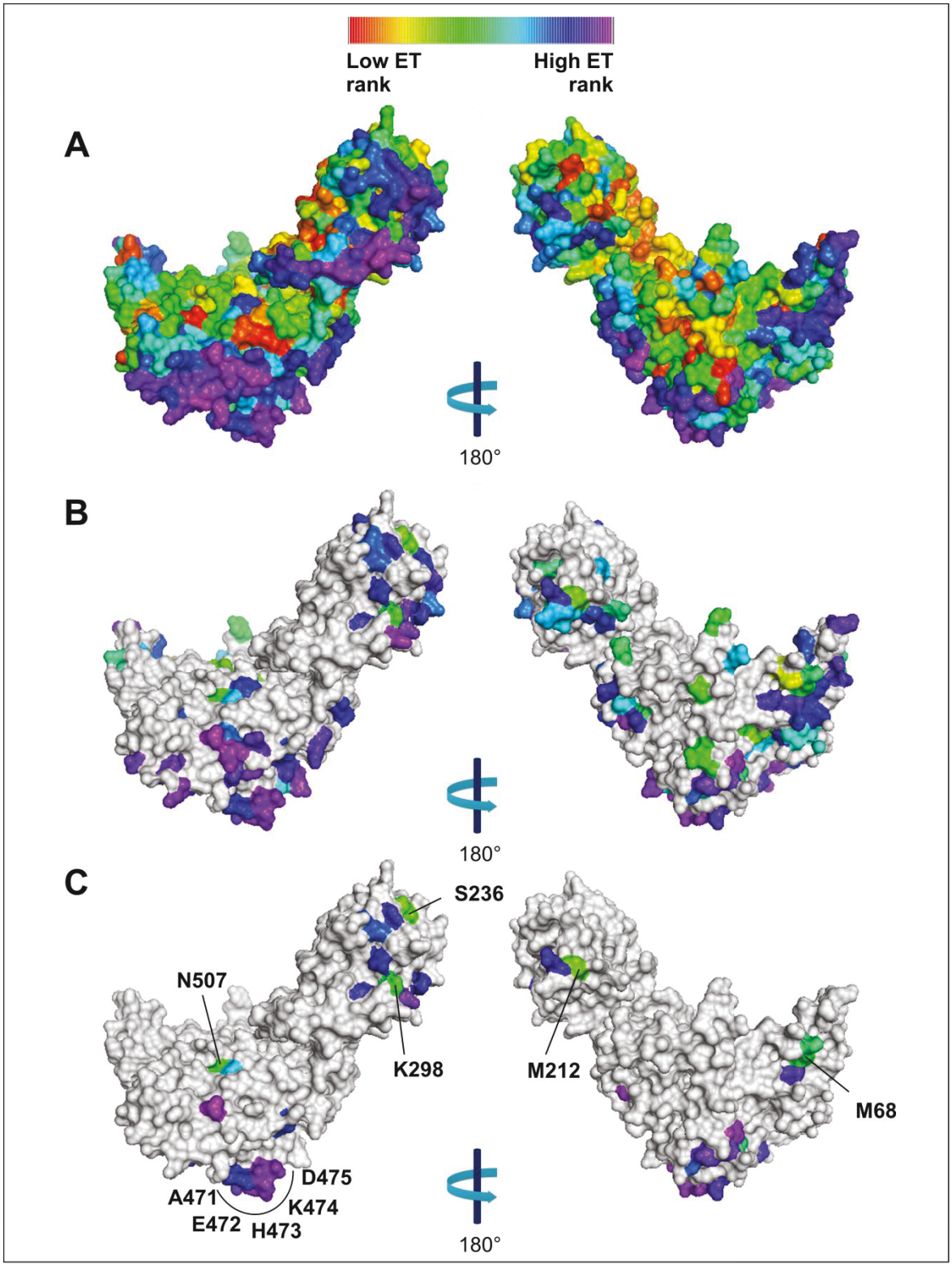

Number of amino acids
0
50
100
150
200
0
50
100
150
200
250
ET rank

Figure 15. Evolutionary trace ranks mapped onto the HtpB structure. Diagram showing HtpB predicted structures as surface representations generated after a 180 degrees rotation of the molecule. **A:** ET ranks obtained from the ET analysis mapped onto HtpB structure. **B:** ET ranks of amino acid positions where substitutions between HtpB and GroEL were found. **C:** From the amino acid ET ranks mapped in B, the less likely to occur substitutions (negative BLOSUM62 scores) were selected and their ET ranks were mapped onto the HtpB structure. Amino acids selected for mutagenesis are labeled: M68, M212, S236, K298, N507 and 471-475. Colors illustrate the ET ranks of the amino acids as indicated in the top bar. Lower ET ranks indicate high evolutionary importance and higher ET ranks indicate low evolutionary importance.


Low ET rank
High ET rank
A
180°
B
180°
C
S236
N507
K298
M212
M68
D475
A471
K474
E472
H473
180°

cyan in figure 15B) meaning that these positions are more evolutionarily important. However, the number of candidate residues (137 substitutions) was still too high to evaluate all of them by mutagenesis. Therefore, the BLOSUM62 matrix was used to narrow down the candidate amino acids by selecting only the less likely substitutions resulting in the selection of 41 residues (Table 6). The ET ranks of these 41 residues are mapped onto the HtpB structure in Figure 15C. The vast majority of less likely substitutions were located in the equatorial domain and have high ET ranks (Table 6). Therefore, from these 41 residues only 5 amino acids with low ET rank were selected for mutagenesis: M68 (ET rank=29.84), M212 (ET rank=21.73), S236 (ET rank=22.91), K298 (ET rank=38.33) and N507 (ET rank=29.53). These residues did not form any obvious cluster, and were scattered on the molecular surface of HtpB's 3D structure. In spite of having high ET ranks (109.10 to 185.86), a short peptide of five HtpB amino acids (471-475) was also selected, mainly because they represent a surface-exposed domain that forms a randomly coiled loop (Figs. 15C and 16) that could be available to interact with other proteins.

Taken together, the results show that the Cpn60 molecular surfaces that do not have any reported yet role in protein-folding are less evolutionary important and more prone to undergo amino acid substitutions. Consequently, most of the substitution between HtpB and GroEL are in positions not related with protein-folding and showed high ET ranks. The ET results coupled with BLOSUM62 analysis strongly support the hypothesis that HtpB moonlighting functions are due to few scattered amino acid substitutions rather than the presence of a specific domain or active site.

3.2.3 HtpB mutants K298A, N507A, H473A and K474A exhibit a partially impaired interaction with hECM29

To evaluate the role of the amino acids predicted by the bioinformatics analysis, single or multiple amino acid replacements were introduced into HtpB (Table 2). All introduced mutations were confirmed by sequencing but, initially, some of the single amino acid substitutions were also confirmed by restriction enzyme digestion. Substitution of methionine 212 by alanine induced loss of a *Fat*I restriction site in *htpB* and substitution of asparagine 212 by alanine generated a new PstI restriction site (Figs. 17 and 18).

Table 6. Less likely substitutions between HtpB and GroEL

Aligment position[a]	HtpB (aa)	GroEL (aa)	Variability[b] (N°)	Variability[c] (aa)	rvET rank[d]	Blosum score
3	M	A	10	AMVTSPGKLD	46.31	-1
19	A	R	15	RKEASVQDNHTILMY	130.61	-1
65	H	D	10	DCNHESQAKG	46.08	-1
68	M	E	13	ERKAQMIPLNHSV	29.84	-2
105	H	L	14	LINAHMVCSFYTQR	66.77	-3
126	L	T	15	TEAKINVDQLSGRHY	210.52	-1
137	K	V	13	VKIQSRHETNDLA	93.13	-2
161	A	K	15	KDNERQSLTAHGVXI	201.68	-1
209	Q	P	14	PSANTRQKGVHLMD	120.55	-1
212	M	G	7	GQMLARS	21.73	-3
214	C	V	7	VACITSG	76.14	-1
218	H	S	11	SDKENRQTHGA	115.53	-1
236	S	P	8	PHSTNGAQ	22.91	-1
295	I	T	7	TIVANCM	113.27	-1
298	K	G	10	GAKNDSQHRE	38.33	-2
300	Q	T	14	TQEIVLSKHRMNDA	112.4	-1
308	K	M	17	MLRISFYKGANVTHCDQ	85.41	-1
312	G	K	12	KNTDSAQGEHRM	116.02	-2
337	E	V	20	VALDNGMSKI.EFTRQHYCP	157.89	-2
340	A	E	17	ETSKPAQGDNVCRH.YL	169.34	-1
342	E	A	18	AQDSNEVKTRHMIGLPFY	187.52	-1
352	A	Q	14	QAKSVTGNRHMILE	151.18	-1
424	Q	A	17	AYISTVLKQREGMFHCN	149.47	-1
426	A	K	19	KPTASVQDYEIGCLNRHMF	173.35	-1
428	D	A	18	ALESTKHDQVGFRN.ICP	191.2	-2
444	L	A	7	AVILFMT	64.13	-1
445	R	L	16	LKIFRMEQVAYTSCGN	122.42	-2
457	T	L	18	LFHAVEIKTYNQSDRGCM	194.97	-1
461	Y	E	19	ELKFYIDQAVGSMTHWRCN	142.12	-2
463	A	P	10	PGASENDTRK	46.27	-1
469	K	T	14	TKRQANESHMIYDG	126.5	-1
471	A	K	14	KRLMAIQESVTGHY	115.7	-1
472	E	G	17	GNHSAE.KQTDRLMCVI	185.86	-2
473	H	G	17	G.RAKSNTEVHDQPLMC	109.1	-2
474	K	D	15	DPASEKQ.TVGNCRH	183.88	-1
475	D	G	20	GAVSLPKERYITDQFNHW.M	184.4	-1
484	G	E	12	EGDNMKLFHRSA	82.24	-2
503	M	S	14	SCTIVMNAYLFQHG	65.83	-1
507	N	Y	11	YNSHDKAFGLT	29.53	-2
530	E	A	15	AKGS.PENDTVQIMH	219.66	-1
536	D	G	15	G.ADQSEMNPYHVTI	118.38	-1

[a] Position of the amino acid (aa) in the multiple sequence alignment.
[b] Number of different aa found at that position in the multiple sequence alignment.
[c] Variability of the aa found at that position in the multiple sequence alignment. Dots are gaps in the alignment.
[d] Evolutionary trace (ET) real value (see methods, section 2.5.1).

Figure 16. Analysis of the secondary structure of peptide 471-475 shows that it is exposed and has a random coil secondary structure. Secondary structure was predicted using VADAR and visualized using PyMol. A: table showing prediction of the secondary structure of the amino acids forming the peptide 471-475. Positions of interest are indicated in red. Residue names (RES. NAME.) are shown as three letters code. Secondary structure of the residue (SCND STRUC) is denoted by the letters, H: alpha helix, C: random coil and B: beta sheet. The first three letters represent the prediction of three independent methods and the fourth separate letter is the consensus secondary structure for the amino acid. Values under the RES. ASA column corresponds to the residue accessible surface area measured in square angstroms. Higher ASA values indicate higher accessibility of the amino acid to be in contact with a water molecule. B: Diagram showing a close-up to the ribbon structure of HtpB showing the spatial organization of the peptide of interest (blue) corresponding to the amino acids shown in panel A.

A

	RES. NAME	SCND STRUC	RES. ASA
	LYS	HHH H	99.9
	VAL	HHH H	9.3
471	ALA	CHH H	52.8
472	GLU	CCC C	83.7
473	HIS	CCC C	127.1
474	LYS	CCC C	173.7
475	ASP	CBB B	98.5
	ASN	BBB B	25.2
	TYR	BBB B	50.5

Figure 17. Substitution of methionine 212 by alanine induces loss of a *Fat*I restriction site in *htpB*. A: schematic representation of PCR amplification and digestion with *Fat*I restriction anzime of a ~800 bp fragment amplified from WT *htpB* and *htpB* M212A. Expected sizes of the digestion products are indicated in base pairs on top of each fragment. Blue bars represent DNA. Red arrows are the primer binding regions for htpB-419_F and htpB-1200_R primers. Dashed lines show the section of the gene that was PCR amplified. Red circles indicate *Fat*I restriction sites. **B:** Photo showing separation of the digestion products from *htpB* M212A and WT *htpB* on a 2% agarose gel. DNA size is indicated on the left side of the gel. First lane is a 100 bp ladder.

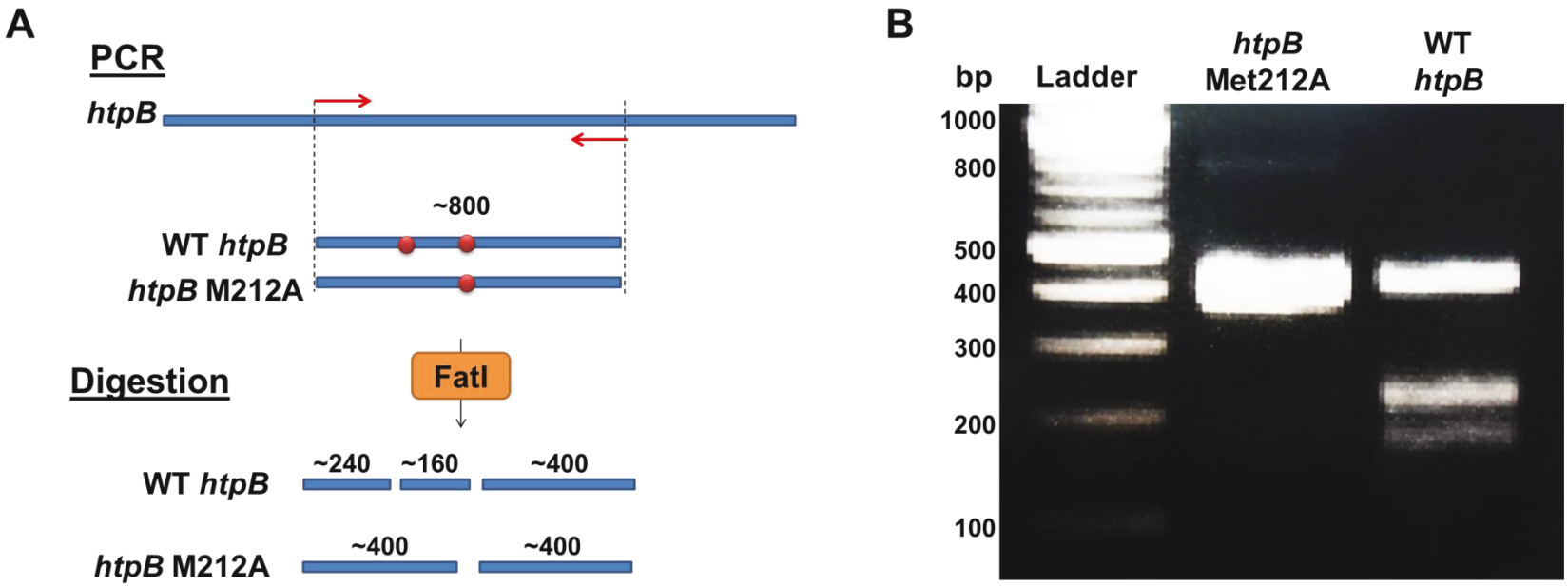
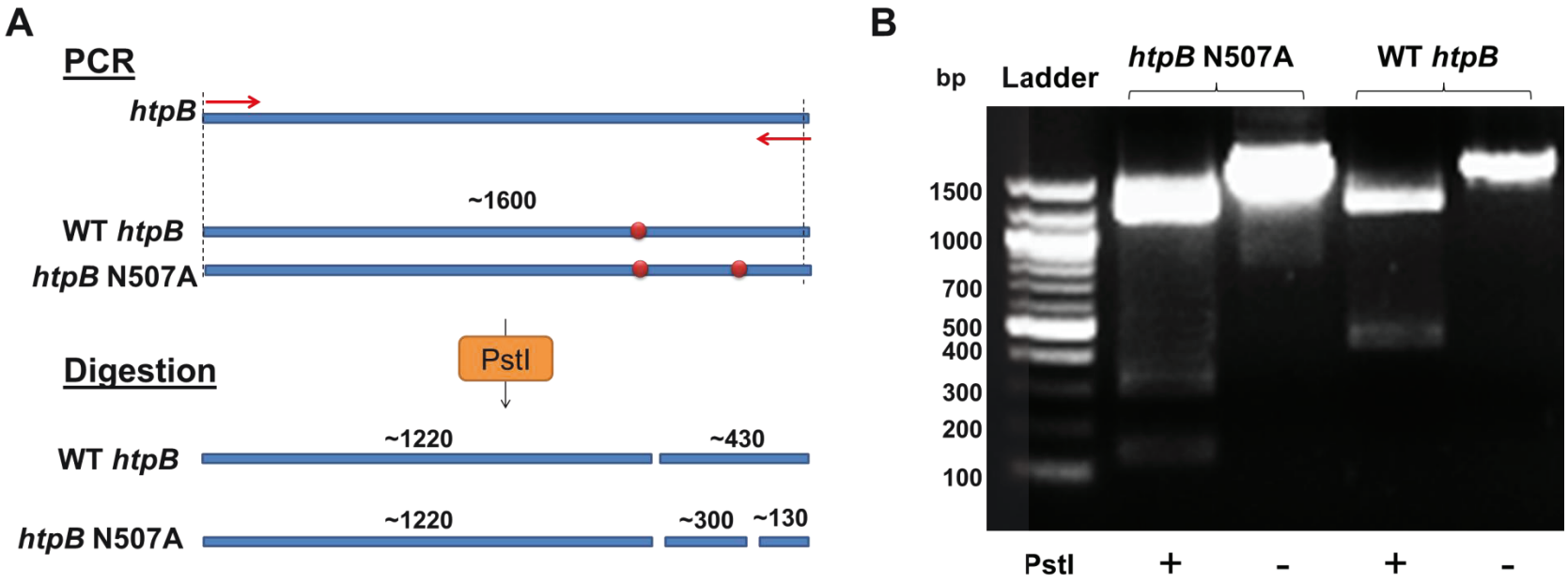

A
PCR
htpB
~800
WT htpB
htpB M212A
Digestion
FatI
~240
~160
~400
WT htpB
~400
~400
htpB M212A
B
bp
Ladder
htpB
Met212A
WT
htpB
1000
800
500
400
300
200
100

Figure 18. Substitution of asparagine 212 by alanine generates a new PstI restriction site in *htpB*. A: schematic representation of PCR amplification of the WT *htpB* and *htpB* N507A genes, followed by digestion with PstI. Expected sizes of the digestion products are indicated in base pairs on top of each fragment. Blue bars represent DNA. Red arrows are primer binding regions for EcoRI-HtpB_F and BamHI-HtpB_R primers. Dashed lines indicate the fragment that was PCR amplified (~1600 bp). Red circles are PstI restriction sites. **B:** Photo showing separation of the digested and undigested (as indicated at the bottom of the gel) PCR products from *htpB* N507A and WT *htpB* on a 2% agarose gel. DNA size is indicated on the left side of the gel. First lane is a 100 bp ladder.


A
PCR
htpB
~1600
WT htpB
htpB N507A
Digestion
PstI
~1220
~430
WT htpB
~1220
~300
~130
htpB N507A
B
bp
Ladder
htpB N507A
WT htpB
1500
1000
700
500
400
300
200
100
PstI
+
-
+
-

Following confirmation of the mutations, thirteen mutant variables of HtpB were tested on their ability to interact with hECM29 by Y2H. *S. cerevisiae* co-transformed with plasmid N° 20 isolated from the Y2H screening expressing hECM29 and WT HtpB, or HtpB mutant variants, were grown on selective QDO/X/A plates (plate assay). When HtpB and hECM29 interacted, activation of the reporters (*AUR1-C, ADE2, HIS3,* and *MEL1*) was induced allowing yeast to grow in the high stringency selection medium (QDO/X/A). Given that GroEL does not interact with hECM29, it was used as a negative control for the interaction. Single amino acids substitutions at positions K298A, N507A, H473A and K474A partially impaired HtpB-ECM29 interaction showing less growth compared with WT-HtpB (Fig. 19). N507, H473 and K474 are in the equatorial domain of HtpB and K298 forms part of the apical domain (Fig. 15C). The strength of the interaction was also affected when multiple amino acids were replaced by alanine in mutants MMSKN/A and 472-475/A. Additionally, when the selected amino acids in HtpB were simultaneously substituted by the corresponding GroEL amino acids mutant EGPGY, but not KGGDG, showed attenuation in the interaction with hECM29 (Fig. 19). Mutations in positions M68A, M212A, SER236A, E472A and D475A did not seem to affect the interaction in the plate assay. However, when yeast growth was quantified in selective QDO broth (broth assay), mutants M68A and M212A also showed diminished interaction with hECM29 (Fig. 20). The broth assay was started with same amount of viable yeast (initial OD600 = 0.8) however adjustment of the yeast quantity used in the plate assays was not performed. Thus, initial cultures of mutants M68 and M212 could have had higher amounts of viable yeast explaining the inconsistency observed between the plate and broth assays (Figs. 19 and 20).

Alpha-galactosidase (alpha-gal) activity was quantified as another method to evaluate the interaction between HtpB and hECM29 in yeast. When HtpB interacts with hECM29 the *MEL1* reporter gene is activated followed by synthesis and secretion of the enzyme alpha-galactosidase. Enzymatic activity quantification allowed detection of increased strength of the HtpB-hECM29 interaction in mutant E472A which was observed as increased production of alpha-galactosidase compared to WT-HtpB (Fig. 21). Consistently, mutant 472-475/A, which contains the mutation E472A, showed an intermediate phenotype compared to WT-HtpB and mutant E472A (Fig. 21), but this result is opposite to the

Figure 19. Substitutions in amino acids K298, N507, H473 and K474 of HtpB affect the interaction with hECM29. Images show yeast growth on plates of selective (QDO/X/A) and non-selective (DDO) media after 5 days of incubation. Mutant and WT versions of HtpB and GroEL (first column on the left) fused to the Gal4 DNA-binding domain were co-expressed with hECM29 fused to the Gal4 activation domain in *S. cerevisiae*. Yeast carrying the corresponding plasmids (e.g., hECM29 x WT-HtpB) were grown overnight in DDO broth and 10 µl of undiluted (10^{-0}) and diluted (10^{-1}, 10^{-2} and 10^{-3}) culture were spotted onto QDO/X/A plates. Additionally, the highest dilution (10^{-3}) was plated on DDO plates to monitor culture viability. The images are representative of at least three independent experiments with the same results.

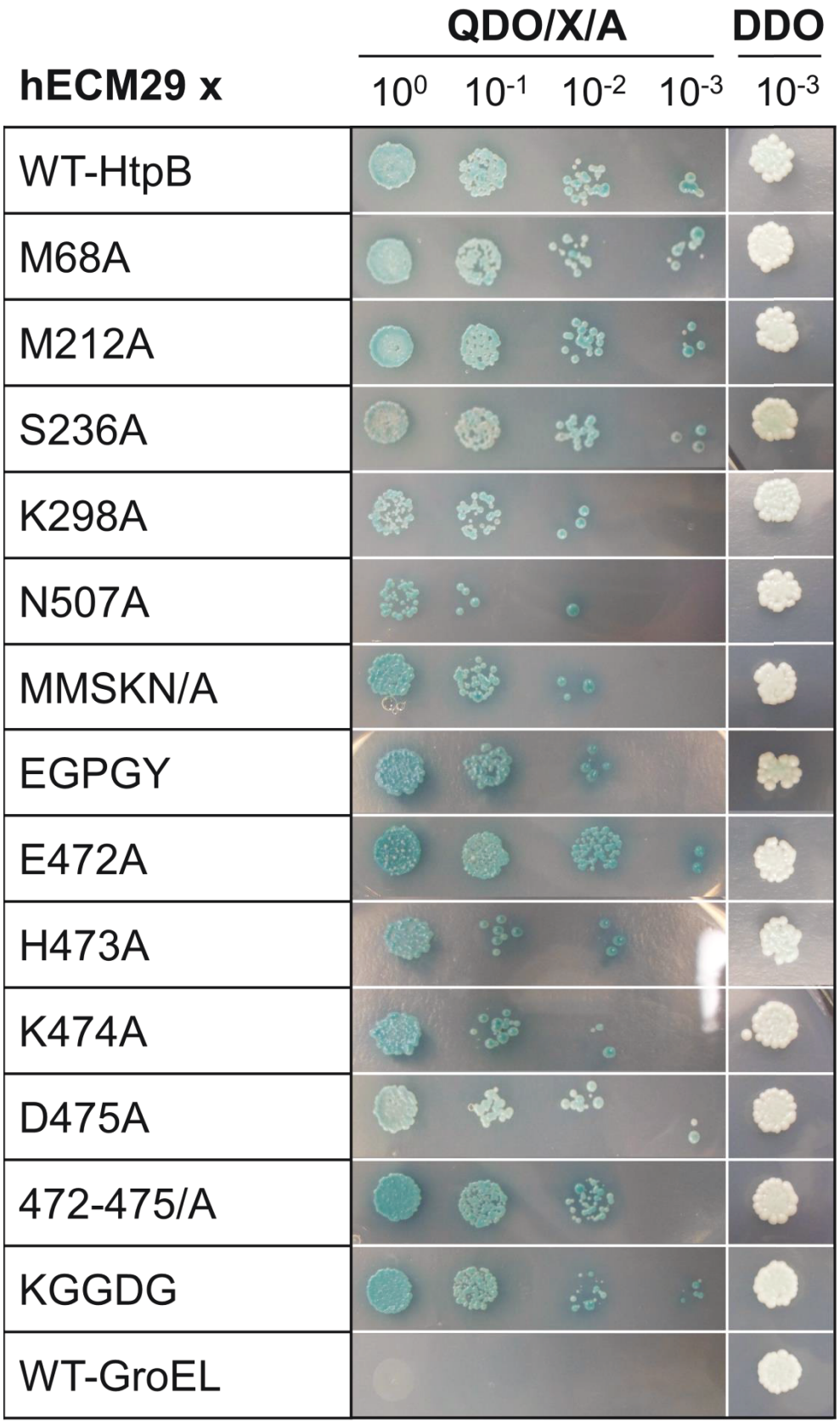

QDO/X/A
DDO
hECM29 x
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-3}
WT-HtpB
M68A
M212A
S236A
K298A
N507A
MMSKN/A
EGPGY
E472A
H473A
K474A
D475A
472-475/A
KGGDG
WT-GroEL

Figure 20. Evaluation of the mutated amino acids role in the HtpB-hECM29 interaction by quantification of growth in QDO broth. Bar graph showing growth of *S. cerevisiae* carrying pGAD:*hECM29* and WT HtpB (HtpB) or mutated versions of HtpB (shown in the X axis of the graph). Yeast were grown overnight in DDO medium and then sub-cultures (initial OD_{600}=0.8) were started in selective QDO broth. After 6 days of incubation at 30°C with shaking, cultures were diluted 1 in 10 and OD_{600} was measured. Y axis is OD_{600} and X axis gives the HtpB variants tested. Height of the bars represents the means from a single experiment run in triplicate (n=3). Error bars are 1 standard deviation (SD) above the mean. Red asterisks indicate significant less growth compared with WT HtpB (p-value < 0.05)

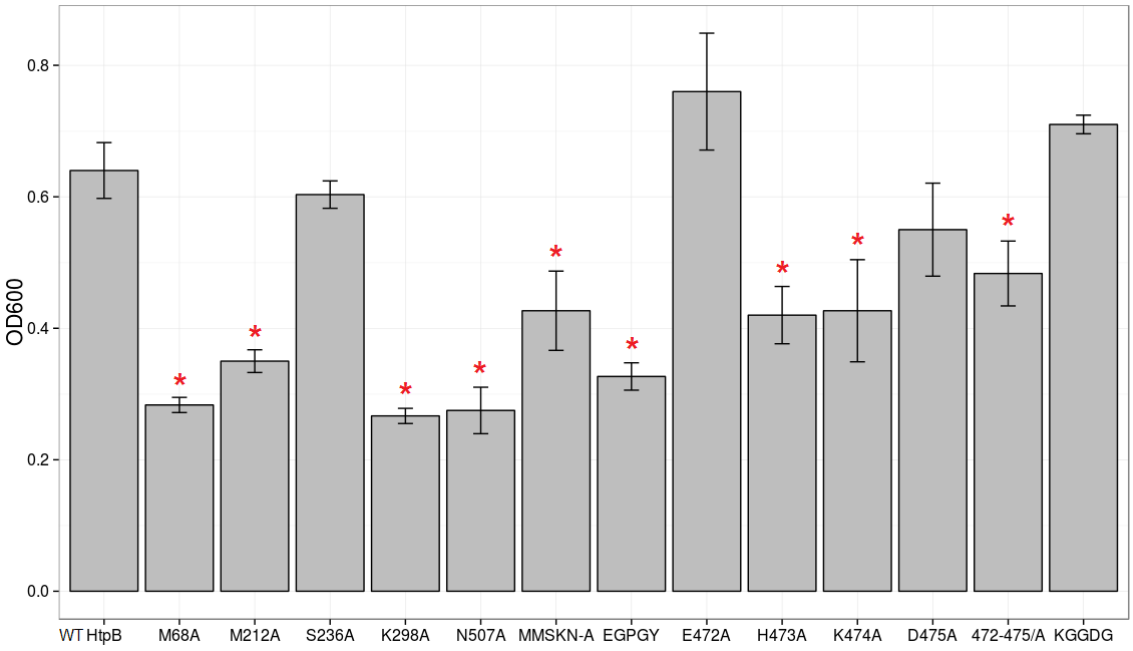
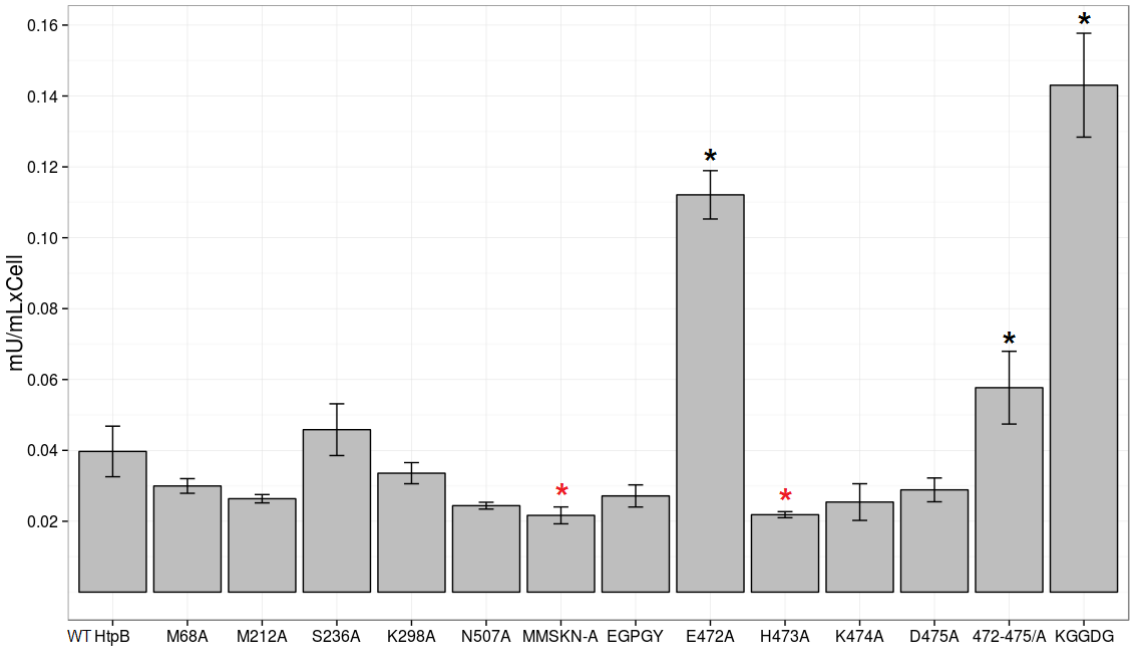

OD600
0.0
0.2
0.4
0.6
0.8
WT HtpB
M68A
M212A
S236A
K298A
N507A
MMSKN-A
EGPGY
E472A
H473A
K474A
D475A
472-475/A
KGGDG

Figure 21. Evaluation of the mutated amino acids role in the HtpB-hECM29 interaction by alpha-galactosidase activity quantification. Bar graph showing alpha-galactosidase activity of *S. cerevisiae* grown in selective medium QDO/A. Yeast carrying WT HtpB (HtpB) or mutated versions of HtpB (shown in the X axis of the graph) were grown overnight in DDO medium and then subcultures were started in selective QDO/A broth and incubated for 6 days at 30°C with shaking. Alpha-galactosidase activity was measured in the supernatant of the cultures. Y axis indicates milliunits of alpha galactosidase activity per mL per cell (mU/mL*cell) and X axis are the HtpB variants tested. Height of the bars represents the means from a single experiment run in triplicate (n=3). Error bars are 1 standard deviation (SD) above the mean. Red asterisks indicate significant lower activity and black asterisks mark significant higher activity compared with WT HtpB (p-value <0.05).


mU/mLxCell
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
WT HtpB
M68A
M212A
S236A
K298A
N507A
MMSKN-A
EGPGY
E472A
H473A
K474A
D475A
472-475/A
KGGDG

growth in plates where mutant 472-475/A showed less growth (impairment of interaction) (Fig. 19). Interestingly, mutant KGGDG also showed statistically significant higher production of the enzyme but not higher growth in the broth assay compared with WT-HtpB (Fig. 20 and 21). Confusingly, no statistically significant less production of alpha-gal was observed in mutants M68A, M212A, K298A, N507A, EGPGY or K474A. Only mutants MMSKN/A and H473A produced less enzyme indicating a negative effect the interaction as observed in both the plate and broth growth assays (Figs. 19-21).

Discrepancies observed between the plate, broth and alpha-gal results could be due to the selectivity of the media used. For example, for the plate assay, QDO/X/A highly stringency medium was used. This medium select for the activation of 4 reporter genes (*AUR1-C*, *ADE2*, *HIS3*, and *MEL1*), but the medium used for the broth and alpha-gal assays was QDO which select for the activation of only 2 of the reporter genes (*ADE2* and *HIS3*). In experiments using QDO/A broth (data no shown), yeast growth was too slow (more than 7 days) and evaporation of the cultures was observed. Thus, to accelerate growth, the antibiotic aureobasidin A was not added to the medium. Neither X-alpha-Gal because activation of the *MEL1* reporter gene induces synthesis of the enzyme alpha-gal that turns the medium blue in presence of X-alpha-Gal altering OD_{600} (broth assay) and enzymatic activity (alpha-gal assay) quantification. The results obtained in the alpha-gal assay were inconsistent with the results observed in plate and broth growth. However, this assay was better at detecting increase of the strength of the HtpB-hECM29 interaction.

Considering the aforementioned, for an amino acid mutation to be categorized as inhibitory of the HtpB-hECM29 interaction, results obtained in both the plate and broth assays had to be consistent. Thus, mutations in positions M68 and M212 were not classified as important for the HtpB-hECM29 interaction. On the other hand, only the alpha-gal results that showed an increased production of the enzyme (increased strength of the interaction) were considered as true alterations of the interaction. Mutant 472-475/A showed opposite results; less growth in the plate and broth assays and increased production of alpha-gal. Since quantification of alpha-gal activity measures only the activation of one of the reporter genes (*MEL1*), it could be possible that mutant 472-475/A produces higher amounts of the enzyme in response to the interaction but still

shows less growth when the additional reported genes are tested (plate and broth assay). This could be due to a masking effect of mutation in position E472A that induces higher production of the enzyme (Fig. 21).

The effect of mutations in GroEL was also evaluated. Based on the MSA used in the ET analysis, the homologous positions in GroEL to the HtpB residues chosen for the mutational analysis were identified and then these amino acids were substituted with the corresponding HtpB amino acids (Table 3). As shown in Figure 22, WT-GroEL does not interact with hECM29, neither when GroEL amino acids at positions 67, 211, 235, 297 and 506 were changed to HtpB amino acids (GroEL-MMSKN) or when the peptide 470-474 was substituted by the HtpB peptide (GroEL-AEHKD). Importantly, when all 10 residues in the selected amino acid positions were simultaneously mutated from the GroEL wild type sequence to the amino acids in HtpB (GroEL-Multi), GroEL was able to consistently interact with ECM29 (Fig. 22) in three independent experiments, though not as strongly as HtpB (Figs. 10 and 19).

Taken these results together, it was confirmed that at least 4 (K298A, N507A, H473A and K474A) out of the 10 predicted amino acids are required for the correct HtpB-hECM29 interaction. Weak interaction of GroEL-Multi mutant with hECM29 suggests that other amino acids, not identified in this study, are also involved.

3.3 PHYLOGENETIC ANALYSIS OF BACTERIAL CHAPERONIN 60

3.3.1 Phylogenetic analysis of chaperonins belonging to the phylum Proteobacteria

The final alignment of the Cpn60 amino acid sequences from phylum Proteobacteria contained a total of 231 sequences belonging to 180 organisms from classes Alpha, Beta, Gamma, Delta and Epsilon proteobacteria. After removal of ambiguous regions, the alignment considered 534 amino acid positions. Generally, the distribution of the chaperonins in the tree followed the 16S RNA distribution (Fig. 23). That is, sequences belonging to the same class of bacteria were clustered together forming clearly separated

Figure 22. GroEL weakly interacts with hECM29 when the selected amino acids are substituted by the corresponding HtpB residues. Images showing yeast growth in selective (QDO/X/A) and non-selective (DDO) media after 5 days of incubation. Mutant and WT versions of GroEL (first column on the left) fused to the Gal4 DNA-binding domain were co-expressed with hECM29 fused to the Gal4 activation domain in *S. cerevisiae*. Yeast carrying the corresponding plasmids (e.g., hECM29 x WT-GroEL) were grown overnight in DDO broth and 10 µl of undiluted (10^{-0}) and diluted (10^{-1}, 10^{-2} and 10^{-3}) culture were spotted onto QDO/X/A plates. Additionally, the highest dilution (10^{-3}) was plated on DDO plates to monitor culture viability. The images are representative of at least three independent experiments with the same results.

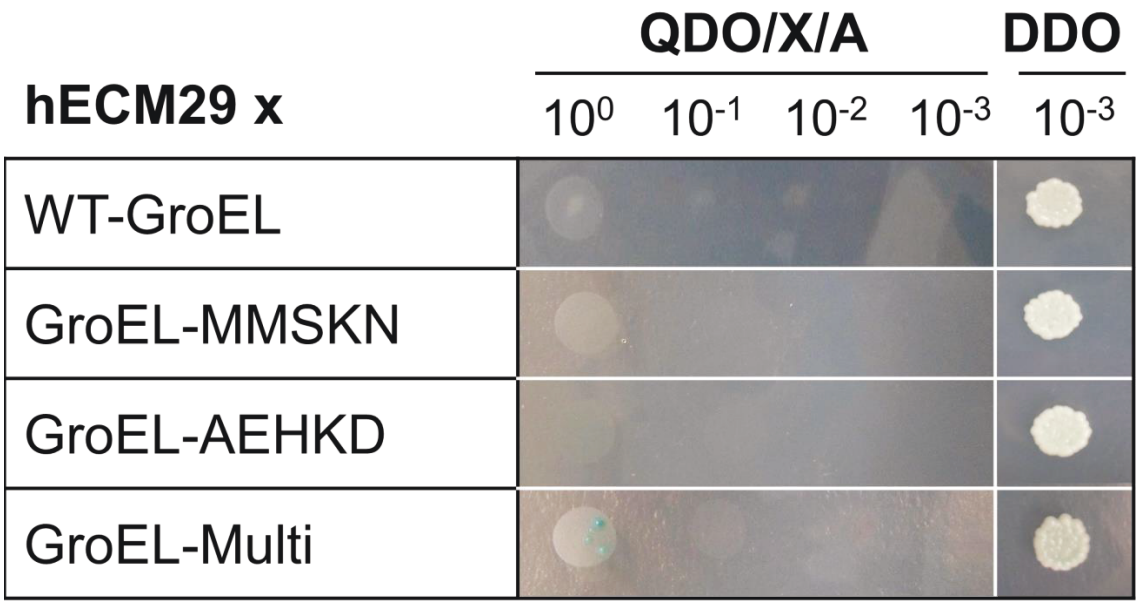
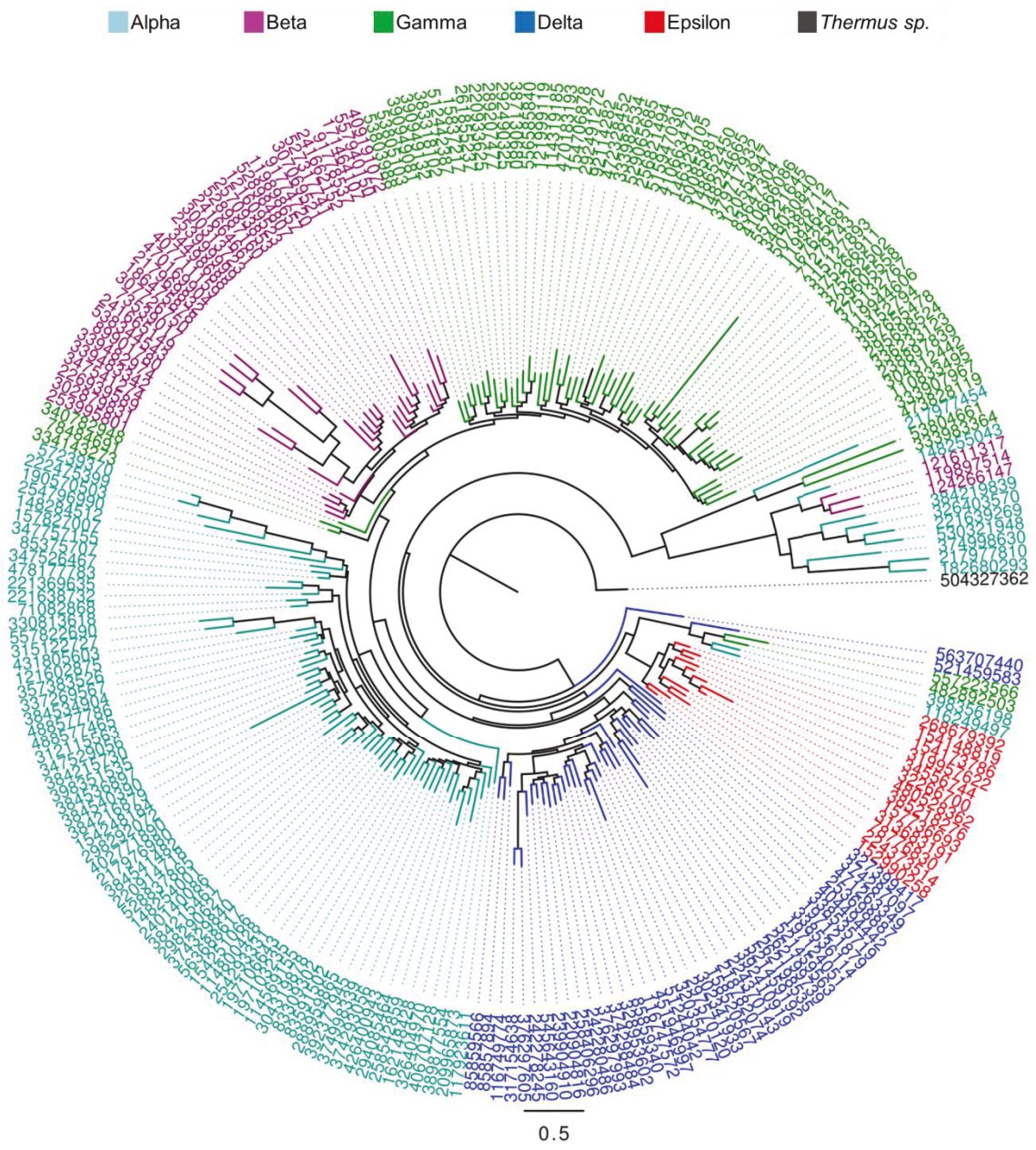

QDO/X/A
DDO
hECM29 x
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-3}
WT-GroEL
GroEL-MMSKN
GroEL-AEHKD
GroEL-Multi

Figure 23. Phylogenetic tree of chaperonin 60 from phylum Proteobacteria. Diagram showing a circular rooted phylogenetic tree of Cpn60 amino acid sequences derived from 180 species. Phylogeny was inferred using the maximum likelihood method. The Cpn60 amino acid sequence of *Thermus sp*. (GI:504327362, in black) was used as the outgroup to root the tree. Clades were colored according to proteobacteria class as per the legend. Tip labels are sequence identification numbers (GIs). Branch length is proportional to the number of substitutions per site (bottom scale).


Alpha
Beta
Gamma
Delta
Epsilon
Thermus sp.
504327362
563707440
521459583
0.5

clades. This indicates that chaperonins have suffered little horizontal gene transfer during evolution. However, some chaperonins were more similar to chaperonins corresponding to different classes rather than to their own class (Figs. 23 and 24).

Two clades were formed by chaperonins belonging to different classes where all organisms have more than one Cpn60. Both clades were well supported by bootstrap values of 76% and 100%, respectively (Fig. 24). One clade was composed by paralogs from bacteria belonging to classes Alpha, Beta and Gamma-Proteobacteria, showing high rates of evolution. It seems that gene transfer of a highly divergent Cpn60 copy from Alpha proteobacteria to the common ancestor of *Verminephrobacter eiseniae*, *Azoarcus* sp. and *Methylibium petroleiphilum* occurred. A similar transference event could explain the topology of the sub-clade formed by *Methylocella silvestris*, *Methylococcus capsulatus* and *Methylomonas methanica.* The second clade was composed by bacteria from classes Alpha, Gamma and Delta-Proteobacteria (Fig. 24). Interestingly, this clade appeared closely related to the Epsilon-Proteobacteria clade; this topology was supported by a bootstrap value of 65%. On the other hand, the single chaperonin of *Delta proteobacterium* seems to be highly divergent and did not form part of any clade (Fig. 24). Unfortunately, the chain of evolutionary events and relationships between members within each class clade could not be inferred due to weak bootstrap support.

3.3.2 Phylogenetic analysis of HtpB and other moonlighting bacterial chaperonins

The phylogenetic analysis performed to the reported moonlighting chaperonins and their paralogs (MoonlightingCpn60 dataset composed by 48 Cpn60 protein sequences belonging to 35 bacterial species) shows high divergence of some chaperonins, events of possible horizontal gene transfer and gene duplications before speciation. The distribution of the chaperonins was generally in agreement with the 16S tree with some exceptions as in *Helicobacter* and *Mycoplasma* (Fig. 25).

The Cpn60 of 2 members of different phyla, *H. pylori* (Proteobacteria) and *Mycoplasma penetrants* (Firmicutes), formed a well-supported clade suggesting horizontal gene

Figure 24. Close up of the phylogenetic tree of chaperonin 60 from phylum Proteobacteria. Diagram shows a close up of the tree in figure 24, emphasizing two clades formed by chaperonins belonging to different classes. Color code: Alpha proteobacteria (cyan), beta proteobacteria (purple), Gamma proteobacteria (green), Delta proteobacteria (blue) and Epsilon proteobacteria (red). The sequence of *Thermus sp*. (GI:504327362, in black) was used as the outgroup to root the tree. GI and name of the organism are indicated at the tip labels. Bootstrap values (%) based on 100 replications are given at branches.

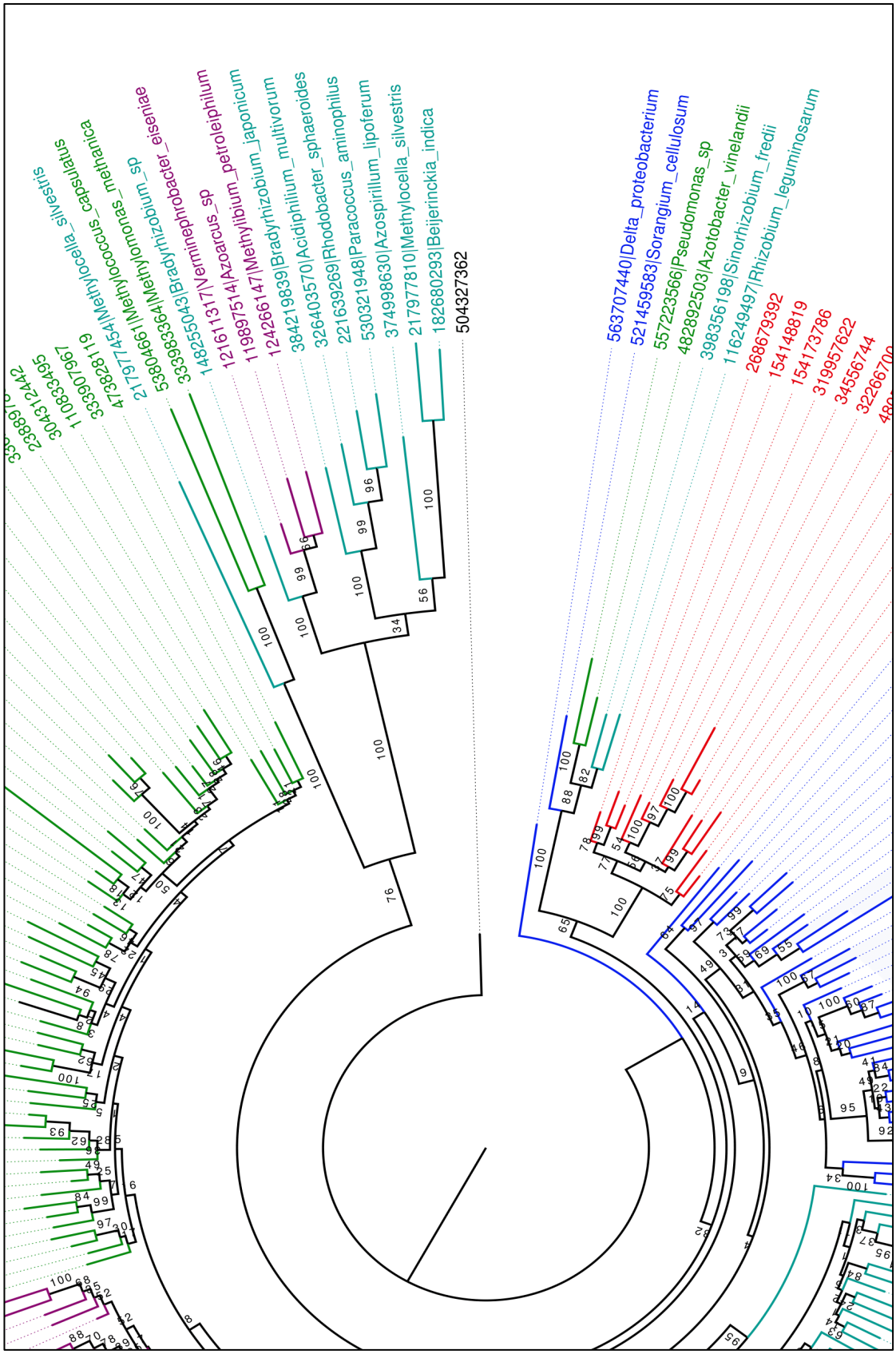
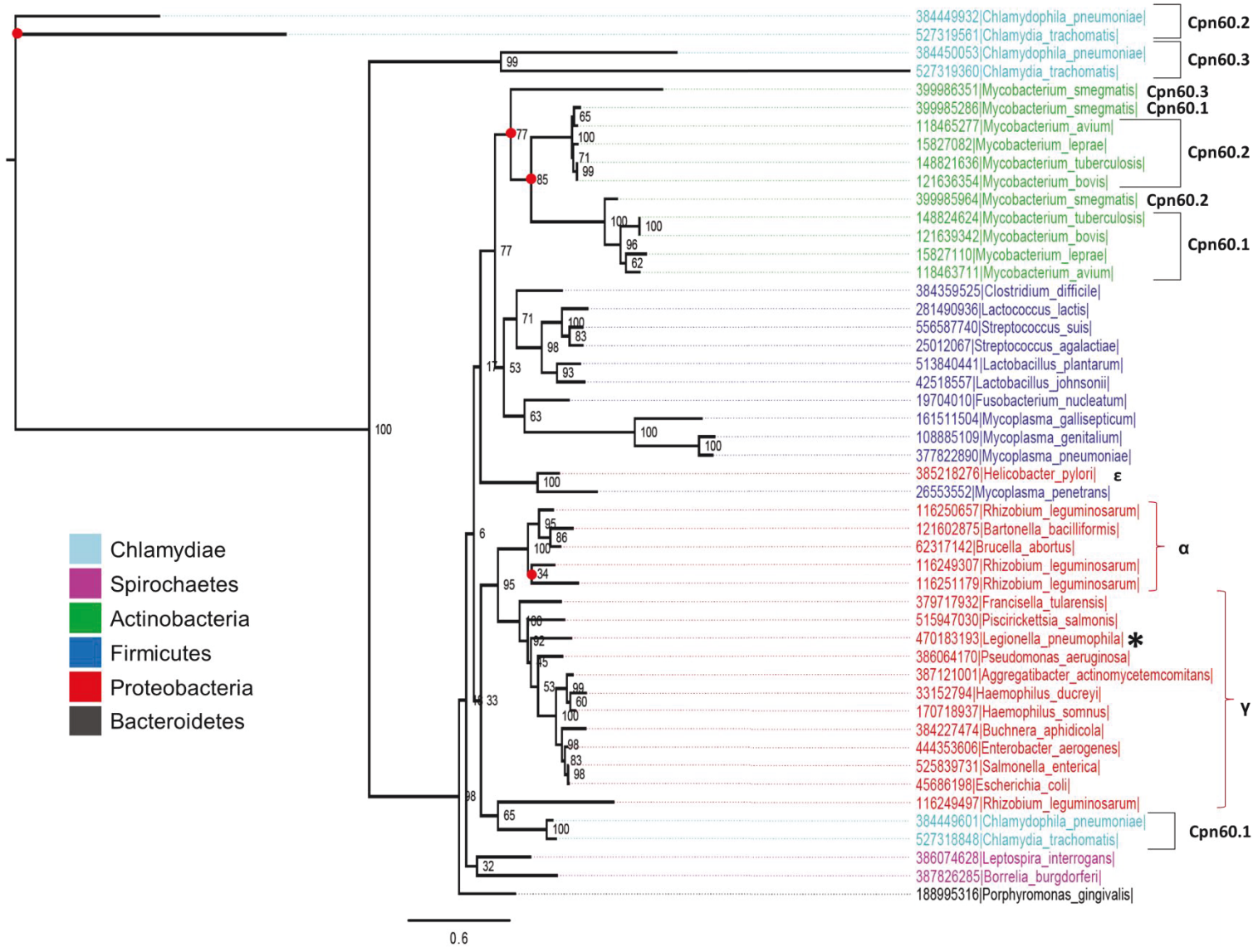

504327362
182680293|Beijerinckia_indica
217977810|Methylocella_silvestris
374998630|Azospirillum_lipoferum
530321948|Paracoccus_aminophilus
221639269|Rhodobacter_sphaeroides
326403570|Acidiphilium_multivorum
384219839|Bradyrhizobium_japonicum
124266147|Methylibium_petroleiphilium
119897514|Azoarcus_sp
121611317|Verminephrobacter_eiseniae
148255043|Bradyrhizobium_sp
533983364|Methylomonas_methanica
517977454|Methylococcus_capsulatus
473828119
533390796
110833495
304312442
238897
56370440|Delta_proteobacterium
521459583|Sorangium_cellulosum
557223566|Pseudomonas_sp
482892503|Azotobacter_vinelandii
398356198|Sinorhizobium_fredii
116249497|Rhizobium_leguminosarum
268679392
154148819
154173786
319957622
345567744
322667700

Figure 25. Phylogenetic tree of moonlighting chaperonin 60s. Diagram showing the unrooted phylogenetic tree of 48 chaperonins belonging to 35 species. Moonlighting chaperonins and their paralogs (if any) were gathered and the phylogeny was predicted using the maximum likelihood method. The tree was colored according to phylum as per the legend. Tips are labelled with Cpn60 identification numbers (GIs) and species names (right side of the diagram). Cpn60 paralogs names are shown on the right side of the tree (e.g., Cpn60.1) as indicated by black square brackets. Proteobacteria classes (α, γ and ε) are marked at the right side of the red curly brackets. The asterisk marks HtpB. Red dots at nodes indicate duplication events. Bootstrap values (%) based on 100 replications are given at nodes. Branch length is proportional to the number of substitutions per site (bottom scale).


384449932|Chlamydophila_pneumoniae|
527319561|Chlamydia_trachomatis|
384450053|Chlamydophila_pneumoniae|
527319360|Chlamydia_trachomatis|
399986351|Mycobacterium_smegmatis|
399985286|Mycobacterium_smegmatis|
118465277|Mycobacterium_avium|
15827082|Mycobacterium_leprae|
148821636|Mycobacterium_tuberculosis|
121636354|Mycobacterium_bovis|
399985964|Mycobacterium_smegmatis|
148824624|Mycobacterium_tuberculosis|
121639342|Mycobacterium_bovis|
15827110|Mycobacterium_leprae|
118463711|Mycobacterium_avium|
384359525|Clostridium_difficile|
281490936|Lactococcus_lactis|
556587740|Streptococcus_suis|
25012067|Streptococcus_agalactiae|
513840441|Lactobacillus_plantarum|
42518557|Lactobacillus_johnsonii|
19704010|Fusobacterium_nucleatum|
161511504|Mycoplasma_gallisepticum|
108885109|Mycoplasma_genitalium|
377822890|Mycoplasma_pneumoniae|
385218276|Helicobacter_pylori|
26553552|Mycoplasma_penetrans|
116250657|Rhizobium_leguminosarum|
121602875|Bartonella_bacilliformis|
62317142|Brucella_abortus|
116249307|Rhizobium_leguminosarum|
116251179|Rhizobium_leguminosarum|
379717932|Francisella_tularensis|
515947030|Piscirickettsia_salmonis|
470183193|Legionella_pneumophila|
386064170|Pseudomonas_aeruginosa|
387121001|Aggregatibacter_actinomycetemcomitans|
33152794|Haemophilus_ducreyi|
170718937|Haemophilus_somnus|
384227474|Buchnera_aphidicola|
444353606|Enterobacter_aerogenes|
525839731|Salmonella_enterica|
45686198|Escherichia_coli|
116249497|Rhizobium_leguminosarum|
384449601|Chlamydophila_pneumoniae|
527318848|Chlamydia_trachomatis|
386074628|Leptospira_interrogans|
387826285|Borrelia_burgdorferi|
188995316|Porphyromonas_gingivalis|
Cpn60.2
Cpn60.3
Cpn60.3
Cpn60.1
Cpn60.2
Cpn60.2
Cpn60.1
ε
α
γ
Cpn60.1
Chlamydiae
Spirochaetes
Actinobacteria
Firmicutes
Proteobacteria
Bacteroidetes
0.6

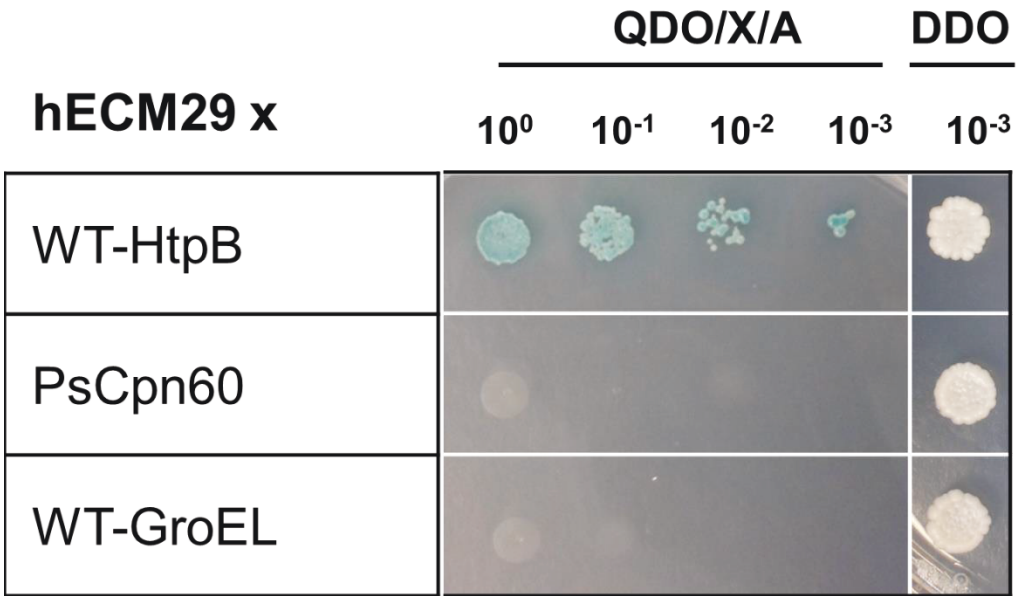
transfer. Similarly, *Chlamydia* chaperonins showed an interesting distribution: Cpn60.1 was closer to the Protebacteria clade than Cpn60.2 or Cpn60.3, and formed a cluster with one of the *Rhizobium leguminosarum* Cpn60s. It seems that gene duplication happened before speciation because the chaperonins of both species studied followed the same distribution (Fig. 25). Paralogy was also observed in the *Mycobacterium* clade (bootstrap = 77%); from the 5 species, forming the clade, 4 have 2 copies of Cpn60 and *Mycobacterium smegmatis* has three copies. The tree topology suggests that gene duplication perhaps happened in the common ancestor of the *Mycobacterium* species. Cpn60.1 and Cpn60.2 formed two separate groups and the third Cpn60 of *M. smegmatis* seems to greatly diverge from the other two copies. *M. smegmatis* Cpn60.1 and Cpn60.2 appeared in the opposite group of paralogs but this could be due to arbitrary annotation of the copies of the gene (Fig. 25). Among the Proteobacteria members, only *R. leguminosarum* has multiple copies of the Cpn60 gene. Three out of 4 copies seem to share the same Cpn60 common ancestor with the rest of the members of the clade showing low rates of evolution. The Proteobacteria cluster was supported by a bootstrap value of 94% and placed HtpB close to other intracellular pathogen chaperonins (*Francisella tularensis* and *Piscirickettsia salmonis*). Lastly, the Spirochaetes and Firmicutes clade was formed only by organisms with one Cpn60 and the chaperonin of the only member of phylum Bacteroidetes included in this study (*Porphyromonas gingivalis*) was not evolutionarily closer to any of the other clades (Fig. 25). Taken together, these results suggest that chaperonin 60 has undergone multiple events of gene duplication and in some cases it is followed by sequence divergence. However, HtpB seems to have evolved exclusively in the context of *L. pneumophila* likely to support the intracellular lifestyle of this bacterium.

3.4. THE CPN60 OF *PISCIRICKETTSIA SALMONIS* DID NOT INTERACT WITH hECM29 IN Y2H ASSAY

The phylogenetic tree topology showed that HtpB is in close proximity with the moonlighting Cpn60s of two intracellular pathogens, *Francisella tularensis* and *Piscirickettsia salmonis*. AT the time, a fellow PhD student in our lab was evaluating the

subcellular localization of the Cpn60 of *P. salmonis* (PsCpn60) and observed that similarly to HtpB, PsCpn60 is found in association with the outer membrane and in outer membrane vesicles (OMVs) of this fish pathogen. Since I had access to genomic DNA of *P. salmonis* I wanted to check whether, according to their phylogenetic relationships, there were some functional similarities between PsCpn60 and HtpB. Therefore, PsCpn60 gene was cloned into pGBKT7 and its ability to interact with hECM29 was tested by Y2H. As shown in Figure 26, PsCpn60 did not interact with hECM29. This result indicates that although the Cpn60 of *P. salmonis* shows cellular location similarities with HtpB, they do not necessarily have common moonlighting functions.

Figure 26. The Cpn60 of *Piscirickettsia salmonis* does not interact with hECM29. Photos of agar plates showing yeast growth in QDO/X/A selective media after 5 days of incubation. WT-HtpB, *P. salmonis* Cpn60 (PsCpn60) and WT-GroEL (first column on the left) fused to the Gal4 DNA-binding domain were co-expressed with hECM29 fused to the Gal4 activation domain in *S. cerevisiae*. Yeast carrying the corresponding plasmids (e.g., hECM29 x WT-HtpB) were grown overnight in DDO broth and 10 µl of undiluted (10^{-0}) and diluted (10^{-1}, 10^{-2} and 10^{-3}) culture were spotted onto QDO/X/A plates. Additionally, the highest dilution (10^{-3}) was plated on DDO medium to ensure culture viability. The images are representative of at least three independent experiments with the same results.


QDO/X/A
DDO
hECM29 x
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-3}
WT-HtpB
PsCpn60
WT-GroEL

CHAPTER 4 DISCUSSION

Bacterial chaperonins are well conserved proteins that function to assist the folding of other proteins. It is emerging that the chaperonin 60 (Cpn60) family has evolved a puzzling variety of additional protein-folding independent biological functions. HtpB, the *Legionella peumophila* chaperonin, has virulence-related moonlighting functions [90], which are not shared with GroEL, the Cpn60 of *E. coli*. Given that chaperonins are highly conserved proteins, I hypothesised that substitutions in only a few amino acid positions in HtpB could explain its functional promiscuity. This study was initiated to address the question of which amino acids could be involved in HtpB moonlighting functions, and to identify an interaction partner that could be used as a functional experimental model to explore the effects of HtpB in the host cell.

Herein, I found a novel interacting partner of HtpB, hECM29, which I confirmed does not interact with GroEL, and further identified some of the HtpB amino acids involved in the interaction. In addition, a phylogenetic study of the bacterial Cpn60 family and some of its moonlighting members revealed multiple events of gene duplication and in some cases high divergence. HtpB seems to have co-evolved only with *L. pneumophila* as its phylogeny correlates well with that based on the 16S rRNA gene sequence of this intracellular bacterium. Suggesting that HtpB has acquired moonlighting functions that support the intracellular lifestyle of *L. pneumophila*.

4.1 HTPB INTERACTS WITH hECM29

Among the moonlighting functions of HtpB, its ability to reach the host cell cytoplasm and modify mitochondria trafficking and microfilament organization is one of the most interesting [97]. Although, the interaction of HtpB with the eukaryotic proteins S-adenosyl methionine decarboxylase (SAMDC) and mitochondrial Hsp10 has been reported [99,101], these do not relate with the alterations in cell signaling observed upon HtpB uptake [97]. SAMDC is an enzyme required for the biosynthesis of polyamines. The more recently described interaction of HtpB with eukaryotic Hsp10 [101] is not surprising given that this protein is homologous to HtpA (HtpB's cognate co-chaperonin 10) which naturally

interacts with HtpB. Here I report the identification of two new interaction partners for HtpB: TOX-4 and the human homolog of ECM29 (hECM29). However, neither human SAMDC nor Hsp10 were identified in the yeast-two-hybrid screening performed in this study. This might be due to the fact that different cDNA libraries were screened; here I used a human library whereas SAMDC was initially identified in a screening performed using a yeast library [99] and Hsp10 was found in a HeLa cell cDNA library [101]. Perhaps, a smaller quantity of cDNAs corresponding to SAMDC and Hsp10 were present in the human library used in my study. Another possibility is that the interaction of HtpB with these proteins is weak and was not detected in my screening since only strong interactions (yeast that grew in the high stringency selection medium) were analyzed.

One of the HtpB partners identified in this study was TOX4 (TOX family member 4 or KIAA0737) which is a member of the high motility group box (HMG-box) protein family [122] initially cloned from human brain [123]. Although its specific function is not yet completely understood, TOX4 seems to be involved in DNA repair [124], regulation of chromatin structure and cell cycle progression [125,126]. HMG-box proteins possess a motif that binds DNA but also appears to be involved in protein-protein interactions [127]. In addition, immunofluorescence experiments show that stably expressed HtpB localizes to the nucleus of CHO cells [97]. Therefore, HtpB could interact with the HMG-box of TOX4 and induce alterations in the expression of genes that are critical for certain physiological processes. This could be a conserved characteristic of chaperonins given that GroEL also interacted with TOX4. However, confirmation of the interaction between HtpB and TOX4 by other methods was not performed; thus, further experiments are needed to evaluate whether this interaction occurs *in vivo* and possible implications in *L. pneumophila* pathogenesis, perhaps by co-immunoprecipitation and co-localization experiments.

The second and more attractive HtpB partner found in this study was the human homolog of ECM29 (KIAA0368). The interaction of HtpB with hECM29 was confirmed by co-immunoprecipitation in yeast and some of the HtpB amino acids involved in the interaction were identified by Y2H experiments. ECM29 (extracellular mutant 29) is a 200-kDa protein initially identified in a transposon mutagenesis screening performed in *S.*

cerevisiae where disruption of the ECM29 gene affected cell wall integrity [128]. Bioinformatics analysis and structural predictions performed on the ECM29 protein sequence showed that it is composed almost entirely of HEAT-like repeats [129], causing it to acquire an elongated and curved shape as observed by electron-microscopy [130]. HEAT repeats in proteins function as flexible domains that can wrap around target substrates helping them to assemble [131]. In fact, studies on the proteins that interact with the proteasome identified ECM29 as one of the major components that assist 26 S proteasome assembly in the cytosol [130,132,133].

The 26 S proteasome is a multi-protein complex formed by the regulatory particle (RP, 19S) that recognises and unfolds ubiquitinated protein substrates and the core particle (CP, 20 S) that degrades substrates into short peptides to then be reused by the cell [134]. ECM29 has been shown to enhance the stability of the interaction between CP and RP [130,132]. The 26 S proteasome is part of the machinery that participates in endoplasmic reticulum (ER) associated protein degradation (ERAD) [135] and also has a role in proteolysis in the centrosome [136]. Interestingly, as shown by immunofluorescence experiments in HeLa-cells, hECM29-proteosome complexes localize to the centrosome, ER and endosomes indicating a role of hECM29 in coupling the proteasome to areas of high protein degradation [137]. This was confirmed by genome-wide two-hybrid screens and mass spectrometry experiments where the interactome of hECM29 was assessed showing that hECM29 subcellular location correlates with the localization of the hECM29-interacting proteins found in the screening. Surprisingly, hECM29 not only binds to the 26 S proteasome but also binds to members of the ERAD pathway, Cep152 (a centrosomal protein), endocytic components and myosins and kinesins [138]. Co-precipitation experiments using full length and truncated versions of hECM29 showed that the N-terminal half of hECM29 binds endocytic components, the central region of hECM29 binds the 26 S proteasome, whereas the C-terminal half of the protein binds to myosins and kinesins [138]. It has been reported that 26 S proteasomes bind to actin filaments and myosin [139,140] probably through an interaction with hECM29; more importantly, inhibition of proteasome activity results in the alteration of actin filaments and microtubule organization [141,142]. The cDNA fragment identified in my Y2H screening corresponds to the C-terminus of hECM29. Therefore, interaction of HtpB with hECM29

could alter its interaction with molecular motors leading to modifications in microfilament organization in the host cell. Thus, HtpB-hECM29 interaction could help to explain why HtpB-, but not GroEL-, coated beads induced F-actin rearrangements in CHO cells [97].

On the other hand, it has been shown that active proteasomes and factors associated with the ERAD pathway are required for optimal *L. pneumophila* replication in *Drosophila* Kc167 cells [143]. Thus, HtpB could be an effector used by this pathogen to hijack the ubiquitination machinery of the host cell through interaction with hECM29. For example, the HtpB-hECM29 interaction could be used by *L. pneumophila* to regulate proteolysis of its own effector proteins as shown by Kubori and co-workers who reported that LubX (a *Legionella* effector that functions as an E3 ubiquitin ligase) targets the bacterial effector protein SidH for degradation by the host cell proteasome [144]. Other *Legionella* type IV secreted effectors, such as AnkB [145] and SidC [146], induce association of ubiquitinated host cell proteins with the LCV membrane which subsequently serve as a nutrient source for *Legionella*, as they are targeted to the proteasome for degradation. Perhaps, binding of HtpB to hECM29-proteasomes promotes the localization of the 26 S proteasome to the LCV promoting protein degradation.

A recent study links hECM29-proteasomes to toll-like receptor 3 (TLR3) signaling and autophagy [147]. TLR3 is a transmembrane receptor found in endosomal compartments that recognizes dsRNA, (viral or from dying cells) and triggers the activation of several intracellular signaling pathways that culminate in the induction of cytokines and other proinflammatory mediators generating anti-viral responses, as well as apoptosis [148]. hECM29 depleted human embryonic kidney (HEK) 293 cells showed increased abundance of TLR3, TLR3 downstream effectors and inhibition of autophagy suggesting that hECM29-proteasomes participate in the degradation of TLR3 and its effectors by promoting autophagy [147]. Although the most characterized function of TLR3 is sensing of viral infection, recent studies have provided new insights into the role of TLR3 in bacterial infection. For example, it has been shown that *Chlamydia muridarum* replicates more efficiently in TLR3 deficient epithelial cells and that TLR3 is required for INF-β production upon *C. muridarim* infection [149,150]. Similarly, artificial activation of TLR3

protects mice from *Francisella tularensis* infection and decreases intracellular replication [151]. Another study shows that TLR3 recognizes dsRNA of intestinal commensal, but not pathogenic, bacteria triggering production of INF-β [152]. Perhaps HtpB interaction with hECM29 enhances hECM29-proteasome activity inducing attenuation of TLR3 signaling by degradation of TLR3 and its downstream effectors. This could be a mechanism that *L. pneumophila* exploits to escape sensing by the cell, allowing the establishment of a replication niche. In addition, increased protein degradation by the hECM29-proteasomes would be also beneficial as it would provide amino acids readily available to support *Legionella* intracellular growth.

Given that TLR3 seems to be involved in the recognition of other intracellular pathogens, the interaction of hECM29 with Cpn60s of intracellular pathogens should be assessed to determine whether this mechanism is conserved among closely related pathogens. Here, I showed the Cpn60 of *Piscirickettsia salmonis* (PsCpn60), an intracellular phylogenetically related fish pathogen, does not interact with hECM29. However, I evaluated the interaction of PsCpn60 with the human homolog of ECM29; thus, it could be possible that PsCpn60 and the fish homolog of ECM29 interact *in vivo*.

Finally, HtpB could be one of the *L. pneumophila* effectors that regulate the establishment of the legionella-containing vacuole (LCV). Soon after phagocytosis, *L. pneumophila* escapes the endocytic pathway [34,153], and induces redecoration of the LCV with ER-derived vesicles [41,154], polyubiquitinated host proteins [155], mitochondria [40,97] and ribosomes [42]. The establishment of a niche that permits optimal intracellular replication of *L. pneumophila* is mostly mediated by Dot/Icm effectors [36]. However, interaction of HtpB with hECM29 could have a role in the alterations in vesicular trafficking observed upon *Legionella* infection. As mentioned before, hECM29 interacts with endosomal constituents and molecular motors. Thus, HtpB could be intercepting secretory vesicles that transit between the ER and the Golgi when they are in close proximity with the LCV, blocking normal trafficking by competing for the binding site in hECM29 with molecular motors.

In summary, the interaction between HtpB and hECM29 could have a variety of roles as hECM29 interacts with many proteins in the host cell. HtpB-hECM29 interaction could

be involved in microfilament re-organization, hijacking of the ubiquitin-proteasome pathway altering protein degradation and TLR3 signaling, and stealing of secretory vesicles from the ER. This interaction could represent a previously undescribed strategy used by *L. pneumophila* to successfully infect mammalian cells. However, confirmation of the HtpB-hECM29 interaction in mammalian cells, *in vivo*, was not assessed in this study and a full characterization of this interaction *in vivo* remains to be completed.

Further experiments will be performed to confirm by co-immunoprecipitation and co-immunolocalization the HtpB-hECM29 interaction in CHO cells stably expressing HtpB (manuscript in preparation). Likewise, the biological significance of the HtpB-hECM29 interaction could be assessed by evaluation of *L. pneumophila* infection progression in hECM29-deficient cells. However, to confirm that any phenotype observed in hECM29 -deficient cells infected with *L. pneumophila* is related to HtpB, an *htpB*Δ strain should be generated. However, given that HtpB is an essential protein, deleting its gene is not possible [103]. An alternative approach would be to generate a tagged mutant version of HtpB that does not interact with hECM29 anymore but keeps its essential folding function intact. Following insertion of the mutant HtpB into the *L. pneumophila* chromosome, the WT-HtpB gene could be disrupted end evaluation of the infectivity of this HtpB mutant compared with the parent strain could be evaluated.

4.2 IDENTIFICATION OF HTPB AMINO ACIDS INVOLVED IN THE INTERACTION WITH hECM29

Proteins that have at least two biochemical functions that reside in one polypeptide chain are known as “moonlighting” proteins, but by which mechanism(s) can a protein perform multiple functions? Studies have shown that some moonlighting proteins undergo conformational changes that expose a different set of amino acids that can then interact with additional substrates or proteins [156,157]. Another possibility is that moonlighting proteins have multiple active sites that are always exposed [158], and so are able to interact with different substrates depending on the cellular conditions or on which intracellular compartment this proteins are located in. In this regard, HtpB moonlighting activity

would probably fit better into the second mechanism. Given that HtpB has no paralogs and it is an essential protein in *L. pnemophila* [103], the requirement for functional native proteins in the bacterium cytoplasm keeps HtpB under pressure to perform its folding function. However, when HtpB reaches other cellular compartments such as the outer membrane of *L. pneumophila* [91], or is translocated into host cell cytoplasm [99], other functional site(s) may become exposed to different substrates, thus inducing one or more moonlighting function(s). Some evidence for the involvement of specific amino acids in the moonlighting activities of other Cpn60s has been reported, and in some cases very few substitutions give a Cpn60 a whole new function. For example, the Cpn60 of *Mycobacterium leprae* displays proteolytic activity and only three amino acids (T375, K409, and S502) form the catalytic group [88] and the toxic effect of a symbiotic strain of *Enterobacter aerogenes*' Cpn60 is mediated by 4 residues (V100, N101, D338 and A471) [87]. Therefore, high conservation in the chaperonin family and the fact that GroEL does not exert any of the reported HtpB moonlighting functions, led me to think that HtpB multi-functionality is due to substitutions in a few specific amino acids which are probably different in GroEL.

As a first step to identify amino acids that are involved in HtpB moonlighting, I performed an evolutionary trace (ET) analysis. ET uses phylogenetic information to predict functional amino acids in proteins based on the hypothesis that substitutions at key amino acid positions should result in functional evolutionary divergences. As expected, the amino acids involved in the multimerization of HtpB or amino acids required for its protein-folding function had low ET ranks indicating evolutionary importance and high conservation among the chaperonins analysed. Similar results were found in a conservation study performed on 43 bacterial Cpn60s where amino acids at the ATP binding site, hydrophobic residues that contribute to substrate binding and intra-ring interacting amino acids showed high conservation [159]. In the same study, less conservation was found among the amino acids that participate in the interaction between rings; this is comparable to my results where some of these amino acids showed high ET ranks (e.g., D12, L15 and N468). Lack of conservation in the amino acids that participate in the inter-ring interactions could indicate that formation of the 14-mer chamber is not essential for folding and single ring structures can have folding activity. There is

convincing evidence that mitochondrial Hsp60 and its co-chaperonin Hsp10 can function as a one ring chaperonin both *in vitro* and *in vivo* [160]. However, it has been demonstrated that only the double ring conformation of GroEL is able to carry out folding. In fact, a chimeric GroEL that can only form single heptameric ring structures cannot perform as a chaperonin *in vivo* [161]. This is because ATP binding to the second GroEL ring induces conformational changes that are required for the release of the GroES lid and the folded substrate from the first ring [162]. Therefore, in single ring conformations, GroES would remain attached to GroEL and the substrate protein would stay sequestered in the folding chamber underneath GroES [63,161].

The ET analysis did not show any evolutionarily important cluster of amino acids other than those related to folding which supports the hypothesis that moonlighting functions of HtpB are due to substitutions in a few rather scattered amino acids that are not involved in folding. Knowing that GroEL does not have most of the moonlighting functions attributed to HtpB [90], it is reasonable to think that differences in the amino acid sequences of HtpB and GroEL could account for the multi-functionality of HtpB. In fact, 137 substitutions were found between HtpB and GroEL and from these, 41 less likely substitutions were analysed. Finally, of the 10 amino acids (out of the 41) selected for mutagenesis based on low ET rank and negative Blosum62 scores, single mutations at amino acids K298, N507, H473 and K474 induced partial impairment of the HtpB-hECM29 interaction. These amino acids are located at the surface of HtpB that faces the outside of the rings and are probably not essential to the protein-folding functions of the 14-mer complex. These results support the idea that the moonlighting activities of HtpB may have evolved from substitutions in amino acids scattered in “unused” solvent-exposed superficial areas. Because these “unused” amino acids are not under evolutionary pressure, they accumulate mutations that can lead to new functions [163].

Conversely, a substitution of glutamic acid 472 (E472A) for alanine increased the strength of the HtpB-hECM29 interaction as demonstrated by elevated production of alpha galactosidase. Glutamic acid is a negatively charged, polar amino acid frequently involved in binding sites whereas alanine is a non-polar amino acid widely used in mutational analysis because of its non-bulky, chemically inert nature. Hence, the

increased affinity of HtpB E472A for hECM29 is unexpected. Interestingly, E472 is a homolog position to A471 of the Cpn60 of *E. aerogenes*, which is also a moonlighting protein. The Cpn60 of *E. aerogenes* is a chaperone and also an insect paralysing toxin; residues V100, N101, D338 and A471 are required for toxicity [87]. Substitution of alanine by glutamic acid has a BLOSUM62 score of 5, which means that this substitution is commonly found in homologous proteins even if these two amino acids do not share chemical characteristics. Therefore, HtpB could initially have had an alanine in position 472, although no explanations other than random mutations could explain why a substitution less favorable for the HtpB-hECM29 interaction took place. One can speculate that at some point in evolution the ancestor of *L. pneumophila* did not have to make use of the HtpB-hECM29 interaction; thus, the amino acids involved in the interaction were not under positive pressure and thus changed. This is assuming that the interaction of Cpn60 with hECM29 is a conserved feature, and that it is advantageous for a given organism.

On the other hand, human Hsp60 activity as an innate immune response modulator is well characterized [164]; recently Habich et al. (2004) showed that the peptide in positions 481-500 of the human Hsp60 (457-476 in HtpB) is responsable for binding to macrophages [165]. This peptide is located in the equatorial domain of human Hsp60 similar to amino acids N507, H473 and K474 in HtpB. Likewise, Cpn60 of the aphid symbiont bacterium *Buchnera* spp., also a moonlighting chaperonin, interacts with viral particles through its equatorial domain [166]. This evidence could indicate that amino acids in the equatorial domain of the Cpn60 protein are more prone to undergo substitutions that result in gain of function(s).

The effect of mutations in GroEL amino acids located at homologous positions to those selected for mutagenesis in HtpB was also evaluated. WT GroEL does not interact with hECM29 but when all the amino acids in the 10 selected positions were substituted in GroEL with the corresponding HtpB residues, the mutant GroEL weakly interacted with hECM29. This confirms that at least 4 amino acids (K298, N507, H473 and K474) out of the 10 tested have a role in the HtpB-hECM29 interaction. However, the effect of substitutions only in the residues that directly participate in the HtpB-hECM29 interaction

was not assessed. It would be interesting to see if replacing only these 4 amino acids increases or diminishes the strength of the interaction between mutant GroEL and hECM29. It is possible that substitutions in 10 amino acids at once could have induced too many conformational changes in GroEL that affected binding to hECM29. A weak interaction between mutant GroEL and hECM29 also indicates that there are other HtpB amino acids forming the molecular surface that interact with hECM29. I selected only the best candidates for mutational analysis, but some other amino acids also had low ET ranks and negative BLOSUM62 scores. For example, G484 (82.24) and M503 (65.83) are very close to N507 and have low ET ranks. Furthermore, there are 4 amino acids surrounding K298 that could also be part of the hECM29 interacting surface (Fig. 15C). In sum, these results support the hypothesis that HtpB's moonlighting functions are due to substitutions in a few scattered amino acids rather than to the existence of a cluster.

In conclusion, the ET trace analysis performed in this study was able to predict residues that participate in the interaction of HtpB with hECM29, namely K298, N507, H473 and K474. Additionally, I was able to convert GroEL into an interacting partner of hECM29 by substituting selected HtpB residues into the GroEL sequence. On the other hand, the interaction between HtpB and hECM29 needs to be further characterized and the role of the amino acids identified should be evaluated *in vivo*. The aforementioned results suggest that multi functionality of chaperonins is more likely due to substitutions in a few residues preferentially located in "unused" areas of the molecule, perhaps the equatorial domain. However, additional experimentation analysing other multifunctional chaperonins should be performed to verify this hypothesis. In addition, ET analysis provides a framework for further experimentation to evaluate the role of the predicted amino acids in other HtpB moonlighting functions. For example, outer membrane localization of tagged mutant versions of HtpB could be assessed by immunogold labelling. Approaches similar to those used in this study could also be performed to find amino acids involved in the interaction of HtpB with SAMDC [99] and mitochondrial Hsp10[101].

4.3 EVOLUTIONARY RELATIONSHIPS BETWEEN MEMBERS OF THE BACTERIAL CHAPERONIN 60 FAMILY

This research initially focused on finding the molecular basis of HtpB moonlighting and the functional analysis performed provided some insights into what makes HtpB a multifunctional protein. The next question that I addressed was if there is any phylogenetic relationship between moonlighting Cpn60s, since many other Cpn60s are also multifunctional proteins. The initial objective was to evaluate the entire bacterial Cpn60 family, but due to the huge number of available sequences the data set had to be reduced to only the Proteobacteria Cpn60s. A separate phylogeny was inferred for moonlighting Cpn60s and their paralogs. As a whole, the results showed high conservation of Cpn60 within the phylum Proteobacteria; events of gene duplication followed by rapid evolution were observed among some organisms that harbor moonlighting Cpn60s. However, no clear relationships between moonlighting Cpn60s were observed, indicating that acquisition of additional functions is more likely an event induced by recent evolutionary constraints.

4.3.1 Phylogeny of Cpn60 (phylum Proteobacteria)

The high conservation of the Cpn60 family has made this protein a viable alternative to the use of 16S RNA in establishing phylogeny. In fact, a previously reported eubacterial phylogeny inferred from Cpn60 sequences is very similar to that obtained with 16S RNA [167]. This is in agreement with the phylogeny results presented here where chaperonins from Proteobacterial classes Alpha, Beta, Gamma and Epsilon were clustered into distinct subdivisions, with the exception of a few Cpn60s belonging to organisms that have multiple copies. Because all Cpn60 copies (paralogs) present in each organism were included, the interpretation of the phylogeny is more complicated. Nevertheless, it could be inferred that an early duplication event occurred in the Alpha proteobacteria and one of the copies greatly diverged. It seems that gene transfer of this divergent Cpn60 copy from the Alpha proteobacteria class to some members of the classes Beta and Gamma may have occurred. Similarly, gene transfer from Epsilon proteobacteria to *Sorangium cellulosum*, *Pseudomonas* sp. and Rhizobiales was observed. Interestingly, all bacteria that formed part of the multi-taxa clades have at least two copies of the Cpn60 gene, and

some of them have up to 6 copies. But, why do some bacteria keep so many copies of a particular gene?[72] And why is the most divergent copy distributed among different lineages?

Species of the order Rhizobiales, which are symbiotic nitrogen-fixing bacteria, often have multiple copies of the Cpn60 gene; for example, *B. japonicum* harbors 6 copies. The requirement of Cpn60 for efficient nitrogen fixation could explain why this legume-root nodulating bacterium keeps many copies of *cpn60* [168]. Surprisingly, at least 5 of the *B. japonicum* Cpn60s can complement a GroEL deficient *E. coli* strain, indicating that they are fully functional chaperonins [168]. Perhaps one or more proteins that mediate nitrogen metabolism in this bacterium are Cpn60 substrates, and having multiple copies of this gene positively impacts bacterial growth. However, it has been reported that only one *cpn60* copy (out of those five) is essential for *Sinorhizobium meliloti* viability and successful symbiosis, suggesting functional redundancy of the other copies [169]. Nonetheless, the possibility of undiscovered moonlighting functions of the other Cpn60s should not be discarded.

With regard to the question of horizontal (lateral) gene transfer observed in α-proteobacteria, one or more copies of the Cpn60 gene have been found encoded on rhizobia plasmids [170]. Therefore, these copies of cpn60 are not essential and can be transferred across species and potentially jump from the plasmid back into the chromosome of a different bacterium, explaining the distribution observed in the multi-taxa clades. A second possibility to take into account is gene conversion, which is a non-reciprocal transfer of genetic information between two gene copies [171]. Gene conversion is similar to homologous recombination with the difference of that the transfer of genetic material from a donor sequence towards a homologous recipient sequence is unidirectional [171]. There is experimental evidence that this process, previously thought to be restricted to eukaryotic organisms, happens in bacteria [172] and accounts for antigenic variation [173]. Thus, it is reasonable to speculate that homologous sections of a Cpn60 gene contained in a chromosome could be transferred to a Cpn60 encoded on a plasmid, or vice versa, leading to alterations in the recipient, but not in the donor, sequence. Gene conversion could explain why sequences of phylogenetically diverse bacteria display

enough similarities with the Rhizobiales to form a clade. Lastly, the most divergent mixed-taxa clade of Cpn60s could be a long-branch attraction (LBA) artifact. LBA is a methodological artifact that erroneously groups highly divergent branches [174]. This is that the sequences forming the most divergent clade found here are not phylogenetically related but they are divergent enough to not be related with any other sequence in the tree, thus forming a separate clade. Although LBA artifacts are more common in Parsimony analysis [175], the maximum likelihood method is not exempt of producing erroneous topologies due to LBA [176]. A possibility to test if the divergent clade found is really formed by phylogenetically related sequences would be to add more members of the taxa (alpha proteobacteria) to break out the clade. Another option would be to use a less related outgroup to root the tree.

Despite what has been previously discussed, the function of any of the highly divergent Cpn60s found in this study has not been reported so far. Thus, it is still unclear why these organisms are retaining additional copies of Cpn60 even if they might have lost their protein-folding function. One attractive hypothesis is the acquisition of a new advantageous function. Analysis of conserved regions (relative to their most closely related orthologs rather than to their paralogs) in the more divergent Cpn60s identified in this study could reveal clues as to their putative new functions. In fact, a recently described computational method was used to identify moonlighting domains in proteins by using putative functional motifs from conserved blocks of amino acids to search for functions in the data base [177]. A similar strategy could be used to predict possible functions of the highly divergent Cpn60s identified in this study.

4.3.2 Phylogeny of bacterial moonlighting Cpn60s

Numerous moonlighting activities have been attributed to bacterial chaperonins and it is intriguing that not all Cpn60s exhibit the same functions, although membrane localization seems to be the most common feature of moonlighting Cpn60s [78]. Herein, phylogenetic relationships between moonlighting Cpn60s were analysed. The bulk of the Cpn60s analysed in this study belong to pathogenic bacteria because pathogenicity related moonlighting functions have been mostly described so far. Overall, phylogenetic relationships between moonlighting Cpn60s and their paralogs correlated with the 16S

RNA phylogeny. However, some events of gene duplication and rapid evolution were found.

Interestingly, it seems that an early gene duplication event, followed by rapid evolution of one copy, together with a second duplication occurred in the Chlamydiae. Similar results were reported by McNally and Fares (2007), who showed that Cpn60.3 greatly diverges from Cpn60.1 whereas an intermediate functional divergence pattern was observed in Cpn60.2 with respect to Cpn60.1 [178]. In the same study, amino acid replacements at folding-related functional regions were detected, suggesting loss of canonical function. Indeed, only the Cpn60.1 of *Chlamydia trachomatis* is able to complement an *E. coli groEL* mutant [179]. Additionally, it has been reported that Cpn60.1, but not Cpn60.2 and Cpn60.3, is associated with the bacterial surface and mediates binding to the host cell [180]. Thus, only Cpn60.1 fulfils the definition of a moonlighting protein (i.e., having two jobs). This could indicate that since Cpn60.1 retained the essential protein-folding function, the other two chaperonins were left free of this essential duty, leading to a higher rate of mutation and possibly the acquisition of new functions. Although no specific functions of Cpn60.2 and Cpn60.3 have been identified, differential expression levels depending on the host cell type invaded by *Chlamydia* have been reported [181].

On the other hand, Mycobacterial chaperonins are more similar to each other than across other lineages, indicating that multiple copies arose after gene duplication. Given that the 3 Cpn60s of this clade have evolved at overall similar rates, one could speculate that all copies are essential, either because they participate in folding or have other advantageous function(s). Perhaps, evolutionary pressure has induced higher conservation. The Cpn60s of *Mycobacterium* spp. exhibit a plethora of functions including adhesion to host cells (*M. tuberculosis* Cpn60.2), biofilm formation (*M. smegmatis* Cpn60.1), proteolytic activity (*M. leprae* Cpn60.2) and alterations of host cell signaling (*M. tuberculosis* Cpn60.1); for details, see Henderson et al. [78]. The phylogenetic distribution of the *Mycobacterium* Cpn60 clade observed in this study suggests that homologs of the Cpn60.2 of *M. tuberculosis* have a tendency to acquire new functions. Therefore, it is possible that only Cpn60.1 has kept the essential folding function whereas Cpn60.2 performs other activities related to pathogenesis. However, studies have shown the opposite: a *M. tuberculosis*

strain lacking Cpn60.1 is viable but fails to induce an inflammatory response in mice whereas Cpn60.2 mutants are not viable [83]. Furthermore, Cpn60.2, but not Cpn60.1, can complement a *groEL*Δ *E.coli* strain [182]. Hence, it is feasible that only Cpn60.2 of *Mycobacterium* is a true moonlighting protein. Faster evolution of Cpn60.1 compared to Cpn60.2 observed in this study supports the idea that Cpn60.1 has lost its protein-folding capabilities.

A possible case of *cpn60* horizontal gene transfer was found in the phylum Firmicutes clade. All the *Mycoplasma* Cpn60s formed part of the same group, except for the *M. penetrans* Cpn60 which is phylogenetically related to the Cpn60 of *Helicobacter pylori*. A similar topology was obtained from a phylogenetic analysis performed by Clark and Tillier [44]. Strikingly, Cpn60 is absent in many species of *Mycoplasma*, indicating that these bacteria may have another heat shock regulatory system and (or) an alternate protein-folding mechanism. Interestingly, the Cpn60 of *M. penetrans* has been implicated in the ability of this bacterium to invade human cells [44]; therefore, it is possible that in this case Cpn60 is only a pathogenic factor.

Lastly, little Cpn60 divergence was observed within the Proteobacteria clade, and HtpB grouped close to other intracellular pathogens. Remarkably, the Cpn60 of *Francisella novicida* and *Francisella philomiragia* has been found in outer membrane vesicles (OMVs) and the Cpn60 of *Francisella tularensis* is released in the cytoplasm of infected cells [183,184]. In addition, the Cpn60 of *Piscirickettsia salmonis* was observed in the membrane of this bacterium by immunogold staining (unpublished results from our lab). HtpB is also found in the outer membrane [91], OMVs [93] and in the host cell cytoplasm [99]. Perhaps the Cpn60 of Gama proteobacteria has an export signal that induces translocation of this otherwise cytosolic protein into extracytoplasmic locations where it performs moonlighting functions.

Taken together, phylogenetic analysis showed that events of gene duplication, gene transfer and possible gene conversion occurred in the Cpn60 family. Some conclusions can be drawn from these findings. For example, having multiple copies of the Cpn60 gene seems to positively impact bacterial fitness as seen in the Rhizobiales. Even though the folding function could have been lost after rapid evolution, it is possible that accessory

copies of Cpn60 at some point acquire a new advantageous function assuring permanent residency in the genome.

It is important to notice that in all the paralogy cases analysed here, the Cpn60 copy that has moonlighting function(s) was the most conserved and probably the original ortholog. This indicates that gene duplication was not induced to resolve the conflict of having two functions in one protein as proposed by Fares (2014) [69]. Contrary to the model proposed by Fares (2014) [69] where the common ancestor of the chaperonin family contained all the functions observed in the orthologs derived from it, I think the all Cpn60s have the potential to acquire new functions depending on the evolutionary constraints it is under without losing its folding activity. This would explain why cases of moonlighting Cpn60s are found in bacteria harboring only one copy of the gene and also would account for the huge variety of functions reported so far.

4.4 CONCLUSION AND SIGNIFICANCE OF STUDY

Herein, an interaction of HtpB with the proteasome-associated protein hECM29 was discovered. In addition, putative key amino acids for moonlighting functions were successfully predicted using the ET bioinformatics method. Among the predicted amino acids, K298, N507, H473 and K474 were determined to be part of the HtpB surface that interacts with hECM29. Furthermore, although phylogenetic analysis of Cpn60s showed cases of gene duplications, variable rates of evolution, and possible horizontal gene transference events, HtpB seems to have co-evolved only with *L. pneumophila*, conserving its essential protein-folding function while acquiring novel moonlighting functions that support the intracellular lifestyle of *L. pneumophila*.

Although the interaction between HtpB and hECM29 needs to be further characterized *in vivo*, I propose that the HtpB-hECM29 interaction is a strategy used by *L. pneumophila* to successfully infect mammalian cells by hijacking the ubiquitin-proteasome/ERAD pathway and altering TLR3 signaling. Further experiments would need to include immunolocalization of HtpB-hECM29-proteasomes in mammalian cells and characterization of *L. pneumophila* infection in hECM29-depleted cells

I can conclude that the ET analysis method is reliable and that the strategy used in this study can be used to predict and characterize amino acids involved in interactions between HtpB and other proteins. Herein, amino acids involved in the HtpB-hECM29 interaction were accurately predicted and from these results I suggest that moonlighting activities can be acquired after substitutions in a few residues preferentially located in "unused" areas of the Cpn60 molecule. Further experimentation with other moonlighting Cpn60s would clarify the mechanism that allows the evolution of additional biologically active sites in chaperonins.

The phylogenetic analysis performed here confirms high conservation in the Cpn60 family. Interestingly, moonlighting Cpn60s are also highly conserved despite of the huge variety of functions attributed to them. Taken these results into consideration, I hypothesise that all Cpn60s are prone to acquire new functions in response to constraints imposed during adaptive evolution and the environmental conditions that surround the bacteria in which these Cpn60s reside.

Finally, this study contributes to a better understanding of *L. pneumophila*'s pathogenesis as the interaction of HtpB with ECM29 could account for some of the events that occur during legionella infection. Also, the evolutionary trace and phylogenetic analysis add to the currently limited knowledge about the evolution of moonlighting.

REFERENCES

1. Brenner DJ, Steigerwalt AG, McDade JE. Classification of the Legionnaires' disease bacterium: *Legionella pneumophila*, genus novum, species nova, of the family Legionellaceae, familia nova. *Ann Intern Med*. 1979;90(4):656-658.

2. Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. Ecological distribution of *Legionella pneumophila*. *Appl Environ Microbiol*. 1981;41(1):9-16.

3. Rowbotham TJ. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Pathol*. 1980;33(12):1179-1183.

4. Rogers J, Dowsett AB, Dennis PJ, Lee JV, Keevil CW. Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl Environ Microbiol*. 1994;60(5):1585-1592.

5. Murga R, Forster TS, Brown E, Pruckler JM, Fields BS, Donlan RM. Role of biofilms in the survival of *Legionella pneumophila* in a model potable-water system. *Microbiology*. 2001;147(Pt 11):3121-3126.

6. Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, Harris J, Mallison GF, Martin SM, McDade JE, Shepard CC, Brachman PS. Legionnaires' disease: Description of an epidemic of pneumonia. *N Engl J Med*. 1977;297(22):1189-1197.

7. Glick TH, Gregg MB, Berman B, Mallison G, Rhodes WW,Jr, Kassanoff I. Pontiac fever. an epidemic of unknown etiology in a health department: I. clinical and epidemiologic aspects. *Am J Epidemiol*. 1978;107(2):149-160.

8. Sanchez JL, Polyak CS, Kolavic SA, Brokaw JK, Birkmire SE, Valcik JA. Investigation of a cluster of *Legionella pneumophila* infections among staff at a federal research facility. *Mil Med*. 2001;166(9):753-758.

9. Timbury MC, Donaldson JR, McCartney AC, Fallon RJ, Sleigh JD, Lyon D, Orange GV, Baird DR, Winter J, Wilson TS. Outbreak of Legionnaires' disease in Glasgow royal infirmary: Microbiological aspects. *J Hyg (Lond)*. 1986;97(3):393-403.

10. Spitalny KC, Vogt RL, Orciari LA, Witherell LE, Etkind P, Novick LF. Pontiac fever associated with a whirlpool spa. *Am J Epidemiol*. 1984;120(6):809-817.

11. Marston BJ, Lipman HB, Breiman RF. Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality. *Arch Intern Med*. 1994;154(21):2417-2422.

12. Farnham A, Alleyne L, Cimini D, Balter S. Legionnaires' disease incidence and risk factors, New York, USA, 2002-2011. *Emerg Infect Dis*. 2014;20(11):1795-1802.

13. Hägele S, Köhler R, Merkert H, Schleicher M, Hacker J, Steinert M. *Dictyostelium discoideum*: A new host model system for intracellular pathogens of the genus legionella. *Cell Microbiol*. 2000;2(2):165-171.

14. Alli OA1, Gao LY, Pedersen LL, Zink S, Radulic M, Doric M, Abu Kwaik Y. Temporal pore formation-mediated egress from macrophages and alveolar epithelial cells by *Legionella pneumophila*. *Infect Immun*. 2000;68(11):6431-6440.

15. Berger KH, Merriam JJ, Isberg RR. Altered intracellular targeting properties associated with mutations in the *Legionella pneumophila* dotA gene. *Mol Microbiol*. 1994;14(4):809-822.

16. Brand BC, Sadosky AB, Shuman HA. The *Legionella pneumophila* icm locus: A set of genes required for intracellular multiplication in human macrophages. *Mol Microbiol*. 1994;14(4):797-808.

17. Segal G, Purcell M, Shuman HA. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc Natl Acad Sci U S A*. 1998;95(4):1669-1674.

18. Vogel JP, Andrews HL, Wong SK, Isberg RR. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science*. 1998;279(5352):873-876.

19. Kubori T, Koike M, Bui XT, Higaki S, Aizawa S, Nagai H. Native structure of a type IV secretion system core complex essential for *Legionella* pathogenesis. *Proc Natl Acad Sci U S A*. 2014;111(32):11804-11809.

20. Burstein D, Zusman T, Degtyar E, Viner R, Segal G, Pupko T. Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathog*. 2009;5(7):e1000508. doi: 10.1371/journal.ppat.1000508.

21. Lifshitz Z, Burstein D, Peeri M, Zusman T, Schwartz K, Shuman HA, Pupko T, Segal G. Computational modeling and experimental validation of the *Legionella* and *Coxiella* virulence-related type-IVB secretion signal. *Proc Natl Acad Sci U S A*. 2013;110(8):E707-15.

22. Zhu W1, Banga S, Tan Y, Zheng C, Stephenson R, Gately J, Luo ZQ. Comprehensive identification of protein substrates of the dot/icm type IV transporter of *Legionella pneumophila*. *PLoS One*. 2011;6(3):e17638. doi: 10.1371/journal.pone.0017638.

23. O'Connor TJ, Adepoju Y, Boyd D, Isberg RR. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. *Proc Natl Acad Sci U S A*. 2011;108(36):14733-14740.

24. O'Connor TJ, Boyd D, Dorer MS, Isberg RR. Aggravating genetic interactions allow a solution to redundancy in a bacterial pathogen. *Science*. 2012;338(6113):1440-1444.

25. McNeil PL, Tanasugarn L, Meigs JB, Taylor DL. Acidification of phagosomes is initiated before lysosomal enzyme activity is detected. *J Cell Biol*. 1983;97(3):692-702.

26. Gorvel JP, Chavrier P, Zerial M, Gruenberg J. Rab5 controls early endosome fusion in vitro. *Cell*. 1991;64(5):915-925.

27. Mu FT, Callaghan JM, Steele-Mortimer O, Stenmark H, Parton RG, Campbell PL, McCluskey J, Yeo JP, Tock EP, Toh BH. EEA1, an early endosome-associated protein. EEA1 is a conserved alpha-helical peripheral membrane protein flanked by cysteine "fingers" and contains a calmodulin-binding IQ motif. *J Biol Chem*. 1995;270(22):13503-13511.

28. Feng Y, Press B, Wandinger-Ness A. Rab 7: An important regulator of late endocytic membrane traffic. *J Cell Biol*. 1995;131(6 Pt 1):1435-1452.

29. Huynh KK, Eskelinen EL, Scott CC, Malevanets A, Saftig P, Grinstein S. LAMP proteins are required for fusion of lysosomes with phagosomes. *EMBO J*. 2007;26(2):313-324.

30. Flannagan RS, Cosio G, Grinstein S. Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nat Rev Microbiol*. 2009;7(5):355-366.

31. Kinchen JM, Ravichandran KS. Phagosome maturation: Going through the acid test. *Nat Rev Mol Cell Biol*. 2008;9(10):781-795.

32. Clemens DL, Lee BY, Horwitz MA. Deviant expression of Rab5 on phagosomes containing the intracellular pathogens mycobacterium tuberculosis and *Legionella pneumophila* is associated with altered phagosomal fate. *Infect Immun*. 2000;68(5):2671-2684.

33. Santic M, Molmeret M, Abu Kwaik Y. Maturation of the *Legionella pneumophila*-containing phagosome into a phagolysosome within gamma interferon-activated macrophages. *Infect Immun*. 2005;73(5):3166-3171.

34. Horwitz MA. The Legionnaires' disease bacterium (*Legionella pneumophila*) inhibits phagosome-lysosome fusion in human monocytes. *J Exp Med*. 1983;158(6):2108-2126.

35. Horwitz MA, Maxfield FR. *Legionella pneumophila* inhibits acidification of its phagosome in human monocytes. *J Cell Biol*. 1984;99(6):1936-1943.

36. So EC, Mattheis C, Tate EW, Frankel G, Schroeder GN. Creating a customized intracellular niche: Subversion of host cell signaling by *Legionella* type IV secretion system effectors. *Can J Microbiol*. 2015:1-19.

37. Machner MP, Isberg RR. Targeting of host Rab GTPase function by the intravacuolar pathogen *Legionella pneumophila*. *Dev Cell*. 2006;11(1):47-56.

38. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR. A bacterial guanine nucleotide exchange factor activates ARF on legionella phagosomes. *Science*. 2002;295(5555):679-682.

39. Kagan JC, Stein MP, Pypaert M, Roy CR. *Legionella* subvert the functions of Rab1 and Sec22b to create a replicative organelle. *J Exp Med*. 2004;199(9):1201-1211.

40. Horwitz MA. Formation of a novel phagosome by the legionnaires' disease bacterium (legionella pneumophila) in human monocytes. *J Exp Med*. 1983;158(4):1319-1331.

41. Tilney LG, Harb OS, Connelly PS, Robinson CG, Roy CR. How the parasitic bacterium *Legionella pneumophila* modifies its phagosome and transforms it into rough ER: Implications for conversion of plasma membrane to the ER membrane. *J Cell Sci*. 2001;114(Pt 24):4637-4650.

42. Gerhardt H, Walz MJ, Faigle M, Northoff H, Wolburg H, Neumeister B. Localization of *Legionella* bacteria within ribosome-studded phagosomes is not restricted to legionella pneumophila. *FEMS Microbiol Lett*. 2000;192(1):145-152.

43. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem*. 2013;82:323-355.

44. Clark GW, Tillier ER. Loss and gain of GroEL in the *Mollicutes*. *Biochem Cell Biol*. 2010;88(2):185-194.

45. Horwich AL, Fenton WA, Chapman E, Farr GW. Two families of chaperonin: Physiology and mechanism. *Annu Rev Cell Dev Biol*. 2007;23:115-145.

46. Horwich AL, Fenton WA. Chaperonin-mediated protein folding: Using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys*. 2009;42(2):83-116.

47. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB. The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature*. 1994;371(6498):578-586.

48. Chandrasekhar GN, Tilly K, Woolford C, Hendrix R, Georgopoulos C. Purification and properties of the GroES morphogenetic protein of *Escherichia coli. J Biol Chem*. 1986;261(26):12414-12419.

49. Hunt JF, Weaver AJ, Landry SJ, Gierasch L, Deisenhofer J. The crystal structure of the GroES co-chaperonin at 2.8 A resolution. *Nature*. 1996;379(6560):37-45.

50. Chaudhry C, Farr GW, Todd MJ, Rye HS, Brunger AT, Adams PD, Horwich AL, Sigler PB. Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: Function, structure and energetics. *EMBO J*. 2003;22(19):4877-4887.

51. Chaudhry C, Horwich AL, Brunger AT, Adams PD. Exploring the structural dynamics of the E.coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. *J Mol Biol*. 2004;342(1):229-245.

52. Bartolucci C, Lamba D, Grazulis S, Manakova E, Heumann H. Crystal structure of wild-type chaperonin GroEL. *J Mol Biol*. 2005;354(4):940-951.

53. Xu Z, Horwich AL. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. (cover story). *Nature*. 1997;388(6644):741.

54. Boisvert DC, Wang J, Otwinowski Z, Horwich AL, Sigler PB. The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat Struct Biol*. 1996;3(2):170-177.

55. Clare DK, Vasishtan D, Stagg S, Quispe J, Farr GW, Topf M, Horwich AL, Saibil HR. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*. 2012;149(1):113-123.

56. Koike-Takeshita A, Arakawa T, Taguchi H, Shimamura T. Crystal structure of a symmetric football-shaped GroEL:GroES2-ATP14 complex determined at 3.8 Å reveals rearrangement between two GroEL rings. *J Mol Biol*. 2014;426(21):3634-3641.

57. Yang D, Ye X, Lorimer GH. Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. *Proc Natl Acad Sci U S A*. 2013;110(46):E4298-305.

58. Fei X, Ye X, LaRonde NA, Lorimer GH. Formation and structures of GroEL:GroES2 chaperonin footballs, the protein-folding functional form. *Proc Natl Acad Sci U S A*. 2014;111(35):12775-12780.

59. Llorca O, Marco S, Carrascosa JL, Valpuesta JM. Symmetric GroEL-GroES complexes can contain substrate simultaneously in both GroEL rings. *FEBS Lett*. 1997;405(2):195-199.

60. Sameshima T, Ueno T, Iizuka R, Ishii N, Terada N, Okabe K, Funatsu T. Football- and bullet-shaped GroEL-GroES complexes coexist during the reaction cycle. *J Biol Chem*. 2008;283(35):23765-23773.

61. Tanaka N, Fersht AR. Identification of substrate binding site of GroEL minichaperone in solution1. *J Mol Biol*. 1999;292(1):173-180.

62. Tyagi NK, Fenton WA, Horwich AL. GroEL/GroES cycling: ATP binds to an open ring before substrate protein favoring protein binding and production of the native state. *Proc Natl Acad Sci U S A*. 2009;106(48):20264-20269.

63. Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, Saibil HR, Fenton WA, Horwich AL. Mechanism of GroEL action: Productive release of polypeptide from a sequestered position under GroES. *Cell*. 1995;83(4):577-587.

64. Campbell RM, Scanes CG. Endocrine peptides 'moonlighting' as immune modulators: Roles for somatostatin and GH-releasing factor. *J Endocrinol*. 1995;147(3):383-396.

65. Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci*. 1999;24(1):8-11.

66. Mani M, Chen C, Amblee V, Liu H, Mathur T, Zwicke G, Zabad S, Patel B, Thakkar J, Jeffery CJ. MoonProt: A database for proteins that are known to moonlight. *Nucleic Acids Res*. 2015;43:D277-82.

67. Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev*. 2003;17(23):2875-2888.

68. Fushinobu S, Nishimasu H, Hattori D, Song HJ, Wakagi T. Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase. *Nature*. 2011;478(7370):538-541.

69. Fares MA. The evolution of protein moonlighting: Adaptive traps and promiscuity in the chaperonins. *Biochem Soc Trans*. 2014;42(6):1709-1714.

70. Williams TA, Codoner FM, Toft C, Fares MA. Two chaperonin systems in bacterial genomes with distinct ecological roles. *Trends Genet*. 2010;26(2):47-51.

71. Copley SD. An evolutionary perspective on protein moonlighting. *Biochem Soc Trans*. 2014;42(6):1684-1691.

72. Lund PA. Multiple chaperonins in bacteria: why so many? *FEMS Microbiol Rev*. 2009;33(4):785-800.

73. Henderson B, Martin A. Bacterial moonlighting proteins and bacterial virulence. *Curr Top Microbiol Immunol*. 2013;358:155-213.

74. Jung DW, Kim WH, Williams DR. Chemical genetics and its application to moonlighting in glycolytic enzymes. *Biochem Soc Trans*. 2014;42(6):1756-1761.

75. Paludo GP, Lorenzatto KR, Bonatto D, Ferreira HB. Systems biology approach reveals possible evolutionarily conserved moonlighting functions for enolase. *Comput Biol Chem*. 2015;58:1-8.

76. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I, Gómez A, Querol E, Cedano J. Bioinformatics and moonlighting proteins. *Front Bioeng Biotechnol*. 2015;3:90.

77. Ensgraber M, Loos M. A 66-kilodalton heat shock protein of *Salmonella typhimurium* is responsible for binding of the bacterium to intestinal mucus. *Infect Immun*. 1992;60(8):3072-3078.

78. Henderson B, Fares MA, Lund PA. Chaperonin 60: A paradoxical, evolutionarily conserved protein family with multiple moonlighting functions. *Biol Rev Camb Philos Soc*. 2013;88(4):955-987.

79. Hickey TB, Ziltener HJ, Speert DP, Stokes RW. *Mycobacterium tuberculosis* employs Cpn60.2 as an adhesin that binds CD43 on the macrophage surface. *Cell Microbiol*. 2010;12(11):1634-1647.

80. Kamiya S, Yamaguchi H, Osaki T, Taguchi H. A virulence factor of *Helicobacter pylori*: Role of heat shock protein in mucosal inflammation after *H. pylori* infection. *J Clin Gastroenterol*. 1998;27 Suppl 1:S35-9.

81. Gonzalez-Lopez MA, Velazquez-Guadarrama N, Romero-Espejel ME, Olivares-Trejo Jde J. *Helicobacter pylori* secretes the chaperonin GroEL (HSP60), which binds iron. *FEBS Lett*. 2013;587(12):1823-1828.

82. Lewthwaite JC, Coates AR, Tormay P, Singh M, Mascagni P, Poole S, Roberts M, Sharp L, Henderson B. *Mycobacterium tuberculosis* chaperonin 60.1 is a more potent cytokine stimulator than chaperonin 60.2 (Hsp 65) and contains a CD14-binding domain. *Infect Immun*. 2001;69(12):7349-7355.

83. Hu Y, Henderson B, Lund PA, Tormay P, Ahmed MT, Gurcha SS, Besra GS, Coates AR. A *Mycobacterium tuberculosis* mutant lacking the GroEL homologue Cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. *Infect Immun*. 2008;76(4):1535-1546.

84. Sakamoto K. The pathology of *Mycobacterium tuberculosis* infection. *Vet Pathol*. 2012;49(3):423-439.

85. Goulhen F, Hafezi A, Uitto VJ, Hinode D, Nakamura R, Grenier D, Mayrand D. Subcellular localization and cytotoxic activity of the GroEL-like protein isolated from *Actinobacillus actinomycetemcomitans*. *Infect Immun*. 1998;66(11):5307-5313.

86. Zhang L, Pelech S, Uitto VJ. Long-term effect of heat shock protein 60 from *Actinobacillus actinomycetemcomitans* on epithelial cell viability and mitogen-activated protein kinases. *Infect Immun*. 2004;72(1):38-45.

87. Yoshida N, Oeda K, Watanabe E, Mikami T, Fukita Y, Nishimura K, Komai K, Matsuda K. Protein function. Chaperonin turned insect toxin. *Nature*. 2001;411(6833):44.

88. Portaro FC, Hayashi MA, De Arauz LJ, Palma MS, Assakura MT, Silva CL, de Camargo AC. The *Mycobacterium leprae* hsp65 displays proteolytic activity. Mutagenesis studies indicate that the *M. leprae* hsp65 proteolytic activity is catalytically related to the HslVU protease. *Biochemistry*. 2002;41(23):7400-7406.

89. Hoffman PS, Butler CA, Quinn FD. Cloning and temperature-dependent expression in *Escherichia coli* of a *Legionella pneumophila* gene coding for a genus-common 60-kilodalton antigen. *Infect Immun*. 1989;57(6):1731-1739.

90. Garduño RA, Chong A, Nasrallah GK, Allan DS. The *Legionella pneumophila* chaperonin - an unusual multifunctional protein in unusual locations. *Front Microbiol*. 2011;2:122.

91. Garduño RA, Faulkner G, Trevors MA, Vats N, Hoffman PS. Immunolocalization of Hsp60 in *Legionella pneumophila*. *J Bacteriol*. 1998;180(3):505-513.

92. Garduño RA, Garduño E, Hiltz M, Hoffman PS. Intracellular growth of *Legionella pneumophila* gives rise to a differentiated form dissimilar to stationary-phase forms. *Infect Immun*. 2002;70(11):6273-6283.

93. Galka F, Wai SN, Kusch H, Engelmann S, Hecker M, Schmeck B, Hippenstiel S, Uhlin BE, Steinert M. Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of outer membrane vesicles. *Infect Immun*. 2008;76(5):1825-1836.

94. Hoffman PS, Houston L, Butler CA. *Legionella pneumophila htpAB* heat shock operon: Nucleotide sequence and expression of the 60-kilodalton antigen in *L. pneumophila*-infected HeLa cells. *Infect Immun*. 1990;58(10):3380-3387.

95. Fernandez RC, Logan SM, Lee SH, Hoffman PS. Elevated levels of *Legionella pneumophila* stress protein Hsp60 early in infection of human monocytes and L929 cells correlate with virulence. *Infect Immun*. 1996;64(6):1968-1976.

96. Garduño RA, Garduño E, Hoffman PS. Surface-associated hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. *Infect Immun*. 1998;66(10):4602-4610.

97. Chong A, Lima CA, Allan DS, Nasrallah GK, Garduño RA. The purified and recombinant *Legionella pneumophila* chaperonin alters mitochondrial trafficking and microfilament organization. *Infect Immun*. 2009;77(11):4724-4739. d

98. Chong A, Riveroll A, Allan DS, Garduño E, Garduño RA. The Hsp60 chaperonin of *Legionella pneumophila*: An intriguing player in infection of host cells. In: *Legionella*: State of the Art 30 years after its recognition (edited by N.P. Cianciotto, Y. Abu Kwaik, P.H. Edelstein, B.S. Fields, D.F. Geary, T.G. Harrison, C.A. Joseph, R.M. Ratcliff, J.E. Stout, M.S. Swanson). Washington: ASM Press; 2006:255-260.

99. Nasrallah GK, Riveroll AL, Chong A, Murray LE, Lewis PJ, Garduño RA. *Legionella pneumophila* requires polyamines for optimal intracellular growth. *J Bacteriol*. 2011;193(17):4346-4360.

100. Garduño RA, Chong A. The *Legionella pneumophila* chaperonin 60 and the art of keeping several moonlighting jobs. In: Henderson B, ed. Vol 7. Springer Netherlands; 2013:143-160.

101. Nasrallah GK. A yeast two-hybrid screen reveals a strong interaction between the *Legionella* chaperonin Hsp60 and the host cell small heat shock protein Hsp10. *Acta Microbiol Immunol Hung*. 2015;62(2):121-135.

102. Fernandez-Moreira E, Helbig JH, Swanson MS. Membrane vesicles shed by *Legionella pneumophila* inhibit fusion of phagosomes with lysosomes. *Infect Immun*. 2006;74(6):3285-3295.

103. Nasrallah GK, Gagnon E, Orton DJ, Garduño RA. The *htpAB* operon of *Legionella pneumophila* cannot be deleted in the presence of the *groE* chaperonin operon of *Escherichia coli*. *Can J Microbiol*. 2011;57(11):943-952.

104. Sadosky AB, Wiater LA and Shuman HA. Identification of *Legionella pneumophila* genes required for growth within and killing of human macrophages. *Infect. Immun*. 1993;61:5361–5373.

105. Sambrook J. *Molecular cloning : A laboratory manual.* 2nd ed. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press; 1989.

106. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol*. 2011;7:539.

107. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-2461.

108. Michener C, Sokal R. A quantitative approach to a problem in classification. *Evolution*. 1957;11(2):130-162.

109. Wilkins A, Erdin S, Lua R, Lichtarge O. Evolutionary trace for prediction and redesign of protein functional sites. *Methods Mol Biol*. 2012;819:29-42.

110. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*. 1992;89(22):10915-10919.

111. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*. 2003;31(13):3316-3319.

112. Wu M, Chatterji S, Eisen JA. Accounting for alignment uncertainty in phylogenomics. *PLoS One*. 2012;7(1):e30288. doi: 10.1371/journal.pone.0030288.

113. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27(8):1164-1165.

114. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008;25(7):1307-1320.

115. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307-321.

116. Belshaw R, de Oliveira T, Markowitz S, Rambaut A. The RNA virus database. *Nucleic Acids Res*. 2009;37:D431-5.

117. Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Horwich AL, Saibil HR. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*. 2001;107(7):869-879.

118. Wang J, Chen L. Domain motions in GroEL upon binding of an oligopeptide. *J Mol Biol*. 2003;334(3):489-499.

119. Wang J, Boisvert DC. Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)14 at 2.0A resolution. *J Mol Biol*. 2003;327(4):843-855.

120. Fenton WA, Kashi Y, Furtak K, Horwich AL. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*. 1994;371(6498):614-619.

121. Chen L, Sigler PB. The crystal structure of a GroEL/peptide complex: Plasticity as a basis for substrate diversity. *Cell*. 1999;99(7):757-768.

122. O'Flaherty E, Kaye J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics*. 2003;4(1):13.

123. Nagase T, Ishikawa K, Suyama M, Kikuno R, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Research*. 1998;5(5):277-286.

124. Puch CBMd, Barbier E, Kraut A, Couté Y, Fuchs J, Buhot A, Livache T, Sève M, Favier A, Douki T, Gasparutto D, Sauvaigo S, Breton J. TOX4 and its binding partners recognize DNA adducts generated by platinum anticancer drugs. *Arch Biochem Biophys*. 2011;507(2):296-303.

125. Lee JH, You J, Dobrota E, Skalnik DG. Identification and characterization of a novel human PP1 phosphatase complex. *J Biol Chem*. 2010;285(32):24466-24476.

126. Lee SJ, Lee JK, Maeng YS, Kim YM, Kwon YG. Langerhans cell protein 1 (LCP1) binds to PNUTS in the nucleus: Implications for this complex in transcriptional regulation. *Exp Mol Med*. 2009;41(3):189-200.

127. Aidinis V, Bonaldi T, Beltrame M, Santagata S, Bianchi ME, Spanopoulou E. The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. *Mol Cell Biol*. 1999;19(10):6532-6542.

128. Lussier M, White AM, Sheraton J, di Paolo T, Treadwell J, Southard SB, Horenstein CI, Chen-Weiner J, Ram AF, Kapteyn JC, Roemer TW, Vo DH, Bondoc DC, Hall J, Zhong WW, Sdicu AM, Davies J, Klis FM, Robbins PW, Bussey H. Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. *Genetics*. 1997;147(2):435-450.

129. Kajava AV, Gorbea C, Ortega J, Rechsteiner M, Steven AC. New HEAT-like repeat motifs in proteins regulating proteasome structure and function. *J Struct Biol*. 2004;146(3):425-430.

130. Leggett DS, Hanna J, Borodovsky A, Crosas B, Schmidt M, Baker RT, Walz T, Ploegh H, Finley D. Multiple associated proteins regulate proteasome structure and function. *Mol Cell*. 2002;10(3):495-507.

131. Neuwald AF, Hirano T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res*. 2000;10(10):1445-1452.

132. Lehmann A, Niewienda A, Jechow K, Janek K, Enenkel C. Ecm29 fulfils quality control functions in proteasome assembly. *Mol Cell*. 2010;38(6):879-888.

133. Tai HC, Besche H, Goldberg AL, Schuman EM. Characterization of the brain 26S proteasome and its interacting proteins. *Front Mol Neurosci*. 2010;3:10.3389/fnmol.2010.00012. doi: 10.3389/fnmol.2010.00012.

134. Bedford L, Paine S, Sheppard PW, Mayer RJ, Roelofs J. Assembly, structure and function of the 26S proteasome. *Trends Cell Biol*. 2010;20(7):391-401.

135. Ng W, Sergeyenko T, Zeng N, Brown JD, Romisch K. Characterization of the proteasome interaction with the Sec61 channel in the endoplasmic reticulum. *J Cell Sci*. 2007;120(Pt 4):682-691.

136. Fabunmi RP, Wigley WC, Thomas PJ, DeMartino GN. Activity and regulation of the centrosome-associated proteasome. *J Biol Chem*. 2000;275(1):409-413.

137. Gorbea C, Goellner GM, Teter K, Holmes RK, Rechsteiner M. Characterization of mammalian Ecm29, a 26 S proteasome-associated protein that localizes to the nucleus and membrane vesicles. *J Biol Chem*. 2004;279(52):54849-54861.

138. Gorbea C, Pratt G, Ustrell V, Bell R, Sahasrabudhe S, Hughes RE, Rechsteiner M. A protein interaction network for Ecm29 links the 26 S proteasome to molecular motors and endosomal components. *J Biol Chem*. 2010;285(41):31616-31633.

139. Ryabova LV, Virtanen I, Olink-Coux M, Scherrer K, Vassetzky SG. Distribution of prosome proteins and their relationship with the cytoskeleton in oogenesis of *Xenopus laevis*. *Mol Reprod Dev*. 1994;37(2):195-203.

140. Arcangeletti C, Sütterlin R, Aebi U, De Conto F, Missorini S, Chezzi C, Scherrer K. Visualization of prosomes (MCP-proteasomes), intermediate filament and actin networks by "instantaneous fixation" preserving the cytoskeleton. *J Struct Biol*. 1997;119(1):35-58.

141. Haarer B, Aggeli D, Viggiano S, Burke DJ, Amberg DC. Novel interactions between actin and the proteasome revealed by complex haploinsufficiency. *PLoS Genet*. 2011;7(9):e1002288. doi: 10.1371/journal.pgen.1002288.

142. Moshe A, Belausov E, Niehl A, Heinlein M, Czosnek H, Gorovits R. The tomato yellow leaf curl virus V2 protein forms aggregates depending on the cytoskeleton integrity and binds viral genomic DNA. *Sci Rep*. 2015;5:9967.

143. Dorer MS, Kirton D, Bader JS, Isberg RR. RNA interference analysis of *Legionella* in drosophila cells: Exploitation of early secretory apparatus dynamics. *PLoS Pathog*. 2006;2(4):e34. doi: 10.1371/journal.ppat.0020034.

144. Kubori T, Shinzawa N, Kanuka H, Nagai H. *Legionella* metaeffector exploits host proteasome to temporally regulate cognate effector. *PLoS Pathog*. 2010;6(12):e1001216. doi: 10.1371/journal.ppat.1001216.

145. Price CT, Al-Khodor S, Al-Quadan T, Abu Kwaik Y. Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of *Legionella pneumophila* within macrophages and amoebae. *Infect Immun*. 2010;78(5):2079-2088.

146. Hsu F, Luo X, Qiu J, Teng YB, Jin J, Smolka MB, Luo ZQ, Mao Y. The *Legionella* effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling. *Proc Natl Acad Sci U S A*. 2014;111(29):10538-10543.

147. Gorbea C, Rechsteiner M, Vallejo JG, Bowles NE. Depletion of the 26S proteasome adaptor Ecm29 increases toll-like receptor 3 signaling. *Sci Signal*. 2013;6(295):ra86. doi: 10.1126/scisignal.2004301.

148. Chattopadhyay S, Sen GC. dsRNA-activation of TLR3 and RLR signaling: Gene induction-dependent and independent effects. *J Interferon Cytokine Res*. 2014;34(6):427-436.

149. Derbigny WA, Johnson RM, Toomey KS, Ofner S, Jayarapu K. The *Chlamydia muridarum*-induced IFN-beta response is TLR3-dependent in murine oviduct epithelial cells. *J Immunol*. 2010;185(11):6689-6697.

150. Derbigny WA, Shobe LR, Kamran JC, Toomey KS, Ofner S. Identifying a role for toll-like receptor 3 in the innate immune response to *Chlamydia muridarum* infection in murine oviduct epithelial cells. *Infect Immun*. 2012;80(1):254-265.

151. Pyles RB, Jezek GE, Eaves-Pyles TD. Toll-like receptor 3 agonist protection against experimental *Francisella tularensis* respiratory tract infection. *Infect Immun*. 2010;78(4):1700-1710.

152. Kawashima T, Kosaka A, Yan H, Guo Z, Uchiyama R, Fukui R, Kaneko D, Kumagai Y, You DJ, Carreras J, Uematsu S, Jang MH, Takeuchi O, Kaisho T, Akira S, Miyake K, Tsutsui H, Saito T, Nishimura I, Tsuji NM. Double-stranded RNA of intestinal commensal but not pathogenic bacteria triggers production of protective interferon-beta. *Immunity*. 2013;38(6):1187-1197.

153. Xu L, Shen X, Bryan A, Banga S, Swanson MS, Luo ZQ. Inhibition of host vacuolar H+-ATPase activity by a *Legionella pneumophila* effector. *PLoS Pathog*. 2010;6(3):e1000822. doi: 10.1371/journal.ppat.1000822.

154. Kagan JC, Stein MP, Pypaert M, Roy CR. *Legionella* subvert the functions of Rab1 and Sec22b to create a replicative organelle. *J Exp Med*. 2004;199(9):1201-1211.

155. Price CT, Al-Quadan T, Santic M, Rosenshine I, Abu Kwaik Y. Host proteasomal degradation generates amino acids essential for intracellular bacterial growth. *Science*. 2011;334(6062):1553-1557.

156. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I, Rösch P. An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell*. 2012;150(2):291-303.

157. Du J, Say RF, Lu W, Fuchs G, Einsle O. Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase. *Nature*. 2011;478(7370):534-537.

158. Jeffery CJ. Molecular mechanisms for multitasking: Recent crystal structures of moonlighting proteins. *Curr Opin Struct Biol*. 2004;14(6):663-668.

159. Brocchieri L, Karlin S. Conservation among HSP60 sequences in relation to structure, function, and evolution. *Protein Sci*. 2000;9(3):476-486.

160. Nielsen KL, Cowan NJ. A single ring is sufficient for productive chaperonin-mediated folding in vivo. *Mol Cell*. 1998;2(1):93-99.

161. Erbse A, Yifrach O, Jones S, Lund PA. Chaperone activity of a chimeric GroEL protein that can exist in a single or double ring form. *J Biol Chem*. 1999;274(29):20351-20357.

162. Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, Sigler PB, Horwich AL. Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. *Nature*. 1997;388(6644):792-798.

163. Tourasse NJ, Li WH. Selective constraints, amino acid composition, and the rate of protein evolution. *Mol Biol Evol*. 2000;17(4):656-664.

164. Quintana FJ, Cohen IR. The HSP60 immune system network. *Trends Immunol*. 2011;32(2):89-95.

165. Habich C, Kempe K, Burkart V, Van Der Zee R, Lillicrap M, Gaston H, Kolb H. Identification of the heat shock protein 60 epitope involved in receptor binding on macrophages. *FEBS Lett*. 2004;568(1-3):65-69.

166. Hogenhout SA, van der Wilk F, Verbeek M, Goldbach RW, van den Heuvel JF. Identifying the determinants in the equatorial domain of *Buchnera* GroEL implicated in binding potato leafroll virus. *J Virol*. 2000;74(10):4541-4548.

167. Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol*. 1994;44(3):527-533.

168. Fischer H-, Schneider K, Babst M, Hennecke H. GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum*. *Arch Microbiol*. 1999;171(4):279-289.

169. Bittner AN, Foltz A, Oke V. Only one of five groEL genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *J Bacteriol*. 2007;189(5):1884-1889.

170. Hernandez-Salmeron JE, Santoyo G. Phylogenetic analysis reveals gene conversions in multigene families of Rhizobia. *Genet Mol Res*. 2011;10(3):1383-1392.

171. Santoyo G, Romero D. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol Rev*. 2005;29(2):169-183.

172. Yamamoto K, Kusano K, Takahashi NK, Yoshikura H, Kobayashi I. Gene conversion in the *Escherichia coli* RecF pathway: A successive half crossing-over model. *Mol Gen Genet*. 1992;234(1):1-13.

173. Zhang QY, DeRyckere D, Lauer P, Koomey M. Gene conversion in *Neisseria gonorrhoeae*: evidence for its role in pilus antigenic variation. *Proc Natl Acad Sci U S A*. 1992;89(12):5366-5370.

174. Bergsten J. A review of long-branch attraction. *Cladistics*. 2005;21(2):163-193. d

175. Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol*. 2005;5:50. doi: 1471-2148-5-50.

176. Kuck P, Mayer C, Wagele JW, Misof B. Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS One*. 2012;7(5):e36593. doi: 10.1371/journal.pone.0036593.

177. Wong A, Gehring C, Irving HR. Conserved functional motifs and homology modeling to predict hidden moonlighting functional sites. *Front Bioeng Biotechnol*. 2015;3:82.

178. McNally D, Fares MA. *In silico* identification of functional divergence between the multiple GroEL gene paralogs in *Chlamydiae*. *BMC Evol Biol*. 2007;7:81. doi: 1471-2148-7-81.

179. Karunakaran KP, Noguchi Y, Read TD, Cherkasov A, Kwee J, Shen C, Nelson CC, Brunham RC. Molecular analysis of the multiple GroEL proteins of Chlamydiae. *J Bacteriol*. 2003;185(6):1958-1966.

180. Wuppermann FN, Molleken K, Julien M, Jantos CA, Hegemann JH. *Chlamydia pneumoniae* GroEL1 protein is cell surface associated and required for infection of HEp-2 cells. *J Bacteriol*. 2008;190(10):3757-3767.

181. Gerard HC, Whittum-Hudson JA, Schumacher HR, Hudson AP. Differential expression of three *Chlamydia trachomatis* hsp60-encoding genes in active vs. persistent infections. *Microb Pathog*. 2004;36(1):35-39.

182. Fan M, Rao T, Zacco E, Ahmed MT, Shukla A, Ojha A, Freeke J, Robinson CV, Benesch JL, Lund PA. The unusual mycobacterial chaperonins: Evidence for *in vivo* oligomerization and specialization of function. *Mol Microbiol*. 2012;85(5):934-944.

183. Lee BY, Horwitz MA, Clemens DL. Identification, recombinant expression, immunolocalization in macrophages, and T-cell responsiveness of the major extracellular proteins of *Francisella tularensis*. *Infect Immun*. 2006;74(7):4002-4013.

184. Pierson T, Matrakas D, Taylor YU, Manyam G, Morozov VN, Zhou W, van Hoek ML. Proteomic characterization and functional analysis of outer membrane vesicles of *Francisella novicida* suggests possible role in virulence and use as a vaccine. *J Proteome Res*. 2011;10(3):954-967.

APPENDIX A

			HtpB			GroEL		
Alig. Pos.	Var. (N°)	Variability (aa)	aa N°	aa	rvET rank	aa N°	aa	rvET rank
15	11	.MKLWNEARYG	1	M	234.92	-	-	-
16	9	M.INVTFQK	2	I	34.94	1	M	34.94
17	10	AMVTSPGKLD	3	M	46.31	2	A	46.31
18	10	ASVTPFGHYI	4	A	51.32	3	A	51.32
19	3	KNR	5	K	1.82	4	K	1.82
20	15	DTNEQLISMVKRHAF	6	E	134.7	5	D	134.7
21	4	VILM	7	L	47.14	6	V	47.14
22	16	KAELTRISYVQFMHND	8	R	123.99	7	K	123.99
23	7	FYHTSLR	9	F	28.73	8	F	28.73
24	10	GDSNAEHKRQ	10	G	99.31	9	G	99.31
25	16	NESAIRDQKHTGVMLP	11	D	90.99	10	N	90.99
26	14	DEKTSQANCHRVIG	12	D	133.91	11	D	133.91
27	7	ASVTLGC	13	A	18.06	12	A	18.06
28	5	RLDQK	14	R	6.33	13	R	6.33
29	15	VRKSIQEANTDLHGM	15	L	113.12	14	V	113.12
30	14	KGAQSRLMHEVNCT	16	Q	104.3	15	K	104.3
31	5	MLIVF	17	M	36.66	16	M	36.66
32	15	LEHVQIFYKMSRAGT	18	L	48.39	17	L	48.39
33	15	RKEASVQDNHTILMY	19	A	130.61	18	R	130.61
34	3	GVA	20	G	1.56	19	G	1.56
35	5	VLIMA	21	V	34.84	20	V	34.84
36	12	NDKERTSAQHLV	22	N	57.93	21	N	57.93
37	12	VAQTIKMLSREH	23	A	125.52	22	V	125.52
38	4	LIVM	24	L	19.05	23	L	19.05
39	10	ASHNEVTDCF	25	A	19.58	24	A	19.58
40	9	DNESKRHQA	26	D	84.6	25	D	84.6
41	5	ATSIV	27	A	26.68	26	A	26.68
42	4	VAIL	28	V	3.74	27	V	3.74
43	9	KRAQGTSMV	29	Q	28.94	28	K	28.94
44	6	VIATSC	30	V	41.66	29	V	41.66
45	1	T	31	T	1	30	T	1
46	6	LIAMVH	32	M	16.1	31	L	16.1
47	1	G	33	G	1	32	G	1
48	1	P	34	P	1	33	P	1
49	7	KREAGCS	35	R	31.74	34	K	31.74
50	1	G	36	G	1	35	G	1
51	7	RQLHKCN	37	R	9.63	36	R	9.63
52	10	NHTEPYLCVS	38	N	9.8	37	N	9.8
53	3	VAI	39	V	5.59	38	V	5.59
54	7	VLIAMGS	40	V	44.13	39	V	44.13
55	5	LIVMF	41	L	39.88	40	L	39.88
56	8	DEQASGNK	42	E	62.78	41	D	62.78
57	5	KQRNE	43	K	24.31	42	K	24.31
58	11	SKETAGPDNRQ	44	S	46.66	43	S	46.66
59	4	FWYA	45	Y	40.51	44	F	40.51

60	8	GTASVEND	46	G	15.15	45	G	15.15
61	9	ASNTPGILV	47	A	72.93	46	A	72.93
62	4	PSAV	48	P	3.21	47	P	3.21
63	14	TLQKRHSEVIDMAN	49	T	78.62	48	T	78.62
64	6	IVTSLM	50	V	27.73	49	I	27.73
65	5	TVIAS	51	T	14.23	50	T	14.23
66	4	KNDH	52	K	6.27	51	K	6.27
67	1	D	53	D	1	52	D	1
68	1	G	54	G	1	53	G	1
69	4	VIYC	55	V	8.84	54	V	8.84
70	4	STKA	56	S	43.99	55	S	43.99
71	3	VIA	57	V	7.41	56	V	7.41
72	3	AMI	58	A	7.68	57	A	7.68
73	4	RKNQ	59	K	23.3	58	R	23.3
74	7	ESADNGQ	60	E	22.17	59	E	22.17
75	4	IVLT	61	I	34.59	60	I	34.59
76	8	EDSKVAQT	62	E	28.78	61	E	28.78
77	5	LFPIV	63	F	14.15	62	L	14.15
78	14	EKADSTPRNQYGLV	64	E	94.94	63	E	94.94
79	10	DCNHESQAKG	65	H	46.08	64	D	46.08
80	15	KPHASRTQYNEIVGM	66	R	86.93	65	K	86.93
81	12	FYHIVLWTMKSN	67	F	70.04	66	F	70.04
82	13	ERKAQMIPLNHSV	68	M	29.84	67	E	29.84
83	10	NKRLAVDSEH	69	N	18.67	68	N	18.67
84	8	MIATQLKV	70	M	33.84	69	M	33.84
85	3	GIA	71	G	2.56	70	G	2.56
86	4	AVCT	72	A	10.89	71	A	10.89
87	10	QEKSNARGTL	73	Q	30.47	72	Q	30.47
88	7	MLVIAFT	74	M	25.65	73	M	25.65
89	7	VLIAMCF	75	V	36.26	74	V	36.26
90	11	KSQARTYNIVL	76	K	37.93	75	K	37.93
91	6	ESDVQT	77	E	10.83	76	E	10.83
92	4	VASI	78	V	13.86	77	V	13.86
93	5	ASTCV	79	A	15.03	78	A	15.03
94	10	SKNTIVQAEF	80	S	33.99	79	S	33.99
95	7	KRSQNHA	81	K	21.07	80	K	21.07
96	5	ATCSQ	82	T	10.25	81	A	10.25
97	9	NDASGQKHE	83	S	29.31	82	N	29.31
99	5	DEGNK	84	D	9.44	83	D	9.44
100	14	AVKITQNLDEHSMG	85	T	96.35	84	A	96.35
101	4	AVST	86	A	8.66	85	A	8.66
102	1	G	87	G	1	86	G	1
103	1	D	88	D	1	87	D	1
104	1	G	89	G	1	88	G	1
105	1	T	90	T	1	89	T	1
106	1	T	91	T	1	90	T	1
107	2	TS	92	T	1.64	91	T	1.64
108	3	ACS	93	A	4.71	92	A	4.71
109	6	TSIVCA	94	T	6.81	93	T	6.81
110	5	VILMC	95	V	32.06	94	V	32.06

111	5	LMIYA	96	L	10.51	95	L	10.51
112	6	ATGVCS	97	A	20.29	96	A	20.29
113	11	QERYHAGWNDS	98	R	20.39	97	Q	20.39
115	10	ASEKGNITVR	99	S	61.05	98	A	61.05
116	7	ILMVFAS	100	I	49.13	99	I	49.13
117	7	IVYALFT	101	L	55.17	100	I	55.17
118	16	TRQSHNKVAEGYMLDI	102	V	116.63	101	T	116.63
119	11	EKRAVQIDNHT	103	E	20.77	102	E	20.77
120	4	GAVL	104	G	4.64	103	G	4.64
121	14	LINAHMVCSFYTQR	105	H	66.77	104	L	66.77
122	7	KRQETVD	106	K	36	105	K	36
123	15	ANLSIYQVHFMGRKT	107	A	58.63	106	A	58.63
124	6	VILKEA	108	V	23.04	107	V	23.04
125	9	ATVNSIQDE	109	A	48.43	108	A	48.43
126	8	ASNKGRTL	110	A	24.42	109	A	24.42
127	4	GQNS	111	G	2.21	110	G	2.21
128	13	MAYSTFRILVHND	112	M	49.9	111	M	49.9
129	9	NTSDEARQK	113	N	22.45	112	N	22.45
130	8	PIATSRGV	114	P	17.68	113	P	17.68
131	9	MLVITSAQN	115	M	68.31	114	M	68.31
132	17	DGAENQISLCVFTKMHY	116	D	84.69	115	D	84.69
133	4	LVIM	117	L	58.48	116	L	58.48
134	4	KRQN	118	K	27.18	117	K	27.18
135	12	RQKITNEHSLDA	119	R	54.77	118	R	54.77
136	1	G	120	G	1	119	G	1
137	5	IMVLA	121	I	40.62	120	I	40.62
138	13	DERQALKSTINHM	122	D	55.56	121	D	55.56
139	15	KTLAINQDHESVRMF	123	K	89.8	122	K	89.8
140	8	ASGTFEVI	124	A	20.63	123	A	20.63
141	11	VSTACIKLRGM	125	V	69.46	124	V	69.46
142	15	TEAKINVDQLSGRHY	126	L	210.52	125	T	210.52
143	18	AKTVEILRFSGDNHCQMY	127	A	154.62	126	A	154.62
144	7	AVLISTG	128	V	84.46	127	A	84.46
145	11	VTISALKNEMC	129	T	78.5	128	V	78.5
146	14	EDKANVRITSMQGL	130	K	197.44	129	E	197.44
147	17	ETQARDNYGKHSIFVCL	131	K	133.23	130	E	133.23
148	5	LIVMF	132	L	46.07	131	L	46.07
149	16	KLHVRSTEQGDIAMNF	133	Q	144.01	132	K	144.01
151	16	AKNETDRQSVHCIGFL	134	A	208.37	133	A	208.37
152	17	LGINSMKQVAYHERFTD	135	M	165.73	134	L	165.73
153	7	SAKTRHQ	136	S	64.65	135	S	64.65
154	13	VKIQSRHETNDLA	137	K	93.13	136	V	93.13
155	16	PEKDGVAQINTLMHSR	138	P	111.25	137	P	111.25
156	8	CVIATLSM	139	C	57.63	138	C	57.63
157	15	SEQANKTDGRHLVIP	140	K	134.93	139	S	134.93
160	15	DTSHNGE.ARLKIQV	141	D	114.15	140	D	114.15
162	18	SKHQNPFITDRYEAGLM.	142	S	95.55	141	S	95.55
163	13	KENSDARTHQGIL	143	K	135.9	142	K	135.9
164	13	AQSEDKHNMTRVG	144	A	58.13	143	A	58.13
165	7	IVTYMKF	145	I	19.68	144	I	19.68

166	14	AETKSRVQGIMYCF	146	A	51.45	145	A	51.45
167	10	QASHRYENKT	147	Q	36.28	146	Q	36.28
168	5	VTICA	148	V	20.48	147	V	20.48
169	2	GA	149	G	21.1	148	G	21.1
170	11	TASGIKERVLC	150	T	77.64	149	T	77.64
171	5	IVLNC	151	I	29.03	150	I	29.03
172	3	STA	152	S	4.14	151	S	4.14
173	3	ASG	153	A	23.45	152	A	23.45
174	9	NGRASQK.D	154	N	33.54	153	N	33.54
175	11	S.NGAHTYQWF	155	S	25.47	154	S	25.47
176	6	DSENAI	156	D	19.57	155	D	19.57
177	18	EQSKPATHDNRMVIFGYL	157	E	208.45	156	E	208.45
178	18	TSEKFNQAYVIDRPLHMG	158	A	117.88	157	T	117.88
179	4	VITL	159	I	42.59	158	V	42.59
180	2	GA	160	G	1.12	159	G	1.12
181	15	KDNERQSLTAHGVXI	161	A	201.68	160	K	201.68
182	15	LYKIMDRFQEVSTAN	162	I	58.66	161	L	58.66
183	3	ILV	163	I	13.19	162	I	13.19
184	2	AS	164	A	18.46	163	A	18.46
185	15	EQKSDNHTVGAFLMR	165	E	126.37	164	E	126.37
186	5	ASCVG	166	A	7.29	165	A	7.29
187	7	MLIFVHT	167	M	17.22	166	M	17.22
188	11	DEQAGKSRNYT	168	E	104.63	167	D	104.63
189	11	KREILATVQCS	169	K	43.08	168	K	43.08
190	3	VAI	170	V	2.34	169	V	2.34
191	5	GTSKN	171	G	16.32	170	G	16.32
192	14	KNQSTPRHEALIDG	172	K	49.85	171	K	49.85
194	9	EDNSKTQHA	173	E	35.5	172	E	35.5
195	1	G	174	G	1	173	G	1
196	5	VSITA	175	V	8.22	174	V	8.22
197	3	IVM	176	I	6.16	175	I	6.16
198	4	TDSN	177	T	9.15	176	T	9.15
199	4	VILT	178	V	16.08	177	V	16.08
200	3	EDG	179	E	8.53	178	E	8.53
201	3	DEQ	180	D	14.63	179	D	14.63
203	4	GSAN	181	G	27.6	180	G	27.6
204	6	TNKRSQ	182	N	43.17	181	T	43.17
205	8	GTNRHSAK	183	G	57.61	182	G	57.61
206	9	LFMAITVRS	184	L	66.68	183	L	66.68
208	14	QGAENSTDHRKVLF	185	E	116.75	184	Q	116.75
209	9	DLTFNYISM	186	N	21.59	185	D	21.59
210	14	EQVDATSYNFLHGK	187	E	73.66	186	E	73.66
211	10	LVTIKQMFSC	188	L	29.09	187	L	29.09
212	12	DESNGKTAVRQI	189	S	80.46	188	D	80.46
213	13	VLFTIYAKRWMEH	190	V	48.74	189	V	48.74
214	3	VTA	191	V	8.18	190	V	8.18
215	8	EKDQVTGL	192	E	17	191	E	17
216	1	G	193	G	1	192	G	1
217	5	MLIYV	194	M	11.07	193	M	11.07
218	9	QRNMEAKSV	195	Q	16.21	194	Q	16.21

219	6	FIYLVW	196	F	11.76	195	F	11.76
220	5	DNKES	197	D	8.1	196	D	8.1
221	4	RKGN	198	R	10.2	197	R	10.2
222	2	GT	199	G	2.28	198	G	2.28
223	3	YFH	200	Y	10.29	199	Y	10.29
224	8	LITQASVM	201	I	64.79	200	L	64.79
225	4	SANT	202	S	6.66	201	S	6.66
226	9	PGSQARLHV	203	P	37.24	202	P	37.24
227	3	YHN	204	Y	1.65	203	Y	1.65
228	3	FML	205	F	14.07	204	F	14.07
229	8	IVSAMECT	206	I	74.41	205	I	74.41
230	3	NTS	207	N	23.26	206	N	23.26
231	7	KDNSQET	208	N	47.69	207	K	47.69
232	14	PSANTRQKGVHLMD	209	Q	120.55	208	P	120.55
233	8	EDQGTKAN	210	Q	67.06	209	E	67.06
234	11	TRKDSNAEGQH	211	N	59.62	210	T	59.62
235	7	GQMLARS	212	M	21.73	211	G	21.73
236	15	AEVSTLIQRKMDHNC	213	S	97.23	212	A	97.23
237	7	VACITSG	214	C	76.14	213	V	76.14
238	14	EVDQANGSHTLIRK	215	E	121.98	214	E	121.98
239	8	LYFMIHVA	216	L	60.47	215	L	60.47
240	9	EDASKNQTG	217	E	117.15	216	E	117.15
241	11	SDKENRQTHGA	218	H	115.53	217	S	115.53
242	6	PAVCTS	219	P	52.27	218	P	52.27
243	16	FYLKRVMATQWIHNCS	220	F	97.92	219	F	97.92
244	4	IVLF	221	I	55.44	220	I	55.44
245	5	LFIMV	222	L	7.75	221	L	7.75
246	6	LIVMCF	223	L	47.99	222	L	47.99
247	12	AVTYFHSNCLIM	224	V	78.01	223	A	78.01
248	9	DSENLGAQT	225	D	30.06	224	D	30.06
249	14	KSNLGRMQHDEAFT	226	K	30.61	225	K	30.61
250	4	KSER	227	K	3.43	226	K	3.43
251	3	IVL	228	V	58.18	227	I	58.18
252	8	SNTAGEQP	229	S	71.53	228	S	71.53
253	11	NTGSHLAIVQM	230	S	105.94	229	N	105.94
254	9	IVFLQMTAN	231	I	58.21	230	I	58.21
255	10	RKQANGSHTE	232	R	32.97	231	R	32.97
256	9	EDPASQNGT	233	E	77.17	232	E	77.17
257	5	MLIFV	234	M	69.38	233	M	69.38
258	5	LVIMA	235	L	31.35	234	L	31.35
259	8	PHSTNGAQ	236	S	22.91	235	P	22.91
260	6	VLIAMT	237	V	124.21	236	V	124.21
261	5	LAMVI	238	L	3.42	237	L	3.42
262	5	EQDNG	239	E	9.22	238	E	9.22
263	15	AKQSLEGPNDRHTIV	240	G	99.91	239	A	99.91
264	10	VISTALMQGC	241	V	72.55	240	V	72.55
265	11	AIVLMQSNHTF	242	A	46.79	241	A	46.79
266	9	KGQERNHPA	243	K	41.5	242	K	41.5
268	18	AQSTLENHMRVICGKFWY	244	S	130.87	243	A	130.87
269	7	GNQ.STA	245	G	63.73	244	G	63.73

272	8	KRANSQGH	246	R	79.05	245	K	79.05
273	13	PKASHNQRGETDC	247	P	51.8	246	P	51.8
274	5	LMIFV	248	L	27.71	247	L	27.71
275	9	LFVMIGAQC	249	L	81.46	248	L	81.46
276	4	IVLM	250	I	28.34	249	I	28.34
277	4	IVLM	251	I	51.66	250	I	51.66
278	3	ASC	252	A	21.28	251	A	21.28
279	5	EDSPG	253	E	14.91	252	E	14.91
280	6	DENASG	254	D	26.66	253	D	26.66
281	5	VIFLM	255	V	72	254	V	72
282	6	EDASTG	256	E	41.75	255	E	41.75
283	8	GESKANTD	257	G	14.73	256	G	14.73
284	5	EDQTG	258	E	12.87	257	E	12.87
285	3	AVG	259	A	4.17	258	A	4.17
286	5	LQVMT	260	L	7.37	259	L	7.37
287	6	ASPTGQ	261	A	33.76	260	A	33.76
288	9	TNGAMVSLI	262	T	14.39	261	T	14.39
289	2	LI	263	L	9.11	262	L	9.11
290	4	VIAL	264	V	19.47	263	V	19.47
291	6	VLITAY	265	V	32.63	264	V	32.63
292	2	NS	266	N	1.49	265	N	1.49
293	7	TKRNSHA	267	N	46.49	266	T	46.49
294	5	MILVA	268	M	44.84	267	M	44.84
295	3	RKQ	269	R	10.8	268	R	10.8
296	7	GASKLCI	270	G	9.81	269	G	9.81
297	12	ITGVSAQNLR.M	271	I	54.8	270	I	54.8
298	5	VFLIM	272	V	30.07	271	V	30.07
299	15	KNRTQDHSIAEGVML	273	K	74.67	272	K	74.67
300	7	VSAITCG	274	V	72.24	273	V	72.24
301	8	AVCSTLIN	275	C	67.07	274	A	67.07
302	5	ASGVC	276	A	12.25	275	A	12.25
303	5	VAITC	277	V	19.91	276	V	19.91
304	2	KN	278	K	2.63	277	K	2.63
305	3	ACS	279	A	6.52	278	A	6.52
306	4	PTNA	280	P	6.22	279	P	6.22
307	4	GEAS	281	G	8.61	280	G	8.61
308	2	FY	282	F	7.64	281	F	7.64
309	1	G	283	G	1	282	G	1
310	3	DEQ	284	D	16.16	283	D	16.16
311	5	RANCK	285	R	3.19	284	R	3.19
312	3	RQK	286	R	2	285	R	2
313	5	KLSQE	287	K	4.14	286	K	4.14
314	9	AENQDSRMT	288	A	44.53	287	A	44.53
315	9	MTAQILYVN	289	M	29.54	288	M	29.54
316	3	LMI	290	L	7.44	289	L	7.44
317	11	QENKRGATDHS	291	Q	78.01	290	Q	78.01
318	1	D	292	D	1	291	D	1
319	4	IMLV	293	I	32.02	292	I	32.02
320	5	ASGRC	294	A	11.18	293	A	11.18
321	7	TIVANCM	295	I	113.27	294	T	113.27

322	6	LSVIMA	296	L	14.01	295	L	14.01
323	4	TSVA	297	T	20.12	296	T	20.12
324	10	GAKNDSQHRE	298	K	38.33	297	G	38.33
325	4	GATS	299	G	30.7	298	G	30.7
326	14	TQEIVLSKHRMNDA	300	Q	112.4	299	T	112.4
328	8	VLAFTMPC	301	V	42.5	300	V	42.5
329	5	IVYAL	302	I	49.07	301	I	49.07
330	10	STDNAFMLKI	303	S	59.16	302	S	59.16
331	9	EDSNKPTQA	304	E	33.58	303	E	33.58
332	6	EDTAQK	305	E	56.53	304	E	56.53
333	11	IVLATRMKQSF	306	I	82.52	305	I	82.52
334	6	GNSAQD	307	G	13.45	306	G	13.45
335	17	MLRISFYKGANVTHCDQ	308	K	85.41	307	M	85.41
336	11	ETKDSGNAQRI	309	S	79.81	308	E	79.81
337	5	LFMIV	310	L	20.76	309	L	20.76
338	9	EKDRAQTNS	311	E	43.03	310	E	43.03
339	12	KNTDSAQGEHRM	312	G	116.02	311	K	116.02
340	8	ATVILSMC	313	A	116.37	312	A	116.37
341	12	TDSNEKARGQVP	314	T	128.26	313	T	128.26
343	12	LVIMTFSAEPQK	315	L	139.8	314	L	139.8
344	13	ESDKATNQGPCHY	316	E	179.61	315	E	179.61
345	16	DLQMASVNKFHECYTG	317	D	94.55	316	D	94.55
346	4	LFMV	318	L	10.11	317	L	10.11
347	1	G	319	G	1	318	G	1
348	16	QKTRSNDHEMIYLFVC	320	S	147.44	319	Q	147.44
349	7	ASCIGVT	321	A	25.16	320	A	25.16
350	13	KRNASGHEDQVTI	322	K	76.99	321	K	76.99
351	10	RKQSTLNMVI	323	R	74.57	322	R	74.57
352	5	VIAFL	324	I	97.31	323	V	97.31
353	14	VITESKHQNRLMAD	325	V	114.72	324	V	114.72
354	8	IVMLSATC	326	V	81.06	325	I	81.06
355	11	NTSDAQGKERV	327	T	89.14	326	N	89.14
356	3	KQR	328	K	2.42	327	K	2.42
357	3	DEN	329	E	47.07	328	D	47.07
358	14	TEDSNGKHAFLMYR	330	N	91.37	329	T	91.37
359	4	TASC	331	T	17.62	330	T	17.62
360	6	TVIMLK	332	T	36.4	331	T	36.4
361	5	IVLMF	333	I	54.85	332	I	54.85
362	4	IVLR	334	I	45.86	333	I	45.86
363	10	DESNGAQTHK	335	D	63.49	334	D	63.49
364	7	GETSDAN	336	G	18.77	335	G	18.77
365	20	VALDNGMSKI.EFTRQHYCP	337	E	157.89	336	V	157.89
366	12	GCNDH.EKPSAR	338	G	17.27	337	G	17.27
367	15	EDNSQKHV.AGRTCL	339	K	208.23	338	E	208.23
371	17	ETSKPAQGDNVCRH.YL	340	A	169.34	339	E	169.34
372	16	ADTENKVHSPRGQILY	341	T	233.92	340	A	233.92
373	18	AQDSNEVKTRHMIGLPFY	342	E	187.52	341	A	187.52
374	5	IVLRA	343	I	34.99	342	I	34.99
375	14	QASEDKNHVTRLMG	344	N	183.54	343	Q	183.54
376	14	GAHDNEKSTQMRIV	345	A	138.14	344	G	138.14

377	4	RYHK	346	R	3.43	345	R	3.43
378	10	VCILSEAQTK	347	I	90.13	346	V	90.13
379	15	ANESLGKQHTDVCIR	348	T	186.96	347	A	186.96
380	14	QNVSLITEFGAHMK	349	Q	68.26	348	Q	68.26
381	3	ILV	350	I	21.99	349	I	21.99
382	9	RKNMHEAQD	351	R	56.94	350	R	56.94
383	14	QAKSVTGNRHMILE	352	A	151.18	351	Q	151.18
384	10	QEKRNSMHAL	353	Q	25.16	352	Q	25.16
385	11	IMLSVRAYWCH	354	M	63.61	353	I	63.61
386	11	EAKGDQSLNVP	355	E	96.62	354	E	96.62
387	15	ENDTKAVSQGLRHIM	356	E	133.23	355	E	133.23
388	8	ASTLINVE	357	T	58.68	356	A	58.68
389	8	TDESNKAL	358	T	79.35	357	T	79.35
390	6	SDEVLN	359	S	10.35	358	S	10.35
391	8	DEKSATQN	360	D	46.25	359	D	46.25
392	3	YFW	361	Y	9.78	360	Y	9.78
393	3	DEA	362	D	4.47	361	D	4.47
394	7	RKSTQAM	363	R	86.83	362	R	86.83
395	6	EDKQSN	364	E	10	363	E	10
396	6	KRLTNS	365	K	6.8	364	K	6.8
397	4	LIMA	366	L	7.43	365	L	7.43
398	11	QKTNMREASLH	367	Q	18.84	366	Q	18.84
399	3	EKN	368	E	2.41	367	E	2.41
400	1	R	369	R	1	368	R	1
401	8	VLIMQKSH	370	V	20.68	369	V	20.68
402	3	AGS	371	A	3.38	370	A	3.38
403	6	KNYHAR	372	K	6.14	371	K	6.14
404	3	LIM	373	L	10.31	372	L	10.31
405	9	ASVTGIRNQ	374	A	72.64	373	A	72.64
406	8	GNQSDVKC	375	G	5.94	374	G	5.94
407	1	G	376	G	1	375	G	1
408	2	VI	377	V	7.33	376	V	7.33
409	5	ACGSP	378	A	5.1	377	A	5.1
410	7	VILKQMR	379	V	14.69	378	V	14.69
411	4	IVLM	380	I	38.1	379	I	38.1
412	12	KRHYQNAESGVM	381	K	50.98	380	K	50.98
413	5	VAISL	382	V	12.5	381	V	12.5
414	1	G	383	G	1	382	G	1
415	4	AGSD	384	A	8.56	383	A	8.56
416	10	APVSGMNTIQ	385	A	35.43	384	A	35.43
417	2	TS	386	T	14.64	385	T	14.64
418	2	ED	387	E	4.19	386	E	4.19
419	9	VTILAMPSF	388	V	57.72	387	V	57.72
420	3	EAD	389	E	8.98	388	E	8.98
421	6	MLVAQI	390	M	22.23	389	M	22.23
422	6	KQRSNE	391	K	4.83	390	K	4.83
423	6	EDANTL	392	E	9.5	391	E	9.5
424	11	KRMLEQISAVT	393	K	31.62	392	K	31.62
425	4	KRQT	394	K	7.07	393	K	7.07
426	10	AHLDYMFSGT	395	A	29.4	394	A	29.4

427	4	RLKH	396	R	3.83	395	R	3.83
428	6	VILFMY	397	V	20.02	396	V	20.02
429	3	EDQ	398	E	15.79	397	E	15.79
430	2	DH	399	D	1.72	398	D	1.72
431	3	ATS	400	A	2.94	399	A	2.94
432	7	LVQIMFK	401	L	10	400	L	10
433	9	HRNASLCQD	402	H	40.95	401	H	40.95
434	4	ANST	403	A	12.51	402	A	12.51
435	3	TAS	404	T	8.04	403	T	8.04
436	6	RKLIAQ	405	R	19.53	404	R	19.53
437	2	AS	406	A	5.19	405	A	5.19
438	2	AG	407	A	3.87	406	A	3.87
439	8	VIAMLSTK	408	V	25.28	407	V	25.28
440	7	EQDATLG	409	E	27.93	408	E	27.93
441	1	E	410	E	1	409	E	1
442	1	G	411	G	1	410	G	1
443	7	VILMYFT	412	I	39.26	411	V	39.26
444	5	VLIGS	413	V	21.73	412	V	21.73
445	8	APISVGTL	414	A	72.05	413	A	72.05
446	1	G	415	G	1	414	G	1
447	1	G	416	G	1	415	G	1
448	2	GE	417	G	2.81	416	G	2.81
449	9	VTLIASCYK	418	V	64.72	417	V	64.72
450	9	ATSGCIVMP	419	A	40.65	418	A	40.65
451	5	LYFIM	420	L	41.46	419	L	41.46
452	8	ILVAMFYT	421	I	106.24	420	I	106.24
453	16	RQSNHTDKELYMAGIW	422	R	67.18	421	R	67.18
454	10	VAICSLTGMF	423	A	80.7	422	V	80.7
455	17	AYISTVLKQREGMFHCN	424	Q	149.47	423	A	149.47
456	18	S.NPHQTEKALDCGVMRI	425	K	203.68	424	S	203.68
458	19	KPTASVQDYEIGCLNRHMF	426	A	173.35	425	K	173.35
459	9	LTIVAS.MG	427	L	70.18	426	L	70.18
460	18	ALESTKHDQVGFRN.ICP	428	D	191.2	427	A	191.2
462	12	.GSDNEATKQVL	429	S	234.92	-	-	-
463	-	-	-	-	-	428	D	211.78
464	19	LEFI.AVNDSWTGKHMYQP	430	L	159.12	429	L	159.12
465	19	RLK.IEDHTVNAQCSGFPM	431	K	172.6	430	R	172.6
477	17	GK.VNLYESTPADFQIM	432	G	111.56	431	G	111.56
478	18	QLATDEV.HSNGKIYWFP	433	D	148.67	432	Q	148.67
479	17	NETKIHQDSLVA.GPCY	434	N	150.01	433	N	150.01
480	17	EGKAPDIQYVSNHTRLF	435	D	146.41	434	E	146.41
481	5	DEGAC	436	D	27.85	435	D	27.85
482	12	QEVIDAFTRKMN	437	Q	53.78	436	Q	53.78
483	15	NAEQLRTKDGVSYIC	438	N	146.33	437	N	146.33
484	14	VTIALYHQMFRSWK	439	M	90.05	438	V	90.05
485	2	GA	440	G	4.35	439	G	4.35
486	14	IARVFGYTLWCSKM	441	I	87.04	440	I	87.04
487	16	KNREAQDMLVTSHGIY	442	N	160.55	441	K	160.55
488	6	VILTSA	443	I	35.67	442	V	35.67
489	7	AVILFMT	444	L	64.13	443	A	64.13

490	16	LKIFRMEQVAYTSCGN	445	R	122.42	444	L	122.42
491	16	RVKESAINLQDHFTYM	446	R	79.93	445	R	79.93
492	5	ASGTV	447	A	38.48	446	A	38.48
493	9	MLICVFATS	448	I	62.2	447	M	62.2
494	18	ESTKQRPALGDVICYFMN	449	E	74.39	448	E	74.39
495	14	AESDYTVQKGRHMF	450	S	71.62	449	A	71.62
496	2	PL	451	P	2.12	450	P	2.12
497	7	LVIAMTC	452	M	74.54	451	L	74.54
498	12	RKASTYMWQFHN	453	R	48.94	452	R	48.94
499	17	QTRMEWVIGHLYCASKN	454	Q	35.56	453	Q	35.56
500	3	ILM	455	I	8.15	454	I	8.15
501	8	VATSIMCG	456	V	51.74	455	V	51.74
502	18	LFHAVEIKTYNQSDRGCM	457	T	194.97	456	L	194.97
503	1	N	458	N	1	457	N	1
504	5	CSAGT	459	A	37.54	458	C	37.54
505	2	GA	460	G	1.11	459	G	1.11
506	19	ELKFYIDQAVGSMTHWRCN	461	Y	142.12	460	E	142.12
508	7	ENSDQAK	462	E	32.73	461	E	32.73
509	10	PGASENDTRK	463	A	46.27	462	P	46.27
510	13	SGAVDYNTKEQIP	464	S	50.97	463	S	50.97
512	10	VIKCLYEDHW	465	V	42.35	464	V	42.35
513	4	VIYA	466	V	68.85	465	V	68.85
514	9	AVCIFLSTY	467	V	83.81	466	A	83.81
515	16	NEQDGSAHIRMVKTYC	468	N	101.41	467	N	101.41
516	14	TKRQANESHMIYDG	469	K	126.5	468	T	126.5
517	7	VLIAMQC	470	V	44.88	469	V	44.88
518	14	KRLMAIQESVTGHY	471	A	115.7	470	K	115.7
519	17	GNHSAE.KQTDRLMCVI	472	E	185.86	471	G	185.86
525	17	G.RAKSNTEVHDQPLMC	473	H	109.1	472	G	109.1
526	15	DPASEKQ.TVGNCRH	474	K	183.88	473	D	183.88
527	20	GAVSLPKERYITDQFNHW.M	475	D	184.4	474	G	184.4
528	13	NGSTHDAERKIPV	476	N	86.73	475	N	86.73
529	18	YHVEIFTQLMWNSRDKAG	477	Y	168.92	476	Y	168.92
530	2	GI	478	G	1.08	477	G	1.08
531	4	YLFW	479	F	55.13	478	Y	55.13
532	2	ND	480	N	41.39	479	N	41.39
533	5	ATVSC	481	A	13.62	480	A	13.62
534	19	AQLYSEKGRDMNIVHTCFW	482	A	88.43	481	A	88.43
535	15	TNRDKASHVELIYFC	483	T	140.73	482	T	140.73
536	12	EGDNMKLFHRSA	484	G	82.24	483	E	82.24
537	13	EVAQTKRNIDSGC	485	E	150.13	484	E	150.13
538	9	YWFKVCHIL	486	Y	59.12	485	Y	59.12
539	13	GEVTKICADQNSM	487	G	78.19	486	G	78.19
541	9	NDESCKYQA	488	D	56.22	487	N	56.22
542	5	MLAFV	489	M	41.1	488	M	41.1
543	14	ILMVFAHYKRQSTC	490	V	124.36	489	I	124.36
544	15	DAEKSNTGQRVIYPH	491	E	164.2	490	D	164.2
545	16	MAQTLVNESKFRDGWH	492	M	113.07	491	M	113.07
546	3	GKR	493	G	1.66	492	G	1.66
547	2	IV	494	I	30.94	493	I	30.94

548	9	LAVITPNMS	495	L	52.16	494	L	52.16
549	3	DVE	496	D	1.4	495	D	1.4
550	2	PA	497	P	2.33	496	P	2.33
551	9	TVALKFCMS	498	T	58.94	497	T	58.94
552	4	KQMR	499	K	5.41	498	K	5.41
553	1	V	500	V	1	499	V	1
554	10	TSVEANGCIL	501	T	40.94	500	T	40.94
555	4	RKLT	502	R	2.25	501	R	2.25
556	14	SCTIVMNAYLFQHG	503	M	65.83	502	S	65.83
557	5	AVTGS	504	A	10.41	503	A	10.41
558	5	LIFMV	505	L	20.65	504	L	20.65
559	13	QELDKGRVITASM	506	Q	26.5	505	Q	26.5
560	11	YNSHDKAFGLT	507	N	29.53	506	Y	29.53
561	2	AS	508	A	6.79	507	A	6.79
562	8	ASCVGTFI	509	A	56.94	508	A	56.94
563	1	S	510	S	1	509	S	1
564	4	VIAL	511	V	46.6	510	V	46.6
565	5	ASGVC	512	A	50.17	511	A	50.17
566	7	GASNTKI	513	S	87.03	512	G	87.03
567	9	LMTSQVIAN	514	L	35.96	513	L	35.96
568	7	MFLIVYA	515	M	55.13	514	M	55.13
569	5	ILVMT	516	L	24.92	515	I	24.92
570	6	TIASML	517	T	4.29	516	T	4.29
571	4	TALS	518	T	7.62	517	T	7.62
572	8	ENDSQAGM	519	E	25.31	518	E	25.31
573	6	CATVSG	520	C	50.52	519	C	50.52
574	9	MVLATICSG	521	M	64.93	520	M	64.93
575	3	VIL	522	V	68.71	521	V	68.71
576	11	TAVNHSYGCIF	523	A	69.18	522	T	69.18
577	7	DNSEATK	524	D	82.05	523	D	82.05
578	17	LKIERVHAQSDTMFPYN	525	L	131.41	524	L	131.41
579	11	PLKVMQAESNG	526	P	44.71	525	P	44.71
580	11	KEDRSAQNTVG	527	K	85.39	526	K	85.39
581	12	NKSEPDQRGAVT	528	K	143.47	527	N	143.47
582	13	DENKV.ASTGPQR	529	E	181.95	528	D	181.95
583	15	AKGS.PENDTVQIMH	530	E	219.66	529	A	219.66
584	17	APSG.ETNKDVLQIMHC	531	G	208.83	530	A	208.83
585	-	-	-	-	-	531	D	234.92
586	19	L.ADPTMVGFNSQHIKERC	532	V	202.64	532	L	202.64
587	-	-	-	-	-	533	G	206.08
590	-	-	-	-	-	534	A	234.92
594	-	-	-	-	-	535	A	234.92
595	12	.DGPAMLVRSEN	533	G	234.92	-	-	-
596	14	.GMADSPHLNYQEV	534	A	234.92	-	-	-
597	14	G.PASDEMQYHRNT	535	G	96.64	536	G	96.64
598	15	G.ADQSEMNPYHVTI	536	D	118.38	537	G	118.38
599	14	M.GTNASEDQPHYF	537	M	114.17	538	M	114.17
600	15	G.DMPSFAHNEVQYT	538	G	101.2	539	G	101.2
601	13	GD.APNMHSYQFE	539	G	98.31	540	G	98.31
602	15	MG.PTAYSVDQNEHF	540	M	99.69	541	M	99.69

603	13	G.ADPYEMFQNHS	541	G	94.44	542	G	94.44
604	13	GD.PMFSNAHEQY	542	G	95.87	543	G	95.87
605	14	MPQD.GYAVNSHFT	543	M	91.66	544	M	91.66
606	13	GDS.MANPEHIYQ	544	G	108.76	545	G	108.76
607	14	GYM.PFDEHNSARL	545	G	101.25	546	G	101.25
608	14	M.GPALFDQHIESY	546	M	73.6	547	M	73.6
609	15	MDG.PRHNEQYFASI	547	G	116.66	548	G	116.66
610	11	.FGYMSDVNHE	548	G	88.4	549	G	88.4
611	9	.MFYDHGIP	549	M	60.41	550	M	60.41
612	9	.MYGHDIPV	550	M	234.92	551	M	234.92

APPENDIX B

BCYE

10 g yeast extract (MO BIO, Cat. No.12110)

1 g ACES (Sigma, Cat. No. A9758)

1g α-ketoglutaric acid (Sigma, Cat. No. K1750)

1.5 g charcoal (Sigma, Cat. No. C9157)

15 g agar

ddH_2O to 1 L

pH adjusted to 6.6 with 6N KOH

Dissolve and autoclave for 15 min at 121˚C, Allow medium to cool to ~ 55°C and then add:

0.4 g L-cysteine (dissolved in a 4 mL volume, pH adjusted to 6.6 with 6N KOH, and filter sterilized)

1 mL, 25% (wt/vol) iron pyrophosphate (filter sterilized and stored in the dark at 4 °C)

BYE broth

The formulation and preparation instructions for BYE are the same as BCYE, with the exception of adding charcoal and agar.

Lysogeny broth (LB) medium

5 g yeast extract

10 g tryptone (Sigma, Cat. No. T7293)

10 g NaCl (Sigma, Cat. No. S7653)

15 g agar (for plates only)

ddH_2O up to 1L

Dissolve and autoclave for 15 min at 121˚C.

Super optimal broth with catabolite repression (SOC) medium

For 100 mL

2 g tryptone

0.5 g yeast extract

60 mg NaCl

2 mg KCl

0.2 g $MgCl_2$

0.12 g $MgSO_4$

100 mL ddH_2O

Autoclave the solution at 121 °C, let cool and then add 4 mL of 20% glucose

20 mM glucose.

Freezing medium

8 g nutrient broth

20% v/v glycerol (Sigma, Cat. No. G6279)

ddH_2O up to 1L

Dissolve and autoclave for 15 min at 121˚C.

Synthetic defined (SD) media

7 g Yeast nitrogen base without amino acids (Sigma, Cat. No.Y0626)

20 g glucose (Sigma, Cat. No.G8270)

15 g agar (for plates only)

1 g of the appropriate drop out mix lacking one or more specific nutrients

Dissolve and autoclave for 15 min at 121˚C.

Drop out mix

Grams	Nutrient	SD/-Leu	SD/-Trp	DDO	QDO	Sigma Cat. No.
2	L-Adenine hemisulfate salt	+	+	+	-	A-5131
2	L-Arginine HCl	+	+	+	+	H-8125
2	L-Histidine HCl monohydrate	+	+	+	-	I-2752
2	L-Isoleucine	+	+	+	+	L-8000
2	L-Leucine	-	+	-	-	L-5626
2	L-Lysine HCl	+	+	+	+	M-9625
2	L-Methionine	+	+	+	+	P-2126

3	L-Phenylalanine	+	+	+	+	T-8625
2	L-Threonine	+	+	+	+	T-0254
2	L-Serine	+	+	+	+	S4500
3	L-Tryptophan	+	-	-	-	T0254
2	L-Tyrosine	+	+	+	+	T-3754
1.2	L-Uracil	+	+	+	+	U-0750
9	L-Valine	+	+	+	+	V-0500

Aureobasidin A Stock Solution

Dissolve 1 mg Aureobasidin A (Clontech, Cat. No. 630466) in 2 ml of absolute ethanol for a stock concentration of
500 µg/ml. Store at 4°C.

X-α-Gal Stock Solution

Dissolve 100 mg of X-α-Gal (Clontech, Cat. No. 630463) in 5 mL of dimethylformamide (DMF). Store X-a-Gal solutions at –20°C in the dark.

DDO/X/A agar plates

Double dropout media (DDO) containing 40 µg/ml X-α-Gal and 200 ng/ml aureobasidin A is prepared as follows:
Prepare 500 ml DDO (SD–Leu/–Trp) with agar, autoclave for 15 min at 121˚C, and cool to 55–60°C in a waterbath.
Add 1 ml of X-a-Gal Stock Solution.
Add 200 µl of Aureobasidin A stock solution.
Mix, pour immediately, and allow to dry.

QDO/X/A agar plates

Quadruple dropout (QDO) media containing 40 µg/ml X-a-Gal and 200 ng/ml aureobasidin A is prepared as follows:
Prepare 500 ml QDO (SD/–Ade/–His/–Leu/–Trp) with agar, autoclave for 15 min at 121˚C, and cool to 55–60 °C in a waterbath.

Add 200 µl of Aureobasidin A stock solution.

Add 1 ml of X-a-Gal Stock Solution.

Mix, pour immediately, and allow to dry.

QDO/A broth

The formulation and preparation instructions for QDO/A broth are the same as QDO/X/A. whit the exception of adding agar and X-a-Gal.

YPD and YPDA media

10 g yeast extract

20 g peptone (BD, Cat. No. 211677)

20 g glucose

For adenine-supplemented YPD (YPDA) medium:

Add 40 mg of adenine hemisulfate

15 g agar (for plates only)

ddH_2O up to 1L

Dissolve and autoclave for 15 min at 121˚C.

2X YPDA (100 mL)

2 g yeast extract

4 g peptone

4 g glucose

8 mg adenine hemisulfate

100 mL ddH_2O

50X TAE buffer

242 g Tris base (Life technologies, Cat. No.15504-020)

57.1 ml glacial acetic acid (Fisher scientific, Cat. No. A38-212)

100 ml 0.5M EDTA, pH 8 (Sigma, Cat. No. E9884)

ddH_2O to 1 liter

1X TAE buffer

10 ml of 50X TAE buffer

ddH_2O to 500 ml

10X TE buffer

1M Tris-HCl (Sigma, Cat. No. T5941)

0.01M EDTA

pH 8

1X TE buffer

100 ml of 10X TE buffer

ddH_20 to 1L ml

1X TE 0.5% SDS buffer

100 ml of 10X TE buffer

ddH_20 to 1L ml

5 g SDS (Sigma, Cat. No. L3771)

PEG/LiAc solution (polyethylene glycol/lithium acetate)

Prepare the following stock solutions:

- 50% PEG 3350 (Polyethylene glycol; Sigma Cat No. P-3640) in deionized H_2O; if necessary, warm solution to 50°C to help the PEG go into solution.
- 10X TE buffer: 0.1 M Tris-HCl, 10 mM EDTA, pH 7.5. Autoclave for 15 min at 121˚C.
- 10X LiAc: 1 M lithium acetate (Sigma Cat No. L-6883) Adjust to pH 7.5 with dilute acetic acid. Autoclave for 15 min at 121˚C.

To prepare 100 mL PEG/LiAc solution:

80 ml of 50% PEG

10 ml of 10X TE

10 ml of 10X LiAc

Mix well and autoclave for 15 min at 121˚C.

1X TE/1X LiAc

Prepare fresh on a sterile 1.5 mL Eppendorf tube

100 µl 10X TE

100 µl 10X LiAc

800 µl sterile ddH_20

Ice-cold lysis buffer for Co-IP

25 mM Tris-HCl pH 7.5

15 mM EGTA

1 mM EDTA

150 mM NaCl

0.1 % Triton X-100

0.2 10% glycerol

0.3 1mM DTT

0.4 mM PMSF

PNP-α-Gal Solution

100 mM (p-nitrophenyl α-d-Galactopyranoside; Sigma Cat No. N0877) in deionized H2O

For 10 ml, dissolve 301.3 mg of PNP-α-Gal in 10 ml of deionized H2O.

Prepare solution fresh before each use.

Keep the p-nitrophenyl α-d-Galactopyranoside solid anhydrous. Store in a desiccator at -20 °C.

10X Stop Solution

1 M Na2CO3 in deionized H2O (Sigma Cat No. S7795)

1X NaOAc

0.5 M sodium acetate, pH 4.5 (Sigma Cat No. S7545)

Assay Buffer for alpha-galactose activity quantification

Prepare Assay Buffer fresh, before each use, by combining 2 volumes 1X NaOAc Buffer with 1 volume PNP-α-Gal Solution [2:1 (v/v) ratio]. Mix well.

Agarose/EtBr gel

50 mL 1X TAE buffer

0.5 (1% gel) or 1 g (2% gel) agarose

Dissolve by heating in the microwave for 30 to 60 seconds

Add 50 µl Ethidium bromide

Stacking gel

For 1 minigel

970 µl ddH_20

420 µl of 0.5 M Tris-HCl 0.4% SDS pH 6.8

283 µl of 30% Acrylamide:Bis-Acrylamide Solution [29:1]

11.7 µl of 10% ammonium persulfate (APS) (stock must be prepared fresh)

3.3 µl of N,N,N´,N´-tetramethylethylenediamine (TEMED) (BioShop Canada Inc. Cat. No. TEM001).

12% resolving gel

For 1 minigel

2.8 mL ddH_20

2 mL of 1.5 M Tris-HCl 0.4% SDS pH 8.8

3.2 mL of 30% Acrylamide:Bis-Acrylamide Solution [29:1] (BioRad)

50 µl of 10% APS

12 µl of TEMED

10 X Ponceau S

For 100 mL

2 g Ponceau S

30 g trichloroacetic acid

30 g sulfosalicylic acid

ddH20 to 100 mL

5X SDS-PAGE gel loading buffer

0.25M Tris-HCl, pH 6.8

15% SDS

50% glycerol

25% β-mercaptoethanol

0.01% bromophenol blue

10X running buffer

25 mM Tris-base

192 mM glycine

0.1% SDS

1X running buffer

100 ml 10X running buffer

ddH20 to 1L

10X Transfer buffer

For 4 L

121.1 g Tris base

576 g glycine

Bring up the volume to 4 L with ddH2O

1X transfer buffer

For 1 L

700 mL cold ddH2O

100 mL 10x Transfer buffer

200 mL methanol

1X phosphate-buffered saline (PBS)

140 mM NaCl

3 mM KCl

8 mM Na_2HPO_4

1 mM KH_2PO_4

Blocking solution

10 mL TTBS

0.2 g skim milk

0.2 mL 10% BSA

10X TBS:

For 2 L

121.1 g Tris base

175.3 g NaCl

37.22 EDTA

Add 2 L ddH2O

Adjust the pH to 7.3 with concentrated HCl

1X TBS

100 mL 10X TBS

ddH20 to 1L

1X TTBS buffer

100 mL 10X TBS

5 mL Tween 20

ddH20 to 1L

AP buffer

For 100 mL

80 mL ddH20

10 mL 1M Tris-base pH 9.5

5 mL 1.5 M NaCl

5 mL 1 M $MgCl_2$

NBT stock solution

75 mg/ml NBT (Nitrotetrazolium Blue chloride, Sigma, Cat. No. N6876) in 70% dimethylformamide and 30 % ddH20

Developing solution

For 10 mL

10 mL AP buffer

0.0016 g BCIP (5-Bromo-4-chloro-3-indolyl phosphate disodium salt, Sigma, Cat. No.B1026)

44 µl NBT