

Dissertation zur Erlangung des Doktorgrades der Fakultät für Chemie und  
Pharmazie der Ludwig-Maximilians-Universität München

Protein Folding in Archaea:  
Analysis of the co-existing Group I and Group II  
chaperonins in *M. mazei*



Angela Maria Hirtreiter

aus Straubing

München, 2006

## Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. F.-Ulrich Hartl betreut.

## Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 15.10.2006

.....

(Angela Hirtreiter)

Dissertation eingereicht am	15.10.2006
1. Gutachter	Prof. F.Ulrich Hartl
2. Gutachter	PD Dr. Konstanze F. Winklhofer
Mündliche Prüfung am	31.01.2007

---

I.	Summary .....	I-1
II.	Introduction.....	II-1
II.1	Proteins.....	II-1
II.1.1	Protein structure.....	II-1
II.1.2	Protein folding: from primary to quaternary structure.....	II-1
II.1.3	The protein folding mechanism.....	II-2
II.2	Protein folding in vivo.....	II-3
II.2.1	Protein folding in the cellular environment.....	II-3
II.2.2	Protein folding upon de novo synthesis .....	II-5
II.3	Molecular chaperones .....	II-5
II.3.1	The chaperone system .....	II-6
II.3.2	Ribosome associated chaperones .....	II-7
II.3.3	The Hsp70s .....	II-8
II.3.4	Prefoldin / GimC.....	II-8
II.3.5	The chaperonins .....	II-9
II.4	The Group I chaperonin system .....	II-11
II.4.1	Structure and function of the Group I chaperonin system GroEL/GroES ..	II-11
II.4.2	Mechanism of the GroEL/GroES system .....	II-12
II.4.3	Substrates of the GroEL/GroES system .....	II-13
II.5	The Group II chaperonin system .....	II-14
II.5.1	Structure and function of Group II chaperonins.....	II-14
II.5.2	Mechanism of Group II chaperonins .....	II-16
II.5.3	Substrates of Group II chaperonins.....	II-17
II.6	Archaea .....	II-18
II.6.1	The Methanogens .....	II-20
II.6.2	The genus Methanosarcina.....	II-21
II.6.3	Methanosarcina mazei Gö1 .....	II-23
II.7	Aim of the project .....	II-24
III.	Materials and Methods .....	III-25
III.1	Materials .....	III-25
III.1.1	Chemicals.....	III-25
III.1.2	Buffers and Media.....	III-27

---

III.2	Anaerobic cultivation.....	III-30
III.2.1	Set up of an anaerobic cultivation system .....	III-30
III.2.2	Anaerobic cultivation.....	III-31
III.2.3	Sterilisation.....	III-33
III.2.4	Light microscopy .....	III-33
III.3	Molecular biochemical methods .....	III-34
III.3.1	Preparation and transformation of E. coli.....	III-34
III.3.2	DNA analytical methods .....	III-34
III.3.3	DNA restriction/ligation methods .....	III-35
III.3.4	Plasmid purification.....	III-35
III.4	Protein biochemical methods .....	III-36
III.4.1	Protein expression and purification .....	III-36
III.4.2	Protein analytical methods .....	III-38
III.4.3	Polyclonal antibodies.....	III-42
III.4.4	Affinity purification of antibodies .....	III-42
III.4.5	Western blotting .....	III-43
III.4.6	Preparation of antibody beads.....	III-43
III.4.7	Analytical gel filtration.....	III-43
III.5	Functional analysis methods.....	III-44
III.5.1	Co-immunoprecipitations.....	III-44
III.5.2	Substrate release experiments .....	III-45
III.5.3	PK digestion .....	III-45
III.6	Biophysical methods.....	III-45
III.6.1	Mass spectrometry of protein spots from 2D PAGE .....	III-45
III.6.2	Coupled liquid chromatography – mass spectrometry system (LC-MS/MS).... .....	III-46
III.7	Bioinformatic methods .....	III-49
III.7.1	Sequence data analysis .....	III-49
III.7.2	Structure data analysis .....	III-50
IV.	Results .....	IV-51

---

IV.1	Production of recombinant proteins and polyclonal antibodies .....	IV-51
IV.2	Both chaperonins and their cofactors are expressed at similar levels and as oligomeric proteins .....	IV-53
IV.2.1	Group II chaperonin: The MmThs consists of three different subunits ....	IV-53
IV.2.2	Group II chaperonin: the supposed co-chaperon prefoldin .....	IV-56
IV.2.3	Group I chaperonin MmGroEL and the Co-chaperone MmGroES .....	IV-59
IV.2.4	Both chaperonins are moderately induced under heat shock conditions	IV-61
IV.3	Substrate identification .....	IV-62
IV.3.1	Co-precipitation of chaperonin-substrate complexes .....	IV-62
IV.3.2	The MmGroEL/GroES complex is stable during the experimental procedure . .....	IV-64
IV.3.3	GroEL/GroES complexes contain mainly encapsulated cis-ring bound substrates .....	IV-65
IV.3.4	The co-precipitated proteins can be released from the corresponding chaperonin in a nucleotide dependent manner .....	IV-67
IV.3.5	Chaperonin-substrate complexes captured under native conditions differ from in vitro reconstituted Chaperonin-substrate complexes .....	IV-71
IV.3.6	Substrate Identification .....	IV-73
IV.3.7	Overview of the Proteomic Data Set .....	IV-81
IV.4	Analysis of the role of chaperonins in the network of <i>M. mazei</i> .....	IV-105
V.	Discussion .....	V-107
V.1	The co-existence of Group I and Group II chaperonins .....	V-107
V.2	Substrates of group I and group II chaperonins .....	V-108
V.2.1	Substrate identification techniques .....	V-109
V.2.2	Substrate sets of MmThs and MmGroEL .....	V-110
V.2.3	Phylogenetic origin of the substrate proteins .....	V-111
V.2.4	Structural features of substrate proteins .....	V-113
VI.	References .....	VI-123
VII.	Appenices .....	132
VII.1	Supplementary Figures .....	132

---

VII.2	Supplementary Tables .....	140
VII.3	Abbreviations .....	VII-186
VII.4	Publications, Poster und Vorträge.....	VII-189
VII.5	Curriculum vitae .....	VII-190
VII.6	Danksagung.....	VII-191

## I. Summary

The correct three-dimensional structure of a protein is essential for its biological function. In the cell, the adoption of this so-called native state must proceed in a rapid and reliable manner. *In vivo*, efficient folding of newly synthesized proteins relies on the assistance of molecular chaperones, which act to prevent aggregation and promote folding.

Chaperonins are a specific class of barrel-shaped chaperones, present in almost all organisms. Newly synthesized proteins encapsulated by the chaperonin can attain their native structure unimpaired by aggregation during repeated cycles of ATP-dependent binding and release. Chaperonins are generally divided into two groups. Group I chaperonins, such as the barrel-shaped GroEL oligomer, are found predominantly in *bacteria* and cooperate with cofactors of the Hsp10 family (*i.e.* GroES). The Group II chaperonins, on the other hand, do not require a Hsp10- cofactor and are found in the eukaryotic cytosol and in *archaea*.

The function of GroEL is understood in great detail and the substrate interaction proteome has been recently identified. In contrast, our knowledge about the natural substrates of Group II chaperonins is deficient and as a consequence, mechanistical studies on Group II chaperonins have been limited to using the eukaryotic model substrates actin and tubulin as well as heterologous model substrates.

In the present study, the complete substrate spectrum of a Group II chaperonin, the thermosome (Ths) of the mesophilic archaeon *Methanosarcina mazei* (*M. mazei*), was analysed for the first time. In addition, the unique coexistence of both the group I and the group II chaperonins in *M. mazei*, which was confirmed in the initial part of the study, provided the opportunity to obtain new insights into how the substrate selection differs between the two chaperonin groups. For these purposes, the chaperonin substrates were isolated by immunoprecipitation of the chaperonin-substrate complexes and identified by liquid chromatography coupled mass spectrometry (LC-MS) using three different approaches: LC-MS after separation of the proteins (i) by classical 2D-PAGE, (ii) by difference gel electrophoresis (Ettan DIGE) and (iii) by 1D-PAGE.

---

Analysis of substrates of both the thermosome (*MmThs*) and GroEL/GroES (*MmGroEL*, *MmGroES*) of *M. mazei* revealed that each chaperonin handles a defined set of substrates, and both chaperonins contribute to the folding of ~17% of the proteins in the archaeal cytosol. Bioinformatic analysis revealed that the chaperonin specificity is governed by specific properties of substrate proteins.

The study demonstrates that the chaperonin-substrate protein interaction is determined by a complex set of structural features of the client protein such as hydrophobicity, net charge and size of the substrate. Substrate selection is likely to be influenced by the initial contacts between the unfolded substrate protein and the accepting apical domains of the chaperonins. The hydrophobicity of the apical domains of GroEL can explain the preference for hydrophobic proteins, whereas the positively charged nature of the apical domains of Ths may explain the preferential interaction of *MmThs* with negatively charged proteins.

The encapsulation mechanism of both chaperonins is expected to limit the size of substrate proteins. As expected, our study revealed that the majority of the *MmGroEL* substrates does not exceed 60 kDa, which corresponds to the theoretical size limit imposed by the folding cage. The upper size limit of *MmThs* substrates, however, is expanded to ~80 kDa. This upshift in size limit might be due to a more flexible closing mechanism of the Group II chaperonin. An aperture-like closure of the Ths cavity by helical protrusions of its apical domains may allow the cage volume of the chaperonin to increase.

Previous work on the *E. coli* GroEL substrate proteome revealed that certain domain folds, such as the c-class domain fold in the Triosephosphate isomerase (TIM)  $\beta/\alpha_8$  barrel, increase the tendency of a protein to interact with the Group I chaperonin. In this study, a prevalent interaction of the TIM $\beta/\alpha_8$  barrel with the Group I chaperonin could also be confirmed for *M. mazei*. In addition, a significant fraction of *MmGroEL* interactors share similar domain folds of the c-class of alpha helical elements and mainly parallel beta sheets, such as the Adenine nucleotide alpha hydrolase-like and the UDP-Glycosyltransferase/glycogen phosphorylase folds. As expected, the *MmThs*-substrate interaction is also influenced by fold motifs. The thermosome associates preferentially with proteins of the ribonuclease H-like domain fold, a c-

---

class domain fold with alpha helical elements, and the OB-fold, which has only beta-elements (b-class).

In *Methanosarcina*, the group I chaperonin as well as a large fraction of other proteins are of bacterial origin, and therefore may preferentially interact. However, our analysis revealed that there was no direct relationship between the phylogeny of the substrate protein and its chaperonin interaction in *M. mazei*. The increased number of bacterial proteins among both chaperonin substrate sets indicates an important role in the acquisition and/or the maintenance of proteins derived from horizontal gene transfer. Interestingly, the analysis of the evolutionary state of the substrates suggests a role of Group II chaperonins in the evolution of protein structures, while the Group I chaperonin might rather contribute to the conservation of a protein. The interaction of a significant number of essential proteins with both chaperonin groups makes both *MmGroEL* as well as *MmThs* likely to be essential for *M. mazei*.

In conclusion, substrate selection of chaperonins is defined by a combination of a variety of physical properties (hydrophobicity, net charge and size), structural features (*i.e.* the domain fold), and less concrete characteristics like the evolutionary status and, in this context, the phylogenetic origin of the substrate.

The use of the identified chaperonin substrates in *in vitro* studies should give further insight into the mechanistic differences between the two chaperonin systems. In the future, *M. mazei* – harbouring not only both chaperonin systems, but also the chaperones NAC, GimC and DnaK/J/E – will provide the opportunity to analyse the interplay of chaperones in a network containing the major components known to assist in *de novo* folding.

## II. Introduction

Proteins play an essential role in virtually all biological processes. They are structural building blocks and execute nearly all cellular functions as catalysts of manifold biochemical reactions. They also serve as active or passive transporters of most molecules across biological membranes - even between different cells.

### *II.1 Proteins*

#### *II.1.1 Protein structure*

The specific biological activity of a protein is determined by its three dimensional structure. The architecture of a protein is generally defined at four structural levels: the primary structure of a protein is described by the amino acid sequence of a polypeptide chain, which is defined by the encoding gene. The secondary structure is referred to as the three-dimensional arrangement of proximal amino acid side chains resulting in structures like the  $\alpha$ -helix or the  $\beta$ -sheet. The tertiary structure of proteins describes the three-dimensional arrangement of secondary elements within a single polypeptide chain, while the quaternary structure gives information on the interaction between different polypeptide chains in oligomeric proteins.

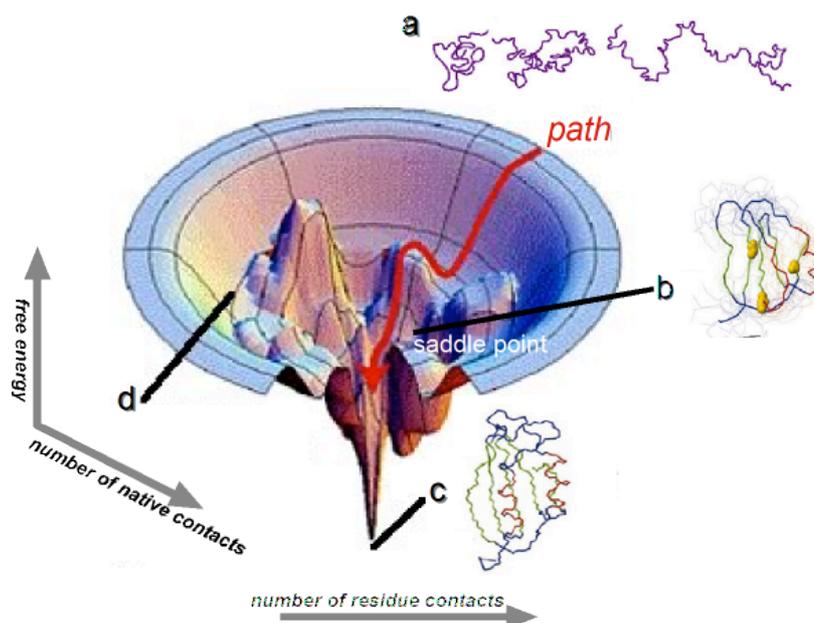
#### *II.1.2 Protein folding: from primary to quaternary structure*

All information determining the native structure, which is generally the polypeptide conformation with the lowest Gibbs free energy of the complete system, is fully contained in the amino acid sequence of a protein. But the process of the formation of a native protein from the polypeptide, which is generally referred to as protein folding, involves a complex arrangement and rearrangement of non-covalent interactions of amino acid residues along the polypeptide chain.

In 1960 Christian Anfinsen showed that this process can occur spontaneously *in vitro* (Schechter *et al.*, 1970). However, considering that a small protein with 150 amino acid residues can theoretically adopt  $\sim 10^{300}$  different conformations and transitions between these different states occur at a maximum rate of  $\sim 10^{11} \text{ s}^{-1}$ , it would take about  $10^{11}$  years to search through the full repertoire of conformations. This so-called Levinthal paradoxon led to the conclusion that a protein can not adopt all possible conformations during the folding process, but the folding rather “is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide. This suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process” (Levinthal, 1969).

### *II.1.3 The protein folding mechanism*

Protein folding seems to follow a certain pathway through folding intermediates (Baldwin, 1996; Baldwin and Rose, 1999; Privalov, 1996), reducing the amount of possible conformations and therefore allowing protein folding to occur in biological relevant time. The folding intermediates possess partially native and stabilized structural elements as well as unstructured non-native elements. The folding process can be illustrated by ‘folding energy landscapes’ (Figure 1; (Dobson, 1998). The energy landscape describes the free energy of a molecule depending on certain variables. In a folding energy landscape, the native state of a protein is defined by the global thermodynamic energy minimum. To reach this structure, the unfolded polypeptide may have to transit a variety of intermediates, which are defined by local energy minima. This reduction of possible conformations enables a protein to fold within milliseconds and resolves the “Levinthal paradoxon” (Schultz, 2000).



**Figure 1 : Schematic energy landscape of protein folding**

The free energy is represented as a function of the topology of its atoms. (a) representative starting structures, (b) transition state, (c) lowest energy structure, native state, (d) collapsed, hydrophobic internal, molten globule (Dobson, 2003, modified).

## II.2 Protein folding in vivo

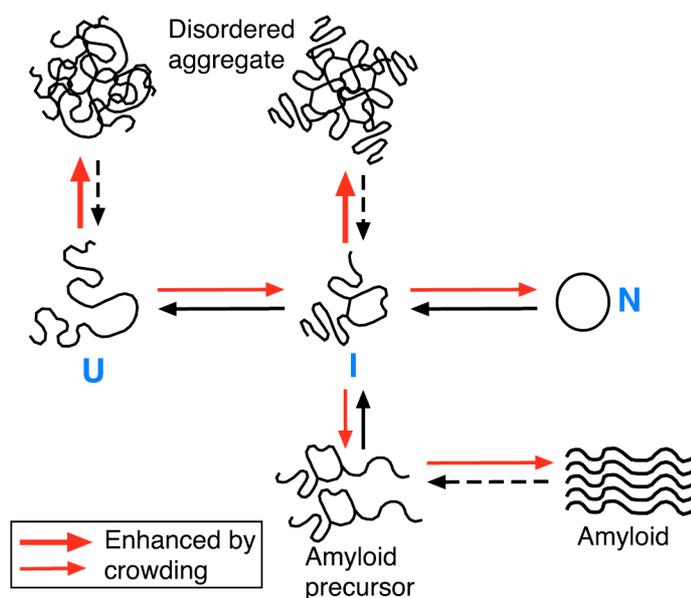
### II.2.1 Protein folding in the cellular environment

In contrast to an *in vitro* folding situation where proteins were shown to be able to fold spontaneously, protein folding inside a cell is not a spontaneous process. In a living cell factors like temperature, pH-value and ion concentrations deviate from optimal conditions that can be provided *in vitro*. Also, in a living cell most proteins have an average size of 40-60 kDa and generally exceed the size of model proteins used in *in vitro* experiments (Dobson, 1998).

More importantly the cellular environment with protein concentrations of 200-300 g/l, varies dramatically from the dilute solution used for refolding *in vitro* (Sakikawa *et al.*, 1999). This high concentration of molecules causes excluded volume effects (macromolecular crowding), leading to increased thermodynamic activity and changes in binding equilibria (Minton, 1983; Zimmerman and Minton, 1993). The excluded volume effect favours compact structures relative to more expanded structures. Therefore, *in vivo* folding intermediates can be formed harbouring

secondary structures, while no stable tertiary structure is adopted yet (*molten globules*). In this state hydrophobic regions are often exposed, which would be hidden inside the molecule in the native state. These hydrophobic residues can lead to unspecific associations or irreversible aggregation of the protein (Ellis and Hartl, 1999).

In some cases, contacts may be formed between regions of a polypeptide that are not associated in the native status, which can lead to a reversible misfolded state of the polypeptide, the so-called “kinetically trapped intermediates” (Pande *et al.* 1998; Dobson and Karplus 1999; Dinner *et al.* 2000). In these misfolded structures hydrophobic residues can be exposed, contributing to an increased aggregation activity (Figure 2; Dobson *et al.* 1998).



**Figure 2 : Aggregation of unfolded polypeptide chains is a side-reaction of protein folding**

Molecular crowding (red arrows) in the cytosol can amplify these reactions as well as folding to the native state. U: Unfolded polypeptide chain, I: partially unstructured intermediate, N: native protein (Hartl and Hayer-Hartl, 2002).

A special case of protein aggregation is the formation of amyloid fibrils, which were shown to be associated with several protein folding diseases, such as Alzheimer’s, Huntington’s, Parkinson’s and Creutzfeldt-Jacob’s disease (Dobson, 1999). These amyloids exhibit a highly

ordered fibrillar structure with a characteristic X-ray fibre diffraction pattern, consistent with a helical array of beta-sheets parallel to the fibre long axis, with the strands perpendicular to this axis. It is expected that all proteins have the potential to form amyloid structures, and such formations are favoured under conditions of extreme pH or elevated temperature. But in contrast, the proteins involved in protein folding diseases assemble into amyloids under physiological conditions.

### *II.2.2 Protein folding upon de novo synthesis*

In contrast to the *in vitro* refolding situation, the complete polypeptide chain is not available for folding during *de novo* synthesis of proteins. Thus, the incomplete polypeptide chain has to be prevented from misfolding and aggregation during translation. The cell overcomes this problem by engaging a group of proteins termed molecular chaperones (Hartl 1996; Netzer und Hartl 1998; Ellis und Hartl 1999; Agashe und Hartl 2000; Feldman und Frydman 2000). These molecular chaperones reversibly bind to the nascent chain or the newly synthesized polypeptide preventing it from misfolding and aggregation, and can even promote folding to the native state.

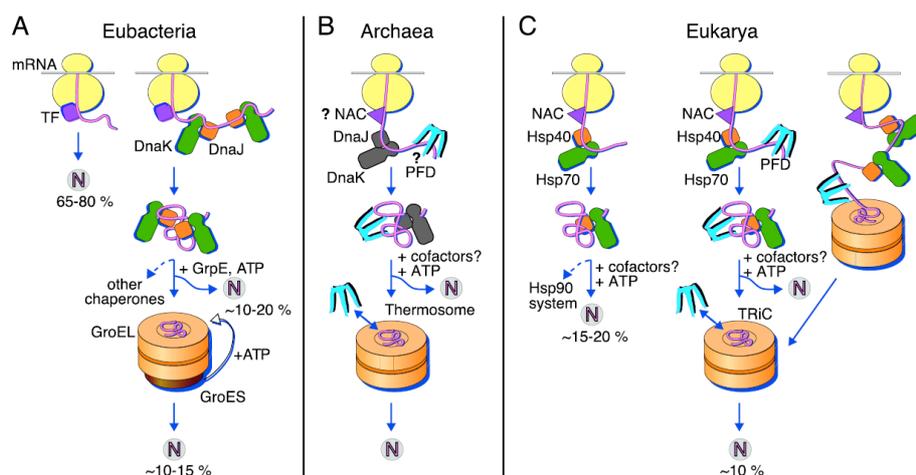
## *II.3 Molecular chaperones*

In a cell polypeptides are generated at the ribosome and released into the cytosol. During this process, the newly synthesized polypeptide exposes unstructured hydrophobic residues to the crowded cellular environment. In order to prevent aggregation of these non-native peptides, nascent chain interacting and downstream acting chaperones are necessary to shield the aggregation prone regions as soon as the peptide chains exits the ribosomal tunnel. Chaperones carry no structural information for the native state of their target protein. They generally bind non native states of newly synthesized or stress denatured proteins and promote their folding by repeated cycles of binding and release, a process often driven by ATP hydrolysis. Under stress conditions, for example elevated temperatures, many proteins tend to

unfold triggering the heat shock response mechanism, which results in a high level induction of chaperones (Lindquist 1986; Morimoto 1998, Ellis *et al.* 1987). Therefore molecular chaperones are also referred to as stress proteins or heat shock proteins.

### II.3.1 The chaperone system

The genes of molecular chaperones are highly conserved and are divided into chaperone groups according to sequence homologies and molecular sizes, such as the Hsp60's or Hsp70s. Among the three domains of life, universal groups of chaperones are found as well as domain specific groups. It is speculated that the chaperones inside a cell are organized in networks and a lot of effort has been put into understanding the interplay of chaperones. In principle there are two major groups of chaperones involved in the folding of newly synthesized proteins in the cytosol.



**Figure 3 : Model for *de novo* protein folding assisted by a network of molecular chaperones in the cytosol of Bacteria, Archaea and Eukarya.**

(A) Many proteins in the Bacterial cytosol fold without further assistance upon release from the ribosome and ribosome-bound TF. DnaK assists the remainder of proteins in folding, and can transfer substrates to the chaperonin system (GroEL/GroES). (B) Only some Archaea contain DnaK/DnaJ. Interaction of PFD with nascent chains and existence of NAC is not experimentally confirmed. (C) In the example of the mammalian cytosol, NAC probably interacts with nascent polypeptide chains, together with Hsp70 and Hsp40. The majority of proteins can fold upon release from these factors. A subset of Hsp70 substrates is transferred to the Hsp90 system. Furthermore, PFD interacts with nascent chains, and transfers these to TRiC, the eukaryotic chaperonin.

N: Natively folded protein, TF: trigger factor, NAC: nascent chain-associated complex, PFD: prefoldin (Hartl and Hayer-Hartl, 2002).

The first contact upon synthesis is provided by the ribosome associated chaperones and the Hsp70 group associate with the nascent polypeptide chain. In the eukaryotic cytosol and in Archaea an additional co-factor termed prefoldin or *GimC* was also found to interact with newly synthesized polypeptides. The prefoldin and the Hsp70s are able to hand over their substrates to second type of chaperones the large cylindrical chaperonins (Hsp60's), which act further downstream in the folding network.

### *II.3.2 Ribosome associated chaperones*

The first chaperones to interact with a nascent polypeptide chain emerging from the ribosome are trigger factor (TF) in *Bacteria* and possibly the nascent chain associated complex (NAC) in *Archaea* and the eukaryotic cytosol. These chaperones bind to the nascent polypeptide and are themselves associated with the ribosome.

The well studied trigger factor has been shown to bind to the protein L23 at the large ribosomal subunit (Kramer *et al.*, 2002). From this docking site, TF supposedly scans the emerging peptide for hydrophobic regions in an ATP-hydrolysis independent manner (Hesterkamp *et al.*, 1996). In addition to the chaperone activity, TF exhibits a peptidyl prolyl *cis-trans* isomerase activity, a function that has been shown to be dispensable for the biological role of TF in protein folding (Genevaux *et al.*, 2004; Kramer *et al.*, 2004).

In the eukaryotic cytosol the function of TF is probably taken over by the non homologous protein NAC, which was shown to be associated with the ribosome and interact with the emerging polypeptide (Wiedmann *et al.*, 1994). However, a proof for the chaperone activity of NAC is still missing. Only recently though, was the presence of NAC homologs recognized in several Archaea. The presence of a ubiquitin-associated domain in the crystal structure of the *Methanothermobacter marburgensis* NAC suggests an additional, yet unidentified role for NAC in the cellular protein quality control system *via* the ubiquitination pathway (Kramer *et al.*, 2004; Spreter *et al.*, 2005).

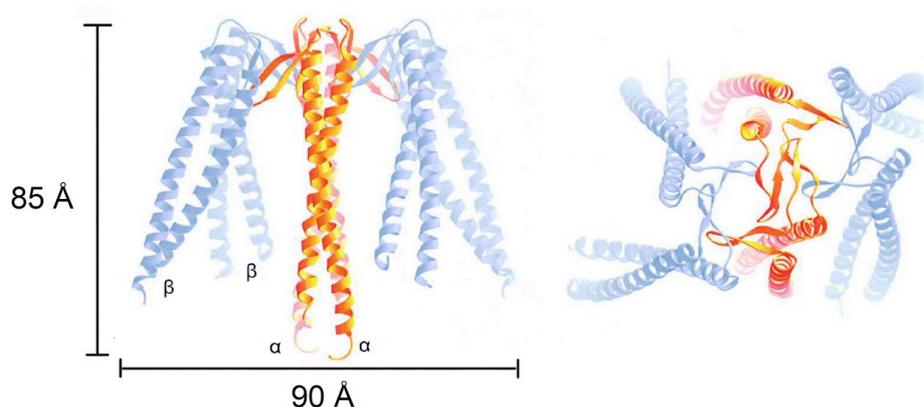
### II.3.3 The Hsp70s

The function of TF overlaps with the further downstream acting Hsp70 system in stabilizing the newly synthesized polypeptide. Even though Hsp70 is not expected to bind directly to the ribosome, it was shown to interact with the nascent chain (Teter *et al.*, 1999). For this reason a combined deletion of TF and DnaK resulted in increased protein misfolding and protein aggregation, leading to a loss of viability of cells at temperatures above 30°C (Deuerling *et al.*, 1999; Genevaux *et al.*, 2004; Teter *et al.*, 1999). The Hsp70- system is not only found in all *Bacteria*, but also in the eukaryotic cytosol and in some *Archaea*, as a result of horizontal gene transfer (HGT) from *Bacteria* (Macario and de Macario, 1999).

The Hsp70s are involved not only in protein folding but also in a variety of functions such as disaggregation and degradation of proteins (Glover and Lindquist, 1998; Goloubinoff *et al.*, 1999; Hohfeld *et al.*, 2001). The 70kDa chaperone action depends on the assistance of a variety of partners. In protein folding, the 40kDa proteins (Hsp40) and nucleotide exchange factors cooperate with Hsp70 (Craig *et al.*, 2006), while interactions with the Hsp104/Cip recruits Hsp70 in protein disaggregation (Glover and Lindquist, 1998). In the eukaryotic cytosol, Hsp70 is also functionally linked to the Hsp90 system (Wegele *et al.*, 2004).

### II.3.4 Prefoldin / GimC

In the eukaryotic cytosol and in *Archaea* an additional chaperone is involved in the folding of newly synthesized proteins. Prefoldin(Pfd)/GimC, in an ATP-hydrolysis independent manner, stabilizes the newly synthesized polypeptide chain to maintain the non native protein in a folding competent state and to subsequently hand the substrate over to the downstream folding machinery. Prefoldin has the appearance of a jellyfish: its body consists of a double beta barrel assembly with six long tentacle-like coiled coils protruding from it. The distal regions of the coiled coils expose hydrophobic patches and are required for multivalent binding of nonnative proteins (Figure 4; Siegert *et al.* 2000).



**Figure 4 : Prefoldin structure from *Methanobacterium thermoautotrophicum*.**

Side view (left) and bottom view (left) of the prefoldin complex. A large central cavity is bordered by six long tentacles formed by rods of  $\alpha$ -helical coiled coils. The coiled coil tentacles point away from a platform consisting of two eight-stranded up and down  $\beta$  barrels. Each prefoldin subunit (pfd a: yellow, pfd b: blue) forms one coiled coil that is tightly anchored to the platform via its proximal end (Siegert *et al.*, 2000).

The distal ends of these N and C-terminal tentacles expose hydrophobic residues that were shown to bind substrate proteins.

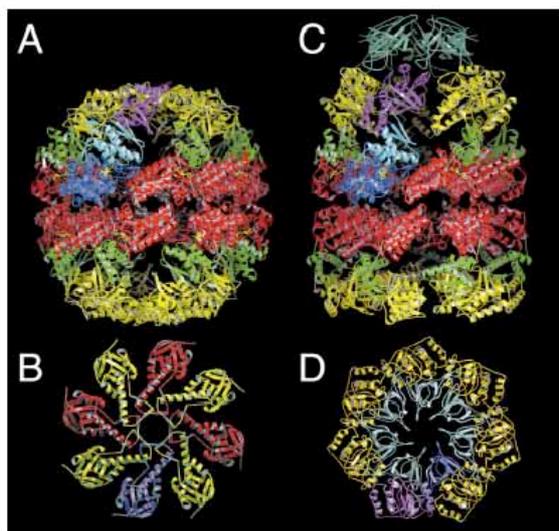
The eukaryotic prefoldin homolog was shown to bind co-translationally to newly synthesized actin and tubulin and is therefore referred to as GimC (Genes involved in microtubule biogenesis). Pfd/GimC subsequently hands over these substrate proteins to the Group II chaperonin TRiC (TCP-1 ring Complex; also termed CCT (chaperonin containing TCP-1)) (Vainberg *et al.* 1998; Hansen *et al.* 1999; Siegers *et al.* 1999). The role of Pfd in archaeal protein folding remains unclear as the *Archaea* do not possess cytoskeletal proteins such as tubulin or actin.

### II.3.5 The chaperonins

The chaperonins are highly conserved and structurally related proteins (Figure 5) and they are present in virtually every living cell. Hemmingsen first identified this specific class of large double-ring oligomeric complexes of ~800 kDa (Figure 5; Hemmingsen *et al.* 1988). Two groups of chaperonins are known, and even though they share only little sequence homology they show a very similar topology. Group I chaperonins, also referred to as Hsp60s, are

exclusively found in *Bacteria* and eukaryotic organelles, which are known to be of bacterial origin. The Group I chaperonins cooperate in protein folding with a member of the Hsp10 family that covers the barrel shaped chaperonin in a lid-like manner. In contrast, the Group II chaperonins fulfil their function independently; they are equipped with helical protrusions that work as a built-in lid. Group II chaperonins are found in *Archaea* as the thermosome and in the eukaryotic cytosol as the so-called TRiC or CCT (Trent *et al.* 1991; Gao *et al.* 1992; Lewis *et al.* 1992; Kubota *et al.* 1995 (Frydman, 2001)).

Chaperonins do not only prevent the aggregation of proteins by shielding exposed hydrophobic residues, they actually encapsulate their substrate into the central cavity of the barrel shaped complex – serving as a so-called folding chamber, where the unfolded protein is protected from non productive interactions with macromolecules of the cellular environment and can undergo folding (Mayhew *et al.*, 1996; Weissman *et al.*, 1995). This typical encapsulation mechanism as well as the downstream position in the chaperone network may contribute to a narrower substrate set for chaperonins than for the more promiscuous upstream chaperones. Only 10-15% of all newly synthesized proteins in the bacterial cytosol were shown to interact with the bacterial Hsp60 typically referred to as GroEL/GroES (Ewalt *et al.*, 1997).



**Figure 5 : Structure of chaperonins**

(A) side view and (B) topview of the closed hexadecamer thermosome (Ditzel *et al.* 1998). (C) side view and (D) topview of the GroEL-GroES-(ADP)<sub>7</sub> complex (Xu *et al.* 1997).

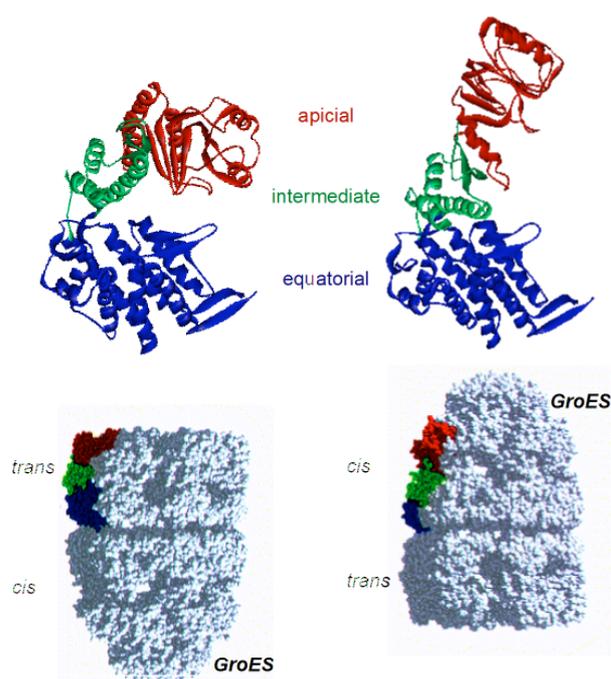
## II.4 The Group I chaperonin system

### II.4.1 Structure and function of the Group I chaperonin system GroEL/GroES

Detailed insights into the structural and mechanistic principles of chaperonins arose from extensive studies on the Group I chaperonin in *E. coli*, referred to as EcGroEL and its co-factor GroES. The double ring structure of EcGroEL, in which two rings of seven identical 57kDa subunits are stacked back to back, was shown in several crystallographic (Boisvert *et al.*, 1996; Braig *et al.*, 1994; Xu *et al.*, 1997) and electron microscopic studies (Ranson *et al.*, 2001; Roseman *et al.*, 1996; Saibil *et al.*, 1991).

The GroEL monomer is divided into three domains (Figure 6). The apical domain exposes hydrophobic residues at the entrance of the cavity for substrate binding and subsequent interaction with the co-chaperone EcGroES (Fenton *et al.* 1994; Hartl 1996; Bukau und Horwich 1998; Sigler *et al.* 1998). It is expected that the substrate binding is mediated by cooperation of apical domains of several subunits within one ring (Farr *et al.* 2000). The equatorial domain contains the nucleotide binding pocket and forms inter-subunit contacts within the GroEL ring and also between the two GroEL rings. Both domains are connected *via* an intermediate domain which passes, in a hinge like manner, allosteric information between the apical and equatorial domain.

Essential for the functionality of GroEL is the 10kDa co-factor GroES, a single homoheptameric dome like structure (Ellis 1996; Hartl 1996; Fenton und Horwich 1997; Bukau und Horwich 1998). GroES binds to the apical domains, covering the central cavity like a lid. This action results in the formation of asymmetric, bullet-shaped GroEL/GroES complexes. The cavity interacting with GroES is defined as the *cis* cavity and the ring opposing is termed as *trans* cavity. The interaction contacts are formed between the so-called mobile loops of GroES at the base of the dome structure and the substrate binding sites of the apical domain of GroEL. Therefore binding of GroES is expected to force the release of apical bound substrates into the central cavity of GroEL (Xu *et al.* 1997; Bukau und Horwich 1998; Richardson *et al.* 1998; Sigler *et al.* 1998), allowing the non native substrate to undergo folding within a shielded environment (Martin *et al.* 1993; Mayhew *et al.* 1996).



**Figure 6 : Hinge rotations in the GroEL subunit upon binding of GroES**

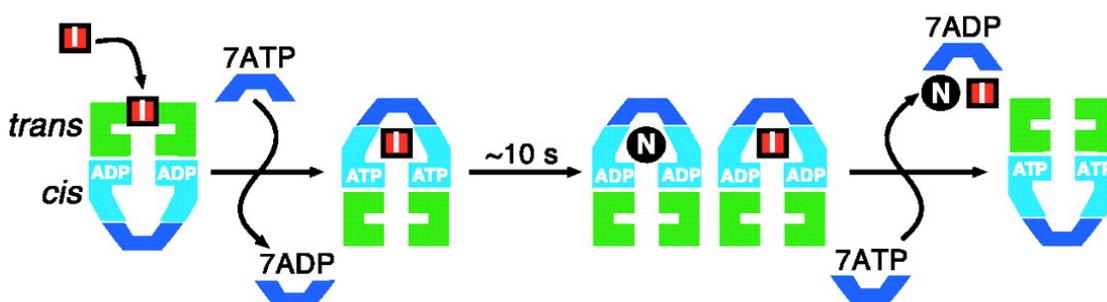
The three domains of the GroEL subunit are coloured individually (apical in red, intermediate in green and equatorial in blue). A schematic representation (upper left) and backbone ribbon of the atomic structure (lower left) of unliganded GroEL are shown. Upon binding of GroES and ADP (GroES not shown), a 60° opening and a 90° rotation of the apical domain is seen (upper right and lower right respectively). Modified from Xu et al., 1997

#### II.4.2 Mechanism of the GroEL/GroES system

The folding mechanism of the bacterial GroEL/GroES system has become well understood within the last years.

The folding process initiates with the binding of the non native substrate to the apical domains of one GroEL ring, subsequent ATP binding induces structural rearrangements of the apical domains in the *cis*-ring. This allows association of GroES and displacement of the substrate into the folding cage. Binding of GroES also results in an increase of the volume of the *cis* cavity (Hayer-Hartl *et al.* 1994; Farr *et al.* 2000; Chen *et al.* 1999). During the next 10-15s of ATP-hydrolysis, the encapsulated protein can undergo folding (Hayer-Hartl *et al.* 1995; Rye

*et al.* 1997). Following hydrolysis of ATP in the *cis* ring, binding of ATP to the *trans* ring causes dissociation of ADP and GroES from the *cis* cavity allowing the substrate protein to leave the folding chamber. Depending on the substrate, cycles of rebinding, folding trials and release are repeated until the protein reaches its native state (Figure 7).



**Figure 7 : Schematic drawing of the GroEL/GroES folding cycle**

Unfolded protein binds (red) to the hydrophobic substrate binding site of GroEL (green). GroES (dark blue) and ATP bind the so-called *cis* ring. This leads to a conformational change in GroEL and the substrate protein gets included into the folding chamber. During hydrolysis of ATP, the substrate undergoes folding inside the GroEL/GroES cavity. Binding of ATP and GroES to the *trans* ring leads to dissociation of ADP and GroES from the *cis* ring, resulting in a release of the substrate. The substrate can rebind and undergo additional folding cycles until it reaches its native state.

#### II.4.3 Substrates of the GroEL/GroES system

Even though much is known about the structure and mechanism of GroEL, information about the natural substrates of GroEL has only recently become available. 10-15% of the cellular protein was shown to interact with the GroEL chaperonin system (Ewalt *et al.*, 1997), but only little was known about their identity and therefore most functional studies used heterologous model proteins, which do not interact with *EcGroEL in vivo*. Several endogenous substrates were identified in 1999 (Houry *et al.*, 1999), and in a recent study, the virtually complete *EcGroEL* interaction proteome was identified revealing three different classes of substrates (Kerner *et al.*, 2005). The non stringent class I substrates occupy the capacity of GroEL only to a very limited extent (2%) and class II substrates are able to switch to other chaperones, such as the bacterial Hsp70 DnaK, for folding assistance. Only the class III

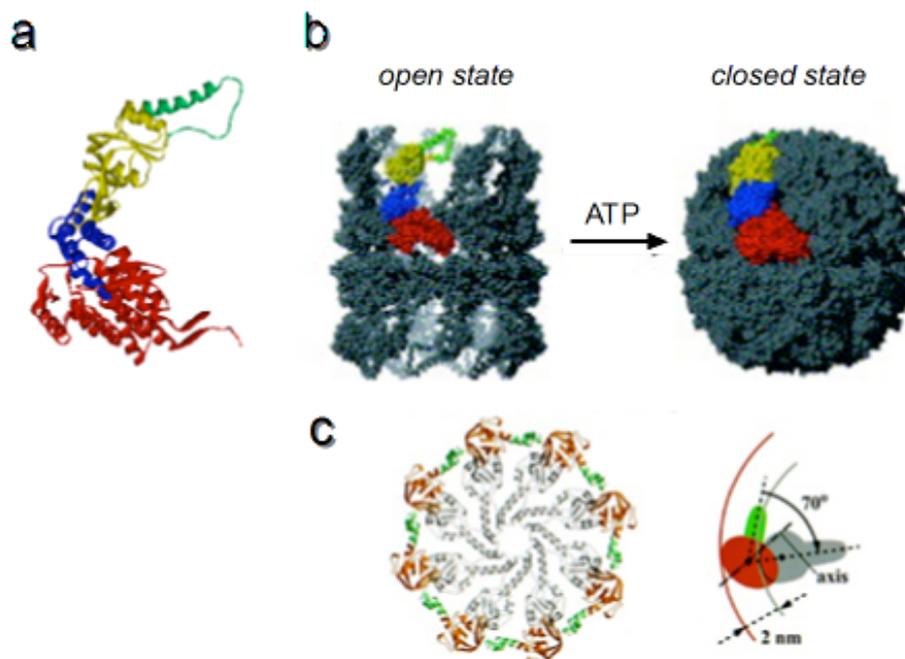
substrates were shown to be absolutely dependent on the chaperonin system to reach their native state. The presence of essential proteins among the class III substrates explains the essential requirement of GroEL itself. Class III preferentially contains proteins with sizes compatible with the available space inside the GroEL/GroES cavity, with an upper limit of approximately 55 – 60 kDa. Bioinformatic analysis of the GroEL interactome revealed a significant enrichment of a particular  $\alpha\beta$  fold (the Triosephosphate isomerase (TIM)  $(\alpha\beta)_8$  barrel) among the stringently GroEL-dependent class III substrates. This suggests that certain TIM barrel proteins tend to populate kinetically trapped folding intermediates that depend on encapsulation by the chaperonin for acquisition of their native state.

## *II.5 The Group II chaperonin system*

### *II.5.1 Structure and function of Group II chaperonins*

Compared to the abundance of information about Group I chaperonins, our knowledge about Group II chaperonins is lagging behind. Electron micrographs and crystallographic studies revealed the typical double-ring structure with each ring offering a folding cavity. In contrast to GroEL, the building blocks are variable, and the rings are formed by up to eight different subunits. Additionally, depending on its phylogenetic origin, the symmetry of the group II chaperonin ring varies from 8-9 fold (Gutsche *et al.*, 1999; Leroux, 2001).

So far the crystal structure of an archaeal Group II chaperonin has been solved (Bosch *et al.*, 2000; Ditzel *et al.*, 1998); Klumpp *et al.* 1997). The thermosome from *Thermoplasma acidophilum* consists of two rings, each composed of eight subunits of alternating  $\alpha$  and  $\beta$  monomers. The rings are arranged in a back to back manner mediated by contacts between identical subunits forming homodimers (Figure 8).



**Figure 8 : General architecture of Group II chaperonins.**

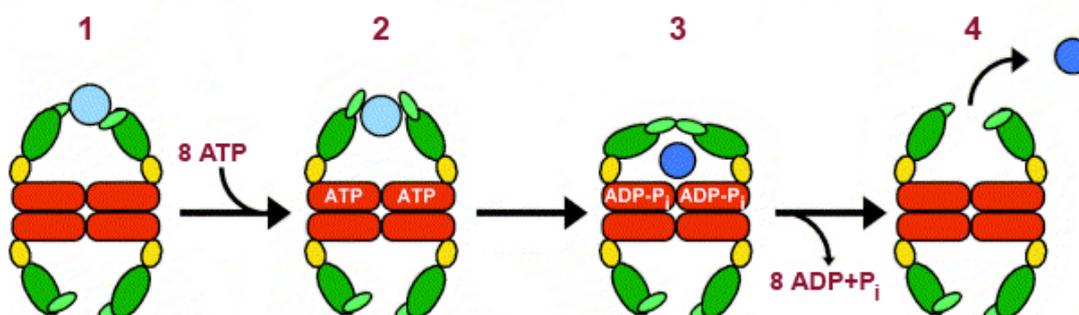
(a) Ribbon diagram of an  $\alpha$  subunit of the thermosome from *Thermoplasma acidophilum*. The equatorial ATPase domain (red) is linked to the substrate-binding apical domain (yellow) by a flexible hinge or intermediate domain (blue). The helical protrusion, which is unique to Group II chaperonins, is in green. (b) Bead models of the ATP-induced transition from the open to closed state for Group II chaperonins. The model of the nucleotide-free, open state (left) is based on electron tomographic studies on the thermosome. The closed state is from the X-ray structure of the thermosome and presumably reflects the ATP-induced state (Spiess *et al.*, 2004). (c) Top view of the apical domains in their ring context in the open (coloured) and closed (grey) conformations. The diagram shows the rotational and translational movement of one apical domain between the open and the closed conformations. The axis of the rotation is tilted by  $60^\circ$  against the paper (or x-y) plane (Gutsche *et al.*, 1999).

Analogous to the GroEL complex, single subunits with a size of  $\sim 60$ kDa are themselves divided into three domains (Figure 8a). The apical domain exposes hydrophobic residues at the entrance of the cavity for substrate binding. The equatorial domain contains the nucleotide binding pocket and forms contacts between the two rings. The intermediate domain functions as a hinge between the apical and equatorial domains passing steric information.

In contrast to the Group I chaperonins, the thermosome and eukaryotic TRiC do not depend on the assistance of a co-chaperonin. Group II chaperonins are thought to contain a built-in lid formed by helical protrusions on the apical domains (Klumpp *et al.* 1997).

### II.5.2 Mechanism of Group II chaperonins

The mechanism of protein folding assisted by Group II chaperonins is only poorly understood. Crystallographic data and neutron diffraction data of thermosome in different nucleotide bound states suggest a mode of action similar to Group I chaperonins (Gutsche *et al.* 2000; Gutsche *et al.* 2000; Figure 9).



**Figure 9 : Folding mechanisms of Group II chaperonins**

(1) The substrate (blue) is bound by the apical domains (green) of the thermosome. Binding of ATP to the equatorial domain (red) leads to rotation of the apical domains (2). Hydrolysis of ATP results in closing of the folding chamber by the helical protrusion (bright green) of the apical domain. The substrate undergoes folding in the closed environment until it gets released by the opening of the apical domains (4). (Model from (Kusmierczyk and Martin, 2001)).

First, the open thermosome, exposing the helical protrusions to the environment, accepts the substrate protein at the apical domain. Binding of ATP to the equatorial domain causes a rotational movement of the apical domains, leading to substrate release into the folding chamber and a simultaneous closure of the cavity by the helical protrusions. After ATP-hydrolysis, ADP is released from the nucleotide binding pocket inducing a conformational change in the apical domain and opening the folding cavity to release the substrate (Ditzel *et al.*, 1998).

It is expected that the helical protrusions of the apical domains replace the lid function of a GroES like co-chaperonin, but proof is still missing. So far it is not clear whether the helical

protrusions and/or hydrophobic residues of the apical domain are responsible for substrate binding (Llorca *et al.*, 2001; Llorca *et al.*, 2000).

The use of a flexible built-in lid instead of an independent lid structure is a major difference between Group II chaperonins from the bacteriotypic Group I homologs. It is speculated that this different closing mechanism enables the thermosome and TRiC to encapsulate substrate proteins exceeding the calculated size of the closed cavity of  $\sim 130\,000\text{ \AA}^3$ , giving space for proteins of sizes up to 50kDa (Ditzel *et al.*, 1998). The fact that the eukaryotic chaperonin was shown to successfully fold firefly luciferase, a 62 kDa protein, suggests that substrates do not have to be completely encapsulated for efficient folding to occur. This might enable the Group II chaperonin to interact co-translationally with substrates *in vivo*, explaining the experimentally observed association of TRiC with nascent polypeptide chains (Frydman *et al.*, 1994). The ability of TRiC to accept a growing nascent polypeptide in the active “closed” state promotes co-translational folding, which is specific to eukaryotic systems (Netzer and Hartl, 1997). The domain-wise folding of polypeptides in a cavity also allows folding of proteins too large to be encapsulated.

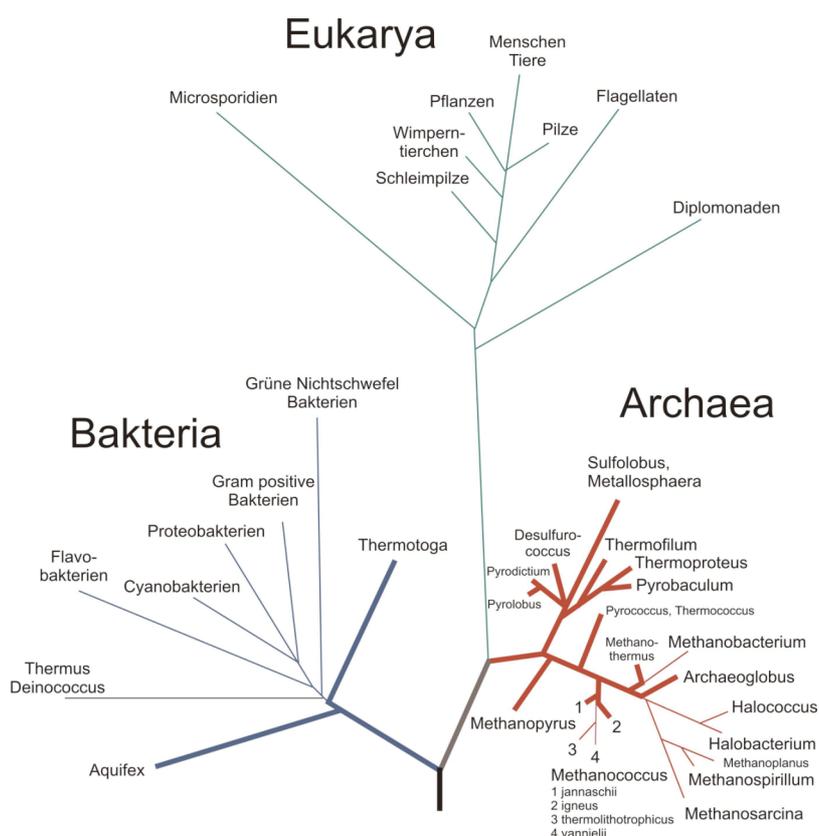
### ***II.5.3 Substrates of Group II chaperonins***

So far no comprehensive study has been performed on endogenous substrates of Group II chaperonin. Several proteins were shown to depend on CCT/TRiC in eukaryotes *in vitro* and *in vivo* (Dunn *et al.*, 2001; Spiess *et al.*, 2004), but no *in vivo* substrates of thermosomes have been identified yet. However, the potential of the archaeal chaperonin to prevent aggregation and to promote refolding of proteins has been shown for several heterologous model substrates (Gutsche *et al.*, 1999). Although, the low efficiency of the thermosome in refolding of these non natural substrate proteins may not allow direct conclusions about the *in vivo* situation.

## II.6 Archaea

It was a revolutionary view, when in 1977 Carl Woese first proposed, based on sequence comparisons of a single molecule, the 16S rRNA, that all life forms can be divided into the three kingdoms: *Eukarya*, *Bacteria* and *Archaea* (Woese and Fox, 1977).

As illustrated in the so-called universal tree of life, *Archaea* represent a clear and separate domain that is as distantly related to *Bacteria* as it is to *Eukarya*.



**Figure 10 : Universal tree of life**

Universal tree of life based on 16SrRNA sequences. *Archaea* (red) are a clearly separate line from *Bacteria* (blue) and *Eukaryotes* (green). Most hyperthermophile organisms (thick lines) are found among *Archaea*, only some among *Bacteria*

According to mostly rRNA information, *Archaea* themselves are subdivided into four subkingdoms: the *Euryarchaeota*, comprising the methanoges, halophiles, thermoacidophiles and thermophiles, the *crenarchaeotes*, a group of only hyperthermophilic members (Woese *et al.*, 1990), the recently discovered *Nanoarchaeota* (Hohn *et al.*, 2002), a probably exclusively

parasitica group of organisms, and the *Korarchaeota* (Barns *et al.*, 1996), which have not been successfully cultivated until now.

The in-between position of *Archaea* in the phylogenetic tree reflects their mixed ancestry. Archaeal genes, which code for metabolic properties, have homologies to bacterial counterparts, while other archaeal gene products clearly resemble the eukaryotic counterparts. It is even speculated that fusion of an archaeal cell with a bacterium gave rise to the eukaryotic cell (Martin and Muller, 1998).

Like *Bacteria*, *Archaea* contain no defined nucleus. Even though *Bacteria* and *Archaea* are very similar in cell size, *Archaea* typically adopt irregular cell shapes. Bacterial and archaeal cells are usually protected by cell walls, but the archaeal envelope is built up from a variety of compounds, depending on the particular organism, and is never composed of peptidoglycans (Kandler and Konig, 1998). Typical for both prokaryotic kingdoms is their enormous repertoire of different metabolism compared to eukaryotes.

The information processing machinery of *Archaea* resembles the eukaryotic system, and this is reflected in its insensitivity to antibiotics impairing transcription or translation (Table 1). Also, archaeal components of the protein folding machinery share strong homology with the eukaryotic system: there are homologues to eukaryotic NAC as well as prefoldin.

As a unique archaeal feature, the cell membrane lipids are ether linked, which is expected to contribute to a higher stability of archaeal membranes (De Rosa and Gambacorta, 1988). Many *Archaea* are exclusively found in extreme environment such as hydrothermal vents with temperatures exceeding the boiling point of water, areas of saturated salinity or pH values below 2.

	<i>Bacteria</i>	<i>Archaea</i>	<i>Eukarya</i>
Cell size	Ø ~ 1 µm	Ø ~ 1 µm	Ø ~ 10 µm
Membrane-bound nucleus	-	-	+
RNA- Polymerases	1 (4 su)	1 (~ 12 su)	3 (12-14 su)
Sensitivity to diphtheria toxin	-	+	+
Ribosomes	70 S	70 S	80 S
Initiator tRNA	F- Met	Met	Met
Sensitivity to kanamycin, streptomycin, chloramphenicol	+	-	-
Chaperonin Group	I	II	II
Membrane lipids	ester linked	ether linked	ester linked

Table 1: Common features among the domains of life

In the past it seemed a typical hallmark of *Archaea* to populate only niches hostile to life, but environmental surveys indicate that the *Archaea* are diverse and abundant not only in extreme environments, but also in soil, oceans and freshwater, where they may fulfil a key role in the biogeochemical cycles of the planet (Robertson *et al.*, 2005).

### II.6.1 *The Methanogens*

The biological production of methane is exclusively carried out by a highly specialized group of *Euryarchaeotes*- the methanogens. Although methane is a relatively minor component of the global carbon cycle, it is of great importance. Most methanogens use CO<sub>2</sub> as their terminal electron acceptor, H<sub>2</sub> generally serves as electron donor. Despite the obligate anaerobiosis and the specialized metabolism of methanogens, they are quite widespread on earth. Although high levels of methanogenesis only occur in anoxic environments, such as swamps and marshes, or in the rumen, the process also occurs in habitats that might be

considered oxic, such as forest and grassland soils. In such habitats, methanogenesis occurs in microenvironments, for example, in the midst of soil crumbs.

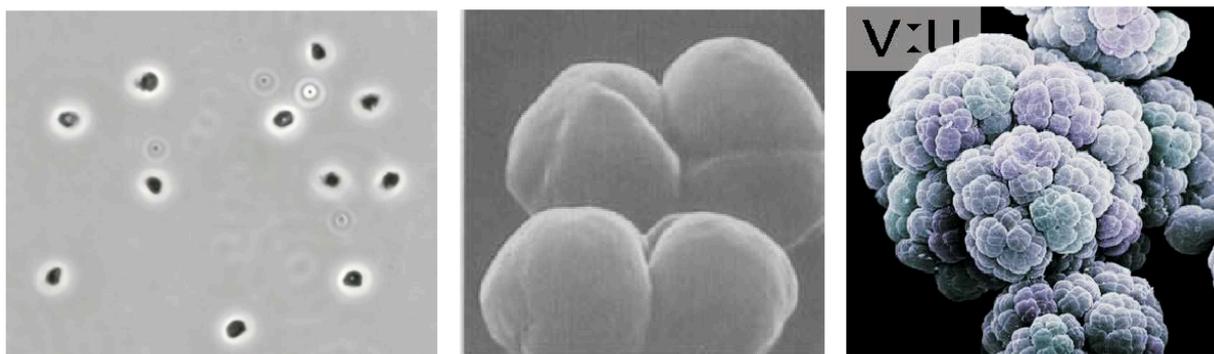
So far a variety of morphological types of methanogenic *Archaea* have been isolated, and studies of their physiology and molecular properties served to classify the methanogens into seven major groups containing a total of 17 *genera*.

Especially now as the need for alternative energy sources becomes eminent, methanogens are gaining more and more importance for industry. Methane production is widely used as a renewable source of energy, and methanogens are also essential components of all urban, agricultural and industrial waste-treatment facilities.

### II.6.2 *The genus Methanosarcina*

The genus *Methanosarcina* was first described by Kluver and van Niel as “irregular spheroid bodies occurring alone or typically in aggregates of cells” (Kluver and Niel, 1936). *Methanosarcina* belong to the subkingdom of *Euryarchaeota* and grow like all methanogens that gain energy from converting CO<sub>2</sub> and H<sub>2</sub> to methane under strictly anaerobic conditions. Among methanogens, the members of the genus *Methanosarcina* can in addition convert short chained carbon sources like acetate, methanol, monomethylamine, dimethylamine, trimethylamine and possibly CO to methane. Members of the genus *Methanosarcina* are found in diverse oxygen free environments, such as rice fields, sea- or fresh water sediment and typically in rennet.

The irregular, non motile cocci have a size of 1-3 μm defined by a proteinous cell wall adjacent to the cell membrane, which may be surrounded by an additional heteropolysaccharide layer- a polymer that is similar to the eukaryotic chondroitin and is therefore referred to as methanochondroitin. The methanochondroitin serves as matrix in which the single cells are embedded, giving rise to the formation of manifold pseudoparenchyma, such as packets (*sarcina*, *lat.*: packet), laminas and cysts (Figure 11; Macario 1995; Lange *et al.* 1997).



**Figure 11 : Morphology of the genus *Methanosarcina***

Phase contrast micrograph of *Methanosarcina* single cells (left), electronmicrographs of *Methanosarcina* tetrads (middle, from the Department of Agricultural and Biological Engineering, University of Florida, [http://www.agen.ufl.edu/~chyn/age4660/lect/lect\\_08x/lect\\_08.htm](http://www.agen.ufl.edu/~chyn/age4660/lect/lect_08x/lect_08.htm) ) and *Methanosarcina* packets (right, from Ralph Robinson / Visuals Unlimited, <http://www.visualsunlimited.com/browse/vu197/vu197368.html> )

The metabolic versatility makes this genus of methanogens an interesting tool to study methanogenesis and to develop techniques to genetically engineer methanogenic *Archaea*. Not only the unusual growth in different life cycles has drawn the attention of many scientists, recent studies revealed the existence of an additional 22<sup>nd</sup> aminoacid pyrrolysine in the genus *Methanosarcina* (Srinivasan *et al.*, 2002). Sequences of the complete genomes of several *Methanosarcina* species revealed the presence of a high percentage of genes of bacterial origin in the unusual big genomes of more than 4 Mbp. Interestingly, the major components of the bacterial folding machinery are present in all genomes of the *Methanosarcina* species that were sequenced so far ((Deppenmeier *et al.*, 2002; Galagan *et al.*, 2002; Klunker *et al.*, 2003; Macario *et al.*, 1991); <http://www.jgi.doe.gov/>).

Recently the genus *Methanosarcina*, represented by the type species *M. barkeri* and *M. mazei*, *M. acetivorans*, *M. frisia* and *M. thermophila* (Maestrojuan *et al.*, 1992) were accreted by three new species: *M. lacustris*, *M. balitca* and *M. semesiae* (Lyimo *et al.*, 2000; Simankova *et al.*, 2001; von Klein *et al.*, 2002).

### II.6.3 *Methanosarcina mazei* Gö1

The mesophilic *M. mazei* strain Gö1 was isolated from a sewage plant in Göttingen and grows optimally at a neutral pH of 6,8 - 7,2 and salinity around 0,5 M (Deppenmeier *et al.*, 1988). The poor energy yield from methanogenesis at moderate temperature results in a relatively long doubling time, depending on the energy source, of 9 hours on CO<sub>2</sub>, 17 hours on acetate and 7-15 h on methanol and methylamines (Baumer *et al.* 2000).

In 2002 the circular chromosome of *M. mazei*, amounting to 4.2 Mbp has been sequenced (Deppenmeier *et al.* 2002). 3371 open reading frames (ORF) could be identified and 2450 ORFs were annotated to a known function. Surprisingly 1043 ORFs making up one third of the total genome, were identified as being of bacterial origin. These bacterial genes probably originated from multiple events of horizontal gene transfers as they are dispersed all over the genome. Many of these genes are thought to contribute to the metabolic versatility of *M. mazei*, as *Methanosarcina* are the only methanogens known to date that can convert short chained carbon sources, such as acetate, methanol and methylamines to methane (Deppenmeier *et al.* 2002). Surprisingly among these bacterial originated proteins of *M. mazei*, the major components of the bacterial folding machinery, the complete *dnaK*, *dnaJ* and *grpE* – and the *groEL/groES* operon, were also identified.

## II.7 Aim of the project

There is a great deal of structural information on the Group II chaperonin thermosome, obtained either from X-ray diffraction (Ditzel *et al.*, 1998; Klumpp *et al.*, 1997) or electron microscopy (Nitsch *et al.*, 1998), revealing that the overall thermosome structure is similar to the Group I chaperonin. But in contrast to detailed functional studies on Group I chaperonins, there is only limited information on the mechanism of Group II chaperonins, especially those of archaeal origin. The proposed function of the thermosome is mostly based on comparisons to eukaryotic and bacterial counterparts; so far, only heterologous substrate proteins were used in *in vitro* refolding experiments (Gutsche *et al.*, 1999) the result of which may not accurately describe thermosome function *in vivo*. Even the function of the thermosome as a structural element in the natural membrane of *Archaea* remains controversial (Trent *et al.*, 2003).

The aim of this study was on the one hand to identify the natural substrate spectrum of a thermosome in order to draw conclusions regarding on the substrate selectivity of Group II chaperonins and to enable functional studies on thermosomes using natural substrates. On the other hand, the unique co-existence of both chaperonin groups in *Methanosarcina mazei*, allowed a direct bioinformatic comparison of their substrate spectra, providing detailed insights into how the Group I and Group II chaperonins select their substrate proteins.

## III. Materials and Methods

### III.1 Materials

#### III.1.1 Chemicals

Chemicals were of *pro analysi* grade and purchased from **Sigma-Aldrich** (Steinheim, Germany) unless stated otherwise.

**Amersham Pharmacia Biotech** (Freiburg, Germany): ECL™ detection kit, ECF™ detection kit; Thiourea; CyDye DIGE Fluor minimal dyes; PlusOne Bind-Silane and Reference Markers.

**BioMol** (Hamburg, Germany): IPTG; HEPES.

**BioRad** (Munich, Germany): ethidiumbromide; AffiGel 10/15 columns.

**Difco** (Heidelberg, Germany): Bacto tryptone, Bacto yeast extract, Casitone Bacto agar.

**Fluka** (Deisenhofen, Germany): activated charcoal, soda lime, Ammoniumchloride, Iodacetamide.

**New England Biolabs** (Frankfurt am Main, Germany): restriction enzymes, T4 DNA ligase.

**Promega** (Hilden, Germany): Wizard Plus SV Miniprep and Midiprep system, Wizard SV Gel and PCR Clean-Up System

**Pierce Biotechnology** (Rockford, USA): DMP (Dimethyl pimelimidate•2 HCl)

**Roche** (Basel, Switzerland): ATP; ADP; benzonase; Complete protease inhibitor; hexokinase; Pefabloc; proteinase K (PK, *Tritriachium album*); shrimp alkaline phosphatase.

**Roth** (Karlsruhe, Germany): Formaldehyde, Calciumchlorid.

**Serva** (Heidelberg, Germany): Pefabloc Protease inhibitor (4-(2-aminoethyl) benzenesulfonyl-fluoride HCl); Resazurin (Diazoresorcinol).

**Schleicher&Schuell BioScience** (Dassel, Germany): protan nitrocellulose transfer membrane

**WWR** (Darmstadt, Germany): ampicillin; Potassiumdihydrogenphosphate, Potassiumhydrogenphosphate, Magnesiumsulfate Calciumchloride, Sodiumchloride, Methanol, Ethanol, Acetic acid, Sodiumacetate.

Gases were purchased from **Linde Gas AG** (Pullach, Germany): Biogon (N<sub>2</sub>/CO<sub>2</sub>= 80/20), Formiergas (N<sub>2</sub>/H<sub>2</sub>=95/5). The gas pipe system was installed by **Draeger** (Lübeck, Germany).

#### Instruments

**Abimed** (Langenfeld, Germany): Gilson Pipetman (2, 10, 20, 100, 200, 1000 µl).

**Alcatel Hochvakkumtechnik** (Ismaning, Germany): Vacuum pump.

**Amersham Pharmacia Biotech** (Freiburg, Germany): electrophoresis power supply EPS 3500; FPLC systems; SMART-System; prepacked chromatography columns: HiPrep Desalting, MonoQ, HiTrap Heparin, Sephacryl S200/S300, Superdex 200, Superose 6; chromatography resins: Q-Sepharose, DE52, Source 30 Q, Source 30 S. IPGphor II IEF Unit; Ettan DALT *twelve* 230 V; Typhoon Variable Mode Imager; Ettan Spot Picker; DeCyder - DIA, Differential In-gel Analysis software, ImageQuant software version 5.2 (Molecular Dynamics).

**Amicon** (Beverly, MA, USA): vacuum filtration unit (0.2 µm); concentration chambers (Centriprep, Centricon).

**Beckmann** (Munich, Germany): DU 640 UV/VIS Spectrophotometer; centrifuges: Avanti J-25, Avanti J20 XP, J-6B, GS-6R, Optima LE-80k ultracentrifuge.

**Biometra** (Göttingen, Germany): T3 PCR-Thermocycler.

**Bio-Rad** (München, Germany): electrophoresis chambers MiniProtean 2 and 3; electrophoresis power supply Power PAC 300.

**Eppendorf** (Hamburg, Germany): centrifuges 5415C and 5417R, Thermomixer Comfort.

**Fisher Scientific** (Schwerte, Germany): pH meter Accumet Basic.

**Fuji** (Tokyo, Japan): FLA 2000 Phosphoimager.

**Hoefler Scientific Instruments** (San Francisco, USA): SemiPhore blotting transfer unit.

**Invitrogen** (Karlsruhe): 4-16 % BisTris-NuPage Novex Gels, NuPAGE MES SDS Running Buffer, Colloidal Blue Staining Kit and XCell SureLock Mini-Cell

**Mettler Toledo** (Gießen, Germany): balances AG285, PB602.

**Millipore** (Eschborn, Germany): deionization system MilliQ plus PF, Millex-HA filters 0.22µm.

**Misonix Inc.** (New York, USA): sonicator Ultrasonic Processor XL.

**Müller + Krempel AG** (Bülach, Swiss): 1000 ml serum bottles, aluminium rings GL45.

**New Brunswick Scientific** (Nürtingen, Germany): orbital shaker and incubator Innova4430.

**Ochs Glasgerätebau** (Bovenden/Lenglern, Germany): 100 ml serum bottles (N20), butyl rubber septa (N20/GL45), aluminium rings (N20), 1ml syringes (Ersta); Thoma counting chamber.

**Oxoid** (Hampshire, UK): oxoid anaerobic jar.

**Raytest** (Straubenhardt, Germany): AIDA gel imaging software version 2.31; LAS-3000 CCD-Imaging System.

**Toepffer Lab systems** (Göppingen, Germany): two person vinyl glove box (Coy Labs).

**VWR** (Darmstadt, Germany): Dispensette® bottle-top dispenser; viton-hoses.

**Zeiss** (Göttingen, Germany): Axiovert 200 M fluorescence microscope; AxioVision software version 3.1.

### *III.1.2 Buffers and Media*

#### *Buffers*

**Competence buffer I:** 100 mM KCl, 30 mM KOAc, 60 mM CaCl<sub>2</sub>, 15 % glycerol; pH 5.8, adjusted with acetic acid. Filter sterilized and stored at 4 °C.

**Competence buffer II:** 10 mM MOPS, 10 mM KCl, 75 mM CaCl<sub>2</sub>, 15 % glycerol; pH 6.8, adjusted with NaOH. Filter sterilized and stored at 4 °C.

**PBS:** 137 mM NaCl, 2.7 mM KCl, 20mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub>, pH 7.4

**TBS:** 25 mM Tris-HCl, 140mM NaCl, 3mM KCl, pH 8.0

**TBST:** TBS + 0.05% (v/v) Tween 20

**TAE:** 40 mM Tris-Acetate, 1mM EDTA, pH 8.3

## Media

Media were prepared with demineralized H<sub>2</sub>O and autoclaved after preparation.

LB medium:

10 g/l Tryptone, 5 g/l Yeast Extract, 5 g/l NaCl, (+ 15 g/l agar for solid medium) are dissolved in  $\text{d}_0\text{H}_2\text{O}$  and adjusted to pH 7.0 with NaOH (Sambrook *et al.*, 1989).

Medium 120a (modified):

K <sub>2</sub> HPO <sub>4</sub>	0,348g
KH <sub>2</sub> PO <sub>4</sub>	),227g
NH <sub>4</sub> Cl	0,5g
MgCl <sub>2</sub> x7H <sub>2</sub> O	0,5g
CaCl <sub>2</sub> x2H <sub>2</sub> O	0,25g
NaCl	2,25g
FeSO <sub>4</sub> x7H <sub>2</sub> O	0,002g
Vitamin solution (141)	10ml
Trace element solution SL-10	1ml
Resazurin	0,001g
NaHCO <sub>3</sub>	3g
2.5 M sodium acetate	10ml
Methanol	10ml
Na <sub>2</sub> Sx9H <sub>2</sub> O	0,5g
$\text{d}_0\text{H}_2\text{O}$	ad 1000 ml

The medium is prepared anaerobically under a N<sub>2</sub>/CO<sub>2</sub> (80:20, 100 kPa) atmosphere, adjusted with acetic acid to pH of 6.5- 6.8. Methanol (50% v/v) and 10 g/l of sodium acetate are heat sterilized separately under N<sub>2</sub> atmosphere in tightly closed tubes. Appropriate volumes of the solutions are injected into the autoclaved medium with hypodermic syringes.

Vitamin Elixir (Balch et al., 1979, modified):

D (+) Biotin	2.00 mg
Folic acid	2.00 mg
Pyridoxamindihydrochloride	10.00 mg
Thiaminedihydrochloride	5.00 mg
Riboflavine	5.00 mg
Nicotinic acid	5.00 mg
Calcium- D- pantothenate	5.00 mg
Cyanocobalamine	0.10 mg
Para-aminobenzoic acid	5.00 mg
DL- Lipoic acid	5.00 mg
$\text{ddH}_2\text{O}$	ad 1000 ml

The vitamin solution is filter sterilized (0.2  $\mu\text{m}$ ) and kept in dark bottles at 4°C.

Trace Element Solution SL-10:

HCl (25%; 7.7 M)	10 ml
$\text{FeCl}_2 \times 4 \text{H}_2\text{O}$	1.5 g
$\text{ZnCl}_2$	70 mg
$\text{MnCl}_2 \times 4 \text{H}_2\text{O}$	100 mg
$\text{H}_3\text{BO}_3$	6 mg
$\text{CoCl}_2 \times 6 \text{H}_2\text{O}$	190 mg
$\text{CuCl}_2 \times 2 \text{H}_2\text{O}$	2 mg
$\text{NiCl}_2 \times 6 \text{H}_2\text{O}$	24 mg
$\text{Na}_2\text{MoO}_4 \times 2 \text{H}_2\text{O}$	36 mg
$\text{ddH}_2\text{O}$	ad 1000 ml

First  $\text{FeCl}_2$  is dissolved in the HCl. After dilution in  $\text{ddH}_2\text{O}$  the other salts are added, finally  $\text{ddH}_2\text{O}$  is added to the final volume.

## *III.2 Anaerobic cultivation*

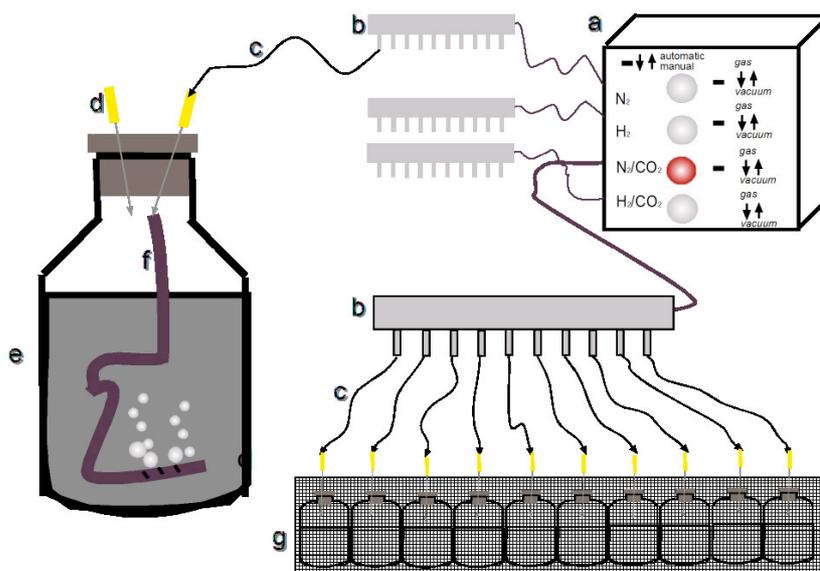
### *III.2.1 Set up of an anaerobic cultivation system*

#### Gas distribution system

For an even distribution of the appropriate gas environment on the media a gas distribution system for 10 gas inflow/evacuation sources was designed (Figure 12). Therefore an electronic relay was constructed to control gas inflow or the connection to a vacuum pump to the medium bottles. The relay allows not only manual switch between gas inflow or vacuum position, the `automatic` option starts a series of three evacuation steps followed by gas inflow in which both steps a variable in time spans from 30s to 5 min. The pressure of the gas inflow is adjusted at the manometer at the gas source tubing and can be checked at the gas rake manometer.

For bubbling of media a rubber hose with a perforated ending is connected through a butyl rubber plug to a needle of the gas distribution (N<sub>2</sub>, 120-140 kPa). The rubber hose is placed into the dissolved medium in 1 l glass bottle and before plugging of the bottle a counter needle is placed into the rubber to provide gas out flow.

To distribute the appropriate gas phase on the media, the sealed serum bottles are placed in to a protective metal cavity and the gas phase is supplied with 200 kPa. An automated series of three 5 min evacuation steps followed by 1 min gas inflow steps guarantees a complete exchange of the appropriate gas against the chamber gas.



**Figure 12 : Gas Distribution System**

The electronic relay (a) provides manual options between gas/vacuum and an automated option for a series of three step of evacuation/gas flow-in. The gas is led through the distribution rake (b) with rubber hoses ending in injection needles (c) either into another rubber hose (f) in a 1 l serum bottle containing the dissolved medium compounds (e) to bubble the solution and is released through the safety needle (d). Or the gas is directly led onto the gas phase of the reduced and pH- adjusted media in sealed bottles, the automated gas exchange program finally provides a homogenous gas phase at the adjusted pressure.

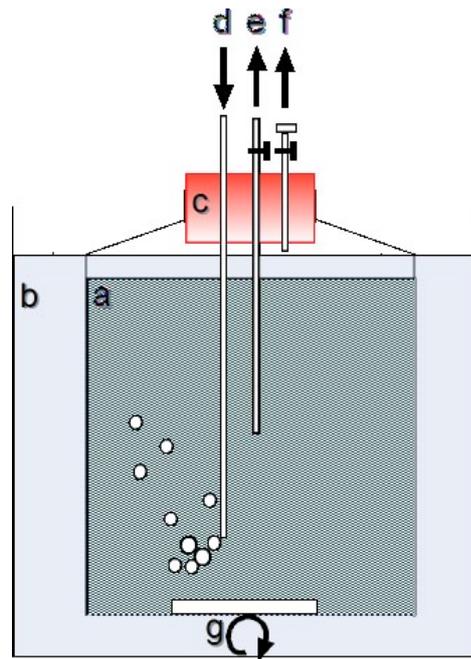
### III.2.2 Anaerobic cultivation

#### *Cultivation in medium bottles*

The dissolved components are bubbled with 120-140 kPa N<sub>2</sub> for 30 min and subsequently reduced with 0.5 mg/l Na<sub>2</sub>S x 7-9 H<sub>2</sub>O. After pH adjusting the medium is aliquoted into serum bottles inside the anaerobic chamber, sealed with butyl rubber plugs (Hungate *et al.*, 1966) and fused with aluminium rings, finally the indicated gas phase is applied. After autoclaving the medium is ready for injection of sterile substrates. Cells were routinely inoculated in a 1:10 - 1:50 ratio.

### *Cultivation in a Glass Fermentor*

The medium components are dissolved in the 10 l glass fermentor, bubbled with N<sub>2</sub> for 30 min, sealed with a rubber plug and autoclaved. After connection to the cultivation system supplying a constant and low gas flow rate, the medium is reduced and adjusted to the final pH. After addition of substrates the cells are inoculated in a 1:10 - 1:50 ratio. Stirring at a speed of 5-10 rpm provides even distribution of gas (Figure 13).



**Figure 13 : Glass Fermentor**

The temperature of the fermentation system is externally regulated by a water bath. To provide anaerobic conditions the glass fermentor (a) is sealed with a butyl rubber. The rubber plug holds perforations for glass capillaries to provide gas inflow from the gas source tubing (d), gas outflow (f) and sample collection (e). g: stirring device

### *Large scale cultivation in a Teflon Fermentor*

Large scale fermentation was performed at the Archaeen Zentrum, University of Regensburg. The medium compounds were dissolved inside the fermentor and bubbled with N<sub>2</sub> during the heating up phase to 121°C, then autoclaved at 121°C/200 kPa for 40 min. After reaching the incubation temperature of 35°C the pH was adjusted through a septum with sterile,

anaerobic acid. Before inoculation with 10-15 l of culture sterile, anaerobic substrate was injected.

Cells were harvested anaerobically after shock cooling to 4°C (Durchlaufkühler, Feichtenschläger, Neusäss) through a flow centrifuge (Type 416, Padberg, Lahr) and a maximal speed of 8 l/h. The cell suspension then was directly used for co-immunoprecipitation experiments or shock frozen in liquid nitrogen and kept at -80°C.

### *III.2.3 Sterilisation*

Media and stock solutions are autoclaved for 20min at 121°C/200 kPa. Syringes, glass equipment etc. are sterilized for 40 min at 121°C/200 kPa. Organic compounds and heat labile substances are filter sterilized (0.2 µm).

### *III.2.4 Light microscopy*

Cell growth was monitored by light microscopy with a phase contrast microscope or by fluorescence microscopy using the filter set FS 01 (excitation at 365/12 nm). If necessary the cell number was determined in a Thoma cell counting chamber (cell depth 0.02 mm, length 0.0025mm).

### *III.3 Molecular biochemical methods*

#### *III.3.1 Preparation and transformation of E. coli*

For preparation of chemically competent *E. coli* cells, LB medium (including antibiotic, if applicable) was inoculated with a single colony and cells were grown to an optical density (OD<sub>600nm</sub>) of 0.25 - 0.5. The culture was then chilled on ice for 15 min and centrifuged at 1500 g for 15 min at 4 °C. The cell pellet was resuspended in one third of the original volume of competence buffer I and incubated on ice. After 1 h the cells were pelleted and resuspended in 1/25 of the original volume competence buffer II, incubated on ice for 15 min and shock-frozen in liquid N<sub>2</sub> in 20-100 µl aliquots.

For transformation, ~50 µl competent cells were mixed with 0.05 - 0.2 µg plasmid DNA or 1-5 µl ligation reaction and incubated on ice for 30 min. The cells were heat shocked at 42 °C for ~ 40 s and subsequently placed on ice for 2 min. 450 µl of LB medium was added and the cells suspension was shaken for 1 h at 37 °C. The cells then were plated on selective solid media and incubated at 37 °C, until colonies had developed (typically 10-16 h).

#### *III.3.2 DNA analytical methods*

DNA concentrations were measured by UV absorption spectroscopy at  $\lambda=260$  nm. A solution of 50 µg/ml of double stranded DNA in H<sub>2</sub>O exhibits approximately  $A_{260nm} = 1$ .

Agarose gel electrophoresis was performed in TAE buffer and 1 – 2 % TAE-agarose gels, supplemented with 1 µg/ml ethidium bromide, at 4 – 6 V/cm.

DNA sequencing was performed by Medigenomix GmbH (Martinsried, Germany) or Sequiserve (Vaterstetten, Germany).

#### PCR amplification

PCR (polymerase chain reaction) mediated amplification of DNA was performed according to the following protocol:

---

DNA template	250 ng or less (bacterial genomic DNA) 20 ng (plasmid DNA)
Primers	20 pmol each
dNTPs	200 $\mu$ M each
<i>Taq</i> or <i>Pfu</i> Polymerase	2.5 U
Polymerase buffer	5 $\mu$ l
Additives	4 % DMSO if GC content was >50 %, 7 % DMSO if GC content was >60 %, MgCl <sub>2</sub> (3 $\mu$ l), if <i>Taq</i> Polymerase was used.
Final volume	50 $\mu$ l

### III.3.3 DNA restriction/ligation methods

DNA restriction was performed according to product instructions of the respective enzymes. Typically, a 50  $\mu$ l reaction contained 1-2  $\mu$ l of each restriction enzyme and 30  $\mu$ l purified PCR product or 1-5  $\mu$ g plasmid DNA in the appropriate reaction buffer. Digested vector DNA was dephosphorylated with shrimp alkaline phosphatase. For ligation, 50-100 ng (~1-2  $\mu$ l) dephosphorylated vector DNA, 200-300 ng (~5-10  $\mu$ l) DNA insert and 1  $\mu$ l (100 U) T4 ligase were incubated in ligase buffer at 25 °C for 1 h or, for increased efficiency, at 16 °C overnight and transformed into competent *E. coli* DH5 $\alpha$  cells.

### III.3.4 Plasmid purification

LB medium containing the appropriate antibiotic was inoculated with a single *E. coli* colony harbouring the DNA plasmid of interest and shaken 8 – 14 h at 37 °C. Plasmids were isolated using the Promega Wizard Plus SV Miniprep or Midiprep according to the instructions.

## III.4 Protein biochemical methods

### III.4.1 Protein expression and purification

For protein expression *E. coli* BL21 (DE3) Gold cells harbouring the indicated expression plasmid, grown at 37 °C in 6 l LB medium containing 100 mg/l ampicillin were induced with 1 mM IPTG at an OD<sub>600</sub> ~0.6 for 5 – 6 h and harvested by centrifugation for 30 min at 2500 g. Cells were resuspended in 50 mM Tris-HCl pH 7.8, 50 mM NaCl, 1mM EDTA and Complete protease inhibitor (1 tablet/25 ml). The suspension was frozen in liquid N<sub>2</sub> and thawed before addition of lysozyme (~0.5 mg/ml) and Benzonase (~200 units). Lysis was achieved by homogenization of the cell suspension in an EmulsiFlex C5 device kept on ice. Cell debris was removed by ultracentrifugation for 30 min at 4 °C and ~100 000 g and subsequent filtration (0.2 µm). All protein purifications steps were performed using a FPLC system at 4 – 8 °C unless stated otherwise. Every column used was equilibrated in the indicated running buffer prior to loading the protein solution. After every separation step on a column the eluted fractions were analyzed by SDS-PAGE visualized by coomassie staining. The purified proteins were aliquoted and shock frozen in liquid nitrogen and kept at -80°C.

#### *MmGroEL*

The cell lysate was applied to a DE52-column (Whatman) and washed with two column volumes of 30mM Tris-HCl pH 7.6, 1 mM DTT. The bound proteins were eluted with a NaCl-gradient in above described buffer. The collected fractions were dialysed against 25 mM Histidin-HCl pH 5.8, 1 mM EDTA, 1 mM DTT o.n. and supplied to a 200ml Source30Q-Säule (Pharmacia). After washing with two column volumes running buffer the bound proteins were eluted by a NaCl-gradient from 10-500 mM NaCl. After an overnight dialysis of the collected fractions against 30 mM Tris-HCl pH 7.8, 1 mM DTT the protein solution was separated on a heparin-HiTrap-column followed by a NaCl gradient (0-1M NaCl) dependent elution in 30 mM Tris-HCl pH 7.8, 1 mM DTT. As a final step the proteins were separated according to their molecular size on a gel-filtration column (S-300 HR 26/60, Pharmacia) in 20 mM MOPS-NaOH

pH 7.4, 100 mM NaCl, 10% Glycerol. Fractions containing *MmGroEL* or *MmGroES* were collected, concentrated using a Centriprep concentrator, frozen in aliquots in liquid N<sub>2</sub> and stored at -80 °C.

**GroEL-D87K (GroEL-Trap)** (Fenton *et al.*, 1994; Weissman *et al.*, 1994) was purified with modifications to the purification protocol of wild type GroEL.

Additional MonoQ anion exchange chromatography steps were introduced between heparin and final size exclusion chromatography. GroEL was eluted from MonoQ in 50 mM Tris pH 7.8 and 7.1, using very shallow NaCl gradients at GroEL elution concentrations.

#### *MmThs* $\alpha$ , $\beta$ , $\gamma$

The clarified lysate was supplied to a Source30Q I<sub>o</sub> exchange column and the bound proteins were eluted in a gradient from 50-500 mM NaCl in 30 mM Tris-HCl, pH 8,0, 2 mM EDTA, 5 mM  $\beta$ -Mercaptoethanol. Monomeric *MmThs* were released from the column at ~180 mM NaCl while oligomeric thermosomes eluted at a ionic strength of ~250 mM NaCl. The elution fractions containing monomeric thermosome subunits were concentrated using a Centriprep concentrator and six-fold diluted in buffer (30 mM Tris-HCl, pH 7,5, 5 mM  $\beta$ -Mercaptoethanol), separated on four consecutive Heparin-Sepharose HiTrap columns and eluted in a NaCl gradient from 0.05-1.0 M in 30 mM Tris-HCl, pH 7,5, 5 mM  $\beta$ -Mercaptoethanol). Fractions containing thermosome subunits were collected, concentrated and supplied to a final purification step on a size exclusion column (S300 HR column) in 30 mM Tris-HCl, pH 7,6, 50 mM NaCl, 10 % Glycerol. Fractions containing thermosome subunits were collected, concentrated using a Centriprep concentrator, frozen in aliquots in liquid N<sub>2</sub> and stored at -80 °C.

#### *MmPrefoldin*

The clear cell lysate was loaded onto a Sepharose-30Q and bound proteins were eluted in a NaCl-Gradienten of 10-800 mM NaCl in running buffer (30 mM Tris-HCl, pH 7,6, 1 mM DTT). The elution fractions containing *MmPfd* subunits were equilibrated overnight against buffer (25

mM Histidin-HCl, pH 5,8, 30 mM NaCl) and put on a Heparin-Sepharose HiTrap column and the bound proteins were eluted in a 10-600 mM NaCl gradient in buffer (30 mM Tris-HCl, pH 7,6, 1 mM DTT). The protein solution was dialysed o.n. in 50 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7,0, 1,5 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and supplied to a Phenylsepharose column where bound proteins were eluted in a (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> gradient from 1,5 M-0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> in buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.0). During the concentration step pure *MmPfd* precipitated and was resuspended over night in buffer (100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10 % Glycerol). Purified *MmPfd* complex was aliquoted and kept at -80°C.

#### *MmDnaK-(His)<sub>6</sub>*

Cell lysate from cells harboring the plasmid pProEX-MmDnaK-N(His)<sub>6</sub> was prepared as described above, but EDTA was omitted from all solutions. The lysate was diluted with running buffer (20 mM HEPES, pH7.4 500mM NaCl) and applied to a 8 ml Ni-NTA column using fresh or regenerated resin. The column was washed with 5 volums of running buffer containing 10 mM imidazole. Elution was achieved by applying a shallow imidazole gradient from 50 to 250 mM. Fractions containing DnaK-N(His)<sub>6</sub> were pooled and applied to a MonoQ ion exchange chromatography column (20 mM HEPES, pH 7.4, NaCl gradient from 0.05 to 0.5 M NaCl) followed by size exclusion chromatography in 20 mM HEPES, 50 mM NaCl and 10 % glycerol and a final concentration step.

### *III.4.2 Protein analytical methods*

#### *Determination of the protein concentration*

Protein concentrations of purified proteins were determined spectroscopically at  $\lambda = 280$  according to the theoretical extinction coefficient (calculated by ProtParam tool at the ExPASy proteomics server <http://www.expasy.org>). The protein concentration of cell lysates or complex protein solutions was estimated according to the Bradford-method (Bradford 1976) using the BioRad Protein Assay.

### *SDS polyacrylamide gel electrophoresis (SDS-PAGE)*

For analysis of proteins and complex protein solutions the proteins were separated electrophoretically according to their molecular weight by SDS-Polyacrylamide gelelectrophoresis (PAGE, Laemmli 1970). While the stacking gel was kept at a concentration of 30%/0.8% - Acrylamide/Bisacrylamide in the separation gel varied according to the molecular size of the proteins of interest between 8 and 16%. SDS-PAGE was performed using a discontinuous buffer system (Laemmli, 1970) in BioRad Mini-Protean II or 3 electrophoresis chambers employing a constant voltage of 160-220 V in 50 mM Tris-Base, 380 mM glycine, 0.1 % SDS (pH 8.3). SDS loading buffer was added to the protein samples. Polyacrylamide gels were fixed and stained in 0.1 % Coomassie brilliant blue R-250, 40 % ethanol, 7 % acetic acid for 1 h or longer and destained in 20 % ethanol, 7 % acetic acid for removal of unspecific background stain.

For a better separation of *MmThs* $\beta$  and *MmThs* $\gamma$  subunits the pH of the separating gel was set to 7.5.

### *2D- polyacrylamide gel electrophoresis (2D-PAGE)*

Complex protein samples were separated by two dimensional PAGE. Therefore the TCA precipitated and acetone washed protein sample is resuspended in a final volume of 340  $\mu$ l 7M urea, 2M thiourea, 4% CHAPS, 0.002% Bromphenol blue, 0.5% IPG buffer (ampholytes), 20mM DTT and soaked for at least 10 hrs into a Immobiline Dry strip, 18cm, pH 3-10, NL (Amersham Biosciences). The proteins were focused to their isoelectric point by applying following voltage steps on the Ettan IPGphor II system (Amersham Biosciences):

step	500V	1hr
gradient	1000V	1hr
gradient	8000V	3hrs
Step	8000V	4hrs

After reduction in 65mM DTT, 6M urea, 1%SDS, 26% glycerol, 50mM Tris, pH 8.8 and acylation in 135mM iodoactedamide, 6M urea, 1%SDS, 26% glycerol, 50mM Tris, pH 8.8 the

immobiline strips were placed on SDS-16% Polyacrylamide gels and fixed by overlaying with 0.5% agarose, 0.002% bromphenol blue in SDS-PAGE running buffer. The size dependent separation was performed on the ETTAN DALT twelve system separation unit (Amersham Biosciences) for 30 min at 2.5 W per gel followed by a 17 W per gel run for four to seven hours.

### *Ettan DIGE*

For quantitative comparison up to three different protein samples were fluorescently labelled with distinct CyDye DIGE Fluor minimal dyes following the instruction manual, pooled and subjected to 2D-GE.

To allow picking of the protein spots by the Ettan Spot Picker the polyacrylamide gel is covalently attached on the glass plates. Therefore the glass plate is evenly coated with 4ml bind-silane solution (0.1% bind-silane, 80% Ethanol, 2% HOAc). The silane solution is dried on the glass plate for 45 min (fume hood) and excess of the solution is removed by careful washing with conventional scavenger and subsequent rinsing with ddH<sub>2</sub>O. To allow orientation of the spot picker reference markers are positioned on the glass plate as indicated in (Figure 14). The silanized glass plates are immediately used for pouring of SDS-PA gels to prevent evaporation of the Bind-Silane from the coated glass plate to the covering glass plate.



**Figure 14 : Position of Reference Markers on Ettan DIGE**

The reference markers (grey spots) are placed at positions on the gel, where they will not interfere with the pattern of the protein spots in the gel (the Immobiline strip is shown in red).

### *Silver staining*

Silver staining of proteins was employed to detect proteins of low amount (1ng-200ng) after SDS-PAGE. Immediately after electrophoresis the gels were placed into fixing solution for 30 min to precipitate the proteins and to allow diffusion of SDS out of the gel. Then gels were transferred into incubation solution for 30 min to oxidize the proteins. After three washing steps for 5 min in <sub>dd</sub>water the silver solution is poured on the gel and incubated for 40 min. Finally the protein spots are visualized in developing solution, the reaction is stopped by addition of 5mM EDTA, pH 8.0. The gels are kept in 10% glycerol at least 30 min and can be dried then.

Fixing solution	400 ml ethanol 100 ml acetic acid ad 1l H <sub>2</sub> O
Incubation solution	300 ml ethanol 68 g sodium acetate x 3 H <sub>2</sub> O 2g sodium thiosulphate x 5 H <sub>2</sub> O ad 1l H <sub>2</sub> O
silver solution	1g silver nitrate ad 1l H <sub>2</sub> O 250 µl formaldehyde, prior to usage
Developing solution	26 g sodium carbonate ad 1l H <sub>2</sub> O 125 µl formaldehyde, prior to usage

### *III.4.3 Polyclonal antibodies*

Rabbit and goat polyclonal antibodies were generated at the animal facilities of the MPI for Biochemistry according to Harlow and Lane (1988). Purified proteins were injected subcutaneously as water in oil emulsion formed out of 1 volume of protein solution (~0.2 – 0.5 mg, rabbit; 0.5-1mg, goat) in PBS and 1 volume Freund's adjuvant (Freund and McDermot, 1942). Complete Freund's adjuvant was used for the initial immunization and incomplete Freund's adjuvant for 4 – 6 succeeding boosts, which were injected at intervals of 4 – 7 weeks. Serum was taken ~10 days post injection.

### *III.4.4 Affinity purification of antibodies*

As described in the manual 9mg of antigen was attached to Affigel 15 (Biorad). Several times 10ml of serum was passed through the antigen column and then washed with 20 bed volumes of 10mM Tris, pH 7.5 and 20 bed volumes of 10mM Tris, pH 7.5, 500 mM NaCl. The bound antigen was eluted in 10 bed volumes of 100mM glycine, pH 2.5 and the eluate was immediately neutralized by addition of 100  $\mu$ l 1M TRIZMA to 1ml. The column was then wash with 10mM Tris/HCl, pH 8.8 followed by 10mM Tris/HCl, pH 7.5 until the pH of the buffer was reached in elution. From this step either again serum was headed onto the column or the column was stored in 10mM Tris/HCl, pH 7.5, 0.01% merthiolate at 4°C. The eluted fractions were analysed by SDS-PAGE and the antibody containing fractions were pooled and concentrated.

Antibodies that were used for co-immunoprecipitations were purified of antibodies that potentially reacted with the "negative antigen" (the  $\alpha$ -*MmGroES*- and the  $\alpha$ -*MmGroEL* antibodies against thermosome  $\alpha, \beta, \gamma$ ; the  $\alpha$ - *MmThs* $\alpha$  and *MmThs*  $\beta$  against *MmGroEL* and *MmGroES*; the  $\alpha$ - FFLuc antibody against any chaperonin subunit). Therefore the serum was passed several times through a "negative antigen" column and the run though was kept as negative-purified antibody serum, the washing steps were neglected and the column was regenerated as described above.

### *III.4.5 Western blotting*

Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes in a semi-dry western blotting unit (SemiPhore) in 25 mM Tris, 192 mM glycine, 20 % methanol, pH 8.4 at constant current of ~0.5 – 0.8 mA/cm<sup>2</sup> gel size for 1.5 h (Towbin *et al.*, 1979).

Nitrocellulose membranes were blocked in 5 % skim milk powder in TBST for 1 h. The membranes were then incubated with a 1:2000 – 1:10000 dilution of primary antibody serum in TBST and extensively washed in TBST before incubation with a 1:5000 dilution of secondary antibody in TBST (Anti-rabbit IgG, whole molecule – horseradish peroxidase conjugate. Antibody produced in goat or *vice versa*. Sigma). After extensive washing, protein bands were detected by incubating the membranes with ECL chemiluminescence solution and exposure to X-ray film or LAS-3000 CCD-Imaging System.

### *III.4.6 Preparation of antibody beads*

Serum was mixed with in H<sub>2</sub>O equilibrated protein A beads and gently rocked for 2 hrs at room temperature. The beads were transferred to a poly-Prep column and solution was passed through by gravity force. The beads were washed with 10 bed volumes of 100mM Tris/HCl, pH 8.0 and equilibrated in 0.2 M Triethanolamine, pH 8.9. The antibody was crosslinked by addition of 10 fold molar excess of dimethylpimelimidate (DMP) for 1 hr at RT. The reaction was stopped by addition by a washing step in 5 bed volumes of 0.1M glycine, pH 2.5 and neutralized by washing TBS. The beads were stored after addition of 0.01% merthiolate at 4°C.

### *III.4.7 Analytical gel filtration*

Protein samples were applied to a size-exclusion column (Superdex 200 or Superose 6, Pharmacia) in aliquots of 50-100 µl, at room temperature with a flow rate of 40 µl/min and fractions of 100 200 µl were collected. The columns were calibrated with thyroglobulin (669 kDa), ferritin (460 kDa), catalase (206 kDa), aldolase (158 kDa), BSA (67 kDa), carbon-anhydrase (29 kDa) und α-lactalbumin (14 kDa) (Pharmacia) as proteins size standards.

### *III.5 Functional analysis methods*

#### *III.5.1 Co-immunoprecipitations*

Exponentially grown *M. mazei* cells were shock cooled by a radiator coil and harvested at 3800 rpm/4°C. The cell pellet was lysed in 25mM Tris/HAc pH 7.4, 20mM glucose, 100 U hexokinase/ml, 0.01% Tween 20, 250 U benzonase, 2 mM Pefabloc and adjusted to 150mM NaCl. To remove cell debris the solution is spun twice at 21000 x g for 15 min at RT. After estimation of the protein concentration using the Bradford method antibody-bead solution is added in a 1:5 ratio of antibody and antigen, assuming a 1% abundance of the chaperonins in the cell lysate. After gentle rocking for 1.5 - 2hrs at RT the beads are washed with 200 bed-volumes of 50 mM Tris/HCl pH 8.0, 1% Triton-X-100, 0.5% DOC, 150mM NaCl, 5mM ADP 200 bed- volumes of 50 mM Tris/HCl pH 8.0, 1% Triton-X-100, 500 mM NaCl, 5mM ADP and 200 bed- volumes of TBS, 5mM ADP. The bound antigen is eluted from the beads in 5 bed-volumes of 100mM glycine pH 2.5.

For immunoprecipitation at denaturing conditions the lysate was denatured in 1% SDS at 95°C for 1 min. After 20 fold dilution into 25mM Tris/HAc pH 7.4, 0.01% Tween 20, 250 U benzonase, 2 mM Pefabloc, adjusted to 150mM NaCl and spun twice at 21.000 x g for 15 min at 4°C. The clarified solution was then subjected to the immunoprecipitation experiment described above, but at 4°C to prevent reassociation of the chaperonin subunits into functional complexes.

For immunoprecipitation from denatured lysate xxx mg of cell pellet was resuspended in 3M guanidine/HCl. After centrifugation the lysate was diluted 50-fold into 25mM Tris/HAc pH 7.4, 0.01% Tween 20, 250 U benzonase, 2 mM Pefabloc containing the indicated recombinant chaperonins. After adjusting to 150mM NaCl and two spins at 21.000 x g for 15 min at RT, the immunoprecipitation protocol above was followed.

### *III.5.2 Substrate release experiments*

The co-immunoprecipitation experiment was performed as described above, but before elution of the antigen the beads were transferred into refolding buffer (100mM KCl, 5mM MgCl<sub>2</sub>, 500mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20mM MOPS/7.4). After addition of 5mM TRAP-EL at 35°C, the experiment was started with 5mM ATP and time points were taken by stopping depleting ATP from the reaction by addition of 100 U/ml hexokinase, 20mM glucose. After collection of the end time point the samples were washed with 100 bed-volumes of each washing buffer and eluted in 5 bed volumes of 0.1M glycine, pH2.5.

### *III.5.3 PK digestion*

*MmGroEL/ES*/substrate complexes were prepared as described above and immediately neutralized by addition of equal amount of 1M Tris/HCl, pH 8.0. After estimation of the protein concentration according to the Bradford method Proteinase K was added at 25 °C in a 1:20 weight ratio of Proteinase K: immunoprecipitate. At indicated times, samples were taken and the digestion was stopped with 10 mM PMSF.

## *III.6 Biophysical methods*

### *III.6.1 Mass spectrometry of protein spots from 2D PAGE*

Protein spots of interest were picked using the Ettan spot picker. The mass spec analysis was performed by Magda Puype and Evy Timmermann at the Department of Medical Protein Research (VIB09), Ghent University.

The gel spots were transferred to a Biopure Eppendorf tube and washed with water (LC-MS graded, Biosolve, Valkenswaard, Netherlands), then with 50% (v/v) acetonitrile (HPLC graded, Baker, Amsterdam, Netherlands) in water and finally with 100% acetonitrile. The gel spots were submerged in 50 mM ammonium bicarbonate at pH 8.0 and 0.05 µg sequencing grade modified trypsin (Promega Corporation, Madison, WI, USA) was added. After overnight digestion at 37°C, the mixture was acidified with formic acid to pH 2-3 to deactivate the trypsin.

The peptide mixture was then centrifuged, transferred to a new Eppendorf tube, dried and re-dissolved in 20  $\mu$ l of 0.1% formic acid in 2/98 (v/v) acetonitrile/water (solvent A). 10  $\mu$ l of this peptide mixture was applied for nano-LC-MS/MS analysis on an Ultimate (Dionex, Amsterdam, The Netherlands) in-line connected to an Esquire HCT mass spectrometer (Bruker, Bremen, Germany). The sample was first trapped on a trapping column (PepMap™ C18 column, 0.3 mm I.D. x 5mm, Dionex (Amsterdam, The Netherlands)). After back-flushing from the trapping column, the sample was loaded on a 75  $\mu$ m I.D. x 150 mm reverse-phase column (PepMap™ C18, Dionex (Amsterdam, The Netherlands)) The peptides were eluted with a linear gradient of 3% solvent B (0.1% FA in water/acetonitrile (3/7, v/v)) increase per minute at a constant flow rate of 0.2  $\mu$ l/min. Using data dependent acquisition, the multiply charged ions with intensities above threshold (adjusted for each sequence according to the noise level) were selected for fragmentation. During MS/MS analysis a scan time of 40 ms was used and MSMS fragmentation amplitude ramped from 30% to 300% of 0.7V.

The fragmentation spectra were converted to mgf files using the Automation Engine software (version 3.2, Bruker) and were searched using the MASCOT database search engine (<http://www.matrixscience.com>) against the *Methanosarcina mazei* Gö1 database derived from the NCBI database. Peptide mass tolerance was set at 2 Da and peptide fragment mass tolerance at 0.5 Da, with the ESI-IT as selected instrument for peptide fragmentation rules. Variable modifications were set to methionine oxidation, pyro-glutamate formation of amino terminal glutamine, acetylation of the N-terminus and propionamide modification of cysteines.

### *III.6.2 Coupled liquid chromatography – mass spectrometry system (LC-MS/MS)*

Mass spectrometry and data analysis of the mass spectrometric output was performed by Morten Kirkegaard (Center for Experimental Bioinformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark – Odense) and Dr. Francesca Forner (Department for proteomics and signal transduction, MPI of Biochemistry, Martinsried).

### *Sample preparation*

200 µg of the respective *MmGroEL/MmGroES*-substrate or *MmThs*-substrate complexes were separated by SDS-PAGE (4-16 % BisTris-NuPage, 1 mm, 200 V for 35 min, Invitrogen). The gel was Coomassie stained (colloidal blue staining kit, Invitrogen) and the entire lanes were cut out and sliced into 5-6 pieces. After in-gel reduction, alkylation the proteins were in gel digested with trypsin as described (Lasonder *et al.*, 2002). After extraction of peptides from gel pieces using 3 % trifluoroacetic acid (TFA) and 30 % acetonitrile, the sample volume was partially reduced by vacuum evaporation and the residual solutions were applied to StageTips to desalt, filtrate and concentrate the peptide samples (Rappsilber *et al.*, 2003).

### *Coupled liquid chromatography – mass spectrometry system (LC-MS/MS)*

For the *in gel* separation approach, 200 µg of proteins (Bradford assay) were loaded onto a 4-12% gradient polyacrylamide pre-cast gels (NuPAGE Novex Bis-Tris gels, 1 mm, Invitrogen) and stained with colloidal Coomassie (Invitrogen). Gel lanes were cut into 6 slices and subjected to *in gel* tryptic digestion essentially as described (Schevchenko *et al.*, 1996). Briefly, following complete de-staining, gel slices were cut, washed with 50 mM ammonium bicarbonate and shrunk with ethanol. Reduction/alkylation of proteins was performed with 10 mM DTT and 55 mM iodoacetamide. After two wash steps with ammonium bicarbonate/ethanol, the gel was dried with ethanol and incubated with 12.5 ng/µl trypsin (Promega) in 50 mM ammonium bicarbonate at 4°C for 15 min. The supernatant was then discarded and replaced with 50 mM ammonium bicarbonate and the reaction allowed proceeding overnight at 37°C. The reaction was stopped with 1% TFA, 0.5% acetic acid and 3% acetonitrile and the supernatant recovered. Additional peptide extraction steps were performed with 30% acetonitrile and 100% acetonitrile. Supernatants were concentrated to low volume and then diluted to approximately 200 µl with 0.5% acetic acid, 3% acetonitrile, 1% TFA. Resulting tryptic peptides were desalted and concentrated on reversed phase C<sub>18</sub> StageTips (Rappsilber *et al.*, 2003). Peptides were resuspended in 10 µl of 3% acetonitrile, 1% trifluoroacetic acid, 0.5% acetic acid before injection. Liquid chromatography was performed on a 15 cm fused silica emitter (75 µm ID from

Proxeon Biosystems, Odense, Denmark) packed in-house with reversed phase ReproSil-Pur C18-AQ 3  $\mu\text{m}$  resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The injection volume was 5  $\mu\text{l}$ , and the flow rate was 250 nL/min after a tee splitter. The experiments were performed on an Agilent 1100 nanoflow system connected to a 7-Tesla Finnigan linear quadrupole ion-trap Fourier transform (LTQ-FT) mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark).

The peptide mixtures were injected onto the column with a flow of 500 nL/min and subsequently gradient-eluted with a flow of 250 nL/min from 5% to 40%  $\text{CH}_3\text{CN}$  in 0.5% acetic acid. Gradients were 140 minutes long. The mass spectrometer was operated in the data dependent mode to automatically switch between MS,  $\text{MS}^2$  and  $\text{MS}^3$  acquisition essentially as described (Olsen and Mann, 2004). Survey full scan MS spectra (from  $m/z$  350 – 1550) were acquired in the ICR cell with resolution  $R=25,000$  at  $m/z$  400 and accumulation to a target value of 5,000,000 charges. The three most intense ions were sequentially isolated for accurate mass measure in the FT-ICR cell (SIM) and fragmented in the linear ion trap by collisionally induced dissociation at a target value of 10,000 charges. For  $\text{MS}^3$ , the three most intense ions in each  $\text{MS}^2$  spectrum were further isolated and fragmented. Target ions already selected for MS/MS were dynamically excluded for 30 seconds. General mass spectrometric conditions were: electrospray voltage, 2.3 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 150°C; collision gas pressure, 1.3 mTorr; normalized collision energy, 35% for  $\text{MS}^2$ . Ion selection threshold was 500 counts for  $\text{MS}^2$ . An activation  $q = 0.25$  and activation time of 30 ms was applied for  $\text{MS}^2$  acquisitions.

Proteins were identified by automated database searching (Mascot Daemon, Matrix Science) against an in-house curated version of the *Methanosarcina Mazei* protein sequence database. The database was obtained from [ftp://ftp.ebi.ac.uk/pub/databases/integr8/fasta/proteomes/89.M\\_mazei.fasta.gz](ftp://ftp.ebi.ac.uk/pub/databases/integr8/fasta/proteomes/89.M_mazei.fasta.gz) and was complemented with frequently observed contaminants (porcine trypsin, achromobacter lyticus lysyl endopeptidase and human keratins). Search parameters specified an initial MS tolerance of 10 ppm and an MS/MS tolerance at 0.4 Da and Trypsin/P+DP specificity allowing for up to 2

missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionines, N-protein acetylation and N/Q deamidation were allowed as variable modifications. The 99 percent significance threshold in the *Methanosarcina mazei* Gö1 database search was a Mascot score of 19. Acceptance criteria for protein identifications were at least one sequenced non redundant peptide with mass accuracy better than 5 ppm, Mascot score higher than 19 and MS<sup>3</sup> event. A total of 4926 unique peptide sequences with more than 5 amino acids in length and mass accuracies better than 5 ppm were identified.

#### Semi-quantitative protein abundance calculation using emPAI scores

To estimate the absolute protein content in the different samples we calculated emPAI scores (Ishihama *et al.*, 2005) based on the number of unique peptides. emPAI is defined as  $10^{\text{PAI}} - 1$ , where PAI is the ratio between the number of observed and observable peptides per protein. It has been shown that there is a roughly linear relationship between protein concentration and emPAI score, therefore the ratios between emPAI scores from different samples is an approximation to the ratio of protein amounts.

## III.7 Bioinformatic methods

### III.7.1 Sequence data analysis

Protein and nucleotide sequences were compared using BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool; [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov); (Altschul *et al.* 1997).

Sequence alignments were performed using the *multialin interface page* „Blosum62-12-2“ (Henikof und Henikof 1992; <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>).

### III.7.2 Structure data analysis

Structural assignment of *M.mazei* lysate proteins according to the SCOP database (Structural Classification Of Proteins,(Lo Conte *et al.*, 2002)) was derived from the HMM library and genome assignment server (Gough *et al.*, 2001) (<http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/index.html>). Domain folds according to the SCOP data base were downloaded from the Protein Data Base (<http://pdbeta.rcsb.org/pdb/Welcome.do>). Proteins containing transmembrane domains are identified Munich Information Center for Protein Sequences (MIPS, <http://pedant.gsf.de/>) and validated using DAS- Transmembrane Prediction server of the Stockholm Informatics center (<http://www.sbc.su.se/~miklos/DAS/maindas.html>).

Functional assignment of *M. mazei* proteins according to the COG database (Clusters of Orthologous Groups of proteins) (Tatusov *et al.*, 1997) was downloaded from the Integrated Microbial Genomes (IMG) system (<http://img.jgi.doe.gov/>) (Markowitz *et al.*, 2006). COGs are based on phylogenetic classification of proteins encoded by multiple complete genomes. 17 distinct functional categories are assigned to COGs, which can be further summarized as subgroups of information storage and processing, cellular processes, metabolism, and poorly characterized proteins.

Physical data, such as hydrophobicity, net charge, pI, and calculation of the evolutionary scope of the *M. mazei* proteins were obtained using the gi|numbers for identification from the Evolutionary and Functional Genomics Server (EMU) at the Institute for Molecular Bioscience at the University of Queensland (<http://emu.imb.uq.edu.au/index.php>).

Essentiality was attributed to proteins from *M. mazei* according to the function of the protein in an essential pathway according to the Kyoto Encyclopedia of Genes and Genomes (KEGG). A protein was judged to be essential if the protein plays a critical role in a pathway indispensable for cell viability and no other functional homolog of this protein was present in the genome of *M. mazei*.

## IV. Results

Analysis of the genome of the mesophilic, methanogenic archaeon *Methanosarcina mazei* Gö1 (*M. mazei*) revealed that this organism contains not only the genes encoding the complete archaeo-typical Group II chaperonin system (thermosome; *MmThs*), but surprisingly also an operon of the bacterio-typical group I chaperonin genes (*MmGroEL*/*GroES*). This offers the unique opportunity to study the function of both chaperonins under the absolute same environmental conditions: the archaeal cell.

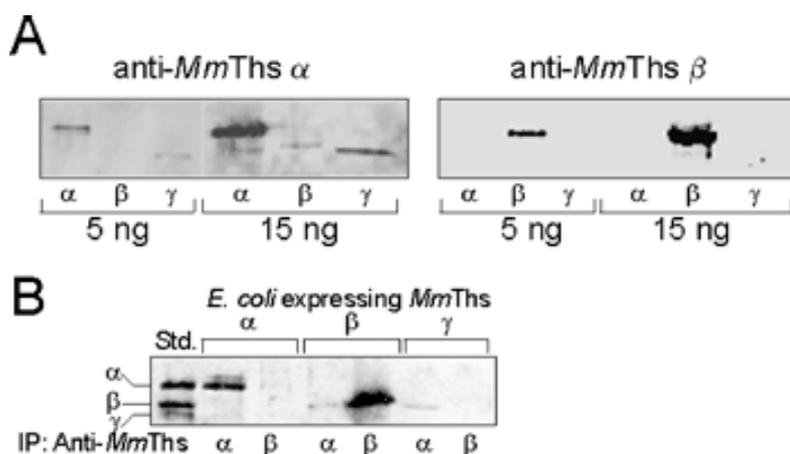
### IV.1 Production of recombinant proteins and polyclonal antibodies

To analyse the chaperones of *Methanosarcina* we cloned the genes encoding for the three thermosome subunits (*MmThs*  $\alpha$ , *MmThs*  $\beta$ , *MmThs*  $\gamma$ ), for the two prefoldin subunits (*MmPfd*  $\alpha$ , *MmPfd*  $\beta$ ), *MmGroEL*, *MmGroES* and *MmDnaK*. The respective proteins were expressed in *E.coli* BL21 and purified as described in the section materials and methods.

Polyclonal antibodies against *MmThs* $\alpha$  and *MmThs* $\beta$ , *MmPfd*  $\beta$  and *MmGroEL* and *MmGroES* were raised in rabbits to detect the corresponding proteins in *M. mazei* lysate. The specificity of the antibodies was tested in immunoblotting experiments and no significant crossreaction was detected (data not shown).

Because of the high sequence homology of the *MmThs*-subunits (*MmThs* $\alpha$  and *MmThs* $\gamma$  are identical in 70% and similar in 85% of their aminoacid sequence; *MmThs* $\alpha$  and *MmThs* $\beta$  are identical in 47% and similar in 69% of their aminoacid sequence; *MmThs* $\beta$  and *MmThs* $\gamma$  are identical in 47% and similar in 68% of their aminoacid sequence), the specificity of the antibodies against the *MmThs* $\alpha$  and *MmThs* $\beta$  was tested on recombinant *MmThs*-subunits under native and denaturing conditions. In immunoblotting experiments no significant cross reaction of anti-*MmThs* $\alpha$  and the recombinant *MmThs* $\beta$  subunit was detected and *vice versa*. In immunoblotting experiments, probably because of the strong homology of *MmThs* $\alpha$  and *MmThs* $\gamma$ , the anti-*MmThs* $\alpha$  showed minor cross reactivity of against the non-native *MmThs* $\gamma$  subunit (Figure 15a). Under native conditions though, when the antibodies were used in

immunoprecipitations from lysate of *E.coli* BL21 expressing a specific *MmThs* subunit, exclusively only the corresponding *MmThs* subunit was precipitated (Figure 15b).



**Figure 15 : Specificity of anti-*MmThs* $\alpha$ - and anti-*MmThs* $\beta$ - antibodies**

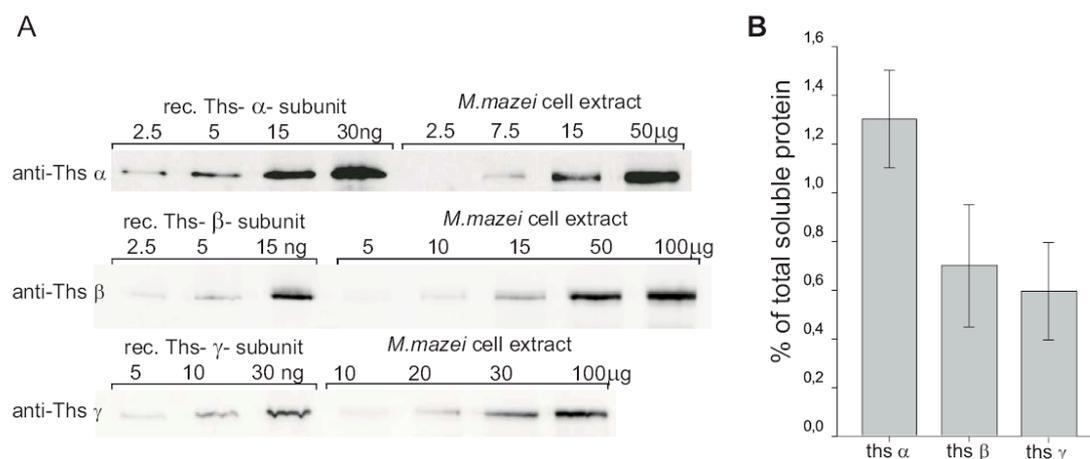
(A) Specificity of antibodies against non native *MmThs* subunits: Purified recombinant Ths  $\alpha, \beta, \gamma$  subunits were separated by 16% SDS-PAGE (pH 6.5) and immunoblotted with either anti-*MmThs* $\alpha$  (left) or anti-*MmThs* $\beta$  (right). (B) Specificity of antibodies against native *MmThs* subunits: soluble extracts from *E. coli* BL21 expressing the indicated *MmThs* subunit were immunoprecipitated with immobilized antibodies either anti-*MmThs* $\alpha$  or anti-*MmThs* $\beta$ . Precipitates were analysed for *MmThs* subunits by immunoblotting using a mixture of the antibodies.

## *IV.2 Both chaperonins and their cofactors are expressed at similar levels and as oligomeric proteins*

### *IV.2.1 Group II chaperonin: The MmThs consists of three different subunits*

All archaeal species analysed so far have Group II chaperonins composed of one to two different subunits. Only members of the family *Methanosacrinaceae* were shown to possess an additional gene coding for a thermosome subunit. Within the genome of *M. mazei* three different thermosome genes exist, termed  $\alpha$ ,  $\beta$  and  $\gamma$  with a molecular mass of 58.9, 58.5 and 58.2 kDa, respectively. Sequence alignment with their respective archaeal and eukaryotic homologs showed homologies of 50-80%. Sequence comparisons of apical domains of the subunits showed a high identity and similarity between the *MmThs* $\alpha$  and *MmThs* $\gamma$  subunit of ~70% and ~85%, the *MmThs* $\beta$  subunit shares only ~35% identity and ~70% similarity with *MmThs* $\alpha$  or *MmThs* $\gamma$  subunit. The *MmThs* $\alpha$  subunit shows a high degree of similarity to Ths  $\alpha$  subunits of other archaeal species and to the TCP-I subunit of the eukaryal group II chaperonin TRiC/CCT.

Immunoblotting-analysis of *M. mazei* lysate confirmed that all three subunits are expressed under standard conditions. Comparison of signal intensities of the respective *MmThs* subunit in *M. mazei* lysate with standard quantities of recombinant *MmThs*-subunits revealed that the thermosome subunits are expressed in a 2:1:1 ratio of  $\alpha$ : $\beta$ : $\gamma$  (Figure 16). The abundance of *MmThs* complex was assessed to a level of about 2% of total soluble proteins.

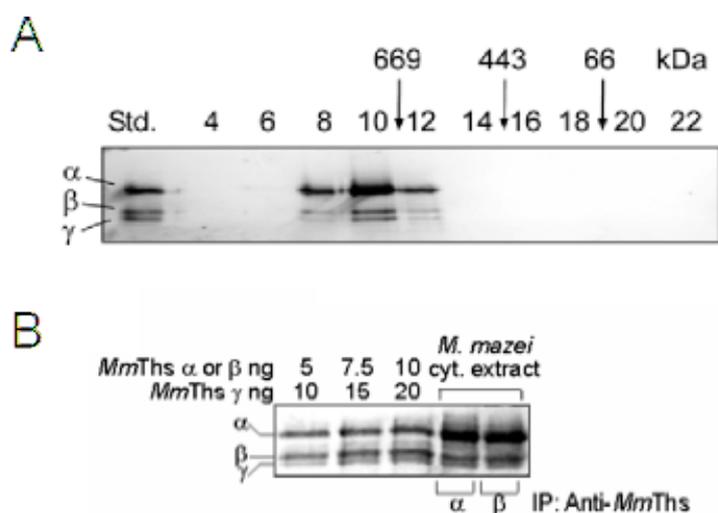


**Figure 16 : expression levels of group II chaperonins in *M. mazei* lysate**

(A) Relative amounts of *MmThs* subunits in *Mm*lysate detected by immunoblotting after separation by 16% SDS-PAGE, signal intensities of endogenous *MmThs* subunits were compared to signal intensities of recombinant *MmThs* $\alpha,\beta,\gamma$  subunits as indicated. (B) Quantitation of immunoblots of three independent experiments.

*MmThs* was also quantified from defined numbers of cells (data not shown). At an average cell size of  $\sim 1 \mu\text{m}$  in diameter, implying a single cell volume of  $4 \times 10^{-15} \text{ l}$ , a concentration of  $60 \mu\text{M}$  of monomer was calculated and  $2 \mu\text{M}$  of the thermosome complex supposing a hexadameric structure. At a theoretical total protein concentration of  $200 \text{ g/L}$  in the archaeal cell, this value is consistent with an abundance of thermosome at 2% of total soluble proteins shown above.

Subsequent immunoblotting of *M. mazei* total cell lysate, fractionated on a size-exclusion chromatography column (superdex 600) showed that all thermosome subunits were exclusively detected within the high molecular weight fractions. This allows to the conclusion that all the subunits are associated as thermosome double ring complexes (Figure. 17A).



**Figure 17 : Analysis of the composition of the group II chaperonin complexes in *M. mazei* lysate**

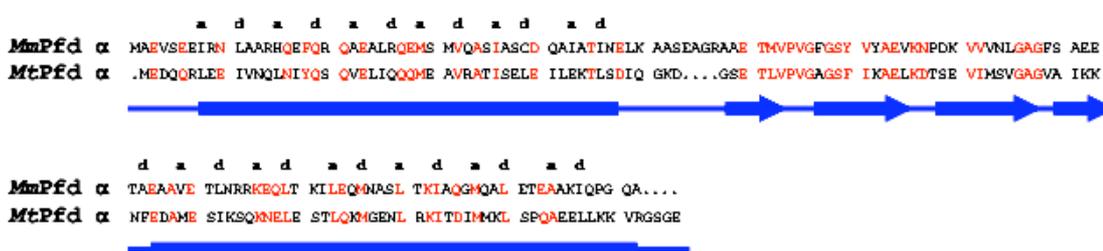
(A) Size exclusion chromatography of *M. mazei* lysate on a Superdex 600 column. After separation by SDS-PAGE (pH 6.5) the *MmThs* subunits were detected by immunoblotting using a mixture of anti-*MmThs* $\alpha$  and  $\beta$ . (B) Analysis of the endogenous *MmThs* complex by co-immunoprecipitation of the *MmThs* subunits. *MmThs*-complexes were precipitated from the lysate using the specific antibodies anti-*MmThs* $\alpha$  or anti-*MmThs* $\beta$ . The subunit composition of the precipitated complexes were analysed after separation by 16% SDS-PAGE (pH 6.5) by immunoblotting against *MmThs* $\alpha$  and *MmThs* $\beta$ , increasing amounts of recombinant subunits were used as standard.

The composition of the endogenous thermosome complexes was estimated by precipitation of the chaperonin complexes from the lysate using immobilized *MmThs* subunit specific antibodies. In both cases, when the *MmThs* complex was precipitated with anti-*MmThs* $\alpha$  or with anti-*MmThs* $\beta$ , similar ratios of the three subunits were found in the pull down experiments (Figure 17B). This similar subunit composition of the precipitated complexes by the different antibodies suggests an overall homogeneous population of the thermosome complexes in the archaeal cytosol. Quantification of the *MmThs* subunits from the co-immunoprecipitations revealed that the endogenous subunits preferentially associate in 2:1:1 ratio of  $\alpha$ : $\beta$ : $\gamma$ , in line with the ratio that was shown for the lysate.

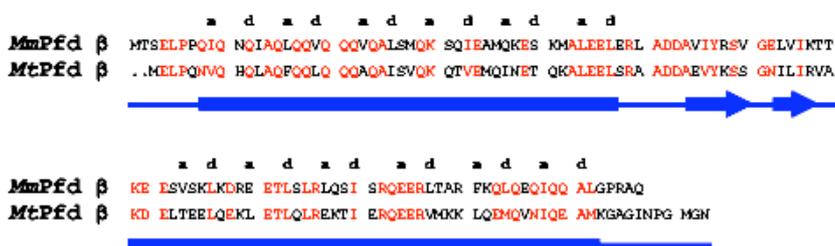
### IV.2.2 Group II chaperonin: the supposed co-chaperon prefoldin

Also genes for the potential cofactor prefoldin were also found to be present in the genome of *M. mazei*- represented by two genes as it is typical for *Euryarchaeotes* (Karlin *et al.*, 2005). The prefoldin  $\alpha$  subunit (*MmPfd* $\alpha$ ) has a predicted mass of 15.3 kDa and the prefoldin  $\beta$  subunit (*MmPfd* $\beta$ ) a mass of 13.5 kDa. Sequence comparisons of *MmPfd*  $\alpha$  with the  $\alpha$  subunit of prefoldin from *Methanobacterium thermoautotrophicum* (*MtPfd* $\alpha$ ) show 29% identity and 58% homology, the *Pfd*  $\beta$  subunits of these organisms are 46% identical and 71% similar.

#### Prefoldin $\alpha$



#### Prefoldin $\beta$

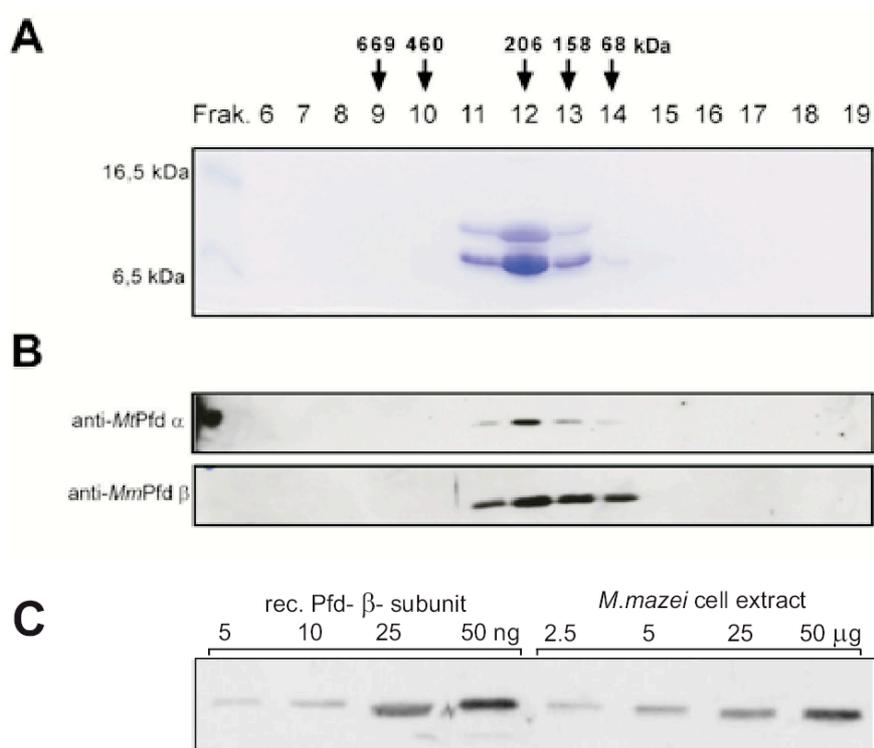


**Figure 18 : Sequence comparison of the *Pfd* subunits from *M. mazei* and *M. thermoautotrophicum***

“Coiled coil” regions are marked as blue squares, a/d represent hydrophobic amino acids typically found in „coiled coil” regions. Beta sheets are marked with blue arrows.

The crystal structure of *MtPfd* showed that the functional *MtPrefoldin* has the appearance of a jellyfish: its body consists of a double beta barrel assembly with six long tentacle-like coiled coils protruding from it. The distal regions of the coiled coils expose hydrophobic patches and are required for multivalent binding of non native proteins (Siegert *et al.*, 2000). From the strong homology of *MmPfd* and *MtPfd*, a similar structure of the complexes can be expected. Additionally, sequence comparisons with known coiled coil structures (Lupas *et al.* 1991) predict this typical coiled coil motif of prefoldin on the distal areas of both subunits.

This assumption was experimentally supported when recombinant *MmPfd* –complexes was run on a size exclusion column. The *MmPfd* $\alpha$  and the *MmPfd* $\beta$  subunit elute simultaneously at an expected size (~200 kDa due to the unusual jelly fish like shape; Vainberg *et al.* 1998; Siegers *et al.* 1999) with the  $\beta$ -subunit as being the more prominent band on a coomassie stained SDS-PAGE (theoretical ratio  $\beta:\alpha = 2:1$ ; Figure 19A). Similarly the endogenous *MmPfd* subunits were detected by immunoblotting in the expected size fractions for the complex (Figure 19B). From quantitative Immunoblotting an abundance of 0.25 - 0.5% of *MmPfd*  $\beta$  in the *Mmlysate* was estimated (Figure 19C). Proposing a protein concentration of 200g/L in the archaeal cell and supposing that *MmPfd* is present as the typical  $\alpha_2\beta_4$  complex a concentration of the *MmPfd* complex around 10  $\mu$ M in the archaeal cell can be expected.



**Figure 19 : Analysis of the subunit composition of recombinant and endogenous MmPfd**

(A) SDS-PAGE analysis of recombinant *MmPfd* after size fractionation. The 85 kDa Pfd complex elutes around 200 kDa due to its unusual shape. (B) Immunodetection of endogenous *MmPfd* subunits in size fractionated *M. mazei* lysate (arrows: Calibration of the size exclusion [kDa] : 669-thyroglobulin, 460-ferritin, 206-catalase, 158-aldolase, 68-albumin). (C) Abundance of *MmPfd*  $\beta$  in *M. mazei* lysate shown by comparative immunoblot analysis of lysate (right) and recombinant protein (left).

### IV.2.3 Group I chaperonin *MmGroEL* and the Co-chaperone *MmGroES*

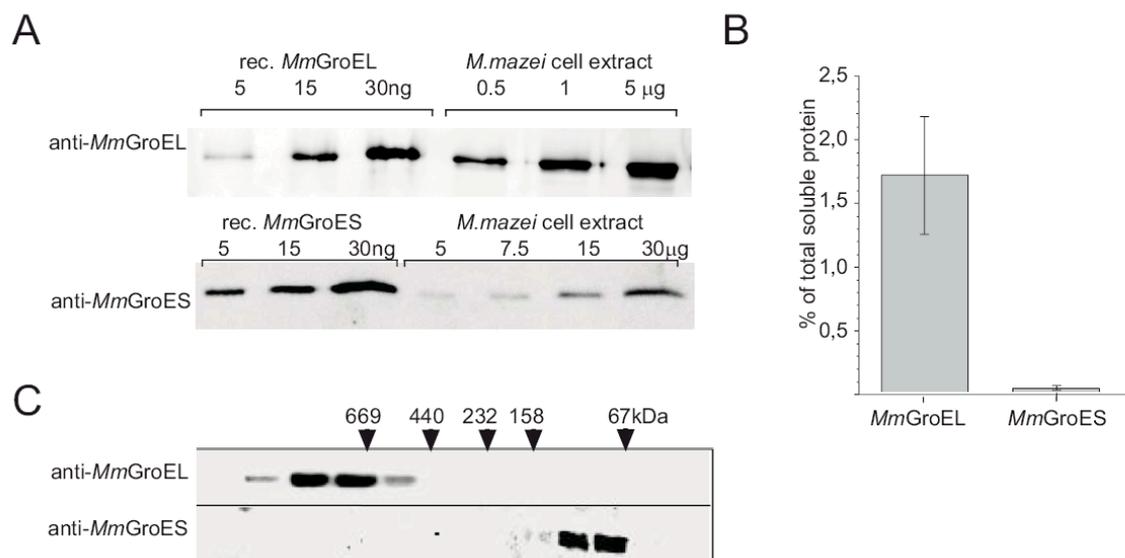
Most surprisingly the analysis of the *M. mazei* genome revealed the presence of a *groE* operon encoding for the complete group I chaperonin *MmGroEL* and its cofactor *MmGroES*. The general structure of the operon resembles the organisation in gram (-) bacteria. However the promoter and the heat shock element are of archaeal origin.

To ensure that the *groE* operon is expressed in the archaeal cell *M. mazei* total soluble protein was separated by SDS-PAGE. In Immunoblotting experiments both the *MmGroEL* and the *MmGroES* could be detected by the specific antibodies at the expected size of 60 kDa for the group I chaperonin and 10 kDa for the co-chaperone.

Comparison of signal intensities of the chaperones from *M. mazei* lysate to standard amounts of the recombinant proteins revealed an expression level of the chaperonin *MmGroEL* of ~ 1% and 0.1- 0.25% of the co-chaperone *MmGroES* was estimated (Figure. 20A, B).

In cell lysate that was fractionated on a size exclusion column *MmGroEL* and *MmGroES* were exclusively detected as oligomeric complexes. Compared to the elution profile of standard proteins *MmGroEL* was detected in fractions from a superdex 600 column corresponding to a size above 669 kDa, *MmGroES* eluted at a size bigger than 67 kDa (Figure 20C).

Taken together and assuming a cellular protein concentration of 200g/L the functional Group I chaperonin complex is present at concentration of 2 $\mu$ M, while the co-chaperon *MmGroES* is present in a up to 3fold excess at 3-6  $\mu$ M.

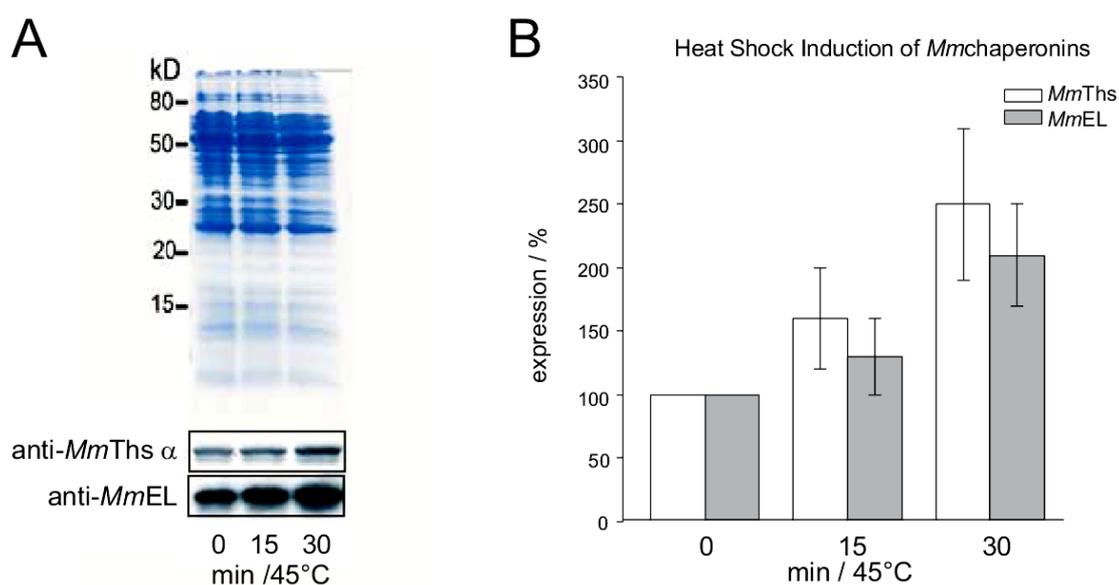


**Figure 20 : Group I chaperonin system in *M. mazei* cells**

(A) Abundance of *MmGroEL* and *MmGroES* in *M. mazei* lysate detected by immunoblotting after separation by 16% SDS-PAGE, signal intensities of the endogenous chaperones were compared to standard amounts of recombinant proteins as indicated. (B) Quantification of immunoblots of three independent experiments. (C) Size exclusion chromatography of *M. mazei* lysate on a superdex 600 column. After separation by SDS-PAGE the *MmThs* subunits were detected by immunoblotting.

#### IV.2.4 Both chaperonins are moderately induced under heat shock conditions

The similar abundance of both chaperonin argues for an equivalent contribution of both systems to protein folding in the cell under standard conditions. To assess the role of both chaperonins in the stress response of *M. mazei* the cells were subjected to heat stress at 45°C. After 15min, increased levels of both chaperonins in the lysate could be detected by immunoblotting, reaching a maximum expression level of 2-3 fold amounts after 30 min (Figure 21 A, B). Under heat stress no significant deviation in the response kinetics or the final level of the group I and the group II chaperonins were detected.



**Figure 21 : Heat shock induction of *M. mazei* chaperonins**

(A) Immunodetection of endogenous amounts of *M. mazei* chaperonins under heat stress. While the heat stress shows no significant effect on the protein pattern of the *M. mazei* lysate on a Coomassie stained 16% SDS-PAGE, increased amounts of *MmThs* and *MmGroEL* can be detected after 15 min at 45°C. (B) Quantification of expression levels of *M. mazei* chaperonins in three independent heat shock experiments.

### IV.3 Substrate identification

Several studies have been performed to identify substrate spectrum of the GroEL/GroES system in *E.coli* (Houry *et al.*, 1999; Kerner *et al.*, 2005) and several proteins have been found to interact with the group II chaperonin in the eukaryotic cell (Dunn *et al.*, 2001). So far no comprehensive study of the interaction proteome of group II chaperonins has been performed (Gomez-Puertas *et al.*, 2004). Analysis of the polypeptide flux through the eukaryotic group II chaperonin TRiC in mammalian cells showed that about 9-15% of the newly synthesized proteins can be expected to interact with the eukaryotic group II chaperonin, similar to *EcGroEL* (Thulasiraman *et al.*, 1999). Our knowledge about the thermosome is lagging behind in terms of its natural substrates and also its mode of action. Thermosomes of some hyperthermophilic *Archaea* have been shown to prevent aggregation and to promote refolding of heterologous model substrates, such as of green fluorescent protein (GFP) or citrate synthase (CS) (Gutsche *et al.*, 1999). However, only the use of natural substrates for biochemical and functional studies will give detailed insights into the mechanism of chaperonin assisted folding in the archaeal cytosol.

Therefore in this study, not only the GroEL/GroES of *M. mazei* Gö1 interacting proteins were identified, for the first time, a comprehensive identification of also the Group II chaperonin substrates was performed.

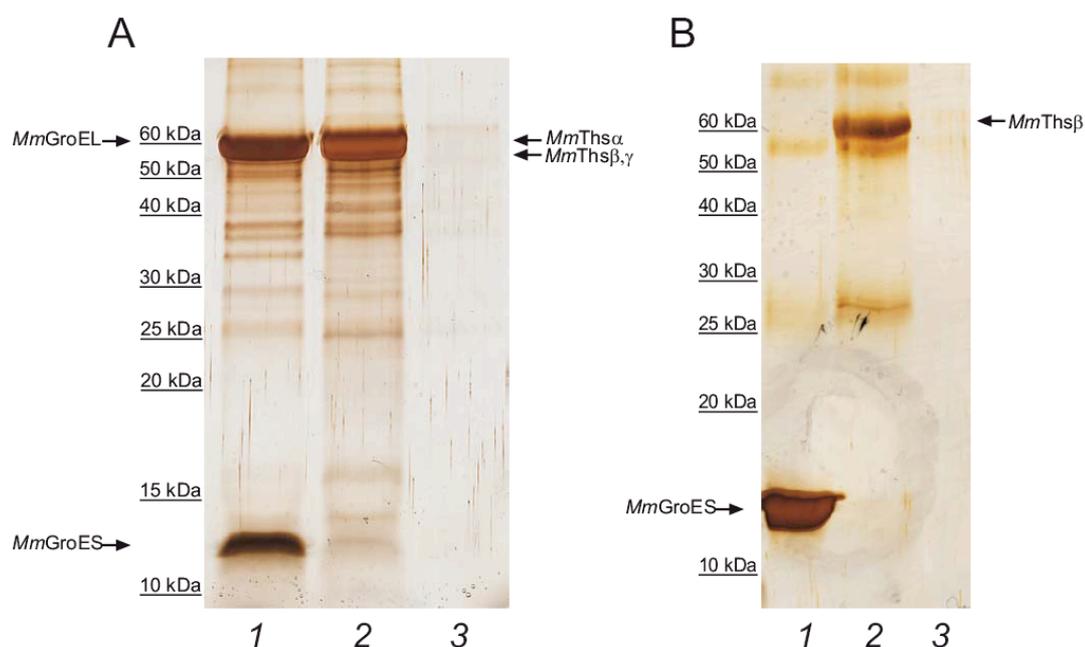
Most importantly the unique co-existence of both chaperonins in the same compartment allows the most differentiated comparison possible between the substrate specificity for either the Group I or the Group II chaperonins.

#### IV.3.1 Co-precipitation of chaperonin-substrate complexes

For a fast and efficient isolation of the chaperonin-substrate complexes single step purifications *via* immunopulldowns were performed. To precipitate the chaperonin-substrate complexes from *M. mazei* Gö1 cell lysate specific antibodies against the Group I and the Group II chaperonin proteins, which were raised in rabbits, were covalently linked to ProteinA – Sepharose Fast Flow beads before use to minimize the protein complexity of the precipitations

for the analysis by mass spectrometry. It was shown by Figueiredo *et al.* that *MmGroES* and *MmGroEL* form stable complexes in the presence of ADP at room temperature (RT). This interaction turned out to be sufficient to allow co-precipitation of the *MmGroEL*/GroES/substrate complex when specific antibodies against the co-chaperone *MmGroES* were used. Thermosome-substrate-complexes were isolated from *M. mazei* Gö1 lysate using specific antibodies against the *MmThs* $\beta$  subunit (anti-*MmThs* $\beta$ , Figure 22A).

Due to stringent washing conditions and specific elution of the chaperonin complexes from the corresponding antibody the background level was efficiently reduced. In the control experiment an unspecific antibody against firefly luciferase (FF-Luc) was used and virtually no proteins could be detected when equivalent amount of the experimental samples was analysed by SDS-PAGE (Figure 22A). Three independent immunoprecipitations from shock-cooled and subsequently lysed *M. mazei* Gö1 cells were performed.



**Figure 22 : Co-Immunoprecipitations form *M. mazei* lysate**

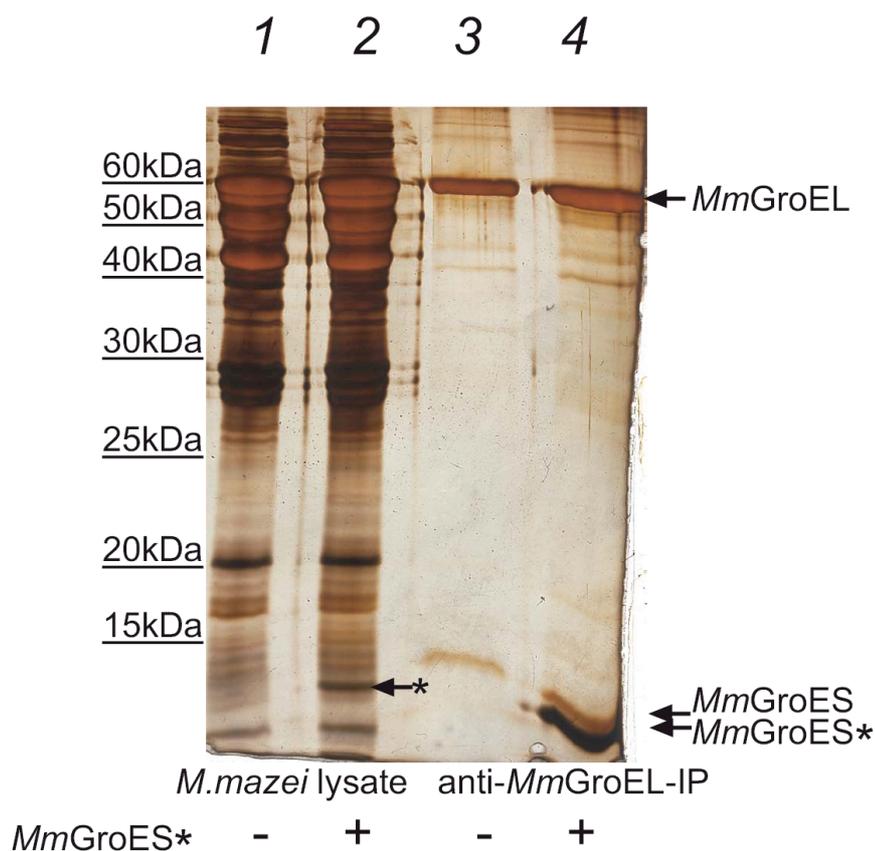
(A) Large scale co-immunoprecipitation of *M. mazei* chaperonins and interacting proteins (lane 1: *MmGroEL*/GroES and substrates, lane 2: *MmThs* and substrates, lane 3: control). (B) After denaturation of proteins from the lysate at 95°C for 1 min in presence of 1% SDS only the corresponding *M. mazei* chaperone subunits were precipitated (lane 1: anti-*MmGroES*, lane 2: anti-*MmThs* $\beta$  and lane 3: anti-FF-Luc), 16% SDS-PAGE, silver stained.

To prove specificity of the employed antibody-beads, the co-immunoprecipitation (co-IP) experiment was carried out on denatured lysate, where the chaperonin complexes are expected to be disrupted. For this purpose the lysate proteins were denatured at 95 °C for 1 min in the presence of 1% SDS. After renaturation by dilution only the  $\beta$ -subunit of the *MmThs* and the co-chaperonin *MmGroES* were precipitated- proving not only the specificity of the antibodies themselves, but also confirming the low background level of the experimental procedure, which has been already shown above by using the unspecific antibody (Figure 21b).

### *IV.3.2 The MmGroEL/GroES complex is stable during the experimental procedure*

Interestingly recombinant *MmGroES* shows a faster migration on SDS-PAGE than endogenous *MmGroES*, which might be due to posttranslational modification in the archaeal cell. This deviation in mobility though, allows differentiation between the recombinant and endogenous *MmGroES*.

To exclude possible rearrangements of the GroEL/GroES complexes and their corresponding substrates during the experimental procedure, recombinant *MmGroES* was added in excess over endogenous *MmGroES* during lysis. When the *MmGroEL/GroES* complex was precipitated in this case by using a specific antibody against *MmGroEL* (anti-*MmGroEL*), no replacement of the endogenous *MmGroES* by the recombinant *MmGroES* was detected, proving the stability of the endogenous *MmGroEL/GroES* complex. The additional appearance of recombinant *MmGroES* is probably due to an association of the recombinant co-chaperone to *MmGroEL* complexes that had endogenous *MmGroES* bound.



**Figure 23 : Stability of MmGroEL/GroES complex**

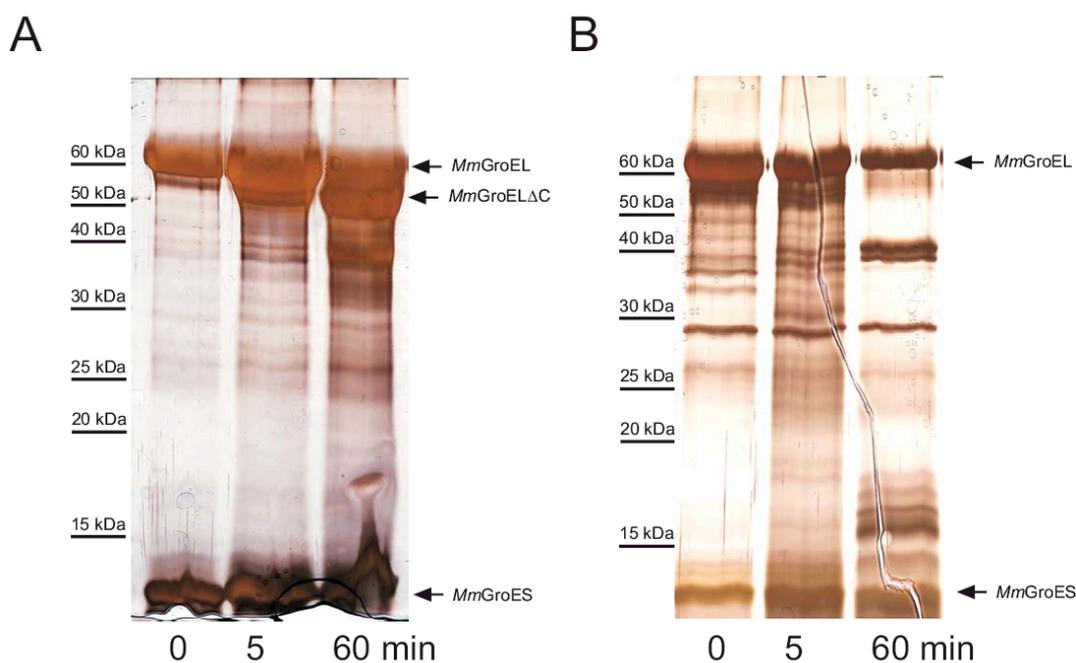
*M. mazei* lysate was subjected to co-IP using an anti-MmGroEL antibody without (lane 1) and after addition of recombinant MmGroES in excess (lane 2, asterisk: recombinant MmGroES) during lysis and influence of the addition of recombinant MmGroES in excess (lane 4) on the amount of endogenous MmGroES in the precipitated MmGroEL/GroES complexes was detected by silver staining of 16% SDS-PAGE.

#### IV.3.3 GroEL/GroES complexes contain mainly encapsulated *cis*-ring bound substrates

In the ADP-bound status of GroEL/GroES complexes, the substrates are expected to be encapsulated mainly in the *cis*-cavity undergoing folding (Figure 24). In addition, some proteins can be expected to be bound to the *trans*-GroEL ring, especially proteins exceeding the volume of the GroEL/GroES cavity. It is speculated by Kerner *et al.* that probably the proteins identified among the *Ec*GroEL substrates with a monomeric size exceeding the estimated cut-off (60 kDa) might interact with the *trans* cavity of the GroEL/GroES complex.



becomes visible, due to the digestion of the carboxy-termini of the trans-cavity by PK (Figure 25A). As a control the PK digestion experiment was performed in presence of ATP. ATP allows the cycling of the co-chaperone on and off GroEL and therefore proteinase K can access both GroEL-rings similarly. Under this condition the bands representing the non native substrate proteins disappear after 60 min as they are completely digested by PK (Figure 25B). Moreover, the *MmGroEL/MmGroEL* $\Delta$ C double band can not be detected as there is no stable formation of the asymmetric GroEL/GroES complex that would lead to a protection of the carboxy-termini in the *cis*-cavity.



**Figure 25 : Nucleotide dependent Proteinase K digestion of *MmGroEL* substrates**

Proteinase K digestion of *MmGroEL*-substrate-complexes was performed in presence of ADP (A) and after addition of ATP (B). Stabilization of the GroEL/GroES complex by ADP led to a protection of the C-termini of the *MmGroEL* *cis*-cavity, resulting in the characteristic double band of *MmGroEL/MmGroEL* $\Delta$ C. The protection of substrates from digestion is due to encapsulation in the *cis* cavity. Addition of ATP allows cycling and abolishes the protective effect of *MmGroES*. This leads to a digestion of all *MmGroEL* C-termini and the *MmGroEL*-GroES-substrates.

#### *IV.3.4 The co-precipitated proteins can be released from the corresponding chaperonin in a nucleotide dependent manner*

Non-native proteins are captured by the chaperonin systems of GroEL/GroES and thermosome followed by their nucleotide dependent, transient enclosure for folding and release

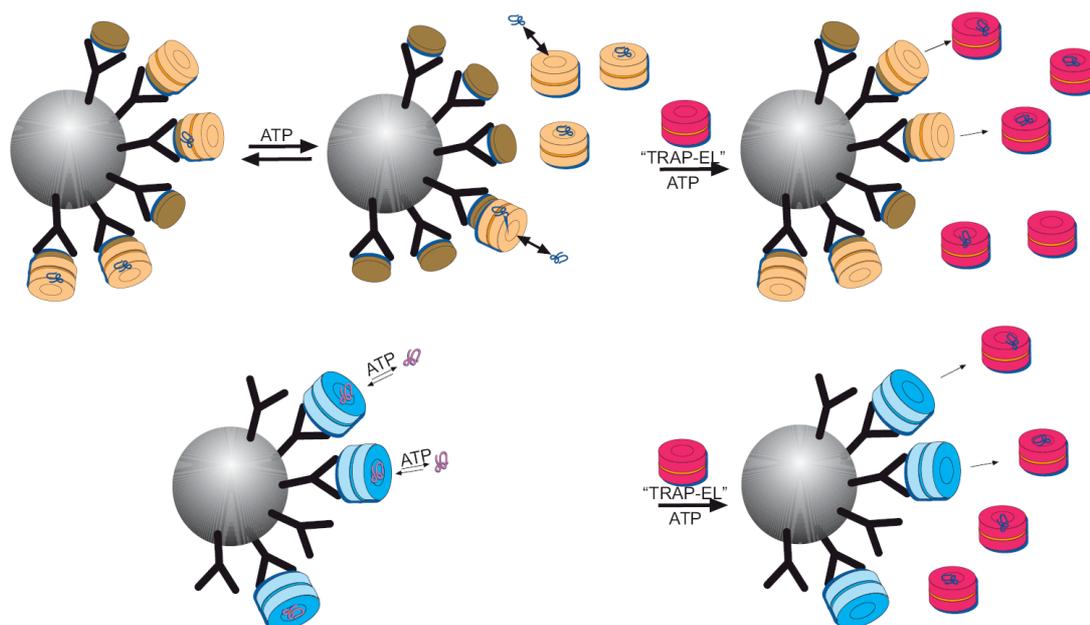
---

into bulk. When the client protein has reached its native and no longer exposes surface motifs for chaperonin interaction, it will not rebind to the chaperonin.

To ensure that the co-precipitated proteins were associated with their corresponding chaperonins in a folding competent state, the nucleotide dependent substrate release from the chaperonin cavities was monitored. Different conditions were tested for the release of substrates. Addition of ATP led only to a very limited release of proteins from the chaperonins (data not shown), which may be due to several reasons. Firstly, the function of the chaperonin could be negatively influenced by the association to the antibody-beads. Secondly, a co-fraction relevant for folding and/or release of the client protein may have been missing. Thirdly, many of the proteins might be members of oligomeric structures and preserve exposed hydrophobic surfaces for binding the corresponding interaction partner. Such proteins will continuously rebind to chaperones until they have associated with their natural interaction partner (Figure 25, left side).

To overcome the rebinding of the substrate proteins can be trapped from the chaperonin upon release by addition of an excess of a mutant version of the *EcGroEL* chaperonin, the so-called TRAP-*EcGroEL* (Figure 25, right side). This mutant GroEL binds proteins, which expose hydrophobic surfaces, with a similar affinity as *EcGroEL*, but is defective in its ATPase activity and is therefore not able to undergo cycling, resulting in a stable binding of the client protein.

Although this experiment does not demonstrate successful folding, it shows a nucleotide dependent interaction of the chaperonin and substrate proteins.

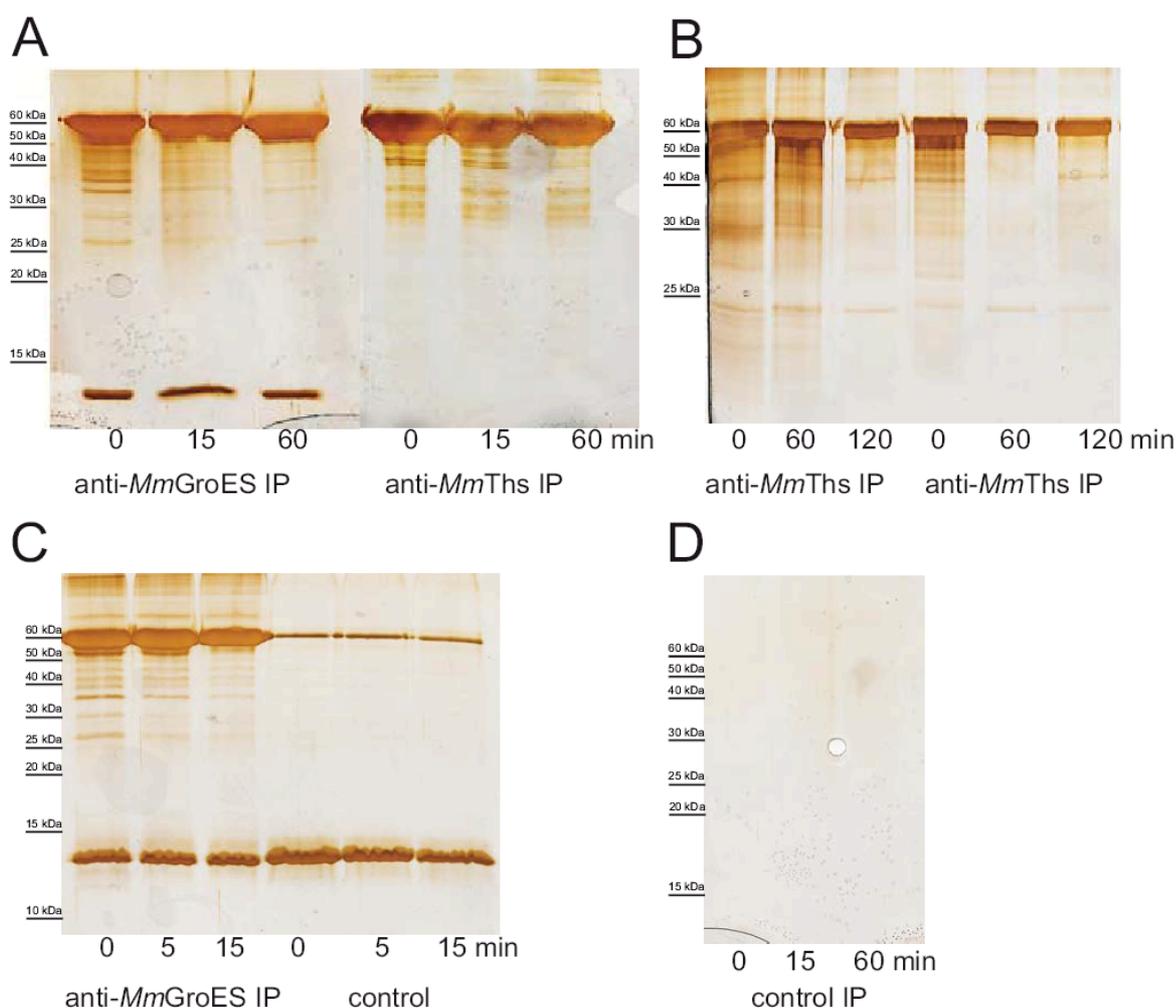


**Figure 26 : Schematic drawing of the “TRAP” EcEL driven substrate release**

Addition of ATP to the antibody bound *MmGroES* (brown)-*MmGroEL* (orange)- substrate (blue) will lead to cycling of the chaperonin system, the substrate will rebind until no hydrophobic surfaces are exposed. As many proteins are part of oligomeric complexes, they will expose hydrophobic surfaces also under native conditions and therefore continuously rebind. A similar scenario can be expected for *MmThs*(blue)- substrate complexes. Rebinding of substrates during cycling of the *M. mazei* chaperonins is prevented by addition of TRAP-EcEL (pink), which can not undergo cycling and therefore stably binds the released substrate.

When *MmGroEL*/*GroES* substrates complexes were incubated with ATP in the presence of 10-fold excess TRAP-*EcGroEL*, an efficient release of the substrate proteins was observed already after 5 min (Figure 27A, lane 1-3 and 27C, lane 1-3), while a considerable amount of substrates was still bound on the thermosome even after 60 min (Figure 27A, lane 4-6).

To rule out if this lower release efficiency was due to a cycling deficiency of the thermosome complex under the experimental conditions or due to a lower affinity of TRAP-*EcGroEL* for the thermosome associated proteins, TRAP-*EcGroEL* was added at increasing amounts. When TRAP-*EcGroEL* was used at 100-fold excess the substrate proteins were released from *MmThs* within 60 min (Figure 27B, lane 4-6), while at a 10-fold excess of TRAP-*EcGroEL* efficient release was only detected after 120 min (Figure.26b, lane 1-3). The inefficient transfer of proteins from the thermosome to TRAP-*EcGroEL* is potentially caused by a limited affinity of TRAP-*EcGroEL* for thermosome substrates.



**Figure 27 : TRAP-EL promoted substrate release**

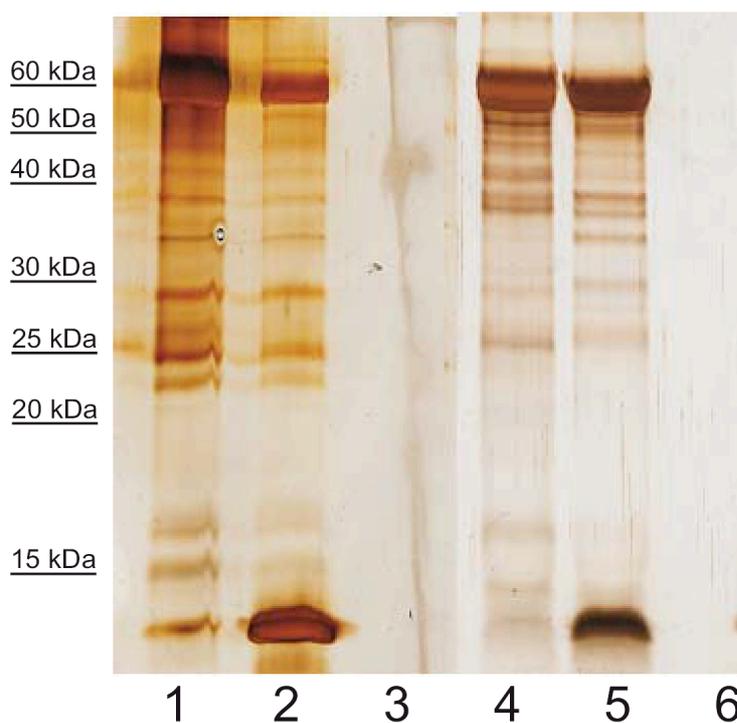
(A) Nucleotide dependent release of substrates from *MmGroEL*/*GroES*- complexes (lane 1-3) and *MmThs* complexes (lane 4-6) in the presence of 10x excess TRAP-*EcGroEL*. (B) Release of substrates from *MmThs* complexes depending on excess of TRAP-*EcGroEL*. Efficient release from *MmThs* can be provoked by 100x excess of TRAP-*EcGroEL* over *MmThs*. (C) The release of substrates (lane 1-3) is not due to replacement of *MmGroEL* by TRAP-*EcGroEL*, which shows only very little and nucleotide independent affinity for the antibody-column bound *MmGroES*. (D) The addition of TRAP-*EcGroEL* to the experimental procedure had no significant effect in the control experiment with the unspecific anti-FF-Luc.

As a control, to confirm the inability of TRAP- *EcGroEL* to bind the co-chaperone *MmGroES*, which excludes that the observed substrate release from *MmGroEL* could be due to exchange of *MmGroEL* against TRAP-*EcGroEL*, TRAP- *EcGroEL* was added to purified recombinant *MmGroES* that was bound to protein A sepharose via the *MmGroES* antibody. Only limited amount of TRAP-*EcGroEL* was found associated with *MmGroES* independent of

the duration of the experiment (Figure 27C, lane 4-6). In the background control experiment using anti-FF-Luc no influence of the addition of TRAP-*EcGroEL* on the immunoprecipitation was detected (Figure 27D).

#### *IV.3.5 Chaperonin-substrate complexes captured under native conditions differ from in vitro reconstituted Chaperonin-substrate complexes*

The co-precipitations showed remarkable differences in the substrate spectra of Group I and Group II chaperonins - as visualized on 1D or 2D- SDS-PAGE. These differences might be due to different affinities of the chaperonin groups for newly synthesized polypeptides, but also other cellular factors- for example upstream acting chaperones- may also play an important role in sorting proteins to the respective chaperonins. To investigate if the chaperonin substrate interaction is driven only by the amino acid sequence of the polypeptide chain, the co-IPs were compared to *in vitro* generated chaperonin and *M. mazei* protein complexes. Chemically denatured *M. mazei* lysate was diluted into a solution containing both recombinant *M. mazei* chaperonins in the presence of 5mM ATP. The formed chaperonin-protein complexes were stabilized by the addition of hexokinase/glucose/ADP, resulting in hydrolysis of ATP to ADP, and then precipitated using the respective antibody beads. Both the *MmThs*- and the *MmGroEL* lysate protein complexes generated *in vitro* (Figure 27, lane 1, 2) significantly differed from complexes precipitated from *M. mazei* cells (Figure 27, lane 4, 5). The *in vitro* formed chaperonin-substrate complexes showed similar protein pattern for both the Group I and the Group II chaperonin and several band of substrates of the *in vivo* precipitated *MmThs* and *MmGroEL* were present on both *in vitro* chaperonin-substrate complexes.



**Figure 28 : *In vitro* formed *M. mazei* chaperonin-substrate complexes compared to endogenous chaperonin substrate complexes**

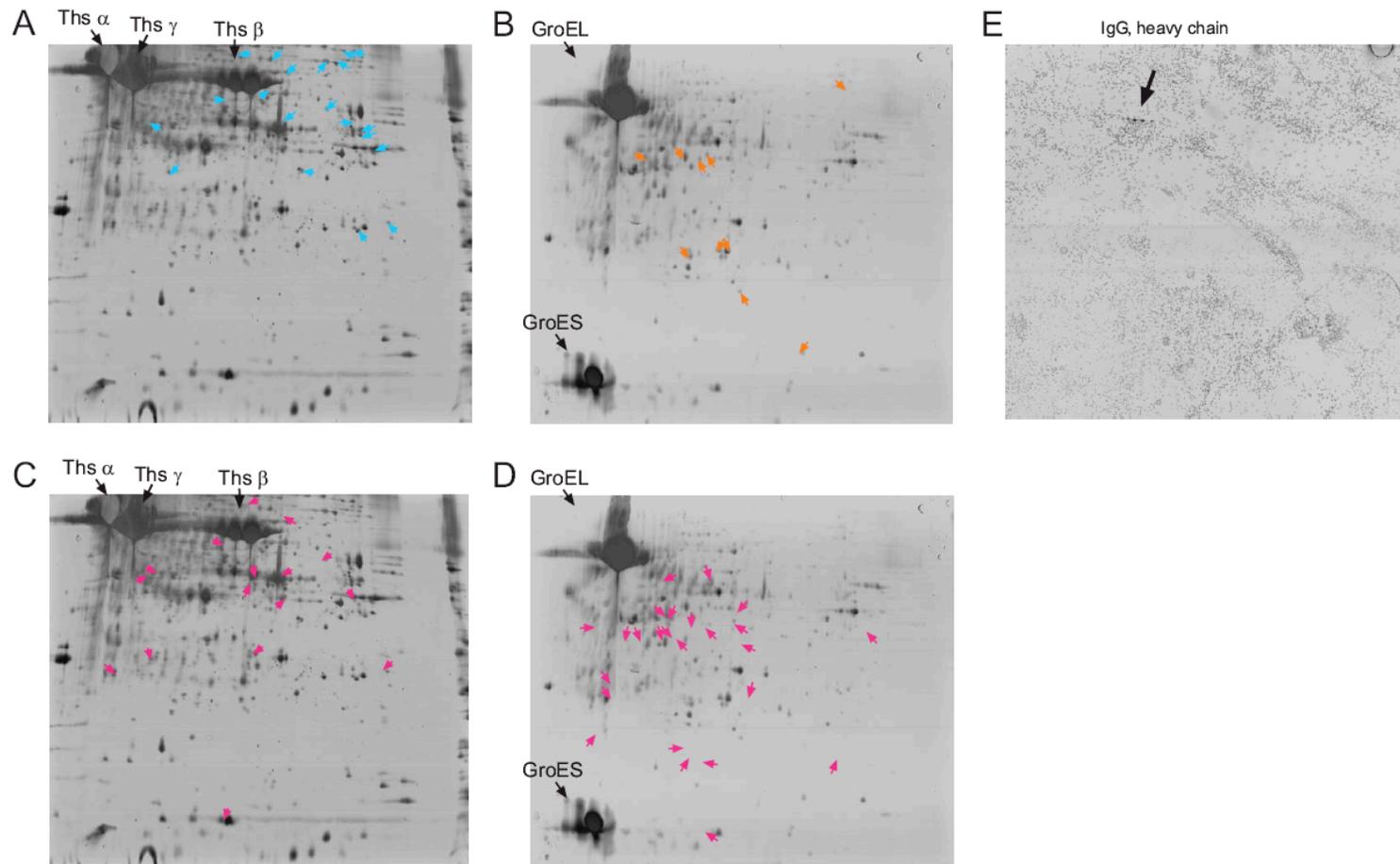
After thermal denaturation at 95°C in the presence of 1% SDS for 1 min the *M. mazei* lysate was diluted into buffer containing both *M. mazei* chaperonins allowing the formation of *MmThs*-substrate (lane 1) and *MmGroEL*/*GroES*-substrate complexes (lane 2). The chaperonin-substrate complexes were co-immunoprecipitated and were compared to complexes precipitated from *M. mazei* cells (lane 4: *MmThs*- substrate complexes, lane 5: *MmGroEL*/*GroES*-substrate complexes; control- immunoprecipitations with FF- antibody under “*in vitro*” (lane3) and *in vivo* conditions (lane 6).

### IV.3.6 Substrate Identification

Routinely three independent, large scale experiments were performed to precipitate *M. mazei* chaperonin-substrate complexes including the corresponding background control. Three different analysis methods were employed to identify the co-precipitated substrate proteins. Firstly, the classical 2D PAGE was performed on the chaperonin-substrate complexes to visualize differences in the substrate spectra, and spots were selected for identification by LC-MS/MS. Secondly, the proteins precipitated in another large scale experiment were analysed by the Ettan-DIGE technique - a more differentiated method of 2D-PAGE – followed by LC-MS/MS of specific protein spots. Thirdly, because of the greater sensitivity, the 1D-LC/MS was chosen as a more comprehensive substrate identification method.

#### *Substrate identification by 2D-PAGE-MS*

On the one-dimensional SDS-PAGE the precipitated *MmGroEL/GroES* substrate- and *MmThs* substrate- complexes showed remarkable deviations in their protein pattern and, for a better visualisation of these differences, the chaperonin – substrate complex were separated by 2D- PAGE (Figure 29). After visual inspection 145 protein spots from the anti-*MmThs* and 107 protein spots from the anti-*MmGroEL/GroES* pull down were picked and analysed by LC-MS/MS (Table 2). In the background control no proteins could be detected and therefore no proteins were picked (Figure 29E). 40 proteins of the thermosome pull down could be identified, 19 (47.5 %) of these were categorized as *MmThs*-specific substrates (Figure 28A) as they were never found in any *MmGroEL/GroES*-IP (including the 1D-LC/MS analysis). Analogously 9 (32 %) of the 28 proteins identified in the anti-*MmGroES* pull down were defined as *MmGroEL/GroES* specific substrates (Figure 29B). The majority of the identified proteins from the *MmThs* pull down (21 proteins, Figure 29C) and from the *MmGroEL* pull down (19 proteins, Figure 29D) were identified in IPs of both chaperonins. These proteins interact with both chaperonin groups and were therefore categorized as overlapping substrates.



**Figure 29 : 2D-PAGE of *M. mazei* chaperonin substrate complexes**

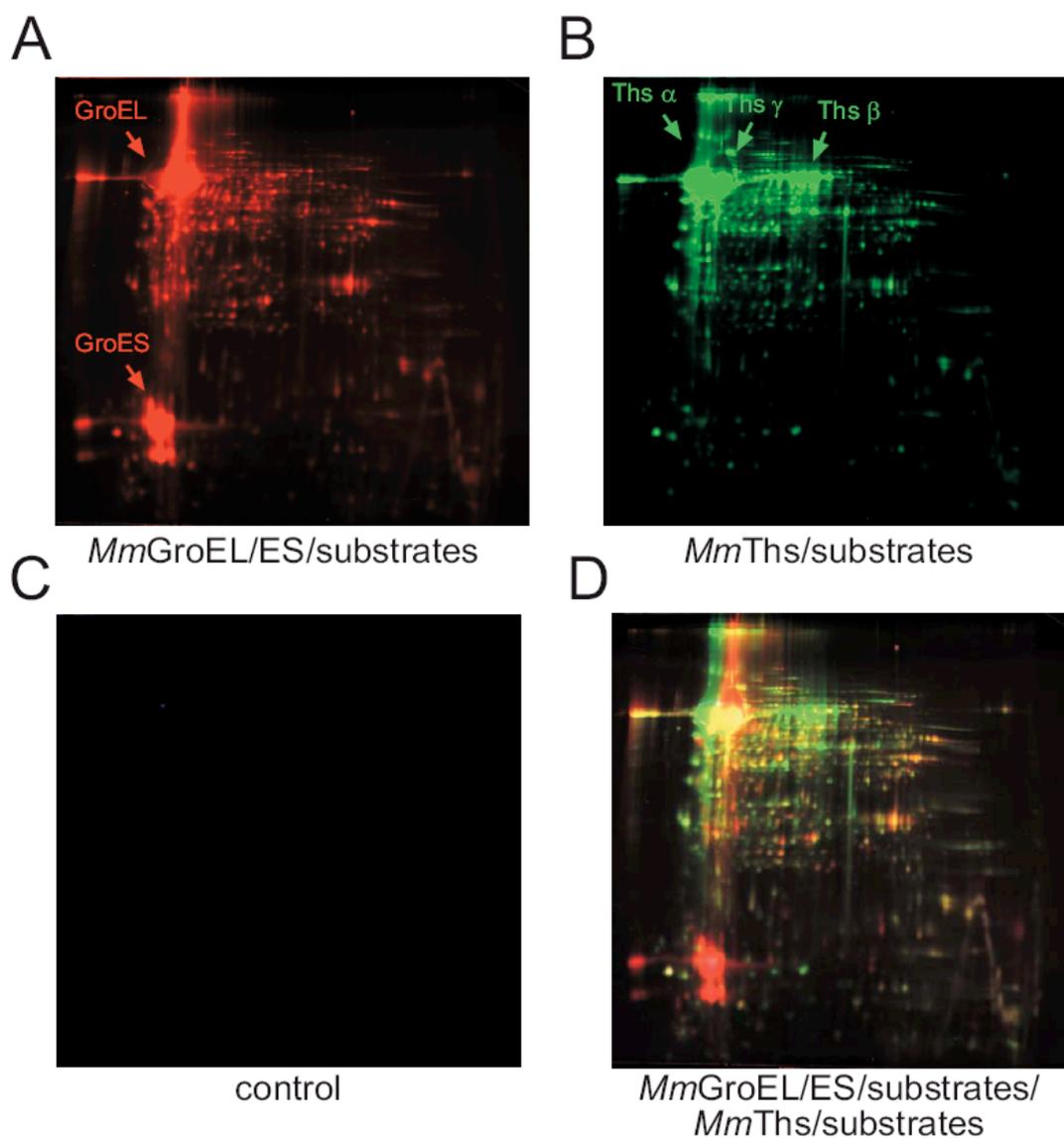
Proteins from the *MmThs* substrate and *MmGroEL/GroES* were focused to their isoelectric point at a range between pH3 and pH10 and then separated by 16%-SDS PAGE. Proteins were detected by silver staining and spots were picked by visual inspection. *MmThs* specific proteins are depicted in turquoise (A), *MmGroEL/GroES* specific proteins in orange (B), overlapping proteins that were present in the *MmThs* complexes (C) as well as in the *MmGroEL/GroES* (D) complexes are labelled in pink. No proteins were detected in the background control (E).

### *Substrate identification by Ettan DIGE-MS*

The chaperonin – substrate complex samples were also subjected to a modified 2D-PAGE analysis, the 2-D Fluorescence Difference Gel Electrophoresis (Ettan DIGE). In this technique up to three samples are labelled with distinct fluorescent dyes (CyDye DIGE Fluors, minimal dyes), which do not affect the electrophoretic properties of the proteins. The mixture of the three labelled samples is then separated by 2D-PAGE and the presence of a protein is detected by its fluorescence signal. This technique allows a more reliable comparison of up to three different protein samples, because the samples are analysed on the identical gel and therefore artificial deviations in the protein pattern are minimized, which are due to deviations of different 2D-PAGE. Additionally, because of the minimal labelling technique, the amount of a given protein can be compared quantitatively between the different samples as the fluorescence intensity is directly correlated to the amount of a protein.

Here the *MmGroES*-substrate complexes and the *MmThs*-substrate complexes were analyzed together with the background control sample (Figure 30) or with the sample containing the total soluble proteins of *M. mazei* (Figure 31). After detection of the present proteins by a Typhoon reader the peak intensities of the protein spots were quantified using the DeCyder software.

Firstly, the *MmGroEL*/*GroES*- substrate (Figure 30A) and the *MmThs*-substrate sample (Figure 29B) were compared. In the visual overlay proteins only present in *MmGroEL* precipitations appear in red, while proteins specifically associated *MmThs* appear in green (Figure 29D). Proteins that are found at similar levels in both precipitations appear in yellow. No proteins were detected in the background control sample (Figure 30C).



**Figure 30 : 2D-DIGE analysis of *M. mazei* chaperonin substrate complexes**

*MmGroEL*/*GroES* substrate complexes (A) labelled with Cy3, *MmThs* substrate complexes (B) labelled with Cy5 and the control sample labelled with Cy2 (C) were pooled, separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison (D). As the fluorescence signal is directly proportional to the amount of a protein, the fluorescence intensities of the proteins were integrated and compared using the DeCyder software. Proteins spots showing a more than 2.5 fold deviation among the IP-samples were defined as proteins specifically present in the corresponding IP. These proteins were picked and analysed by mass spectrometry. This way 61 protein spots were defined as specifically associated with *MmGroEL*/*GroES* and 58 with *MmThs* (Table 2).

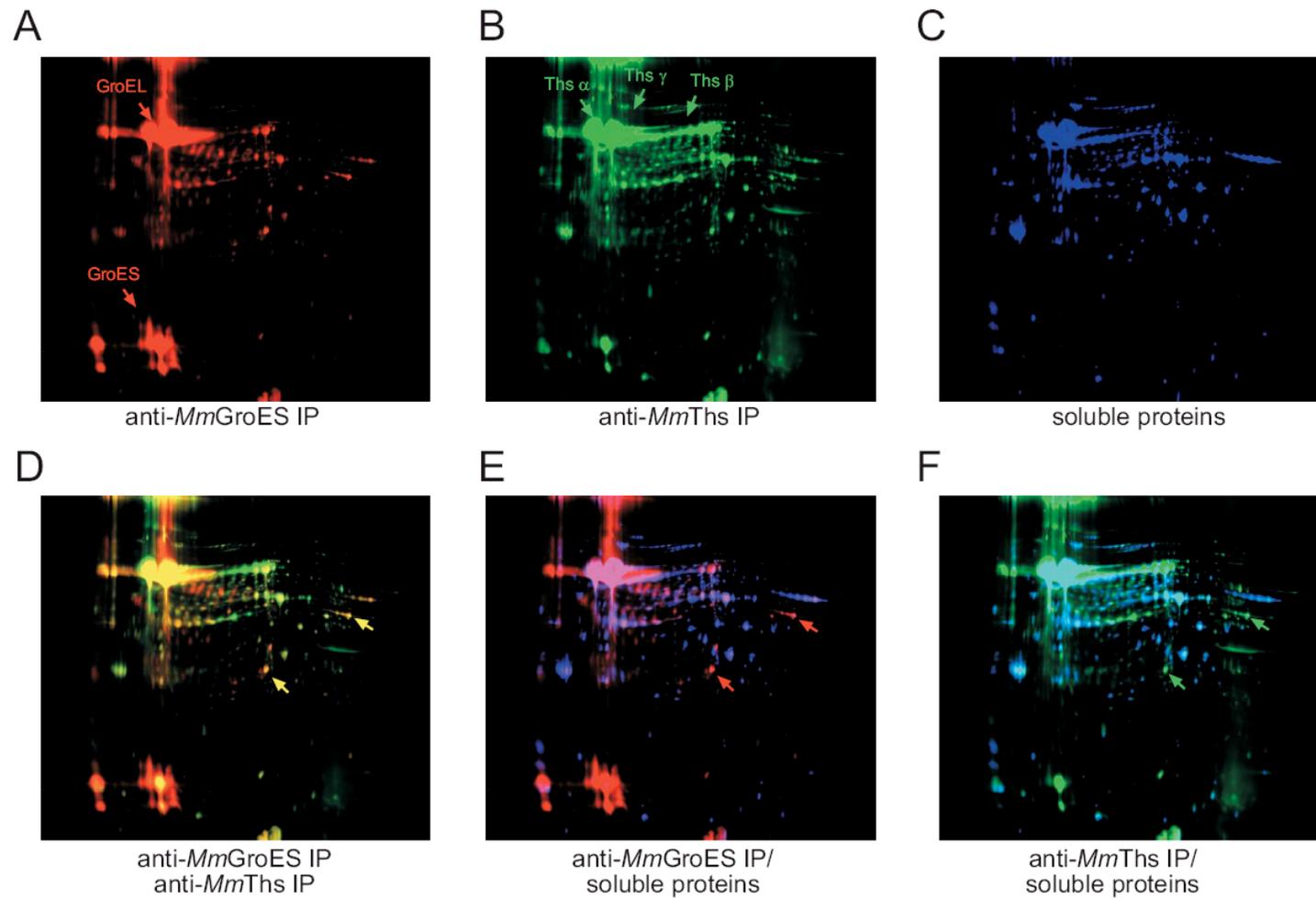
To prove the specific association of overlapping substrates with both chaperonin groups, the enrichment of these proteins on the chaperonins compared to their presence in the lysate was demonstrated. Therefore the fluorescence intensity of these proteins in the *M. mazei* chaperonin-substrate samples (Figure 31A, B) was compared to the fluorescence intensity of these proteins from an equivalent amount of total soluble protein of *M. mazei* (Figure 31C). Proteins at similar levels among both chaperonin-substrate complexes appear in yellow in the overlap of the *MmGroEL*-IP and the *MmThs*-IP (Figure 30D, *i.e.* indicated by the yellow arrow). Overlapping substrates appear in both overlays of the IPs and lysate in the colour corresponding to the IP sample (Figure 30 E, F). This allows the conclusion that these proteins were enriched in both IP samples compared to the lysate. These proteins can be expected to specifically interact with both of the chaperonins and were therefore defined as overlapping substrates. When the fluorescence intensities were compared using the DeCyder software, 31 proteins had more than 2.5 fold higher intensities in both precipitations than in the lysate (Table 2).

Table 2: *M. mazei* chaperonin substrates identified from 2D-PAGE

<b>2D-PAGE, silverstained</b>	<b><i>MmThs</i></b>	<b><i>MmGroEL/ES</i></b>
Interacting	40	28
Specifically interacting with the corresponding chaperonin group	19	9
Picked spots	145	107

<b>Ettan DIGE</b>	<b><i>MmThs</i></b>	<b><i>MmGroEL/ES</i></b>
Interacting (number of picked spots)	12 (89)	15 (92)
Specifically interacting with the corresponding chaperonin group	8 (58)	11 (61)

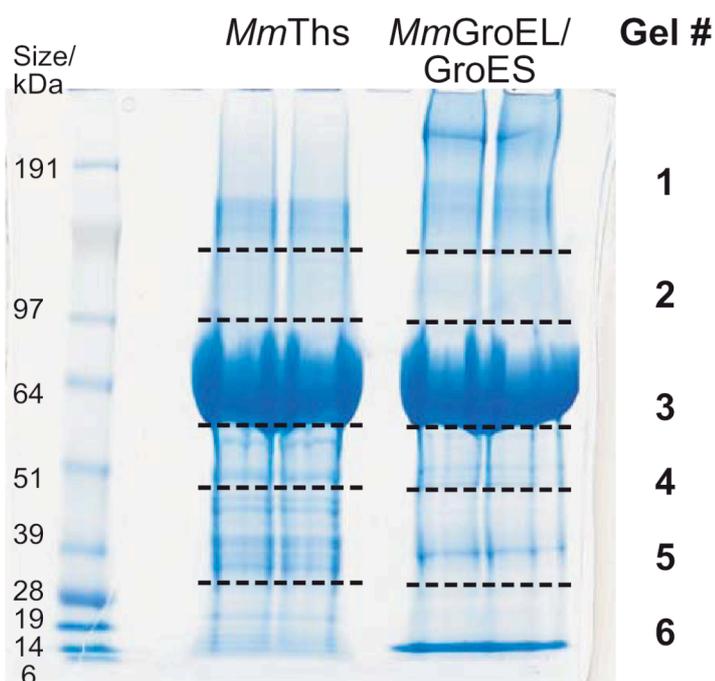


*Mm*GroEL/GroES substrate complexes (A) labelled with Cy5, *Mm*Ths substrate complexes (B) labelled with Cy3 and an equivalent amount of *M. mazei* lysate labelled with Cy2 (C) were pooled, separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison (D-F). Examples of overlapping substrates are indicated by arrows.

**Figure 31 : 2D-DIGE analysis of *M. mazei* chaperonin substrate complexes and *M. mazei* lysate**

### Substrate identification by 1D-LCMS

150- 200ug of each sample of all three large scale experiments were separated on one dimensional SDS-PAGE and each one was cut into 5-6 gel-fragments (Figure 32). After in-gel trypsin digestion the peptides were separated by reverse phase chromatography and identified by a coupled QSTAR Pulsar quadrupole TOF tandem mass spectrometer (Lasonder *et al.*, 2002).



**Figure 32 : 1D-SDS-PAGE of large scale IP of *M. mazei* chaperonins**

Before in-gel trypsin digestion 150-200  $\mu$ g of the large scale IPs, an equivalent volume on the background control or 20-50  $\mu$ g of *M. mazei* lysate (not shown) were separated by SDS-PAGE (4-16% BisTris-NuPage) and visualized by staining with colloidal Coomassie Brilliant Blue. The lanes of the *MmGroEL* and the *MmThs* IP were separated and cut as indicated.

Proteins that were identified in at least two of the three independent co-IP experiments with the corresponding chaperonin and absent in both analyzed background controls were defined as substrates. In total 454 proteins from *M. mazei* were identified as interacting with the chaperonins (Table 3). Out of these chaperonin-dependent proteins 315 proteins were found to be associated with the group II chaperonin *MmThs*, 227 proteins with the group I chaperonin *MmGroEL*. Each chaperonin substrate set is composed of a fraction of specific substrate

proteins, which were exclusively identified in context with the corresponding chaperonin, and the fraction of overlapping substrate proteins, which interact with both chaperonin groups. Among both chaperonin substrate sets about 2/3 of the proteins are found to interact with both chaperonins (Table 3).

Deviations in the number of proteins in the overlap fraction of *MmThs* and *MmGroEL* proteins (198 overlapping proteins in the *MmThs* substrate set, 153 overlapping proteins in the *MmGroEL* substrate set) are due to the evaluation of identified proteins. Several proteins from the *MmThs* overlap fraction were identified in only one of the *MmGroEL* samples and therefore not categorized as *MmGroEL* substrates and *vice versa*.

Table 3: *M. mazei* chaperonin substrates identified from 1D-LCMS

	<i>Mmchaperonin</i>	<i>MmThs</i>	<i>MmGroEL</i>
<b>Interacting</b>	454	315	227
(% of total soluble protein)	(17%)	(12%)	(9%)

	<i>MmThs</i>	<i>MmGroEL</i>
<b>Specifically interacting</b>	117	74
(% of corresponding substrates)	(37%)	(35%)
<b>Overlapping</b>	198	153
(% of corresponding substrates)	(63%)	(65%)

### IV.3.7 Overview of the Proteomic Data Set

A total of 454 proteins were reproducibly identified as specifically associated with either of the chaperonins. LC-MS/MS also identified 711 of the 2546 predicted proteins in the *M. mazei* lysate.

To define basic rules that lead to chaperonin dependence of a protein for folding, the sets of *MmThs* and *MmGroEL* interactors were compared with proteins in *M. mazei* lysate by a variety of criteria. The proteins were compared according to their relative abundance, which is indicated by the corresponding emPAI score (exponentially modified Protein Abundance Index; (Lasonder *et al.*, 2002). As the emPAI score is a relative unit of a protein within one sample, the emPAI scores were normalized.

The substrate sets of both chaperonins show a considerable overlap and therefore four different substrate sets were defined for the bioinformatic analysis: all proteins that were found to interact with *MmThs* or *MmGroEL* were subsumed to the ***MmThs* substrate** set or the ***MmGroEL* substrate** set, respectively. Additionally, to refine the differences between the two chaperonin groups, the specific substrate set were compared. The ***MmThs* specific** and the ***MmGroEL* specific substrate** set comprises of proteins that are only found on the respective chaperonin.

#### *Contribution of the specific substrates to the total chaperonin substrate set*

As mentioned above, the substrate sets of both chaperonins show a high degree of overlap. 263 (60%) of the 454 identified substrate proteins were identified as interacting with the group I and group II chaperonin. Among both the *MmThs* substrates and the *MmGroEL* substrates about 30% were specific substrates (Table 4).

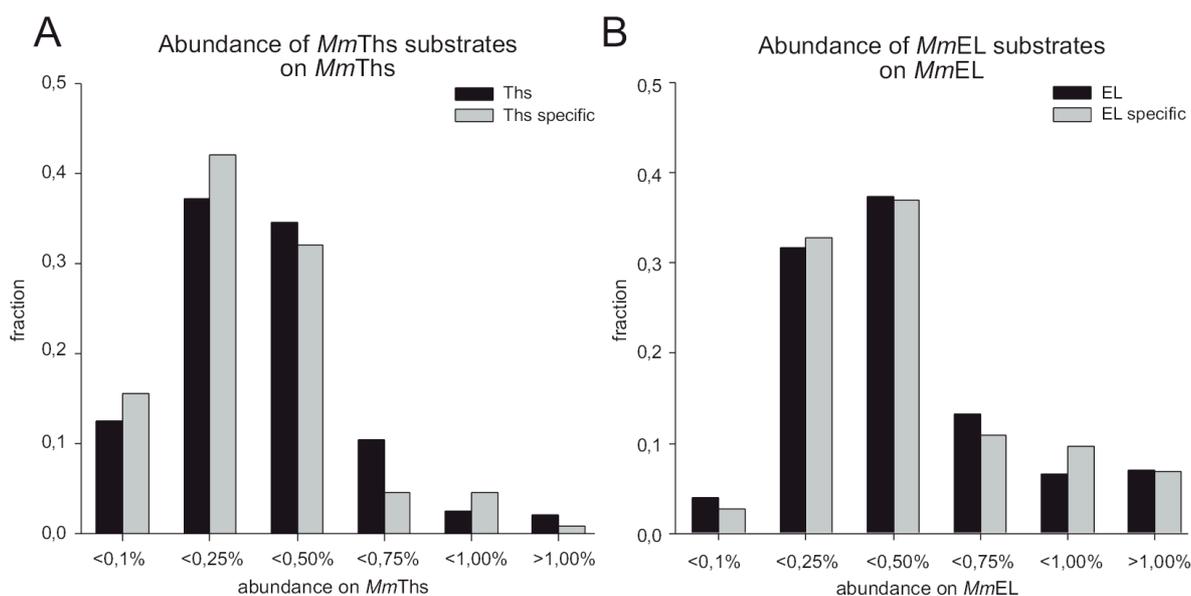
To assess the contribution the substrate specificity of each chaperonin to the protein folding in the cell, the abundance of specific substrates within the corresponding substrate set was compared. The abundance of a protein was calculated from the normalized emPAI score from each substrate sample. Similar to the absolute number of specific substrates among the

complete substrate set, specific substrates contribute ~ 30% of the total mass of the corresponding substrate set (Table 4).

To rule out if this similarity is due to an average frequency of specific substrates throughout the total substrate set, the abundance of the distinct specific proteins was assessed (Figure 32). Specific substrates were present at all levels of abundance of the substrate set.

*Table 4 : Distribution of specific and overlapping substrates of M. maze chaperonins*

	<i>MmThs</i>	<i>MmGroEL</i>
Abundance of specific substrates	30%	33%
Abundance of overlapping substrates	70%	67%
Number of specific substrates	117 (37%)	74 (35%)
Number of overlapping substrates	198 (63%)	153 (65%)



**Figure 33 : Abundance of specific substrates on the *M. mazei* chaperonins**

Comparison of the frequency of specific *MmThs* (A) and specific *MmGroEL* (B) to the complete substrate sets. Specific substrates are evenly distributed among the complete substrate sets.

### *Essentiality of Substrates.*

GroEL and GroES are generally essential in *bacteria*, this is also the case for TRiC in *Eukarya* and can be expected for the thermosome in *Archaea* (Kapatai *et al.*, 2006). However, due to the lack of genetic tools to manipulate *M. mazei*, it can only be assumed that both chaperonins are essential for survival of *M. mazei*.

The essentiality of GroEL/GroES in *E. coli* cells can be explained by its requirement for efficient folding of several (13) essential proteins. Due to the lack of information about essentiality of proteins in *M. mazei* all *MmThs* specific and *MmGroEL* specific substrates were manually checked for their potential indispensability for cell survival. First all the proteins were compared with essential proteins of bacteria. But, because of the limited homologies between many archaeal and bacterial proteins, all substrate proteins of known function were assigned to their pathway according to the Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>). If a protein was judged to play an irreplaceable role in a pathway important for cell viability, this protein was categorized as an essential protein. This way 17 (14.5%) of the 117 *MmThs* specific substrates and 14 (19%) of the 74 *MmGroEL* specific substrates were classified as potentially essential (Table 5A, B).

Two of the 14 potentially essential *MmGroEL* specific substrates, a PPIase (gj|20907465) and the suppressor protein suhB (gj|20907414), are homologous to stringent GroEL interactors in *E.coli* (Kerner *et al.*, 2005). The interaction of both chaperonin groups with a significant number of essential proteins suggests that both proteins are required for cell survival, yet it can not be excluded that the folding of these essential proteins could be taken over by the other chaperonin group.

Interestingly, two of the potentially essential proteins are of bacterial origin. These proteins, MutS (gj|20906191) and the ribosomal protein S6 modification protein (gj|20906929), interact specifically only with the bacterial group I chaperonin. Essential proteins that interact specifically with *MmThs* are exclusively of archaeal origin.

Table 5A: Essential Proteins among MmGroEL Specific Substrates

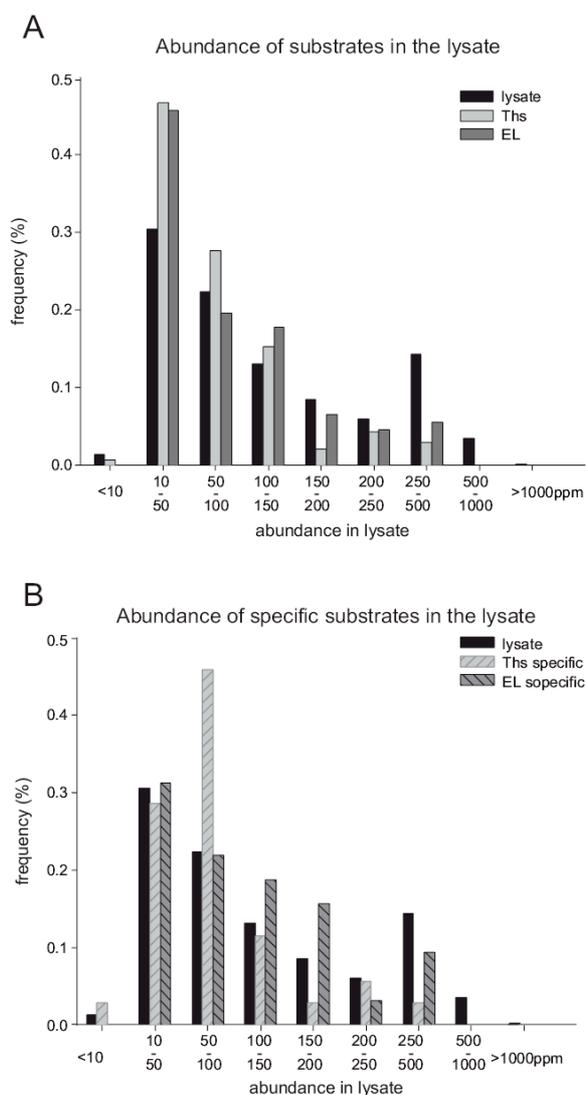
<i>gi number</i>	<i>Description</i>	<i>Gene Symbol</i>	<i>function</i>	<i>enrichment factor relative to lysate</i>	<i>abundance among EL specific substrates</i>	<i>abundance among EL substrates</i>
20904390	Tungsten formylmethanofuran dehydrogenase subunit D (EC:1.2.99.5)		methanogenesis	0.09	0.6%	0.2%
20904820	NH(3)-dependent NAD(+) synthetase (EC:6.3.1.5)	nadE	nitrogen/nicotinamide metabolism	0.7	1.6%	0.5%
20907414	Suppressor protein SuhB homolog (EC:3.1.3.25)		carbohydrate metabolism	2	0.9%	0.3%
20906036	Tetrahydromethanopterin S-methyltransferase subunit B (EC:2.1.1.86)	mtrB	methanogenesis	2.4	3.4%	1.2%
20907091	Biotin synthase (EC: 5.3.1.16)		Biotin metabolism	6.6	1.1%	0.4%
20905556	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC:5.3.1.16)		purine/histidine biosynthesis	9.1	9.7%	3.1%
20904436	Serine/threonine protein phosphatase (EC:3.1.3.16)		signal transduction	n.d.	0.6%	0.2%
20904704	Amidophosphoribosyltransferase (EC:2.4.2.14)		purine metabolism	n.d.	1.3%	0.4%
20906191	DNA mismatch repair protein mutS	mutS	replication, recombination and repair	n.d.	5.8%	1.8%
20906323	Aspartate aminotransferase (EC:2.6.1.1)		Aminoacid metabolism	n.d.	1.5%	0.5%
20906589	Aspartate aminotransferase (EC:2.6.1.1)		Aminoacid metabolism	n.d.	1.8%	0.7%
20906929	Ribosomal protein S6 modification protein		coenzyme transport and metabolism	n.d.	1.1%	0.3%
20907999	Threonyl-tRNA synthetase (EC:6.1.1.3)	thrS	translation	n.d.	0.8%	0.3%
					30.2%	9.8%

Table 5B: Essential Proteins among MmThs Specific Substrates

<i>gi number</i>	<i>Description</i>	<i>Gene Symbol</i>	<i>function</i>	<i>enrichment factor relative to lysate</i>	<i>abundance among Ths specific substrates</i>	<i>abundance among Ths substrates</i>
20907953	Coenzyme F420 hydrogenase beta subunit (EC:1.12.99.1)		Methanogenesis	n.d.	1.2%	0.32%
20907921	Thiamine-phosphate pyrophosphorylase (EC:2.5.1.3)	thiE	thiamin metabolism	n.d.	0.5%	0.12%
20906844	DNA-directed RNA polymerase beta chain (EC:2.7.7.6)		Transcription	0.7	3.4%	0.89%
20906821	Tetrahydromethanopterin S-methyltransferase, subunit A (EC:2.1.1.86)		Methanogenesis	2.1	0.7%	0.20%
20906609	Histidinol-phosphate aminotransferase		Aminoacid metabolism	n.d.	0.2%	0.05%
20906060	Molybdenum formylmethanofuran dehydrogenase subunit (EC:1.2.99.5)	fmdA	Methanogenesis	n.d.	0.1%	0.04%
20905956	Tryptophanyl-tRNA synthetase (EC:6.1.1.2)	trpS	Translation	n.d.	0.5%	0.12%
20905876	DNA-directed RNA polymerase subunit M (EC:2.7.7.6)		Transcription	n.d.	1.1%	0.31%
20905409	CoB--CoM heterodisulfide reductase 1 subunit B (EC:1.8.98.1)	hdrB	Methanogenesis	n.d.	1.1%	0.27%
20905396	Coenzyme F420 hydrogenase beta subunit (EC:1.12.99.1)		Methanogenesis	1.1	1.9%	0.50%
20905283	Methionyl-tRNA synthetase (EC:6.1.1.10)	metG	Translation	1	1.3%	0.34%
20905197	Geranyltranstransferase		sterole biosynthesis	n.d.	1.5%	0.36%
20905038	3-isopropylmalate dehydratase small subunit 1 (EC:4.2.1.33)	leuD1	Aminoacid metabolism	1.3	1.1%	0.29%
20904988	30S ribosomal protein S27ae	rps27Ae	Translation	2.5	3.2%	0.94%
20904984	DNA-directed RNA polymerase subunit E' (EC:2.7.7.6)		Transcription	0.2	0.6%	0.16%
20904454	CTP synthase (EC:6.3.4.2)	pyrG	nucleotide transport and metabolism	6.3	0.6%	0.15%
					19.0%	5.0%

### Concentration of substrate proteins in the cytosol

In order to assess the quantitative contribution of the chaperonins to the protein folding in the archaeal cell, the substrates were analysed for their relative abundance in the cytosol. The abundance was calculated according to the abundance of the substrate proteins in the *M. mazei* lysate.



**Figure 34 : Abundance of *M. mazei* chaperonin substrate proteins in the cytosol**

The abundance of substrate proteins from the chaperonin interacting (A) and specifically interacting was assessed and analysed according to the abundance of these proteins in the corresponding substrate set

The majority of substrates of both chaperonins are found at low expression levels between 10 and 100 ppm in the cell (Figure 33A). *MmGroEL* specific substrates are present at levels

below 500 ppm at similar frequency as the average of the lysate proteins, the majority of *MmThs* specific substrates are expressed below 100 ppm of the soluble proteins (Figure 33B).

More than half of each substrate set could not be identified in the cytosol fraction (Table 6). This might be due to a very low abundance of these proteins and underline the low abundance of *M. mazei* chaperonin substrates in the cell. On the other hand it can not be excluded that proteins were not found because the lysate was analysed only once and only 711 (20%) of the 3370 annotated gene products were identified.

Table 6: *M. mazei* chaperonin substrates - not identified in the *M. mazei* lysate

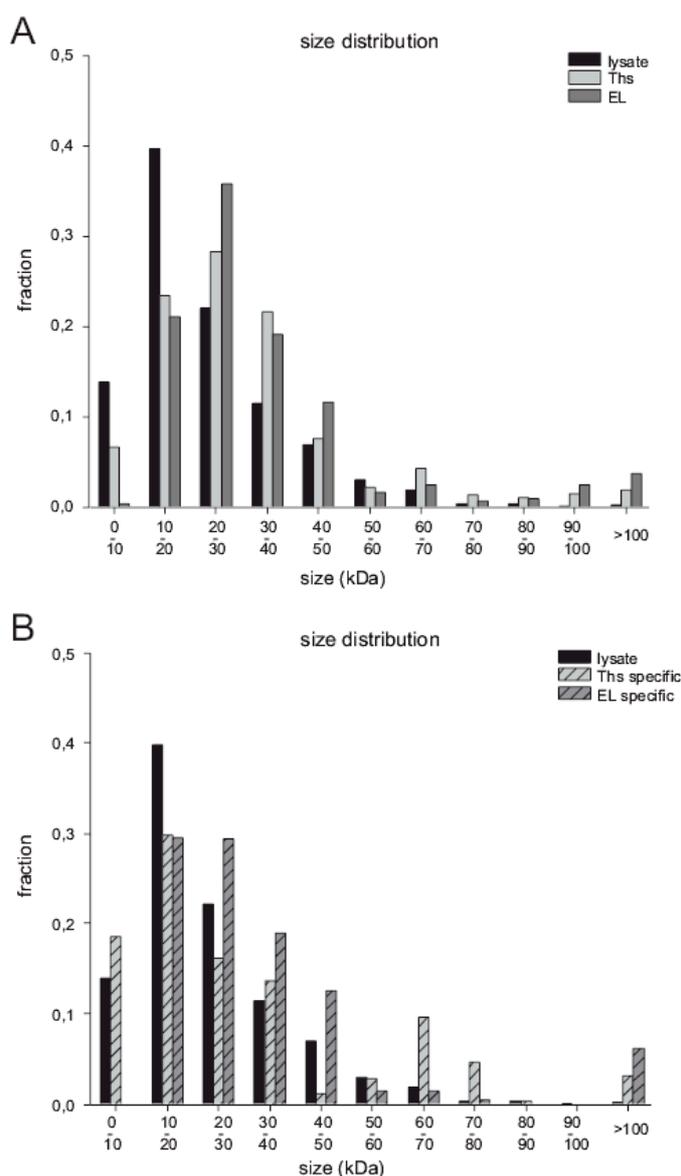
	<i>Ths</i>	<i>EL</i>	<i>Ths</i> specific	<i>EL</i> specific
not identified in the lysate	178 (56 %)	120 (53%)	80 (68%)	42 (57%)

### *Significance of structural properties for chaperonin dependence*

#### **Size distribution among chaperonin substrates**

The GroEL/GroES folding cage has a volume capacity of for proteins up to 60 kDa in size (Sakikawa *et al.*, 1999). Such a size limitation is not expected for *Ths* substrates because of the different closing mechanism of the cavity by the helical protrusions. It is speculated that proteins exceeding the theoretical cavity size could eventually be accommodated in part, with the remaining of their sequence protruding from the cage.

The molecular mass distribution of both the group I and the group II chaperonin substrate sets are shifted towards larger sizes compared to that of the total lysate proteins. The majority of the substrates is found at sizes below 50kDa, preferentially between 20 kDa and 30 kDa, consistent with the dependence on an encapsulation mechanism for folding assistance (Figure 35A). Proteins smaller than 10 kDa are thought to fold mostly spontaneous and are rarely found as *M. mazei* chaperonin interactors.



**Figure 35 : Size distribution among chaperonin substrates**

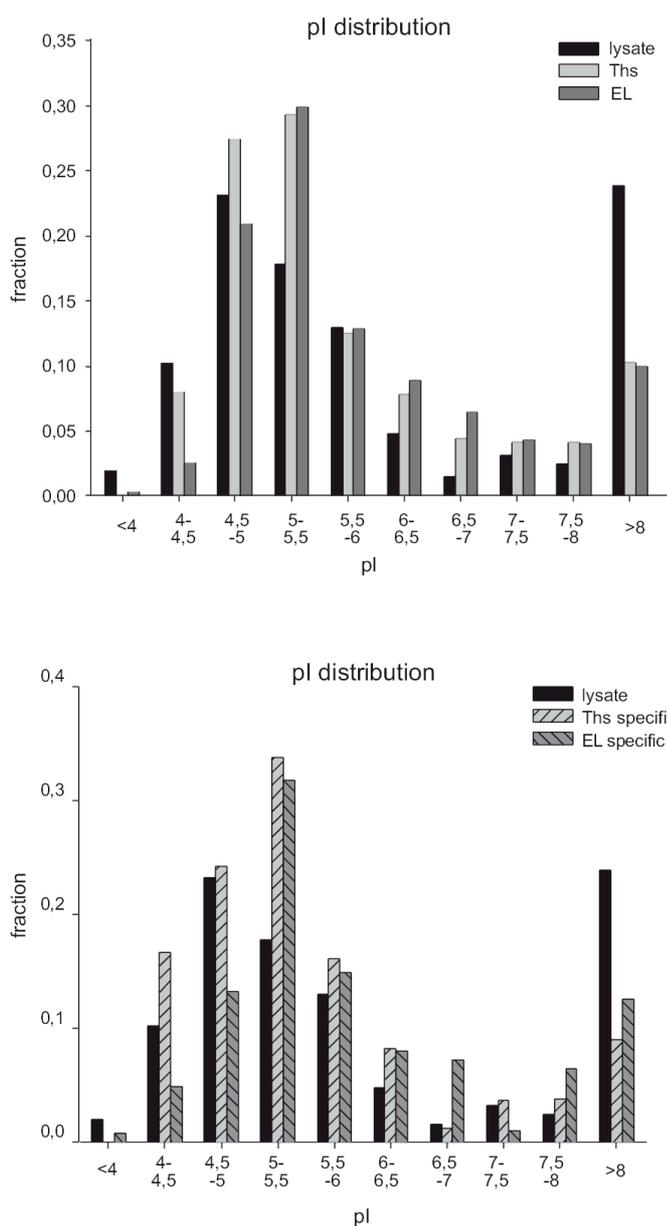
Size distribution of proteins interacting with *MmThs* or *MmGroEL* (A) and of specifically interacting substrates (B) according to their relative abundance compared with relative abundance in *M. mazei* lysate.

Differences on the size of the chaperonin substrates become apparent among the specific substrate sets. There is no significant change in the size preference of *MmGroEL* specific substrates and the complete *MmGroEL* substrate set (between 10 and 40 kDa). But in contrast to the complete substrate set, a relevant fraction of *MmThs* specific substrates is found at a range between 60 and 80 kDa and also below 10 kDa. Proteins above 100 kDa show no preferential interaction with the group II chaperonin *MmThs* (Figure 35B).

### ***Isoelectric point distribution of substrates***

The isoelectric point (pI) is defined as the pH at which a protein carries no net electrical charge. In biological systems, where prevalently neutral pH values are given, a isoelectric point (pI) value close to pH 7 indicates that a protein is mostly uncharged. The pIs of the annotated genes of *M. mazei* were downloaded from <http://www.neurogadgets.com>.

Compared to the average pI of lysate proteins, the pIs of both the substrates of group I and the group II chaperonin show a shift of 0.5 towards a neutral value (Figure 36). It is likely that the cytosolic pH of *M. mazei* is like in most organisms close to pH 7.0 and therefore at physiological conditions these proteins exhibit a lower overall net charge than an average cellular protein, a property that is expected to enhance the tendency of proteins to aggregate at non native folded states (Chiti *et al.*, 2002). This effect is more pronounced for *MmGroEL* substrates than for *MmThs* substrates, especially for the *MmGroEL* specific substrates.



**Figure 36 : pI distribution among *M. mazei* chaperonin substrates**

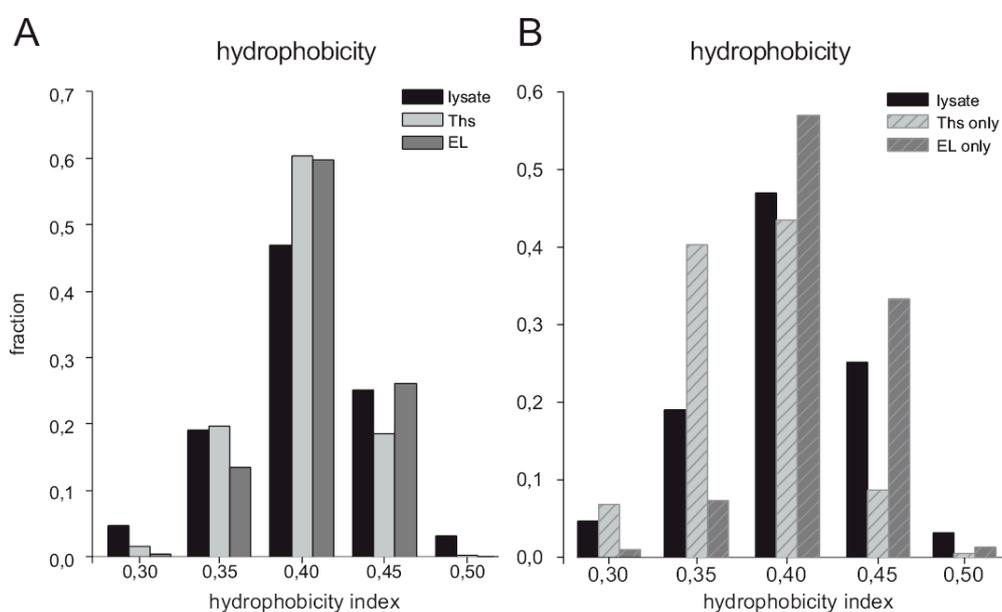
pI distribution of substrates interacting with *M. mazei* chaperonins (A) and *MmThs/MmGroEL* specific substrates (B) compared to lysate proteins. Both chaperonin substrate set are shifted towards neutral pI.

### **Relative hydrophobicity of substrates**

GroEL is known to bind substrates via hydrophobic interactions. Therefore it may be expected that *MmGroEL* interacts preferentially with proteins characterized by a higher content of hydrophobic aminoacid side chains than average. The hydrophobicity of a protein is

indicated by the product proportion of hydrophobic amino acids, these being Ala, Ile, Leu, Met, Phe, Pro, Trp, Val of an ORF. The hydrophobicity index of *M. mazei* proteins were obtained from <http://www.neurogadgets.com>.

The total of the chaperonin substrates displays no significant preference for proteins with a higher hydrophobicity index compared to the lysate fraction (Figure 37). However substrates, which interact specifically only with *MmGroEL*, exhibit a significantly higher hydrophobicity than the average lysate proteins. In contrast, proteins interacting specifically with the *MmThs* showed a reduced hydrophobicity (Figure 37B).

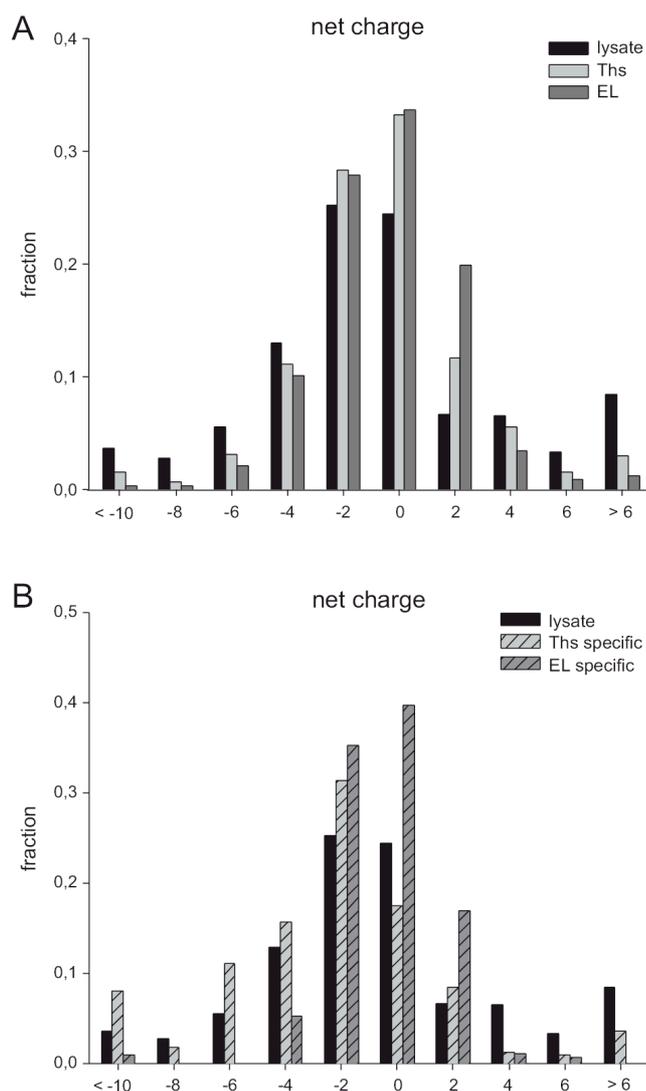


**Figure 37 : Hydrophobicity of *M. mazei* chaperonin substrate proteins**

Relative hydrophobicity of substrates interacting with *M. mazei* chaperonins (A) and *MmThs*/*MmGroEL* specific substrates (B) compared to lysate proteins. *MmGroEL* specific proteins show an increased hydrophobicity than *MmThs* (specific) substrates and lysate proteins.

## Net charge of substrate proteins

In contrast to the group I chaperonins that are supposed to bind substrate proteins via hydrophobic interactions, group II chaperonins are thought to bind proteins via polar interactions, at least in part. Analysis of the substrate binding domains of the eukaryotic group II chaperonin TRiC revealed that relevant positions are occupied by amino acids that are positively charged at physiological pH (Pappenberger *et al.*, 2002).



**Figure 38 : Net Charge Distribution among *M. mazei* chaperonin Substrates**

Net charge distribution among substrates interacting with *M. mazei* chaperonins (A) and *MmThs*/*MmGroEL* specific substrates (B) compared to lysate proteins. A relevant fraction of *MmThs* specific substrates show high negative charge. This is not found for *MmGroEL* (specific) substrates, which are mostly uncharged.

---

Intriguingly, the comparison of the net charge distribution (obtained from <http://www.neurogadgets.com>) of *Mm*Ths specific substrates shows that a substantial fraction of these proteins is negatively charged under physiological pH, which is not seen for the average lysate proteins or GroEL interacting proteins (Figure 38B).

### ***Transmembrane domains***

Proteins that are associated with membranes usually use the so-called translocon system for insertion or transport through the membrane (Eichler and Duong, 2004). Neither the group I nor the group II chaperonin system is expected to be generally involved in these machineries. Though, productive interaction of GroEL with membrane proteins has been reported for some membrane proteins (Bochkareva *et al.*, 1998; Goulhen *et al.*, 2004; Gozu *et al.*, 2002; Hanninen *et al.*, 1997; Sun *et al.*, 2005).

Therefore it is not surprising that only a very minor fraction of *M. mazei* chaperonin substrates exhibit transmembrane domains (obtained from <http://pedant.gsf.de/index.html>, Figure 39A). Three among *Mm*GroEL specific and two *Mm*Ths specific proteins contain predicted membrane-spanning domains, the role of the interaction of the chaperonins with this proteins will have to be further elucidated (Table 7).

Table 7A: *MmGroEL* Substrates exhibiting Transmembrane Motifs

<i>gi number</i>	<i>Description</i>	<i>Gene Symbol</i>	<i>function</i>	<i>enrichment factor relative to lysate</i>	<i>abundance among GroEL specific substrates</i>	<i>abundance among GroEL substrates</i>
20904704	Amidophosphoribosyl-transferase (EC:2.4.2.14)		nucleotide transport and metabolism	n.d.	1.0%	1%
20905590	Glycosyltransferase (EC:2.4.1.-)		cell wall biogenesis	n.d.	0.3%	0%
20906036	Tetrahydromethanopterin S-methyltransferase subunit B (EC:2.1.1.86)	mtrB	coenzyme transport and metabolism	2.4	3.0%	2%
					4.3%	3%

Table 7B: *MmThs* Substrates exhibiting Transmembrane Motifs

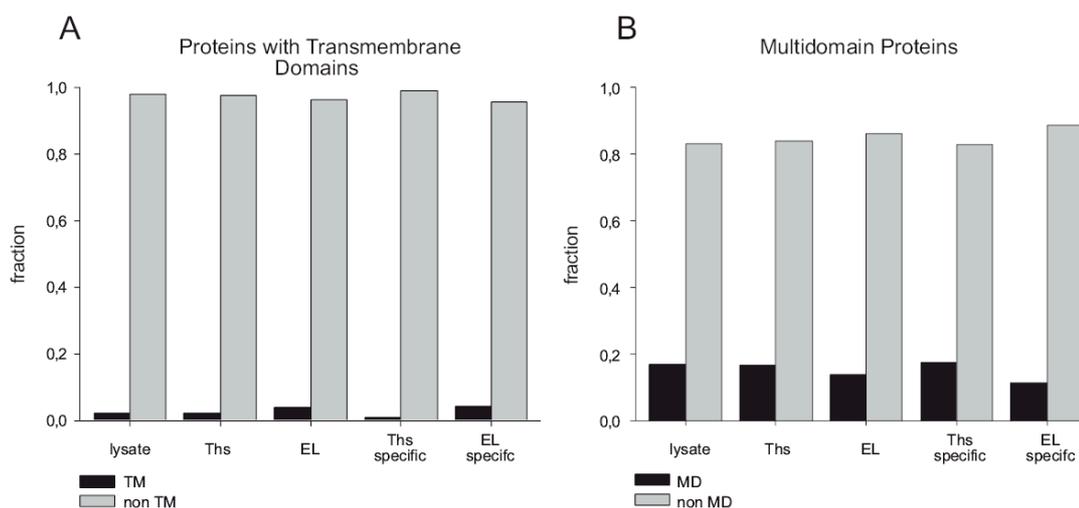
<i>gi number</i>	<i>Description</i>	<i>Gene Symbol</i>	<i>function</i>	<i>enrichment factor relative to lysate</i>	<i>abundance among Ths specific substrates</i>	<i>abundance among Ths substrates</i>
20904970	conserved protein			n.d.	1.1%	0.3%
20905880	Phosphatidylserine decarboxylase proenzyme (EC:4.1.1.65)		lipid transport and metabolism	n.d.	0.3%	0.1%
					1.4%	0.4%

Table 7C : Overlapping Substrates exhibiting Transmembrane Motifs

<i>gi number</i>	<i>Description</i>	<i>Gene Symbol</i>	<i>function</i>	<i>enrichment factor relative to lysate</i>	<i>abundance among GroEL substrates</i>	<i>abundance among Ths substrates</i>
20905395	Glutamate synthase, large chain (EC 1.4.1.13). Methanosarcina mazei			4.6 on <i>MmGroEL</i> 5.2 on <i>MmThs</i>	3,3%	2,6%
20905193	V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I). Glycosyltransferase (EC 2.4.-.-)	AHA1, ATPI	Energy production and conversion genes	n.d.	0,6%	0,3%
20905045			Cell wall/membrane biogenesis	n.d.	0,6%	0,3%

## Multidomain Proteins

Because of the different closing mechanism of the group II chaperonins *via* the helical protrusion this chaperonin has been speculated to be more suited for folding of multidomain proteins. The closing mechanism could enable the group II chaperonin to encapsulate the protein domains sequentially and allow independent folding of these protein modules. However, analysis of the substrate sets of *MmThs* showed no preference for multidomain proteins (obtained from <http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY>) (Figure 39B).



**Figure 39 : Distribution of Proteins harbouring Transmembrane Domains and Multidomain Proteins**

Proteins exhibiting transmembrane domains are rarely found to interact with chaperonins, a significant number of these proteins interact with *MmGroEL* (A). Proteins consisting of more than one domain fold show no preferential binding to any *M. mazei* chaperonin (B).

### **Domain fold distribution among substrates**

A recent study by Kerner *et al.* concluded that fold types play an important role in the substrate selection by the group I chaperonin in *E. coli*. It was shown that proteins containing domains with the TIM barrel are overrepresented among GroEL substrates compared to the total lysate proteins.

Analysis of the domain fold distribution of *M. mazei* proteins that interact with the chaperonins (obtained from <http://supfam.mrc-lmb.cam.ac.uk>) gives a more differentiated picture of fold type preferences of the chaperonins (Figure 40). In contrast to other organisms, where only one chaperonin is available, proteins have the choice between both chaperonins and therefore proteins of archaeal origin can also interact with the bacterial chaperonin and *vice versa*.

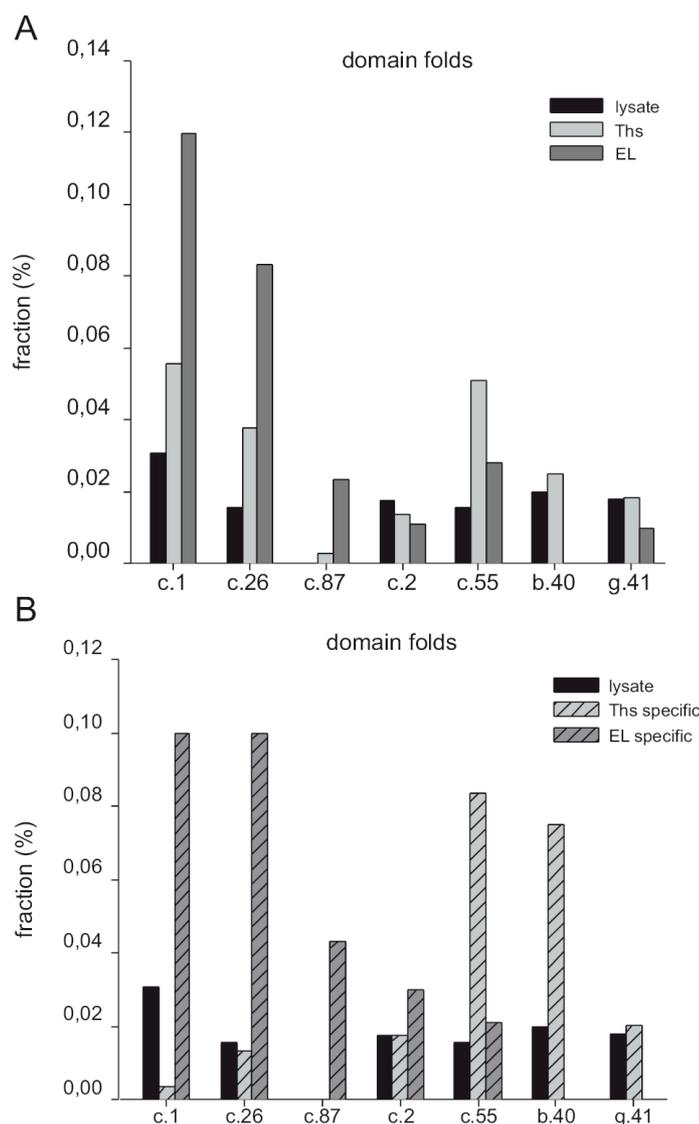
The fold-categorisation of the identified chaperonin substrates in *M. mazei* according to the SCOP data base confirmed that the TIM barrel fold represents the largest fraction of interactors with the bacterial group I chaperonin (12% of total GroEL interactors and 10 % if specific GroEL interactors according to their relative abundance). This fraction of GroEL interactors comprises

19 proteins of all GroEL substrates and 3 proteins of the GroEL specific substrates. A significant number of TIM barrel proteins (20) interact also with the Group II chaperonin, but in this case the TIM barrel domain proteins present only 5% of the fraction of all known folds, insignificantly higher than the average abundance of this fold in the lysate fraction (3%).

In addition to the TIM barrel fold, structurally related members of c-class folds, defined as alpha and beta proteins ( $\alpha/\beta$ ) of mainly parallel beta sheets ( $\beta/\alpha/\beta/$  units) are of high relative abundance among *MmGroEL* (specific) substrates: the adenine nucleotide alpha hydrolase-like (c.26; core: 3 layers,  $\alpha/\beta/\alpha$ ; parallel beta-sheet) and the UDP-glycosyltransferase/glycogen phosphorylase fold (c.87; two non-similar domains with 3 layers ( $\alpha/\beta/\alpha$ ) each domain parallel beta-sheets).

The fraction of proteins with the c.26 domain fold (8% of the total GroEL substrates and 10% of the specific GroEL substrates by relative abundance) comprises 14 proteins of the total GroEL substrates and 4 proteins of the specific GroEL substrates. Again the c.26 domain fold is also found among *MmThs* substrates. The c.26 domain fold is found among 12 *MmThs* interactors and at similar frequency as among lysate proteins (4% and 1.5% among *MmThs* and *MmThs* specific substrates, respectively and 1.5% in the lysate).

8 proteins and 4 proteins sharing the c.87 domain fold represent 2% of the total GroEL interactors and 4% of the specific GroEL interactors, respectively. Proteins with the c.87 domain fold are underrepresented among *MmThs* substrates (2 proteins, 0.3% of *MmThs* substrates) and not found among *MmThs* specific substrates. This domain fold has also not been identified in the lysate fraction, suggesting a low abundance of proteins sharing this domain fold in the archaeal cell.



**Figure 40 : Fold distribution of *M. mazei* chaperonin substrates according to the SCOP database**

Domain folds of *MmGroEL* and *MmThs* substrates (A) and specific *MmGroEL* and *MmThs* substrates (B). Members of the c - class folds c.1, c.26 and c.87 show preferential binding to *MmGroEL*, while the c-class fold c.55 and the b-class fold b.40 accumulate on *MmThs*. c.1: TIM  $\beta/\alpha$  barrel; c.2: NAD(P)-binding Rossmann-fold domains; c.26: Adenine nucleotide alpha hydrolase-like; c.87: UDP-Glycosyltransferase/glycogen phosphorylase; c.2; c.55: Ribonuclease H-like motif; b.40: OB-fold; g.41: Rubredoxin-like; d.37: CBS-domain.

Intriguingly, among the folding types of *MmThs* substrates, members of the ribonuclease H-like fold, which includes the family of actin ATPase like- domain superfamily (c.55.1). 9 of the 11 *MmThs* substrate proteins belonging to this fold class interact exclusively with the group II

---

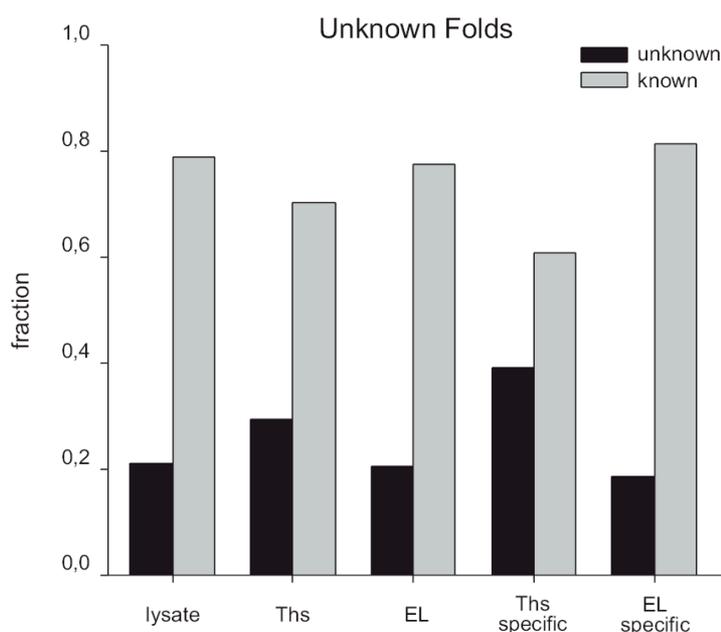
chaperonin. All proteins identified in the lysate sharing this fold were also identified in complex with *MmThs*.

In contrast to the group I chaperonin substrates members of the class-b folds (only beta) are also prominent members of the group II chaperonin substrates. The OB-fold (b.40, beta-barrel, closed or partly opened) is frequently found among *MmThs* substrates, but excluded from the *MmGroEL* substrates. Likewise, proteins of the nucleic-acid-binding protein-superfamily (b.40.4) occupy only 2% of the total *MmThs* capacity, but represent 7.5% of the mass of specific *MmThs* interactors.

### *Unknown folds*

Analysis of the fold distribution in *M. mazei* is limited by the number of soluble proteins with structural information, 933 (~35%) of the 2610 predicted soluble proteins in the genome of *M. mazei* show no homology to any known domain fold.

The abundance of proteins that show no homology to any known structure among *MmGroEL* and *MmGroEL* specific is found at similar levels as among the lysate fraction (20%). Interestingly, this fraction of proteins without any homology known structures is significantly increased among the *MmThs* substrates and highest (40%) among *MmThs* specific substrates (Figure 41).



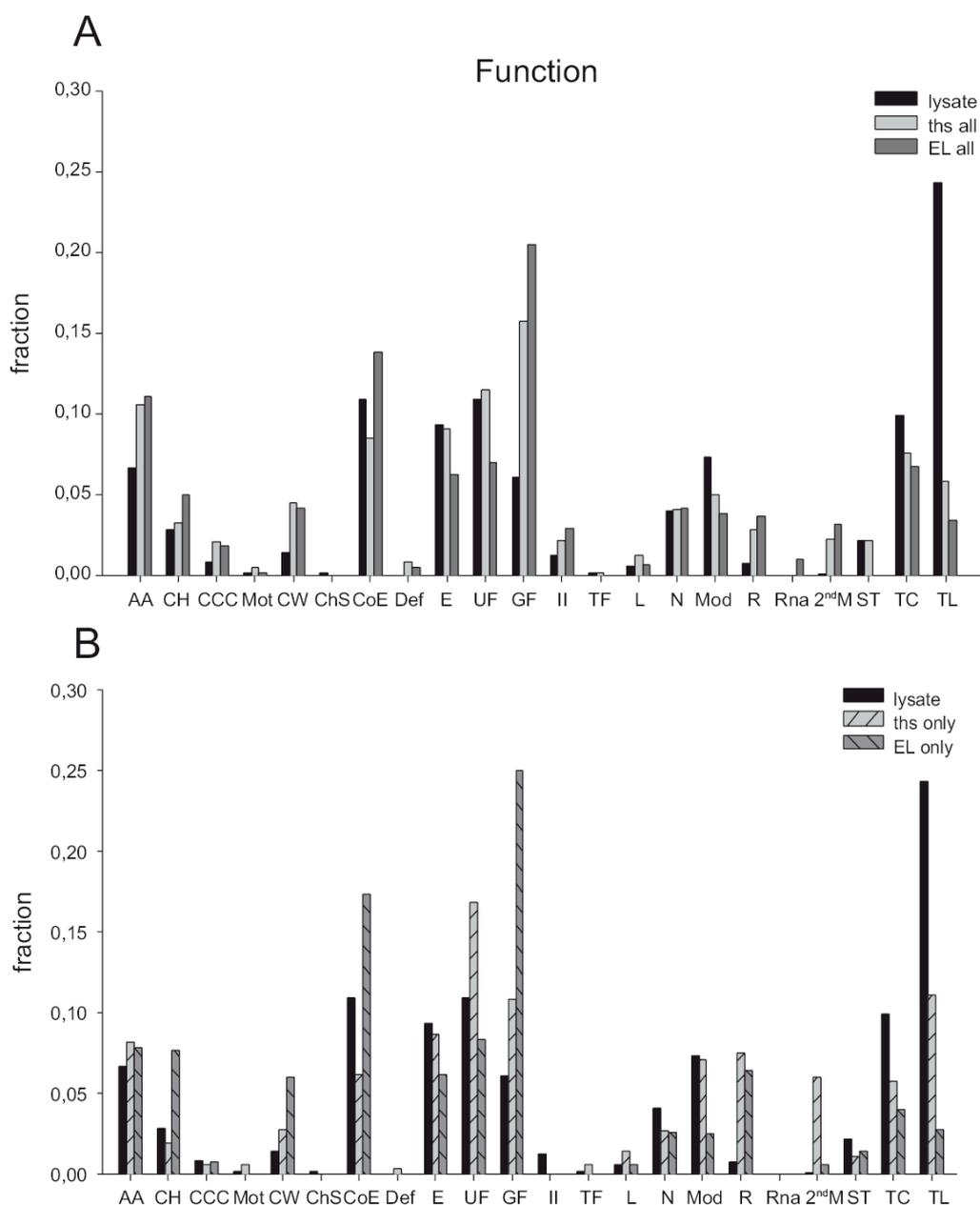
**Figure 41 : Distribution of unknown folds**

35% of the soluble proteins in identified in the genome *M. mazei* show no structural homology to any known domain fold, 20% of the protein mass in the lysate fulfils this criterion. Among *MmThs*, but not *MmGroEL* substrates this value increases two-fold.

### *Function of substrates*

It is not expected that the function of a substrate protein could define the specificity of the client protein for the chaperone type as the function is not influencing the structural features of a protein. When the function distribution of substrate proteins were analysed according to the Clusters of Orthologous Groups (COGs) (obtained from The Integrated Microbial Genomes (IMG) system, <http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>) no obvious trend of *M. mazei* chaperonin substrate proteins to be involved in a specific function could be detected.

The majority of the substrates fulfil functions that can only be defined by general prediction, among *MmThs* specific substrates most of the proteins can not be assigned to any known function at all. Proteins that are involved in translation are significantly reduced among *M. mazei* chaperonin substrates compared to lysate proteins (Figure 42).

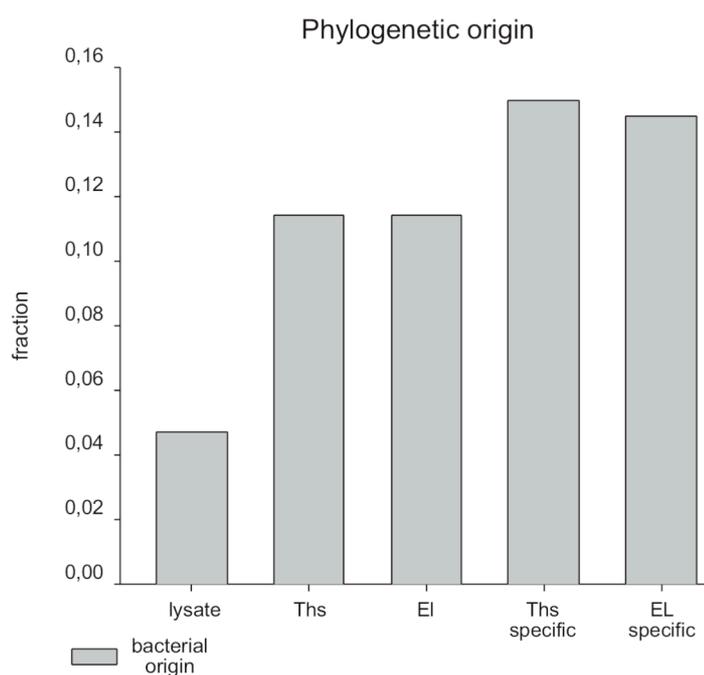


**Figure 42 : Function of *M. mazei* chaperonin substrates**

Aminoacid metabolism (AA), carbohydrate metabolism (CH), cell cycle control (CCC), cell motility (Mot), cell wall biogenesis (CW), chromatin structure (ChS), coenzyme transport and metabolism (CoE), defense mechanism (Def), energy production and conversion (E), function unknown (UF), general function prediction only (GF), inorganic ion transport and metabolism (II), intracellular trafficking, secretion and vesicular transport (TF), lipid transport and metabolism (L), nucleotide transport and metabolism (N), posttranslational modification, protein turnover, chaperones (Mod), replication, recombination and repair (R), rna processing and modification (Rna), secondary metabolites (2<sup>nd</sup> M), signal transduction (ST), transcription (TC), translation (TL). For most *M. mazei* chaperonin substrates only a general function can be assigned, most *MmThs* specific substrate can not be assigned to a known function.

### Phylogenetic origin

Analysis of the genome of *M. mazei* revealed an unusual high frequency of genes of bacterial origin that were acquired via horizontal gene transfer (HGT). Therefore it could have been expected that these originally bacterial proteins would preferentially interact with the group I chaperonin, which is also of bacterial origin. However, bacterial proteins occur at similar levels among *MmGroEL* substrates as among *MmThs* substrates (Figure 43). Interestingly, there seems to be a strong tendency of proteins derived from HGT to interact with chaperonins. In contrast to the high frequency of 30 % of bacterial proteins annotated in the genome of *M. mazei* only 5% of the lysate proteins were identified as originally bacterial proteins. This can be explained on the one hand by a high contribution of transposons to the fraction of bacterial genes on the archaeal chromosome. On the other hand the presence of a gene gives no information on its expression level and many bacterial genes might be not expressed under standard condition.



**Figure 43 : Phylogenetic origin of *M. mazei* chaperonin substrates**

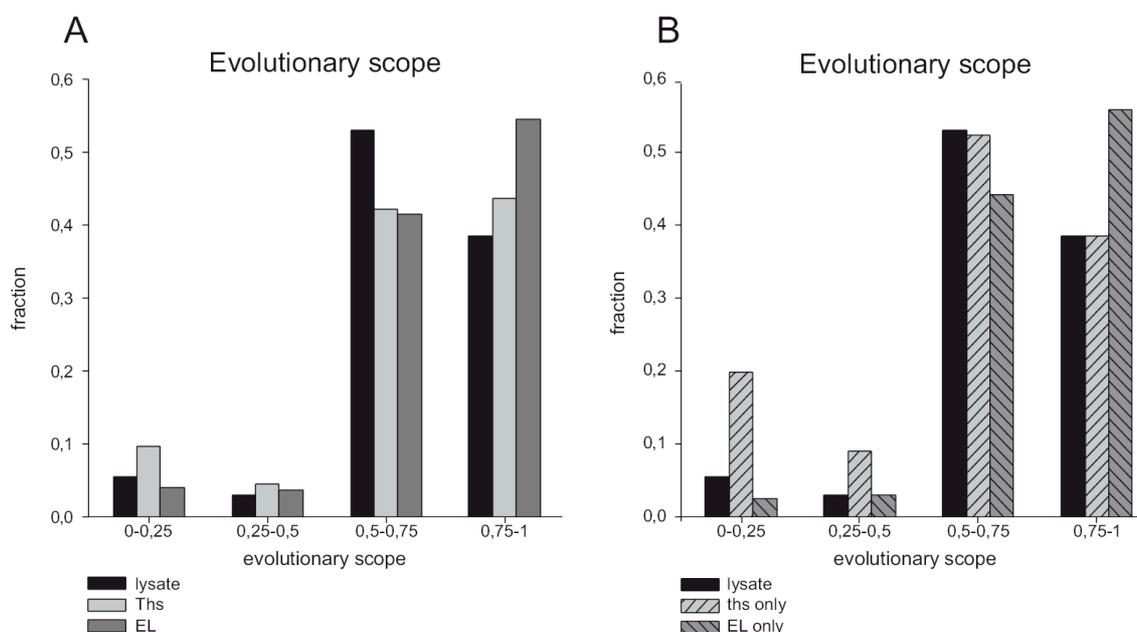
Abundance of bacterial proteins in the lysate fraction and among *M. mazei* chaperonin substrates. Bacterial proteins show no preferential interaction with the bacterial chaperonin *MmGroEL*.

### *Evolutionary scope*

Recently, it was shown by Cowen L.E. and Lindquist S. that the eukaryotic chaperone Hsp90 may influence evolution by supporting genetic variations of proteins in response to environmental change. There are so far no Hsp90 homologues found in *Archaea* and thus the question arose whether chaperonins could play a role in supporting evolution of proteins in the archaeal cell.

The evolutionary scope value of a protein (obtained from <http://www.neurogadgets.com>) gives the mean distance between members of order *Methanosarcinales*, if a phylogenetic tree was based on the sequence of this protein. Proteins have a low evolutionary scope that if they show a higher homology to proteins from *Methanosarcinales* than to proteins from other organisms. So a low evolutionary scope indicates a relatively high frequency of mutation in a protein. Conserved proteins have a high evolutionary scope, because they artificially elongate the phylogenetic distance between the species.

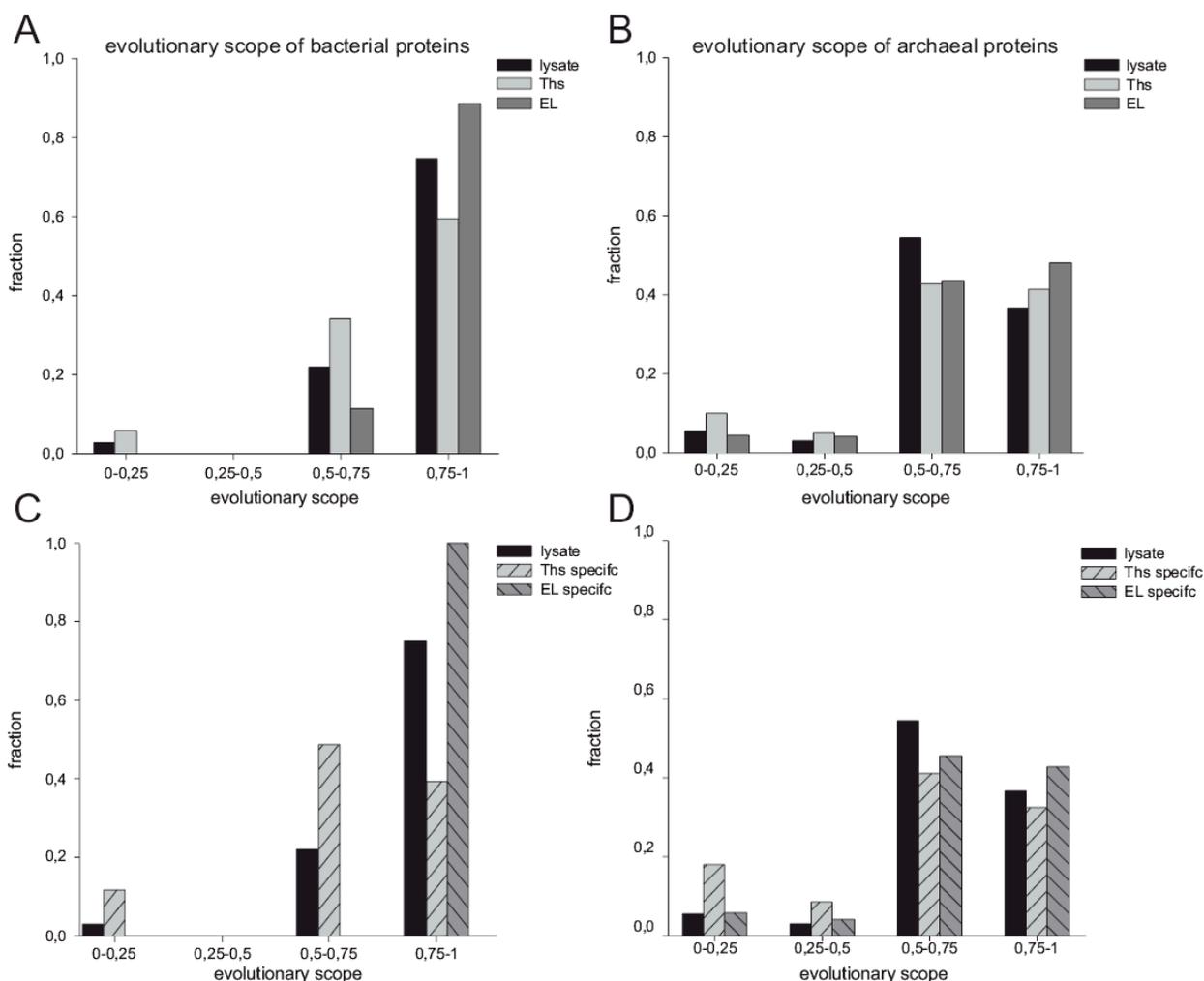
When the *M. mazei* chaperonin substrates were analysed by their evolutionary scope, a significant fraction of *MmThs* substrates and especially *MmThs* specific substrates showed an unusually low evolutionary scope value below 0.25 compared to the average lysate proteins. Proteins that interact with the group I chaperonin can be expected to be highly conserved during evolution (Figure 44).



**Figure 44 : Evolutionary scope of *M. mazei* chaperonin substrates**

Evolutionary scope of *M. mazei* chaperonin substrates (A) and *Mm*Ths and *Mm*GroEL specific substrates (B). A low evolutionary scope indicates a recent development of the protein, while a highly conserved protein is indicated by a high evolutionary scope. A relevant fraction of *Mm*Ths (specific) substrates shows a recent development, *Mm*GroEL interacts preferentially with highly conserved proteins.

Because proteins of bacterial origin are found at similar frequency with both chaperonin groups, the evolutionary scope of these bacterial proteins was of special interest. Analysis of the conservation of the bacterial proteins that interact with *M. mazei* chaperonins revealed that the bacterial proteins that interact with *Mm*GroEL show an even higher conservation than it has been shown for all *Mm*GroEL substrates above (Figure 45A, C). Bacterial proteins that interact with *Mm*Ths are at a similar evolutionary state as the average of *Mm*Ths substrates. Substrate proteins of archaeal origin show no deviation in the distribution of the value for the evolutionary scope from the complete substrate sets (Figure 45B, D).

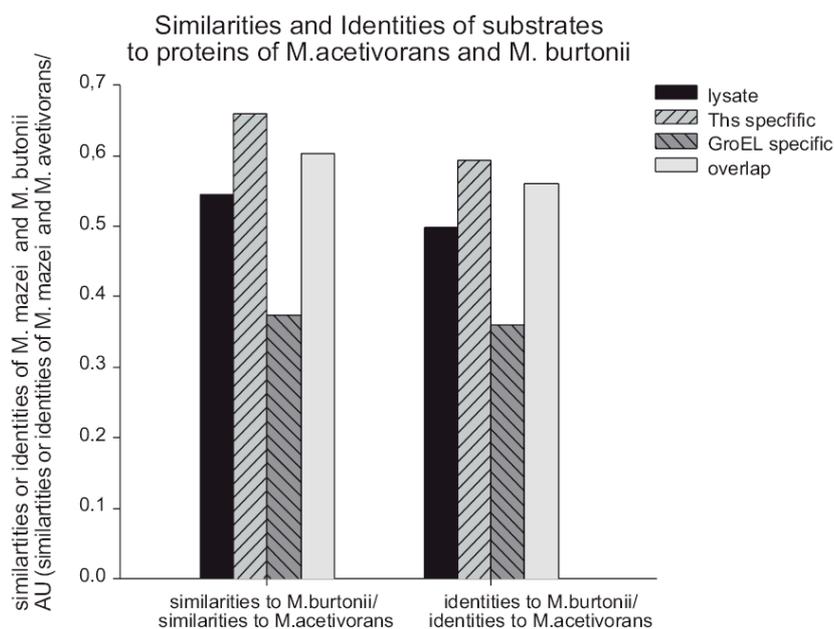


**Figure 45 : Evolutionary scope of bacterial and archaeal substrate proteins**

Substrates of *M. mazei* chaperonins were analysed according to their phylogenetic origin for their evolutionary scope. *MmGroEL* proteins of bacterial origin (A, C) show an unusual high conservation, while bacterial *MmThs* substrate change at similar frequency as average *MmThs* substrates and even faster than average bacterial lysate proteins. Proteins of archaeal origin (B, D) develop at a similar rate independent of the chaperonin dependency.

Even though the evolutionary scope of Group I interacting proteins showed an overall high conservation of these proteins compared to all available sequences, the evolutionary distance of the specific substrate proteins was analysed within the members of *Methanosarcinales* separately. It would be expected that especially proteins that depend on the Group I chaperonin in *M. mazei* are more different to homologous proteins from a relative that holds no GroEL (*Methanococcoides burtonii*) than to other *Methanosarcina*, which also have GroEL.

Therefore the specific substrate proteins were blasted against the genomes of *Methanosarcina acetivorans* and *M. burtonii*. Proteins that interact with GroEL specifically showed significantly reduced similarities and identities in the aminoacid sequence to *M. burtonii* than to *M. acetivorans* (Figure 46).



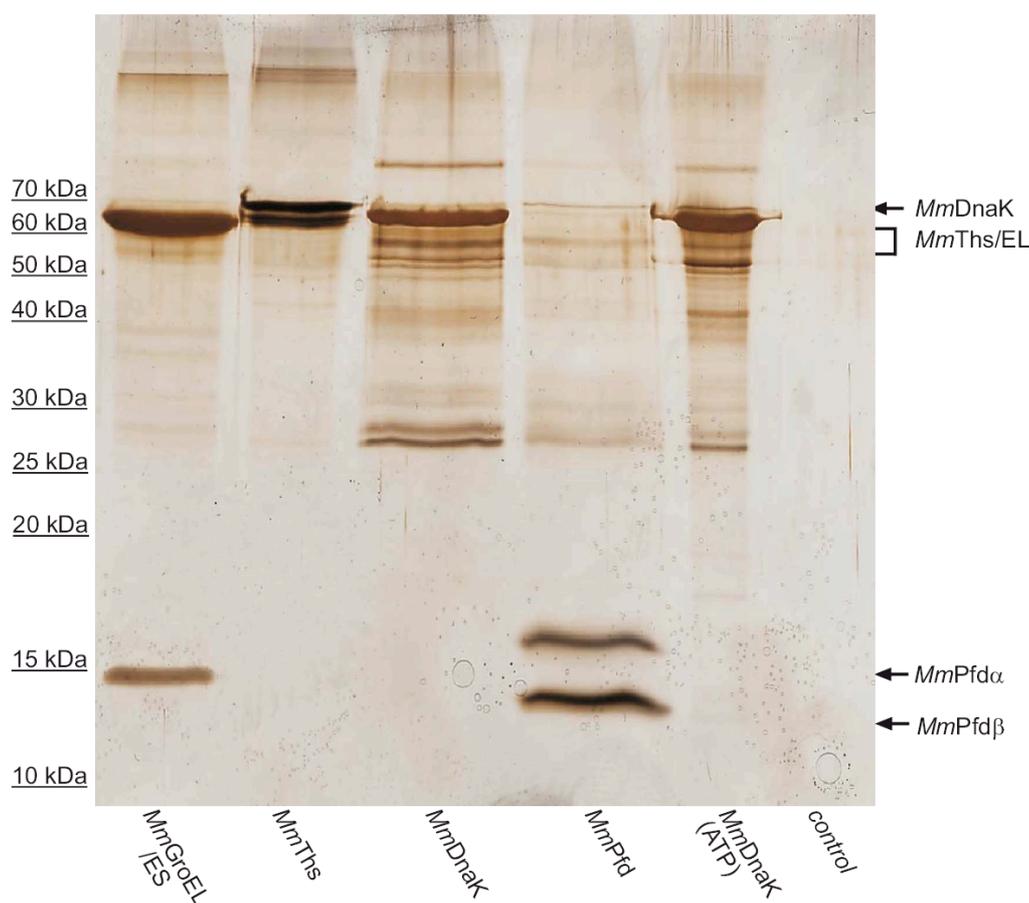
**Figure 46 : Similarities and Identities of identified proteins to homologs in *M. acetivorans* and *M. burtonii***

Aminoacid sequences of proteins identified in the lysate and in complex with chaperonins were compared to homologous sequences from other members of *Methanosarcinales*. *Methanococcoides burtonii* was chosen a closed relative holding no GroEL and *Methanosarcina acetivorans* as closed member possessing GroEL. Compared to lysate proteins, Ths interactors and overlapping substrates are GroEL specific substrates more distantly related to proteins from *M. burtonii* than to *M. acetivorans*.

#### IV.4 Analysis of the role of chaperonins in the network of *M. mazei*

Due to a high degree of horizontal gene transfer, *M. mazei* handles all major players in the folding of newly synthesized proteins. Thus *Methanosarcina* is not only the perfect model to study differences in the substrate specificity of group I and group II chaperonins, it also offers the possibility to investigate the interplay with other chaperones.

In order to analyse the connections within the chaperone network of *M. mazei* an substrate chaperone complexes of *MmPfd* and *MmDnaK* were isolated in addition to the *M. mazei* chaperonin substrate complexes (Figure 47).



**Figure 47 : Co-Immunoprecipitations from *M. mazei* lysate**

(A) Large scale co-IP of *M. mazei* chaperonins and interacting proteins (lane 1: *MmGroEL/GroES* and substrates, lane 2: *MmThs* and substrates, lane 4: *MmPfd* and substrates, lane 3, 5: *MmDnaK* and substrates, lane 6: control). The cells were lysed in the presence of Hexokinase/glucose and ADP to stabilize chaperone-substrate interactions. *MmDnaK* complexes were also precipitated from cells in presence of ATP (lane 5). The nucleotide had no influence on the substrate pattern of *MmDnaK*.

Analysis by SDS-PAGE showed a significant overlap of then protein pattern of *MmPfd* and *MmDnaK*. Similarities can be found between the *MmPfd* or *MmDnaK* co-precipitations and both *M. mazei* chaperonin substrate pattern. Identification of the substrate sets by mass spectrometry will reveal detailed insights about the interplay within the chaperone network of *M. mazei*. This model system will elucidate how chaperones are arranged for the folding of newly synthesized proteins in the cell.

## V. Discussion

In a living cell, the folding of many newly synthesized proteins is dependent on a special class of assisting proteins, the so-called chaperones. The specific set of chaperones, which is thought to be organized in a network-like manner, is one important hallmark of the three domains of life. Extensive work has been performed on the analysis of the bacterial chaperone network - especially the *E. coli* chaperones. Also the eukaryotic system with a main focus on *S. cerevisiae* has been investigated in detail. The archaeal folding machinery is thought to reflect a more ancient and therefore simpler version of the eukaryotic system. Structural analysis of archaeal chaperones, *i.e.* the Group II chaperonin thermosome, revealed details of these components of the archaeal as well as in the eukaryotic folding machinery.

The major components in the *de novo* folding system in the cell are represented by (i) the ribosome associated elements like trigger factor (TF) in *Bacteria* and probably the nascent chain associated complex (NAC) in *Archaea* and *Eukarya*, (ii) the downstream operating Hsp 70/40 system, which is found in *Bacteria* as well as in *Eukarya*; Hsp70 components identified in *Archaea* originated from horizontal gene transfer of the bacterial hsp70/40-proteins and (iii) the barrel-shaped chaperonins. The major difference in the chaperone content among the domains of life is the presence of the Group I chaperonin-system in *Bacteria* and the Group II chaperonins in *Archaea* and *Eukarya*.

### V.1 *The co-existence of Group I and Group II chaperonins*

It was extraordinarily surprising when the genes encoding for the group I chaperonin GroEL (*MmGroEL*) and its co-chaperone GroES (*MmGroES*) were identified in the genome sequence of *Methanosarcina mazei* - in addition to the thermosome that is usually found in *Archaea*. This frontier crossing was established in the genus *Methanosarcina* by horizontal gene transfer (HGT), probably before the formation of the different *Methanosarcina* species, because this operon is found in all members of this genus.

Biochemical analysis of *M. mazei* cells proved the actual co-existence of the complete group I (*MmGroEL/MmGroES*) and the group II chaperonin system (*MmThs*) at similar levels of ~1% of total cellular protein of both chaperonins under standard growth conditions.

Immunoblotting analysis of size fractionated lysate revealed that both proteins are present in their functional oligomeric state of ~ 800 -1000 kDa. Additionally, immunoprecipitations of the endogenous chaperonin complexes demonstrated that *MmGroEL* forms homo-oligomeric complexes and the endogenous thermosome complex of *M. mazei* is composed of three different subunits *MmThs* $\alpha$ , *MmThs* $\beta$  and *MmThs* $\gamma$ . The ratio of the thermosome subunits was assessed by quantitative immunoblotting analysis of immunoprecipitated thermosome complexes to be 2:1:1 of *MmThs* $\alpha$ : *MmThs* $\beta$ : *MmThs* $\gamma$  and a homogeneous population of this thermosome complex is expected. An independent study (L. Figueiredo, B. Haas) suggests that *Ths* complex formation is likely coordinated with transcriptional and/or translational levels in the cell, as spontaneous complex formation of the three subunits occurs to only a very limited extent *in vitro*.

It might have been expected that there is a specific response of each chaperonin to environmental stress, but surprisingly, thermal stress lead to an 2.5 fold increase in expression of both *MmGroEL/ES* as well as *MmThs*. No other stress conditions were tested, but based on the similarity of the promoter regions it is likely that both chaperonins contribute in a similar manner to protein folding in the cell.

Unfortunately, because genetic manipulations of *M. mazei* are not possible, it is only possible to speculate that both chaperonins are essential, because of their similar and high expression levels.

## ***V.2 Substrates of group I and group II chaperonins***

A similar, principal mechanism for the bacterial and the eukaryotic/archaeal chaperonins is proposed, but analysis of their respective substrate spectra will reveal detailed insights into differences between the bacterial and the eukaryotic/archaeal system.

So far most effort for investigation of natural substrates has been focused on GroEL/ES client proteins in *E.coli* (Houry et al.,1999; Kerner et al., 2005) and a virtually complete set of

substrates has been identified and qualitatively analyzed by their dependency on GroEL for successful folding. In the case of the group II chaperonins, there are several proteins known as natural substrates of the eukaryotic CCT/TRiC, this compilation is still far from completeness. In the case of the archaeal group II chaperonin, no natural substrates are known to date

Thus, the unique combination of both chaperonin systems in the methanosarcinal cell provides not only the possibility to gain a more complete picture of group II chaperonin substrates, but also the opportunity to analyze defining features of substrate specificity of those chaperonins. For several million years, substrate proteins had accessibility to both chaperonin systems and interactions to the optimal chaperonin may have evolved. Analysis of the substrate specificity in *M. mazei* may reveal fundamental differences in substrate selection that would otherwise be hidden in comparisons of substrates of group I and group II chaperonins from different organisms.

The study presented here provides extensive insights into the substrates specificity of group I and group II substrates.

### *V.2.1 Substrate identification techniques*

In the study, three different techniques were employed to demonstrate differences in the substrate specificity of Group I and Group II chaperonins. As expected the more sensitive technique, 1D- LCMS, resulted in a higher number of identified proteins. Thus, from the incongruity of the number of overlapping substrates in the *MmThs* substrate set (198) and the *MmGroEL* substrate set (153), the identification is not complete yet. This discrepancy is due to the validation of substrate proteins, which had to be identified at least twice in a sample. 68 substrates from the *MmThs* overlap were identified only once in a *MmGroEL* sample and are therefore not valid *MmGroEL* substrates. *Vice versa* 17 proteins from in the *MmGroEL* overlap were identified only once in the *MmThs* samples. Additional mass spectrometrical analysis of the samples would eliminate this problem.

Despite little information about the identity of the substrates from the 2D-PAGE techniques, these methods, especially the Ettan-DIGE, are more suitable for visualizing differences in the substrate spectra. Several proteins identified as substrates in 2D-PAGE approaches had to be

excluded as false positives, as they were detected in the background control sample by 1D-LCMS. These unspecifically precipitated proteins were apparently below the detection limit of the 2D techniques. Although some of these proteins might actually be substrate proteins, a specific interaction above background level could not be demonstrated.

The Ettan-DIGE technique represents a clear improvement of classical 2D-PAGE. Due to the possibility to directly compare samples the same gel, less false positive proteins were identified. In addition, the detection of proteins by fluorescence allows quantitative comparison of a given protein between the different samples.

There is no discrepancy of the results from the 2D-PAGE techniques and the 1D-LCMS.

### V.2.2 *Substrate sets of MmThs and MmGroEL*

Similar to *E. coli* (Ewalt et al., 1997) and mammalian cells (Thulasiraman et al., 1999), about 17% of the total cellular proteins were associated with chaperonins in *M. mazei*. These 454 proteins show different affinities for the two co-existing chaperonins. While 73 proteins were exclusively associated with *MmGroEL*, 117 proteins interact exclusively with the *MmThs*. About 263 of the identified proteins are found associated with both chaperonins, but again most of these show a significant preference for one of the two chaperonins.

Thus, in the methanosarcinal cell there is a balanced cooperation between the bacterially-derived and the archaeally-derived folding machinery, without a strong bias towards the “original” archaeal chaperonin, the thermosome. The similar impact of chaperonins on folding in the archaeal cell can be due to a principal commutability of both chaperonin systems to a high degree, as 45% of the substrate proteins show a promiscuous chaperonin dependency. On the other hand the methanosarcinal proteome may have adapted, to some extent, to the bacterial chaperonin over several million years. An extensive adaptation of the client proteins to the two folding machineries may also be reflected in the similar distribution of the proteins from archaeal and bacterial origin among the substrate sets of both chaperonins.

### V.2.3 Phylogenetic origin of the substrate proteins

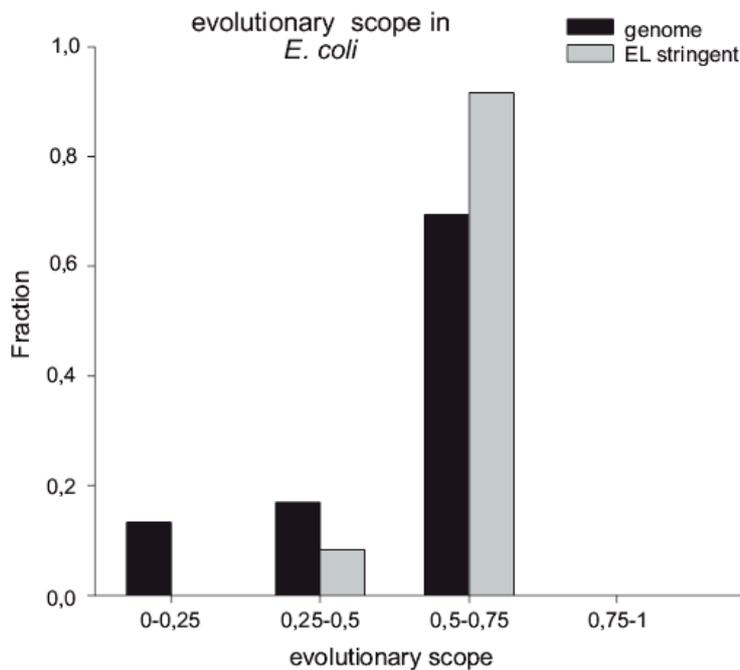
From the fact that *M. mazei* acquired 15-30% of its total genome by horizontal gene transfer (HGT), one might expect that especially these bacterially-derived proteins are the most likely candidates for *MmGroEL* substrates proteins. However, analysis of the distribution of bacterial proteins showed no increased fraction of these bacterial proteins among *MmGroEL* interactors compared to *MmThs* substrates. As mentioned above a co-evolution of the bacterial proteins in *M. mazei* with the co-existing archaeal chaperonin would erase a potential “bacterial protein- *MmGroEL*“-dependency and *vice versa*. This hypothesis is supported by the fact that the bacterial proteins, which show a strict dependency on *MmGroEL*, are highly conserved proteins, according to their evolutionary scope and a significant number of the bacterial proteins, which interact specifically only with the archaeal chaperonin, are fast developing proteins.

In addition, the dependence of bacterial proteins on the archaeal chaperonin (*MmThs*) might be explained by HGT prior to the acquisition of *MmGroEL*. In this case proteins that were depending on a chaperonin would have had to interact with the *MmThs* and might have adapted to this interaction before the transfer of GroEL.

Interestingly, the fraction of bacterial proteins among substrates interacting with both chaperonins is increased compared to the fraction of bacterial proteins in the *M. mazei* lysate. This might implement an important role of the chaperonins in maintenance of new functions derived from other organisms, which might then lead to an evolutionary advantage. While *MmGroEL* would be predestined to assist the folding of newly transferred bacterial components in the new archaeal environment; *MmThs* on the other hand may assist adaptation of bacterial proteins to the new conditions.

Assistance in the evolution of a protein can be especially expected from the group II chaperonin, because 20% of the specific *MmThs* substrates show an unusual low evolutionary scope (<0.25), which means that these proteins are evolving fast. And indeed, when these proteins were compared to known sequences using the global BLAST program, only similarities to fragments of mostly eukaryotic proteins could be identified outside the family Methanosarcina.

*MmGroEL* on the other hand apparently favours interaction with conserved proteins. This becomes even clearer among the 39 bacterial chaperonin substrate proteins out of which 22 proteins were found to interact with the Group I chaperonin. Bacterial proteins that interact specifically with the *MmGroEL* are exclusively highly conserved proteins. This role of a keeper of the *status quo* of a protein for *MmGroEL* might be a basic feature of the Group I chaperonins, because the class III substrates in the *EcGroEL* interactome show also a significantly higher conservation than the average lysate of *E. coli* (Figure 47).



**Figure 48 : Evolutionary scope of stringent GroEL substrates in *E. coli***

The evolutionary scope of stringent GroEL interacting substrates (class III) is increased compared to the evolutionary scope of proteins in the lysate

Despite the conservation of *MmGroEL* substrates, *MmGroEL* specific substrates exhibit significantly more frequent changes of the amino acid sequences compared to the homologues of GroEL lacking relatives of *Methanosarcinales*, such as *M. burtonii*. The presence of GroEL in *Methanosarcina* might have facilitated these sequence changes, but more likely the lack of GroEL in the relative *M. burtonii* might have caused the changes in the amino acid sequence at relevant positions to enable the cell to handle those proteins.

Most importantly, the substrate chaperonin interaction is probably driven mostly by physical features of a protein that might not be stringently reflected by the phylogenetic origin of the substrate protein.

#### *V.2.4 Structural features of substrate proteins*

The simultaneous analysis of the *MmGroEL*-substrate proteome along with the *MmThs*-substrate proteome allows not only deducing certain features of the distinct chaperonin substrate proteins to be deduced and, the natural co-existence of both chaperonins sheds light on the specific need for group I and group II chaperonins.

#### *Size dependence*

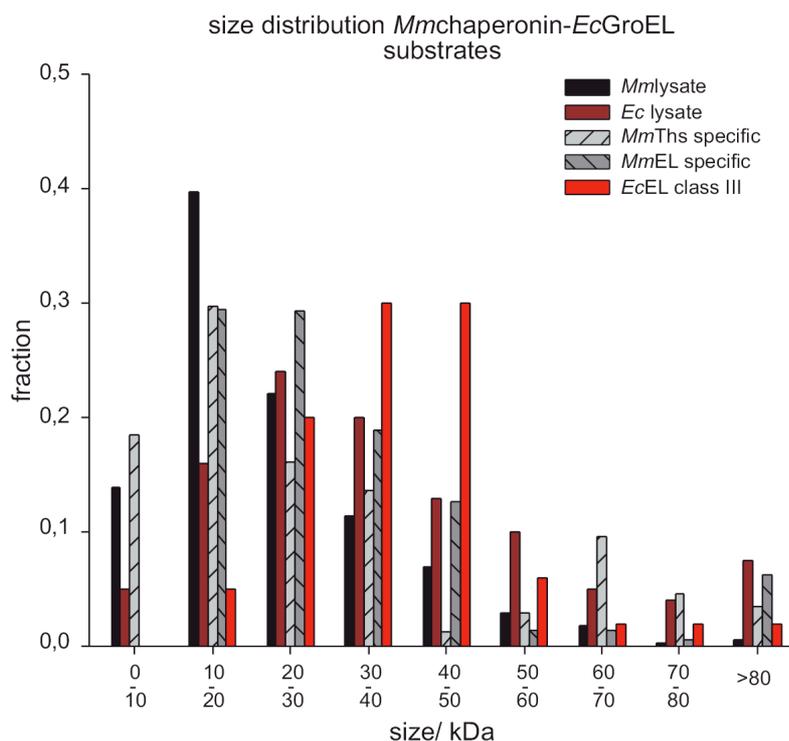
A principal feature of the folding mechanism of chaperonins is the encapsulation of the client protein. The folding chamber formed by the seven member of the GroEL ring under the GroES lid provides limited space to accommodate proteins up to a size ~ 60 kDa. It has been shown that proteins exceeding this size can be assisted by the *trans* cavity (Farr *et al.*, 2003), but because this mechanism is thought to be of minor impact for *in vivo* folding, large proteins are thought to interact with other chaperones like the Hsp 70/40 system. It is not clear whether the size limitation is also valid for group II chaperonins. It has even speculated that due to the closing mechanism by the helical protrusions, proteins could be encapsulated in part, which would allow folding of large proteins *via* encapsulation. This speculation is driven by the fact that the group II chaperonin assists folding in the eukaryotic system and thus, has to deal with proteins of larger size on average. On the other hand, it has been shown by Thulasiraman *et al.* that the majority of proteins found associated with the TRiC are not exceeding the theoretical cavity size ~60 kDa and larger proteins preferentially interact with the eukaryotic Hsp70/40 system (Hsc70). In addition, the majority of proteins that have been shown to interact with the eukaryotic TRiC are also found at a size range between 40kDa and 50 kDa (Dunn *et al.*, 2001) .

Most of the substrates of *MmThs* and *MmGroEL* are found at sizes between 10 kDa and 40 kDa, clearly below the theoretical excluding volume of 60 kDa of the chaperonin cavity. Compared to the lysate, the average size of substrate proteins is larger. Surprisingly a substantial number of proteins smaller than 10 kDa are found among the *MmThs* specific

---

substrates, proteins of such low sizes would be expected to fold spontaneously. Unexpectedly, a significant number of *MmThs* specific substrates have masses between 60 kDa and 80 kDa, which exceeds the theoretical size limit of the cage. It has been shown previously that the eukaryotic group II chaperonin TRiC is able to assist folding of proteins in this size range (Frydman *et al.*, 1992). These proteins would also be too large to fit the folding cavity, suggesting the same dimensions of the group II and the group I chaperonins. This up-shift of the size limit for substrates to 80 kDa raises the question of whether the specific closing mechanism might rather contribute to an enlargement of the cage volume, than to an “infinite” capacity by the domain-wise folding theory. Proposing a closing mechanism, which is similar to an optical aperture, the volume of the chamber volume might be regulated to some extent. However, the flexibility to encapsulate proteins only in part can not be confirmed from this study. Very large proteins (>100 kDa) and also multi-domain proteins show no preferential interaction with *MmThs*, which would be expected if a partial encapsulation was possible.

Similarly to the situation among class III *EcGroEL* substrates, the size distribution pattern of *MmGroEL* interactors differs significantly from the size distribution of proteins in the lysate and shows a broad maximum ~10-20 kDa above the sharp maximum peak maximum of the size distribution of lysate proteins. In contrast, the size distribution pattern of *MmThs* reflects in principle the pattern of the size distribution in the lysate.



**Figure 49 : Size distribution of Chaperonin substrates and EcGroEL class III substrates**

Proteins between 30 kDa and 50 kDa are about two fold more abundant on the group I chaperonins compared to lysate. Group II interacting proteins show deviation at the range between 60kDa-80 kDa from the average size of lysate proteins.

### *Hydrophobicity of substrate proteins*

In contrast to other chaperones like Hsp 70, which is thought to recognize sequence motifs of client proteins, the interaction of chaperonins is probably driven by more global structural features of the substrate. The group I chaperonins are expected to bind hydrophobic surfaces via hydrophobic residues in the apical domain (Fenton and Horwich, 2003). Though, the hydrophobicity is not significantly stronger among *Mm*GroEL substrates compared to *Mm*Ths substrates and lysate proteins. *Mm*GroEL specific substrates however, show a significant shift to higher hydrophobicity, which is not the case for group II specific substrates. This is also supported by an overall pI up-shift of 0.5 units of *Mm*GroEL substrates towards a neutral pH and an overall net charge distribution around 0. Also the binding of *Mm*GroEL to some proteins harbouring trans-membrane domains might be driven by these hydrophobic interactions. GroEL is not expected to play a major role in the folding of trans-membrane proteins, but a principle

---

possibility of GroEL to interact with transmembrane proteins was shown *in vitro* (Sun et al., 2005), however, the role of GroEL in maturation of membrane proteins *in vivo* requires further investigation.

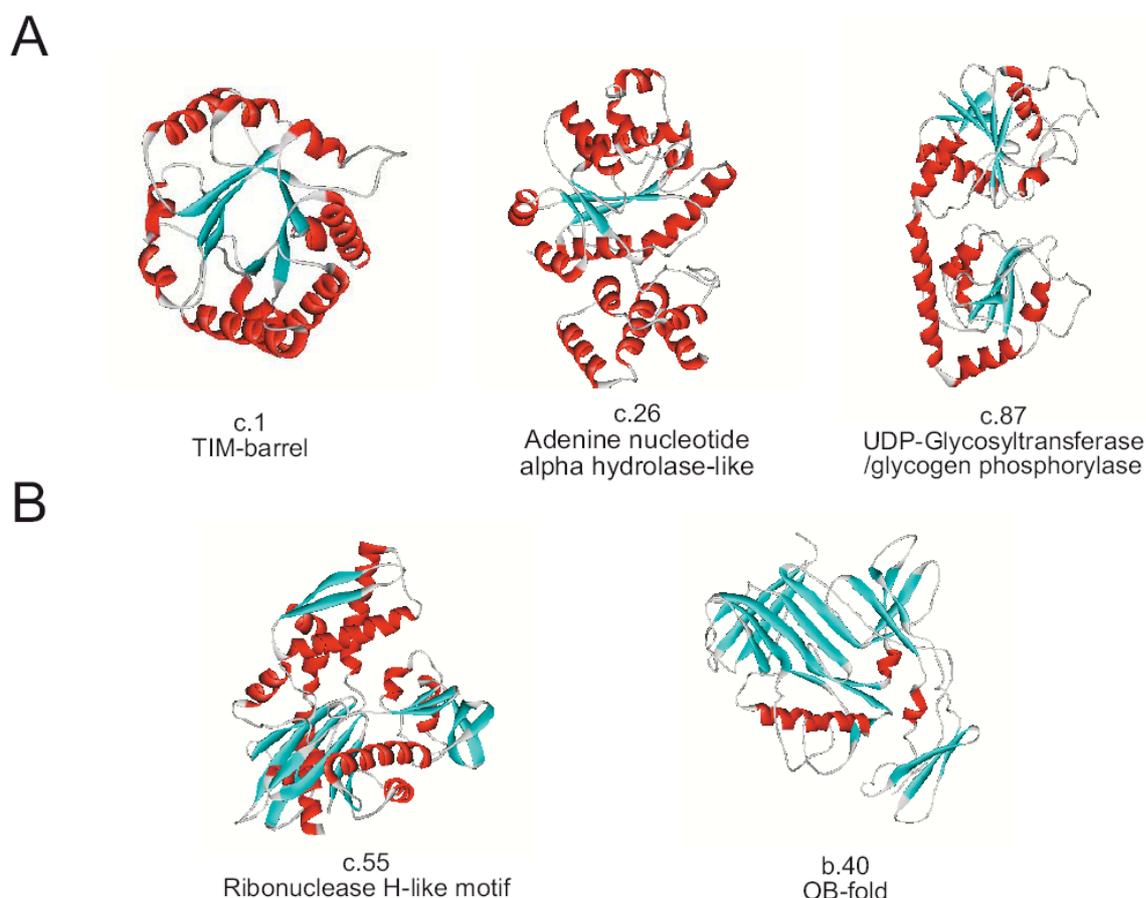
### *Net charge of substrate proteins*

Unlike GroEL positively charged amino acids are frequently found at relevant positions of the substrate binding domain of the group II TRiC (Pappenberger *et al.*, 2002). These positively charged amino acids are thought to contribute to the interaction of group II chaperonins and substrates. Analysis of both substrate sets identified no preference of both chaperonins for charged proteins. Analysis of the specific interacting proteins though, revealed a high frequency of negative charges among only the *MmThs* specific substrates. The overrepresentation of negatively charged amino acid side chains among *MmThs* specific interactors might explain their preferential binding to the group II chaperonin.

### *Structural determinants for chaperonin interaction*

#### Domain fold

It has been shown that the domain fold play a major role in substrate selection by GroEL in *E. coli* and also the analysis of the domain fold distribution according to the SCOP database revealed a clear impact of the domain fold on the substrate spectra of the group I and the group II chaperonins (Figure 48).



**Figure 50 : Representative structure images of domain folds of chaperonin substrates**

(A) A significant number of substrate proteins of the group I chaperonin exhibit c-class domain folds (mainly  $\alpha/\beta$  folds). The most abundant are the TIM  $\beta/\alpha$  barrel fold (c.1; parallel beta-sheet barrel), Adenine nucleotide alpha hydrolase-like (c.26; core : 3 layers, a/b/a; parallel beta sheet) and UDP-Glycosyltransferase/glycogen phosphorylase fold (c.87; 3 layers, a/b/a; parallel beta sheet of 6 strands). (B) Proteins with the ribonuclease H- like motif (c.55; a/b/a; mixed beta-sheet) and the OB-fold domain of mainly beta sheets (class b) show a strong dependency on the group II chaperonin for folding assistance.

In *M. mazei* several members of the c-class fold, mainly  $\alpha/\beta$  folds, showed a significantly increased interaction with *MmGroEL*. Most prevalent among these *MmGroEL* dependent c-class folds (mainly  $\alpha/\beta$  folds) was the TIM  $\beta/\alpha$  barrel (c.1) fold- consistent with the study on the *EcGroEL* interaction proteome. Some representatives of this fold that were identified in the lysate were not found to interact with a chaperonin. This makes it unlikely that the TIM  $\beta/\alpha$  barrel fold is a sufficient criterion to determine interaction of a protein with *MmGroEL*. The c.1 fold is classified into a several superfamilies, and the most prominent superfamily found among *MmGroEL* interactors is the ribulose-phosphate binding barrel (c.1.2) superfamily (6.0% of

*MmGroEL*-, 2.4% of *MmThs* substrates, 1% in lysate); only one protein sharing the c.1.2. superfamily domain fold was identified in the lysate but not found among the *MmGroEL* interactors.

All members of the c-class fold adenine nucleotide alpha hydrolase-like (c.26) that were found in the lysate were also identified on *MmGroEL*, some were also found to interact with the group II chaperonin. Substrates sharing the c.26 domain fold are mainly represented by two superfamilies: the nucleotidyl transferase (c.26.1) superfamily and the adenine nucleotide alpha hydrolases-like (c.26.2) superfamily. No correlation of the c.26 fold or its superfamilies with any other structural features (i.e. net charge, hydrophobicity, molecular weight) could be detected that could determine interaction with *MmGroEL*. Depending on the individual protein, overlapping substrates sharing the c.26 domain fold show preference for either of the two chaperonins.

The third c-class domain fold, which is frequently found with *MmGroEL*, is categorized as the UDP-glycosyltransferase/glycogen phosphorylase fold. This fold is represented by only one superfamily in *M. mazei*: the UDP-Glycosyltransferase/glycogen phosphorylase (c.87.1) superfamily. Proteins sharing this domain fold might be of a low abundance in the lysate, because they could only be detected among *MmGroEL*/substrate complexes and not in the lysate. Two proteins with this fold domain were also identified on *MmThs*. These proteins (gi|0905621 and gi|20906666) were found at lower abundance on *MmThs* compared to *MmGroEL*. Interestingly these proteins are negatively charged (-2.8168 and -0.5864) and have a low hydrophobicity index (0.4085 and 0.39) compared to strict *MmGroEL* interactors (0.43 at average), such features are typically found among *MmThs* (specific) substrates.

The most prominent domain fold found to interact with *MmThs* exhibits the Ribonuclease H-like motif (c.55). With the exception of two proteins, all members of this fold class that were identified in the lysate interact with *MmThs*: (i) the archaeal "conserved protein" (gi|20905336) was detected only in one of the *MmThs* substrate samples and was therefore not categorized as *MmThs* substrate and (ii) the DNA mismatch repair protein mutS (gi|20906191), which is a highly conserved protein (evolutionary scope of 0.77) and interacts specifically with *MmGroEL*. Interestingly, all but one of the c.55 *MmThs* substrates are negatively charged, which should promote their binding to *MmThs*. The most prominent superfamily found is the Actin-like

ATPase domain (c.55.1) superfamily (9 of the 13 substrates) and these proteins exclusively interact with the group II chaperonin, consistent with actin being an obligate substrate of the eukaryotic group II chaperonin TRiC. Surprisingly, 10 of the *MmThs* interacting proteins sharing the c.55 fold are of bacterial origin, but do not show a high conservation with evolutionary scope values below 0.75 - typically for bacterial proteins that interact with the archaeal chaperonin. Additionally 4 of these 6 bacterial proteins exhibit a size between 62 kDa and 74 kDa that is not favoured by the *MmGroEL* interactors, but more frequently found among *MmThs* (specific) substrates.

*MmThs* substrates also include members of the b-class domain folds, such as the OB-fold (b.40). Most proteins of *M. mazei* exhibiting the b.40 domain fold are found among the superfamily of nucleic acid-binding proteins (b.40.4) and interact strictly with *MmThs*, with the exception of the Lysyl-tRNA synthetase2 (gi|20906448). This protein is highly conserved (evolutionary scope: 0.77) and the only OB fold protein in *M. mazei* of bacterial origin. Only one protein with the b.40 domain fold that was identified in the lysate, the inorganic pyrophosphatase (gi|20905923), which shows no association with any chaperonin. This has a fold of the superfamily inorganic pyrophosphatases (b.40.5). Interestingly, proteins sharing the OB fold are strongly charged, but in contrast to the finding above, that suggests a preferential interaction of *MmThs* with negatively charged proteins, 4 of the 10 proteins are positively charged.

#### Unknown folds

The structural analysis is based on a limited number of the substrate proteins, as structural data about 30% of the total soluble proteome is still missing. Interestingly comparison of the abundance of these unknown folds between the lysate fraction and the substrate sets revealed a strong bias to unknown folds among *MmThs* specific substrates. Proteins of unknown folds are found twice as often in the *MmThs* substrate set as in the lysate or in the *MmGroEL* substrate set. This imbalance might have established because of technical problems in protein preparation for crystallisation. For purification, proteins are routinely over-expressed in bacterial cells and thus the overproduction of protein with a strong dependence on group II chaperonins would be doomed to fail in the bacterial host. This problem is less pronounced for *MmGroEL* dependent proteins, as GroEL/ES is available at least to a limited extent or it might even be co-

---

overexpressed. In addition, the large scale crystallisation attempts using both *E. coli* as well as *S. cerevisiae* for high through-put purification of proteins might be again limited by the availability of the group II chaperonin in the yeast system (0.3  $\mu$ M).

This limited accessibility for proteins that depend on the assistance of a group II chaperonin by heterologous expression might also be a major cause for the substantial fraction of proteins among the *MmThs* substrates that can not be assigned to any known function. The inability to produce these proteins efficiently in the bacterial host is a major problem in studying the function of these proteins. Apart from this technically caused correlation, the function of a protein seems to have no direct influence on its chaperonin requirement.

### *Essentiality of the chaperonin substrates*

Chaperonins are the only chaperones that are known to be essential for the cell under all conditions. But because of the lack of tools to genetically manipulate *M. mazei*, it can only be speculated whether both chaperonins are essential in the archaeal cell. The similar expression level of both chaperonins under stress and non-stress conditions in the cell along with the significant number of potentially essential proteins that interact specifically with only one of the chaperonins suggests that both chaperonins are indispensable for viability of archaeal cell.

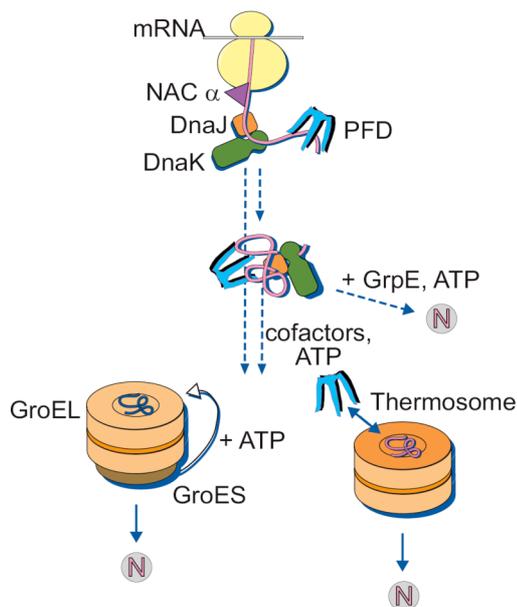
### *Importance of the study in M. mazei*

Identification of the substrates of the thermosome from *M. mazei* revealed for the first time a comprehensive picture of the substrate selectivity of group II chaperonins. But even more importantly, many differences in the substrate specificity of Group I and Group II chaperonins, based on features such as hydrophobicity or net charge of the substrate, became apparent. Our results emphasise the importance of performing such a study in an organism, where both chaperonin groups co-exist. Comparison of isolated substrate sets from different organisms, *i.e.* the interaction proteome of *EcGroEL* and a group II chaperonin interactome - if there will be one available in the future- probably would not provide insights into the differences in substrate specificity of the two chaperonins systems.

The study revealed that the affinity of proteins for either chaperonin is influenced by a combination of structural and physical properties, such as the domain fold, net charge, hydrophobicity and size. In addition, the chaperonin interaction is influenced by the evolutionary state of the substrate and, to some extent, by the phylogenetic origin of a protein - or rather the chaperonin interaction might influence the evolutionary state of the substrate protein. As an example, the protein MutS (gi|20906191) shares the c.55 domain fold with many *MmThs* substrates and might therefore be expected to bind preferentially to *MmThs*. The net charge of -2.11 and a relative hydrophobicity of 4.11 should not influence the chaperonin selection. However, the combination of high conservation (evolutionary scope >0.75) and bacterial origin makes this protein a likely candidate for interaction with the Group I chaperonin. The size of 100 kDa should enforce this tendency. In agreement with these predictions, MutS is one of only two proteins with a c.55 fold domain that interacts with *MmGroEL*.

For future studies: the chaperone system of *M. mazei* provides all major components of *de novo* protein folding of the three domains of life, and this makes *M. mazei* an attractive model for investigation of possible chaperone networks (Figure 51).

## *Methanosarcina*



**Figure 51 : Chaperonin system of *Methanosarcina***

Main components of the machinery for folding of newly synthesized proteins. The monomeric Prefoldin and DnaK (together with its co-chaperone DnaJ and the nucleotide exchange factor GrpE) welcome the newly synthesized polypeptide and assist for folding or hand the non native protein over to the further downstream acting chaperonins GroEL/ES or Ths

Interestingly, initial pull down experiments of the *MmGimC* and *MmDnaK* substrate complexes along with the chaperonin substrate complexes also revealed an interaction of *MmDnaK* and *MmGimC*. The Identification of the client proteins of *MmDnaK* and *MmGimC* will likely elucidate the basis of the interaction of these chaperones.

In addition, comparison of substrate set of all *M. mazei* chaperones will elucidate the flux of proteins though the chaperone network.

## VI. References

- Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Fold Des* 1, R1-8.
- Baldwin, R. L., and Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci* 24, 77-83.
- Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* 93, 9188-9193.
- Bochkareva, E. S., Solovieva, M. E., and Girshovich, A. S. (1998). Targeting of GroEL to SecA on the cytoplasmic membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A* 95, 478-483.
- Boisvert, D. C., Wang, J., Otwinowski, Z., Horwich, A. L., and Sigler, P. B. (1996). The 2.4 Å crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat Struct Biol* 3, 170-177.
- Bosch, G., Baumeister, W., and Essen, L. O. (2000). Crystal structure of the beta-apical domain of the thermosome reveals structural plasticity in the protrusion region. *J Mol Biol* 301, 19-25.
- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371, 578-586.
- Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G., and Dobson, C. M. (2002). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci U S A* 99 *Suppl* 4, 16419-16426.
- Craig, E. A., Huang, P., Aron, R., and Andrew, A. (2006). The diverse roles of J-proteins, the obligate Hsp70 co-chaperone. *Rev Physiol Biochem Pharmacol* 156, 1-21.
- De Rosa, M., and Gambacorta, A. (1988). The lipids of archaebacteria. *Prog Lipid Res* 27, 153-175.

Deppenmeier, U., Blaut, M., Jussofie, A., and Gottschalk, G. (1988). A methyl-CoM methylreductase system from methanogenic bacterium strain Go 1 not requiring ATP for activity. *FEBS Lett* 241, 60-64.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., *et al.* (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* 4, 453-461.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* 400, 693-696.

Ditzel, L., Lowe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R., and Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* 93, 125-138.

Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* 24, 329-332.

Dobson, C. M., Sali A., Karplus M. (1998). Proteinfaltung aus theoretischer und experimenteller Sicht. *Angewandte Chemie* 110, 908-935.

Dunn, A. Y., Melville, M. W., and Frydman, J. (2001). Review: cellular substrates of the eukaryotic chaperonin TRiC/CCT. *J Struct Biol* 135, 176-184.

Eichler, J., and Duong, F. (2004). Break on through to the other side--the Sec translocon. *Trends Biochem Sci* 29, 221-223.

Ellis, R. J., and Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Curr Opin Struct Biol* 9, 102-110.

Ewalt, K. L., Hendrick, J. P., Houry, W. A., and Hartl, F. U. (1997). In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* 90, 491-500.

Farr, G. W., Fenton, W. A., Chaudhuri, T. K., Clare, D. K., Saibil, H. R., and Horwich, A. L. (2003). Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes. *Embo J* 22, 3220-3230.

Fenton, W. A., and Horwich, A. L. (2003). Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q Rev Biophys* 36, 229-256.

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

Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P., and Hartl, F. U. (1992). Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *Embo J* 11, 4767-4778.

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

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., *et al.* (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* 12, 532-542.

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P. S., Hartl, F. U., and Georgopoulos, C. (2004). In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep* 5, 195-200.

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

Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T., and Bukau, B. (1999). Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc Natl Acad Sci U S A* 96, 13732-13737.

Gomez-Puertas, P., Martin-Benito, J., Carrascosa, J. L., Willison, K. R., and Valpuesta, J. M. (2004). The substrate recognition mechanisms in chaperonins. *J Mol Recognit* 17, 85-94.

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313, 903-919.

Goulhen, F., De, E., Pages, J. M., and Bolla, J. M. (2004). Functional refolding of the *Campylobacter jejuni* MOMP (major outer membrane protein) porin by GroEL from the same species. *Biochem J* 378, 851-856.

Gozu, M., Hoshino, M., Higurashi, T., Kato, H., and Goto, Y. (2002). The interaction of beta(2)-glycoprotein I domain V with chaperonin GroEL: the similarity with the domain V and membrane interaction. *Protein Sci* 11, 2792-2803.

---

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

Hanninen, A. L., Bamford, D. H., and Bamford, J. K. (1997). Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. *Virology* 227, 207-210.

Hartl, F. U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852-1858.

Hesterkamp, T., Hauser, S., Lutcke, H., and Bukau, B. (1996). Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proc Natl Acad Sci U S A* 93, 4437-4441.

Hohfeld, J., Cyr, D. M., and Patterson, C. (2001). From the cradle to the grave: molecular chaperones that may choose between folding and degradation. *EMBO Rep* 2, 885-890.

Hohn, M. J., Hedlund, B. P., and Huber, H. (2002). Detection of 16S rDNA sequences representing the novel phylum "Nanoarchaeota": indication for a wide distribution in high temperature biotopes. *Syst Appl Microbiol* 25, 551-554.

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

Hungate, R. E., Smith, W., and Clarke, R. T. (1966). Suitability of butyl rubber stoppers for closing anaerobic roll culture tubes. *J Bacteriol* 91, 908-909.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4, 1265-1272.

Kandler, O., and König, H. (1998). Cell wall polymers in Archaea (Archaeobacteria). *Cell Mol Life Sci* 54, 305-308.

Kapatai, G., Large, A., Benesch, J. L., Robinson, C. V., Carrascosa, J. L., Valpuesta, J. M., Gowrinathan, P., and Lund, P. A. (2006). All three chaperonin genes in the archaeon *Haloferax volcanii* are individually dispensable. *Mol Microbiol* 61, 1583-1597.

Karlin, S., Mrazek, J., Ma, J., and Brocchieri, L. (2005). Predicted highly expressed genes in archaeal genomes. *Proc Natl Acad Sci U S A* 102, 7303-7308.

---

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F. U. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122, 209-220.

Klumpp, M., Baumeister, W., and Essen, L. O. (1997). Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* 91, 263-270.

Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D. J., Pfeifer, G., Muller, V., Deppenmeier, U., Gottschalk, G., Hartl, F. U., and Hayer-Hartl, M. (2003). Coexistence of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. *J Biol Chem* 278, 33256-33267.

Kluyver, and Niel, v. (1936). *Methanosarcina*. *The Prokaryotes*, 2199-2205.

Kramer, G., Patzelt, H., Rauch, T., Kurz, T. A., Vorderwulbecke, S., Bukau, B., and Deuerling, E. (2004). Trigger factor peptidyl-prolyl cis/trans isomerase activity is not essential for the folding of cytosolic proteins in *Escherichia coli*. *J Biol Chem* 279, 14165-14170.

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

Kusmierczyk, A. R., and Martin, J. (2001). Chaperonins--keeping a lid on folding proteins. *FEBS Lett* 505, 343-347.

Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G., and Mann, M. (2002). Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419, 537-542.

Leroux, M. R. (2001). Protein folding and molecular chaperones in archaea. *Adv Appl Microbiol* 50, 219-277.

Levinthal, C. (1969). How to fold graciously. Paper presented at: Mossbauer spectroscopy in biological systems (eds. P. DeBrunner et al.).

Llorca, O., Martin-Benito, J., Gomez-Puertas, P., Ritco-Vonsovici, M., Willison, K. R., Carrascosa, J. L., and Valpuesta, J. M. (2001). Analysis of the interaction between the eukaryotic chaperonin CCT and its substrates actin and tubulin. *J Struct Biol* 135, 205-218.

Llorca, O., Martin-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G. M., Willison, K. R., Carrascosa, J. L., and Valpuesta, J. M. (2000). Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *Embo J* 19, 5971-5979.

Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res* 30, 264-267.

Lyimo, T. J., Pol, A., Op den Camp, H. J., Harhangi, H. R., and Vogels, G. D. (2000). *Methanosarcina semesiae* sp. nov., a dimethylsulfide-utilizing methanogen from mangrove sediment. *Int J Syst Evol Microbiol* 50 Pt 1, 171-178.

Macario, A. J., and de Macario, E. C. (1999). The archaeal molecular chaperone machine: peculiarities and paradoxes. *Genetics* 152, 1277-1283.

Macario, A. J., Dugan, C. B., and Conway de Macario, E. (1991). A dnaK homolog in the archaeobacterium *Methanosarcina mazei* S6. *Gene* 108, 133-137.

Maestrojuan, G. M., Boone, J. E., Mah, R. A., Menaia, J. A., Sachs, M. s., and Boone, D. R. (1992). Taxonomy and Halotolerance of Mesophilic Strains, Assignment of Strains to Species, and Synonymy of *Methanosarcina mazei* and *Methanosarcina frisia*. *International Journal of Systematic Bacteriology* 42, 561-567.

Markowitz, V. M., Korzeniewski, F., Palaniappan, K., Szeto, E., Werner, G., Padki, A., Zhao, X., Dubchak, I., Hugenholtz, P., Anderson, I., *et al.* (2006). The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34, D344-348.

Martin, W., and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37-41.

Mayhew, M., da Silva, A. C., Martin, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1996). Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* 379, 420-426.

Minton, A. P. (1983). The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem* 55, 119-140.

Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388, 343-349.

---

Nitsch, M., Walz, J., Typke, D., Klumpp, M., Essen, L. O., and Baumeister, W. (1998). Group II chaperonin in an open conformation examined by electron tomography. *Nat Struct Biol* 5, 855-857.

Olsen, J. V., and Mann, M. (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* 101, 13417-13422.

Pappenberger, G., Wilsher, J. A., Roe, S. M., Counsell, D. J., Willison, K. R., and Pearl, L. H. (2002). Crystal structure of the CCTgamma apical domain: implications for substrate binding to the eukaryotic cytosolic chaperonin. *J Mol Biol* 318, 1367-1379.

Privalov, P. L. (1996). Intermediate states in protein folding. *J Mol Biol* 258, 707-725.

Ranson, N. A., Farr, G. W., Roseman, A. M., Gowen, B., Fenton, W. A., Horwich, A. L., and Saibil, H. R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107, 869-879.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75, 663-670.

Robertson, C. E., Harris, J. K., Spear, J. R., and Pace, N. R. (2005). Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* 8, 638-642.

Roseman, A. M., Chen, S., White, H., Braig, K., and Saibil, H. R. (1996). The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* 87, 241-251.

Saibil, H., Dong, Z., Wood, S., and auf der Mauer, A. (1991). Binding of chaperonins. *Nature* 353, 25-26.

Sakikawa, C., Taguchi, H., Makino, Y., and Yoshida, M. (1999). On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J Biol Chem* 274, 21251-21256.

Schechter, A. N., Chen, R. F., and Anfinsen, C. B. (1970). Kinetics of folding of staphylococcal nuclease. *Science* 167, 886-887.

Shevchenko, V. A., Akayeva, E. A., Yeliseyeva, I. M., Yelisova, T. V., Yofa, E. L., Nilova, I. N., Syomov, A. B., and Burkart, W. (1996). Human cytogenetic consequences of the Chernobyl accident. *Mutat Res* 361, 29-34.

Schultz, C. P. (2000). Illuminating folding intermediates. *Nat Struct Biol* 7, 7-10.

Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U., and Moarefi, I. (2000). Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* 103, 621-632.

Simankova, M. V., Parshina, S. N., Tourova, T. P., Kolganova, T. V., Zehnder, A. J., and Nozhevnikova, A. N. (2001). *Methanosarcina lacustris* sp. nov., a new psychrotolerant methanogenic archaeon from anoxic lake sediments. *Syst Appl Microbiol* 24, 362-367.

Spies, C., Meyer, A. S., Reissmann, S., and Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol* 14, 598-604.

Spreter, T., Pech, M., and Beatrix, B. (2005). The crystal structure of archaeal nascent polypeptide-associated complex (NAC) reveals a unique fold and the presence of a ubiquitin-associated domain. *J Biol Chem* 280, 15849-15854.

Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296, 1459-1462.

Sun, J., Savva, C. G., Deaton, J., Kaback, H. R., Svrakic, M., Young, R., and Holzenburg, A. (2005). Asymmetric binding of membrane proteins to GroEL. *Arch Biochem Biophys* 434, 352-357.

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* 278, 631-637.

Teter, S. A., Houry, W. A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F. U. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* 97, 755-765.

Thulasiraman, V., Yang, C. F., and Frydman, J. (1999). In vivo newly translated polypeptides are sequestered in a protected folding environment. *Embo J* 18, 85-95.

Trent, J. D., Kagawa, H. K., Paavola, C. D., McMillan, R. A., Howard, J., Jahnke, L., Lavin, C., Embaye, T., and Henze, C. E. (2003). Intracellular localization of a group II chaperonin indicates a membrane-related function. *Proc Natl Acad Sci U S A* 100, 15589-15594.

---

von Klein, D., Arab, H., Volker, H., and Thomm, M. (2002). *Methanosarcina baltica*, sp. nov., a novel methanogen isolated from the Gotland Deep of the Baltic Sea. *Extremophiles* 6, 103-110.

Wegele, H., Muller, L., and Buchner, J. (2004). Hsp70 and Hsp90--a relay team for protein folding. *Rev Physiol Biochem Pharmacol* 151, 1-44.

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

Wiedmann, B., Sakai, H., Davis, T. A., and Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. *Nature* 370, 434-440.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74, 5088-5090.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87, 4576-4579.

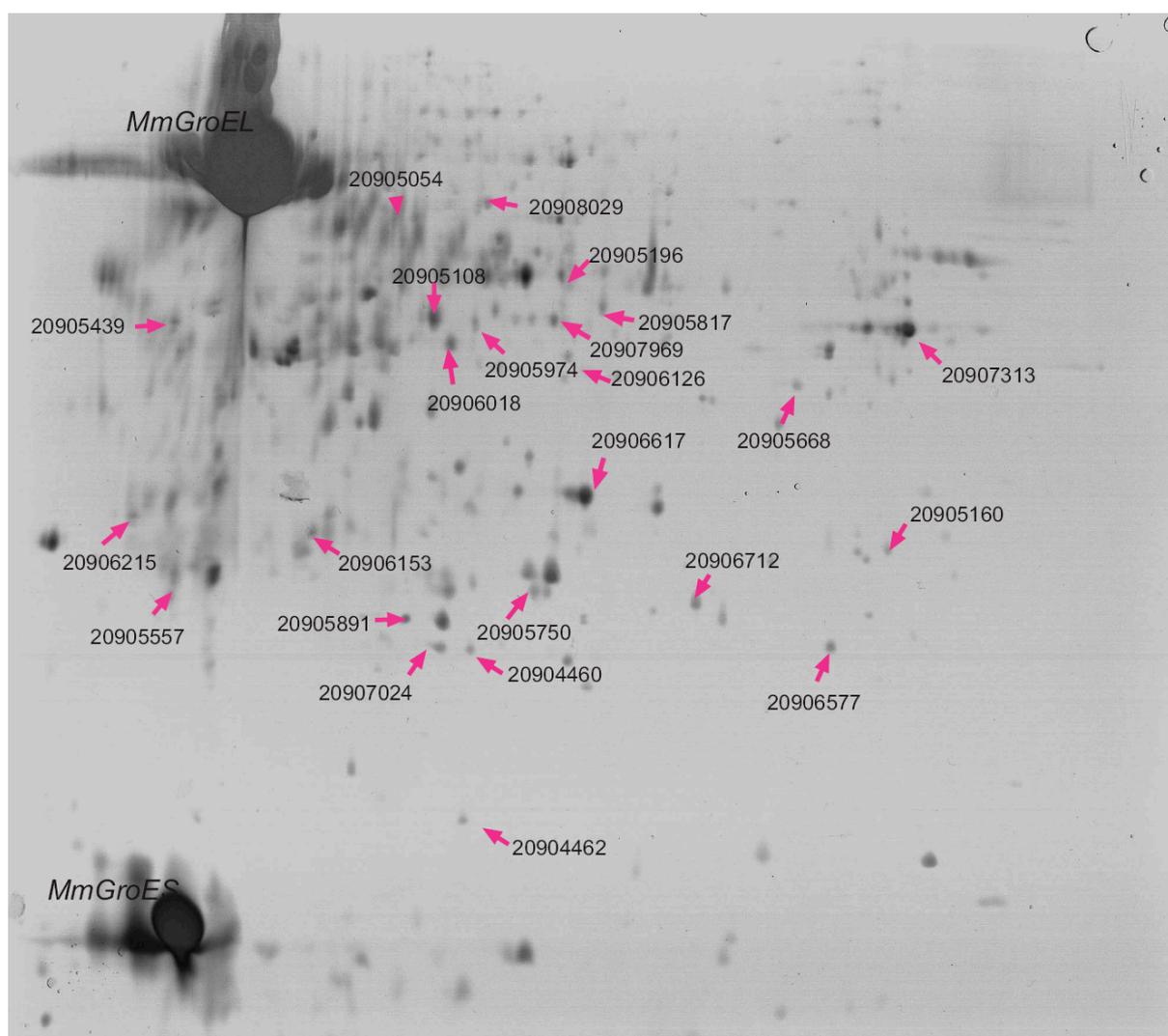
Xu, Z., Horwich, A. L., and Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)<sub>7</sub> chaperonin complex. *Nature* 388, 741-750.

Zimmerman, S. B., and Minton, A. P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22, 27-65.

## VII. Appenices

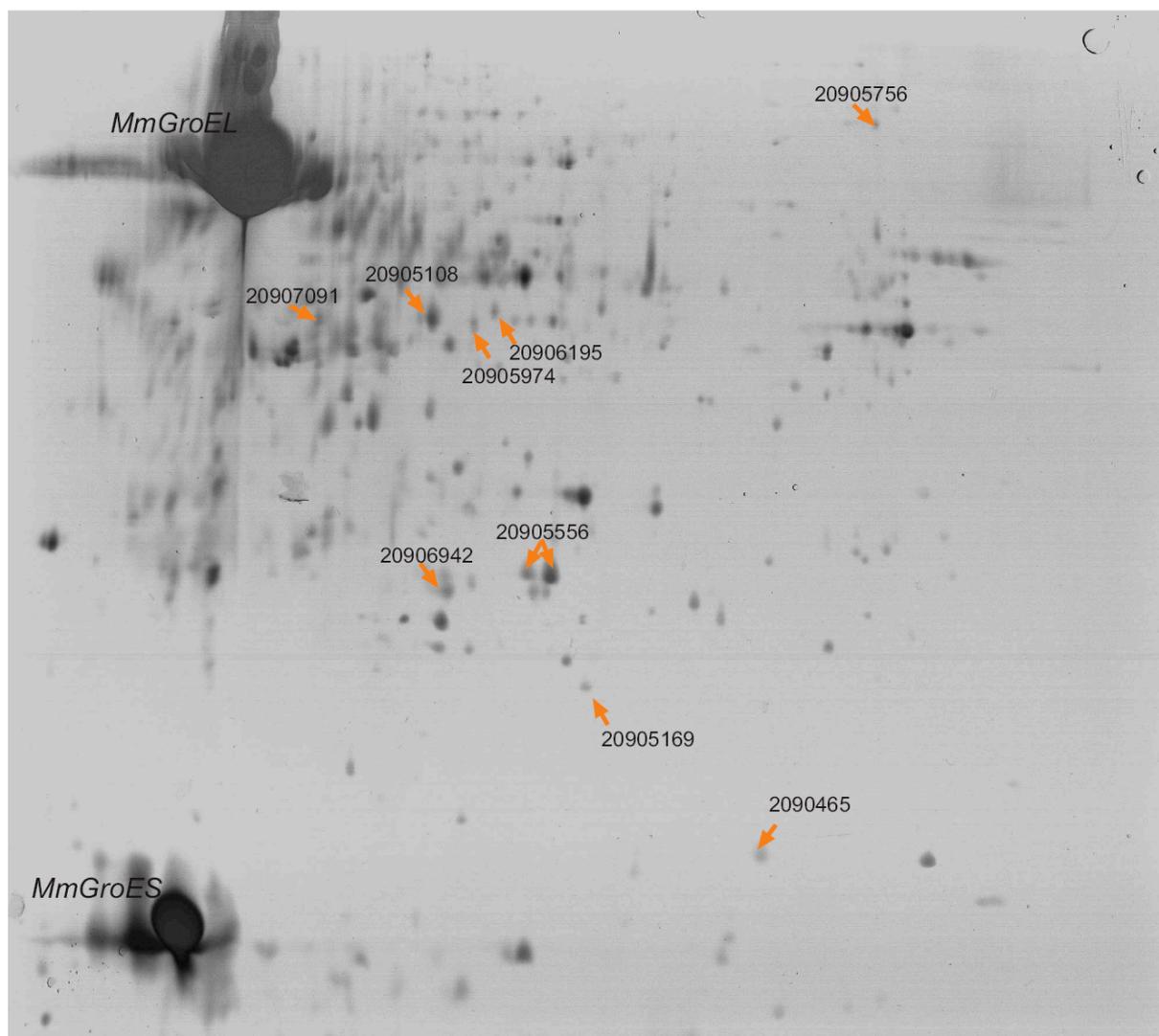
### VII.1 Supplementary Figures

**Figure S1: GroEL- overlapping substrates, 2D-PAGE, silverstained**



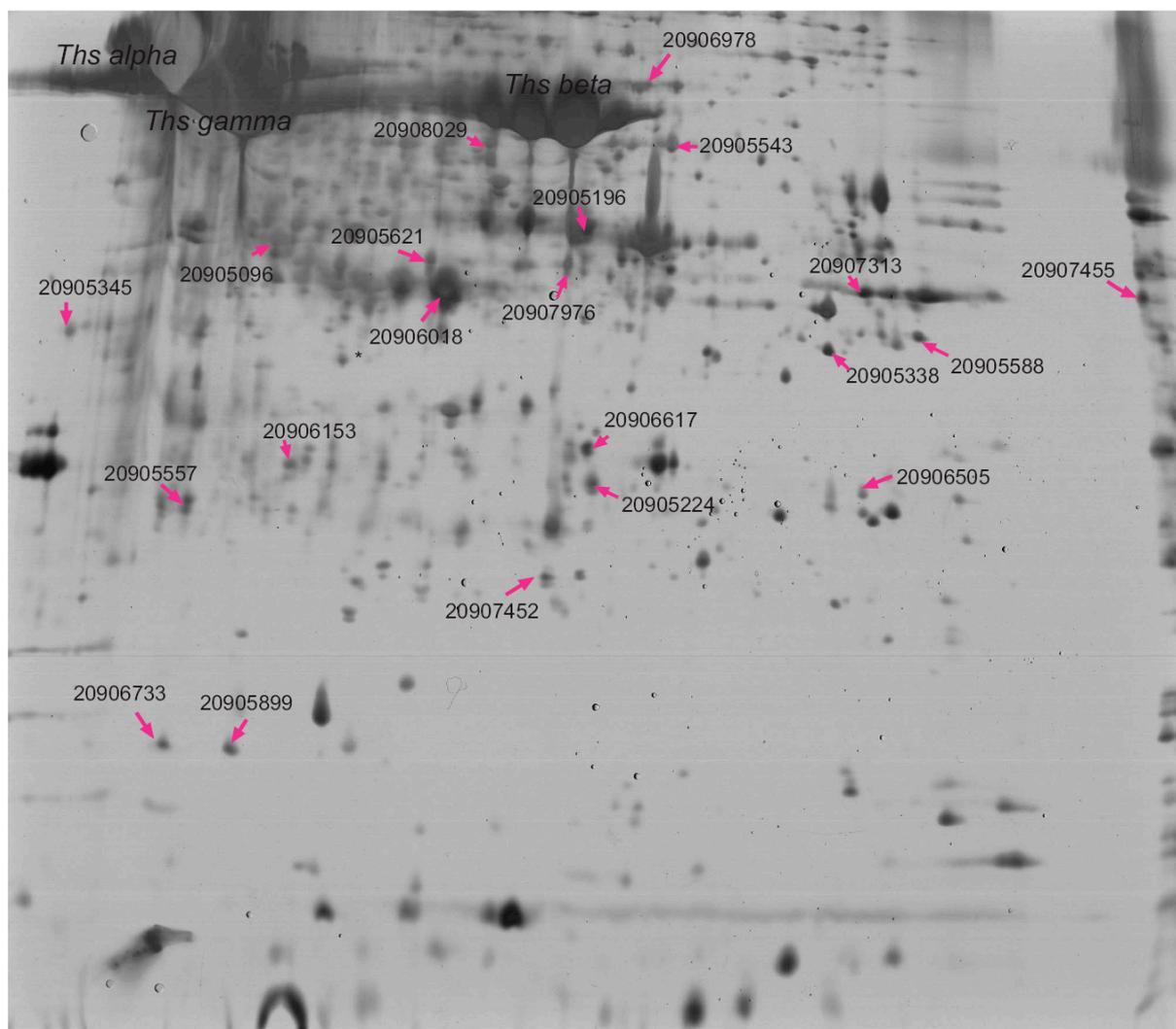
Proteins from the *MmGroEL*/*GroES* immunoprecipitation were focused to their isoelectric point at a range between pH3 and pH10 and then separated by 16%- SDS PAGE. Proteins were detected by silver staining and spots were picked by visual inspection. Proteins that were identified by mass spectrometry in *MmGroEL* immunoprecipitations and *MmThs* immunoprecipitations are with pink arrows and labelled with the corresponding gi-accession number; these proteins are listed in table S1.

**Figure S2: GroEL- specific substrates, 2D-PAGE, silverstained**



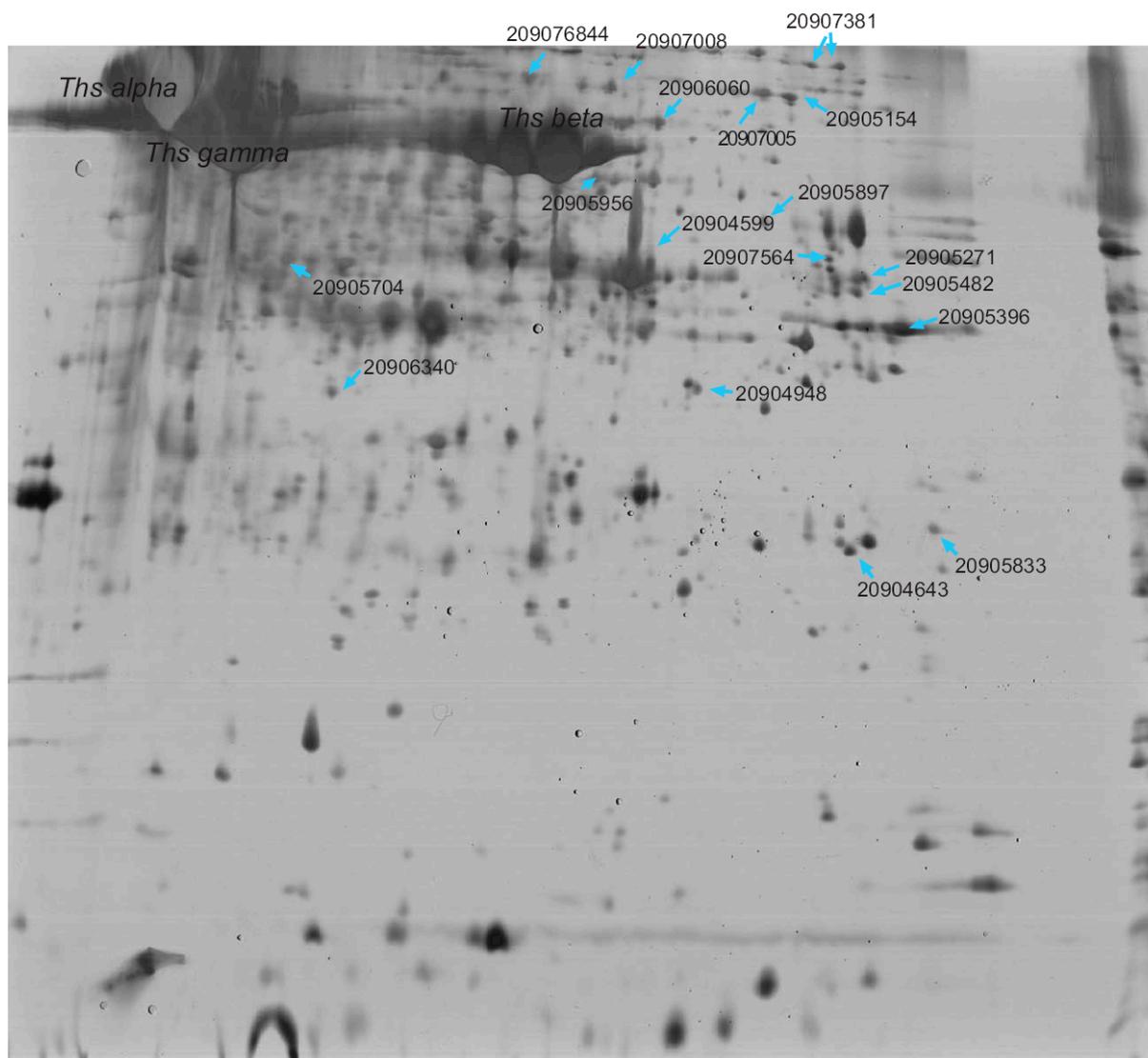
Proteins from the *MmGroEL*/GroES immunoprecipitation were focused to their isoelectric point at a range between pH3 and pH10 and then separated by 16%- SDS PAGE. Proteins were detected by silver staining and spots were picked by visual inspection. Proteins that were identified by mass spectrometry only in *MmGroEL* immunoprecipitations are with orange arrows and labelled with the corresponding gi-accession number; these proteins are listed in table S2.

**Figure S3: *Ths*- overlapping substrates, 2D-PAGE, silverstained**



Proteins from the *MmThs* immunoprecipitation were focused to their isoelectric point at a range between pH3 and pH10 and then separated by 16%- SDS PAGE. Proteins were detected by silver staining and spots were picked by visual inspection. Proteins that were identified by mass spectrometry in *MmGroEL* immunoprecipitations and *MmThs* immunoprecipitations are with pink arrows and labelled with the corresponding gi-accession number; these proteins are listed in table S3.

**Figure S4: *Ths*- specific substrates, 2D-PAGE, silverstained**



Proteins from the *MmThs* immunoprecipitation were focused to their isoelectric point at a range between pH3 and pH10 and then separated by 16%- SDS PAGE. Proteins were detected by silver staining and spots were picked by visual inspection. Proteins that were identified by mass spectrometry only in *MmThs* immunoprecipitations are marked with turquoise arrows and labelled with the corresponding gi-accession number; these proteins are listed in table S4.

**Figure S5: GroEL- overlapping substrates, Ettan-DIGE**



*Mm*GroEL/GroES substrate complexes were labelled with Cy3 and mixed with of *Mm*Ths substrate complexes labelled with Cy5 and the control sample labelled with Cy2. The sample was separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison. Proteins spots showing a more than 2.5 fold stronger signal in the *Mm*GroEL IP compared to the *Mm*Ths IP were marked with red arrows, proteins with similar signal intensities from both co-IPs, but a 2.5 fold higher intensity than in the lysate sample are marked with lilac arrows. Proteins that were identified by mass spectrometry as overlapping substrates are labelled with the corresponding gi-accession number; these proteins are listed in table S5.

**Figure S6: GroEL- specific substrates, Ettan-DIGE**



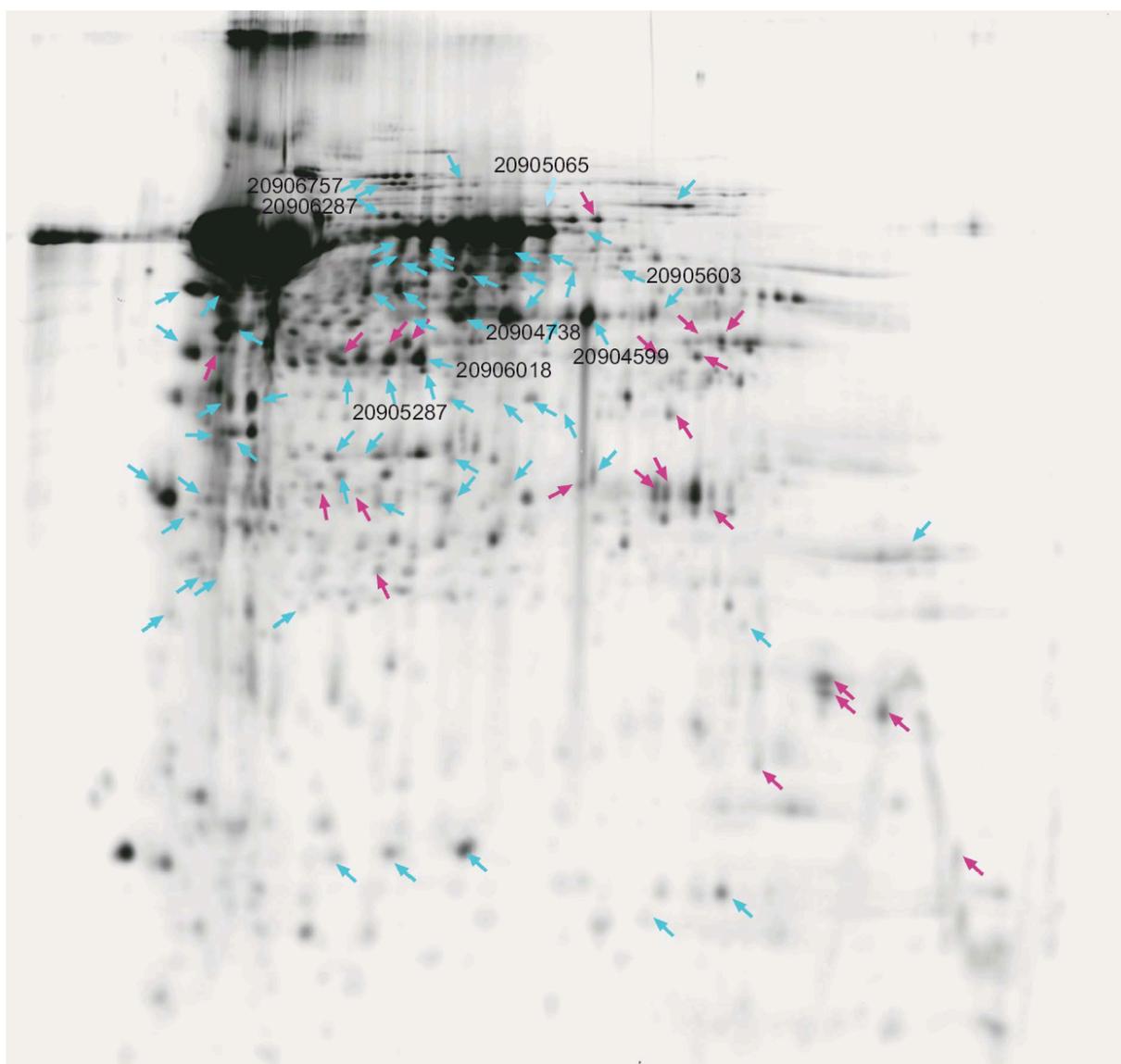
*MmGroEL*/*GroES* substrate complexes were labelled with Cy3 and mixed with of *MmThs* substrate complexes labelled with Cy5 and the control sample labelled with Cy2. The sample was separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison. Proteins spots showing a more than 2.5 fold stronger signal in the *MmGroEL* IP compared to the *MmThs* IP were marked with red arrows, proteins with similar signal intensities from both co-IPs, but a 2.5 fold higher intensity than in the lysate sample are marked with lilac arrows. Proteins that were identified by mass spectrometry as *MmGroEL* specific substrates are labelled with the corresponding gi-accession number; these proteins are listed in table S6.

**Figure S7: *Ths*- overlapping substrates, Ettan-DIGE**



*MmThs* substrate complexes were labelled with Cy5 and mixed with of *Mm* GroEL/GroES substrate complexes labelled with Cy3 and the control sample labelled with Cy2. The sample was separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison. Proteins spots showing a more than 2.5 fold stronger signal in the *MmThs* IP compared to the *MmGroEL* IP were marked with turquoise arrows, proteins with similar signal intensities in both co-IPs, but a 2.5 fold higher intensity than in the lysate sample are marked with lilac arrows. Proteins that were identified by mass spectrometry as overlapping substrates are labelled with the corresponding gi-accession number; these proteins are listed in table S7.

**Figure S8: *Ths*- specific substrates, Ettan-DIGE**



*MmThs* substrate complexes were labelled with Cy5 and mixed with of *Mm* GroEL/GroES substrate complexes labelled with Cy3 and the control sample labelled with Cy2. The sample was separated on 2D-PAGE and visualized by fluorescence on a Typhoon reader and intensities were overlaid for direct comparison. Proteins spots showing a more than 2.5 fold stronger signal in the *MmThs* IP compared to the *MmGroEL* IP were marked with turquoise arrows, proteins with similar signal intensities in both co-IPs, but a 2.5 fold higher intensity than in the lysate sample are marked with lilac arrows. Proteins that were identified by mass spectrometry as *MmThs* specific substrates labelled with the corresponding gi-accession number; these proteins are listed in table S8.

## VII.2 Supplementary Tables

Table S1: GroEL overlapping substrates, 2D-PAGE, silverstained

gi number	name	pI	MW	hydrophobicity	net charge
20904460	Universal stress protein	5.33	16447	0.451	-1.9607
20904462	Universal stress protein	5.41	16296	0.4527	-2.0269
20905054	Mannose-6-phosphate isomerase (EC5.3.1.8)(EC2.7.7.22)	4.99	48988	0.4092	-4.1378
20905160	conserved protein	6.58	28653	0.3961	-0.3921
20905196	metallo cofactor biosynthesis protein	5.66	44810	0.4502	-1.7412
20905439	Cell division protein	4.61	39359	0.4251	-4.278
20905557	Hypothetical UPF0264 protein MM1114	4.73	24942	0.4658	-2.9914
20905668	Aspartate carbamoyltransferase EC 2.1.3.2 Aspartate transcarbamyl	6.3	35096	0.4434	-0.9708
20905750	Hypothetical protein MM1286 EC 3.1.2.6	5.58	23223	0.3944	-2.8168
20905817	Fe-S oxidoreductase	7.53	41968	0.411	0.274
20906018	Citrate synthase EC 4.1.3.7	5.33	40216	0.4394	-2.5351
20906126	Conserved protein	5.73	36119	0.426	-0.9062
20906153	Conserved protein	5.05	30806	0.4297	-3.0417

gi number	name	pI	MW	hydrophobicity	net charge
20906215	Imidazole glycerol phosphate synthase subunit hisF EC 4.1.3.- IGP	4.54	29691	0.4212	-5.4944
20906617	IronIII dicitrate transport ATP-binding protein	5.8	29715	0.4684	-1.8586
20906712	Cytidylate kinase EC 2.7.4.14 CK Cytidine monophosphate kinase CM	5.73	20587	0.392	-1.1363
20907024	Glutamine amidotransferase subunit pdxT EC 2.6.-.- Glutamine amid	5.35	21425	0.4778	-2.463
20907313	Hypothetical protein MM2693	7.06	40291	0.3846	0
20907976	Transcriptional regulator	5.68	42892	0.4354	-2.1107
20908029	Conserved protein EC 1.3.-.-	5.41	54567	0.4215	-2.2726

Table S2: GroEL specific substrates, 2D-PAGE, silverstained

gi number	Name	pI	MW	hydrophobicity	net charge
20905108	Hypothetical UPF0284 protein MM0708	5.34	37006	0.467	-2.0056
20905169	Conserved protein	7.81	15867	0.4338	0
20905556	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide iso	5.63	26068	0.4667	-1.2499
20905756	Archaeosine tRNA-ribosyltransferase EC 2.4.2.-	6.63	69097	0.4222	-0.3209
20905974	Methyltransferase EC 2.1.1.-	5.94	43319	0.4567	-0.7873

gi number	Name	pI	MW	hydrophobicity	net charge
20906195	iron-sulfur binding reductase	5.53	42148	0.3796	-1.3088
20906942	Conserved protein	5.31	25144	0.4444	-3.111
20907091	FO synthase subunit 1 EC 2.5.1.-	4.94	38308	0.4364	-3.1791
20907465	Peptidyl-prolyl cis-trans isomerase EC 5.2.1.8	6.08	17083	0.4129	-1.2902

Table S3: This overlapping substrates, 2D-PAGE, silverstained

gi number	name	pI	MW	hydrophobicity	net charge
20905096	Cell division protein	4.96	42234	0.4184	-2.806
20905196	Metallocofactor biosynthesis protein	5.66	44810	0.4502	-1.7412
20905224	Probable peroxiredoxin (EC 1.11.1.15)	7.09	24946	0.4566	0
20905338	Conserved protein	6.36	34476	0.4502	-0.9645
20905345	Agmatinase (EC 3.5.3.11)	4.31	32335	0.4467	-7.56
20905543	Archaeosine tRNA-ribosyltransferase (EC 2.4.2.-)	6.06	55032	0.4367	-1.2244
20905557	Hypothetical UPF0264 protein MM1114	4.73	24942	0.4658	-2.9914
20905588	Glycosyltransferase (EC 2.4.1.-)	6.89	35654	0.3698	-0.3214

gi number	name	pl	MW	hydrophobicity	net charge
20905621	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)	5.3	39514	0.4085	-2.8168
20905899	Small heat shock protein	4.82	17814	0.3856	-3.9215
20906018	Citrate synthase (EC 4.1.3.7)	5.33	40216	0.4394	-2.5351
20906153	Conserved protein	5.05	30806	0.4297	-3.0417
20906505	Hypothetical protein MM1967	6.62	28659	0.4217	-0.4015
20906617	Iron(III) dicitrate transport ATP-binding protein	5.8	29715	0.4684	-1.8586
20906733	F420-nonreducing hydrogenase II (Hydrogenase expression/formation protein)	4.7	17385	0.4845	-4.9688
20906978	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	5.78	61380	0.4257	-2.0182
20907313	Hypothetical protein MM2693	7.06	40291	0.3846	0
20907452	Anthranilate synthase, component II (EC 4.1.3.27)	6.35	24072	0.4326	-0.9301
20907455	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)	8.2	39203	0.4405	0.5405
20907976	Transcriptional regulator	5.68	42892	0.4354	-2.1107
20908029	conserved protein	5.41	54567	0.4215	-2.2726

Table S4: The specific substrates, 2D-PAGE, silverstained

gi number	Name	pI	MW	hydrophobicity	net charge
20904599	Serine-pyruvate aminotransferase (EC 2.6.1.51)	6.14	42359	0.455	-0.7711
20904643	Conserved protein	6.55	24843	0.4352	-0.4629
20904948	Hypothetical protein MM0563	6.01	33339	0.3746	-1.0308
20905154	Dihydropyrimidinase (EC 3.5.2.2)	6.25	69827	0.4408	-1.0902
20905271	Conserved protein	7.18	45552	0.4706	0
20905396	Coenzyme F420 hydrogenase beta subunit (EC 1.12.99.1)	7.31	37194	0.4215	0.2907
20905482	GTP-binding protein	6.64	40758	0.3901	-0.5494
20905704	conserved protein	4.93	45077	0.4331	-2.9196
20905833	ABC transporter, ATP-binding protein	6.64	29748	0.4144	-0.3801
20905897	Aspartate kinase (EC 2.7.2.4)	6.36	51914	0.4447	-0.835
20905956	Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (TrpRS)	5.71	55728	0.4121	-2.0201
20906060	Molybdenum formylmethanofuran dehydrogenase subunit (EC 1.2.99.5)	5.92	65094	0.4007	-1.8835
20906340	Conserved protein	5.16	31647	0.4296	-3.78
20906844	DNA-directed RNA polymerase beta chain (EC 2.7.7.6)	5.6	67023	0.4023	-2.6489

gi number	name	pI	MW	hydrophobicity	net charge
20907005	Type II DNA topoisomerase VI, subunit B	6.26	68760	0.4477	-0.9661
20907008	DNA gyrase, subunit B	5.76	71068	0.3864	-1.7349
20907381	conserved protein	8.12	83634	0.3796	0.5442
20907564	Oxidoreductase	6.44	45114	0.3945	-0.5024

Table S5: GroEL substrates, DIGE

gi number	name	pI	MW	hydrophobicity	net charge
20905108	Hypothetical UPF0284 protein MM0708	5.34	37006	0.467	-2.0056
20905395	Glutamate synthase, large chain (EC.1.4.1.13)	6.43	26637	0.4211	-0.8096
20905425	Precorrin-3B C17-methyltransferase (EC 2.1.1.131)	4.9	29087	0.4377	-2.6414
20905439	Cell division protein	4.61	39359	0.4251	-4.278
20905556	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide iso	5.63	26068	0.4667	-1.2499
20905771	Conserved protein	7.58	29025	0.433	0
20906126	Conserved protein	5.73	36119	0.426	-0.9062
20906195	Iron-sulfur binding reductase	5.53	42148	0.3796	-1.3088
20906605	Putative nucleotidyltransferase (EC 2.7.7.-)	5.38	22360	0.4433	-0.9851
20906942	Conserved protein	5.31	25144	0.4444	-3.111
20906945	Shikimate kinase (EC 2.7.1.71)	4.94	30952	0.4437	-2.389

Table S6: Ths substrates, DIGE

gi number	name	pI	MW	hydrophobicity	net charge
20904599	Serine-pyruvate aminotransferase (EC 2.6.1.51)	6.14	42359	0.455	-0.7711
20904738	MoxR-like ATPase	5.24	36517	0.4321	-1.8518
20905065	Acetolactate synthase large subunit (EC 4.1.3.18)	5.67	61858	0.4415	-1.7729
20905287	Hypothetical UPF0219 protein MM0871	5.39	37070	0.4441	-1.7191
20905603	NDP-N-acetyl-D-galactosaminuronic acid dehydrogenase (EC 1.1.1.-)	6.13	51826	0.4414	-0.8367
20906018	Citrate synthase (EC 4.1.3.7)	5.33	40216	0.4394	-2.5351
20906287	Pyruvate, phosphate dikinase (EC 2.7.9.1)	5.27	97107	0.4389	-2.0361
20906757	DNA double-strand break repair rad50 ATPase	5.23	121981	0.3701	-2.5233
20907177	methyltransferase	4.73	45591	0.4216	-4.4117

Table S7: overlapping substrates, DIGE

gi number	name	pI	MW	hydrophobicity	net charge
20906121	Divalent cation transport protein	5.24	41395	0.4075	-2.4128
20906505	Hypothetical protein MM1967	6.62	28659	0.4217	-0.4015
20907313	Hypothetical protein MM2693	7.06	40291	0.3846	0
20908039	Hypothetical protein MM3340	4.46	31481	0.461	-6.4406

Table S8: GroEL substrates, 1D-LCMS, proteins are sorted according to their relative abundance on *MmGroEL*, specific substrates are indicated with bold and italic *gi|accession* numbers

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
<b>20905556</b>	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16).			A	0.753229	0.4667	-1.2499	5.63	26068	c.1	2.62%
20907313	Hypothetical protein MM2693.			A	0.641863	0.3846	0	7.06	40291	x	2.54%
<b>20906942</b>	Conserved protein.			A	0.702047	0.4444	-3.111	5.31	25144	c.26	2.21%
20905395	Glutamate synthase, large chain (EC 1.4.1.13).			A	0.718357	0.4211	-0.8096	6.43	26637	b.80	1.77%
20906605	Putative nucleotidyltransferase (EC 2.7.7.-).			A	0.701416	0.4433	-0.9851	5.38	22360	c.68	1.74%
20906287	Pyruvate, phosphate dikinase (EC 2.7.9.1).			A	0.769966	0.4389	-2.0361	5.27	97107	d.142. c.8	c.1. 1.57%
<b>20906191</b>	DNA mismatch repair protein mutS.	MUTS		B	0.77053	0.4111	-2.111	5.37	100504	c.55	1.56%
20906617	Iron(III) dicitrate transport ATP-binding protein.			A	0.770927	0.4684	-1.8586	5.8	29715	c.37	1.49%
20905557	Hypothetical UPF0264 protein MM1114.			A	0.627721	0.4658	-2.9914	4.73	24942	c.1	1.43%
20904462	Universal stress protein.		Signal transduction mechanisms genes	A	0.763365	0.4527	-2.0269	5.41	16296	c.26	1.37%
20906153	Conserved protein.			A	0.479189	0.4297	-3.0417	5.05	30806	a.4	1.28%
20906505	Hypothetical protein MM1967.			A	0.667214	0.4217	-0.4015	6.62	28659	c.55	1.25%
<b>20905174</b>	Glycosyltransferase (EC 2.4.1.-).			B	0.780968	0.4019	1.9293	8.94	36580	x	1.18%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
<b>20905898</b>	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1) (Phosphoglyceromutase) (BPG-independent PGAM) (aPGAM).	APGM		A	0.722395	0.4423	-3.8461	5.01	43175	c.76	1.10%
20907980	Conserved protein			A	0.753149	0.4615	-2.0512	4.97	21622	b.45	1.03%
20905318	Conserved protein.		Transcription genes	A	0.479189	0.4368	-1.5325	6	30471	a.4	1.03%
20905439	Cell division protein.	FTSZ		A	0.769563	0.4251	-4.278	4.61	39359	c.32. d.79	0.99%
<b>20906036</b>	Tetrahydromethanopterin S-methyltransferase subunit B (EC 2.1.1.86) (N5-methyltetrahydromethanopterin--coenzyme M methyltransferase subunit B).	MTRB		A	0.501014	0.4722	-3.7036	4.38	11741	x	0.98%
20905160	Conserved protein.		Posttranslational modification, protein turnover, chaperones genes	A	0.759662	0.3961	-0.3921	6.58	28653	x	0.97%
20905992	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase (EC 5.3.1.16)	HISA		A	0.7643	0.4553	-3.6584	4.67	25713	c.1	0.94%
20904675	Valyl-tRNA synthetase (EC 6.1.1.9).		Translation genes	A	0.770927	0.4304	-3.1069	5.05	99130	b.51. a.27 c.26.	0.94%
<b>20905108</b>	Hypothetical UPF0284 protein MM0708.		Coenzyme transport and metabolism genes	A	0.703438	0.467	-2.0056	5.34	37006	c.39	0.93%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on EL
20907454	N-(5'-phosphoribosyl)anthranilate isomerase 1 (EC 5.3.1.24) (PRAI 1).	TRPF1		A	0.763212	0.4658	-2.564	5.01	25506	c.1	0.90%
20905094	Beta-phosphoglucosmutase (EC 5.4.2.6).		General function prediction only genes	A	0.769139	0.4912	-4.4247	4.71	25436	c.108	0.89%
20904821	Cell division control protein.		Posttranslational modification, protein turnover, chaperones genes	A	0.770927	0.4432	-3.7878	4.9	88391	c.37. d.31. c.37. b.52	0.88%
<b>20907691</b>	Lactoylglutathione lyase (EC 4.4.1.5).			A	0.774758	0.4538	-0.7691	6.62	14796	d.32	0.86%
<b>20907267</b>	Hypothetical protein MM2652.			B	0.501014	0.4968	-1.2902	5.28	16940	x	0.81%
20904852	Putative inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205).		General function prediction only genes	A	0.756163	0.4675	-4.7336	4.73	18926	d.37. d.37	0.79%
<b>20905629</b>	Glycosyltransferase (EC 2.4.1.-).			A	0.742385	0.4105	1.6529	9.12	41657	c.87	0.78%
<b>20905169</b>	Conserved protein.		Coenzyme transport and metabolism genes	A	0.575975	0.4338	0	7.81	15867	d.190	0.78%
<b>20907465</b>	Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8).			A	0.770251	0.4129	-1.2902	6.08	17083	b.62	0.78%
<b>20904729</b>	Hypothetical protein MM0363.		Carbohydrate transport and metabolism genes	A	0.702079	0.4554	-1.7856	6.09	12682	b.82	0.74%
20904460	Universal stress protein.		Signal transduction mechanisms genes	A	0.760552	0.451	-1.9607	5.33	16447	c.26	0.73%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20907621	Isoleucyl-tRNA synthetase (EC 6.1.1.5).			A	0.770927	0.4319	-2.93	5.1	120411	a.27. c.26. b.51	0.72%
20906733	F420-nonreducing hydrogenase II (Hydrogenase expression/formation protein).	VHTD		A	0.706262	0.4845	-4.9688	4.7	17385	c.56	0.72%
20907976	Transcriptional regulator.			A	0.750851	0.4354	-2.1107	5.68	42892	x	0.72%
20906712	Cytidylate kinase (EC 2.7.4.14) (CK) (Cytidine monophosphate kinase) (CMP kinase).	CMK		A	0.661161	0.392	-1.1363	5.73	20587	c.37	0.70%
20905732	GMP synthase [glutamine-hydrolyzing] subunit A (EC 6.3.5.2) (Glutamine amidotransferase).	GUAAA		A	0.768004	0.4497	-8.4655	4.73	21138	c.23	0.69%
<b>20907635</b>	Hypothetical protein MM2979.			B	0.764286	0.427	-3.3057	5.22	42680	x	0.68%
20905090	Transcription initiation factor IIE, alpha subunit.		Transcription genes	A	0.641088	0.4085	-3.0487	5.15	19145	a.4	0.67%
20906635	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase) (OMPDCase) (OMPdecase).	PYRF		A	0.764643	0.4955	-3.1817	4.79	23503	c.1	0.67%
<b>20905771</b>	Conserved protein.			A	0.760666	0.433	0	7.58	29025	d.37. d.37. d.37. d.37	0.65%
20907600	Conserved protein.			A	0.220293	0.4296	2.963	9.62	15721	x	0.64%
20906258	Probable porphobilinogen deaminase (EC 2.5.1.61) (PBG) (	HEMC		A	0.768779	0.4146	-3.1645	5.04	34847	c.94. d.50	0.63%
20907024	Glutamine amidotransferase subunit pdxT (EC 2.6.-.-) (Glutamine amidotransferase glutaminase subunit pdxT).	PDXT		B	0.742677	0.4778	-2.463	5.35	21425	c.23	0.63%
<b>20905985</b>	Conserved protein.			A	0.694439	0.5158	-2.1052	4.93	10315	x	0.62%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20906215	Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase)	HISF		A	0.764654	0.4212	-5.4944	4.54	29691	c.1	0.59%
<b>20906589</b>	Aspartate aminotransferase (EC 2.6.1.1).			A	0.768609	0.4427	-4.3256	4.74	43475	c.67	0.58%
20906126	Conserved protein.			A	0.712044	0.426	-0.9062	5.73	36119	x	0.57%
20905106	Thioredoxin 2.		General function prediction only genes	A	0.672582	0.4112	-4.2055	4.73	24120	x	0.56%
20906904	Ech Hydrogenase, Subunit.	ECHF		A	0.761782	0.4286	7.1429	9.13	14014	d.58	0.55%
20907459	Indole-3-glycerol phosphate synthase (EC 4.1.1.48).			B	0.764627	0.4682	-4.4943	4.84	29395	c.1	0.54%
<b>20904631</b>	Hypothetical protein MM0275.			A	0.746497	0.3833	-0.3332	6.66	67664	x	0.54%
20904632	Superfamily II DNA and RNA helicase.		Replication, recombination and repair genes	A	0.755825	0.3929	-1.5685	5.46	144748	c.37	0.54%
<b>20904671</b>	Conserved protein.			A	0.406275	0.3829	1.5823	8.51	36898	d.144	0.54%
20905036	Conserved protein.			A	0.220293	0.3953	-4.054	4.87	33303	x	0.53%
20906018	Citrate synthase (EC 4.1.3.7).			A	0.768742	0.4394	-2.5351	5.33	40216	a.103	0.53%
20906553	Phosphate import ATP-binding protein pstB (EC 3.6.3.27) (Phosphate-transporting ATPase) (ABC phosphate transporter).	PSTB			0.770927	0.4147	1.1628	8.55	28977	c.37	0.53%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20905231	Conserved protein.		Function unknown genes	A	0.562052	0.4328	1.4925	8.73	14955	x	0.53%
20906877	Hypothetical UPF0173 metal-dependent hydrolase MM2300.			B	0.754257	0.4478	-4.7825	5.1	24933	d.157	0.52%
<b>20905425</b>	Precorrin-3B C17-methyltransferase (EC 2.1.1.131).			A	0.765161	0.4377	-2.6414	4.9	29087	c.90	0.51%
20907745	Putative transcriptional regulator.	SYRB		A	0.760653	0.4392	0.6757	9.09	15986	c.26	0.49%
<b>20906195</b>	Iron-sulfur binding reductase.			A	0.76049	0.3796	-1.3088	5.53	42148	c.96. d.58	0.49%
20907062	Metallo cofactor biosynthesis protein.			A	0.739277	0.4147	0	7.09	24558	x	0.49%
20907086	F420H2 dehydrogenase subunit FpoB.	FPOB		A	0.761565	0.413	0	7.06	20732	e.19	0.48%
20905224	Probable peroxiredoxin (EC 1.11.1.15).		Posttranslational modification, protein turnover, chaperones genes	A	0.768515	0.4566	0	7.09	24946	c.47	0.47%
<b>20904778</b>	Hypothetical protein MM0408.		Secondary metabolites biosynthesis, transport and catabolism genes	A	0.487257	0.4388	-1.4387	5.3	15512	b.45	0.47%
20904330	Dipeptide ABC transporter, ATP-binding protein.	DPPF	Amino acid transport and metabolism genes	B	0.770927	0.4039	1.9704	9.02	22741	c.37	0.47%
20905115	Fructose-bisphosphate aldolase (EC 4.1.2.13).		Carbohydrate transport and metabolism genes	A	0.748102	0.4304	-5.1779	4.77	34055	c.1	0.47%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20905723	Xanthine-guanine phosphoribosyltransferase (EC 2.4.2.22).			A	0.743353	0.4118	-2.745	5.32	29374	c.61	0.46%
20905667	Aspartate carbamoyltransferase regulatory chain.	PYRI		A	0.733567	0.4231	-0.6409	6.36	17011	g.41. d.58	0.44%
20905668	Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate transcarbamylase) (ATCase).	PYRB		A	0.76727	0.4434	-0.9708	6.3	35096	c.78. c.78	0.44%
<b>20904817</b>	Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2).		Nucleotide transport and metabolism genes	B	0.772021	0.4455	-0.4949	6.63	22428	c.65	0.44%
20905780	Hypothetical UPF0200 protein MM1313.			A	0.662347	0.4171	-1.6042	5.46	20893	c.37	0.44%
20905096	Cell division protein.	FTSZ	Cell cycle control, mitosis and meiosis genes	A	0.768938	0.4184	-2.806	4.96	42234	d.79. c.32	0.43%
20908029	Conserved protein (EC 1.3.-.-).			A	0.762771	0.4215	-2.2726	5.41	54567	c.3	0.43%
20906666	Conserved protein.			B	0.625699	0.39	-0.5864	6.64	39363	c.87	0.43%
<b>20904820</b>	NH(3)-dependent NAD(+) synthetase (EC 6.3.1.5).	NADE	Coenzyme transport and metabolism genes	A	0.767239	0.4219	-3.1249	4.9	27786	c.26	0.42%
20906945	Shikimate kinase (EC 2.7.1.71) (SK).	AROK		A	0.657089	0.4437	-2.389	4.94	30952	d.58. d.14	0.42%
<b>20907311</b>	Hypothetical protein MM2691.			B	0.703115	0.4962	-2.2555	4.98	14631	b.45	0.42%
20905716	CdcH protein.			A	0.770927	0.4274	-2.7962	5.22	83906	c.37. d.31. b.52. c.37	0.42%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20907557	Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).	SERC		A	0.759299	0.4324	-1.6215	5.49	41596	c.67	0.42%
20904757	Heterodisulfide reductase subunit HdrB (EC 1.97.1.-).		Energy production and conversion genes	A	0.70752	0.4013	-1.5923	6.06	35102	x	0.41%
20906956	Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) (PRA-PH).	HISE		A	0.763686	0.3925	-6.542	4.9	12150	x	0.41%
20907603	Hypothetical protein MM2951.			A	0.220293	0.4351	-4.5454	4.62	17806	x	0.40%
20905906	Replication factor C subunit.			A	0.769859	0.4252	-2.9325	4.96	38349	c.37. a.80	0.40%
20904876	Iron-sulfur flavoprotein.		General function prediction only genes	A	0.741766	0.4063	1.0417	7.85	21070	c.23	0.40%
<b>20906323</b>	Aspartate aminotransferase (EC 2.6.1.1).			A	0.768389	0.4391	-3.2994	5.17	43578	c.67	0.40%
20905065	Acetolactate synthase large subunit (EC 4.1.3.18).		Amino acid transport and metabolism genes	A	0.76593	0.4415	-1.7729	5.67	61858	c.31. c.36. c.36.	0.40%
20907017	Methylthiol:coenzyme M methyltransferase.	MTSA		B	0.763392	0.4659	-3.5421	4.81	40192	c.1	0.40%
20906966	Conserved protein.			A	0.220293	0.406	-2.2555	5.32	14935	x	0.40%
<b>20905331</b>	Potassium/copper-transporting ATPase (EC 3.6.1.36).			A	0.766431	0.4699	-1.1277	5.47	29177	c.108	0.39%
20906433	Hypothetical protein MM1902.			A	0.766222	0.3988	-0.6134	6.62	18551	d.108	0.39%
20906188	Precorrin-6Y C5,15-methyltransferase (EC 2.1.1.132).			A	0.744608	0.4564	-2.0512	5.19	21018	c.90	0.39%
<b>20906560</b>	Nascent polypeptide-associated complex protein.	NAC		B	0.639726	0.4417	-5.8332	4.38	12839	a.5	0.39%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20904986	Hypothetical UPF0218 protein MM0598.		Function unknown genes	A	0.641997	0.4308	-1.5384	5.66	21900	x	0.38%
20907452	Anthranilate synthase, component II (EC 4.1.3.27).			A	0.768004	0.4326	-0.9301	6.35	24072	c.23	0.37%
20907237	DNA-directed RNA polymerase subunit P (EC 2.7.7.6).	RPOP	Transcription genes	B	0.617192	0.2667	13.3333	9.57	5379	g.41	0.37%
20906240	Heme biosynthesis protein.	NIRD		B	0.706346	0.4486	-2.7026	5.23	21443	a.4	0.37%
20905196	Metallo cofactor biosynthesis protein.		General function prediction only genes	B	0.753541	0.4502	-1.7412	5.66	44810	d.115	0.36%
20906586	Conserved protein.			A	0.220293	0.4074	3.7037	9	11891	x	0.36%
<b>20905062</b>	Nitroreductase family protein.		Energy production and conversion genes	A	0.763627	0.45	-1.111	5.69	20152	d.90	0.35%
<b>20906081</b>	Iron-sulfur flavoprotein.			A	0.702682	0.4	-2.7026	4.95	20310	c.23	0.35%
<b>20904704</b>	Amidophosphoribosyltransferase (EC 2.4.2.14).		Nucleotide transport and metabolism genes	A	0.768037	0.4144	-1.0308	6.21	53117	d.153. c.61	0.34%
20904947	Hypothetical protein MM0562.			A	0.7327	0.4353	-0.862	5.96	26949	x	0.34%
<b>20904594</b>	Glycerol-3-phosphate cytidyltransferase (EC 2.7.7.39).		Cell wall/membrane biogenesis genes	A	0.760285	0.4702	0	7.25	16859	c.26	0.34%
<b>20906125</b>	Conserved protein.			A	0.220293	0.3028	-13.761	3.84	12288	x	0.33%
20905017	Nicotinamide-nucleotide adenyltransferase (EC 2.7.7.1) (NAD(+) pyrophosphorylase) (NAD(+) diphosphorylase) (NMN adenyltransferase).		Coenzyme transport and metabolism genes	A	0.704948	0.4365	0	7.64	19970	c.26	0.33%

gi number	Name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
<b>20905546</b>	Oxidoreductase.			A	0.744452	0.4208	1.039	8.68	42509	c.2. c.3	0.33%
20905338	Conserved protein.			A	0.501014	0.4502	-0.9645	6.36	34476	x	0.33%
20905817	Fe-S oxidoreductase.			A	0.677452	0.411	0.274	7.53	41968	x	0.33%
20905345	Agmatinase (EC 3.5.3.11).		Amino acid transport and metabolism genes	A	0.758841	0.4467	-7.56	4.31	32335	c.42	0.32%
<b>20907091</b>	FO synthase subunit 1 (EC 2.5.1.-)	COFG		A	0.741995	0.4364	-3.1791	4.94	38308	x	0.32%
<b>20905034</b>	Isocitrate dehydrogenase (EC 1.1.1.42).		Amino acid transport and metabolism genes	A	0.764949	0.4606	-2.1211	5.18	35867	c.77	0.32%
20905045	Glycosyltransferase (EC 2.4.-.-).		Cell wall/membrane biogenesis genes	A	0.76637	0.3779	2.6059	9.27	35839	c.68	0.31%
20906241	Heme biosynthesis protein.	NIRH		A	0.725238	0.4114	-0.6328	6.51	18042	d.58. a.4	0.31%
20905750	Hypothetical protein MM1286 (EC 3.1.2.6).			A	0.761081	0.3944	-2.8168	5.58	23223	d.157	0.31%
20905588	Glycosyltransferase (EC 2.4.1.-).			A	0.766298	0.3698	-0.3214	6.89	35654	c.68	0.31%
20905899	Small heat shock protein.			A	0.74964	0.3856	-3.9215	4.82	17814	b.15	0.30%
20907312	Hypothetical protein MM2692.			B	0.358138	0.3996	-1.7278	5.16	52553	c.3	0.30%
20905193	V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).	AHAI,ATPI	Energy production and conversion genes	A	0.736091	0.4702	-2.29	5.6	71964	x	0.30%
20907052	Methyltransferase (EC 2.1.1.-).			A	0.760958	0.3589	-3.1358	4.8	33057	c.66	0.30%
20905368	Protein pcrB homolog.	PCRB		A	0.701676	0.4777	-2.0242	5.31	25888	c.1	0.30%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
<b>20906929</b>	Ribosomal protein S6 modification protein.			A	0.764358	0.4419	-0.3321	6.52	33690	d.142	0.29%
<b>20905843</b>	DNA repair protein.			A	0.77048	0.371	2.4194	9.08	27333	c.37	0.29%
20905941	Conserved protein.			A	0.726229	0.4269	-1.1791	5.47	47156	x	0.29%
<b>20906466</b>	Methylcobalamin:coenzyme M methyltransferase.	MTBA		B	0.760187	0.4721	-5.5717	4.41	36882	c.1	0.29%
<b>20906257</b>	Probable glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)	HEML		A	0.767014	0.441	-4.9527	4.73	46306	c.67	0.28%
<b>20905630</b>	Glycosyltransferase (EC 2.4.1.-).			A	0.756837	0.4425	1.5345	8.46	44748	c.87	0.28%
<b>20905560</b>	Polysaccharide deacetylase (EC 3.5.1.-).			A	0.757551	0.4053	-1.3288	6.23	34521	c.6	0.28%
<b>20907623</b>	Conserved protein.			A	0.644291	0.4648	-3.2863	4.84	23537	c.100	0.28%
20908007	Acetyltransferase (EC 2.3.1.-).			A	0.76462	0.3857	-1.4285	6.1	24393	d.108	0.27%
20906121	Divalent cation transport protein.			A	0.767248	0.4075	-2.4128	5.24	41395	c.37	0.27%
<b>20905921</b>	Methylcobamide:CoM methyltransferase mtbA (EC 2.1.1.-)	MTBA		B	0.770577	0.4572	-3.2447	4.87	36582	c.1	0.27%
20907866	Ornithine decarboxylase (EC 4.1.1.17)			B	0.767731	0.4154	-0.7691	6.46	44775	b.49. c.1	0.27%
20905158	Conserved protein.		General function prediction only genes	A	0.762402	0.4156	-6.0605	4.5	25460	c.26	0.27%
20907455	Anthranilate phosphoribosyltransferase (EC 2.4.2.18).	TRPD		A	0.764138	0.4405	0.5405	8.2	39203	c.27. a.46	0.27%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
<b>20905770</b>	Conserved protein.			A	0.761116	0.4591	-0.7116	6.52	31384	d.37. d.37. d.37. d.37	0.27%
20905621	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14).			A	0.765718	0.4085	-2.8168	5.3	39514	c.87	0.26%
<b>20904983</b>	Conserved protein.		General function prediction only genes	A	0.639625	0.4711	4.1322	10.03	13660	c.120	0.26%
20904640	Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).	LEUS	Translation genes	A	0.770927	0.4253	-2.4716	5.32	110957	b.51. a.27. c.26	0.26%
20906869	Type I restriction-modification system specificity subunit.			A	0.768421	0.4097	-2.4751	5.39	91505	c.66	0.26%
<b>20904706</b>	Hypothetical protein MM0342.		Function unknown genes	A	0.751103	0.4015	-4.5454	4.91	14765	x	0.25%
20905077	Conserved protein.		Function unknown genes	A	0.710184	0.3655	-7.1065	4.58	22214	x	0.25%
20905340	4-carboxymuconolactone decarboxylase (EC 4.1.1.44).			A	0.734486	0.4121	-5.4544	4.74	18664	a.152	0.25%
20907577	Hypothetical protein MM2928.			B	0.722121	0.3256	-1.7441	5.43	19191	x	0.25%
20905832	Hypothetical protein MM1359.			A	0.501543	0.4507	5.6338	10.08	15284	x	0.25%
20904975	Conserved protein.		Cell wall/membrane biogenesis genes	A	0.765766	0.4053	0.4796	7.87	47730	c.87	0.25%
20906506	Hypothetical protein MM1968.			A	0.604417	0.439	-6.0975	4.62	28035	b.82	0.25%
<b>20906080</b>	NADPH-flavin oxidoreductase (EC 1.6.99.-).			A	0.757984	0.4713	0.5747	7.97	19485	d.90	0.25%
<b>20907414</b>	Suppressor protein SuhB homolog (EC 3.1.3.25).			A	0.767086	0.4157	-2.2471	5.46	28822	e.7	0.25%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20906820	Conserved protein.			B	0.220293	0.4151	0.9434	7.65	11810	x	0.24%
20904956	ABC transporter, ATP-binding protein.		Inorganic ion transport and metabolism genes	B	0.770927	0.4301	-2.2058	5.64	30186	c.37	0.24%
20905831	Putative methyltransferase.			A	0.71642	0.4412	-7.3528	4.35	7543	x	0.24%
20907109	Hypothetical protein MM2510.			A	0.697179	0.4245	-4.4024	4.68	34498	d.142	0.24%
<b>20904668</b>	Uroporphyrinogen-III synthase (EC 4.2.1.75).		Coenzyme transport and metabolism genes	A	0.748376	0.4349	-2.6021	5.09	29542	c.113	0.24%
<b>20905619</b>	DTDP-4-dehydrorhamnose reductase (EC 1.1.1.133).			A	0.766347	0.4052	-2.6021	5.34	30180	c.2	0.24%
20906782	Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24).			B	0.766903	0.4234	-0.8064	6.14	27249	c.68	0.24%
<b>20905003</b>	Conserved protein.		Cell cycle control, mitosis and meiosis genes	A	0.699645	0.4116	-1.1593	5.82	39275	c.26	0.24%
20904985	DNA-directed RNA polymerase subunit E' (EC 2.7.7.6).		Transcription genes	A	0.648493	0.4262	-1.6392	5.63	6881	x	0.24%
<b>20905574</b>	Conserved protein.			A	0.63646	0.4368	0.2747	7.97	41029	c.87	0.23%
20905403	Conserved protein.			A	0.603238	0.419	-1.6759	5.65	20296	x	0.23%
20905178	Conserved protein.			A	0.220293	0.437	1.4815	9.05	15618	x	0.23%
20905901	F420-0:gamma-glutamyl ligase (EC 6.3.2.-).	COFE		A	0.732841	0.4449	-2.7558	4.95	27741	x	0.23%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20904725	Sulfite reductase, assimilatory-type (EC 1.8.-.-).		Energy production and conversion genes	B	0.763421	0.4545	1.2987	8.53	24912	d.134. d.58	0.22%
20907682	ABC transporter, ATP-binding protein.			A	0.770927	0.411	-1.2711	5.82	26237	c.37	0.22%
20906404	Purine phosphoribosyltransferase (EC 2.4.2.7).			A	0.766566	0.4497	-3.7036	4.73	20679	c.61	0.22%
20905510	Conserved protein.			B	0.768361	0.4	-2.3376	5.71	42368	c.37	0.22%
<b>20907999</b>	Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--tRNA ligase) (ThrRS).	THRS		B	0.770927	0.4274	-1.7349	5.69	72765	d.104. c.51	0.22%
20905756	Archaeosine tRNA-ribosyltransferase (EC 2.4.2.-).			A	0.65145	0.4222	-0.3209	6.63	69097	c.1. b.122. d.17	0.22%
20906591	Hypothetical protein MM2043.			A	0.627596	0.4477	-2.166	5.65	32631	x	0.22%
<b>20905974</b>	Methyltransferase (EC 2.1.1.-).			A	0.747981	0.4567	-0.7873	5.94	43319	c.66. c.66	0.20%
20905279	Glucoamylase (EC 3.2.1.3).		Carbohydrate transport and metabolism genes	B	0.736606	0.4119	-2.2795	5.45	75932	a.102	0.20%
20907082	F420H2 dehydrogenase subunit.	FPOI		A	0.762579	0.4011	1.6949	7.85	19459	d.58	0.19%
<b>20904722</b>	Glutamate dehydrogenase (EC 1.4.1.3).		Amino acid transport and metabolism genes	A	0.765583	0.484	-0.802	6.03	40534	c.2. c.58	0.19%
<b>20905697</b>	Methyltransferase.			A	0.641873	0.411	-1.5336	5.73	36419	c.66	0.19%
20904329	Dipeptide ABC transporter, ATP-binding protein.	DPPD	Amino acid transport and metabolism genes	B	0.770858	0.4123	0.9231	8.23	35955	c.37	0.19%
20904993	Conserved protein.			A	0.241334	0.395	-0.8402	6.15	13497	x	0.19%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20905543	Archaeosine tRNA-ribosyltransferase (EC 2.4.2.-).			A	0.767381	0.4367	-1.2244	6.06	55032	d.17. c.1	0.18%
20907135	Hypothetical UPF0217 protein MM2534.			A	0.65633	0.4028	-3.7914	5.27	23249	x	0.18%
<b>20904531</b>	Short chain dehydrogenase/reductase (EC 1.1.-.-).		Secondary metabolites biosynthesis, transport and catabolism genes	A	0.770637	0.4431	-0.7842	5.48	27556	c.2	0.18%
20906133	Iron-sulfur flavoprotein.			A	0.730466	0.4415	-1.0637	5.66	21070	c.23	0.18%
20906251	Heme biosynthesis protein.	NIRJ		A	0.761061	0.4069	-1.146	6.03	38802	x	0.18%
<b>20904436</b>	Serine/threonine protein phosphatase (EC 3.1.3.16).		Signal transduction mechanisms genes	A	0.736753	0.4322	-3.2966	5.07	31139	d.159	0.18%
20905894	Conserved protein.			A	0.501014	0.4342	-1.9736	5.74	16878	c.115	0.17%
20905578	UDP-glucose 6-dehydrogenase (EC 1.1.1.22).			A	0.765484	0.4436	-1.6786	5.15	45605	c.2. a.100 c.26.	0.17%
<b>20906169</b>	Conserved protein.			A	0.220293	0.3697	0.6061	7.5	19088	x	0.17%
<b>20906959</b>	Conserved protein.			A	0.617818	0.479	-1.7963	5.76	18299	d.58	0.17%
20905755	(S)-2-hydroxy-acid dehydrogenase (EC 1.1.3.15).			A	0.767026	0.4587	-1.3042	5.52	49551	d.145. d.58	0.16%
20904798	Rnase L inhibitor.		General function prediction only genes	A	0.770927	0.4082	-1.7006	5.58	65396	c.37. d.58. c.37	0.16%
<b>20904390</b>	Tungsten formylmethanofuran dehydrogenase subunit D (EC 1.2.99.5).		Energy production and conversion genes	A	0.519204	0.3939	1.5152	9.31	14681	b.52	0.16%
<b>20905267</b>	Conserved protein.		General function prediction only genes	A	0.762284	0.4309	-3.2519	5.87	27743	d.157	0.15%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20906289	Transcription initiation factor IIB (TFIIB).	TFB		A	0.67914	0.3858	2.9674	9.24	38290	g.41. a.74.	0.15%
20904389	Tungsten formylmethanofuran dehydrogenase subunit B (EC 1.2.99.5).		Energy production and conversion genes	A	0.57674	0.4251	-0.7245	6.04	45473	c.81	0.15%
20905700	Methyl-coenzyme M reductase operon protein C.			A	0.501014	0.4223	1.4563	8.59	22427	x	0.14%
20907428	Conserved protein.			A	0.763768	0.4139	1.9868	8.35	33668	c.26	0.14%
<b>20907920</b>	Hydroxyethylthiazole kinase (EC 2.7.1.50) (4-methyl-5-beta-hydroxyethylthiazole kinase) (Thz kinase) (TH kinase).	THIM		A	0.756852	0.4559	-1.9156	5.35	27555	c.72	0.14%
<b>20906937</b>	Probable diphthine synthase (EC 2.1.1.98) (Diphthamide biosynthesis methyltransferase).	DPHB		A	0.67914	0.4248	-2.6315	5.49	29397	c.90	0.14%
<b>20907487</b>	Conserved protein.			A	0.220293	0.3615	-1.8778	6.08	25503	x	0.14%
20908039	Hypothetical protein MM3340.			B	0.220293	0.461	-6.4406	4.46	31481	x	0.14%
<b>20905616</b>	DTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13).			A	0.766065	0.3934	-0.5463	6.39	21004	b.82	0.14%
20906978	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3).			A	0.770339	0.4257	-2.0182	5.78	61380	e.23	0.13%
20906189	Putative cobalt-precorrin-6A synthase [deacetylating] (EC 2.1.1.-).	CBID		A	0.740599	0.4277	-2.0648	5.46	36315	x	0.13%
20907456	Tryptophan synthase alpha chain (EC 4.2.1.20).	TRPA		B	0.765011	0.4649	-3.6899	4.69	29069	c.1	0.13%
20905791	Chemotaxis protein.	CHEC		A	0.720472	0.4528	-5.1886	4.42	23371	x	0.13%
20906041	Conserved protein.			A	0.714573	0.3978	-4.8326	4.99	29514	c.1	0.13%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on EL
20906063	Molybdenum formylmethanofuran dehydrogenase subunit (EC 1.2.99.5).	FMDB		A	0.751098	0.4345	-1.839	5.78	47740	c.81	0.13%
<b>20906451</b>	Hypothetical protein MM1918.			A	0.199251	0.3849	-2.0021	5.83	105187	x	0.12%
<b>20905385</b>	Nitrogen regulatory protein P-II.			A	0.764323	0.4174	1.7391	8.76	12861	d.58	0.12%
20905043	Mannosyltransferase (EC 2.4.1.-).		Cell wall/membrane biogenesis genes	A	0.76544	0.4249	2.5496	9.43	40610	c.87	0.11%
20906891	F420-nonreducing hydrogenase I, large subunit (EC 1.12.2.-).	VHOA		A	0.7398	0.4365	-2.7072	5.17	64243	e.18	0.11%
20904486	Iron-sulfur flavoprotein.		General function prediction only genes	A	0.703765	0.4025	0.8475	7.91	26247	c.23	0.11%
20908032	Monomethylamine corrinoid protein 2 (MMCP 2).	MTMC2		A	0.765737	0.4633	-5.9632	4.42	23270	c.23. a.46	0.11%
20904497	Type IIS restriction enzyme (EC 3.1.21.4) (EC 2.1.1.72).		Defense mechanisms genes	B	0.685957	0.4137	-2.9184	5.16	134098	c.66	0.11%
20906239	Heme biosynthesis protein.	NIRJ		A	0.761123	0.4	-1.4084	5.95	39379	x	0.10%
20905526	Metal dependent hydrolase.			A	0.741609	0.4009	-4.9549	5.05	24425	d.157	0.10%
<b>20905769</b>	Conserved protein.			A	0.742733	0.4414	1.8018	9.9	36798	c.80. d.37	0.10%
<b>20905590</b>	Glycosyltransferase (EC 2.4.1.-).			A	0.762993	0.4529	2.5445	9.35	43734	c.87	0.10%
<b>20904818</b>	Putative HTH-type transcriptional regulatory protein MM0444.		Transcription genes	A	0.648493	0.422	-2.1406	5.22	36296	a.35	0.09%
20905469	Amino-acid acetyltransferase (EC 2.3.1.1).			B	0.760691	0.4403	0	6.75	18041	d.108	0.09%
20905408	CoB--CoM heterodisulfide reductase 1 iron-sulfur subunit C (EC 1.8.98.1).	HDRC		A	0.752359	0.3851	1.2422	7.74	18120	a.1	0.09%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on EL
20906412	Conserved protein.			B	0.624778	0.4511	-1.0869	5.43	19947	x	0.08%
20904811	Methylenetetrahydrofolate reductase (EC 1.5.1.20).		Amino acid transport and metabolism genes	B	0.761607	0.4384	-1.3698	5.87	31927	c.1	0.07%
<b>20907450</b>	Metal dependent hydrolase.			A	0.705694	0.4393	-2.4999	5.69	31160	d.157	0.07%
20907129	N(2),N(2)-dimethylguanosine tRNA methyltransferase (EC 2.1.1.32) (tRNA(guanine-26,N(2)-N(2)) methyltransferase) (tRNA -2,2 dimethylguanosine-26 methyltransferase) (tRNA(m(2,2)G26)dimethyltransferase).	TRM1		A	0.69755	0.4459	-0.5154	6.75	42784	c.66	0.04%
20905871	Probable molybdenum cofactor biosynthesis protein A.	MOAA		A	0.761142	0.3952	1.1976	8.06	37691	x e.10	0.04%
20906757	DNA double strand break repair rad50 ATPase	RAD50		A	0.766547	0.3701	-2.5233	5.23	12198	a.2 c.37	0.04%

Table S9: Ths substrates, 1D-LCMS, proteins are sorted according to their relative abundance on *MmThs*, specific substrates are indicated with bold and italic **gi|accession** numbers

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pI	MW	scop code	on Ths
20906018	Citrate synthase (EC 4.1.3.7).			A	0.768742	0.4394	-2,5351	5,33	40216	a.103	1,67%
20905557	Hypothetical UPF0264 protein MM1114.			A	0.627721	0.4658	-2,9914	4,73	24942	c.1	1,39%
20907313	Hypothetical protein MM2693.			A	0.358138	0.3846	0	7,06	40291	x	1,37%
20906896	Probable hydrogenase nickel incorporation protein hypA.	HYP A		B	0.737554	0.3869	-10,2189	4,42	15135	x	1,28%
20905439	Cell division protein.	FTSZ		A	0.769563	0.4251	-4,278	4,61	39359	d.79, c.32	1,26%
20905395	Glutamate synthase, large chain (EC 1.4.1.13).			A	0.718357	0.4211	-0,8096	6,43	26637	b.80	1,20%
<b>20905020</b>	Zinc finger protein.		Function unknown genes	A	0.465544	0.3816	-5,2631	5,36	8614	x	1,17%
20906287	Pyruvate, phosphate dikinase (EC 2.7.9.1).			A	0.769966	0.4389	-2,0361	5,27	97107	d.142, c.8, c.1	0,99%
<b>20904665</b>	conserved protein.			A	0.220293	0.3837	-4,6511	4,56	9945	x	0,92%
<b>20906894</b>	Hydrogenase expression/formation protein.	HYP C		A	0.735556	0.4302	-11,6278	4,16	9336	b.40	0,91%
<b>20905369</b>	Methyl coenzyme M reductase system, component A2.			A	0.770927	0.4259	-3,0663	5,27	66215	c.37, c.37	0,90%
20905899	Small heat shock protein.			A	0.74964	0.3856	-3,9215	4,82	17814	b.15	0,89%
<b>20906454</b>	Hypothetical protein MM1921.			A	0	0.3571	-7,1428	4,52	12973	x	0,81%
<b>20904988</b>	30S ribosomal protein S27ae.	RPS27AE	Translation genes	A	0.670238	0.3265	12,2449	9,6	5559	x	0,78%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20904876	Iron-sulfur flavoprotein.		General function prediction only genes	A	0.741766	0.4063	1,0417	7,85	21070	c.23	0,75%
20905716	CdcH protein.			A	0.770927	0.4274	-2,7962	5,22	83906	b.52, c.37, d.31, c.37	0,75%
20907441	Phosphoribosylformylglycinamide synthase I (EC 6.3.5.3) (FGAM synthase I).	PURQ		B	0.762958	0.4181	-3,8792	4,88	25615	c.23	0,74%
20907531	conserved protein.			A	0.757405	0.4143	-5,7142	4,79	8241	x	0,73%
20905065	Acetolactate synthase large subunit (EC 4.1.3.18).		Amino acid transport and metabolism genes	A	0.76593	0.4415	-1,7729	5,67	61858	c.31, c.36, c.36	0,73%
20907600	Conserved protein.			A	0.220293	0.4296	2,963	9,62	15721	x	0,70%
20906153	Conserved protein.			A	0.479189	0.4297	-3,0417	5,05	30806	a.4	0,70%
20906605	Putative nucleotidyltransferase (EC 2.7.7.-).			A	0.701416	0.4433	-0,9851	5,38	22360	c.68	0,70%
<b>20904524</b>	conserved protein.			A	0.220293	0.3955	-10,4477	4,27	15559	a.87	0,69%
20905090	Transcription initiation factor IIE, alpha subunit.		Transcription genes	A	0.641088	0.4085	-3,0487	5,15	19145	a.4	0,68%
20906891	F420-nonreducing hydrogenase I, large subunit (EC 1.12.2.-).	VHOA		A	0.7398	0.4365	-2,7072	5,17	64243	e.18	0,68%
20906746	TATA-box binding protein 3 (TATA-box factor 3) (TATA sequence-binding protein 3) (TBP 3) (Box A binding protein 3)	TBP3		A	0.67914	0.4541	-3,7837	4,66	20176	d.129, d.129	0,68%
20905231	Conserved protein.		Function unknown genes	A	0.562052	0.4328	1,4925	8,73	14955	x	0,64%
20905732	GMP synthase [glutamine-hydrolyzing] subunit A (EC 6.3.5.2) (Glutamine amidotransferase).	GUAAA		A	0.768004	0.4497	-8,4655	4,73	21138	c.23	0,63%
20907976	Transcriptional regulator.			A	0.762434	0.4354	-2,1107	5,68	42892	x	0,63%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20905096	Cell division protein.	FTSZ	Cell cycle control, mitosis and meiosis genes	A	0.768938	0.4184	-2,806	4,96	42234	d.79, c.32	0,63%
20907024	Glutamine amidotransferase subunit pdxT (EC 2.6.-.-) (Glutamine amidotransferase glutaminase subunit pdxT).	PDXT		B	0.765737	0.4778	-2,463	5,35	21425	c.23	0,61%
<b>20907861</b>	Ketoisovalerate oxidoreductase subunit.	VORC		A	0.75875	0.378	-2,4389	5,55	9247	a.1	0,61%
20905160	Conserved protein.		Posttranslational modification, protein turnover, chaperones genes	A	0.759662	0.3961	-0,3921	6,58	28653	x	0,61%
20907452	Anthranilate synthase, component II (EC 4.1.3.27).			A	0.765785	0.4326	-0,9301	6,35	24072	c.23	0,60%
20906433	Hypothetical protein MM1902.			A	0.766222	0.3988	-0,6134	6,62	18551	d.108	0,59%
<b>20905673</b>	ATP-dependent DNA helicase (EC 3.6.1.-).	RECQ		B	0.767816	0.3817	-1,0044	5,9	101234	c.37, c.37	0,59%
20905017	Nicotinamide-nucleotide adenyltransferase (EC 2.7.7.1) (NAD(+) pyrophosphorylase) (NAD(+) diphosphorylase) (NMN adenyltransferase).		Coenzyme transport and metabolism genes	A	0.704948	0.4365	0	7,64	19970	c.26	0,55%
20906188	Precorrin-6Y C5,15-methyltransferase (EC 2.1.1.132).			A	0.744608	0.4564	-2,0512	5,19	21018	c.90	0,54%
20906733	F420-nonreducing hydrogenase II (Hydrogenase expression/formation protein).	VHTD		A	0.706262	0.4845	-4,9688	4,7	17385	c.56	0,54%
<b>20905763</b>	Probable RNA 2'-phosphotransferase (EC 2.7.-.-).	KPTA		A	0.718464	0.3865	-2,4154	5,51	23845	x	0,53%
20906617	Iron(III) dicitrate transport ATP-binding protein.			A	0.770927	0.4684	-1,8586	5,8	29715	c.37	0,53%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20907052	Methyltransferase (EC 2.1.1.-).			A	0.742677	0.3589	-3,1358	4,8	33057	c.66	0,53%
<b>20905173</b>	conserved protein.			A	0.220293	0.3884	-10,7437	4,23	13652	x	0,52%
20905365	Adenylosuccinate lyase (EC 4.3.2.2).			A	0.768895	0.4425	-2,876	5,52	50305	a.127	0,52%
20905224	Probable peroxiredoxin (EC 1.11.1.15).		Posttranslational modification, protein turnover, chaperones genes	A	0.768515	0.4566	0	7,09	24946	c.47	0,51%
20905992	1-(5-phosphoribosyl)-5-[[5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase (EC 5.3.1.16) (Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase).	HISA		A	0.7643	0.4553	-3,6584	4,67	25713	c.1	0,51%
20906553	Phosphate import ATP-binding protein pstB (EC 3.6.3.27) (Phosphate-transporting ATPase) (ABC phosphate transporter).	PSTB			0.770927	0.4147	1,1628	8,55	28977	c.37	0,50%
20907456	Tryptophan synthase alpha chain (EC 4.2.1.20).	TRPA		B	0.764138	0.4649	-3,6899	4,69	29069	c.1	0,50%
20905106	Thioredoxin 2.		General function prediction only genes	A	0.672582	0.4112	-4,2055	4,73	24120	x	0,49%
<b>20904409</b>	Hypothetical protein MM0077.		General function prediction only genes	A	0.75124	0.3879	-3,8792	5,35	25929	d.159	0,49%
20906258	Probable porphobilinogen deaminase (EC 2.5.1.61) (PBG) (Hydroxymethylbilane synthase) (HMBS) (Pre-uroporphyrinogen synthase).	HEMC		A	0.768779	0.4146	-3,1645	5,04	34847	d.50, c.94	0,48%
20905094	Beta-phosphoglucomutase (EC 5.4.2.6).		General function prediction only genes	A	0.769139	0.4912	-4,4247	4,71	25436	c.108	0,48%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20905723	Xanthine-guanine phosphoribosyltransferase (EC 2.4.2.22).			A	0.743353	0.4118	-2,745	5,32	29374	c.61	0,47%
<b>20907801</b>	Hypothetical protein MM3127.			B	0.757805	0.45	-6,9999	4,36	11496	x	0,47%
20905832	Hypothetical protein MM1359.			A	0.501543	0.4507	5,6338	10,08	15284	x	0,47%
20906289	Transcription initiation factor IIB (TFIIB).	TFB		A	0.67914	0.3858	2,9674	9,24	38290	a.74, a.74, g.41	0,47%
<b>20905647</b>	DTDP-glucose 4,6-dehydratase (EC 4.2.1.46).			A	0.767599	0.4685	-0,7873	6,34	27840	c.2	0,47%
20904823	SAM-dependent methyltransferases.		Cell wall/membrane biogenesis genes	B	0.747675	0.4322	-6,3558	4,6	27076	c.66	0,47%
20905287	Hypothetical UPF0219 protein MM0871.		Lipid transport and metabolism genes	A	0.76643	0.4441	-1,7191	5,39	37070	c.95, c.95	0,47%
<b>20904399</b>	N-methylhydantoinase (ATP-hydrolyzing) (EC 3.5.2.14).		Amino acid transport and metabolism genes	B	0.748048	0.4526	-1,5788	5,41	61663	c.55	0,46%
<b>20905154</b>	Dihydropyrimidinase (EC 3.5.2.2).		Amino acid transport and metabolism genes	B	0.746053	0.4408	-1,0902	6,25	69827	c.55, c.55, c.55	0,46%
20906945	Shikimate kinase (EC 2.7.1.71) (SK).	AROK		A	0.657089	0.4437	-2,389	4,94	30952	d.14, d.58	0,45%
<b>20906087</b>	conserved protein.			A	0.220293	0.3581	-2,7026	5,34	15269	x	0,45%
<b>20906928</b>	Methyltransferase (EC 2.1.1.-).			A	0.751607	0.4086	-3,5841	5,09	32482	c.66	0,45%
20905588	Glycosyltransferase (EC 2.4.1.-).			A	0.766298	0.3698	-0,3214	6,89	35654	c.68	0,44%
<b>20906120</b>	conserved protein.			A	0.35631	0.3824	-7,843	4,76	11932	x	0,44%
20907237	DNA-directed RNA polymerase subunit P (EC 2.7.7.6).	RPOP	Transcription genes	B	0.67914	0.2667	13,3333	9,57	5379	g.41	0,43%
<b>20906269</b>	conserved protein.			A	0.525073	0.4767	0	7,26	19428	x	0,43%
20905750	Hypothetical protein MM1286 (EC 3.1.2.6).			A	0.761081	0.3944	-2,8168	5,58	23223	d.157	0,43%
<b>20906845</b>	DNA-directed RNA polymerase beta chain (EC 2.7.7.6).			A	0.770899	0.4068	-4,8963	4,81	60584	e.29	0,43%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20907603	Hypothetical protein MM2951.			A	0.220293	0.4351	-4,5454	4,62	17806	x	0,42%
<b>20907978</b>	Hypothetical protein MM3286.			A	0.750851	0.3171	-7,317	4,45	14524	x	0,42%
<b>20905680</b>	conserved protein.			A	0.501014	0.3241	-3,7036	4,83	12239	x	0,42%
<b>20906359</b>	conserved protein.			A	0.646017	0.4271	-7,2916	4,67	10667	x	0,42%
20906240	Heme biosynthesis protein.	NIRD		B	0.706346	0.4486	-2,7026	5,23	21443	a.4	0,42%
<b>20905417</b>	conserved protein.			A	0.645071	0.3884	-2,4792	4,8	12269	c.55	0,42%
20906505	Hypothetical protein MM1967.			A	0.667214	0.4217	-0,4015	6,62	28659	c.55	0,42%
<b>20905396</b>	Coenzyme F420 hydrogenase beta subunit (EC 1.12.99.1).			A	0.690724	0.4215	0,2907	7,31	37194	d.58	0,41%
<b>20907708</b>	Coenzyme F420 hydrogenase, delta subunit (EC 1.12.99.1).			A	0.220293	0.4691	-8,0246	4,24	17734	c.56	0,41%
20904330	Dipeptide ABC transporter, ATP-binding protein.	DPPF	Amino acid transport and metabolism genes	B	0.770927	0.4039	1,9704	9,02	22741	c.37	0,41%
20905345	Agmatinase (EC 3.5.3.11).		Amino acid transport and metabolism genes	A	0.758841	0.4467	-7,56	4,31	32335	c.42	0,41%
20905578	UDP-glucose 6-dehydrogenase (EC 1.1.1.22).			A	0.765484	0.4436	-1,6786	5,15	45605	c.26, a.100, c.2	0,41%
20905115	Fructose-bisphosphate aldolase (EC 4.1.2.13).		Carbohydrate transport and metabolism genes	A	0.748102	0.4304	-5,1779	4,77	34055	c.1	0,41%
20904640	Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).	LEUS	Translation genes	A	0.770927	0.4253	-2,4716	5,32	110957	c.26, a.27, b.51	0,41%
20905338	Conserved protein.			A	0.501014	0.4502	-0,9645	6,36	34476	x	0,41%
20906877	Hypothetical UPF0173 metal-dependent hydrolase MM2300.			B	0.754257	0.4478	-4,7825	5,1	24933	d.157	0,40%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20907157	Acetyltransferase (EC 2.3.1.-).			B	0.65633	0.3871	-3,6865	5,04	25671	d.108	0,40%
20905416	conserved protein.			A	0.682655	0.3719	-5,785	5,1	13025	c.55	0,40%
20906904	Ech Hydrogenase, Subunit.	ECHF		A	0.761782	0.4286	7,1429	9,13	14014	d.58	0,40%
20906655	conserved protein.			A	0.771936	0.416	0,3817	7,72	30683	x	0,39%
20906402	LPPG:FO 2-phospho-L-lactate transferase (EC 2.7.1.-).	COFD		A	0.736632	0.4422	-5,6105	4,49	33357	x	0,39%
20904643	Conserved protein.			A	0.387076	0.4352	-0,4629	6,55	24843	x	0,39%
20905612	UDP-N-acetylglucosamine 4-epimerase.			A	0.767303	0.4012	-2,3951	5,2	36853	c.2	0,39%
20907454	N-(5'-phosphoribosyl)anthranilate isomerase 1 (EC 5.3.1.24) (PRAI 1).	TRPF1		A	0.768004	0.4658	-2,564	5,01	25506	c.1	0,39%
20907086	F420H2 dehydrogenase subunit FpoB.	FPOB		A	0.762579	0.413	0	7,06	20732	e.19	0,38%
20904852	Putative inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205).		General function prediction only genes	A	0.756163	0.4675	-4,7336	4,73	18926	d.37, d.37	0,38%
<b>20907125</b>	Hypothetical protein MM2525.			A	0.58989	0.4024	-4,0649	5,01	27663	a.4	0,38%
20905318	Conserved protein.		Transcription genes	A	0.479189	0.4368	-1,5325	6	30471	a.4	0,37%
20904462	Universal stress protein.		Signal transduction mechanisms genes	A	0.763365	0.4527	-2,0269	5,41	16296	c.26	0,37%
<b>20905258</b>	Hypothetical protein MM0844.		Transcription genes	B	0.220293	0.3511	-5,3434	4,74	14433	a.4, g.39	0,37%
20906947	GTP-binding protein.			A	0.774113	0.426	-0,8968	6,35	24560	c.37	0,37%
20904725	Sulfite reductase, assimilatory-type (EC 1.8.-.-).		Energy production and conversion genes	B	0.763421	0.4545	1,2987	8,53	24912	d.134, d.58	0,36%
<b>20907433</b>	conserved protein.			A	0.66689	0.3846	-3,5502	4,96	18714	x	0,36%
20905506	Two-component response regulator.			A	0.770727	0.4797	-7,317	4,69	14175	c.23	0,36%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20905077	Conserved protein.		Function unknown genes	A	0.710184	0.3655	-7,1065	4,58	22214	x	0,36%
20906820	Conserved protein.			B	0.220293	0.4151	0,9434	7,65	11810	x	0,36%
20905526	Metal dependent hydrolase.			A	0.741609	0.4009	-4,9549	5,05	24425	d.157	0,36%
<b>20905516</b>	conserved protein.			B	0.747624	0.4456	-1,9373	5,5	73552	c.55, c.55	0,36%
20907459	Indole-3-glycerol phosphate synthase (EC 4.1.1.48).			B	0.765223	0.4682	-4,4943	4,84	29395	c.1	0,35%
20906586	Conserved protein.			A	0.220293	0.4074	3,7037	9	11891	x	0,35%
20906380	Archaeal flavoprotein.			A	0.741419	0.4142	-1,6735	5,42	26491	c.34	0,34%
20906917	Translation initiation factor 1A-2 (aIF-1A-2).	EIF1A2		A	0.674859	0.3774	9,434	10,74	12288	b.40	0,34%
20906216	conserved protein.			A	0.751544	0.5	-4,5454	4,78	17553	x	0,34%
20906572	Hypothetical protein MM2026.			A	0	0.3795	-1,0255	5,39	21559	x	0,34%
<b>20906288</b>	Hypothetical protein MM1771.			A	0.56294	0.4382	8,9888	10,94	9972	x	0,34%
<b>20905840</b>	Ribonucleoside-triphosphate reductase activating enzyme.			A	0.757979	0.4183	-2,2813	5,31	29249	x	0,34%
<b>20906954</b>	GTP-binding protein.			A	0.770927	0.4438	-2,9585	5,01	38289	c.37	0,34%
20905410	Iron-sulfur binding protein.			A	0.677115	0.3803	-1,4084	5,62	15992	x	0,33%
20905158	Conserved protein.		General function prediction only genes	A	0.762402	0.4156	-6,0605	4,5	25460	c.26	0,33%
<b>20906724</b>	Hydrogenase expression/formation protein.	HYPC		A	0.736329	0.4302	-13,9534	4	9384	b.40	0,33%
20906635	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase) (OMPDCase) (OMPdecase).	PYRF		A	0.764643	0.4955	-3,1817	4,79	23503	c.1	0,33%
<b>20905517</b>	conserved protein.			B	0.747624	0.4424	-0,1557	6,91	70382	c.55	0,33%
20905543	Archaeosine tRNA-ribosyltransferase (EC 2.4.2.-).			A	0.767381	0.4367	-1,2244	6,06	55032	d.17, c.1	0,33%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20907236	LSU ribosomal protein L37AE.			A	0.746644	0.3404	18,0851	11,13	10447	g.41	0,33%
20906757	DNA double-strand break repair rad50 ATPase.	RAD50		A	0.766547	0.3701	-2,5233	5,23	121981	a.2, c.37, e.10	0,32%
20905812	Pyruvate synthase delta subunit (EC 1.2.7.1).			A	0.76101	0.3	0,9091	7,37	12595	d.58	0,32%
20905675	Fe-S oxidoreductase.			A	0.706853	0.4133	-0,867	5,57	38575	x	0,31%
20907745	Putative transcriptional regulator.	SYRB		A	0.722586	0.4392	0,6757	9,09	15986	c.26	0,31%
<b>20907005</b>	Type II DNA topoisomerase VI subunit B (EC 5.99.1.3) (TopoVI-B).	TOP6B		B	0.691206	0.4477	-0,9661	6,26	68760	d.14, a.156, d.122	0,31%
<b>20905141</b>	conserved protein.		Function unknown genes	A	0.673186	0.3701	-2,3621	5,09	13553	x	0,31%
<b>20906844</b>	DNA-directed RNA polymerase beta chain (EC 2.7.7.6).			A	0.770899	0.4023	-2,6489	5,6	67023	e.29	0,31%
20904947	Hypothetical protein MM0562.			A	0.7327	0.4353	-0,862	5,96	26949	x	0,31%
20906152	conserved protein.			A	0.479189	0.4258	-5,4687	4,65	29854	a.4	0,30%
20905604	conserved protein.			B	0.55017	0.3822	-0,2873	6,8	41066	c.6	0,30%
20905036	Conserved protein.			A	0.220293	0.3953	-4,054	4,87	33303	x	0,30%
<b>20906988</b>	conserved protein.			A	0.241334	0.3676	-7,3528	4,3	7615	x	0,30%
20905791	Chemotaxis protein.	CHEC		A	0.720472	0.4528	-5,1886	4,42	23371	x	0,30%
20906121	Divalent cation transport protein.			A	0.767248	0.4075	-2,4128	5,24	41395	c.37	0,30%
<b>20905197</b>	Geranyltransterferase (EC 2.5.1.-).		Coenzyme transport and metabolism genes	A	0.768678	0.4271	-5,7626	4,76	32747	a.128	0,30%
20906869	Type I restriction-modification system specificity subunit.			A	0.768421	0.4097	-2,4751	5,39	91505	c.66	0,29%
20905894	Conserved protein.			A	0.501014	0.4342	-1,9736	5,74	16878	c.115	0,29%
<b>20905283</b>	Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase) (MetRS).	METG	Translation genes	A	0.770927	0.4139	-3,0707	5,5	79964	b.40, c.26, a.27, g.41	0,28%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20906966	Conserved protein.			A	0.220293	0.406	-2,2555	5,32	14935	x	0,28%
<b>20906328</b>	Protein translation initiation factor 2 subunit alpha.			A	0.768752	0.3985	0,369	7,95	30416	b.40	0,28%
20905603	NDP-N-acetyl-D-galactosaminuronic acid dehydrogenase (EC 1.1.1.-).			A	0.765484	0.4414	-0,8367	6,13	51826	c.2, c.26, a.100	0,28%
20907183	Hypothetical protein MM2577.			A	0.718043	0.4505	4,3956	10,01	10038	b.34	0,28%
20905845	Archaeal protein translation initiation factor 2B subunit 1.			A	0.74415	0.4766	-3,5087	4,69	37371	x	0,27%
20905403	Conserved protein.			A	0.603238	0.419	-1,6759	5,65	20296	x	0,27%
20905196	Metallo cofactor biosynthesis protein.		General function prediction only genes	B	0.753541	0.4502	-1,7412	5,66	44810	d.115	0,27%
<b>20907953</b>	Coenzyme F420 hydrogenase beta subunit (EC 1.12.99.1).			A	0.766801	0.3818	1,0135	7,69	32412	d.58, d.134, d.58	0,26%
20906712	Cytidylate kinase (EC 2.7.4.14) (CK) (Cytidine monophosphate kinase) (CMP kinase).	CMK		A	0.661161	0.392	-1,1363	5,73	20587	c.37	0,26%
20905667	Aspartate carbamoyltransferase regulatory chain.	PYRI		A	0.733567	0.4231	-0,6409	6,36	17011	g.41, d.58	0,26%
<b>20905876</b>	DNA-directed RNA polymerase subunit M (EC 2.7.7.6).			A	0.67982	0.3271	0,9346	7,5	12259	g.41, g.41	0,26%
<b>20905061</b>	Putative nickel-responsive regulator 1.		Transcription genes	A	0.725492	0.3699	-6,1643	5,05	16601	a.43	0,26%
20904675	Valyl-tRNA synthetase (EC 6.1.1.9).		Translation genes	A	0.770927	0.4304	-3,1069	5,05	99130	b.51, c.26, a.27	0,26%
20907577	Hypothetical protein MM2928.			B	0.764291	0.3256	-1,7441	5,43	19191	x	0,25%
20907082	F420H2 dehydrogenase subunit.	FPOI		A	0.739277	0.4011	1,6949	7,85	19459	d.58	0,25%
20908006	Transcriptional regulator, AraC family.			A	0.722475	0.4545	-2,3922	5,14	23172	c.23	0,25%
<b>20904970</b>	conserved protein.			A	0.386683	0.4053	-0,7575	6,26	29731	x	0,25%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20907008</b>	DNA gyrase, subunit B (EC 5.99.1.3).			A	0.774877	0.3864	-1,7349	5,76	71068	d.14,d.122, e.11	0,25%
20907062	Metallo cofactor biosynthesis protein.			A	0.760958	0.4147	0	7,09	24558	x	0,24%
<b>20905038</b>	3-isopropylmalate dehydratase small subunit 1 (EC 4.2.1.33) (Isopropylmalate isomerase 1) (Alpha-IPM isomerase 1)	LEUD1	Amino acid transport and metabolism genes	A	0.763699	0.4438	-1,1833	5,92	18489	c.8	0,24%
<b>20904800</b>	conserved protein.		Translation genes	A	0.710122	0.413	-2,3598	5,9	37434	d.157	0,24%
20907018	Methylthiol:coenzyme M methyltransferase.	MTSB		A	0.763392	0.4436	-5,0908	4,68	29878	a.46, c.23, c.1	0,24%
20907455	Anthranilate phosphoribosyltransferase (EC 2.4.2.18).	TRPD		A	0.763212	0.4405	0,5405	8,2	39203	a.46, c.27	0,24%
<b>20908004</b>	conserved protein.			B	0.358138	0.4184	-4,6024	5,17	27919	c.69	0,24%
20905755	(S)-2-hydroxy-acid dehydrogenase (EC 1.1.3.15).			A	0.767026	0.4587	-1,3042	5,52	49551	d.58, d.145	0,24%
20905408	CoB--CoM heterodisulfide reductase 1 iron-sulfur subunit C (EC 1.8.98.1).	HDRC		A	0.752359	0.3851	1,2422	7,74	18120	a.1	0,24%
<b>20905088</b>	conserved protein.		Translation genes	A	0.65145	0.4266	-2,8248	5,11	39681	x	0,24%
20905924	Hypothetical protein MM1442.			B	0.220293	0.4318	-2,2726	4,9	28289	c.5	0,23%
20908029	Conserved protein (EC 1.3.-.-).			A	0.765262	0.4215	-2,2726	5,41	54567	c.3	0,23%
20905482	GTP-binding protein.			A	0.770927	0.3901	-0,5494	6,64	40758	c.37	0,23%
20905833	ABC transporter, ATP-binding protein.			A	0.770927	0.4144	-0,3801	6,64	29748	c.37	0,23%
20905704	conserved protein.			A	0.638281	0.4331	-2,9196	4,93	45077	g.38	0,23%
20906041	Conserved protein.			A	0.714573	0.3978	-4,8326	4,99	29514	c.1	0,23%
20907621	Isoleucyl-tRNA synthetase (EC 6.1.1.5).			A	0.220293	0.4319	-2,93	5,1	120411	c.26, a.27, b.51	0,22%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20905409</b>	CoB--CoM heterodisulfide reductase 1 subunit B (EC 1.8.98.1).	HDRB		A	0.742551	0.4135	-3,205	5,05	33068	x	0,22%
20905054	Mannose-6-phosphate isomerase (EC 5.3.1.8) (EC 2.7.7.22).		Cell wall/membrane biogenesis genes	A	0.763988	0.4092	-4,1378	4,99	48988	b.82, c.68	0,22%
<b>20905863</b>	Hypothetical protein MM1386.			A	0.520626	0.3676	-5,3921	4,78	23051	x	0,22%
20905469	Amino-acid acetyltransferase (EC 2.3.1.1).			B	0.760691	0.4403	0	6,75	18041	d.108	0,22%
20905700	Methyl-coenzyme M reductase operon protein C.			A	0.501014	0.4223	1,4563	8,59	22427	x	0,22%
20907682	ABC transporter, ATP-binding protein.			A	0.762626	0.411	-1,2711	5,82	26237	c.37	0,21%
20904460	Universal stress protein.		Signal transduction mechanisms genes	A	0.760552	0.451	-1,9607	5,33	16447	c.26	0,21%
20906241	Heme biosynthesis protein.	NIRH		A	0.725238	0.4114	-0,6328	6,51	18042	a.4, d.58	0,21%
20905901	F420-0:gamma-glutamyl ligase (EC 6.3.2.-).	COFE		A	0.732841	0.4449	-2,7558	4,95	27741	x	0,21%
20905941	Conserved protein.			A	0.726229	0.4269	-1,1791	5,47	47156	x	0,21%
20905668	Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate transcarbamylase) (ATCase).	PYRB		A	0.76727	0.4434	-0,9708	6,3	35096	c, c.78.78	0,21%
20905621	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14).			A	0.765718	0.4085	-2,8168	5,3	39514	c.87	0,21%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20904518	Methanol corrinoid protein.	MTAC	General function prediction only genes	A	0.765737	0.4654	-6,923	4,49	28549	c.23	0,21%
20905334	conserved protein.			A	0.501014	0.415	-1,3604	5,76	16496	x	0,20%
20905405	F420-dependent NADP reductase (EC 1.6.8.-).			A	0.665591	0.4809	-1,2765	5,84	25188	c.2	0,20%
20906506	Hypothetical protein MM1968.			A	0.604417	0.439	-6,0975	4,62	28035	b.82	0,20%
<b>20906963</b>	5'-methylthioadenosine phosphorylase (EC 2.4.2.28).			A	0.757515	0.4109	-1,1627	6,19	28545	c.56	0,20%
<b>20907529</b>	Putative metalloendopeptidases.			A	0.686473	0.393	-2,985	5,19	22832	b.29	0,20%
20904599	Serine-pyruvate aminotransferase (EC 2.6.1.51).		Amino acid transport and metabolism genes	A	0.759611	0.455	-0,7711	6,14	42359	c.67	0,20%
20907557	Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).	SERC		A	0.220293	0.4324	-1,6215	5,49	41596	c.67	0,19%
20906215	Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase)	HISF		A	0.764654	0.4212	-5,4944	4,54	29691	c.1	0,19%
20906956	Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) (PRA-PH).	HISE		A	0.763686	0.3925	-6,542	4,9	12150	x	0,19%
20906067	conserved protein.			A	0.489556	0.406	-0,5969	6,44	38311	d.108	0,19%
<b>20906340</b>	conserved protein.			A	0.613732	0.4296	-3,78	5,16	31647	x	0,19%
<b>20905389</b>	conserved protein.			A	0.220293	0.3333	-7,9999	5,29	8604	x	0,19%
20904811	Methylenetetrahydrofolate reductase (EC 1.5.1.20).		Amino acid transport and metabolism genes	B	0.761607	0.4384	-1,3698	5,87	31927	c.1	0,19%
<b>20907381</b>	conserved protein.			A	0.641863	0.3796	0,5442	8,12	83634	x	0,19%
20904993	Conserved protein.			A	0.241334	0.395	-0,8402	6,15	13497	x	0,19%
20905363	Putative flagella related protein H.			A	0.714888	0.395	-2,1007	5,19	26637	c.37	0,18%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20906412	Conserved protein.			B	0.624778	0.4511	-1,0869	5,43	19947	x	0,18%
20908007	Acetyltransferase (EC 2.3.1.-).			A	0.760626	0.3857	-1,4285	6,1	24393	d.108	0,18%
20907457	Tryptophan synthase beta chain 1 (EC 4.2.1.20).	TRPB1		A	0.765011	0.4094	-2,3621	5,53	43753	c.79	0,18%
<b>20906960</b>	conserved protein.			A	0.220293	0.3846	1,2146	8,4	27903	c.108	0,18%
20905353	Nitrogen regulatory protein P-II.		Amino acid transport and metabolism genes	B	0.7761	0.3917	-1,6666	5,27	12921	d.58	0,18%
20905415	conserved protein.			A	0.693806	0.4683	-3,521	4,73	30760	c.37	0,18%
20906280	Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase) (Isopentenyl pyrophosphate isomerase).	FNI		A	0.746633	0.4685	-4,9314	4,57	38789	c.1	0,18%
20905164	NifB protein.	NIFB	General function prediction only genes	A	0.758838	0.4179	-2,388	5,43	37767	x	0,18%
20904821	Cell division control protein.		Posttranslational modification, protein turnover, chaperones genes	A	0.770927	0.4432	-3,7878	4,9	88391	c.37, d.31, c.37, b.52	0,17%
<b>20904948</b>	Hypothetical protein MM0563.		Replication, recombination and repair genes	B	0.764293	0.3746	-1,0308	6,01	33339	x	0,17%
<b>20905464</b>	Chromosome partition protein.			A	0.769171	0.3855	-3,4042	4,94	133838	d.215, c.37	0,17%
20905335	conserved protein.			A	0.501014	0.4412	-4,7058	4,66	19010	g.41	0,17%
20905059	Glutamate synthase, large chain (EC 1.4.1.13).			B	0.766164	0.4158	-1,7057	5,39	50566	x	0,17%
<b>20907435</b>	Hypothetical protein MM2802.			A	0.552797	0.3665	-4,7119	4,64	21861	a.118	0,17%
<b>20905897</b>	Aspartate kinase (EC 2.7.2.4).			A	0.767979	0.4447	-0,835	6,36	51914	c.73	0,17%
20908039	Hypothetical protein MM3340.			B	0.765737	0.461	-6,4406	4,46	31481	x	0,17%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20906857</b>	DNA/pantothenate metabolism flavoprotein.			A	0.766541	0.412	-1,502	5,88	50459	c.34, c.5	0,17%
20905194	A1AO H+ ATPase subunit H.		Energy production and conversion genes	A	0.625384	0.4	-5,5555	4,74	10157	x	0,17%
<b>20906821</b>	Tetrahydromethanopterin S-methyltransferase, subunit A (EC 2.1.1.86).			A	0.568873	0.4286	-3,2966	4,79	19646	x	0,17%
<b>20904378</b>	conserved protein.		Signal transduction mechanisms genes	A	0.770079	0.4198	-6,1068	4,41	14776	c.23	0,16%
<b>20905796</b>	Chemotaxis protein.	CHEY		A	0.770725	0.5106	-2,8368	5,24	15791	c.23	0,16%
20905906	Replication factor C subunit.			A	0.769859	0.4252	-2,9325	4,96	38349	a.80, c.37	0,16%
<b>20905452</b>	Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63) (6-O- methylguanine-DNA methyltransferase) (MGMT) (O-6-methylguanine-DNA-alkyltransferase).	OGT		B	0.767784	0.4241	4,4304	9,57	17996	a.4, c.55	0,16%
20904798	Rnase L inhibitor.		General function prediction only genes	A	0.770927	0.4082	-1,7006	5,58	65396	c.37, c.37, d.58	0,16%
<b>20906853</b>	Chlorohydrolase family protein (EC 3.8.1.-).			A	0.759931	0.4311	-3,2822	5,28	50014	b.92, c.1	0,15%
<b>20907177</b>	Methyltransferase (EC 2.1.1.-).			A	0.765131	0.4216	-4,4117	4,73	45591	a.4, c.66	0,15%
<b>20907204</b>	Hypothetical protein MM2596.			A	0.75597	0.4316	-1,9548	5,37	73324	c.55	0,15%
20906361	RNA methylase.			A	0.761657	0.4208	-0,7721	6,27	28514	c.116	0,15%
20906220	conserved protein.			A	0.658251	0.3778	0,4444	8,42	25317	a.4	0,15%
20904985	DNA-directed RNA polymerase subunit E' (EC 2.7.7.6).		Transcription genes	A	0.648493	0.4262	-1,6392	5,63	6881	x	0,15%
20907135	Hypothetical UPF0217 protein MM2534.			A	0.582928	0.4028	-3,7914	5,27	23249	x	0,15%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20906864	Hypothetical protein MM2289.			B	0.753787	0.4533	-5,1401	5,25	23218	c.6	0,15%
20907094	Hypothetical protein MM2497.			A	0.761164	0.4279	-4,3268	4,8	23051	x	0,15%
<b>20905322</b>	Putative methyltransferase.		Secondary metabolites biosynthesis, transport and catabolism genes	A	0.742501	0.405	-3,2257	5,04	32193	c.66	0,15%
20905045	Glycosyltransferase (EC 2.4.-.-).		Cell wall/membrane biogenesis genes	A	0.76637	0.3779	2,6059	9,27	35839	c.68	0,14%
20905831	Putative methyltransferase.			A	0.71642	0.4412	-7,3528	4,35	7543	x	0,14%
20905193	V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).	AHAI,ATPI	Energy production and conversion genes	A	0.736091	0.4702	-2,29	5,6	71964	x	0,14%
<b>20906377</b>	conserved protein.			A	0.666506	0.4359	-5,1281	4,72	38520	x	0,14%
<b>20907690</b>	Deoxycytidylate deaminase (EC 3.5.4.12).			A	0.770927	0.4128	-2,9069	5,02	18831	c.97	0,14%
20905931	Hypothetical UPF0251 protein MM1448.			A	0.674599	0.3559	2,8249	8,61	18017	a.4	0,14%
20905780	Hypothetical UPF0200 protein MM1313.			A	0.662347	0.4171	-1,6042	5,46	20893	c.37	0,13%
20906126	Conserved protein.			A	0.712044	0.426	-0,9062	5,73	36119	x	0,13%
<b>20906260</b>	Probable dihydroorotate dehydrogenase electron transfer subunit.	PYRK		A	0.755612	0.4324	-1,5443	5,44	28187	b.43, c.25	0,13%
<b>20907483</b>	Hypothetical protein MM2845.			A	0.764627	0.3955	-3,7312	4,64	15133	b.85	0,13%
20907593	conserved protein.			B	0.722121	0.4211	-4,3859	4,62	12674	x	0,13%
20907866	Ornithine decarboxylase (EC 4.1.1.17)			B	0.709727	0.4154	-0,7691	6,46	44775	b.49, c.1	0,13%
<b>20904984</b>	DNA-directed RNA polymerase subunit E' (EC 2.7.7.6).		Transcription genes	A	0.736541	0.4381	-4,1236	4,84	21432	b.40, d.230	0,13%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20905524</b>	Hypothetical protein MM1084.			A	0.489393	0.3697	3,6364	9,23	18292	g.41	0,13%
<b>20904454</b>	CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).	PYRG	Nucleotide transport and metabolism genes	A	0.770553	0.4026	-2,8089	5,33	59718	c.23, c.37	0,13%
20906063	Molybdenum formylmethanofuran dehydrogenase subunit (EC 1.2.99.5).	FMDB		A	0.751098	0.4345	-1,839	5,78	47740	c.81	0,12%
<b>20905271</b>	conserved protein.		Function unknown genes	A	0.648582	0.4706	0	7,18	45552	x	0,12%
20904956	ABC transporter, ATP-binding protein.		Inorganic ion transport and metabolism genes	B	0.770927	0.4301	-2,2058	5,64	30186	c.37	0,12%
<b>20906920</b>	Translation initiation factor 1A.	EIF-1A		A	0.674859	0.4237	5,0847	9,96	7017	b.40	0,12%
20907790	Transcriptional regulator, MarR family.			A	0.606889	0.3813	2,1583	8,41	16105	a.4	0,12%
<b>20906059</b>	Molybdenum formylmethanofuran dehydrogenase subunit (EC 1.2.99.5).	FMDF		A	0.764086	0.3931	-2,6011	5,11	37380	d.58, d.58	0,12%
<b>20908026</b>	conserved protein.			A	0.763817	0.406	-3,4187	5,36	27156	c.23	0,11%
20904632	Superfamily II DNA and RNA helicase.		Replication, recombination and repair genes	A	0.755825	0.3929	-1,5685	5,46	144748	c.37	0,11%
<b>20905412</b>	conserved protein.			A	0.640527	0.3911	-1,9801	5,07	22879	x	0,11%
<b>20905263</b>	Bacterioferritin comigratory protein.		Posttranslational modification, protein turnover, chaperones genes	A	0.768702	0.375	0,625	7,98	18168	c.47	0,11%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20905167</b>	conserved protein.		Defense mechanisms genes	A	0.241334	0.46	-9,9999	4,26	11546	x	0,11%
20905817	Fe-S oxidoreductase.			A	0.677452	0.411	0,274	7,53	41968	x	0,11%
<b>20905076</b>	Hypothetical UPF0285 protein MM0679.		Function unknown genes	A	0.501014	0.4492	-3,0768	5,33	34772	c.55	0,11%
20904385	CoB--CoM heterodisulfide reductase 1 iron-sulfur subunit A (EC 1.8.98.1).	HDRA	Energy production and conversion genes	A	0.743471	0.425	-2,3959	5,37	86803	d.58, c.3	0,11%
20907312	Hypothetical protein MM2692.			B	0.617192	0.3996	-1,7278	5,16	52553	c.3	0,11%
<b>20906761</b>	conserved protein.			A	0.220293	0.317	1,1321	8,07	30920	x	0,11%
<b>20908027</b>	Phosphoglycolate phosphatase (EC 3.1.3.18).			A	0.762592	0.4224	-1,724	5,85	26041	c.108	0,10%
20906978	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3).			A	0.770339	0.4257	-2,0182	5,78	61380	e.23	0,10%
<b>20907410</b>	Hypothetical protein MM2780.			A	0.746342	0.4403	-5,4607	4,6	32588	x	0,10%
20906700	LSU ribosomal protein L19E.			A	0.67914	0.3399	16,3399	11,04	17334	a.94	0,10%
<b>20906082</b>	Transcriptional regulator, MarR family.			A	0.764996	0.3984	2,3438	8,82	14584	a.4	0,10%
20905436	Zinc finger protein.			A	0.674353	0.3602	2,4845	8,88	18477	a.35	0,10%
<b>20905956</b>	Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase) (TrpRS).	TRPS		A	0.68265	0.4121	-2,0201	5,71	55728	c.26	0,10%
<b>20906786</b>	Hypothetical protein MM2219.			A	0.764126	0.3897	3,0769	8,96	22233	c.66	0,10%
<b>20907921</b>	Thiamine-phosphate pyrophosphorylase (EC 2.5.1.3) (TMP pyrophosphorylase) (TMP-PPase) (Thiamine-phosphate synthase).	THIE		A	0.655175	0.3708	-0,8332	5,98	25817	c.1	0,10%
<b>20906558</b>	Translation initiation factor 1A-1 (aIF-1A-1).	EIF1A1		A	0.674859	0.4286	8,7912	10,71	12798	b.40	0,09%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20907817</b>	Type I restriction-modification system restriction subunit (EC 3.1.21.3).			A	0.220293	0.4092	-1,9936	5,43	110297	c.37, c.37	0,09%
<b>20905391</b>	Glutamine synthetase (EC 6.3.1.2).			A	0.766261	0.4295	-2,237	5,59	50593	d.15, d.128	0,09%
20905178	Conserved protein.			A	0.220293	0.437	1,4815	9,05	15618	x	0,09%
20906591	Hypothetical protein MM2043.			A	0.627596	0.4477	-2,166	5,65	32631	x	0,09%
<b>20907199</b>	Hypothetical protein MM2592.			A	0.746459	0.4004	-3,5299	4,94	213324	c.37	0,09%
20907093	FO synthase subunit 2 2 (EC 2.5.1.-) (7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase subunit 2 2).	COFH2		A	0.761565	0.4031	-1,0203	5,94	43094	c.1	0,09%
<b>20904794</b>	conserved protein.		Replication, recombination and repair genes	A	0.220293	0.3944	0,7042	8,5	16554	x	0,08%
20904862	Hypothetical protein MM0484.		Function unknown genes	A	0.711894	0.4271	-5,025	4,71	21907	x	0,08%
20907629	Hypothetical protein MM2974.			B	0.770927	0.4147	-7,834	4,38	25163	x	0,08%
<b>20907692</b>	conserved protein.			A	0.773725	0.3626	1,0989	7,72	21425	x	0,08%
20907771	Hypothetical protein MM3100.			A	0.760653	0.3808	-1,6735	5,77	28102	x	0,08%
20904986	Hypothetical UPF0218 protein MM0598.		Function unknown genes	A	0.641997	0.4308	-1,5384	5,66	21900	x	0,07%
20904471	Threonine synthase (EC 4.2.99.2).		Amino acid transport and metabolism genes	A	0.742634	0.4327	-3,6057	4,75	45396	c.79	0,07%
<b>20907188</b>	Hypothetical protein MM2582.			A	0.656604	0.4059	-2,9411	5,52	19676	d.113	0,07%
<b>20905880</b>	Phosphatidylserine decarboxylase proenzyme (EC 4.1.1.65)	PSD		A	0.753647	0.4519	1,4423	9,07	23416	b.84	0,07%
20908032	Monomethylamine corrinoid protein 2 (MMCP 2).	MTMC2		A	0.762771	0.4633	-5,9632	4,42	23270	c.23, a.46	0,06%
20905368	Protein pcrB homolog.	PCRB		A	0.701676	0.4777	-2,0242	5,31	25888	c.1	0,06%
<b>20907871</b>	Hypothetical UPF0204 protein MM3190.			A	0.767731	0.4108	-3,0302	5,49	32782	x	0,06%

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
20904645	Probable deoxyhypusine synthase 2 (EC 2.5.1.46) (DHS 2).	DYS2	Posttranslational modification, protein turnover, chaperones genes	A	0.710433	0.4298	-3,7248	5,38	38042	c.31	0,06%
<b>20904373</b>	Hypothetical ANK-repeat protein MM0045.		General function prediction only genes	B	0.775062	0.4167	-4,1666	4,58	39044	d.211, d.211	0,06%
20907645	conserved protein.			A	0.519884	0.4393	-0,6556	6,36	34765	c.37	0,06%
20904486	Iron-sulfur flavoprotein.		General function prediction only genes	A	0.703765	0.4025	0,8475	7,91	26247	c.23	0,06%
20906666	Conserved protein.			B	0.625699	0.39	-0,5864	6,64	39363	c.87	0,05%
<b>20905402</b>	Xaa-Pro aminopeptidase (EC 3.4.11.9).			B	0.770471	0.4275	-1,9999	5,41	44634	c.55, d.127	0,05%
20908010	N-(5'-phosphoribosyl)anthranilate isomerase 2 (EC 5.3.1.24) (PRAI 2).	TRPF2		A	0.76462	0.4466	0,3953	7,88	24760	c.1	0,05%
20904497	Type IIS restriction enzyme (EC 3.1.21.4) (EC 2.1.1.72).		Defense mechanisms genes	B	0.685957	0.4137	-2,9184	5,16	134098	c.66	0,05%
20904339	Methyltransferase (EC 2.1.1.-).		Secondary metabolites biosynthesis, transport and catabolism genes	A	0.765499	0.44	-1,9999	5,57	22759	c.66	0,05%
20906239	Heme biosynthesis protein.	NIRJ		A	0.761123	0.4	-1,4084	5,95	39379	x	0,04%
20907564	Oxidoreductase, ALDO/KETO reductase family.			A	0.759299	0.3945	-0,5024	6,44	45114	a.1, c.1	0,04%
<b>20906609</b>	Histidinol-phosphate aminotransferase.			A	0.768389	0.4194	-3,4273	5,03	55532	c.67	0,04%
20905047	Glycosyltransferase involved in cell wall biogenesis (EC 2.4.-.-).		Cell wall/membrane biogenesis genes	B	0.764405	0.3871	0	6,9	29073	c.68	0,04%
20905871	Probable molybdenum cofactor biosynthesis protein A.	MOAA		A	0.761142	0.3952	1,1976	8,06	37691	x	0,04%
20904389	Tungsten formylmethanofuran dehydrogenase subunit B (EC 1.2.99.5).		Energy production and conversion genes	A	0.57674	0.4251	-0,7245	6,04	45473	c.81	0,04%

---

gi number	name	genname	function	Phylogeny	evolutionary scope	hydrophobicity	net charge	pl	MW	scop code	on Ths
<b>20906060</b>	Molybdenum formylmethanofuran dehydrogenase subunit (EC 1.2.99.5).	FMDA		A	0.71511	0.4007	-1,8835	5,92	65094	b.92	0,03%
<b>20906574</b>	Hypothetical protein MM2028.			A	0.605772	0.3926	-0,7991	6,25	114238	a.118, d.160	0,02%

### VII.3 Abbreviations

Units are expressed according to the international system of units (SI), including outside units accepted for use with the SI.

Amino acids are abbreviated with their one or three letter symbols.

Protein names are abbreviated according to their SWISSPROT database entries.

ADP	adenosine 5'-diphosphate
Amp	ampicillin
APS	ammonium peroxodisulfate
ATP	adenosine 5'-triphosphate
BLAST	Basic Local Alignment Search Tool
$\beta$ -NADH	$\beta$ -nicotinamide adenine dinucleotide
BSA	albumin bovine serum
COG	Clusters of Orthologous Groups of proteins
DIGE	Differentiated Gel Electrophoresis
DNA	deoxyribonucleic acid
DnaJ	bacterial Hsp40 chaperone
DnaK	bacterial Hsp70 chaperone
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
emPAI	exponentially modified ( $\rightarrow$ ) PAI
FAD	flavine adenine dinucleotide
FPLC	Fast Protein Liquid Chromatography
g	acceleration of gravity, 9.81 m/s <sup>2</sup>
GdnHCl	guanidinium hydrochloride
GimC	Genes involved in microtubule biogenesis C
GroEL	bacterial Hsp60 chaperonin
GroES	bacterial Hsp10 cochaperonin

---

GTP	guanosin 5'-triphosphate
Hsp	heat shock protein
Hr	hour
IPTG	isopropyl- $\beta$ -D-1-thiogalactopyranoside
LB	Luria Bertani
LCMS	coupled Liquid Chromatography - Mass Spectrometry
MES	2-morpholinoethanesulfonic acid
<i>Mm</i>	<i>Methanosarcina mazei</i> Gö1
<i>M. mazei</i>	<i>Methanosarcina mazei</i> Gö1
<i>Mt</i>	<i>Methanobacterium thermoautotrophicum</i>
MOPS	3-(N-morpholino)propanesulfonic acid
MS	mass spectrometry
NADPH	$\beta$ -nicotinamide adenine dinucleotide 2'-phosphate
OD	optical density
PAGE	PolyAcrylamide Gel Electrophoresis
PAI	Protein Abundance Index
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank. Repository for processing and distribution of 3-D structure data of proteins and nucleic acids. <a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
PEDANT	Protein Extraction, Description and ANalysis Tool
Pefabloc	4-(2-aminoethyl)benzenesulfonylfluoride HCl
SCOP	database of protein structures (Structural Classification of Proteins)
SDS	sodiumdodecylsulfate
<i>Ta</i>	<i>Thermoplasma acidophilum</i>
TCA	trichloroacetic acid
TCP-1	tailless complex polypeptide
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
TF	trigger factor
TFA	trifluoroacetic acid

---

Ths	Thermsome
TOF	time Of Flight. Mass spectrometry ion detector
TRiC	TCP-1 containing ring complex
Tris HCl	tris(hydroxymethyl)aminomethane hydrochloride
Triton X-100	t-Octylphenoxy polyethoxyethanol
Tween 20	Polyoxyethylen-Sorbitan-Monolaurat

---

## VII.4 Publications, Poster und Vorträge

### Vorträge/Poster

- Sep 2004 Posterpräsentation während des GBM Meetings , Münster  
2004  
zum Titel: Analysis of the Co-Existing Group I and Group II  
Chaperonins in the Archaeon *Methanosarcina mazei*
- Dez 2004 Posterpräsentation während der Structural Genomics  
Conference, Barcelona 2004 zum Titel: Functional  
Characterisation of the Group I and II Chaperonins in the  
Archaeon *Methanosarcina mazei*
- Nov 2005 Vortrag im Rahmen des SFB 594 „Molekulare Maschinen“  
zum Thema „Analysis of the co-existing Group I and Group  
II chaperonins in M.mazei“

### Publikationen

- Aug 2003 Klunker D, Haas B, Hirtreiter A, Figueiredo L, Naylor DJ,  
Pfeifer G, Muller V, Deppenmeier U, Gottschalk G, Hartl  
FU, Hayer-Hartl M. “Coexistence of group I and group II  
chaperonins in the archaeon *Methanosarcina mazei*”, J Biol  
Chem. 2003 Jun 9 [Epub ahead of print]

## VII.5 Curriculum vitae

### **Angela Maria Hirtreiter**

Geboren am 20. Mai 1975 in Straubing

### **Wissenschaftlicher Werdegang**

#### **Dissertation**

Okt. 2001– heute      Experimentelle Doktorarbeit in der Abteilung Zelluläre Biochemie von Prof. Dr. F. Ulrich Hartl am Max-Planck-Institut für Biochemie, Martinsried, zum Thema “Protein folding in Archaea: Analysis of the co-existing Group I and Group II chaperonins in *M. mazei*”

#### **Hochschulstudium**

Mai 2001                      Diplom in Biologie  
Okt 2000-Mai 2001      Diplomarbeit unter der Leitung von K.O. Stetter zum Thema: Untersuchungen zur archaeellen Vielfalt in heißen Quellen im Atacama Hochland, Chile  
Juni 2000                      Diplom Hauptprüfung  
Fächer: Mikrobiologie, Genetik, Organische Chemie  
Sep 1995- Mai 2001      Studium der Biologie an der Universität Regensburg

#### **Hochschulreife:**

Juni 1994:                      Allgemeine Hochschulreife, Anton Bruckner Gymnasium, Straubing

## *VII.6 Danksagung*

Mein herzlicher Dank gilt meinem Doktorvater Prof. Dr. F. Ulrich Hartl für die umfangreiche Betreuung dieser Arbeit, die immerwährende Unterstützung, die zahlreiche Diskussionen und essentiellen Anregungen. Des Weiteren geht mein besonderer Dank an Dr. Manajit Hayer-Hartl für die hilfreiche Diskussionen, wertvolle Ratschläge und auch das freundliche Arbeitsklima.

Die Zusammenarbeit mit anderen Arbeitsgruppen war essentiell für das gute Gelingen dieser Arbeit. So danke ich Prof. Joel Vandekerckhove, Magda Pupyte und Evy Timmermann (Universität Gent) für die massenspektrometrische Identifikation der Proteinspots von 2D-PAGE. Prof. Matthias Mann, Dr. Francesca Forner (MPI Martinsried) und Morten Kirkegaard (Universität Odensee) ich sehr dankbar für die äußerst erfolgreiche Kooperation an diesem Projekt und ihre stete Bereitschaft zu weiteren zusätzlichen Analysen. Ganz besonders möchte ich Dr. Burghardt Scheibe (Amersham, München) danken für die freundschaftliche, umfassende und unbedingte Hilfe bei der DIGE.

Mein besonderer Dank gilt Daniel und Luis für die hervorragende Zusammenarbeit und ihre freundliche und hilfsbereite Art. Bei allen Kollegen der Arbeitsgruppe möchte ich mich für die sehr gute Zusammenarbeit, die konstruktiven Diskussionen und das wirklich freundliche Arbeitsklima bedanken. Insbesondere danke ich Grandma, Shrutski, La Astola, Juli, Tobi, Michi, Ulrike, Penny, Dean und Anna, Kautschuk, Giulia, Blueplasticbarral, Kaiserchen und Später-Peter und natürlich Behrenschen.