Characterisation of the RNA Polymerase Subunits E and F from the Antarctic Archaeon Methanococcoides Burtonii

Author:
De Francisci, Davide

Publication Date:
2010

DOI:
https://doi.org/10.26190/unsworks/14946

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **De Francisci**

First name: **Davide** Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **BABS** Faculty: **Science**

Title: **Characterisation of the RNA Polymerase subunits E and F from the Antarctic archaeon *Methanococcoides burtonii***

Abstract 350 words maximum: (PLEASE TYPE)

The archaeal RNA polymerase (RNAP) is composed of eleven or twelve subunits and is closely related eucaryal RNAPs (particularly RNAP II). Nine or ten of these subunits form the core holoenzyme, and the remaining two subunits, E and F, form a dissociable heterodimer whose functional contribution to transcription remains poorly understood. In yeast, the corresponding heterodimer, Rpb4/7, functions in nuclear transcription-coupled transport of specific mRNA species to the cytoplasm, enabling RNA degradation. While the ability of this heterodimer to bind RNA has been demonstrated, whether it can recognise specific RNA targets has not been determined.

In this study the recombinant archaeal heterodimer from the Antarctic archaeon *Methanococcoides bur* (*Mb*RpoE/F) has been characterised by comparison with its thermophilic homologue from *Methanocaldococcus jannaschii* (*Mj*RpoE/F). The structural analysis showed that *Mb*RpoE/F is far less s and more flexible than *Mj*RpoE/F. Crystallising conditions were also found for *Mb*RpoE/F and a complet house 3.2 Å data set was collected, although it was impossible to solve the structure.

Functional comparison of these two heterodimers was performed by electrophoretic mobility shift assa (EMSA), and provided a preliminary indication that the archaeal RpoE/F heterodimers have the capacity bind ssDNA in a sequence-dependent manner.

Following the EMSA findings, it became essential to determine the nucleic-acid binding characteristics *Mb*RpoE/F, with the aim of learning about the nucleic-acid targets it was capable of interacting with. His tagged *Mb*RpoE/F was incubated with whole cell RNA, and complexes were purified using Ni-NTA. The RNA was eluted, labeled and hybridized to a high-density *M. burtonii* microarray, and 118 genes were identified (4% of the total genome). The genes were divided into distinct categories: methanogenesis, nucleotide metabolism, cofactors biosynthesis, transcription, translation, import/export and others; mo importantly, for each category the genes identified appear to code for key regulatory enzymes in the sa category. This suggested that *Mb*RpoE/F is indeed capable of binding RNA in a specific manner and tha polymerase heterodimer also has a regulatory effect within the physiology of the archaeon.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

05/10/10

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed

Date

CHARACTERISATION OF THE RNA POLYMERASE SUBUNITS E AND F FROM THE ANTARCTIC ARCHAEON *METHANOCOCCOIDES BURTONII*

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Davide De Francisci

School of Biotechnology and Biomolecular Sciences
Faculty of Science
The University of New South Wales
Sydney, Australia
2010

Declaration of originality

I, Davide De Francisci, hereby declare that this submission is my own work and to the best of my knowledge contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signature_____ Date / /

Abstract

The archaeal RNA polymerase (RNAP) is composed of eleven or twelve subunits and is closely related to the eucaryal RNAPs (particularly RNAP II). Nine or ten of these subunits form the core holoenzyme, and the remaining two subunits, E and F, form a dissociable heterodimer whose functional contribution to transcription remains poorly understood. In yeast, the corresponding heterodimer, Rpb4/7, functions in nuclear transcription-coupled transport of specific mRNA species to the cytoplasm, enabling RNA degradation. While the ability of this heterodimer to bind RNA has been demonstrated, whether it can recognise specific RNA targets has not been determined.

In this study the recombinant archaeal heterodimer from the Antarctic archaeon *Methanococcoides burtonii* (*Mb*RpoE/F) has been characterised by comparison with its thermophilic homologue from *Methanocaldococcus jannaschii* (*Mj*RpoE/F). The structural analysis showed that *Mb*RpoE/F is far less stable and more flexible than *Mj*RpoE/F. Crystallising conditions were also found for *Mb*RpoE/F and a complete in-house 3.2 Å data set was collected, although it was impossible to solve the structure.

Functional comparison of these two heterodimers was performed by electrophoretic mobility shift assay (EMSA), and provided a preliminary indication that the archaeal RpoE/F heterodimers have the capacity to bind ssDNA in a sequence-dependent manner.

Following the EMSA findings, it became essential to determine the nucleic-acid binding characteristics of *Mb*RpoE/F, with the aim of learning about the nucleic-acid targets it was capable of interacting with. His-tagged *Mb*RpoE/F was incubated with whole cell RNA, and complexes were purified using Ni-NTA. The bound RNA was eluted, labeled and hybridised to a high-density *M. burtonii* microarray, and 118 genes were identified (4% of the total genome). The genes were divided into distinct categories: methanogenesis, nucleotide metabolism, cofactors biosynthesis, transcription, translation, import/export and

others; most importantly, for each category the genes identified appear to code for key regulatory enzymes in the same category. This suggested that *Mb*RpoE/F is indeed capable of binding RNA in a specific manner and that the polymerase heterodimer also has a regulatory effect within the physiology of the archaeon.

Due to the high level of structural and functional similarities between the E/F subunits with their eucaryal homologues, the implication of these findings may be that an important function of the eucaryal and archaeal heterodimers is to directly regulate the abundance of specific classes of cellular mRNA via this specific binding.

List of Publications

Siddiqui K. S., Poljak A., Guilhaus M., De Francisci D., Curmi P. M., Feller G., D'Amico S., Gerday C., Uversky V. N. & Cavicchioli R. 2006. Role of lysine versus arginine in enzyme cold-adaptation: modifying lysine to homo-arginine stabilises the cold adapted alpha-amylase from *Pseudoalteramonas haloplanktis*. *Proteins.* Aug 1;64(2):486-501.

Siddiqui K. S., De Francisci D., & Thomas T. 2005. A novel approach for enhancing the catalytic efficiency of a protease at low temperature: Analysis of unfolding-refolding patterns and stabilities of proteins by transverse urea gradient electrophoresis (TUG-GE). In: Protein structures: Methods in protein structures and stability analysis (eds. Uversky VN. & Permyakon AE.) Nove Science Publishers, Inc. Hauppauge, NY. USA.

Siddiqui K. S., Parkin, D.M., Curmi, P.M.G., De Francisci, D., Poljack, A., Barrow, K., Noble, M. H. Trewhella, J. & Cavicchioli, R. 2009. A novel approach for enhancing the catalytic efficiency of a protease at low temperature: reduction in substrate inhibition by chemical modification. *Biotechnol Bioeng.* 2009 Jul. 1;103(4):676-86.

Allen, M. A., Lauro, F. M., Williams, T. J., Burg, D., Siddiqui, K. S., De Francisci, D., Chong, K. W., Pilak, O., Chew, H. H., De Maere, M. Z., Ting, L., Katrib, M., Ng, C., Sowers, K. R., Galperin, M. Y., Anderson, I. J., Ivanova, N., Dalin, E., Martinez, M., Lapidus, A., Hauser, L., Land, M., Thomas, T., & Cavicchioli, R. 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J.* Sep;3(9):1012-35.

To be submitted

De Francisci, D., Kornfeld, G. Campanaro, S., Siddiqui, K. S., Ertan, H., Treu, L., Pilak, O., Williams, T. J., Harrop, S. J., Curmi, P. G. M. & Cavicchioli R. The RNA polymerase subunits E/F from an archaeon bind to specific classes of mRNA.

Campanaro, S., Williams, T. J., De Francisci, D., Treu, L., Lauro, F. M. & Cavicchioli, R. Thermal regulation of gene and operon expression in the Antarctic archaeon *Methanococcoides burtonii.*

Acknowledgements

This thesis would not have been possible without the contributions of many people who have helped and supported me throughout the various stages of this candidature.

I must thank my supervisor, Prof. Rick Cavicchioli, for giving me the opportunity to carry out my PhD degree in his laboratory at UNSW, and for his support throughout the years. I hope this quest was as rewarding for you as it was for me.

I'd like to thank Prof. Paul Curmi for providing constant support and making always time for me when I was in trouble.

To Dominic, Charmaine and Sabine, thank you for your friendship and support throughout the years. To Sheree, thank you for being close to me in these last months.

I must thank Tim and Haluk for the help in the last months with the methanogenesis, and for the wonderful discussions we had.

To Sohail, Geoff and Stefano: I will never be able to reciprocate what you have done for me.

On a more personal note, thank you to Nadeem and Jeyran for the help in these last days. It is something that I will never forget.

Table of Contents

List of Figures

List of Tables

Common Abbreviations used

ATP	adenosine triphosphate
Å	angstrom
ALA	5-aminolavulenic acid
Bp	basepair
BLAST	basic local alignment search tool
CD	circular dichroism
COG	cluster of orthologous genes
CSP	cold shock protein
Da	dalton
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSC	differential scanning calorimetry
DFQ	dynamic fluorescence quenching
dH_2O	ultra-pure water (Milli-Q)
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EF2	elongation factor 2
ER	evidence rating
Fd	ferredoxin
GO	gene ontology
GSSG	glutathione (oxidised)
GSH	glutathione (reduced)
g	gravitational force
Fe-S	Iron-Sulfur
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICAT	isotope coded affinity tag
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMG	integrated microbial genomes

JGI	Joint genome institute
J	joule
K	kelvin
kDa	kilodalton
LB	Luria Bertani
mRNA	messenger ribonucleic acid
MFM	Methanococcoides trimethylamine media (rich, complex)
MW	molecular weight
NAD(P)H	NADH or NADPH
NCBI	National Center for Biotechnology Information
NADH	nicotinamide adenine dinucleotide-reduced form
NADPH	nicotinamide adenine dinucleotide phosphate-reduced form
OB	oligonucleotide/oligosaccharide binding
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
PMF	proton motive force
Ppase	proton translocating inorganic pyrophosphatase
PPi	pyrophosphate
Q-PCR	quantitative – polymerase chain reaction
RT-PCR	reverse transcriptase – polymerase chain reaction
RNA	ribonucleic acid
rpm	rounds per minute
SAM	S-adenosylmethionine
sec	second
ss	single stranded
SDS	sodium dodecyl sulfate
S	Svedgerg unit
H_4SPT	tetrahydrosarcinapterin

Tris	Tris(hydroxymethyl)aminomethane
TAE	Tris-Acetate-EDTA
tRNA	transfer ribonucleic acid
2DE	two-dimensional electrophoresis
U	units
UV	ultra-violet
v/v	volume per volume
w/v	weight per volume

Chapter 1: General Introduction

1.1 The Archaea

Ever since the initial studies by Woese and Fox (1977), compelling evidence has supported the hypothesis that the archaea represent the third Domain of life, separated from bacteria and eucarya.

Features unique to archaea include: the structure of phospholipids (Koga *et al.*, 1993, 2007), the cell envelopes, the absence of chlorophyll-based photosynthesis (Kletzin, 2007), the ability to perform methanogenesis and the synthesis of a number of cofactors that are exclusive of this metabolic pathway (Ferry and Kastead, 2007; Gao and Gupta, 2007). Comparative genome analysis identified more than 300 archaeal genes that are conserved in all genome sequences of archaea and do not have bacterial or eucaryal homologues (Makarova and Koonin, 2003).

Some of the features that are shared by archaea and bacteria are: cell structure, the lack of a cell nucleus and nuclear membranes, the small size of the genome usually arranged in a single circular chromosome where most of the genes are organised in operons, the absence of membrane-bound organelles, and the small size of their cells (Brown *et al.*, 1997).

Similarities between archaea and eucarya are concentrated in the transcription and translation systems (Bell and Jackson, 2001; Bell *et al.*, 2001; Reeve, 2003). Similar to eucarya, the core RNA polymerase from archaea contains numerous subunits in contrast to just four in bacteria. Recognition of promoter elements in bacteria is facilitated by sigma factors which associate with the core RNA polymerase to form the holoenzyme prior to DNA attachment. In contrast, in archaea transcription occurs only after transcription factors, such as TATA-binding protein (TBP), have bound to promoter regions, which in turns sequester RNA polymerase to bind DNA and to initiate transcription. In these respects, archaea transcription follows the eucaryotic paradigm. Messenger RNA (mRNA)

of archaea are not 5'- end-capped, usually have only short polyadenylated 3'-end tails, and contain Shine-Dalgarno type ribosome binding sites thereby following a bacterial-like structure; (Bell and Jackson, 1998; Leigh, 1999; Soppa, 1999; Hickey *et al.*, 2002; Thomm, 2007). Also similar to bacteria is the organisation of functionally related genes in polycistronic transcription units (operons) (Bell and Jackson, 1998; Thomm, 2007).

Archaeal ribosomes share the 70S structure found in bacteria including the presence of 5, 16 and 23 S ribosomal RNA (rRNA). While many of the genes for ribosomal proteins, translation initiation, elongation and termination factor proteins are organised in operons resembling the one found in bacteria, the structure of the proteins is more closely related to eucarya (Londei, 2007).

Traditionally archaea have been associated with life under extreme conditions and have been divided into methanogens, extreme halophiles, thermo- and hyper-thermophiles and thermoacidophiles, according to their corresponding habitat. However archaea have also been identified in abundance in non extreme environments such as ocean waters (Karner *et al.*, 2001; DeLong, 2005), freshwater (Pernthaler *et al.*, 1998; MacGregor *et al.*, 1997) and soils (Barns *et al.*, 1994, 1996; Sliwinski and Goodman, 2004; Schleper *et al.*, 2005). In some cases, particular types of archaea populate a wide variety of habitats (e.g., methanogens).

The domain archaea is divided into four kingdoms: the euryarchaeota and the crenarchaeota (Woese, 1990; Chaban, 2006), which represent the two largest groups of archaea, the korarchaeota first described by Barns *et al.*, (1996), and the nanoarchaeota (Huber *et al.*, 2002). Within these four branches of the archaea exists a wide diversity of physiologies, environmental tolerance and metabolic processes.

1.2 Methanogens

Methanogens belong to the kingdom of euryarchaeota, and are classified into six different orders: *Methanopyrales*, *Methanococcales*, *Methanobacteriales*, *Methanomicrobiales*,

Methanocellales and *Methanosarcinales* (Shima *et al.*, 2002; Ferry and Kastead, 2007; Sakai *et al.*, 2008). Members of these six orders differ in cell morphology and motility, lipid content and growth requirements, making the methanogens a very diverse group of microorganisms.

The name methanogen reflects their ability to synthesise methane as the final product of their energy and biomass production. The most common methanogenesis substrates include carbon dioxide/molecular hydrogen, acetate, formate, methanol, methylamines, methylthiols and dimethylsulfide (Ferry and Kastead, 2007). These compounds are the final products of the various fermentation processes, and this could explain the variety of habitats of methanogens, which have been isolated from almost every anaerobic environment on the planet, including freshwater and marine sediments, digestive and intestinal tracts of animals, and anaerobic waste digesters. In particular, the range of growth temperatures (-2 to 110), NaCl concentrations (freshwater to hypersaline) and pH (moderately acidic to above 9) is extremely variable in methanogens ecosystems (Chaban *et al.*, 2006). Those isolated from Antarctica include *Methanococcoides burtonii* and *Methanogenium frigidum*, and those from hydrothermal vents include *Methanopyrus kandleri*, *Methanothermus fervidus*, *Methanothermus sociabilis*, *Methanocaldococcus jannaschii* and *Methanococcus igneus*.

1.3 Cold Adaptation

Archaea represent an important fraction of the microbial population of cold environments such as marine waters, freshwater and saline lakes, sea ice, freshwater and marine sediments (Cavicchioli, 2006).

Microorganisms have originally been divided into psychrotrophs and psychrophiles according to the range of temperature they can tolerate (Morita, 1975). Although this terminology is still used, the recently coined terms eurypsychrophiles (from Greek euros =

broad) and stenopsychrophiles (from Greek stenos = narrow) are more appropriate for this purpose (Feller and Gerday, 2003; Bakerman and Nealson, 2004; Cavicchioli, 2006).

Despite the abundance of archaea in the cold environments, culturable isolates are few. Most of what is known about cold adaptation mechanisms in archaea has been gained from studies on the Antarctic methanogen *Methanococcoides burtonii* (Thomas and Cavicchioli, 1998, 2002; Lim *et al.*, 2000; Noon *et al.*, 2003; Goodchild *et al.*, 2004, 2005; Nichols *et al.*, 2004; Allen *et al.*, 2009; Williams *et al.*, 2010), but also on the psychrophilic archaeon *Cenarchaeum symbiosum* (Schleper *et al.*, 1997, 1998), *Halorubrum lacusprufundi* (Gibson *et al.*, 2005) and *Methanogenium frigidum* (Saunders *et al.*, 2003, Cavicchioli *et al.*, 2006; Giaquinto *et al.*, 2007).

Studies on bacteria indicate that some of the main problems related to cold growth are: the loss of membrane fluidity, a lowering in the rates of enzymatic and transport processes, the stabilisation of nucleic acid secondary structures and the subsequent inhibitory effects on DNA replication, transcription and the translation of mRNA, and the formation of crystalline ice and its associate damage to cellular structures if the temperature goes below the freezing temperature of the cytoplasm (Jones and Inouye, 1994; Grauman *et al.*, 1996; Yamanaka, 1999; Phadtare *et al.*, 1999).

As temperature decreases, the membranes of organisms tend to become more rigid (increased gel phase). This leads to a decrease in the permeability to small molecules, nutrient uptake, electron transport, and to an increased susceptibility to cellular damage by ice crystal formation (Yamada *et al.*, 2002; Russell, 2007). The maintenance of membrane fluidity in bacteria is achieved by an increase in the proportion of unsaturated fatty acids (Tarpgaard *et al.*, 2006). In some organisms this is achieved by increasing the desaturase enzyme activity (Wada *et al.*, 1987) or by *de novo* fatty acid biosynthesis (Russell, 1990; Russell and Fukunaga, 1990). This maintains the membrane in a liquid crystalline state and has been reported for not only psychrophilic bacteria but also psychrophilic archaea (Nichols and Franzmann 1992; Nichols *et al.*, 2004; Gibson *et al.*, 2005).

Psychrophiles features also include the capacity to take up and produce osmolytes and antifreeze proteins. Trehalose and proline are known to protect yeast cells from low temperatures (Terao *et al.*, 2003; Kandror *et al.*, 1997), and cryoprotectants have also been identified in organisms that need to survive in cold environments (Methe *et al.*, 2005; Riley *et al.*, 2008); antifreeze proteins have been identified in Antarctic bacteria (Gilbert *et al.*, 2004).

However, mechanisms of protein adaptations to low temperature are the most studied aspect of psychrophiles. The reason for this lies in their high activity and catalytic efficiency in the cold, which translates to reduced incubation temperatures and energy consumption. This property is of course of major interest for industrial applications (Cavicchioli *et al.*, 2002).

Genomic comparison between the archaea *M. burtonii* and *M. frigidum* and non-psychrophilic archaea revealed that overall the proteins from the psychrophiles have a higher content of non-charged polar amino acids and a lower content of hydrophobic amino acids (Saunders *et al.*, 2003). Such changes are thought to be involved in increasing the flexibility of the proteins and therefore the catalytic efficiency at low temperatures.

In the last years several reviews have been written on this subject (D'Amico *et al.*, 2002, Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007). These works focus on structural analyses based on homology modeling or X-ray-crystal structures aimed to the identification of unique features for psychrophilic proteins when compared to their mesophilic or thermophilic homologues. As a result of these studies, the following features have been suggested to be determinant for the properties of low-temperature adapted proteins: reduction in number of salt bridges, reduction of aromatic interactions, reduced hydrophobic clustering, reduced proline content, reduced arginine content, additional loop-structures, more solvent interaction.

It is important to note that not all of these types of changes have been found in a single low-temperature adapted protein. In general, the above mentioned features are thought to

destabilise the three-dimensional structure of the protein and therefore lead to a higher flexibility at low temperatures.

1.4 *Methanococcoides burtonii*

Methanococcoides burtonii was isolated from the anaerobic, methane and H_2S rich depths of Ace Lake, which is located in the Vestfold Hills region, in the Australian Antarctic Territory (Franzmann *et al.*, 1992). The lake has a permanent temperature of 1 – 2 °C and is covered with ice for ~11 months of the year.

M. burtonii is an obligately methylotrophic methanogen, utilising only methylamines and methanol as substrates for growth. It is flagellated, motile and is classified as eurypsychrophile as it can tolerate temperatures ranging from -2 °C to 29 °C, with maximum growth rate at 23 °C (Cavicchioli, 2006; Allen *et al.*, 2009) (Fig. 1.1).

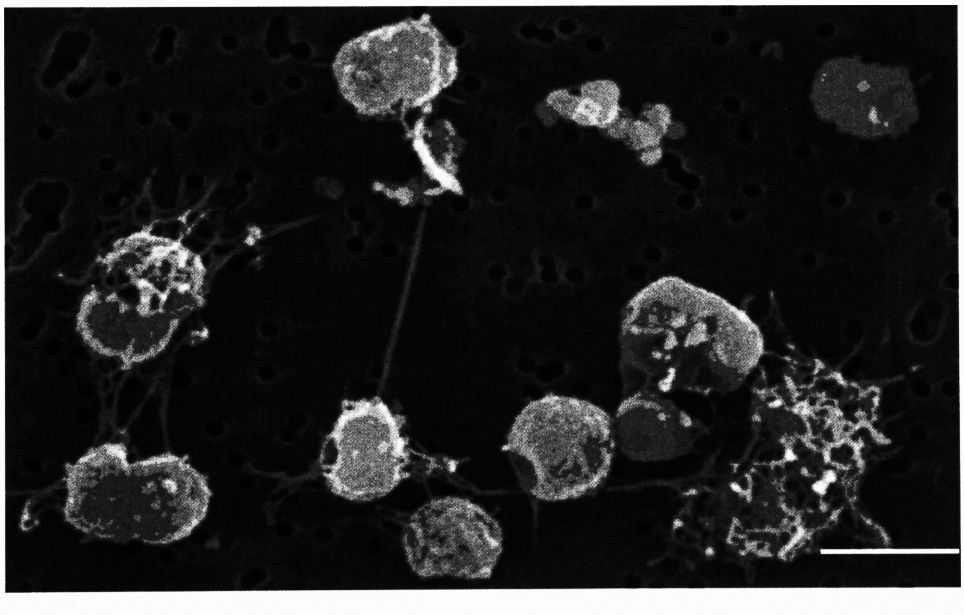


Figure 1. 1: Electron micrograph of *M. burtonii* taken by Dr Dominic Burg at UNSW. Scale bar represents 1 µm.

M. burtonii is one of the most characterised of all psychrophilic microorganisms and has been utilised as a model for a number of studies on cold adaptation.

The genome of *M. burtonii* has recently been closed and this has allowed for comprehensive genomic analysis (Allen *et al.*, 2009). The analysis highlighted the immense adaptive potential of the organism, evidenced by an over-representation of signal transduction genes, several *Bacteria*-like central metabolism genes and a large complement of transposases. Proteomic analyses have also identified two of these transposases as being expressed in the cell, indicating active genome rearrangement events under cold conditions (Goodchild *et al.*, 2004a). This ability for genome re-arrangement may be reflective of the relatively recent geographical isolation of Ace Lake, and hence *M. burtonii*.

For example, the molecular adaptation of the elongation factor 2 (EF2) protein from *M. burtonii* showed that at low temperatures the enzyme was active and relatively thermolabile. Comparison with the homologous protein from the thermophilic methanogen *Methanosarcina thermophila* revealed that the cold adapted protein has a higher GTPase activity at low temperature and requires lower activation energy for the GTP hydrolysis (Thomas and Cavicchioli, 2002). These properties were linked to a higher flexibility and thus a lower thermostability of the *M. burtonii* EF2 structure (Thomas and Cavicchioli, 2002) and these characteristics were attributed to a lower number of predicted salt bridges, less packed hydrophobic cores, and a reduction of proline residues in surface loops compared to the thermophilic homologue (Siddiqui *et al.*, 2002). The studies on the *M. burtonii* EF2 proteins proved that the molecular determinants for the cold adaptation of an archaeal protein are very similar to that of bacterial proteins.

M. burtonii has also been used as a model organism to study the influence of temperature on tRNA modification (Noon *et al.*, 2003). It was found that *M. burtonii* had the lowest levels of modification of any organism studied, and that there was a high content of dihydrouridine, which would confer flexibility to the tRNA molecules (Dalluge *et al.*, 1997).

A DEAD box helicase from *M. burtonii* was found to be expressed exclusively under cold conditions through mRNA analysis (Lim *et al.*, 2000). DEAD box helicases have a variety of roles in the cell including the stabilising of RNA secondary structures and transcriptional and translational accessory functions.

Only one crystal structure of an *M. burtonii* protein (a proliferating cell nuclear antigen) has been published so far (Byrne-Steele *et al.*, 2009). The protein in question is a DNA-clamping protein involved in DNA replication and repair.

Extensive proteome analyses have been performed on *M. burtonii*. Comparative proteomics was first achieved through the use of two-dimensional electrophoresis (2DE), stable isotope labeling using isotope coded affinity tags (ICAT), and through one-dimensional electrophoresis of the secreted proteins (Goodchild *et al.*, 2004b, 2005). The 2DE analysis led to the identification of 54 spots with > 2 fold changes in abundance at 4 °C vs. 23 °C. Many of these proteins were identified as differentially abundant at the transcript level in complementary work using RT-PCR.

This comparative study revealed that transcription, protein folding and central metabolism appear to be deeply involved in *M. burtonii* cold adaptation. In particular, RNA polymerase subunit E was found to be more abundant at 4 °C *vs* 23 °C and therefore has been proposed to be linked to the adaptation of *M. burtonii* to the cold. This data regarding this subunit has been confirmed by a transcriptomic analysis performed in the same conditions, where the transcript coding for this specific subunit was also found to be more abundant in the same conditions (Campanaro *et al.*, paper in preparation).

The efficient generation of energy at cold temperatures is crucial for psychrophiles. Methanogens are able to generate ATP through the use of proton/ion transporting ATPases, which use electrochemical gradients as the driving force behind ATP production.

Goodchild *et al.*, (2004b), found several methanogenesis proteins related to generation of proton motive force (PMF) as differentially abundant in the cell. The authors argued that the cells prefer to generate energy via a PMF rather than a sodium motive force (SMF) at

cold temperatures. However, there is no evidence to suggest that *M. burtonii* uses a SMF for ATP production under any situation. A global proteomics investigation that enriched for secreted and membrane proteins as well as cytoplasmic proteins (Williams *et al.*, 2010) found that the cell envelope of *M. burtonii* underwent profound changes during cold growth, with many surface proteins showing higher expression. Also, many proteins associated with transcription and translation were higher at cold temperatures, including a family of single-domain TRAM proteins that were proposed as a new class of RNA chaperones.

1.5 RNA Polymerase Subunits E and F

The archaeal and the eucaryal transcription systems share a number of features that strongly reinforce theory that these kingdoms have a common ancestry apart from bacteria (Langer and Zillig, 1995). Archaea have a single RNA polymerase (RNAP) that is constituted of 11 or 12 subunits. The crystal structure of *S. solfataricus* RNAP (Hirata et al. 2008), reveals two distinct structures: a nine-subunit core responsible for the catalytic activity and a dissociable heterodimeric subcomplex (subunits E and F), whose functional contribution to the transcription process is poorly understood. The archaeal RNAP is closely related to the eucaryotic RNA polymerases (especially RNAP II) in terms of number and structure of the individual subunits. Likewise, the basic transcription initiation apparatus is modulated in archaea by a TATA-binding protein (TBP) and the general transcription factor TFB, both highly conserved in both sequence and function to the eucaryotic TBP and TFIIB (Bell and Jackson, 1998). The archaeal RNAP system can therefore be considered as an ideal and simple model for understanding the molecular basis of eucaryotic transcription.

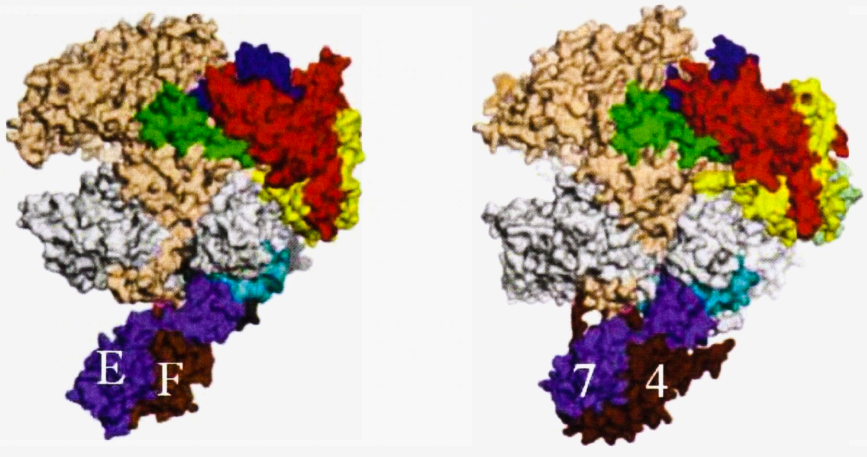


Figure 1. 2: RNAP structures from archaea and eucarya. Surface representations of the RNAP structures from *S. solfataricus* (left) and from *S. cerevisiae* (RNAPII, right). Orthologous subunits are depicted by the same colour. The archaeal subunits E anf F and the eucaryal subunits 4 and 7 are indicated. The architecture of the two RNAP structures is extremely similar. The two heterodimers are in the same position and present a very similar structure as well. The figure has been adapted from Hirata *et al.*, 2008b.

The archaeal RpoE/F heterodimer has a high level of sequence identity and structural similarity to the eucaryal Rpa14/Rpa43 (RNAP I), Rpb4/Rpb7 (RNAP II) and Rpc17/Rpc25 (RNAP III) pairs (Choder, 2004; Jasiak *et al.*, 2006; Kuhn *et al.*, 2007; Meka *et al.*, 2005; Todone *et al.*, 2001). These sequence and structural properties underscore strong functional similarities.

In *S. cerevisiae* Rpb4/7 contacts the core RNAP II complex near the transcript-exit groove and is positioned to be able to interact with GTFs TFIIB and TFIIF, fulfilling roles in transcription initiation, elongation and termination (Choder, 2004; Goler-Baron *et al.*, 2008).

The first crystal structure of an E/F heterodimer came from the archaeon *Methanocaldococcus jannaschii* which revealed that RpoE contains a truncated RNP-fold that includes conserved residues involved in interactions with RNA and a domain that resembles an OB-fold that is also associated with non-specific binding of single stranded RNA (Todone *et al.*, 2001; Meka *et al.*, 2003; Choder 2004). These findings led to the hypothesis that RpoE/F interacts with the emerging RNA transcript through the RNA

binding domains – domains that are highly conserved, from Archaea to humans. The location of this subcomplex on the structure of the RNAP holoenzyme from yeast (Armache *et al.*, 2005) and Archaea (Hirata *et al.*, 2008b), places it near the exit channel for the newly transcribed RNA, supporting its purported role in RNA binding. Electrophoretic mobility shift assays (EMSA) with *M. jannaschii* RpoE/F, and yeast and human Rpb4/7, confirmed the ability of the heterodimer to bind RNA, and ssDNA (Todone *et al.*, 2001; Orlicky *et al.*, 2001; Meka *et al.*, 2003, 2005).

One of the most obscure characteristics of this heterodimer is that both in eucarya and in archaea it appears to be involved in the cellular response to stress. Mutagenesis studies have revealed that in *S. cerevisiae* Rpb7 is an essential protein while Rpb4 is dispensable under optimal growth conditions (Choder, 2004) however it becomes essential at non-optimal temperatures, under oxidative stress, and in starvation. Importantly, Rpb4/7 association with the core of the RNAP complex is dependent on environmental conditions: during exponential phase only 20% of the pol II complexes are bound to Rpb4/7, while in stationary phase, where cells are starved of nutrients, Rpb4/7 binds Pol II in a stoichiometric manner (Choder, 1993). The situation is very similar in archaea: a mutant of *T. kodakarensis* lacking RpoF is thermosensitive, while it has proven to be impossible to isolate mutants lacking RpoE (Hirata *et al.*, 2008). Furthermore, work with *P. furiosus* showed that RpoE is essential for transcription initiation at low temperature *in vitro* (Naji *et al.*, 2007).

In yeast, Rpb4/7 mediates the RNAP II control of the two major mRNA decays in the cytoplasm: during transcription, Rpb4/7 interacts with the transcript and the Rpb4/7–RNA complex migrates out of the nucleus. This feature is apparent only during stress (Farago *et al.*, 2003). At this point Rpb4/7 promotes the binding of Pat1 to the mRNA, the poly(A) tail is shortened, both major pathways of mRNA degradation are stimulated, and degradation of the transcript occurs in the P bodies (Lotan *et al.*, 2005; 2007). In this scenario, it has also been proposed that Rpb4 mediates the decay of a specific class of genes (Choder 2005) which correlates with the finding that Rpb4 affects expression of a

small yet significant fraction of the genome in both stress and normal conditions (Pillai *et al.*, 2003).

1.6 Aims

Methanococcoides burtonii has proven to be a model organism for studying cold adaptation and the biology of archaea, being the subject of a broad range of targetted studies and global analyses of gene content, function and expression. One of the strongest themes that has emerged from these studies is the important role that nucleic-acid binding proteins play in cold adaptation, through functions they perform in regulating gene expression and interacting with RNA.

In the beginning of this study it was hypothesised that the structure of *Mb*RpoE/F could present a number of molecular features responsible for the adaptation of this heterodimer to the cold. In order to individuate these molecular features an attempt was undertaken to generate a crystal structure of this heterodimer for comparison with the one already solved of the homologue from the hyperthermophile *Methanocaldococcus jannaschii* (*Mj*RpoE/F). Other techniques were also used to structurally and functionally compare the two heterodimers and find differences that could be linked to thermal adaptation.

The results from the functional comparison, performed by electrophoretic mobility shift assay (EMSA) led to the most interesting aspect of my research, as it appeared that the heterodimer showed a preferential binding for some ssDNA sequences. These results led to a hypothesis never formulated before: that the heterodimer could bind nucleic acid in a specific manner and that this specificity was the key to understand the real function of the E/F heterodimer in the RNA polymerase machinery. This hypothesis was tested via a novel method involving the mixing of the *Mb*RpoE/F with the total RNA extracted from *M. burtonii* cultures, followed by affinity purification of the target RNA molecules and microarray identification."

In summary, the characterisation of *Mb*RpoE/F was achieved through the following specific aims:

- Develop and optimise recombinant expression and purification methods for the *Mb*RpoE/F and its thermophilic homologue from *Methanocaldococcus jannaschii* (*Mj*RpoE/F).
- Generate a crystal stucture of *Mb*RpoE/F to be compared with that of *Mj*RpoE/F in order to identify molecular features responsible for thermal adaptation.
- Perform a structural and functional comparison between *Mb*RpoE/F and *Mj*RpoE/F via differential scanning calorimetry (DSC), dynamic fluorescence quenching (DFQ) and electrophoretic mobility shift assay (EMSA).
- Develop a system that used affinity purification followed by microarray identification of the target RNA molecules for *Mb*RpoE/F.

Chapter 2: Overexpression and Purification of the RpoE/F Proteins

2.1 Introduction

In this chapter the strategies and methodologies developed for the production of recombinant *M. burtonii* and *M. jannaschii* RpoE/F will be discussed. Large amounts of these proteins were necessary for their characterisation, especially in the case of the crystallisation trials for *Mb*RpoE/F.

The approach initially taken was to express, purify and characterise the *M. burtonii* subunit E alone, because proteomic studies had proved that only this subunit and not F was present in higher abundance at low temperatures relative to high temperatures (Goodchild *et al.*, 2004). These data are supported by transcriptomic studies conducted on *M. burtonii* in the same conditions that have proven that the RpoE transcript is more abundant at 4 °C than at 23 °C (Campanaro *et al.*, to be submitted). It was not possible however to produce a soluble recombinant E subunit.

It has previously been reported that the subunits RpoE and RpoF are not soluble when expressed separately but are soluble when co-expressed (Thomm, 2007). Therefore it was decided to express the two *M. burtonii* subunits together and characterise the recombinant heterodimer. Besides overcoming the practical limitations above, such a strategy is also more biologically relevant, since the subunits are found as the heterodimer *in vivo*.

A standardised strategy for the co-expression of a recombinant full-length *Mj*RpoE/F has been described by Werner *et al.* (2000) and involves the cloning of the two genes into the pGEX-2TK vector (Pharmacia) in order to produce an F subunit fused with glutathione S transferase via a linker which contains a thrombin cleavage site, useful for separating the two proteins in one of the final steps of the purification (Werner *et al.*, 2000; Todone *et al.*, 2001). A different system was developed for the co-expression of the RNA polymerase E/F

from *M. burtonii*. This system involved the pETDuet™-1 vector (Novagen) which is also designed for cloning and expression of two target genes, but in this case one fusion protein will have an N-terminal His-Tag (from Multiple Cloning Site 1) and the other will have a C-terminal S-Tag sequence (from Multiple Cloning Site 2) (Stebbings *et al.*, 1999). In this specific case the subunit F was cloned in the MSC1 to give a His-tagged F, while the subunit E was cloned in the MCS2. The pETDuet™-1 vector was preferred to the pGEX-2TK for the cloning and expression of the *M. burtonii* RpoE/F genes in order to avoid the thrombin cleavage step in the purification process.

With this strategy it was possible to successfully co-express and co-purify the recombinant *Mb*RpoE and F from *M. burtonii* in a soluble form. It was also possible to obtain the same recombinant pGEX-2TK clone described in Werner *et al.* (2000) and Todone *et al.* (2001) from Prof. Werner and the same purification strategy was followed. With these two recombinant homologues it was possible to start comparative stability, structural, and functional analyses alongside the crystallisation trials for *Mb*RpoE/F.

2.2 Materials and Methods

2.2.1 Growth of *M. burtonii* cultures

Anaerobic culture media was prepared using equipment and techniques described in Sowers and Noll (1995). Vitamin and mineral stock solutions (100x concentration) were prepared as described by Goodchild (2004).

A modification of MGM (Franzmann *et al.*, 1992) known as MFM, was prepared as described by Thomas and Cavicchioli (2000). Firstly 1.2 l of dH_2O was heated in a microwave on 100% power for 3 min. The dH_2O was poured into a sidearm flask and a Teflon coated magnetic stir-bar inserted. The upper flask opening was plugged with a rubber bung, the solution placed onto a magnetic stirrer and de-gassed using a Biorad HydroTech vacuum pump (BioRad, Hercules, CA, USA) until gas bubbles were no longer being evacuated. While the dH_2O was being prepared, two Schott bottles (a 500 ml and a 2

L) were thoroughly washed with dH_2O to remove any traces of residual detergents. These bottles were flushed with high purity nitrogen gas (BOC gases, North Ryde, NSW): the 500 ml bottle with a straight unbeveled needle, and the 2 l with a gassing stone using a manifold as described by Sowers and Noll (1995). A 500 ml volume of the de-gassed dH_2O was poured into the 2 l bottle, which was placed onto a magnetic stirrer and a Teflon coated stir-bar added, while continually bubbling with nitrogen gas. The remainder of the de-gassed dH_2O was added to the 500 ml bottle and was bubbled with nitrogen. The following salts were then added to the 2 l bottle: 0.335 g KCl; 6 g $MgCl_2.6H_2O$; 1 g $MgSO_4.7H_2O$; 0.25 g NH_4Cl (AnalaR, BDH, Poole, England); 0.14 g $CaCl_2.2H_2O$; 23.37 g NaCl; 2 mg $[NH_4]_2$ [FE][SO_4]$_2$.$6H_20$(Sigma, St Louis, MO, USA); and 1 mg of the redox indicator rezazurin (Sigma, St Louis, MO, USA). The solution, which now had a purple colour, was left to stir while being bubbled with nitrogen gas for 10 –15 min. The following chemicals were added in order, making sure each was dissolved before the addition of the next: 5 g trimethylamine HCl (ICN, Aurora, OH, USA); 2 g yeast extract (Oxoid, Hampshire, England); 10 ml of the vitamin solution and 10 ml of the mineral salts solution, both prepared as described in Goodchild (2004); 0.1 g sodium acetate; and 0.14 g K_2HPO_4. Food grade CO_2 gas (BOC gases, North Ryde, NSW) was introduced to a ratio of 80:20, N_2:CO_2 as measured using a gas proportioner. The solution was bubbled for 10 min or until the colour of the solution changed from purple to pink indicating a change in pH. Next, 0.5 g of the reducing agent cysteine HCl (Sigma, St Louis, MO, USA) was pre-dissolved in 1 ml dH_2O and added to the solution followed by 2.52 g Na_2CO_3, which with the CO_2 gas, acted to buffer the solution. The volume was adjusted to 1 l using the water from the 500 ml flask. The solution was left to stir, with continuous gas bubbling, for 30 min, or until the solution was pale yellow in colour (indicating reduced conditions). The pH of the solution was adjusted to 6.8 using 10.1 M HCl and the stir-bar turned off. Serum bottles (120 ml) (Wheaton, Millville, NJ, USA), were rinsed with dH_2O and flushed with the N_2:CO_2 gas mixture using straight un-beveled syringes for 20 – 30 min. The media was transferred to the serum bottles using a 60 ml syringe (BD Scientific, Singapore), and a 25 cm needle large gauge. The syringe was first filled with gas from the headspace of the 2 l bottle containing the medium, ejected and re-filled with gas three times. The medium was

transferred to a serum bottle, and this operation was repeated until all serum bottles contained 100 ml of medium. The serum bottles were bubbled with the gas mixture for 45 min or until the solution was pale yellow in colour. The bottles were plugged with butyl rubber stoppers (Bellco Glass, Vineland, NJ, USA), sealed with aluminium crimp seals (Bellco Glass, Vineland, NJ, USA), and autoclaved at 121 °C for 15 min. After autoclaving the media was placed on a shaker overnight to dissolve any precipitate that had formed. 1 ml of 2.5% Na_2S solution (a reducing agent, prepared as described below) was injected into each of the serum bottles and these were allowed to equilibrate for 2 hours before inoculation, or, stored at 4 °C for up to 2 months if not needed immediately. A stock solution of 2.5% Na_2S was prepared by first heating and degassing 120 ml of dH_2O as described above. A 100 ml aliquot of this de-gassed dH_2O was transferred to a serum bottle and bubbled with nitrogen using the gas manifold for 30 min. The serum bottle was temporarily plugged with a butyl rubber stopper and set aside. Working in a fume hood, slightly more than 2.5 g of Na_2S crystals (Sigma, St Louis, MO, USA) were weighed out, washed with dH_2O, dried with a paper towel and re-weighed for a final mass of 2.5g. The stopper was removed from the serum bottle and the crystals rapidly transferred before re-inserting the rubber stopper and crimp sealing the serum bottle. This 2.5% Na_2S solution was autoclaved at 121°C for 15 min, and stored at 4 °C until needed.

M. burtonii cultures were prepared from laboratory stock cultures. Cultures were inoculated 1:100 from cells grown under the same conditions.

2.2.2 Genomic DNA Extraction

M. burtonii cultures (50 ml) were harvested at late logarithmic phase (optical density at 620 nm of 0.25) by centrifugation at 3200 x *g* for 35 min at 4 °C. The cell pellet was resuspended in 2 ml of XS-buffer (1% Potassium ethyl xanthogenate, 800 mM NH_4OAc, 100 mM Tris-HCl, pH 7.4, 20 mM EDTA, 1% SDS) and incubated for 1 hour at 70 °C. The

suspension was briefly vortexed, incubated for 30 min on ice and then centrifugated for 10 min at 21000 x *g* at 4 °C.

The supernatant was transferred into a fresh tube and an equal volume of phenol:chloroform:isoamylalcool 25:24:1 (v/v/v), saturated with 10 mM Tris, pH 8.8, 1 mM EDTA, (Sigma) was added, mixed thoroughly and centrifuged for 5 min at 12000 x *g* at room temperature. The aqueous upper layer was transferred into a fresh tube and the phenol/ chloroform/ isoamylalcohol extraction was repeated until a clear interface was obtained.

The nucleic acids present in the final aqueous layer were precipitated using a standard ethanol precipitation procedure. The DNA pellet was dissolved in 50 µl TE buffer. Residual RNA was removed by RNAse treatment (1 U/µl final concentration of RNAse A and RNAse T). The sample was incubated for 30 min at 37 °C. The concentration and the purity of DNA were verified on a 1% (w/v) agarose gel in TAE-buffer (60 mM Tris-acetate, 1 mM EDTA) for 1 h at 100 V.

2.2.3 Primer Design and PCR Amplification

Forward and reverse primers were designed for *M. burtonii* RpoE and RpoF genes. The primers for the RNA polymerase subunit F gene were designed to add *Bam*HI and *HinD*III restriction sites at the extremities of the amplified gene, while the primers for the subunit E gene were designed to add *Nde*I and *Kpn*I restriction sites at the extremities of the amplified gene and a stop codon at the end of the orf. The primers are indicated in Table 2.1.

Table 2. 1: Primer sequences used to amplify the *M. burtonii* RpoE and RpoF genes.

***M. burtonii* gene name**	**Primer name**	**Primer sequence 5'→ 3'**
RpoE	**Forward**	TTTCATATGTATAAAAGGATGAAACTT
	Reverse	GGTACCTTATTCATCTTTATTCTTAGA

RpoF	**Forward**	AAGGATCCAATGTTAAACATATCTAATT
	Reverse	AAAAGCTTTTATTCCATTGCCTCG

The reaction mix for PCR amplifiction of the constructs contained 20 pmol of forward and reverse primer (Invitrogen), 200 μM of each dNTP, 2.5 mM $MgCl_2$ (Promega), 1x Taq DNA Polymerase buffer (Promega), 2.5 U of Taq DNA Polymerase (Promega) and 100 ng of genomic DNA. The PCR amplifications were carried out in a Thermal Cycle PCR machine (Hybaid). A scheme of the amplification cycles for the two genes is summarised in Table 2.2.

Table 2. 2: Scheme of the thermal cycles used to amplify the *M. burtonii* RpoE and RpoF genes

***M. burtonii* RpoE**	***M. burtonii* RpoF**
95 °C – 3 min	**95 °C – 3 min**
40 °C – 1 min	**40 °C – 1 min**
72 °C – 45 sec x 1	**72 °C – 30 sec x 1**
95 °C – 1 min	**95 °C – 1 min**
55 °C – 1 min	**55 °C – 1 min**
72 °C – 45 sec x 29	**72 °C – 30 sec x 29**
72 °C – 5 min	**72 °C - 5 min**
4 °C – hold temperature	**4 °C – hold temperature**

2.2.4 Electrophoresis on Agarose

The PCR reactions were mixed with 10 µl of 6x loading dye composed of 0.03% xylene cyanol FF, 0.03% bromophenol blue, 15% Ficoll, 10 mM Tris-HCl pH 7.5, 50 mM EDTA. The samples were loaded onto 1% (w/v) agarose gel (AppliChem), electrophoresed in TAE buffer at 100V and stained in TAE buffer containing 0.5 µg/ml of ethidium bromide. The DNA was visualised using a UV transilluminator at 254 nm (Gel Doc 2000, BIORAD). PCR products were extracted from the agarose gel and purified using the QIAquick® gel Extraction Kit (QIAGEN). The extracted DNA was quantified usinga nanodrop (BIOLAB)

2.2.5 Classical Cloning

The pETDuet™-1 vector (Fig 2.1) contains two expression units, each controlled by a T7*lac* promoter (Fig. 2.2). The two promoters are followed by ribosome binding sequences and multiple cloning sites (MCS1 and MCS2). The vector also carries the ampicillin resistance (Ap) gene (bla) and is compatible with the Rosetta™ host strain (Novagen) which contains the pRARE plasmids.

The first multiple cloning region (MCS1) contains a 6-aa His-Tag coding sequence and several restriction sites (including *Bam*H I and *Hin*DIII). The second multiple cloning site (MCS2) begins with an *Nde* I (CATATG) site at the ATG (Met) translation initiation site, followed by other restriction sites including *Kpn*I and a sequence encoding the 15-aa S-Tag peptide. Therefore one of the expressed protein products will have one an N-terminal His-Tag sequence, and the other a C-terminal S-Tag sequence.

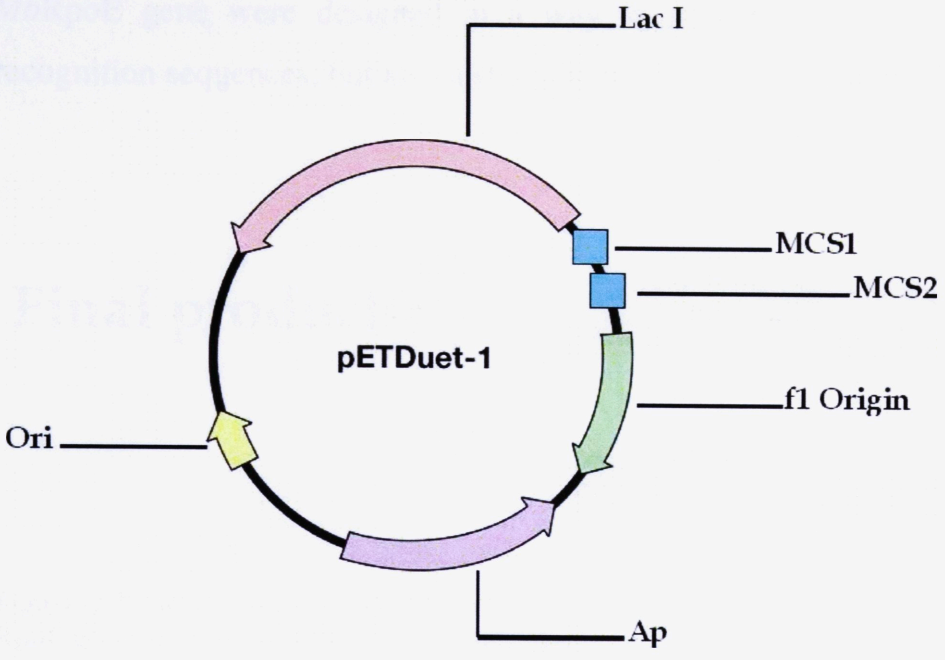


Figure 2. 1: Graphical representation of the pETDuet™-1 vector.

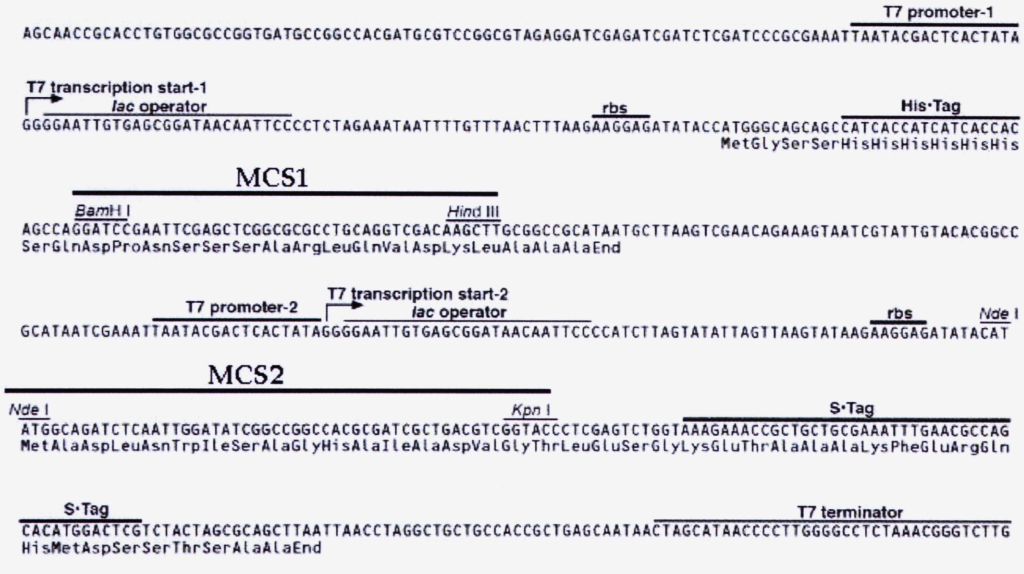


Figure 2. 2: pETDuet™-1 cloning/expression regions.

The amplified gene for *Mb*RpoF was cloned into the pETDuet™-1 vector via the *Bam*HI and *HinD*III restriction site of MCS1, while the amplified gene for *Mb*RpoE was cloned via the *Nde*I and *Kpn*I restriction sites of MCS2. The primers for the amplification of the

*Mb*RpoE gene were designed in a way to not only add the two restriction enzymes recognition sequences, but also a stop codon at the end of the gene (Invitrogen).

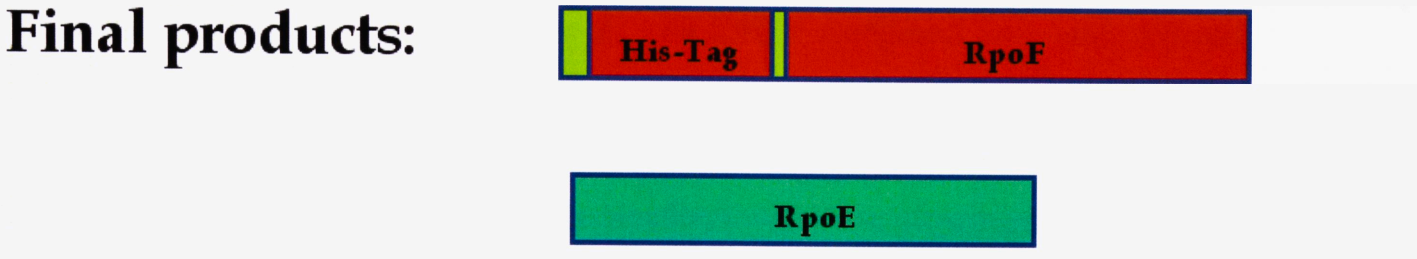


Figure 2.3: Scheme of the protein products deriving from the cloning of the *M. burtonii* RpoE and RpoF genes into the pETDuet™-1 as described in 2.2.5.

The pETDuet™-1 vector was digested with *Bam*HI and *Hin*DIII (NEB). The same enzymes and conditions were used to separately digest the RpoF gene amplicons. The digestion reactions were incubated at 37 °C for 2 h and analysed by agarose gel electrophoresis.

The digested DNA fragments of interest were extracted from the agarose gel using the QIAquick® gel Extraction Kit.

Vector DNA (50 ng) and gene inserts (14 ng of insert corresponding to 1:3 vector:insert molar ratio) were ligated using 1.5 U of T4 DNA Ligase (Promega), 1 x T4 ligation buffer and 1 mM ATP. A ligation control reaction was also prepared to assess self-religation of the vector. The reactions were incubated at 4 °C overnight and transformed into *E. coli* DH5α competent cells. Transformants in the pETDuet™-1 vector were selected on LB + ampicillin (100 μg/ml).

Plasmid DNA was extracted from the *E. coli* cells using the *AccuPrep*® Plasmid Mini Extraction Kit (Bioneer). Recombinant clones were selected by endonuclease digestion and

colony PCR (using the primers listed in Table 2.1 for the amplification of the RpoF gene) and stored at –80 °C in glycerol stocks.

The pETDuet™-1 derived plasmid containing the RpoF insert was digested with *Nde*I and *Kpn*I (NEB). The same enzymes in the same conditions were used to digest the RpoE gene fragments. Ligation, transformation into *E. coli* DH5α competent cells and selected by endonuclease digestion and colony PCR were performed as above (the primers for the amplification of the RpoE gene were used, Table 2.1). Positive clones were stored at –80 °C in glycerol stocks.The constructs were verified by DNA sequencing (ABI 3730xl DNA sequencer, Ramaciotti Centre for Gene Function Analysis, The University of New South Wales).

2.2.6 *E. coli* Competent Cells and Transformation

Bacterial cells were streaked on LB agar plates and incubated overnight at 37 °C. A single colony was inoculated in liquid LB medium and grown overnight at 37 °C under agitation. The pre-culture was diluted 100 times in fresh LB and incubated at 37 °C for 2-3 h until it reached an OD between 0.5 and 0.6. At this point the culture was incubated on ice for 15 min and the cells pelleted at 1000 x *g* for 5 min at 4 °C.

The supernatant was discarded and the cell pellet was resuspended in ½ culture volume of ice cold sterile 100 mM $CaCl_2$ and incubated 30 min in ice. A further centrifugation step was performed in the same condition and the pellet was dissolved in 1/10 culture volume of the same ice cold 100 mM $CaCl_2$.

The bacterial suspension was dispensed in 300 µl aliquots and each of these aliquots was mixed with 5 µl of ligation reaction, incubated in ice for 30 min, heat shocked at 42 °C for 2 min and cooled down in ice again for 60 sec.

To the cells aliquots was added 1 ml of LB and the mix was incubated at 37 °C for 45 minutes. Transformants were selected on LB plates incubated at 37 °C overnight.

The transformation efficiency was calculated by transforming 300 µl of competent cells with 1 ng of control plasmid and plating different volumes on LB agar.

2.2.7 Dye-Terminator Sequencing

DNA sequencing reactions were carried out using the BigDye™ Terminator kit version 3.1 (Applied Biosystems; Scoresby, VIC, Australia). Unincorporated dye was removed by ethanol precipitation. The separation of labeled fragments of DNA was performed with an Automated DNA Sequence Analyser ABI 3730xl (Applied Biosystems) at the Ramaciotti Centre for Gene Function Analysis,, University of New South Wales. Sequencing data were analysed using the FinchTV v1.4 program available from www.geospiza.com.

2.2.8 Overexpression and Purification

***Mb*RpoE/F:** *E.coli* Rosetta carrying the *Mb*RpoE/F genes were grown in 500 ml cultures at 37 °C in LB medium in the presence of 1 mM ampicillin and 1 mM chloramphenicol. When the OD_{600} reached a value between 0.5 – 0.6 the growth temperature was lowered to 15 °C and protein expression induced with the addition of 1 mM IPTG. The cells were harvested after 4 hours by centrifugation at 4000 x *g* for 10 min at 4 °C and resuspended in 30 ml lysis buffer (20 mM Tris, pH 7.4, 500 mM NaCl, 10 mM imidazole). Cells were lysed using a French Press. Cell-free extracts were separated from insoluble fractions by centrifugation at 23000 x *g* for 20 min at 4 °C. The cell-free extracts were filtered through a 0.45 µm Millex™ Syringe Driven Filter Unit (Millipore) and incubated at 4 °C for 30 min in QIAGEN Ni-Nta resin previously equilibrated in lysis buffer. *Mb*RpoE/F was eluted in lysis buffer with increasing concentrations of imidazole (10 mM, 50 mM, 100 mM and 500

mM). *Mb*RpoE/F typically eluted between 50 mM and 100 mM. The purity of the proteins was analysed by 12 % (w/v) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The fractions containing *Mb*RpoE/F proteins were pooled, concentrated, aliquoted and flash-frozen in liquid nitrogen.

***Mj*RpoE/F:** *E.coli* BL21 Rosetta 2 carrying *Mj*RpoE/F genes were grown, induced and harvested in the same way described above, except that the temperature was kept at 37° C after the induction. The harvested cells were resuspended in 30 ml lysis buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, and 1.8 mM KH_2PO4) and lysed with a French Press. Cell-free extracts were separated from insoluble fractions by centrifugation at 23000 x *g* for 20 min at 4 °C. The cell-free extract was filtered through a 0.45 µm Millex™ Syringe Driven Filter Unit and incubated at 4 °C for 30 min with Glutathione-Sepharose Beads (GE-Healthcare) previously equilibrated in lysis buffer. After extensive washing with lysis buffer, *Mj*RpoE/F was eluted from the beads with the addition of 30U of bovine thrombin (SIGMA). The eluted protein was collected and analysed as described above.

2.2.9 Size Exclusion Chromatography

Size exclusion chromatography was performed on Superdex 200 10/300 GL column (Amersham Biosciences) fitted to an ÄKTA Basic 900 series Fast Protein Liquid Chromatography System. The system was controlled by Unicorn Software and the column was equilibrated in 50 mM HEPES/NaOH, pH 7.4, 100 mM NaCl, 1 mM DTT at a flow rate of 0.4 ml min-1 using a 0.2 ml sample loop. The column was calibrated with protein standards of known molecular weight and approximately 100 µg of *Mb*RpoE/F and *Mj*RpoE/F were applied in each independent run. Eluted proteins were detected at 280 nm.

2.2.10 SDS-PAGE

Proteins were qualitatively and quantitatively analysed by discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli (1970).

Samples were mixed with sample buffer (final concentration 0.375 M Tris, pH 6.8, 0.01% SDS, 20% glycerol, 40 mg ml^{-1} SDS, 31 mg ml^{-1} DTT, 1 µg ml^{-1} bromophenol blue). Samples were denatured by heating at 95 °C in a heating block for 5 min prior to loading. For every electrophoresis experiment, a standard high range molecular weight ladder (Mark12™ Unstained Standard, Invitrogen) was used. Proteins were separated on 12% SDS-PAGE gels by electrophoresis for 2 h at 100 V. The gels were stained with Coomassie staining solution.

2.2.11 Circular Dichroism

The far-ultraviolet (UV; 190–260 nm) CD spectra of *Mb* and *Mj*RpoE/F were recorded using a JASCO-Europe J-810CD spectropolarimeter under constant nitrogen flow, connected to a Peltier temperature controller (JASCO-Europe) set at 10 °C. Spectra were recorded using a 0.1 cm path length and a protein concentration of 0.1 mg ml^{-1}. All measurements were carried out at 10 °C in a solution containing 40 mM Hepes, pH 7.5. Spectra were averaged over 10 scans and corrected for buffer signal. Raw data were expressed in terms of the mean residue ellipticity (θ) using the following parameters: molecular weight, concentration, cell path length, and number of residues.

2.2.12 Protein Concentration

Protein concentrations of the purified *Mb* and *Mj*RpoE/F were determined from the UV absorbance at 280 nm as described by Pace *et al.* (1995).

2.3 Results and Discussion

2.3.1 PCR Amplification

M. burtonii genes RpoE and RpoF were amplified by Polymerase Chain Reaction (PCR) with the set of primers in order listed in Table 2.1 in order to generate restriction endonuclease sites for the cloning into the pETDuet™-1 vector.

The PCR reactions were analysed by electrophoresis on agarose gel (Fig. 2.4) to verify the size of the amplified products (575 and 390 bp for RpoE and RpoF, respectively).

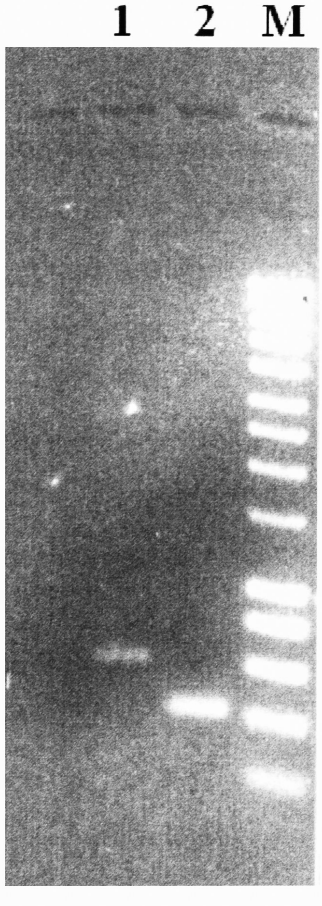


Figure 2. 4: PCR amplification of the *M. burtonii* genes. RpoE (lane 1) and RpoF (lane 2) were amplified with a set of primers for cloning them into pETDuet™-1 vector. DNA marker (M) bands: 10, 8, 6, 5, 4, 3, 2.5, 2, 1.5, 1, 0.8, 0.6, 0.4, 0.2 kb.

2.3.2 Cloning, Expression and Protein Purification

The classical cloning strategy described in 2.2.1 was successful and positive clones were produced. The sequencing of the RpoE/F genes in the pETDuet™-1 vector verified that the PCR reaction did not incorporate any mutations resulting from *Taq* polymersase errors. The recombinant vector was then transformed into the *E. coli* Rosetta (DE_3) strain which carries the plasmid pRARE. This plasmid encodes tRNA genes that supplement codons that are rarely used in *E. coli* (Sharp and Li, 1987; Sharp *et al.*, 1988; Sharp and Matassi, 1994).

The expression of *Mb*RpoE/F in Rosetta cells produced a soluble protein product that was easily purified with a Ni-Nta column as described in 2.2.8 (Fig. 2.6). The overall strategy was successful and the co-expression of the two subunits resulted in the production of a soluble protein product. The temperature was lowered to 15 °C after the induction, which has been proven to increase protein levels and limit proteolytic degradation of recombinant low-temperature adapted proteins (Thomas *et al.*, 2002).

Despite the induction temperature being considerably lower than normal, it did not contribute negatively on the yield of the protein product and the average quantity of *Mb*RpoE/F purified was 20 mg from a litre of *E. coli* culture.

The pGEX-2TK vector containing the *M. jannaschii* RpoE and RpoF genes was also transformed into the *E. coli* Rosetta (DE_3) strain and the recombinant heterodimer was expressed and purified as described in 2.2.8 (Fig.2.5). The quantity of protein product obtained from each purification was approximately 10 mg per liter of *E. coli* culture.

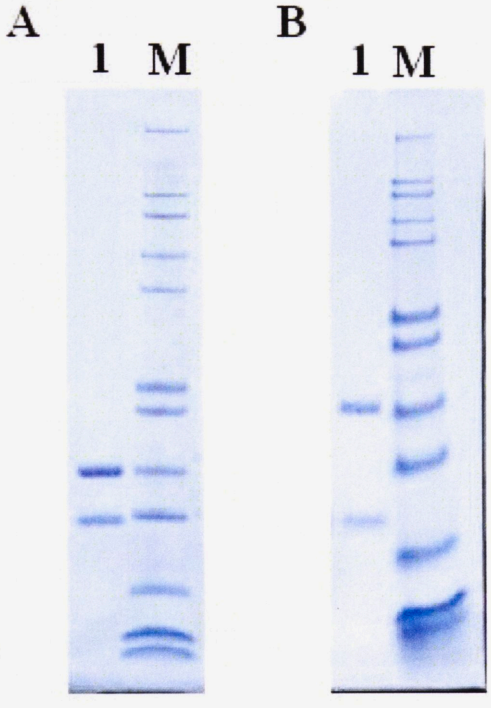


Figure 2. 5: SDS-PAGE of purified *Mb*RpoE/F (panel A, lane 1) and *Mj*RpoE/F (panel B, lane 1). Protein marker (M) bands: 200, 116.3, 97.4, 66.3, 55.4, 36.5, 31, 21.5, 14.4, 6, 3.5, 2.5, 0.4, 0.2 kDa.

2.3.3 Size Exclusion Chromatography

Gel filtration analysis showed that both heterodimers eluted as a single peak at an identical elution time, corresponding to a globular protein of approximate M_R of 35 kDa (Fig. 2.6). The theoretical molecular weight of the two homologues, based on their amino acid sequences, is 38 and 34 kDa for *Mb* and *Mj*RpoE/F, respectively. The elution times are compatible with these values considering that the elongated shape of the heterodimers may affect their elution characteristics on the gel filtration column. The fact that both *Mb* and *Mj*RpoE/F eluted as a single peak proves that the single monomers are properly folded and that their interactions together are strong and stable.

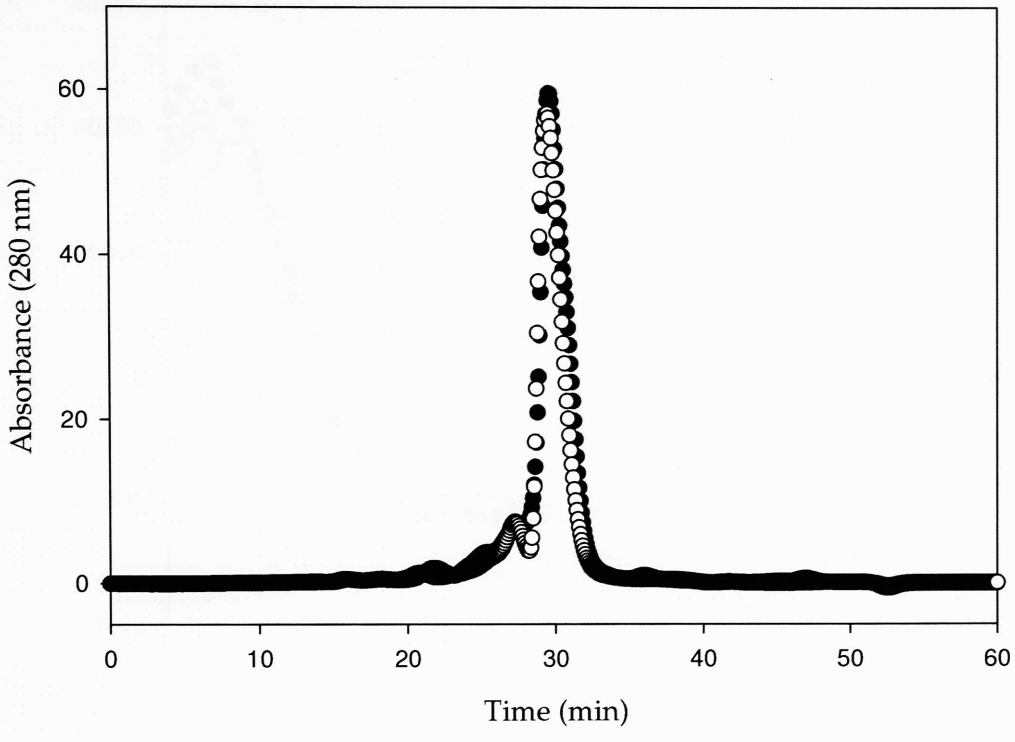


Figure 2. 6: Gel filtration profile of *Mb*RpoE/F (•) and *Mj*RpoE/F (◦). The proteins eluted as a single peak at an identical time corresponding to the apparent MW of 35 kDa, compatible with the theoretical MW of the heterodimeric form of the two subunits.

2.3.4 Circular Dichroism

The Far-UV CD spectra for both heterodimers at 10 °C displayed a maximum peak at 195 nm and two minimum peaks at 205 and 222 nm. These characteristics confirm that the heterodimers were successfully purified in a folded state and also that they present α-helical and β-sheet secondary structures (Siddiqui *et al.*, 2005) (Fig. 2.7). The spectra are almost superimposible, suggesting high structural similarity.

Circular dichroism was not used to study the thermal unfolding of the two homologues because the instrument could not operate at temperatures above 100 °C and therefore was not suitable for the analysis of *Mj*RpoE/F. The thermal denaturation of the both heterodimers was therefore studied by differential scanning calorimetry (DSC, Section 3.2.1).

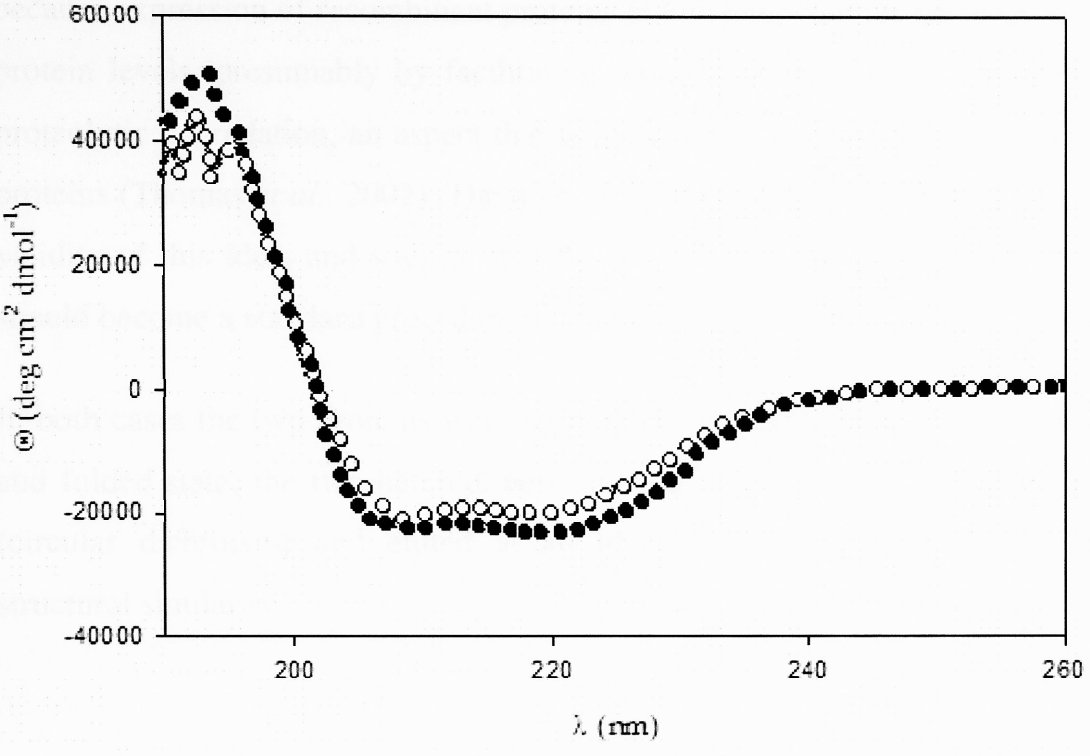


Figure 2. 7: CD spectra in the far-UV at 10 °C. *Mb*RpoE/F (•) and *Mj*RpoE/F (◦).

2.4 Conclusion

M. burtonii RpoE and F genes were successfully cloned into the pETDuet™-1 expression vector: the gene for the subunit F was cloned in the MCS1 in order to give a His-Tagged protein product, while the subunit E was cloned in the MCS2 and the fusion to the S-tag was avoided. For the expression of *Mj*RpoE/F the same recombinant pGEX-2TK vector used by Werner *et al.* (2000) was employed.

Both recombinant plasmids were expressed in *E. coli* Rosetta (DE_3) strain, which contains the pRARE plasmid, which over-expresses the tRNAs for Arg, Leu, and Pro codons rarely used by *E. coli.*

The expression of *Mb*RpoE/F was induced at 15 °C and not at 37 °C as for *Mj*RpoE/F because expression of recombinant proteins at low temperatures has been shown to increase protein levels, presumably by facilitating protein folding and by reducing the amount of proteolytic degradation, an aspect that is particularly relevant for low-temperature adapted proteins (Thomas *et al.*, 2002). The high yield of soluble *Mb*RpoE/F produced confirms the validity of this idea, and suggest that the lowering of the temperature after the induction should become a standard procedure when expressing a cold-adapted recombinant protein.

In both cases the two proteins were successfully co-expressed and co-purified in a soluble and folded state; the two heterodimers showed a very similar secondary structure profile (circular dichroism) and eluted at an identical time (gel filtration), suggesting high structural similarity.

Chapter 3: Structure and Functional Analysis of the RpoE/F Proteins

3.1 Introduction

Once the recombinant *Mb* and *Mj*RpoE/F were successfully purified in a soluble state, it was possible to perform comparative stability, structural, and functional analyses on both heterodimers and obtain insights into their adaptation to different thermal constrains. *Mj*RpoE/F was chosen for this comparison against *Mb*RpoE/F in order to maximise thermal and minimise phylogenetic differences. (*M. burtonii* can tolerate sub-zero temperatures, whereas *M. jannaschii* is still viable above 100 °C and both are methanogenic archaea). In this way all the structural and functional differences that may arise from this study are more likely to reflect thermal adaptation rather than other effectors of natural selection.

The stability of both homologues was studied by **differential scanning calorimetry (DSC)**, and the accessibility of their single tryptophan was tested by **dynamic fluorescence quenching (DFQ)**. The tryptophan accessibility to a quencher can be viewed as an index of protein permeability and thus flexibility.

The data obtained by these analyses confirmed what is generally known about thermally adapted proteins: the psycrophilic homologue is far less stable and more flexible than its thermophilic counterpart (D'Amico *et al.*, 2002, Georlette *et al.*, 2004 Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007).

In order to explain the different thermal properties of the two homologues, a thorough cristallisation strategy was employed to obtain *Mb*RpoE/F crystals suitable for X-ray crystallography with the goal of comparing this structure to that of *Mj*RpoE/F to find structural determinants at the atomic level.

The formation of well-diffracting crystals is generally less successful for low-temperature adapted proteins due to their intrinsic flexibility. High structural flexibility translates into less ordered packing thus hindering cristallisation (Russell, 2000). For these reasons only a few proteins from psychrophilic organisms have been cristallised so far, including one very recently from *M. burtonii* (Byrne-Steele *et al.*; 2009) as compared to the ever growing database of protein structures from mesophiles and thermophiles.

Although it was possible to generate large diffracting crystals of *Mb*RpoE/F, the nature of these crystals did not allow us to solve the structure with reasonable resolution. Nevertheless it was possible to generate a model for *Mb*RpoE/F and infer some conclusions about the structural similarities and differences between the two homologues.

The most interesting findings regarding the two homologues derive from functional comparison. **Electrophoretic mobility shift assay** (EMSA) was used to assess the nucleic binding capability of the two homologues. The results showed that the heterodimers bind to different ssDNA homopolymer sequences with differing specificity. From the above mentioned data it was demonstrated that both homologues could bind single strand nucleic acid in a sequence-specific manner.

3.2 Materials and Methods

3.2.1 Differential Scanning Calorimetry (DSC)

Freshly purified RpoE/F proteins were concentrated to 1 mg ml^{-1} in 40 mM HEPES/NaOH, pH 7.4, 100 mM NaCl, 1 mM β-merchaptoethanol, 1 M 3-(1-pyridinio)-1-propanesulfonate (NSDB) using Amicon® Ultra centrifugal filter devices (15 ml capacity, 10 kDa cut-off). The flow-through at the bottom of the centricon tube was collected and kept as a reference for the measurements (Ashutosh *et al.*, 2000).

NDSB is a non-detergent sulfobetaine that prevents proteins from aggregating. If aggregation takes place, endothermic unfolding and exothermic aggregation/ precipitation occur simultaneously, resulting in a lowering of the apparent melting point *T*m. Addition of NDSB delays or even abolish this effect and thereby improves the accuracy of the measurements (Collins *et al.*, 2006).

Thermograms of both samples were recorded using a Microcal VP-DSC with a cell volume of 0.625 ml and a pressure of 27-28 psi. Prior to scanning, all solutions were centrifuged at 21000 x *g* for 10 minutes at 4 °C to remove particulate matter and then degassed with stirring under vacuum for 15 min at 10 °C using Thermovac Unit (MicroCal).

The cells of the microcalorimeter were prepared by repeated washing with Milli-Q water and reference buffer. Protein sample and buffer reference were loaded into the cells using a Hamilton syringe and particular attention was paid to avoid the introduction of air bubbles.

Cell volumes in both cells were leveled using a Hamilton syringe with a collar and the cells were subsequently sealed under pressure with a screw cap.

Run parameters were controlled using the VP-DSC software (MicroCal).

For each run the calorimeter was set to perform a series of scans with identical parameters (from 10 °C to 120 °C and then down to 10 °C before starting the following scan), with an initial temperature equilibration step of 15 minutes) and a scan rate of 1 Kelvin per minute. In this way, with the exception of the first scan, whose cycle had started from room temperature, all the others would have had the same thermal history.

Thermograms were analysed according to a non-two-state model using Origin Software.

Data analysis was performed with the Origin VP-DSC analysis software. Buffer baseline was subtracted from the sample data and the data were normalised with respect to the protein concentration.

The deconvolution of the thermal transition curves was performed with different transition models incorporated in the Origin Software package and the enthalpy of the unfolding process was determined by integration of the transition peak. Maximum heat capacity (C_p) values and their corresponding temperatures T_m were determined manually with the selection tools of the Origin software.

In order to further analyse the denaturation process a simple kinetic model for the thermal transition of the RpoE/F proteins was used, where the native protein N undergoes an endothermic and irreversible step to a denatured state D with a first order rate constant k

$$N \xrightarrow{k} D$$

According to this model the rate constant (k_{unfol}) of the reaction at a given temperature t can be calculated by the formula:

$$k_{unfol} = (v * C_p)\ (Q - Q_t)$$

where v is the scan rate, C_p is the excess heat capacity at a given temperature t, Q_t being the heat evolved at a given t and Q being the total heat evolved during the process. The values of k (as ln k) were plotted against the inverse of the absolute temperature (in Kelvin) resulting in a linear relationship. According to the Arrhenius equation ($k_{unfol} = A\ e^{(-E/RT)}$), the slope of the straight line corresponds to –E/R, where E is the activation energy of the unfolding process, R is the Universal Gas Constant ((8.314 J K^{-1} mol^{-1}) and A an arbitrary factor assumed to be 1.

Half-life of unfolding ($t_{1/2\text{-}unfol}$) = $\ln 2/k_{unfol}$

The thermodynamic activation parameters for the kinetic stability of both homologues were determined, using

$\Delta G^{\#}$ (free energy of activation) = $-RT \ln \{(k_{unfol} \times h)/(k_B \times T)\}$

where T is the absolute temperature, k_{unfol} is the first-order rate constant of unfolding determined by DSC, *h* is the Planck Constant (6.63×10^{-34} J s), k_B is the Boltzman Constant (1.38×10^{-23} J K^{-1}).

$\Delta H^{\#}$ (enthalpy of activation) = E_a –RT

where E_a is the activation energy determined from the slope of an Arrhenius plot by plotting ln k_{unfol} *versus* 1/T.

$\Delta S^{\#}$ (entropy of activation) = $(\Delta H^{\#} - \Delta G^{\#})/ T$

3.2.2 Dynamic Fluorescence Quenching (DFQ)

Experiments were performed on a Perkin-Elmer Luminescence Spectrometer LS-50B at 10 °C and 40 °C. Both proteins were excited at 280 nm and the fluorescence intensities emitted at 330 nm were recorded in the presence of increasing amounts of acrylamide (Georlette *et al.*, 2003; Cavicchioli *et al.*, 2006; Siddiqui and Cavicchioli, 2006). Excitation and emission slit widths were set at 1 nm and 4 nm respectively. Samples were prepared in 40 mM HEPES/NaOH, pH 7.6 buffer, 100 mM NaCl. Measurements were taken following consecutive additions of a 5 M acrylamide stock solution to 2 ml protein solution in order to increase acrylamide concentration by ~ 0.3 mM steps.

The Stern-Volmer quenching constants K_{SV} were calculated using the following equation:

$$F_0/ F = 1 + K_{SV} [Q]$$

where F is the fluorescence intensity in the presence of the molar concentration of the quencher Q and F_0 is the fluorescence intensity in its absence. The plot of F_0 /F against the acrylamide concentration is a straight line whose slope equals K_{SV} (Siddiqui *et al.*, 2006).

3.2.3 Cristallisation and Data Collection

In order to cristallise, a protein must be brought to the metastable condition, where solubility is maintained but the molecules are thermodynamically more stable in the crystal form. Under right conditions nucleation occurs followed by crystal growth. For every protein, exact conditions (pH, initial protein concentration, type of solvent and precipitant, presence of additives, temperature, buffer, etc) must be found empirically and are impossible to predict *a priori*. A common simple approach to find the right conditions is the sparse matrix screening method. This approach helps to minimise the amount of protein needed for obtaining good crystals for diffraction.

Screening usually gives the first "hits" and then conditions can be optimised in order to produce high quality crystals.

The addition of an appropriate precipitant and/or salt can assist this process by competing for water molecules to a point of supersaturation or a non-equilibrium state. Return to non equilibrium state may promote crystal growth.

Cristallisation trials of *Mb*RpoE/F were screened using the hanging drop vapour diffusion method with an 80 µl reservoir in Hampton 96 well plates. Drops consisted of 0.2 µl of protein and 0.2 µl of reservoir solution. Extensive optimisation trials were undertaken using a variety of techniques including additive screens, seeding and the addition of microporous glass.

Optimisation trials were screened with a 500 µl reservoir in Hampton 24 well plate. Drops consisted of 3 µl of protein at 12.7 mg ml^{-1} and 3 µl of reservoir solution.

3.2.4 Sequence Alignment and Homology Modeling

The amino-acid sequences were aligned using the program CLUSTALW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Aiyar, 2000). Protein modelling was performed with the automated modelling program of the SWISS-MODEL (Schwede *et al.*, 2003) server (http://swissmodel.expasy.org/). The Protein Database Brookhaven (PDB) entry file used to generate the model was 1GO3 corresponding to *Mj*RpoE/F. The *Mb*RpoE/F model and the *Mj*RpoE/F structure were visualised and studied using the programs Swiss-PdbViewer and Pymol.

For the analysis of the hydrophobic and electrostatic interaction WHATIF (http://swift.cmbi.kun.nl/whatif/) was used (Vriend, 1990), for aromatic-aromatic interactions PIC (http://crick.mbu.iisc.ernet.in/~PIC) was used (Rodriguez *et al.*, 1998; Tina *et al.*, 2007) and for Pi-cation interactions the programs CaPTURE (http://capture.caltech.edu) was used (Gallivan and Dougherty, 1999).

3.2.5 Electrophoretic Mobility Shift Assay (EMSA)

Varying amounts of the two heterodimers were incubated for 30 min on ice with 5 fmol of ^{32}P end-labeled oligonucleotides in 10 µl of buffer containing 40 mM HEPES/NaOH, 100 mM potassium acetate and 10% glycerol, pH 7.4. The ssDNA sequences that were tested under these conditions were 25-mers of poly-dA, poly-dC, poly-dG, poly-dT. Bound and free probes were resolved by electrophoresis in a 5% polyacrylamide gel containing 10% glycerol and 0.5x TBE. Radiolabeled DNA was visualised by autoradiography (Sinclair *et al.*, 1994).

3.3 Results and Discussion

3.3.1 DSC

The stability of the purified proteins was studied by **differential scanning calorimetry (DSC)**, a thermoanalytical technique in which the amount of heat required to increase the temperature of a sample and its reference are measured during a controlled temperature scan. With this technique it is therefore possible to directly measure the heat capacity during temperature ramping and to determine enthalpy change as a result of protein unfolding. At the scan rate of 1 K per minute symmetric endothermic thermal transitions were observed (Fig. 3.1).

For *Mb*RpoE/F the thermal transition was initiated around 40.2 °C and the protein was completely unfolded at 62.1 °C. *Mj*RpoE/F showed the initiation of unfolding at 89.1 °C with the transition being completed at 103.1 °C. The temperature values corresponding to the maximum excess heat capacity (T_m) were 50.9 °C and 96.1 °C for *Mb* and *Mj*RpoE/F, respectively. The rescan of the protein samples after they were heated beyond the transition peak showed no increase in heat capacity (data not shown). This indicated that the unfolding of the two RpoE/F proteins from both organisms is an irreversible process.

In order to further analyse the denaturation process, a simple kinetic model for the thermal transition of the proteins was used, where the native protein N undergoes an endothermic and irreversible step to a denatured state with a first order rate constant k. The values of k (evenly distributed over the thermal transition curve before T_m) were calculated and then plotted according to Arrhenius (Fig. 3.2).

The data gave a linear relationship and therefore allowed for the determination of the activation energy of the unfolding process (E_a) as described in Materials and Methods.

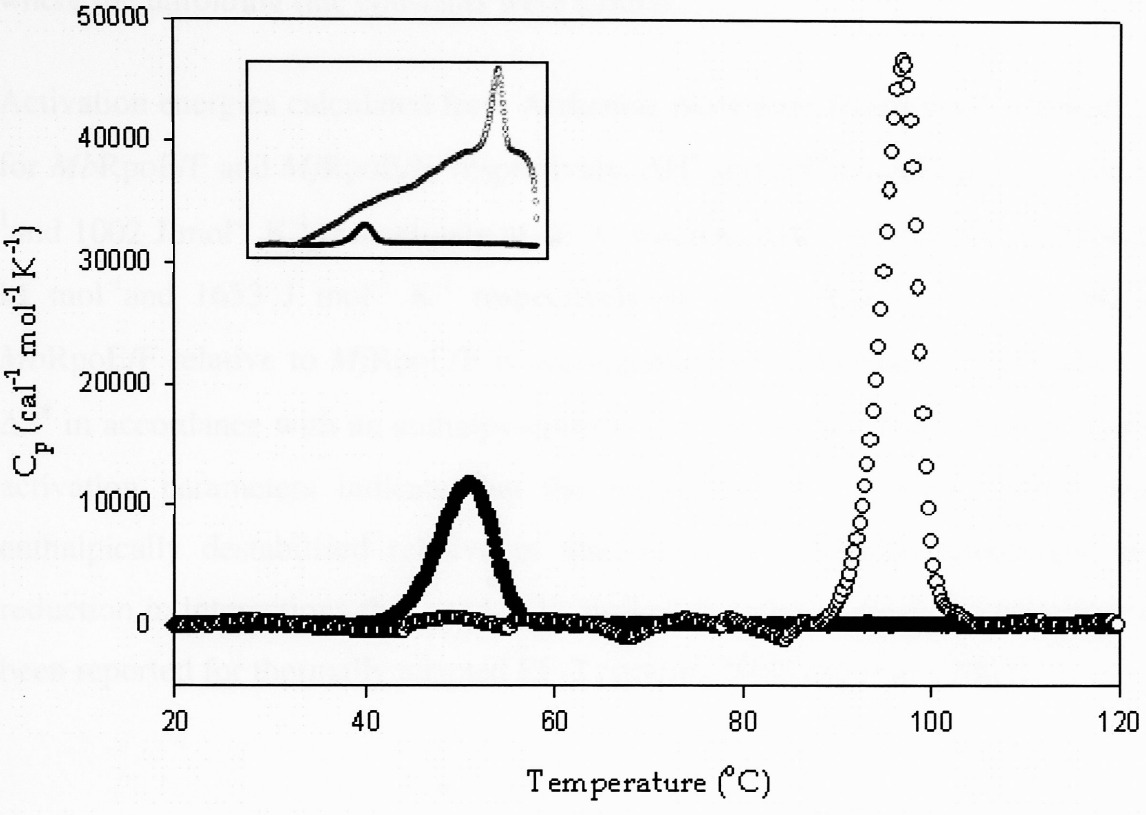


Figure 3. 1: Excess heat capacity (C_p) of the RpoE/F proteins from *M. burtonii* (•)and *M. jannaschii* (◦) versus temperature at the scan rate of 1 K per min. Measurements were performed in 40 mM HEPES/NaOH, pH 7.4 100 mM NaCl, 1 mM β-merchaptoethanol, 1 M NDSB. Inlet graph: thermograms before baseline correction.

Values for E_a of 420 kJ mole^{-1} and 720 kJ per mole^{-1} were calculated for *M. burtonii* and *M. jannaschii*, respectively, which indicates that *Mb*RpoE/F requires less heat to form partially unfolded transition-state as compared to its thermophilic homologue from *M. jannaschii* (D'Amico *et al.*, 2003; Sanchez-Ruiz *et al.*, 1998).

To examine whether the lower kinetic stability of *Mb*RpoE/F (for k, 0.05 min^{-1} and $t_{1/2}$, 13.5 min at 46 °C) relative to *Mj*RpoE/F (for k, 0.05 min^{-1}and $t_{1/2}$, 13.5 min at 93 °C) is attributed to a reduction in activation enthalpy ($\Delta H^{\#}$) or activation entropy ($\Delta S^{\#}$), thermodynamic activation parameters of unfolding were derived from the DSC thermograms. Due to the fact that both proteins unfolded in completely different

temperature ranges, these parameters were determined close to their respective T_m values where the unfolding rate constants were similar.

Activation energies calculated from Arrhenius plots were found to be 420 and 720 kJ mol^{-1} for *Mb*RpoE/F and *Mj*RpoE/F, respectively. $\Delta H^{\#}$ and $\Delta S^{\#}$ for *Mb*RpoE/F were 417 kJ $mol^{-1}$and 1002 J mol^{-1} K^{-1} respectively at 46 °C whereas $\Delta H^{\#}$ and $\Delta S^{\#}$ for *Mj*RpoE/F were 717 kJ $mol^{-1}$and 1653 J mol^{-1} K^{-1} respectively at 93 °C. The decrease in the stability of *Mb*RpoE/F relative to *Mj*RpoE/F is accompanied by a decrease in the values of $\Delta H^{\#}$ and $\Delta S^{\#}$ in accordance with an enthalpy-entropy compensation. The values of thermodynamic activation parameters indicate that the transition-state of psychrophilic homologue is enthalpically destabilised relative to that of the thermophilic homologue suggesting a reduction in interactions that need to be broken in order to reach the transition state as has been reported for thermally adapted EF-2 proteins (Siddiqui *et al.*, 2002).

3.3.2 DFQ

The accessibility of the single tryptophan in *Mb*RpoE/F and *Mj*RpoE/F was tested by **dynamic fluorescence quenching (DFQ)**. In this technique increasing concentrations of a small quencher molecule (in this case acrylamide) decreases the fluorescence of the protein molecule. This effect reflects the ability of the quencher to penetrate inside the protein molecule and can be viewed as an index of protein permeability and thus flexibility.

The Stern-Volmer plots for both heterodimers at 10 °C and 40 °C (Fig. 3.3) clearly show a higher permeability for *Mb*RpoE/F compared to *Mj*RpoE/F at both temperatures.

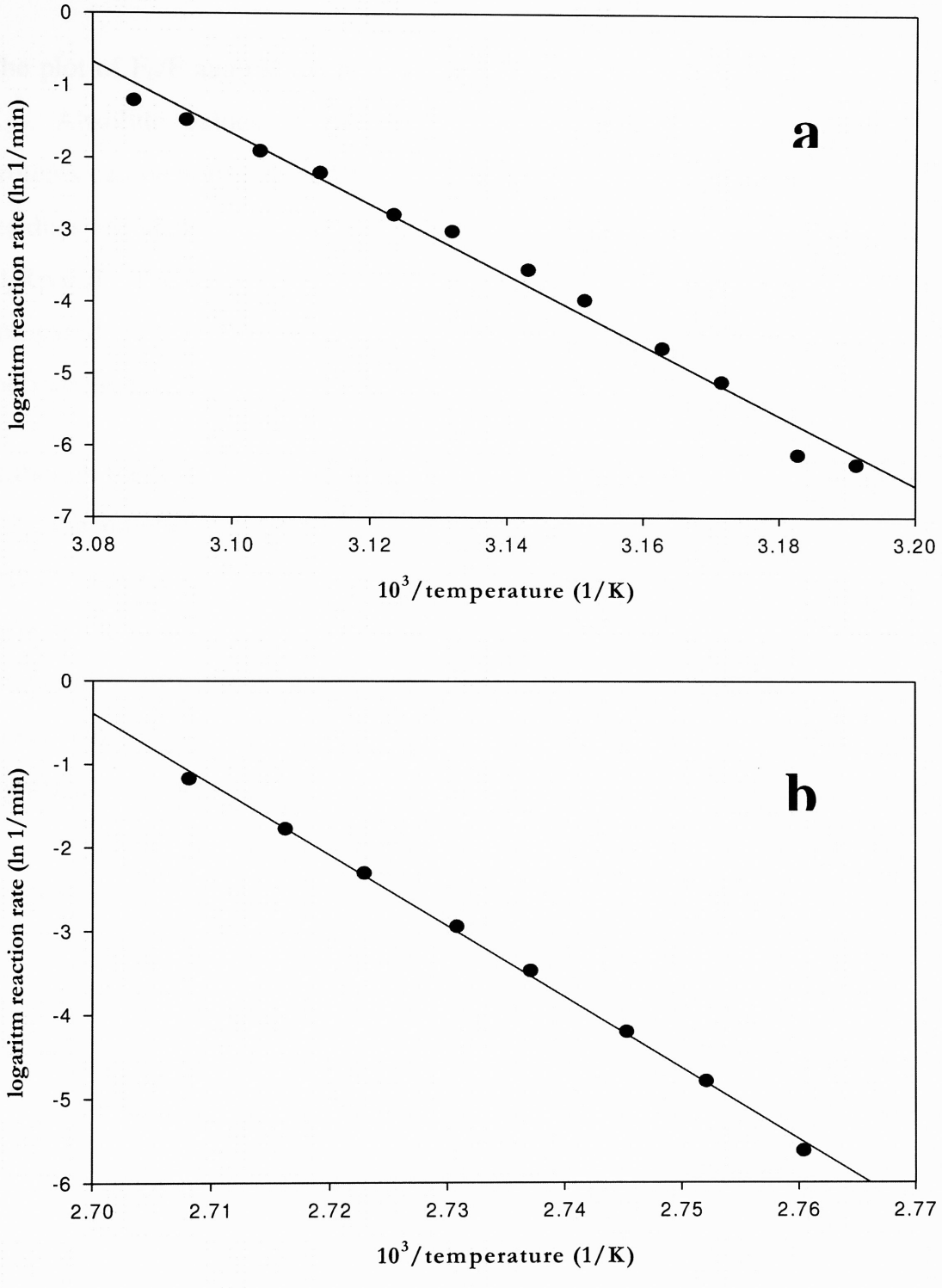


Figure 3. 2: Arrhenius plot for the reaction rate of thermal denaturation for *Mb*RpoE/F (a) and *Mj*RpoE/F (b).

The plot of F_0/F against the acrylamide concentration is a straight line whose slope equals K_{SV}. Absolute values of the Stern-Volmer quenching constants (K_{SV}) from different proteins can be compared only if the number, location and environment of the tryptophane residues are identical in each protein, conditions that are met in the case of *Mb*RpoE/F and *Mj*RpoE/F. The difference in permeability between the two proteins is higher at 40 °C as compared with 10 °C, indicating that the psychrophilic protein is more flexible than its thermophilic homologue at higher temperatures (Fig. 3.4).

Although the lone Trp residue was in identical position in *Mb*RpoE/F and *Mj*RpoE/F (Fig. 3.3), the number and positions of Tyr residues were very different in both proteins. In order to normalise the contribution of Tyr residues so that the acrylamide penetration truly reflects the flexibility of both proteins, the regression lines of Stern-Volmer plots at higher temperature (40 °C) were subtracted from the regression lines of Stern-Volmer plots at lower temperature (10 °C). The results clearly show *Mb*RpoE/F (0.02 mM^{-1}) to be significantly more flexible than *Mj*RpoE/F (0.009 mM^{-1}).

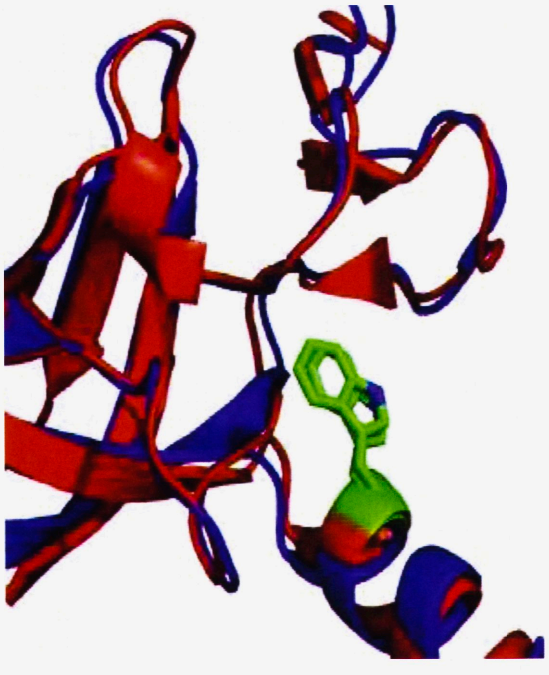


Figure 3. 3: Detail of the superimposition of the *Mj*RpoE/F structure (red) and the *Mb*RpoE/F homology model (blue). Trp 175 (E subunit) is in the identical position in both homologues.

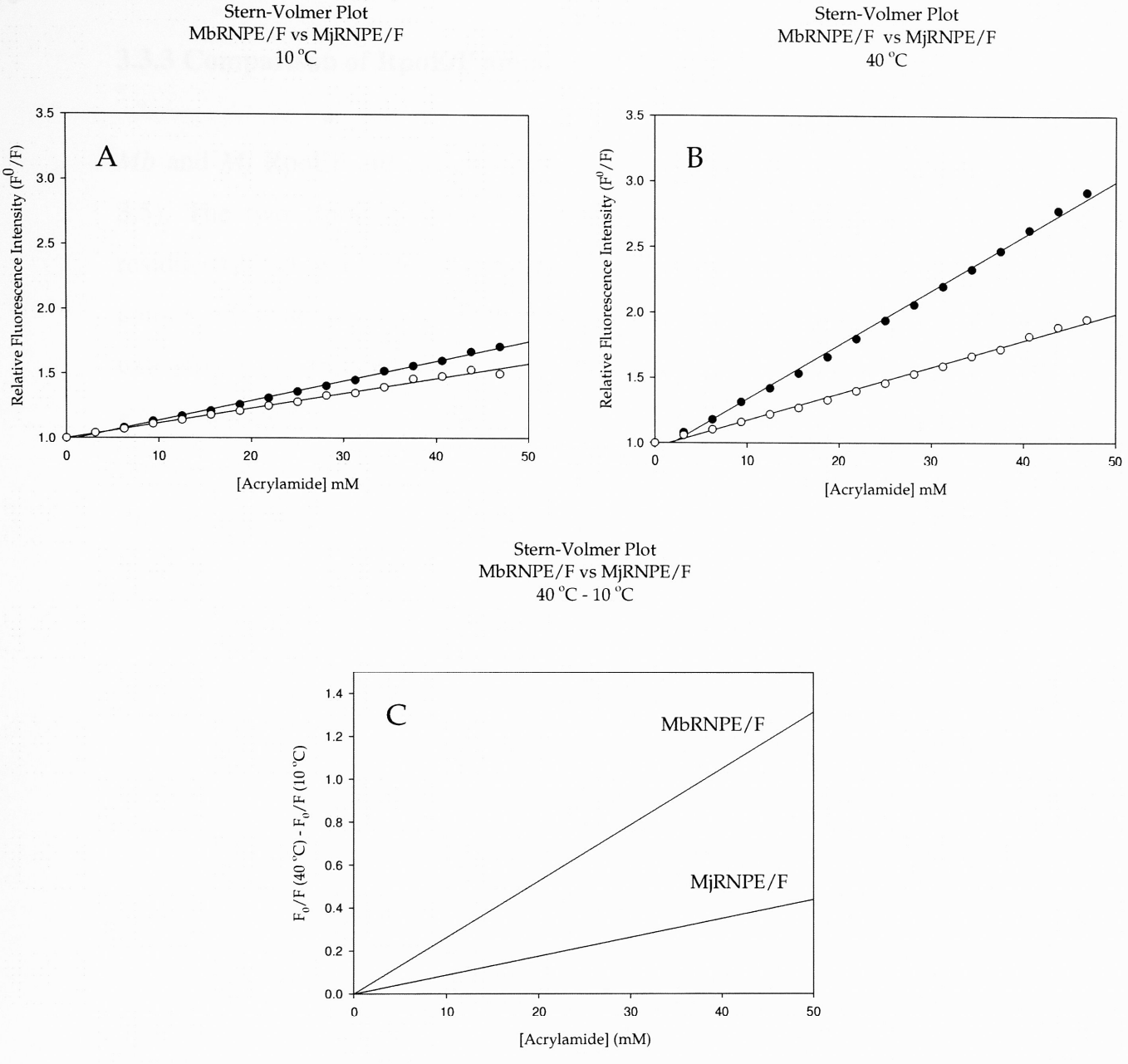


Figure 3.4: Stern-Volmer plots of fluorescence quenching by acrylamide for *Mb*RpoE/F and *Mj*RpoE/F at 10 and 40 °C. Fluorescence quenching values at 10 °C (*A panel*) and at 40 °C (*B panel*) for *Mb*RpoE/F (◦) and *Mj*RpoE/F (•) are shown. *C panel*: variation of fluorescence quenching between 10 and 40 °C obtained by subtracting the regression lines of Stern-Volmer plots at individual temperatures.

The corresponding quenching constant (K_{SV}) are 0.015 and 0.011 mM^{-1} at 10 °C and 0.0416 and 0.0204 mM^{-1} at 40 °C for *Mb*RpoE/F and *Mj*RpoE/F, respectively.

3.3.3 Comparison of RpoE/F amino-acid sequences

Mb and *Mj* RpoEF are closely related as shown by amino-acid sequence alignment (Fig. 3.5). The two RpoE homologues share 89 identical residues (48%) and 48 conserved residues (25%), while the two RpoF homologues share 38 identical residues (36%) and 30 conserved residues (28%). *Mb*RpoF shows an extra N-terminal short sequence and extensive gaps. In RpoE, the identical residues are concentrated in the S1 motif and in the RNP truncated region, which are the putative nucleic acid binding sites (Fig. 5.6). This implies that the direct interaction of the RNA with the protein has not been affected by thermal adaptation. *Mb*RpoE and *Saccharomyces cerevisiae* Rbp7 share 28% identical residues, again concentrated in the nucleic acid binding site.

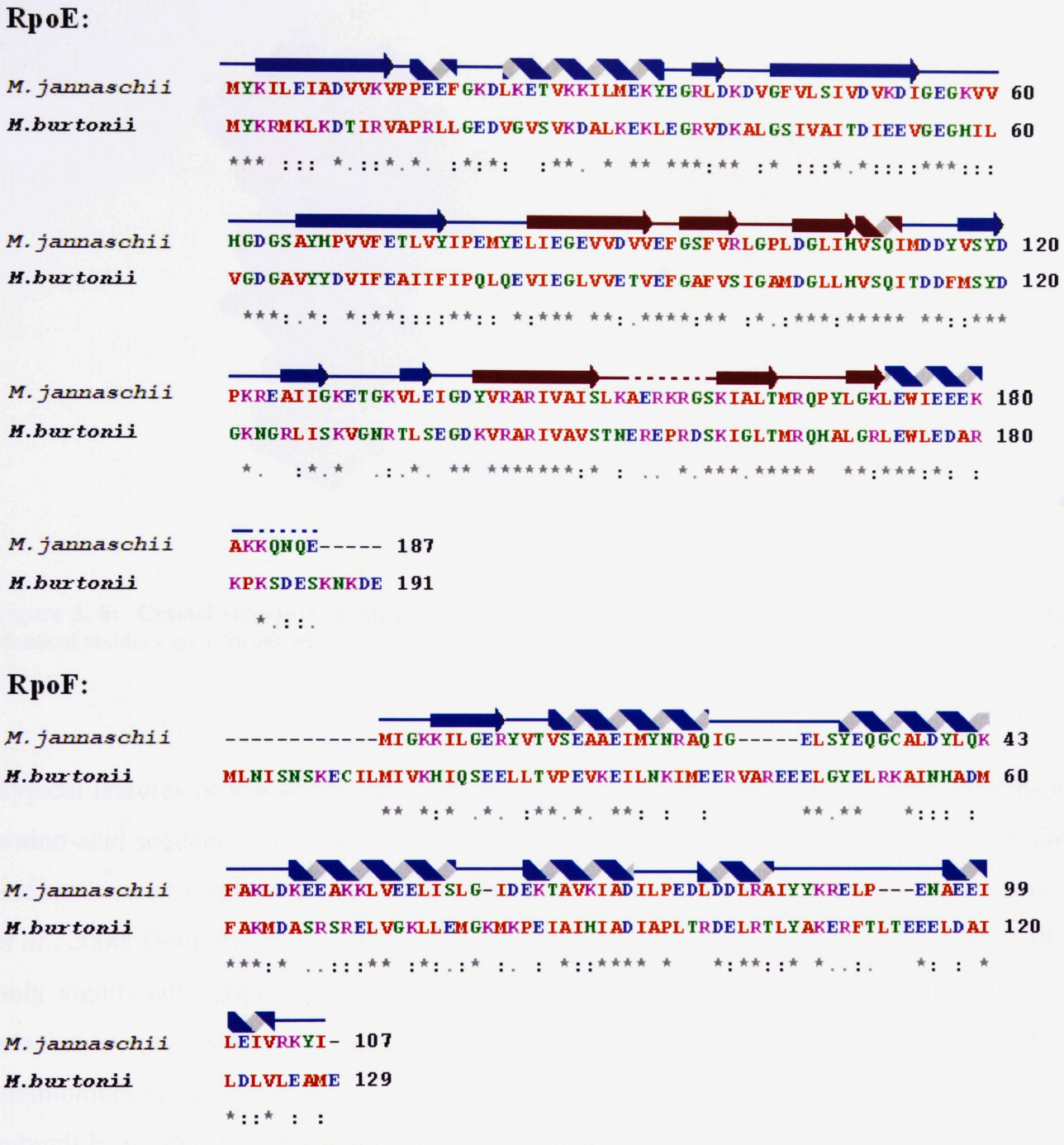


Figure 3. 5: Alignment of the RpoE/F amino-acid sequences from *M. burtonii* and *M. jannaschii*. The position of the secondary structure elements in the *M. jannaschii* E and F subunits are shown above the sequence. The elements of secondary structure in brown are proposed to be involved in the nucleic acid binding. A dashed line indicates residues omitted from the final model.

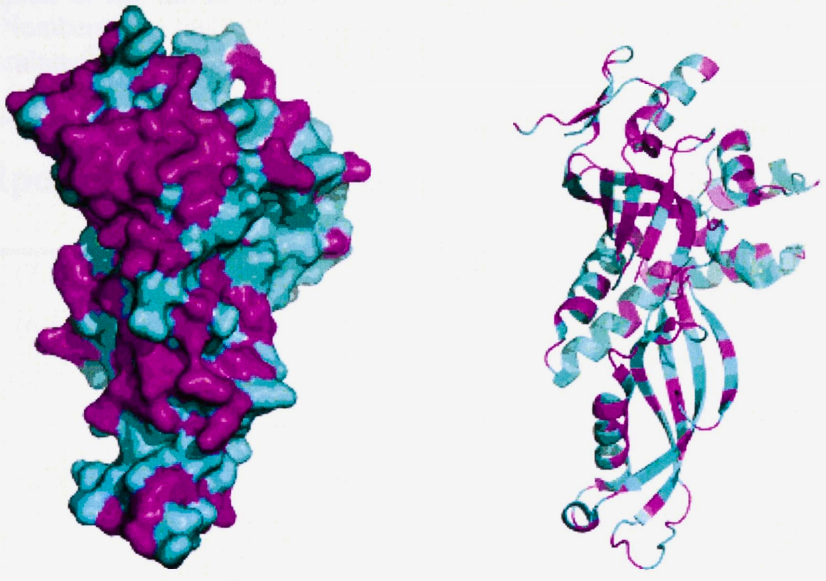


Figure 3. 6: Crystal structure of *Mj*RpoE/F. Identical residues *Mb*RpoE/F are highlighted in pink. The identical residues are concentrated in the two RNA binding motifs.

Typical features of low-temperature adapted proteins derived from the comparison of their amino-acid sequences with mesophilic and thermophilic homologues include more glycine and methionine residues, a low Arg/Lys and a high (Glu + Asp) / (Lys + Arg) ratio (Smalas *et al.*, 2000; Georlette *et al.*, 2004; Leiros *et al.*, 2000; Siddiqui and Cavicchioli, 2006). The only significant change in the amino-acid composition that could be linked to thermal adaptation was found to be methionine (Table 3.1). *Mb* and *Mj* RpoE/F have 13 and 7 methionines respectively; 3 are conserved (1 and 165 in *Mb* and *Mj*RpoE, the third is in the subunit F: residue 13 for *Mb* and 1 for *Mj*). *Mb*RpoF presents 6 more methionines (13, 37, 60, 64, 79, 128 and 82) than *Mj*RpoF. The high entropy and the lack of interacting groups (branches, charge or dipoles) associated with this residue might produce a more flexible protein better able to function at low temperature.

Table 3.1: Comparison of the amino-acid composition of the RpoE/F heterodimers from *M. burtonii* and *M. jannaschii*. Numbers give total amino-acid residues and values in brackets represent the percentage of the total number of amino-acid residues.

Residue	*Mb*RpoE/F	*Mj*RpoE/F	*Mb*RpoE	*Mj*RpoE	*Mb*RpoF	*Mj*RpoF
Ala A	23 (7.2)	18 (6.1)	12 (6.2)	8 (4.3)	11 (8.5)	10 (9.3)
Arg R	21 (6.5)	13 (4.4)	13 (6.3)	8 (4.3)	8 (6.2)	5 (4.6)
Asn N	8 (2.5)	3 (1)	4 (2.1)	1 (0.5)	4 (3.1)	2 (1.8)
Asp D	22 (6.8)	20 (6.8)	16 (8.3)	13 (6.9)	6 (4.6)	7 (6.5)
Cys C	1 (0.3)	1 (0.3)	0	0	1 (0.8)	1 (0.9)
Glu E	39 (12.1)	38 (13)	18 (9.3)	22 (6.1)	21 (16)	16 (14.9)
Gln Q	5 (1.5)	7 (2.4)	4 (2.1)	4 (6.1)	1 (0.8)	3 (2.8)
Gly G	21 (6.5)	21 (7.1)	18 (9.2)	16 (8.5)	3 (2.3)	5 (4.6)
His H	6 (1.9)	3 (1)	3 (1.6)	3 (1.6)	3 (2.3)	0
Ile I	27 (8.4)	28 (9.5)	15 (7.8)	16 (8.5)	12 (9.3)	12 (11.2)
Leu L	35 (10.9)	28 (9.5)	17 (8.8)	15 (8)	18 (14)	13 (12.1)
Lys K	25 (7.8)	31 (10.5)	15 (7.8)	20 (10.7)	10 (7.7)	11 (10.2)
Met M	13 (4)	7 (2.4)	5 (2.6)	5 (2.7)	8 (6.2)	2 (1.8)
Phe F	7 (2.2)	9 (3)	5 (2.6)	5 (2.7)	2 (1.5)	1 (0.9)
Pro P	7 (2.2)	9 (3)	4 (2.1)	7 (3.7)	3 (2.3)	2 (1.8)
Ser S	16 (5)	10 (3.4)	11 (5.7)	7 (3.7)	5 (3.9)	3 (2.8)
Thr T	12 (3.7)	6 (2)	7 (3.6)	4 (2.1)	5 (3.9)	2 (1.8)
Trp W	1 (0.3)	1 (0.3)	1 (0.5)	1 (0.5)	0	0
Tyr Y	6 (1.9)	16 (5.4)	4 (2.1)	9 (4.8)	2 (1.5)	7 (6.5)
Val V	26 (8.1)	28 (9.5)	20 (10.4)	23 (12.3)	6 (4.6)	5 (4.6)
Total	**321**	**294**	**192**	**187**	**129**	**107**

Another significant change regards the increased number of arginine residues in *Mb*RpoE/F. This trend is the opposite to what is expected in a psychrophilic protein, as this residue is thought to enhance enzyme rigidity (hence thermostability) by facilitating more ionic (two salt bridges and five H bond) interactions via its guanidino group (Mrabet *et al.*, 1992; Georlette *et al.*, 2000; Siddiqui *et al.*, 2006).

Table 3.2: Comparison of some amino-acid ratios of the RpoE/F heterodimers from *M. burtonii* and *M. jannaschii*.

Parameter	*Mb*RpoE/F	*Mj*RpoE/F	*Mb*RpoE	*Mj*RpoE	*Mb*RpoF	*Mj*RpoF
Asp + Glu	61	58	34	35	27	23
Arg + Lys	46	44	28	28	18	16
Asp + Glu / Arg + Lys	1.3	1.3	1.2	1.2	1.5	1.4
Arg / Lys	0.84	0.42	0.87	0.4	0.8	0.45

However, no consistent trend towards an increase or decrease of negatively or positively charged amino-acids with the growth temperature of the parent organism was observed (Table 3.1 and 3.2).

3.3.4 Crystallisation

X-ray crystallography is a powerful method that involves the growth of well-ordered diffracting crystals, the measurement of the direction and intensities of x-rays diffracted by the crystals, and the reconstruction of the three-dimensional image of the protein molecule by computation analysis of these data.

Over 1000 different cristallisation conditions were tested with different buffers and temperatures, and with protein concentrations between 10 mg/ml and 20 mg/ml. Although the majority of the conditions either caused rapid precipitation of the protein or had no effect, a few conditions showed signs of crystal growth and were considered worth investigating further.

Of particular interest was one buffer condition comprising 0.1 M HEPES/NaOH, pH 7.5, containing 20% PEG200 and a crystal growth temperature of 20 °C. This condition gave well formed bipyramidal shaped crystals albeit of extremely small size. Improvement was obtained with a grid screening optimisation around the initial condition. When the protein and PEG200 concentrations were increased to 80 mg ml^{-1} and 40% respectively crystals of approximately 50 µm x 50 µm were obtained (Fig. 3.7).

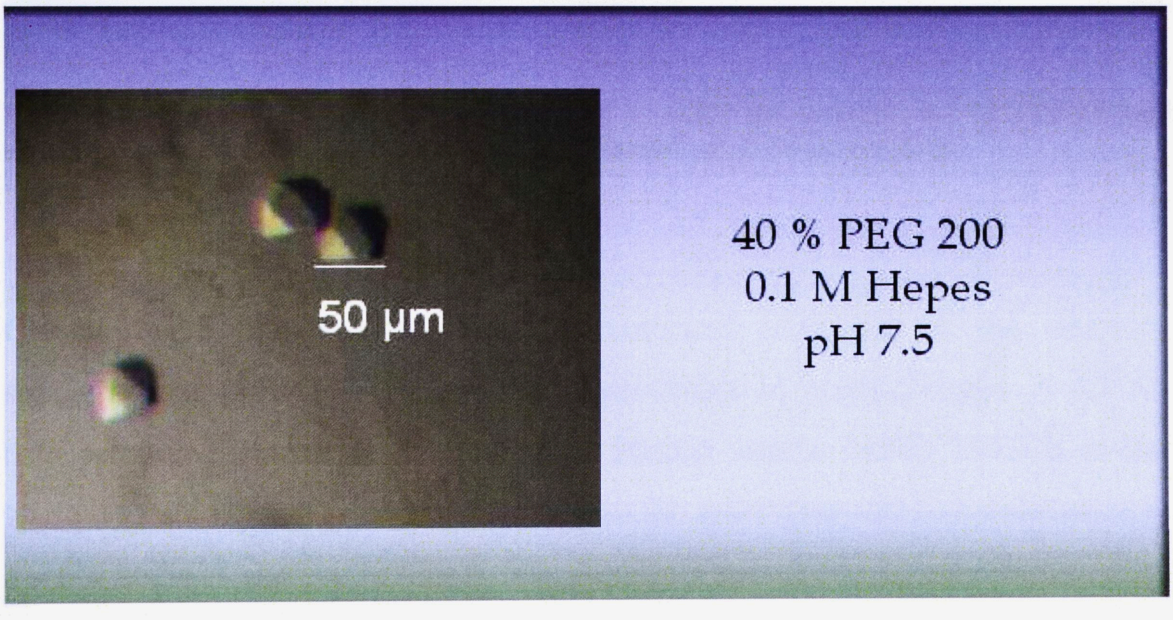


Figure 3. 7: Crystals of *Mb*RpoE/F grown at 20 °C under the condition 0.1 M HEPES/NaOH, pH 7.5, 40% PEG200.

When this specific condition was supplemented with 2 mM L-glutathione reduced/L-glutathione oxidised crystal size showed a further and significant improvement. The

crystals used in diffraction experiments grew to a size of 200 μm x 200 μm x 200 μm, over a period of two days (Fig. 3.8).

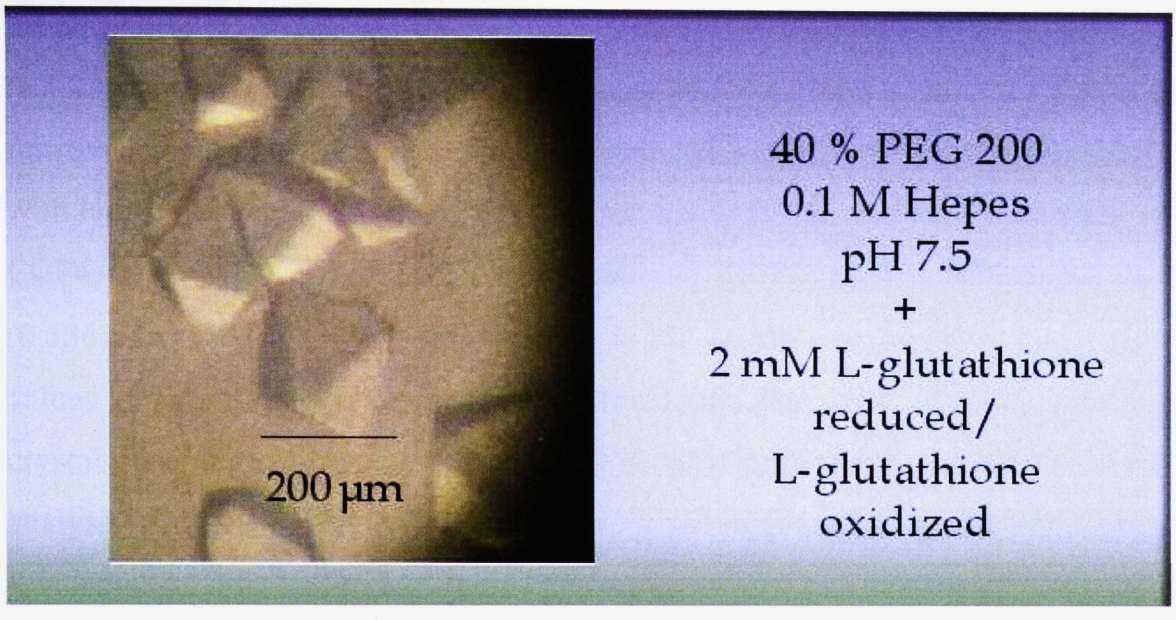


Figure 3.8: Crystals of *Mb*RpoE/F grown at 20 °C under the condition 0.1 M HEPES/NaOH, pH 7.5, 40% PEG200 and with 2 mM L-glutathione reduced/L-glutathione oxidised.

Crystals were transferred into a cryoprotectant consisting of the reservoir solution supplemented with 15 % glycerol and flash-cooled in liquid nitrogen. A 3.7 Å resolution data set was collected at the Advanced Photon Source (APS) 23-ID-B at the Argonne National Lab, USA. The crystal belonged to the space group $P4_32_12$ with lattice parameters: a = b = 86.54 Å and c = 107.39 Å. The data were reduced using the MOSFLM and SCALA from the CCP4 software package (CCP4, 1994) (18113 reflections; 2559 unique reflections; R_{merge} 0.121 (0.471); I/σ = 8.3 (3.3); completeness 95.6% (98%) multiplicity 4.1 (4.0), where the values in parentheses refer to the outer resolution bin: 3.90 - 3.70 Å) (Table 3.3). Phases were estimated by molecular replacement using the program PHASER (McCoy *et al.*, 2007) with the structure of *Mj*RpoE/F as a search model (Todone *et al.*, 2001). A unique solution was obtained (confirming the choice of space group) with a single RpoE/F heterodimer in the asymmetric unit. The resulting packing diagram showed good crystal packing with no steric clashes. Although the limited resolution of the data did

not warrant refinement of *Mb*RpoE/F, the quality of the molecular replacement solution shows that there are no significant structural differences between the two RpoE/F orthologues at this resolution.

A 3.2Å data set was collected using a rotating anode generator (Nonius FR591) with Osmic mirrors using a Mar345 imaging plate detector. The crystal belonged to the space group $P4_3$ with lattice parameters: a = b = 120.86Å and c = 105.19Å. The data were reduced using the CCP4 software package (185530 reflections; 26787 unique reflections; R_{merge} 0.155 (0.536); I/σ = 11.8 (3.8); completeness 99.8% (100%) multiplicity 7.4 (7.4), where the values in parentheses refer to the outer resolution bin: 3.37 - 3.20 Å) (Table 3.3). The crystals had high mosaicity (0.86°) and showed perfect merohedral twinning based on analysis using the PHENIX package.

Table 3. 3: Data collection and refinement statistics for the *Mb*RpoE/F crystals.

Type of crystal	Space group $P4_32_12$	Space group $P4_3$
Lattice dimensions, Å	*a = b = 86.54 Å* *c=107.39 Å*	*a = b = 120.9 Å* *c =105.2 Å*
Number of crystals	*1*	*1*
Number of measured reflections	*18113*	*185530*
Number of unique reflections	*2559*	*26787*
Maximum resolution	*3.7 Å (3.90 -3.70)*	*3.2 Å (3.37 – 3.2)*
Completeness of data	*95.6% (98%)*	*99.8% (100%)*
R_{merge}	*0.121 (0.471)*	*0.149 (0.67)*
I/σ	*8.3 (3.3)*	*11.8 (3.8)*

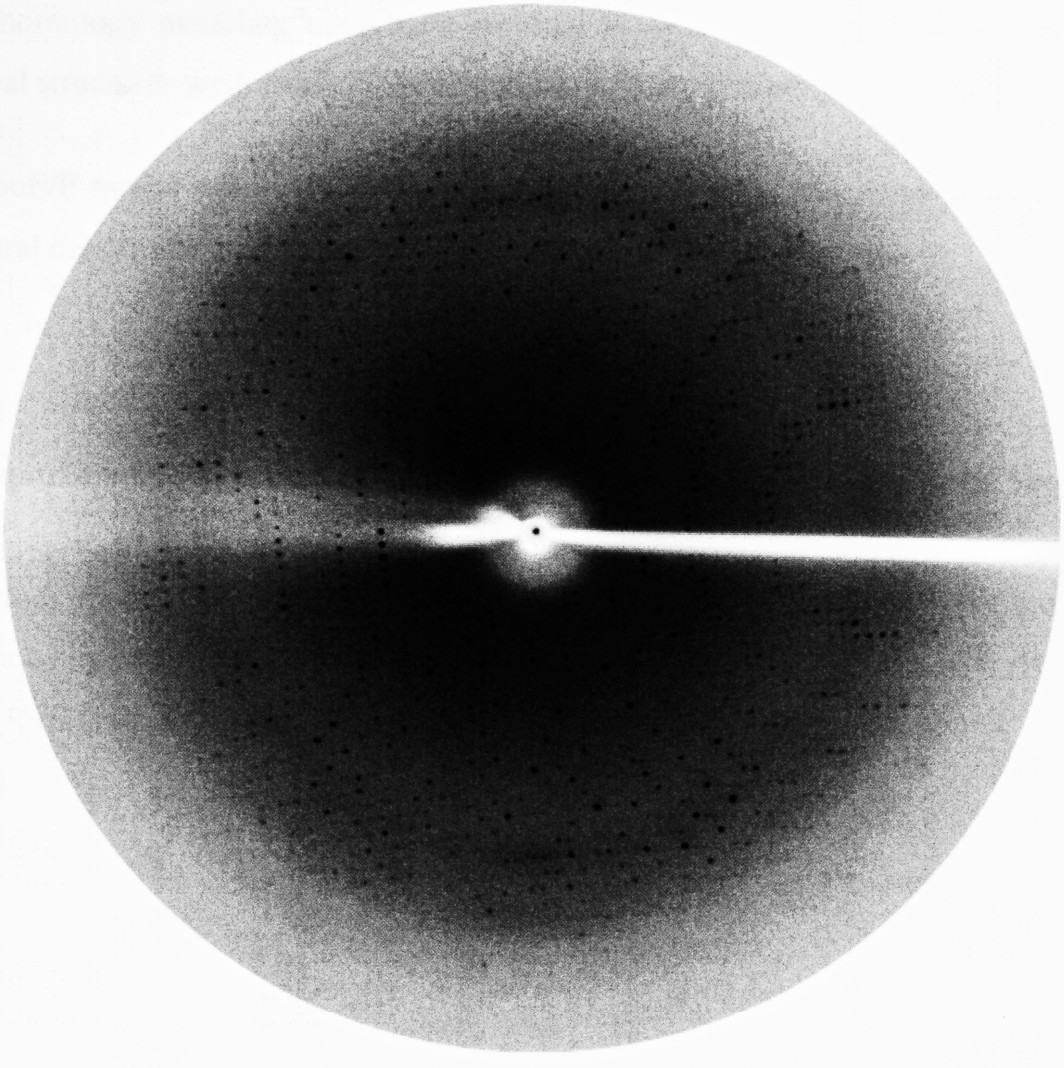


Figure 3. 9: A diffraction image of *Mb*RpoE/F crystal used for data collection. Resolution at the top of middle of the detector is 3.2 Å.

No more time was invested to improve the quality of the *Mb*RpoE/F crystals, as already the work described here had taken more than one year. At the same time the data coming from the functional characterisation appeared extremely interesting and therefore it was preferred to focus attention and resources on that part of the work. In order to further the structural study, an homology model was built for *Mb*RpoE/F to be compared with the structure of *Mj*RpoE/F. Comparative studies on the structural adaptation of cold-adapted enzymes

based on homology modeling have been reported in cases where no experimental three-dimensional structures were available (Gianese *et al.*, 2001; Paiardini *et al.*, 2003).

The *Mb*RpoE/F model and the *Mj*RpoE/F crystal structure were compared in order to find the structural determinants of the thermal adaptation of the two homologues.

3.3.5 Comparison between the *Mb*RpoE/F Model and the *Mj*RpoE/F Crystal Structure

The homology model built for *Mb*RpoE/F confirmed what had been already suggested by the sequence alignments: the two homologues have the same architecture, with the model generated for *Mb*RpoE/F being almost super-imposable on the entire *Mj*RpoE/F structure (Fig. 3.10).

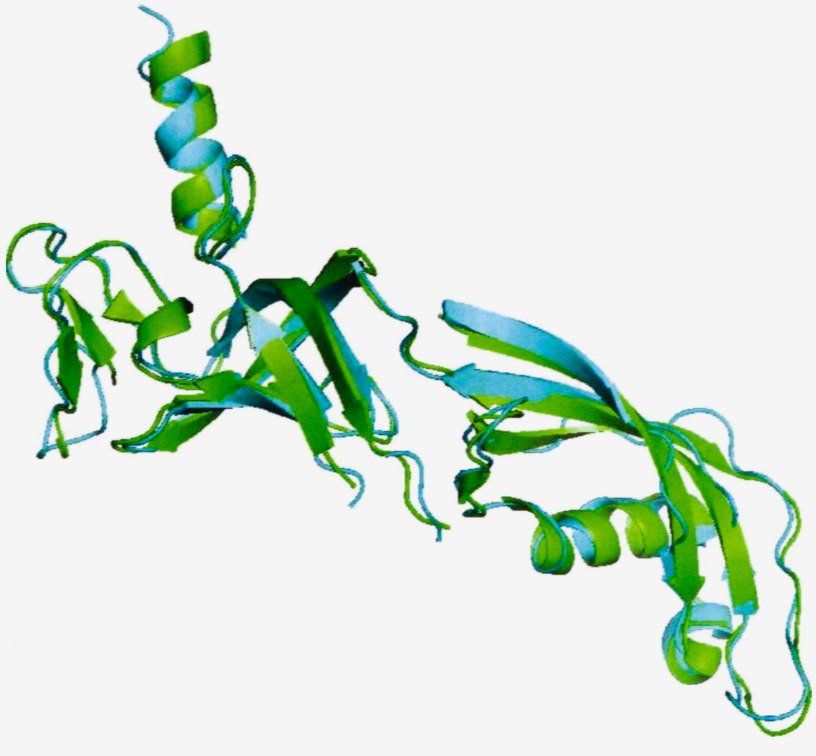


Figure 3.10: Superimposition of the *Mj*RpoE structure (green) and the *Mb*RpoE homology model (light blue).

Typical features of low adapted proteins derived from the comparison of models and crystal structures with mesophilic and thermophilic homologues include less hydrophobic-hydrophobic and electrostatic interactions (H-bonds, salt-bridges, aromatic-aromatic and Pi-cation interactions), decreased core hydrophobicity, increased surface hydrophobicity and hydrophilicity, weaker intersubunit/interdomain contacts, more and longer loops as well as less proline residues in loops (Gianese *et al.*, 2001; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007).

The comparative analysis was performed manually or with the help of several web-based tools (sees Materials and Methods). The results are discussed below.

- **Hydrophobic interactions**

Both homologues contain approximately the same number of hydrophobic residues. However *Mj*RpoE/F showed higher number and stronger hydrophobic interactions than *Mb*RpoE/F. The trend is similar if we consider the interactions between buried or exposed residues, inter-domains or inter-subunit. Weaker hydrophobic-hydrophobic interactions destabilise proteins by increasing the movements of the internal groups whereas solvent-exposed hydrophobic residues destabilise proteins by decrease in entropy of the water molecules which are immobilised on these residues (Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007).

- **Electrostatic interactions:**

Both homologues contain approximately the same total number of hydrogen bonds, but the number of hydrogen bonds in the intersubunit region is significantly higher in *Mj*RpoE/F than *Mb*RpoE/F (40 against 14, respectively). Regarding salt bridges, the total number is slightly higher in *Mb*RpoE/F compared to *Mj*RpoE/F, but again, the number of salt bridges

in the intersubunit region is significantly higher in the thermophilic homologue (19 against 8). A lower number of salt bridges is supposed to be present in psychrophilic homologues compared to their thermophilic counterparts, but several studies have pointed out that low-temperature adaptation is more strongly correlated with a reduced number of intersubunit interactions rather than the number of salt bridges *per se* (D'Amico *et al.*, 2002, Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007).

- **Aromatic interactions:**

Aromatic-aromatic interactions form a fourth type of force (after H-bonds, salt-bridges, and hydrophobic interactions) that stabilise protein structure (D'Amico *et al.*, 2002; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006). These interactions are defined as pairs of phenyl ring centroids separated by ~ 4.5 to 7 Å, dihedral angles within 30° – 90° and free energies of formation between -2.5 and -5.4 kJ mol^{-1} (Burley and Petsko, 1985). The aromatic-aromatic (Pi-Pi) interactions between the residues in position 2/77 and 95/98 of the subunit E are conserved in both homologues.

*Mj*RpoE/F contains extra interactions in the subunit E (positions 34/72), in the subunit F (positions 40/44) and three interactions in inter-subunit regions (positions E2/F40, E82/F88 and E141/F44).

When a cation sidechain (Arg or Lys) is in close proximity to an aromatic sidechain (Phe, Tyr or Trp), the geometry is predisposed towards **cation-pi interaction**. Among amino acid sidechains Arg and Trp are more likely to be involved in these interactions as compared to other sidechains. A favourable cation-pi interaction has been estimated to occur after every 77 residues in a protein and contribute towards the thermostability as much as conventional interactions such as H-bonds, salt-bridges and hydrophobic interactions (Gallivan and Dougherty, 1999; D'Amico *et al.*, 2002; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006)

The cation-Pi interaction between the residues in position 134/175 of the subunit E is also conserved in the two heterodimers. In *Mb*RpoE/F Trp 175 forms another interaction with Arg 172.

*Mj*RpoE/F illustrates four extra cation-Pi interactions: 2 in the subunit F (positions 87/90 and 88/89) and 2 in the intersubunit regions (position 3/11 and 145/88). It is a general trend for low-temperature adapted proteins to have less aromatic-aromatic and aromatic-cation interactions compared to their mesophilic and thermophilic homologues. Interestingly, the extra interactions in *Mj*RpoE/F are often intersubunit interactions, confirming a trend already seen for the electrostatic interactions.

- **Loops:**

*Mb*RpoF contains 3 loops (22-48, 79-83 and 104-116) that are longer compared to the corresponding ones in *Mj*RpoF (Fig. 3.11). Longer loops are a general feature of cold adapted proteins as they contribute to a higher degree of flexibility (D'Amico *et al.*, 2002; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006).

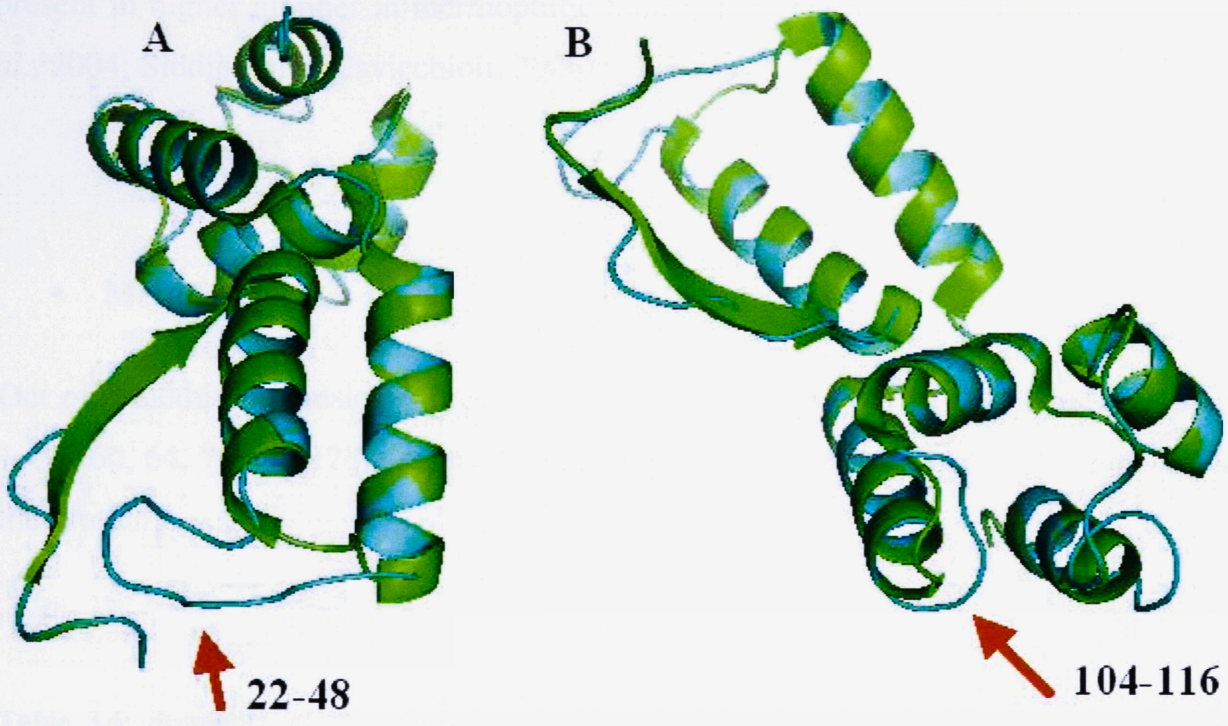


Figure 3.11: Superimposition of the *Mj*RpoF structure (green) and the *Mb*RpoF homology model (light blue). The arrows indicate loop 22-48 in panel A and the loop 104-116 of panel B, both longer in *Mb*RpoF.

- **Proline residues in α-helices vs proline residues in loops:**

Only three proline residues are conserved in the two homologues; 2 are in positions 15 and 79 of the subunit E, and the third is in position *Mb*95/*Mj*77 of the subunit F.

*Mj*RpoE/F presents 7 proline residues inside loops (positions 14, 79, 103, 122, 168 in E and 77 and 93 in F), one inside an α-helix (position 15) and one in a β-strand (position 69). *Mb*RpoE/F presents only 3 proline residues in loops (positions 79 and 156 in E and 95 in F) and 4 in α-helices (15, 182 in E, 27 and 84 in F).

It is evident that the number of proline residues in loops is lower in *Mb*RpoE/F compared to *Mj*RpoE/F. Proline residues in loops restrict conformational flexibility and therefore are

present in higher number in thermophilic homologues (D'Amico *et al.*, 2002; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006).

- **Methionine:**

Out of 6 methionine residues present in *Mb*RpoF but not in *Mj*RpoF, the ones at positions 1, 37, 60, 64, 79 and 128 are surface exposed, while the one at position 82 is buried inside the protein.

Table 3.4: Summary of the structural and compositional elements deduced from the comparison between *Mb* and *Mj*RpoE/F. Codes used: +, presence of structural or compositional feature in *Mb*RpoE/F compared to *Mj*RpoE/F; -, absence of structural or compositional feature in *Mb*RpoE/F compared to *Mj*RpoE/F; =, structural or compositional feature not appreciably different between *Mb* and *Mj*RpoE/F.

Characteristic found in cold adapted proteins	Occurrence
Decreased total hydrophobicity	+
Less total charged residues	=
High (Glu + Asp) / (Lys + Arg) ratio	=
Low Arg / Lys ratio	=
Less aromatic interactions	+
Less salt bridges	-
Weaker intersubunit/interdomain contacts	+
More / longer loops	+
Less Pro in loops	+
More Pro in α-helices	+
More Met	+

All the features associated with higher thermolability and flexibility present in *Mb*RpoE/F compared to its thermophilic homologue from *M. jannaschii* are summarised in Table 3.4. No cold-adapted protein investigated so far shows all of the above-mentioned features together, but a combination of few of them. It is also quite common that some features appear to be the opposite of what is expected for a thermally adapted protein.

It was found that most of the non-covalent interactions are concentrated in the intersubunit regions, which represent a major stabilizing feature associated with the adaptation of enzymes to extreme temperatures (Karlström *et al.*, 2006; Vetriani *et al.*, 1998, Rice *et al.*, 2006). This also suggests that unfolding may be initiated via monomerisation of the dimer.

This nicely confirms our earlier observation that it was not possible to express the subunit E independently from the subunit F in a soluble form.

3.3.6 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assay (EMSA) is a very powerful electrophoresis technique used to study protein-DNA or protein-RNA interactions. To test the binding of the two thermally adapted homologues for different single-strand nucleic acid sequences the assays were performed with four different radiolabeled ssDNA probes. It has been extensively reported for *M. jannaschii* RpoE/F, yeast and human Rbp4/7 that this heterodimer binds single-stranded nucleic acid in a saturable and reversible manner (Todone *et al.*, 2001; Orlicky *et al.*, 2001; Meka *et al.*, 2003, 2005).
In this case it was decided to investigate the binding specificity of the heterodimer for different oligo sequences which had not been addressed previously. The ssDNA sequences tested were 25-mers of poly-dA, poly-dC, poly-dG, poly-dT. According to our data, both heterodimers bind the poly-dT sequence but do not bind poly-dA; poly-dC is only bound by M*b*RpoE/F and poly-dG is only bound by MjRpoE/F (Fig 3.11, 3.12,

3.13).

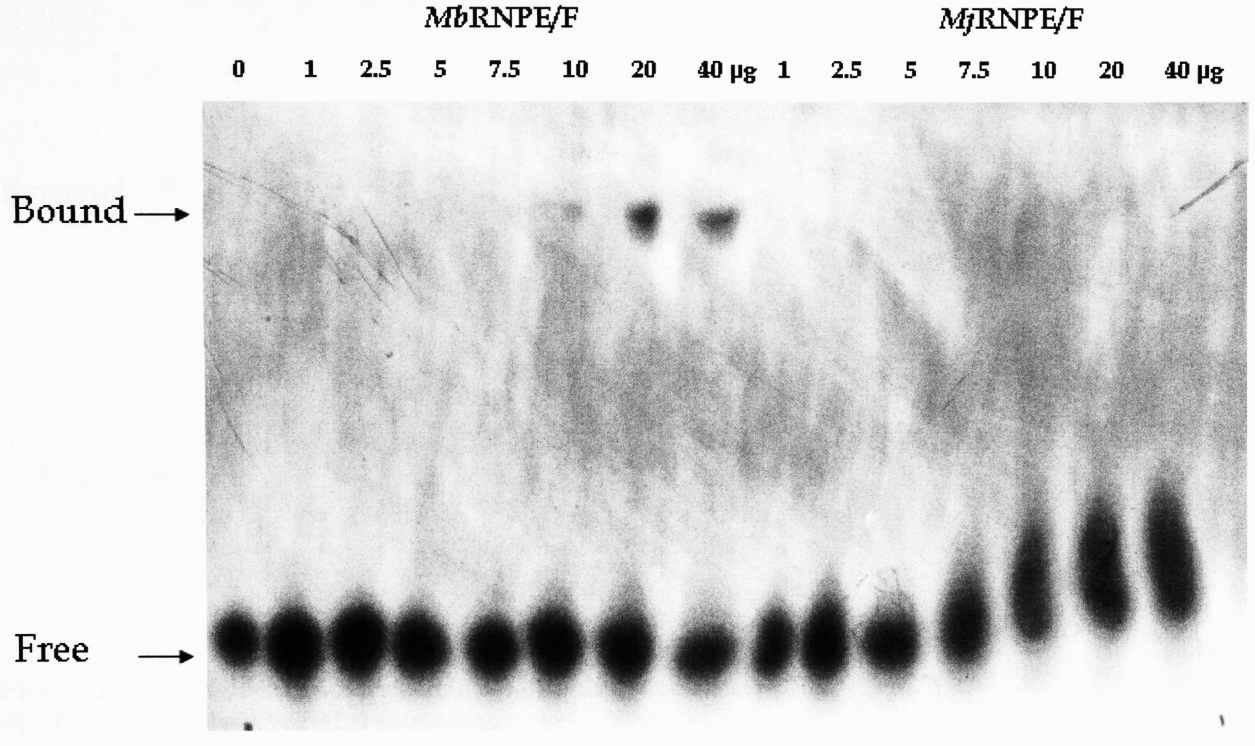


Figure 3. 12: Single-stranded nucleic acid binding specificity of *Mb*RpoE/F and *Mj*RpoE/F for ss-poly-dC (25mer) sequence. The radiolabeled probe was incubated with indicated amounts (micrograms) of purified *Mb*RpoE/F and *Mj*RpoE/F as described in materials and methods. Resulting complexes were resolved from unbound DNA by electrophoresis in a nondenaturing polyacrylamide gel. *Bands* formed by free DNA and protein-DNA complexes (*arrows*) were detected by autoradiography.

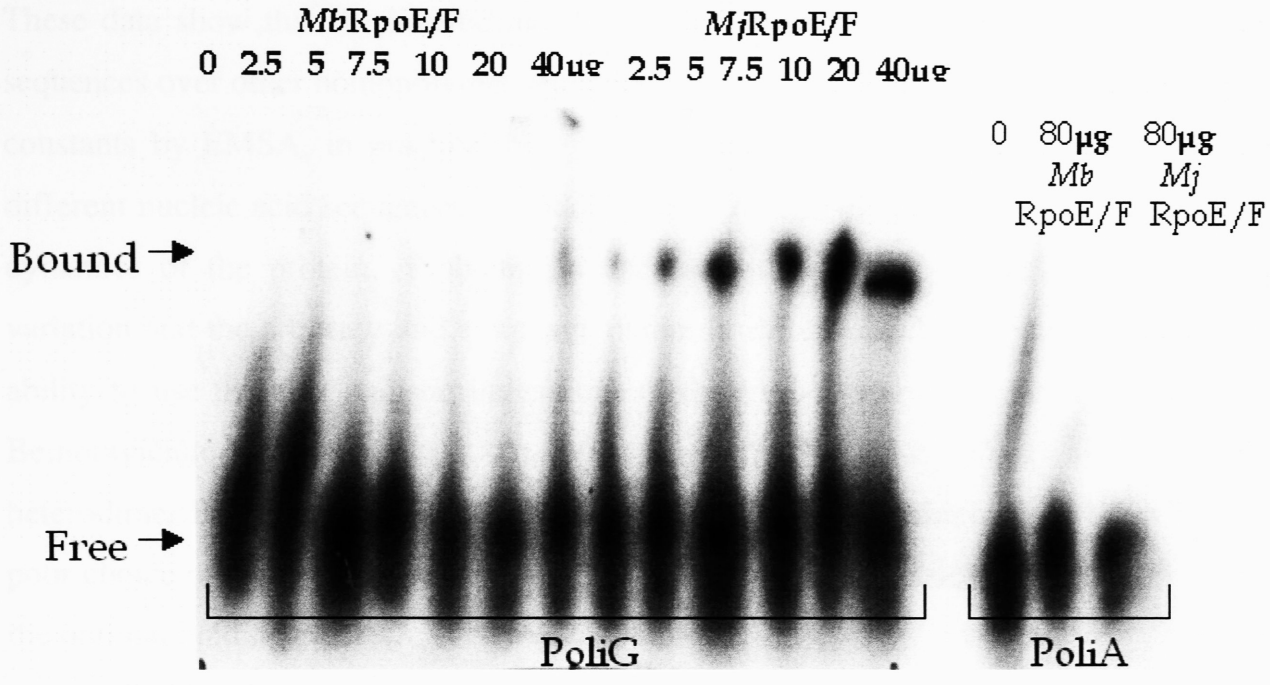


Figure 3. 13: Single-stranded nucleic acid binding specificity of *Mb*RpoE/F and *Mj*RpoE/F for ss-poly-dG (left) and poly-dA (right) sequence.

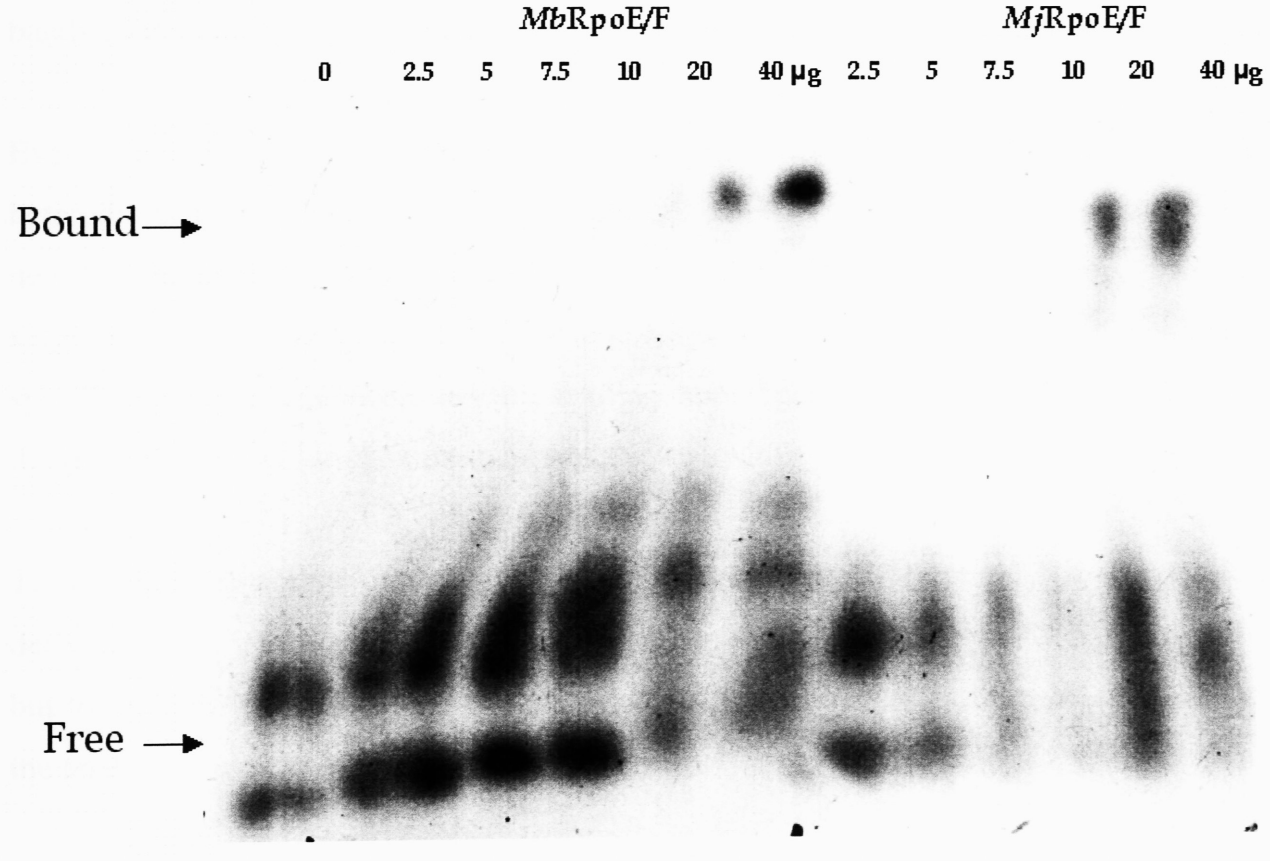


Figure 3. 14: Single-stranded nucleic acid binding specificity of *Mb*RpoE/F and *Mj*RpoE/F for ss-poly-dT sequence.

These data show that the heterodimer has a binding preference for certain homopolymer sequences over other homopolymer sequences. Although it is possible to determine binding constants by EMSA, in practice these are only useful for comparative analysis between different nucleic acid sequences, or modifications of a single protein to look at the binding dynamics of the protein. Problems in accurate quantitation of EMSA gels, gel-to-gel variation and the arbitrary and simplistic nature of the *in vitro* buffer system preclude any ability to use these as realistic indications of the true, *in vivo* Kb value (Li *et al.*, 2004; Beinoravičiūtė-Kellner *et al.*, 2005; Hellman *et al.*, 2007). Kb values calculated for the heterodimer:DNA interactions were very high (above 50 μM) which could be indicative of poor choice of assay conditions or could be indicative that the nucleic acids used were not the optimal binding partner.

Nevertheless, they were proof that that the heterodimer bound DNA and that the interaction was tight, since loose interactions or movement of the nucleic acid would not result in sharp bands on the gel. Most importantly, the interaction showed some sequence preference.

Even if the homopolymer sequences cannot be considered the optimal target of the heterodimer in the conditions tested, they significantly contributed to the formulation of a new hypothesis in the characterisation of *Mb*RpoE/F. The binding of the heterodimer to single strand nucleic acid may not be non-specific as usually postulated, but it could be specific for some sequences, and this binding could be the basis of a regulatory function of the E/F subunits within the transcription machinery.

To validate the ability of the heterodimer to recognise different mRNA targets, it was decided not to investigate the binding properties of small ssDNA or RNA oligos further, but to test full-length RNA transcripts extracted from growing cultures. In order to do so, the analysis was restricted to *Mb*RpoE/F.

3.4 Conclusion

The initial data from the comparative study performed on both thermally adapted homologues confirmed that they are in a native and folded form. *Mb*RpoE/F was found to be less stable and more flexible than *Mj*RpoE/F, and to have a lower activation energy of denaturation. The comparison of models and crystal structures between thermally adapted proteins illustrate thermolability and flexibility to be the result of a combination of several features such as: more glycine and methionine residues, a low Arg/Lys and a high (Glu + Asp) / (Lys + Arg) ratio, less hydrophobic-hydrophobic and electrostatic interactions (H-bonds, salt-bridges, aromatic-aromatic and Pi-cation interactions), decreased core hydrophobicity, increased surface hydrophobicity and hydrophilicity, weaker intersubunit/interdomain contacts, more and longer loops as well as less proline residues in loops (Gianese *et al.*, 2001; Georlette *et al.*, 2004; Siddiqui and Cavicchioli, 2006; Tronelli *et al.*, 2007). In order to perform a similar study on both homologues, cristallisation of the proteins was attempted to determine the crystal structure of *Mb*RpoE/F. Despite the initial promising results from the cristallisation trials, it was not possible to obtain a structure of reasonable quality. Therefore a homology model was generated for the purpose of the comparison with the crystal structure available for *Mj*RpoE/F. It was possible to suggest a number of features that may contribute to the flexibility and thermolability of *Mb*RpoE/F (Table 4).

Functional characterisation of the nucleic acid binding properties of the two homologues was performed using EMSA on a range of ssDNA templates. These data indicated that *Mb*RpoE/F bound specifically to oligo-dC and oligo-dT while *Mj*RpoE/F bound preferentially to oligo-dG and oligo-dT. This is the first published report in either archaea or eucaryotes, that the E/F heterodimer can bind nucleic acid in a sequence-specific manner.

Chapter 4: Total RNA Binding Experiment and Microarray Data Analysis

4.1 Introduction

The functional characterisation of the nucleic acid binding properties of *Mb*RpoE/F and *Mj*RpoE/F, described in Chapter 3, showed that the two homologues have different specificities for ssDNA homopolymer sequences. Although the Kb values indicate that these sequences are possibly due to an unspecific binding, they are also proof that the interaction showed some sequence preference.
This finding was totally unexpected as it has been extensively reported that the archaeal RpoE/F and its eucaryotic homologue, Rbp4/7, bind single strand nucleic acid in a nonspecific manner.

The possibility that the heterodimer could instead bind different nucleic acid targets with different affinities was therefore further investigated for *Mb*RpoE/F.

His-tagged *Mb*RpoE/F was incubated with whole cell RNA and complexes were purified using a Ni-NTA column. RNA was then eluted, labeled and hybridised to a high-density *M. burtonii* microarray.

The EMSA approach was completely abandoned (apart for testing the RNA binding ability of *Mb*RpoE/F) for a number of reasons: the microarray analysis is a more cost- and time-effective solution to screen hundreds if not thousands of different sequences against the few sequences per experiment of the EMSA; it is more biologically relevant as full-length transcripts (rather than short RNA oligos) are the real target of RpoE/F in the cell, and also because they obviously present elements of secondary structure that would likely maximise any RNA-protein interaction.

Another important consideration is that the microarray binding experiment was performed *in vitro,* mixing only the affinity-purified heterodimer and purified total *M. burtonii* RNA, and this excluded the possibility that another partner protein could be involved in the binding process and be responsible for the selectivity.

It may seem a contradiction to say that the microarray experiment was designed taking in consideration that the binding could depend on RNA secondary structure, when the homopolymer sequences tested in the EMSA experiments obviously present no secondary structure. The homopolymers of the EMSA experiments and the total RNA of the microarray experiment are two completely separated entities, tested in completely different conditions and modalities. It is very likely that those sequences are not the real target of the heterodimer, and were never expected to be determinants of the binding within the full length transcripts. Nevertheless, the analysis of the homopolymer sequences was a turning point in this study, as it was from the EMSA experiments that it was hypothesised that a feature of *Mb*RpoE/F could be to bind RNA in a sequence-dependant manner.

The microarray data show that *Mb*RpoE/F bound to a narrow subset of *M. burtonii* transcripts.

4.2 Materials and Methods

4.2.1 EMSA

Varying amounts of *Mb*RpoE/F were incubated for 30 min on ice with 5 fmol of ^{32}P end-labeled RNA oligonucleotide in 10 µl of buffer containing 40 mM HEPES/NaOH, 100 mM potassium acetate and 10% glycerol, pH 7.4 and 0.5 units of RNAse inhibitor (Promega). The random RNA sequence that was tested under these conditions was 5'–AAGGCGGUGCUGAA-3'. Bound and free probe were resolved by electrophoresis in a

7.5% polyacrylamide gel containing 10% glycerol and 0.5 x TBE. Bands formed by free RNA and protein-RNA complexes were detected by autoradiography.

4.2.2 Harvesting of *M. burtonii* Cultures and RNA Extraction

M. burtonii cultures were grown in 100 ml MFM, an enriched medium with trimethylamine (TMA) as the methylated substrate, at 4 °C or 23 °C under a gas phase of 80:20 as described in section 2.2.1. Cultures were inoculated 1:100 from cells grown under the same conditions. Cultures (50 mL) were harvested at late logarithmic phase (optical density at 620 nm of 0.25) by centrifugation at 3200 x *g* for 35 min at a temperature corresponding to their growth temperature (4 °C or 23 °C). RNA was extracted using the SV Total RNA Isolation System (Promega) and RNA concentration was determined using the Nanodrop Spectrophotometer ND-1000 (BIOLAB). From each RNA sample 100 ng was analysed using an Agilent Nano chip to verify RNA integrity.

4.2.3 Total RNA Binding Experiment

The total RNA extracted was then incubated in batch with His-tagged *Mb*RpoE/F and RNA-protein complexes were purified using a Ni-NTA resin (Fig. 4.1).

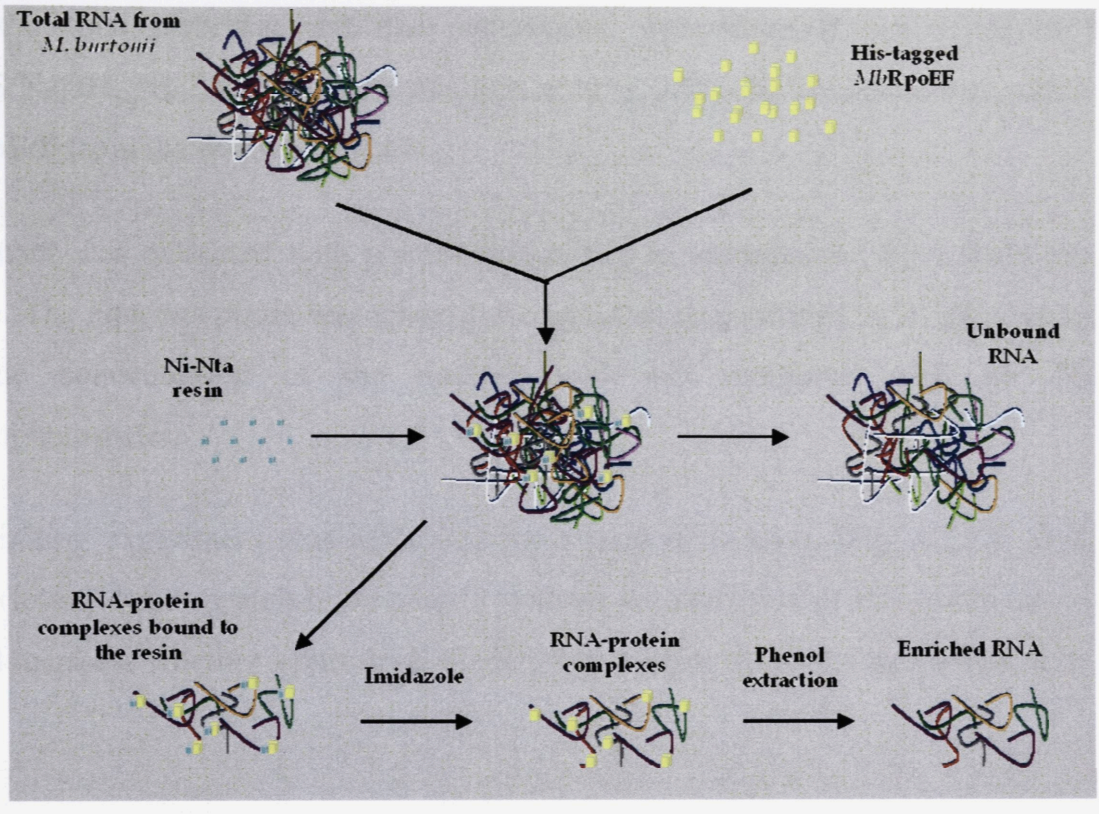


Figure 4. 1: Summary of the strategy used for the total RNA binding experiment. *M. burtonii* total RNA was incubated with His-tagged *Mb*RpoE/F. Ni-Nta resin was added to the mixture and incubated for a further hour. This suspension was extensively washed via a series of centrifugation - supernatant removal - binding buffer addition in order to remove the RNA that was not bound by the heterodimer. After the final washing steps, the protein-RNA complexes were eluted with imidazole and the RNA separated from the protein with a standard phenol extraction followed by ethanol precipitation.

100 µg of total RNA was resuspended in 12 ml of binding buffer (40 mM HEPES/NaOH, pH 7.4, 100 mM potassium acetate) and incubated for 1 h with 12 mg of previously purified *Mb*RpoE/F.

Ni-Nta resin (250 µl, QIAGEN) was added to the mixture and incubated for a further hour. This suspension was centrifuged at 500 × *g* for 10 min to sediment the resin and the supernatant gently removed and replaced with 12 ml of fresh binding buffer. This last step (centrifugation - supernatant removal - binding buffer addition) was repeated 3 times to wash the unbound RNA from the resin.

After the final centrifugation, the supernatant was removed and 1 ml of 40 mM HEPES/NaOH, pH 7.4, 100 mM potassium acetate, 250 mM imidazole was added to elute *Mb*RpoE/F from the resin (Fig. 4.1).

The eluant was extracted with phenol-chloroform to separate the RNA from the protein sample. The aqueous phase was ethanol precipitated, resuspended in 20 µl of milliQ water and the concentration of the nucleic acid was measured with the NanoDrop Spectrophotometer.

This binding experiment was replicated for a total of 8 times (Fig. 4.2): 6 as described above (total RNA + protein), the other 2 without the addiction of the protein as a negative control to assess whether a part or all the RNA had a specificity for the Ni-Nta resin.

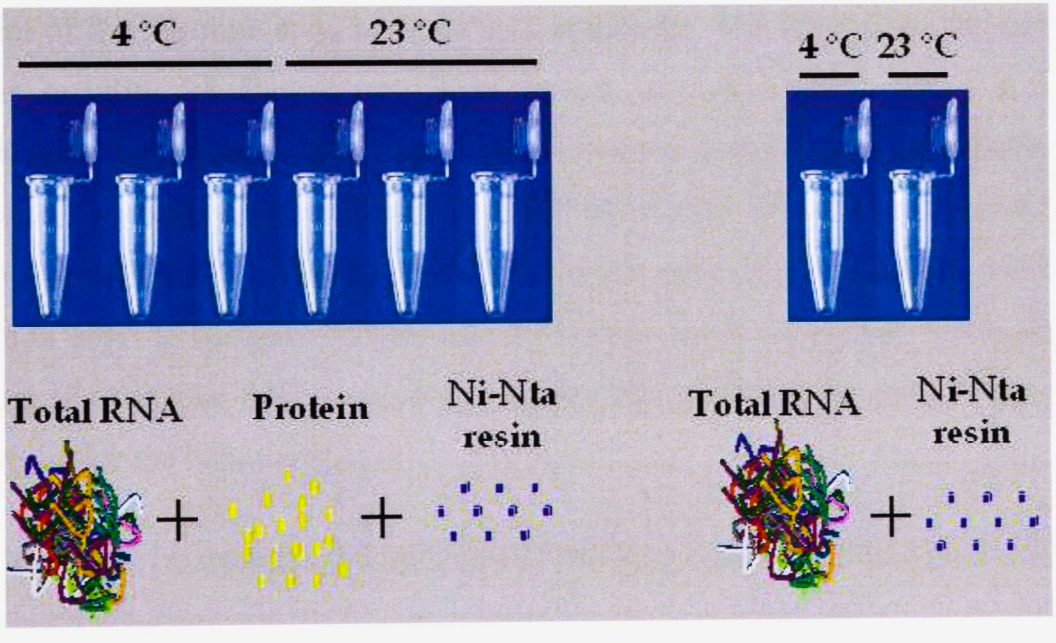


Figure 4. 2: Scheme of the replicas of the total RNA binding experiment. 6 replicas reflected the scheme summarised in Fig. 4.1. The other two experiments did not include the addition of the His-tagged protein and acted as a negative control. In three of the replicates and one control total RNA from *M. burtonii* cultures grown at 4 °C were used and the remaining three replicates and the control used RNA isolated from *M. burtonii* cultures grown at 23 °C.

Because *M. burtonii* has a different expression profile at 4 °C and 23 °C (Campanaro *et al.*, 2010, manuscript to be submitted), the binding experiment was designed in such a way to have an equal number of replicates with RNA isolated from cultures grown at these two temperatures. This would determine whether the different expression profiles would be reflected in the *Mb*RpoE/F binding (Fig. 4.2).

4.2.4 RNA Labeling and Hybridisation

One microgram of each sample was used for direct labeling using ULS™ Fluorescent Labeling Kit for Agilent arrays (with Cy3 and Cy5) (Kreatech). Direct labeling of RNA, rather than cDNA, was chosen as the better means of maintaining proportionality between fluorescence signal and transcript abundance, and therefore to obtain a more accurate assessment of the absolute abundance of each transcript. The RNA concentration and the picomolar quantity of fluorophore incorporated was determined using a Nanodrop spectrophotometer. On average 400 ng of RNA obtained from the binding experiments was labeled with 30 pmoles of fluorophore and 1400 ng of total RNA labeled with 130 pmoles of fluorophore (Fig. 4.3). These two samples were used in a competitive hybridisation procedure in order to compare on the same microarray the RNA obtained from the binding experiment and the total RNA used as a reference, thus enabling the total-RNA to be used as a reference for the bound-RNA.

Labeled RNA was fragmented and hybridised following Agilent standard protocols.

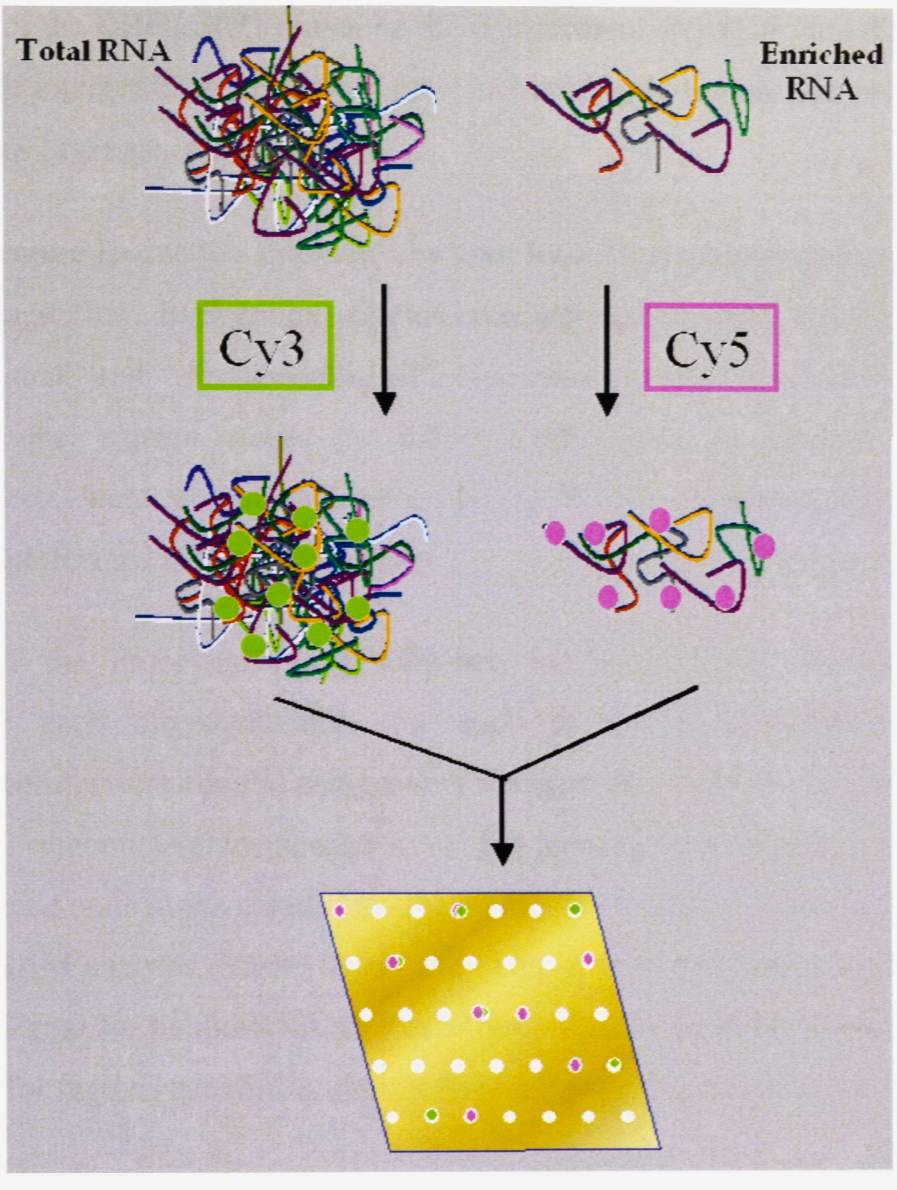


Figure 4. 3: Scheme of the labeling and hybridisation of the total and enriched RNA.

4.2.5 Microarray Data Analysis

Gene expression analysis was performed using Agilent 8 x 15K custom gene expression microarrays containing 15128 probes designed on the *M. burtonii* genome (See Appendix 1). Oligo sequences were aligned on *M. burtonii* genome sequence using BLAST with parameters specific for short sequences and only those having a single perfect match (14531) were considered for further analyses. Of these, 10153 oligos are specific for the

coding region of the ORFs, 3671 match on the complement strand of the ORFs and 707 are localised on the intergenic regions. Each gene and intergenic region was covered by 1 to 6 oligonucleotides (average of four per gene).

The Agilent Feature Extraction software was used for image analysis and processing of the microarray image files. Background signals were subtracted from raw signals to obtain "gProcessedSignal" and "rProcessedSignal". Dye biases created by differences in the red and green channel signals caused by different efficiencies in labeling and intensity-dependent effects were normalised using the LOWESS algorithm. Further statistical analyses were performed as described below.

Log_2 ratios of the Processed Signal obtained for bound-RNA and total-RNA were calculated for each oligonucleotide, for each of the 6 competitive hybridisation experiments. Results were filtered using a very stringent threshold. To be considered bound by *Mb*RpoE/F, oligonucleotides needed to exhibit a mean fluorescence value higher than 1000 fluorescence units in the channel for the bound-RNA relative to the total-RNA, have a q-value from SAM analysis (Tusher *et al.*, 2001) equal to 0, and have a log2 ratio between the ProcessedSignal for bound-RNA vs total-RNA > 0.7. In addition, these criteria needed to be satisfied for regions containing at least three adjacent oligonucletoides.

The microarray design and data analysis was performed by Stefano Campanaro (University of Padua, Italy).

4.2.6 Bionformatic Analysis

Operons were predicted to belong to the same transcriptional unit using the software developed by Price *et al.*, (2005). The list of operons is available at http://microbesonline.org/operons/OperonList.html

To attempt to identify motifs associated with the genes bound by *Mb*RpoE/F (*e.g.* putative binding sites), the overrepresentation of DNA motifs composed of 8, 10 or 12 bases were assessed using Weeder (Pavesi *et al.*, 2004). Two independent analyses were performed: the first approach, examined the entire region bound by the heterodimer, which in many cases was composed of more than one gene; the second approach focused only on the intergenic regions (thereby including the 5'-UTR of genes within multi-gene regions) and 100 bases upstream of single genes or the first gene in a multi-gene region. Six different background models were generated using custom PERL scripts thanks to Stefano Campanaro (University of Padua, Italy), three for the regions covered by the first approach and three for the second approach, in order to calculate the frequency of all the possible 8mers, 10mers and 12-mers. The background model files are required in Weeder to consider the frequency of all the possible oligonucleotides composed by 8, 10 or 12 bases in the regions bound by *Mb*RpoE/F.

4.3 Results and Discussion

4.3.1 EMSA

Electrophoretic mobility shift assay (EMSA) was used to test the binding affinity of *Mb*RpoE/F for RNA. A random sequence (5'–AAGGCGGUGCUGAA-3') was tested. The data (Fig. 4.5) shows that the heterodimer has a binding affinity for RNA.

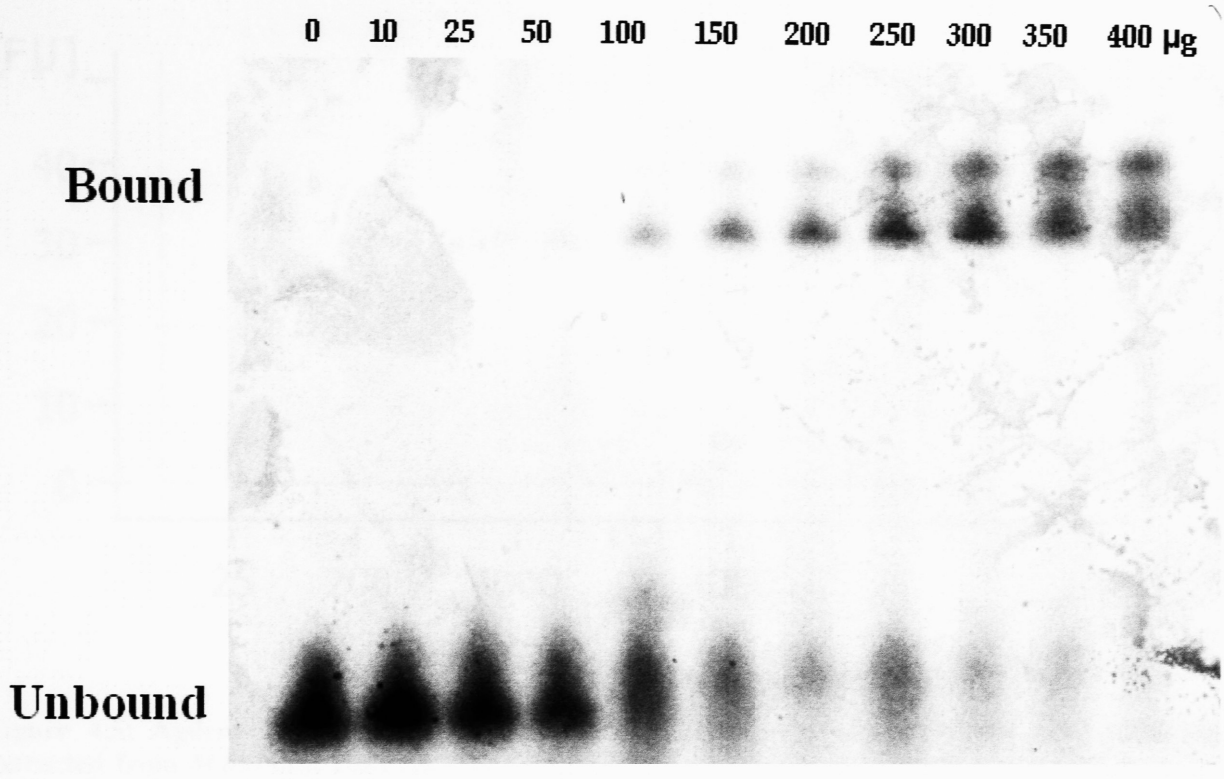


Figure 4. 4: RNA-binding activity of *Mb*RpoE/F. A radiolabeled RNA probe was incubated with indicated amounts (micrograms) of purified *Mb*RpoE/F as described under Materials and Methods. Resulting complexes were resolved from unbound RNA by electrophoresis in a nondenaturing polyacrylamide gel.

4.3.2 RNA Extraction

Approximately 200 µg of RNA was obtained from each culture and 100 ng was analysed using an Agilent Nano chip to verify RNA integrity. For each sample the signals corresponding to the 16S and 23S were the only clearly visible peaks (Fig. 4.6), which is a proof of the high quality of the RNA.

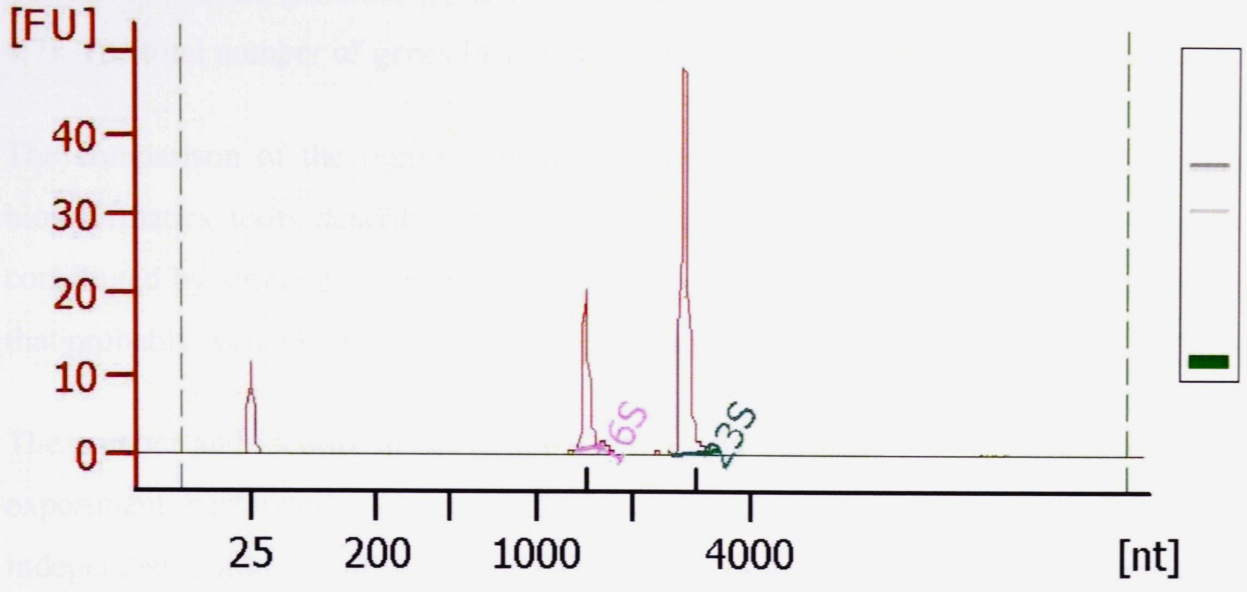


Figure 4.5: Agilent 2100 Bioanalyser Data. Electropherogram of one of the total RNA samples extracted from *M.burtonii* cultures. The first peak in each sample corresponds to the Agilent RNA 6000 Nano Marker and the main ribosomal RNA peaks are indicated. The 16S and 23S peaks are clearly visible. Beside is a gel-like image of the electropherogram.

4.3.3 Total RNA Binding Experiment

Each affinity purification gave approximately 1 µg of enriched RNA. No RNA was obtained from the two controls, meaning that no detectable amount of RNA bound to the Ni-Nta resin in a non-specific way, and therefore that the RNA eluted in the other six experiments was derived from the direct binding with the His-tagged *Mb*RpoE/F.

4.3.4 Identification of the Transcripts Bound by *Mb*RpoE/F

From the microarrays experiments 588 oligos were identified that hybridised with the transcripts that were preferentially bound by *Mb*RpoE/F and that satisfy the statistical criteria described in Materials and Methods. A negligible number of these oligos are distributed

"randomly" in the genome, while the vast majority is localised in 48 "distinct" regions (Fig. 4.7). The total number of genes identified in these 48 regions is 118 (Table 4.1).

The comparison of the regions identified with *M. burtonii* operons, predicted using the bioinformatics tools described in Materials and Methods, indicates that 22 regions are constituted by single-gene transcripts and 27 by operons constituted by two or more genes that probably were bound by the protein as a single mRNA molecule.

The number and identity of the genes is essentially the same for all the six replicas of the experiment performed (See Appendix 1), each with the total RNA coming from an independent culture of *M. burtonii*. This means that the experiment was reproducible, and that the different abundance of the individual transcripts in the two culture conditions (4 °C and 23 °C, Fig. 4.2) did not result in different binding profiles. This may be a consequence of the *in vitro* nature of the binding experiment.

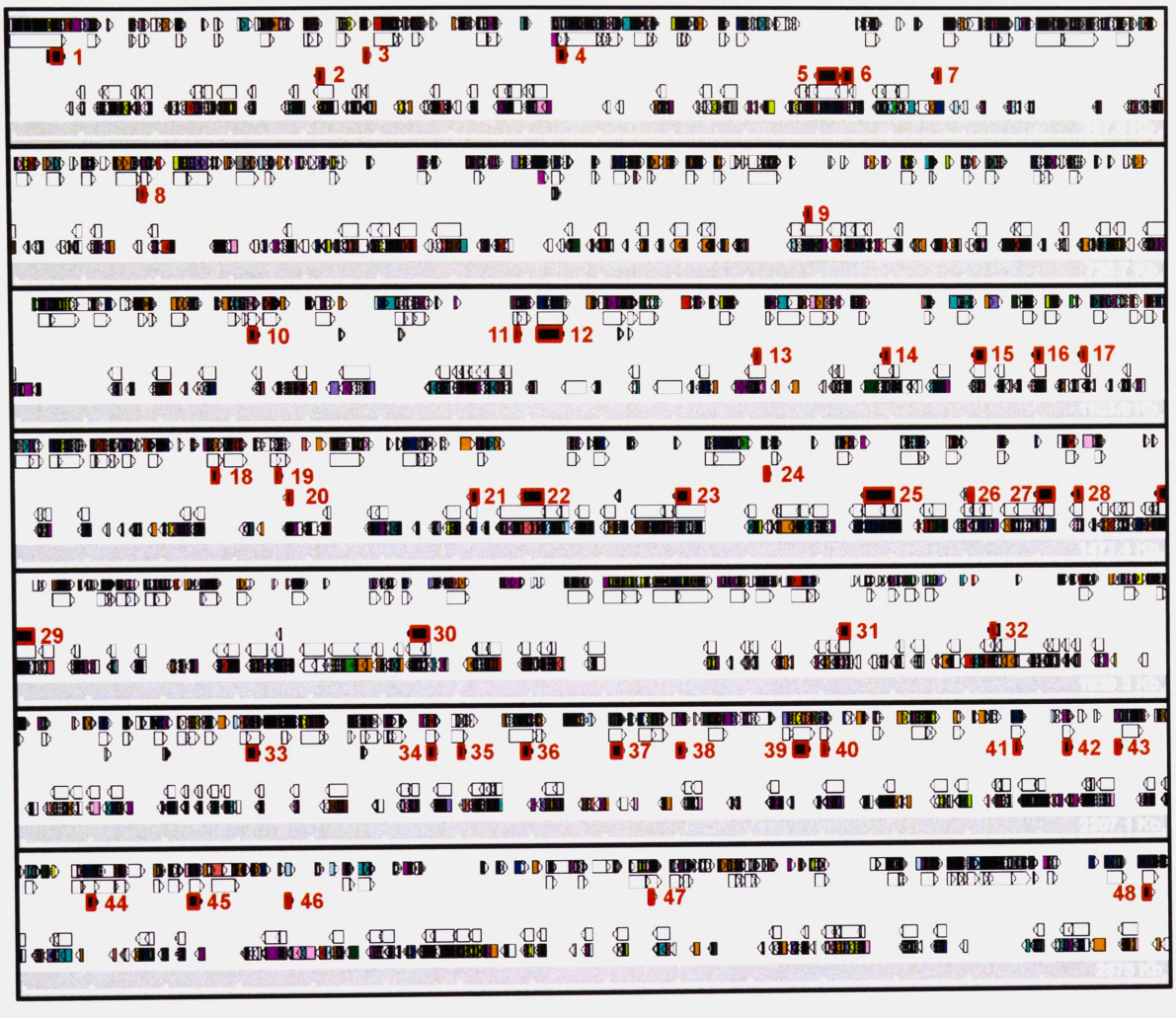


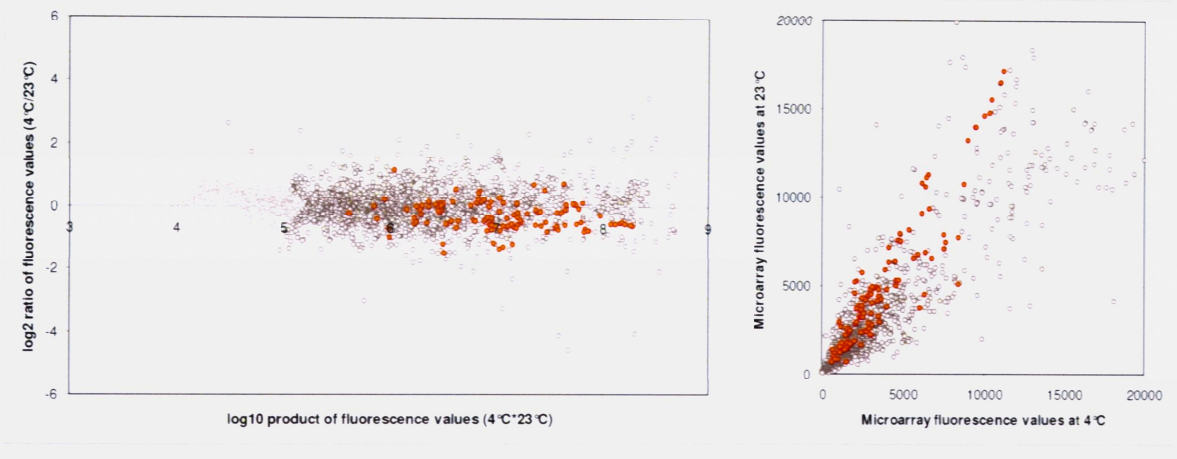
Figure 4. 6: Schematic representation of the *M. burtonii* genome sequence divided in seven sections of 368 Kb each (positions are reported in the gray bar) and localisation of the regions bound by *Mb*RpoE/F. *M. burtonii* genes (lines 1 and 6 of each section) are coloured depending on their COG category. White arrows (lines 2 and 5) represent operon organisation in the genome determined using genome-specific distance models, combined with comparative genomics and functional features and downloaded from (http://microbesonline.org/operons/OperonList.html) (Price *et al.*, 2005). Single microarray oligos that satisfy statistics criteria were reported (lines 3 and 4) and regions that were considered were highlighted in red.

Table 4. 1: List of the genes whose transcripts were bound by *Mb*RpoE/F. The second column, named "ratio" represents the ratio between the fluorescence signal of the labelled RNA obtained from the binding experiment and the "control" RNA. This value is generally proportional to the "strength" of the binding between the heterodimer and each specific transcript. This value was generally very similar for the oligos belonging to the same region and was therefore averaged in order to fit in the table.

Regions	Ratio	Name	Product name
1	1.71	Mbur_0024	Preprotein translocase SecY subunit
		Mbur_0025	Adenylate kinase
		Mbur_0026	Integral membrane protein DUF106
		Mbur_0027	Cytidylate kinase
		Mbur_0028	tRNA pseudouridine synthase B
2	1.80	Mbur_0110	Signal recognition docking protein FtsY
3	2.54	Mbur_0129	Hypothetical protein
4	1.65	Mbur_0190	Prolyl-tRNA synthetase
		Mbur_0191	N^1-α-phosphoribosyltransferase
5	2.14	Mbur_0280	FdhD
	1.74	Mbur_0281 to 0285	Formylmethanofuran (Mo) dehydrogenase subunits B, D, C, A and F
6	1.80	Mbur_0288	Fe-S cluster assembly protein
		Mbur_0289	Cysteine desulfurase
7	1.66	Mbur_0316	Aconitate hydratase 2
8	1.88	Mbur_0425	Sodium/solute symporter
9	1.82	Mbur_0605	Pyruvate synthase γ / β -like protein
10	1.67	Mbur_0772	Aspartyl-tRNA synthetase
		Mbur_0773	D-Ribose-5-phosphate isomerase
11	2.03	Mbur_0850	Hypothetical protein
12	2.00	Mbur_0858 to 0863	ACDS subunits α, ε, β, Ni, δ and γ 2, respectively
13	4.59	Mbur_0904	DNA-directed RNA polymerase subunit F
14	1.69	Mbur_0939	FAD dependent oxidoreductase
15	1.83	Mbur_0961	tRNA (guanine-N(1)-)-methyltransferase
		Mbur_0962	Lysine export domain protein
		Mbur_0963	Transcription initiation factor IIB
16	1.80	Mbur_0981	ABC transporter, permease component
17	1.91	Mbur_0994	Pyrophosphatase
18	1.72	Mbur_1087	Homoserine dehydrogenase
		Mbur_1088	DNA ligase 1, ATP-dependent
19	1.84	Mbur_1105	Uroporphyrinogen-III C-methyltransferase
20	1.95	Mbur_1108	Hypothetical protein
21	1.71	Mbur_1159	Formylmethanofuran (W) dehydrogenase subunit E like protein
		Mbur_1160	Adenylosuccinate lyase (PurB)
22	2.00	Mbur_1176 to 1179	DNA-directed RNA polymerase subunits A'', A', B' and B''
23	2.14	Mbur_1224 and 1225	Dihydroorotate dehydrogenase electron subunits K and D
		Mbur_1226	Porphobilinogen deaminase
		Mbur_1227	Glutamate-1-semialdehyde 2,1-aminomutase
24	2.08	Mbur_1252	Hypothetical protein

25	2.91	Mbur_1286 to 1293	F420H2 dehydrogenase subunits O, N, M, L, K, I and H
26	1.99	Mbur_1324	Threonyl-tRNA synthetase
27	1.68	Mbur_1344	Hypothetical protein
		Mbur_1345	Phosphoglycerate kinase
		Mbur_1346	Protein with DRTGG domain
		Mbur_1347	Acetyl-CoA synthetase (ADP-forming), α and β subunits
28	1.84	Mbur_1353	Radical SAM family protein
		Mbur_1354	Fructose 1-6-bisphosphatase
29	1.93	Mbur_1378	UPF0132 family protein, part of Rnf complex
		Mbur_1379 to 1384	Rnf electron transport complex proteins B, A, E, G, D, and C
30	1.93	Mbur_1518 to 1525	H4SPT-methyltransferase subunits H, G, F A, B, C, D and E
31	1.79	Mbur_1654 and 1655	Asn /Gln –tRNA amidotransferase subunit B and A
32	1.92	Mbur_1699	ABC transporter cobalt-binding protein
		Mbur_1700	Cobalamin biosynthesis protein
33	1.62	Mbur_1813	Serine hydroxymethyltransferase
		Mbur_1814	Methylene-THF dehydrogenase/cyclohydrolase
		Mbur_1815	Dihydropteroate synthase
34	1.82	Mbur_1872	Ferrous iron transport protein B
35	1.66	Mbur_1883	Homoaconitase large subunit
36	1.68	Mbur_1906	2-Hydroxyglutaryl-CoA dehydratase component A
		Mbur_1907	Hypothetical protein
		Mbur_1908	Protein of unknown function UCP019164
37	1.81	Mbur_1932	3-Hydroxy-3-methylglutaryl-CoA synthase
		Mbur_1933	Acetoacetyl-CoA thiolase
		Mbur_1934	DUF35-domain protein
		Mbur_1935	Adenylyl cyclase-family protein
38	2.19	Mbur_1952	Sodium/solute symporter
39	1.93	Mbur_1986	Preprotein translocase SecF subunit
		Mbur_1987	Protein-export membrane protein (SecD)
		Mbur_1988	Replication factor C small subunit
		Mbur_1989	Elongator complex protein 3
40	1.71	Mbur_1994	FAE/HPS bifunctional enzyme
		Mbur_1995	Triosephosphate isomerase
41	1.91	Mbur_2051	Phosphoribosylformylglycinamidine synthase II
42	1.67	Mbur_2069	Protein of unknown function UPF0142
		Mbur_2070	Hypothetical protein
43	1.87	Mbur_2087	L-threonine O-3-phosphate decarboxylase
44	1.85	Mbur_2125	Anaerobic ribonucleoside-triphosphate reductase DNA
		Mbur_2126	Anaerobic ribonucleoside-triphosphate reductase activating protein
45	1.83	Mbur_2155 to 2158	Pyruvate synthase subunits γ, δ, β and α
46	2.01	Mbur_2182	HMP kinase / HMP-P kinase
47	1.72	Mbur_2280	DNA-directed RNA polymerase subunit E'
48	1.72	Mbur_2436 and 2437	CoB-CoM heterodisulfide reductase, subunits E and D

The 118 genes represent only 4.7% of all *M. burtonii* genes (2495 in total) and most importantly, do not correspond to the most abundant transcripts in the cell, but cover the entire transcriptional range from low abundance to very high abundance transcripts. This was determined by a simple comparison with the abundance level of the same transcripts in the total RNA: only 27 transcripts are in common between the most abundant transcripts in the total RNA and the 118 transcripts bound by the protein (Fig. 4.8). Furthermore, the transcripts coding for the ribosomal proteins, which are among the most abundant in *M. burtonii* at any condition tested so far, are not bound by the heterodimer at all; on the other hand the strongest binding was for the *Mb*RpoF transcript itself, which is expressed at moderate levels.



(A) (B)

Figure 4. 7: Abundance of the transcripts bound by *Mb*RpoE/F in the total RNA from *M. burtonii*. Fluorescence level obtained from microarrays at 4°C and 23°C for all the *M. burtonii genes* are compared with the those of the 118 genes bound by *Mb*RpoE/F. (A) M/A plot: on y axes are reported the $\log_{10}$ products of the fluorescence values at 4°C and 23°C, this is roughly proportional to expression values of the genes and on the x axes are reported the $\log_2$ ratios of the fluorescence values at 4°C and 23°C. Genes bound by RpoE/F are colored in orange, genes that have very low expression values are evidenced in light grey and are on the left in the graph. (B) Scatter plot: for each gene the mean fluorescence value at 23°C is reported on y axes and the mean fluorescence value at 4°C is reported on the x axes.

Extremely relevant to this result is the fact that, beside the transcript of *Mb*RpoF, the transcript for *Mb*RpoE is bound by the heterodimer. Of particular interest is that, even in such a small number of genes, there are some that have the same function, but are in completely different regions of the genome, like Mbur_0425 and Mbur_1952 (regions 8 and 38, respectively) both coding for a sodium/solute symporter, or Mbur_0605 and Mbur_2155-8 (regions 9 and 47, respectively) coding for pyruvate synthase subunits, and also Mbur_0281-5 and Mbur_1159 (regions 5 and 21, respectively) corresponding to formylmethanofuran dehydrogenase. These data strongly reinforce the idea that the heterodimer has an affinity for particular sequences, and that its intrinsic function is to regulate a subset of genes, possibly encoding a defined subset of functions.

Unfortunately, the bioinformatic analysis described in section 4.2.6 could not identify any motifs associated with the genes bound by *Mb*RpoE/F.

4.1 Conclusion

The hypothesis that *Mb*RpoE/F could bind to RNA molecules in a sequence-dependent manner was validated by affinity purification followed by microarray identification of the target RNA molecules. The data deriving from this approach show that the heterodimer is capable of selecting a narrow subset of transcripts corresponding to only 118 genes. The result is reproducible, it is not due to unspecific binding of RNA molecules to the Ni-Nta resin and it is not the direct reflection of the abundance of the transcripts in the total RNA.

The transcript that is the most strongly bound by *Mb*RpoE/F is the transcript for *Mb*RpoF. The transcript for *Mb*RpoE is also bound by the heterodimer. These data strongly confirm the hypothesis that a feature of RpoE/F could be the ability to bind RNA molecules with an affinity for certain sequences, and that this feature could be linked to a regulatory role for the heterodimer. The biological significance of the microarray data is deeply investigated in

Chapter 5, where the 118 genes were extensively scrutinised in order to infer insights to this regulatory role.

Chapter 5: Gene ontology

5.1 INTRODUCTION

From the microarray experiments 118 genes were identified that satisfied the statistical criteria described in section 4.2.5 and therefore were considered preferentially bound by *Mb*RpoE/F.

A gene ontology analysis was performed in order to understand the biological role of these genes and distinguish any biological processes statistically that are over-represented in the gene list. The result is summarised in the Table 5.11. The 118 genes could easily be divided in distinct categories according to their function: methanogenesis (44 genes); cofactor biosynthesis (14 genes); nucleotide metabolism (10 genes); transcription (8 genes); translation (7 genes); import/export (7 genes); and hypothetical/unknown function (12 genes).

5.2 Materials and Methods

The genome of *M. burtonii*, isolated as described in section 2.2.2, has been sequenced by the Joint Genome Institute (JGI). Critica, Generation and Glimmer were used to identify putative coding regions and proteins were automatically annotated as described for other JGI genomes (Medigue *et al.*, 2005; Chain *et al.*, 2006; Klotz *et al.*, 2006; Ivanova *et al.*, 2007). These automated annotations were manually verified for accuracy by all of the members of Cavicchioli research group using the JGI's integrated microbial genomes (IMG) expert review platform (http://imgweb.jgi-psf.org/cgi-bin/img_er_v260/main.cgi) as described in Allen *et al.* (2009).

Genes bound by *Mb*RpoE/F were first evaluated to identify enrichment of COG (clusters of orthologous groups) categories relative of what would have been expected by chance alone and a hypergeometric distribution was used to determine whether these categories were statistically over-represented. This analysis identifies the probability of observing the number of genes annotated within a particular COG category among the selected group of orthologs. Hypergeometric distribution was calculated using the program *R* in the *R* statistical package **(R Development Core Team: R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing; 2006)** and the significance threshold considered is 0.05.

Genes bound by RpoE/F were also analysed using GoMiner (Zeeberg *et al.*, 2003) in order to find GO classes statistically enriched with genes bound. To better understand the biological role of the transcripts bound by RpoE/F, the genes identified in microarray experiment were assigned to functional categories according to COG and GO annotations. The evidence of genes enrichment in COG specific groups was calculated using hypergeometric distribution, furthermore we used the Fisher exact test (implemented in GoMiner) to analyse the GO in order to give an overview of the biological processes at a higher level of detail, considering for example specific biological mechanisms. In fact COGs give a more general idea of the processes involved, while GO give also information regarding very specific processes (for example "GO:0006814_sodium_ion_transport" or "GO:0030001_metal_ion_transport"). Despite COG analysis is very useful to describe specific processes in bacteria and archaea, we could probably consider GO analysis an useful complementation of COG.

A manual and more meticulous analysis followed in order to group the genes into more specific categories in accordance to their function. The protein sequence for each of these genes was again searched against the Swiss-Prot and Protein Data Bank. BLAST matches were searched for direct experimental verification of the function and matching proteins not functionally characterised were ignored. InterPro and Pfam domains were checked for their presence in the *M. burtonii* homologues to confirm that all identifiable functional domains were conserved.

After this extensive verification it was possible to assign functional categories for all proteins identified and therefore easily infer biology out of the whole dataset.

5.3 Results and Discussion

5.3.1 Functional Categories

According to the hypergeometric distribution the following COG categories were found to be statistically over-represented: energy production and conversion; coenzyme transport and metabolism; nucleotide transport and metabolism; chromatin structure and dynamics; lipid transport and metabolism; intracellular trafficking; secretion; and vesicular transport (see Appendix 2). The results of the GoMiner analysis are located in Appendix 3.

The refined analysis that was then performed resulted in the subdivision of the 118 genes into the following categories: methanogenesis, nucleotide metabolism, cofactor biosynthesis, transcription, translation, import/export, central metabolism, other pathways and hypothetical genes.

5.3.1.1 Methanogenesis

M. burtonii is an obligately methylotrophic methanogen, and therefore obtains energy from the oxidation and reduction of methyl groups (derived from methanol or methylamines) to CO_2 and methane, respectively.

The process is initiated by methyltransferases that demethylate the substrates and transfer the methyl group to coenzyme M (CoM) via their respective corrinoid binding proteins.

Methyl-CoM is then reduced with coenzyme B (CoB) to form methane and the heterodisulfide CoM-S-S-CoB by the enzyme methyl-CoM reductase (Mcr). This heterodisulfide is then reduced back to CoB and CoM by heterodisulfide reductase (Hdr), and the electrochemical proton gradient that is generated by this process is the driving force for ATP synthesis (Ferry and Kastead, 2007).

The electrons necessary for the reduction of CoM-S-S-CoB derive from the complete oxidation of the methyl moiety of methyl-CoM to CO_2. Thus, under this proportional dismutation pathway, the complete oxidation of C1 compounds generates six electrons that can be used to reduce up to three methyl groups to methane. The C1 intermediates are bound to C1-carrier molecules (tetrahydrosarcinapterin [H_4SPT] and methanofuran). At two steps in the oxidative branch of methanogenesis, reducing power is generated in the form of reduced coenzyme F_{420} ($F_{420}H_2$). $F_{420}H_2$ dehydrogenase (Fpo) reoxidises the $F_{420}H_2$, and the electrons transferred to Hdr via methanophenazine, a hydrophobic cofactor. The terminal step in the oxidative branch is mediated by the formylmethanofuran dehydrogenase complex (Fmd or Fwd, molybdenum- and tungsten-dependent, respectively), which produces CO_2 with the concomitant reduction of ferredoxin. The latter is thought to be reoxidised by the Rnf complex (which in conjunction with Hdr participates in energy conservation), although reduced ferredoxin may be utilised by biosynthetic processes (Ferry and Kastead, 2007).

Carbon assimilation for biomass production is initiated by the condensation of a methyl group (from methyl- H_4SPT) with CO_2, and is mediated by the acetyl-CoA decarbonylase/synthase (ACDS) complex, generating acetyl-CoA. Reducing power and CO_2 required by ACDS and by the subsequent step in biomass production, mediated by the pyruvate synthase (POR) complex, are provided by the final oxidation step in Methanogenesis (Ferry and Kastead, 2007).

Of the 118 genes whose transcript was bound by *Mb*RpoE/F, 44 are involved in the methanogenesis pathway. The genes, mostly organised in large operons, are listed in Table 5.1.

Table 5. 1: Methanogenesis genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0280	FdhD
Mbur_0281 to 0285	Formylmethanofuran (Mo) dehydrogenase subunits B, D, C, A and F
Mbur_0605	Pyruvate synthase γ / β -like protein
Mbur_0858 to 0863	ACDS subunits α, ε, β, Ni, δ and γ 2
Mbur_0994	Pyrophosphatase
Mbur_1159	Formylmethanofuran (W) dehydrogenase subunit E like protein
Mbur_1286 to 1293	$F_{420}H_2$ dehydrogenase subunits O, N, M, L, K, I and H
Mbur_1378	UPF0132 family protein, part of Rnf complex
Mbur_1379 to 1384	Rnf electron transport complex proteins B, A, E, G, D, and C
Mbur_1518 to 1525	H_4SPT-methyltransferase subunits H, G, F A, B, C, D and E
Mbur_2155 to 2158	Pyruvate synthase subunits γ, δ, β and α
Mbur_2436 and 2437	CoB-CoM heterodisulfide reductase, subunits E and D

The functions of the identified genes are as following:

- **Mbur_0280-85 and Mbur_1159: Formylmethanofuran dehydrogenase**

Formylmethanofuran dehydrogenases are found in methanogenic and sulfate-reducing archaea; they are composed by several subunits and contain Fe-S clusters and a pterin dinucleotide cofactor. These enzymes catalyse the reversible reduction of CO_2 and methanofuran via N-carboxymethanofuran to formylmethanofuran. In *M. burtonii*, formylmethanofuran dehydrogenase catalyses the final step in the oxidative branch of methanogenesis, which generates CO_2.

There are two isoenzymes of formylmethanofuran dehydrogenase: a tungsten-containing (Fwd) and a molybdenum-containing isoenzyme (Fmd) (Hochheimer *et al.*, 1996; Vorholt *et al.*, 1996; de Poorter *et al.*, 2003; Ferry and Kastead, 2007).

The gene Mbur_0280 codes for a protein (FdhD) of unknown function, that possibly plays a role in the activity of the Fmd complex.

- **Mbur_0605 and Mbur_2155-58: Pyruvate ferredoxin/ oxidoreductase**

Pyruvate ferredoxin oxidoreductase (POR) or pyruvate synthase is a multiprotein complex that catalyses the formation of pyruvate from acetyl-CoA and CO_2. POR therefore catalyses the second step (after acetyl-CoA synthase/decarbonylase) in the assimilation of carbon for biomass production (Kletzin and Adams, 1996; Ferry and Kastead, 2007).

- **Mbur_0858-63: Acetyl-CoA decarbonylase/synthase (ACDS)**

Acetyl-CoA decarbonylase/synthase (ACDS) is a multienzyme complex which carries out the first step in the assimilation of carbon by *M. burtonii*. Using a methyl group (bound to tetrahydrosarcinapterin) and CO_2 both derived from the oxidative branch of methanogensis, ACDS generates acetyl-CoA. ACDS contains a 4Fe-4S ferredoxin domain (Grahame *et al.*, 2005; Ferry and Kastead, 2007).

- **Mbur_0994: Pyrophosphatase**

Membrane-bound proton-pumping pyrophosphatases (H(+)-PPase) are proteins that hydrolyse inorganic pyrophosphate (PPi) to couple the energy released by this hydrolysis to proton movement across biological membranes. Pyrophosphatases have been proposed to play two crucial roles in the metabolism of methylotrophic methanogens: the first is to reduce the amount of free PPi in the cytoplasm which pulls biosynthetic reactions toward completion; the second is the conversion of pyrophosphate bonds into energy to transport

protons against the gradient to generate ATP (Baltscheffsky *et al.*, 1999; Baumer *et al.*, 2002; Ferry and Kastead, 2007).

- **Mbur_1286-93: $F_{420}H_2$ dehydrogenase**

The $F_{420}H_2$ dehydrogenase (Fpo) is a membrane bound multi-subunit complex which is part of the energy conserving electron transport system of methanogens. During growth on methylated substrates, some of the methyl groups of the substrates are oxidised to CO_2, and reducing equivalents are transferred to F_{420}. The Fpo complex reoxidises the reduced coenzyme $F_{420}H_2$ to coenzyme F_{420} via an Fe-S cluster. The electrons transfer between the enzymes is mediated by methanophenazine, a hydrophobic cofactor. The electrochemical proton gradient generated by this process drives ATP synthesis from ADP and Pi (Baumer *et al.*, 2000; Ferry and Kastead, 2007).

- **Mbur_1378-84: Rnf electron transport complex**

Rnf was first discovered in *Rhodobacter capsulatus*, where the six-subunit complex, containing several Fe-S clusters, oxidises reduced nicotinamide adenine dinucleotide (NADH) and reduces the ferredoxin that supplies electrons to nitrogenase, necessary for nitrogen fixation. In methanogens, Rnf homologues are thought to oxidise reduced ferredoxins produced during the production of CO_2 from formylmethanofuran by formylmethanofuran dehydrogenase (Fmd). The electrons gained are then passed in the membrane through methanophenazine to CoB-CoM heterodisulfide reductase (Hdr). This electron transfer is thought to generate a Na^+ gradient that is high outside the membrane that is coupled to ATP synthesis energy-requiring processes (Li *et al.*, 2006; Ferry and Kastead, 2007).

- **Mbur_1518-25: Tetrahydromethanopterin S-methyltransferase**

Tetrahydromethanopterin S-methyltransferase (Mtr) is a membrane- associated enzyme composed of eight subunits, MtrA-H, that in archaea is involved in the energy conservation step of methanogenesis. In *M. burtonii*, this enzyme catalyses the transfer of a methyl group from coenzyme M to the tetrahydrosarcinapterin (H_4SPT) carrier, leading to the formation of methyltetrahydrosarcinapterin (CH_3-H_4SPT). This process involves the translocation of sodium ions into the cell, and therefore consumes energy (Hippler and Thauer, 1999; Welander and Metcalf, 2005; Ferry and Kastead, 2007).

- **Mbur_2436/7: CoB--CoM heterodisulfide reductase**

CoB--CoM heterodisulfide reductase (Hdr) is a heme- and Fe-S-containing complex found in methanogenic archaea. Hdr is composed of three subunits and regenerates coenzyme M and coenzyme B from heterodisulfide (CoM-S-S-CoB), which is formed by methylcoenzyme M reductase together with methane. The Fpo and Rnf complexes each transfer their electrons (via methanophenazine) to Hdr, which links the oxidative and reductive branches of methanogenesis (Bäumer *et al.*, 2000; Ferry and Kastead, 2007).

5.3.1.2 Nucleotide metabolism

Inosine monophosphate (IMP), the nucleotide of the base hypoxanthine, is the first purine derivative to be synthesised in the purine *de novo* biosynthesis pathway. The starting substrate for biosynthesis of the ribosyl moiety is ribose-5-phosphate (R5P) which in archaea is a product of the ribulose monophosphate pathway (RuMP). This substrate is first activated by phosphorylation to form phosphoribosyl pyrophosphate (PRPP) and eventually converted to IMP after a series of enzymatic reactions. IMP is then converted to adenosine monophosphate (AMP) or guanosine monophosphate (GMP) (Becerra and Cazlano, 1996; Zhang *et al.*, 2003; Kato *et al.*, 2006).

The pyrimidine *de novo* biosynthesis pathway is centered on the formation of orotic acid and its subsequent reaction with PRPP to yield orotidine-5-monophosphate (OMP). OMP is decarboxylated to form uridine monophosphate (UMP), which is the substrate for the formation of UTP that also leads to CTP (Becerra and Cazlano, 1996). Of the 118 genes whose transcript was bound by *Mb*RpoE/F, 10 are involved in the *de novo* synthesis of purine and pyrimidine. They are listed in table 5.2.

Table 5. 2: Nucleotide metabolism genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0025	Adenylate kinase
Mbur_0027	Cytidylate kinase
Mbur_0773	D-Ribose-5-phosphate isomerase
Mbur_1160	Adenylosuccinate lyase (PurB)
Mbur_1224 and 1225	Dihydroorotate dehydrogenase electron subunits K and D
Mbur_1994	FAE/HPS bifunctional enzyme
Mbur_2051	Phosphoribosylformylglycinamidine synthase II (PurL)
Mbur_2125	Anaerobic ribonucleoside-triphosphate reductase DNA
Mbur_2126	Radical SAM protein

The functions of the identified genes are as following:

- **Mbur_0025: Adenylate kinase**

Adenylate kinases (ADK) are phosphotransferases that catalyse the reversible reaction AMP + ATP = ADP + ADP

and therefore are essential for the interconversion of adenine nucleotides and play a pivotal role in cellular energy homeostasis (Kath *et al.*, 1993).

- **Mbur_0027: Cytidylate kinase**

Cytidylate kinases catalyse the phosphorylation of cytidine 5-monophosphate to cytidine-5-diphosphate in the presence of ATP or GTP. UMP and dCMP can also act as acceptors (Briozzo *et al.*, 1998; Yan *et al.*, 1999).

- **Mbur_0773: D-Ribose-5-phosphate isomerase**

Ribose 5-phosphate isomerase, also known as phosphoriboisomerase, catalyses the reversible conversion between ribose-5-phosphate and ribulose-5-phosphate, which in archaea derives from the ribulose monophosphate pathway (RuMP) (Zhang *et al.*, 2003; Kato *et al.*, 2006).

- **Mbur_1160: Adenylosuccinate lyase (PurB)**

Adenylosuccinate lyase (or adenylosuccinase) is an enzyme that converts adenylosuccinate to AMP and fumarate.

This enzyme plays a critical role in the *de novo* purine biosynthetic pathway as it is the only enzyme in this pathway to catalyse two separate reactions, enabling it to participate in the addition of a nitrogen at two different positions in adenosine monophosphate (Toth *et al.*, 2000).

- **Mbur_1224 and 1225: Dihydroorotate dehydrogenase**

Dihydroorotate dehydrogenase is an enzyme that catalyses the fourth step in the *de novo* biosynthesis of pyrimidine: the conversions of dihydroorotate to orotate (Nagy *et al.*, 1992).

- **Mbur_1994: FAE/HPS bifunctional enzyme**

Hexulose phosphate synthase (HPS) catalyses the step of the ribulose monophosphate pathway (RuMP) which leads to the formation of ribulose 5-phosphate and formaldehyde (Kato *et al.*, 2006; Orita *et al.*, 2005). Formaldehyde-activating enzyme (FAE) detoxifies formaldehyde by catalysing the condensation of formaldehyde and H_4SPT to form methylene-H_4SPT, an intermediate in the oxidative branch of methanogenesis. In *M. burtonii*, HPS and FAE constitute domains within a single bifunctional enzyme.

- **Mbur_2051: phosphoribosylformylglycinamidine synthase II**

Phosphoribosylformylglycinamidine synthase II catalyses the conversion of formylglycinamide ribotide (FGAR) to formylglycinamide ribotide (FGAM) in the *de novo* purine nucleotide biosynthetic pathway (Gu *et al.*, 1992; Mathews *et al.*, 2006).

- **Mbur_2125 and Mbur_2126: Anaerobic ribonucleoside-triphosphate reductase DNA**

Ribonucleotide reductases (RNRs) are the sole enzymes responsible for the reduction of ribonucleotides to de-oxyribonucleotides, catalysing the substitution of the 2'OH group of a ribonucleotide with a hydrogen by a mechanism involving protein radicals. The mechanism of anaerobic ribonucleotide reductase (class III) (NrdD) involves a glycine-centred radical,

a C-terminal zinc binding site, and a set of conserved active site cysteines and asparagines. It also requires an activating component, a radical-SAM domain containing enzyme (NrdG) which utilises S-adenosyl methionine, an Fe-S cluster protein and a reductant (dihydroflavodoxin) to produce a glycine-centred radical in NrdD (Nordlund and Reichard, 2006).

5.3.1.3 Cofactor Biosynthesis

Among the transcripts bound by *Mb*RpoE/F, 14 are involved in the synthesis of a number of essential cofactors: cofactor F_{430}, heme, siroheme, cobalamin (vitamin B_{12}), thiamine, Fe-S clusters and tetrahydrofolate. The first four derive from the same intermediate (precorrin-2) and share many biosynthetic enzymes. The list of genes is in Table 5.3.

Table 5. 3: Cofactors biosynthesis genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0191	N^1-α-phosphoribosyltransferase
Mbur_0288	Fe-S cluster assembly protein
Mbur_0289	Cysteine desulfurase
Mbur_0981	ABC transporter, iron permease component
Mbur_1105	Uroporphyrinogen-III C-methyltransferase (CysGA)
Mbur_1226	Porphobilinogen deaminase (HemC)
Mbur_1227	Glutamate-1-semialdehyde 2,1-aminomutase (HemL)
Mbur_1699	ABC transporter cobalt-binding protein (CbiN)
Mbur_1700	Cobalamin biosynthesis protein (CbiM)
Mbur_1813	Serine hydroxymethyltransferase
Mbur_1814	Methylene-THF dehydrogenase/cyclohydrolase (FolD)
Mbur_1815	Dihydropteroate synthase (FolP)
Mbur_1872	Ferrous iron transport protein B
Mbur_2182	HMP kinase / HMP-P kinase (ThiD)

The functions of the identified genes are the following:

- **Mbur_0191: N1-α-phosphoribosyltransferase**

Nicotinate mononucleotide (NaMN):5,6-dimethylbenzimidazole (DMB) phosphoribosyltransferase (CobT) plays a pivotal role in the synthesis of α-ribazole-5'-phosphate, an intermediate for the lower ligand of cobalamin and the assembly of the nucleotide loop. This enzyme catalyses the transfer of the phosphoribosyl moiety of NaMN onto DMB to yeald α-ribazole5'-phosphate (Trzebiatowski *et al.*, 1994).

- **Mbur_0288: Fe-S assembly protein NifU-like**

U-type scaffold proteins (N-terminal domain of NifU) provide an intermediate site for Fe–S cluster assembly in the NIF system. These Fe-S cluster precursors are then transferred to the apo-form of Fe-S proteins for their maturation. They all contain three conserved cysteine residues that are required for cluster incorporation (Johnson *et al.*, 2005).

- **Mbur_0289: Cysteine desulfurase NifS-like**

NifS is a cysteine desulfurase whose function is to provide the inorganic sulfide necessary for the synthesis of the Fe-S clusters. It has been shown to be a pyridoxal-phosphate-containing homodimer that catalyses the specific desulfurisation of L-cysteine (Johnson *et al.*, 2005).

- **Mbur_0981: Citochrome c assembly protein**

CcmC is a six transmembrane domain protein necessary for transport of heme for cytochrome c assembly in the periplasm (Kranz *et al.*, 2009).

- **Mbur_1105: Uroporphyrin-III C-methyltransferase**

This homodimer catalyses the transfer of two methyl groups, each donated by S-adenosylmethionine (SAM), to the C2 and C7 position of uroporphyrinogen III (Raux *et al.*, 2000).

- **Mbur_1226: Porphobilinogen deaminase**

Porphobilinogen deaminase catalyses the condensation of four porphobilinogen molecules into the tetrapyrrole structure, preuroporphyrinogen, with the concomitant release of four molecules of ammonia (Raux *et al.*, 2000).

- **Mbur_1227: Glutamate-1-semialdehyde-2,1-aminomutase**

Glutamate-1-semialdehyde (GSA) aminotransferase catalyses a transamination reaction to produce 5-aminoaevulinic acid during the first stage of tetrapyrrole biosynthesis by the C5 pathway. This enzyme has a high degree of similarity to amino acid transaminases, and is classed as a class III aminotransferase (Raux *et al.*, 2000).

- **Mbur_1699-1700: Cobalt transport binding protein CbiM and CbiN**

CbiM and CbiN are component of a cobalt-transport complex found in bacteria and archaea involved in uptake of cobalt necessary for cobalamin biosynthesis. CbiM is an integral membrane protein while CbiN is a small membrane-bound component (Rodionov *et al.*, 2006).

- **Mbur_1813: Serine hydroxymethyltransferase (GlyA)**

Serine hydroxymethyltransferases are pyridoxal phosphate (PLP) dependent enzymes that generate glycine and methylenetetrahydrofolate from serine, in a reversible reaction (Christensen and MacKenzie, 2006).

- **Mbur_1814: Methylene-THF dehydrogenase/cyclohydrolase (FolD)**

NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase 5,10 methenyltetrahydrofolate cyclohydrolase is a bifunctional enzyme responsible for the interconversion between the C1 carriers methyl, methylene and formyl tetrahydrofolate (Christensen and MacKenzie, 2006).

- **Mbur_1815: dihydropteroate synthase (FolP)**

The enzyme dihydropteroate synthase catalyses the condensation of 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate to para-aminobenzoic acid to form 7,8-dihydropteroate. This is an essential step in the synthesis of tetrahydrofolate (Christensen and MacKenzie, 2006).

- **Mbur_1872: Ferrous iron transport protein B**

The major pathway for uptake of the soluble Fe^{2+} in prokaryotes is the Feo system which has an important role in the iron supply of the cell in anaerobic conditions. The key component, FeoB, is a transmembrane protein which either works itself as a ferrous iron transporter or mediates other processes that eventually result in Fe^{2+} uptake (Kammler *et al.*, 1993; Cartron *et al.*, 2006; Eng *et al.*, 2008, Köster *et al.*, 2009).

- **Mbur_2182: Phosphomethylpyrimidine kinase**

Thiamine contains a pyrimidine and a thiazole ring linked by a methylene bridge. The thiazole and pyrimidine moieties are synthesised separately and then assembled together. Phosphomethylpyrimidine kinase (ThiD) is involved in the synthesis of the pyrimidine ring (Jungerson *et al.*, 2009).

5.3.1.3.1 Tetrapyrroles synthesis pathway

Tetrapyrroles are indispensable metabolites that are found in all the domains of life, from archaea to plants to animals. They serve as pigments and cofactors in a broad spectrum of essential biological reactions (Table 5.4).

Table 5.4: List of modified tetrapyrroles found in nature.

Tetrapyrrole	Metal Ion	Function	Protein complex
Chlorophyll, Bacteriochrorophyll	Mg	Light harvesting	Chlorophyll binding proteins
Heme	Fe	Respiration,	Cytocrome, Catalase,

		phosphorylation, removal of reactive oxygen species, detoxification, N_2 fixation, O_2 transport, storage	Peroxidase, Leghemoglobin, Hemoglobin, Myoglobin,
Phycobiline	-	Light harvesting	Phycobilisomes
Phytochromobiline	-	Light perception	Phytochrome
Siroheme	Fe	Sulfate and nitrate riduction	Sulfite and nitrite reductase
Coenzyme F_{430}	Ni	Methanogenesis	Methylcoenzyme M reductase
Corrinoids	Co	Coenzyme vitamin B_{12}	Methylmalonyl coenzyme A mutase

The biosynthetic pathway of tetrapyrroles is one of the most conserved metabolic pathways known, with most reactions being common to all organisms investigated so date.

The first stage of the pathway is the synthesis of an unusual aminoacid, 5-aminoaevulinic acid (ALA), from compounds of primary metabolism. The structural backbone of all tetrapyrroles is assembled from eight ALA molecules (Raux *et al.*, 2000).

In α-proteobacteria, fungi and mammals ALA is synthesised from the condensation of succinyl coenzyme A and glycine catalysed by ALA synthase (ALAS) as part of the Shemin pathway (Raux *et al.*, 2000).

In all other bacteria, archaea and plants ALA derives from the C5-skeleton of glutamate using tRNA-bound glutamate as substrate. This substrate is first charged with tRNA by the enzyme glutamyl-tRNA synthetase (GltX), and then reduced to glutamate-1-semialdehyde (GSA) by glutamyl-tRNA reductase (HemA). GSA is then transaminated to ALA by glutamate-1-semialdehyde 2,1-aminomutase (HemL) (Raux *et al.*, 2000).

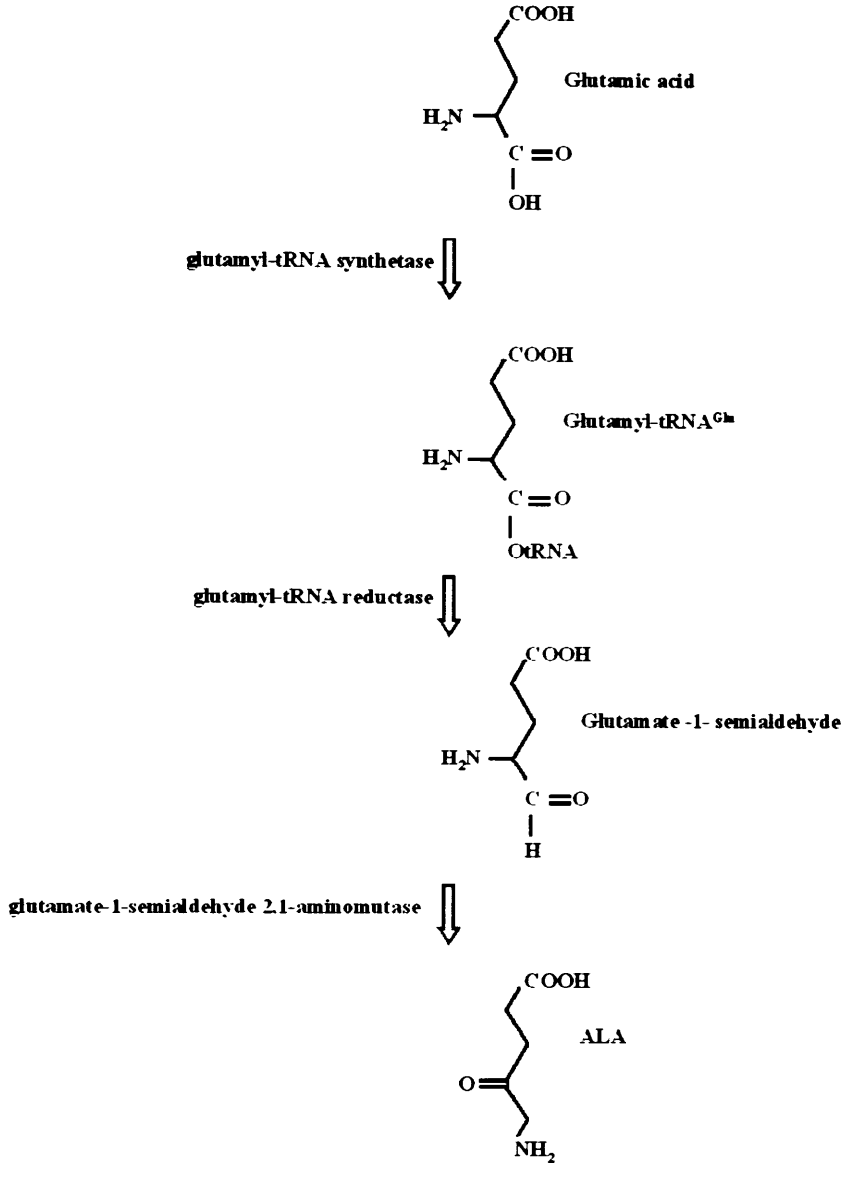


Figure 5. 1: Synthesis of 5-aminoaevulinic acid (ALA) from glutamate in archaea. (Raux *et al.*, 2000).

The enzyme porphobilinogen (PBG) synthase (or ALA dehydratase, HemB) catalyses the condensation of two ALA molecules to form the pyrrole porphobilinogen (PBG).

Four PBG molecules are subsequently condensed by the enzyme porphobilinogen deaminase (commonly referred to as hydroxymethylbilane synthase or HMBS) to form preuroporphyrinogen, a linear tetrapyrrole, which is then cycled and rearranged into uroporphyrinogen III by the enzyme uroporphyrinogen III synthase (HemD).

Uroporphyrinogen III is the first branch point of the pathway. To synthesise heme and chlorophyll, uroporphyrinogen III is generally decarboxylated into coproporphyrinogen III by the action of uroporphyrinogen III decarboxylase (UROD) (Raux *et al.*, 2000).

To synthesise cobalamin (vitamin B_{12}), siroheme, and coenzyme F_{430}, uroporphyrinogen III needs to be methylated to produce precorrin-2 by the action of S-adenosyl-L-methionine uroporphyrinogen III methyltransferase (SUMT). This homodimer catalyses the ordered transfer of two methyl groups, each donated by S-adenosylmethionine, to the C2 and C7 position of uroporphyrinogen III, respectively (Raux *et al.*, 2000).

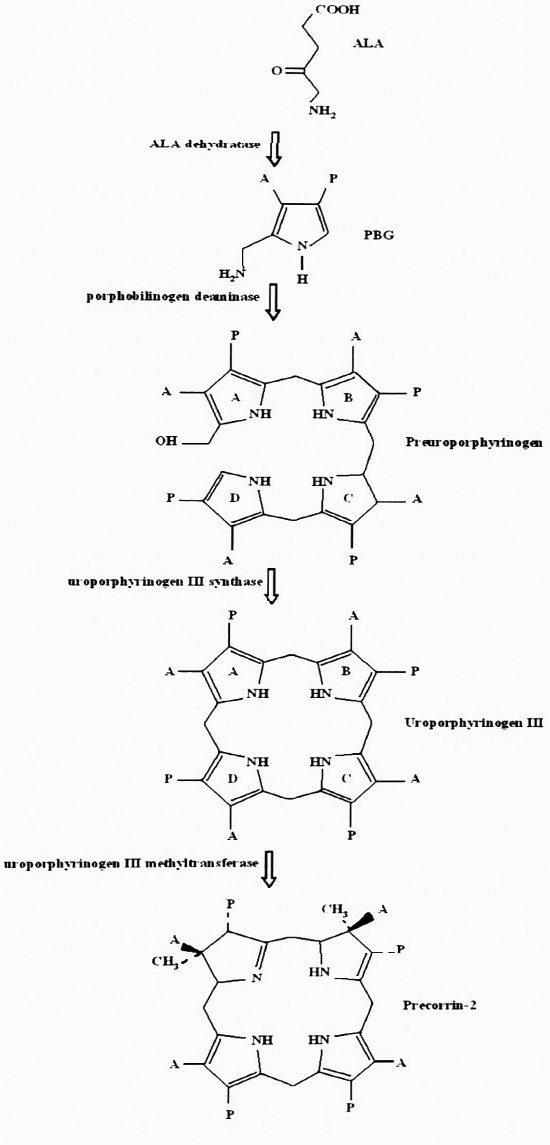


Figure 5. 2: Synthesis of precorrin-2 from ALA in archaea. (Raux *et al.*, 2000).

Many archaea, *M. burtonii* included, contain heme proteins such as cytochromes and catalases, but the way in which they synthesise this heme is still unclear, as their genomes lack homologs of the known genes involved in the pathway of protoheme biosynthesis of eucarya and most bacteria. Studies on the anaerobe *D. vulgaris* have shown that the synthesis of protoheme in this microorganism involves precorrin-2 as an intermediate (Ishida *et al.*, 1998). More recent studies on *M. barkeri* have shown that this pathway is most likely to be also operative in this methanogenic archaeon (Buchenau *et al.*, 2006).

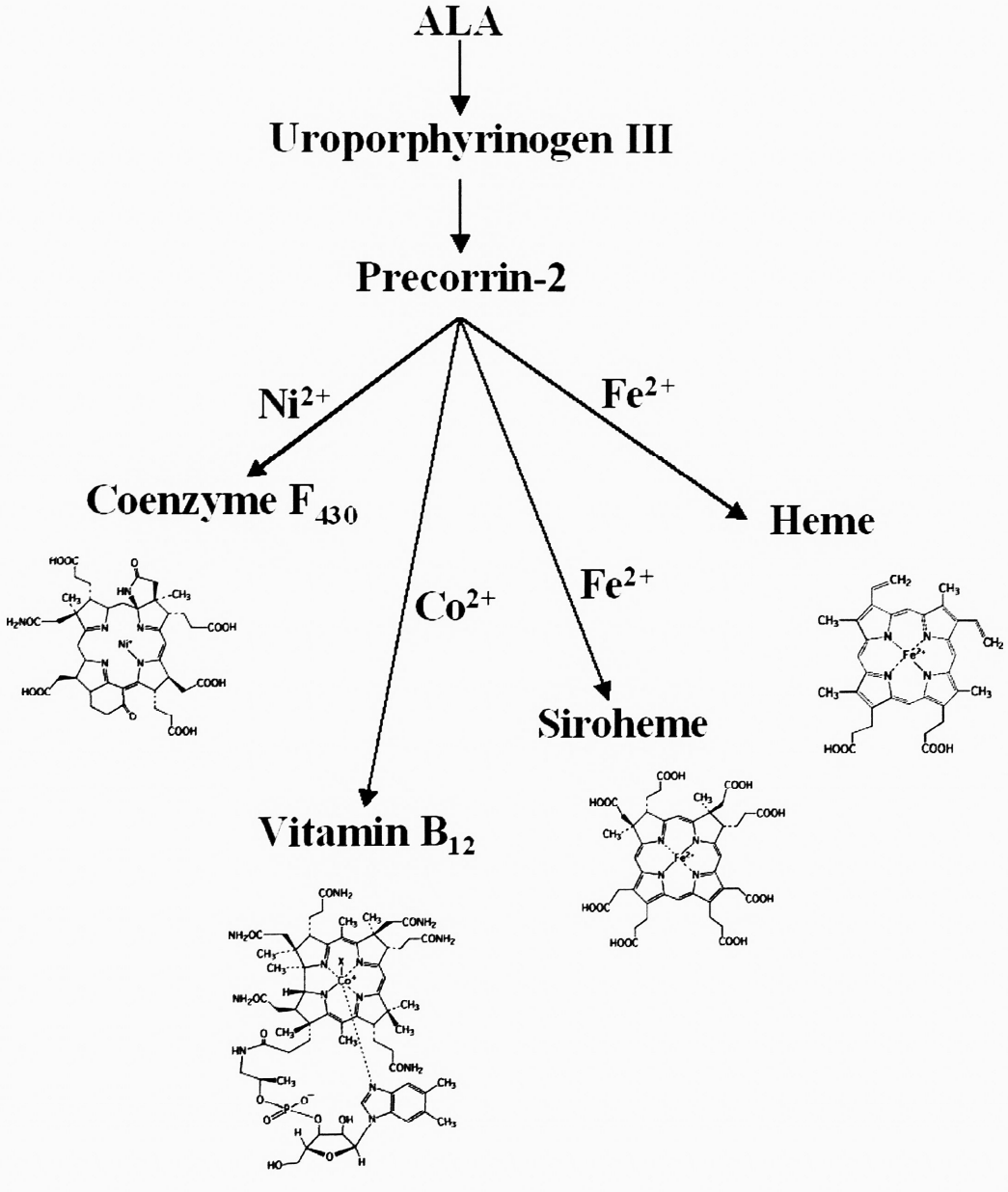


Figure 5.3: Branched biosynthesis of the modified tetrapyrroles. The branched biosynthesis of the modified tetrapyrroles is outlined, revealing how precorrin-2 acts as the last common intermediate for

coenzyme F_{430}, vitamin B_{12}, heme and sirohaem synthesis, while protoporphyrin acts as the last common intermediate in the synthesis of haem and chlorophyll. The pathway for the synthesis of haem via precorrin-2 has recently been described for methanogens (Raux *et al.*, 2000; Buchenau *et al.*, 2006).

* Vitamin B_{12}, first isolated in 1948, is the largest, most complicated vitamin yet discovered. Although its biochemical role has not been completely elucidated, vitamin B_{12} - also known as cobalamin - has been shown to be required as a cofactor by several key metabolic enzymes, such as methionine synthase and methylmalonyl CoA mutase. It is synthesised only by archaea and some bacteria; no eucaryote is capable of synthesizing vitamin B_{12} *de novo*.

The cobalamin molecule has three parts: a central corrin ring with a cobal atom at its center (cobinamide), an upper ligand (5'deoxyadenosyl, methyl and cyano group the most common), and a nucleotide loop (Raux *et al.*, 2000).

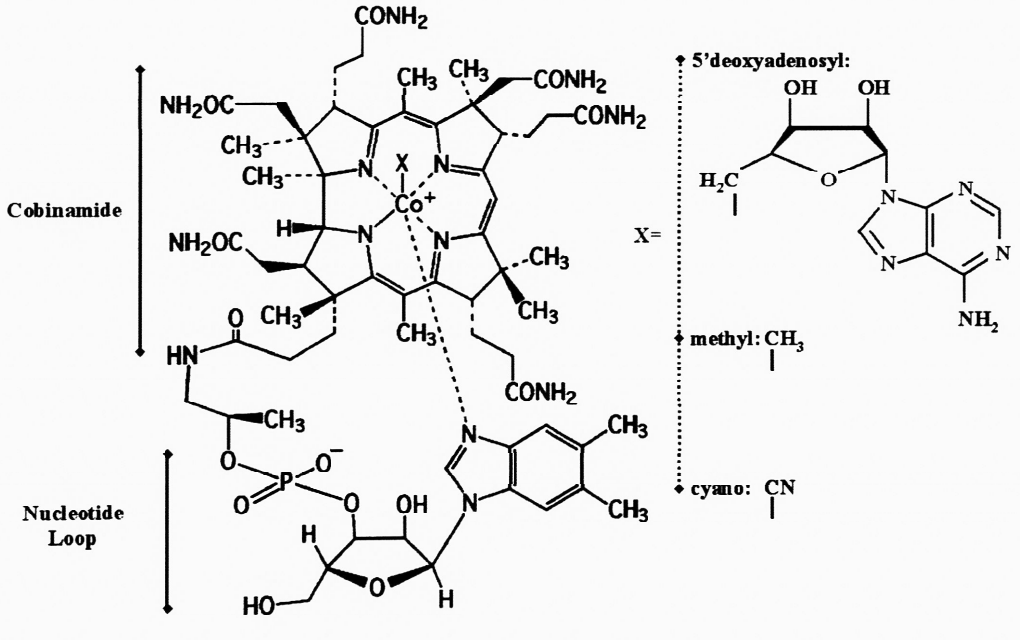


Figure 5. 4: Cobalamin with the upper axial ligand noted with X (-CN in the case of cyanocobalamin, not found in nature; -CH_3 in methylcobalamin, an active form of vitamin B_{12}; -deoxyadenosyl in adenosylcobalamin, another active form of vitamin B_{12}) (Graham *et al.*, 2007).

The biosynthetic pathway that leads to the synthesis of cobalamin from uroporphyrinogen III is highly complex, requiring a large number of enzyme-mediated steps to account for the peripheral methylation, ring contraction, decarboxylation, cobalt insertion, amidations, lower nucleotide loop assembly and attachment of the upper ligand. A total of at least 19 enzyme-mediated reactions are required for this process, by at least two independent pathways: one anaerobic and the other aerobic. These pathways exhibit different chemical approaches, using different set of enzymes, to arrive at the same end product; the main different between the two is that in the anaerobic pathway cobalt is added early in the biosynthetic sequence, while in the aerobic this happens at a comparatively late stage. Cobalt required for this process is up taken via a cobalt-transport complex, CbiMNQO, consisting of two transmembrane components (CbiM and CbiQ), a small membrane-bound component (CbiN) and an ATP-binding protein (CbiO) (Graham *et al.*, 2007).

In the final steps of the pathway the nucleotide loop that ties the lower ligand base to the ring macrocycle is assembled. Nicotinate mononucleotide (NaMN):5,6-dimethylbenzimidazole (DMB) phosphoribosyltransferase (CobT) plays a central role in the synthesis of α-ribazole-5'-phosphate, an intermediate for the lower ligand of cobalamin and the assembly of the nucleotide loop. This enzyme catalyses the transfer of the phosphoribosyl moiety of NaMN onto DMB to yield α-ribazole5'-phosphate (Escalante-Sementera *et al.*, 2007).

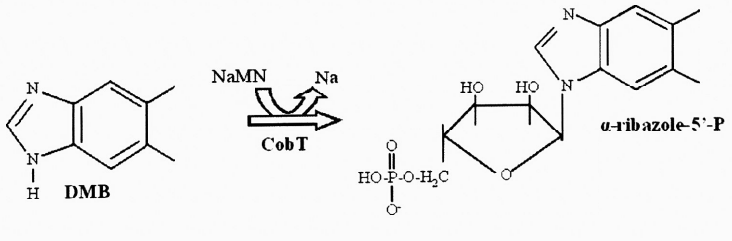


Figure 5. 5: Reaction catalysed by the enzyme (NaMN):5,6-dimethylbenzimidazole (DMB) phosphoribosyltransferase, also known as N1-α-phosphoribosyltransferase (CobT). The enzyme catalyses the transfer of the phosphoribosyl moiety of NaMN onto DMB to yeald α-ribazole5'-phosphate (Trzebiatowski *et al.*, 1994; Escalante-Sementera *et al.*, 2007).

Our knowledge on cobalamin biosynthesis derives from studies performed on bacteria; it is known that some archaea synthesise and require cobalamin for survival, especially the methanogens, as they require this cofactor for methanogenesis. So far, every sequenced archaeal genome appears to encode several orthologs to genes required for nucleotide loop assembly and several have orthologs to the genes required for the *de novo* pathway (Graham *et al.*, 2009).

* Coenzyme F_{430} is the most extensively reduced tetrapyrrole found in nature, with a centrally chelated nickel ion. It is the prosthetic group of methyl-coenzyme M reductase (MCR), the key enzyme in methanogenesis, which catalyses the reduction of methyl-coenzyme M (CH_3-S-CoM) with coenzyme B (HS-CoB) to CH_4 and the mixed disulfide CoM-S-S-CoB. All methanogenic species investigated so far contain a methyl-coenzyme M reductase and coenzyme F_{430} (Warren *et al.*, 2007).

* Heme cofactors are found in a wide range of proteins, where they play various roles in lipid synthesis, energy transduction, gene regulation, cellular signaling, and oxygen transport. The diversity of physiological functions performed by hemoproteins means that heme is among the most versatile of protein cofactors. Aside from electron transferase functions observed in respiratory cytochromes, several hemoprotein sensory or catalytic functions are recognised that involve the binding of gaseous ligands to the heme iron and/or the dissociation or switching of amino acid side chains as axial ligands to the iron (Munro *et al.*, 2007).

* Siroheme is a heme-like prosthetic group first identified in sulfite reductases and subsequently in nitrite reductases, enzymes that catalyse the six-electron reduction of sulfite and nitrite to sulfide and ammonia, respectively (Warren *et al.*, 2007).

5.3.1.3.2 Fe-S cluster synthesis

Fe-S clusters are ubiquitous prosthetic groups present in several different types of enzymes or proteins mainly involved in electron transfer, in redox and non-redox catalysis. The biosynthetic machineries for Fe–S cluster biogenesis are widely conserved in all three kingdoms of life. Fe-S cluster assembly is a complex process involving the mobilisation of Fe and S atoms from storage sources, their cysteine desulfurase-mediated assembly into [Fe-S] form, their transport to specific cellular locations, and subsequent transfer of the clusters to apoproteins. Three distinct Fe-S assembly machineries have been described: ISC (Fe-S cluster), SUF (sulfur assimilation), and NIF (nitrogen fixation) systems.

Nif proteins involved in the formation of Fe-S clusters can also be found in organisms that do not fix nitrogen (Hwang *et al.*, 1996).

In these systems, NifS and NifU are required for the formation of metalloclusters of nitrogenase and maturation of other Fe-S proteins. NifS is a cysteine desulfurase which provides the inorganic sulfide required for the synthesis of the Fe-S clusters, while NifU in involved in the synthesis of a Fe-S cluster precursor (Johnson *et al.*, 2005).

5.3.1.3.3 Thiamine synthesis

Thiamine (vitamin B_1) is an essential vitamin whose phosphate derivatives are involved in many cellular processes. The most common and best characterised form is thiamine pyrophosphate, an indispensable coenzyme in many enzyme complexes, including POR.

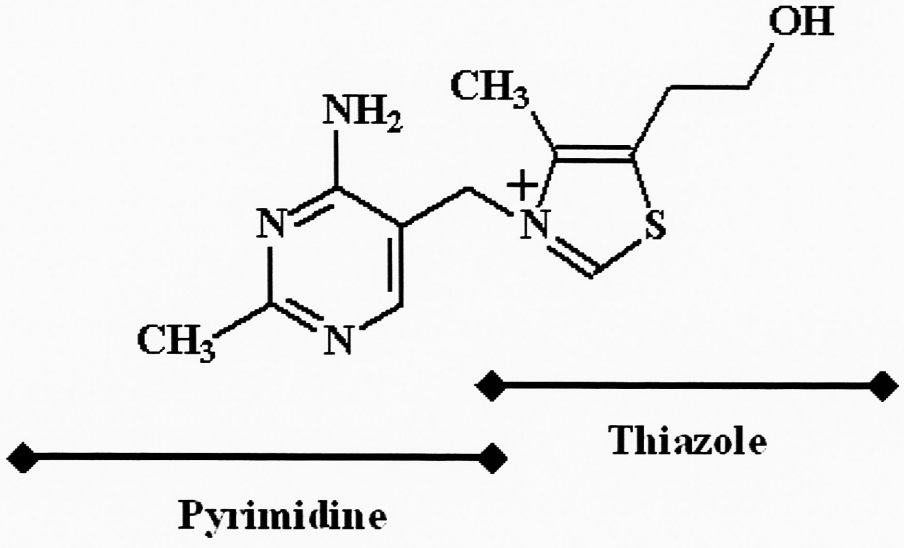


Figure 5. 6: The chemical structure of thiamine, with the pyrimidine and thiazole moieties.

The thiazole and pyrimidine moieties of thiamine are separately synthesised: thiazole phosphate is formed from glycine, cysteine and deoxy-xylulose-5-phosphate, pyrimidine phosphate is formed from aminoimidazole ribonucleotide. Pyrimidine phosphate is then phosphorylated and coupled with thiazole phosphate to give thiamine phosphate. A final phosphorylation gives thiamine pyrophosphate, the active form of the cofactor (Jungerson *et al.*, 2009).

5.3.1.3.4. Tetrahydrofolate synthesis and conversion

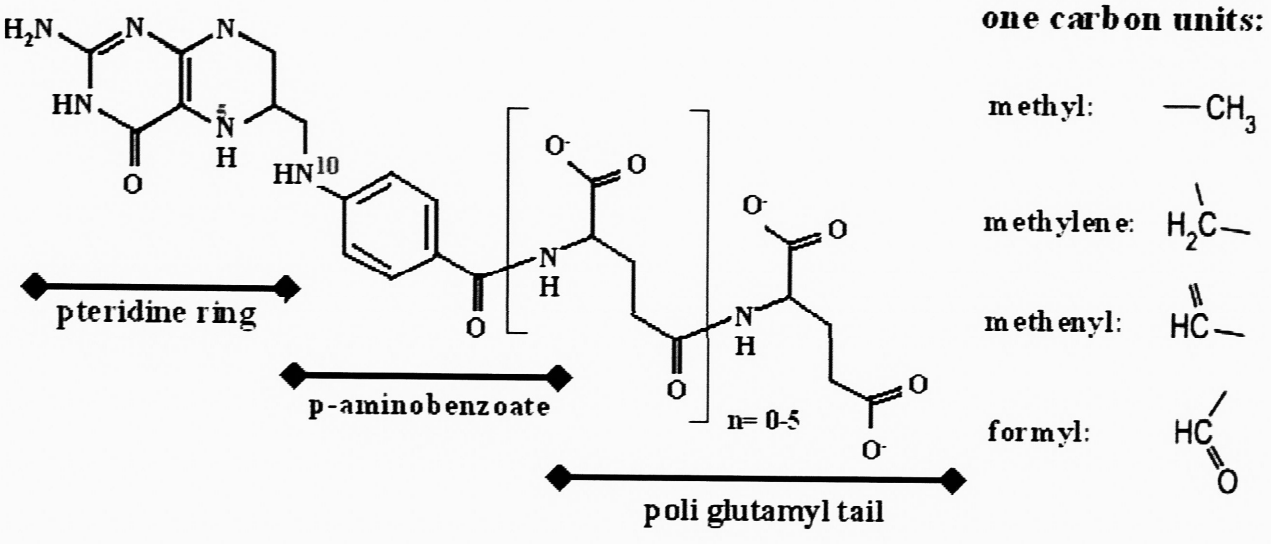


Figure 5. 7: The chemical structure of tetrahydrofolate. The molecule is made up of three elements: a pteridine ring, a p-aminobenzoic acid moiety and a polyglutamate tail. One-carbon units, such as methyl, methylene, methenyl and formyl groups, are added at positions N^5 and N^{10} of the pteridine ring, shown in red.

Folates are acceptors and donors of one carbon (C_1) units for all oxidation levels of carbon except CO_2. They are required in several biological processes, like the synthesis of purines and thymidylate, the support of cellular methylation reactions via recycling of homocysteine to methionine, and the synthesis of S-adenosylmethionine. The active form is tetrahydrofolate (THF) and its derivatives forms, such as methyl, methylene and formyl-THF, are involved in the transfer of C_1 units in multiple oxidation states (Christensen and MacKenzie, 2006).

The THF biosynthesis pathway is centered on the condensation of dihydropterin (synthesised from GTP) and para- α benzoate (*p*-ABA, synthesised from chorismate) to form dihydropteroate. This reaction is catalysed by the enzyme dihydropteroate synthase (folP). Dihydropteroate is then converted into THF via a series of reactions. The C_1 unit carried by THF can come from formate, glycine or serine and the enzymes responsible for this interconversions are methylene-THF dehydrogenase, methenyl-THF cyclohydrolase

and formyl-THF synthetase. One-carbon folates are required for a number of metabolically important pathways, such as the synthesis of purines, thymidylate and methionine (Christensen and MacKenzie, 2006).

5.3.1.4 Transcription

*Mb*RpoE/F bound to 8 transcripts corresponding to genes involved in the transcription system; 6 of these are subunits of the RNAP and among them are the genes coding for *Mb*RpoE and *Mb*RpoF. They are listed in Table 5.5.

Table 5. 5: Transcription genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0904	DNA-directed RNA polymerase subunit F
Mbur_0963	Transcription initiation factor B
Mbur_1176 to 1179	DNA-directed RNA polymerase subunits A'', A', B' and B''
Mbur_1989	Elongator complex protein 3 (Elp3)
Mbur_2280	DNA-directed RNA polymerase subunit E'

Function of the identified genes:

- **Mbur_0904, Mbur_1176-79 and Mbur_2280:** DNA-directed RNA polymerase subunit F, A'', A', B', B'' and E respectively.

See section 1.5

- **Mbur_0963: Transcription initiation factor B**

Transcription factor B is of key importance in the archaeal transcription apparatus, as it recognises the AT-rich promoter element (TATA box), which is the first step in transcription initiation. (Bell and Jackson, 1998; Thomm, 2007).

- **Mbur_1989: Elongator complex protein 3 (Elp3)**

The Elongator complex was first identified in *Saccharomyces cerevisiae* as an N-terminal acetyltransferase that co-purified with hyperphosphorylated RNAPII in high-salt conditions. It is a component of the RNA polymerase II holoenzyme, designated Elp3p for Elongator Protein 3, and is involved in transcription elongation (Paraskevopoulou *et al.*, 2006).

5.3.1.5 Translation

*Mb*RpoE/F bound to 7 transcripts corresponding to genes involved in the translation system; four of these are involved in the synthesis of aminoacyl-tRNA synthetases and 2 are related to the production of modified tRNA bases. They are listed in Table 5.6.

Table 5. 6: Translation genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0028	tRNA pseudouridine synthase B (TruB)
Mbur_0190	Prolyl-tRNA synthetase
Mbur_0772	Aspartyl-tRNA synthetase
Mbur_0961	tRNA (guanine-N(1)-)-methyltransferase
Mbur_1324	Threonyl-tRNA synthetase
Mbur_1654 and 1655	Asn /Gln -tRNA amidotransferase subunit B and A

Function of the identified genes:

- **Mbur_0028: tRNA pseudouridine synthase B**

Pseudouridine synthases catalyse the conversion of uridine to pseudouridine (the most abundant modified nucleotide found in all RNAs) in a variety of RNA molecules, and may function as RNA chaperones (Hoang *et al.*, 2001).

- **Mbur_0190, Mbur_0772, Mbur_1324: Aminoacyl-tRNA synthetases (prolyl, aspartyl and threonlyl, respectively)**

Aminoacyl-tRNA synthetases catalyse the attachment of a specific amino acid or its precursor to one of all its compatible cognate tRNA molecules to form an aminoacyl-tRNA; Prolyl-tRNA synthetase, Aspartyl tRNA synthetase, Threonyl-tRNA synthetase (Londei, 2007).

- **Mbur_0961: tRNA (guanine-N(1)-)-methyltransferase**

tRNA (guanine-N(1)-)-methyltransferase tRNA (guanine-N1-)-methyltransferases catalyse the methylation of guanosine to methylguanosine in tRNA molecules. The presence of this modified base improves cellular growth rate (Hagervall et al., 1993).

- **Mbur_1654 and 1655: Asn /Gln -tRNA amidotransferase subunit B and A**

Glutamyl-tRNA(Gln) amidotransferase catalise the production of correctly charged Gln-tRNA(Gln) through the transamidation of mis-acylated Glu-tRNA(Gln) in organisms which lack glutaminyl-tRNA synthetase (Londei, 2007).

5.3.1.6 Import/Export

*Mb*RpoE/F bound to 7 transcripts corresponding to genes involved in import/export; 4 of these are components of the Sec pathway. They are listed in Table 5.7.

Table 5. 7: Import/export genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0024	Preprotein translocase SecY subunit
Mbur_0110	Signal recognition docking protein FtsY
Mbur_0425	Sodium/solute symporter
Mbur_0962	Lysine export domain protein
Mbur_1952	Sodium/solute symporter
Mbur_1986	Preprotein translocase SecF subunit
Mbur_1987	Protein translocase SecD subunit

Function of the identified genes:

- **Mbur_0024, Mbur_1986/7: Preprotein translocase subunit SecY, sec F and sec D, respectively**

The secY subunit is one of the core component of the sec translocon, the first to interact with the signal sequences of the secretory proteins presented by the FtsY protein. SecY forms a channel-like structure for the translocation of the nascent polypeptide. SecD and SecF proteins form a membrane complex that cooperates with the core translocon (Bolhuis, 2004).

- **Mbur_0110: Signal recognition docking protein FtsY**

Protein targeting by the signal recognition particle (SRP) is universally conserved and starts with the recognition of a signal sequence in the context of a translating ribosome. FtsY, a multidomain protein with guanosine triphosphatase (GTPase) activity, is an SRP receptor (SR), that is responsible for directing the SRP–ribosome nascent chain complex to the available Sec translocation channels (Bolhuis, 2004).

- **Mbur_0425 Mbur_1952: Sodium/solute symporter**

Sodium/substrate symporters are involved in the transport of a solute across the cytoplasmic membrane. The electrochemical sodium gradient is used to transport the solute against its concentration gradient (Jung, 2002).

- **Mbur_0962: Lysine export domain protein**

Lysine exporters are involved in the efflux of excess L-lysine as a control for intracellular levels of L-lysine (Eggeling and Sahm, 2003).

The sec pathway

In all domains of life, the Sec pathway is a commonly used protein translocation pathway for passage of secretory proteins to their extracellular destinations (Bolhuis, 2004; Pohlschröder *et al.*, 2004). The Sec pathway involves transport of proteins from the cytosol across the plasma membrane via the Sec translocon, which is a complex of integral membrane proteins: a core component (SecY, SecE and Sec61γ) and two additional membrane proteins (SecD and SecF) that appear to be involved in the release of the mature

peptide into the periplasm (Mori & Ito, 2001; Ring & Eichler, 2004). Proteins destined for secretion need to be targeted to the membrane embedded Sec translocon pore (Eichler & Moll, 2001). Translocation of proteins can occur co-translationally with the help of the universally conserved signal recognition particle (SRP) and a signal recognition particle docking protein (FtsY) or post-translationally with the help of certain chaperones. The signal sequence located at the N-terminus of the preprotein determines the targeting pathway employed (Pohlschröder *et al.*, 2004).

5.3.1.7. Central Metabolism, Other Pathways and Hypothetical Genes

Central metabolism. The heterodimer bound the transcripts coding for a number proteins that can be linked to central carbon metabolism, with many proteins involved in tricarboxylic acid (TCA) cycle (truncated in *M. burtonii*), glycolysis and gluconeogenesis serving a pivotal role. As well as being a TCA enzyme, aconitase B has a role in post-transcriptional regulation of gene expression; ADP-forming acetyl-CoA synthetase represents the major energy-conserving reaction during pyruvate and sugar conversion to acetate in archaea; phosphoglycerate kinase represents the first ATP-generating step in glycolysis (and is therefore subject to intense regulation); and fructose 1-6-bisphosphatase catalyses an irreversible step in gluconeogenesis. Other transcripts bound by *Mb*RpoE/F can be tied to amino acid biosynthesis, and a variety of other metabolic processes.

Table 5. 8: Methanogenesis genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0316	Aconitate hydratase 2 (aconitase B) (AcnB)
Mbur_1345	Phosphoglycerate kinase (Pgk)
Mbur_1995	Triosephosphate isomerase (TpiA)
Mbur_1354	Fructose 1-6-bisphosphatase (GlpX)

Mbur_1347	Acetyl-CoA synthetase (ADP-forming), α and β subunits
Mbur_1087	Homoserine dehydrogenase (Hom)
Mbur_1883	Homoaconitase large subunit (HacA)
Mbur_2087	L-threonine O-3-phosphate decarboxylase (CobD)

Table 5. 9: Miscellaneous genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_1906	2-Hydroxyglutaryl-CoA dehydratase component A
Mbur_0939	FAD dependent oxidoreductase
Mbur_1088	DNA ligase 1, ATP-dependent (Dnl1)
Mbur_1353	Radical SAM family protein
Mbur_1932	3-Hydroxy-3-methylglutaryl-CoA synthase
Mbur_1933	Acetoacetyl-CoA thiolase
Mbur_1935	Adenylyl cyclase-family protein (ATP to 3',5'-cyclic AMP)
Mbur_1988	Replication factor C small subunit

Table 5. 10: Hypothetical genes whose transcripts are bound by *Mb*RpoE/F.

Gene name	Product name
Mbur_0026	Integral membrane protein DUF106
Mbur_0129	Hypothetical protein
Mbur_0850	Hypothetical protein
Mbur_1108	Hypothetical protein
Mbur_1252	Hypothetical protein
Mbur_1344	Hypothetical protein
Mbur_1346	Protein with DRTGG domain
Mbur_1907	Hypothetical protein
Mbur_1908	Protein of unknown function UCP019164
Mbur_1934	DUF35-domain protein
Mbur_2069	Protein of unknown function UPF0142

5.3.2 Significance of the Microarray Data

The vast majority of the 118 genes whose transcript was bound by *Mb*RpoE/F could easily be divided in very precise and distinct categories: methanogenesis, nucleotide metabolism, cofactors biosynthesis, transcription, translation, import/export; most important, for each category the genes identified appear to code for the key regulatory enzymes in the same category.

The methanogenesis genes code exclusively for the major complexes of the pathway: Fpo (Mbur_1286 to 1293, region 25), Rnf (Mbur_1378 to 1384, region 29), Hdr (Mbur_2436 and 2437, region 48), Mtr (Mbur_1518 to 1525, region 30), Fmd/Fwd (Mbur_0280 to 0285, region 5; Mbur_1159, region 21), ACDS (Mbur_0858 to 0863, region 12) and POR (Mbur_2155 to 2158, region 45; Mbur_0605 region 9). The transcripts coding for the enzyme methyl-CoM reductase (Mcr), pivotal in methanogenesis, are not bound by *Mb*RpoE/F. However, the heterodimer binds to the mRNA coding for a number of enzymes involved in the synthesis of cofactor F_{430}, essential for the activity of Mcr.

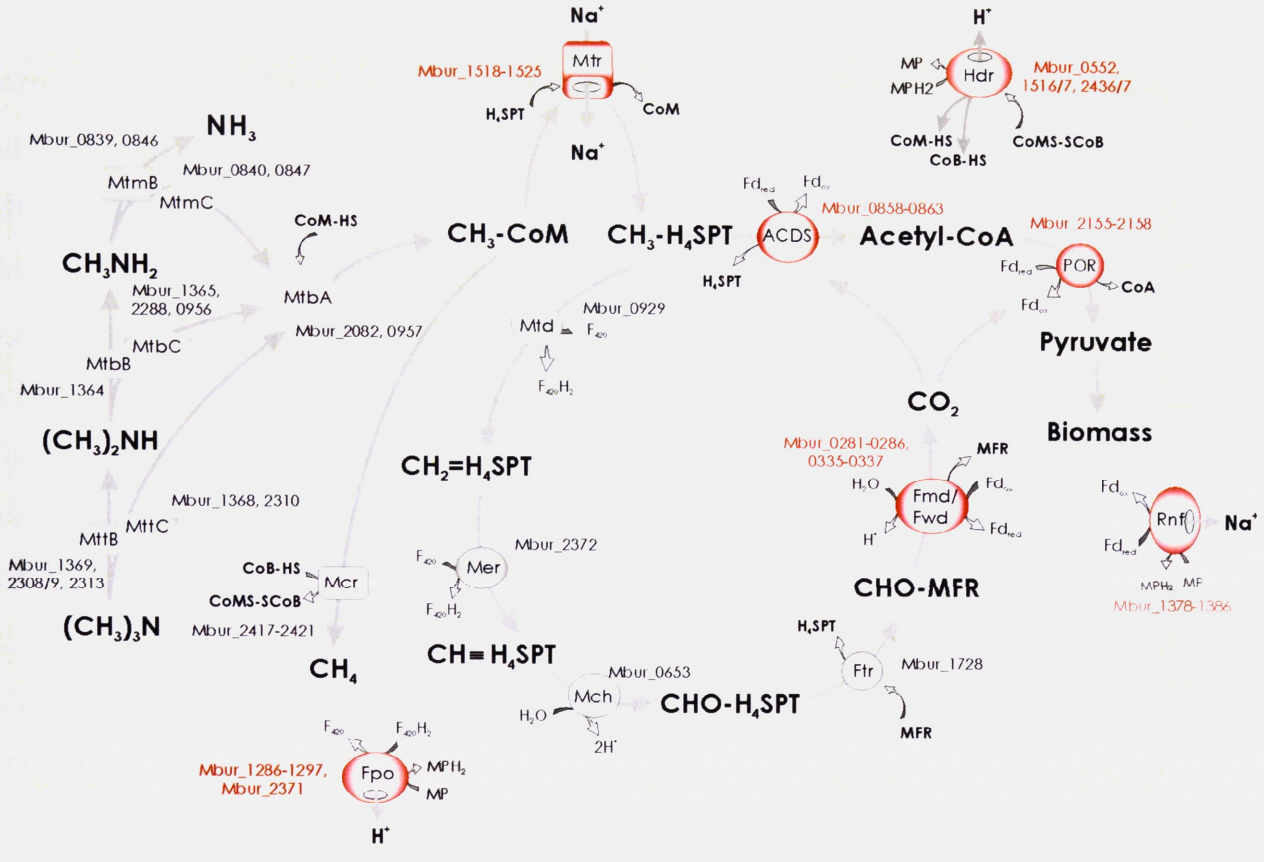


Figure 5. 8: Methanogenesis and biomass production in *M. burtonii*. Energy generation involves a proportionate dismutation of methylated substrates, by which the oxidation of one methyl group to CO_2 liberates electrons that can be used to reduce up to three methyl groups to methane. Electrons are transferred via the cofactors F_{420} and ferredoxin (Fd) to the membrane-bound $F_{420}H_2$ dehydrogenase (Fpo) and Rnf electron transfer complex, respectively. Electron flow between each of these complexes and the membrane-bound heterodisulfide reductase (Hdr) translocates protons across the cell membrane; the resulting gradient is the driving force behind ATP synthesis. Within the oxidative branch, tetrahydrosarcinapterin S-methyltransferase (Mtr) and formyl-methanofuran dehydrogenase (Fmd or Fwd) catalyse the two energy transduction steps, and generate methyl-tetrahydrosarcinapterin (CH_3-H_4SPT) and CO_2, respectively, as products. CO_2 can be excreted from the cell, or used for biosynthesis. Condensation of the methyl moiety of methyl-H_4SPT with CO_2 is the first step in carbon assimilation for biosynthesis, and is catalysed by acetyl-CoA decarbonylase/synthase (ACDS). The subsequent carboxylation of acetyl-CoA, the second step in carbon assimilation, is mediated by pyruvate synthase (POR) (Ferry and Kastead, 2007). The transcripts of all these key complexes are bound by *Mb*RpoE/F and are in red. Other abbreviations are as follows: CH_3OH, methanol, $(CH_3)_3N$, trimethylamine; $(CH_3)_2NH$, dimethylamine; CH_3NH_2, monomethylamine; CoM, coenzyme M; CH_3-CoM, methyl-coenzyme M; CoB, coenzyme B; MePh, oxidised methanophenazine; $MePhH_2$, reduced methanophenazine; F_{420}, coenzyme F_{420}; $F_{420}H_2$, reduced coenzyme F_{420}; H_4SPT, tetrahydrosarcinapterin; CH_3-H_4SPT, methyl-tetrahydrosarcinapterin; CH_2=H_4SPT, methylene-tetrahydrosarcinapterin; CH≡H_4SPT, methenyl-tetrahydrosarcinapterin; CHO-H_4SPT, formyl-tetrahydrosarcinapterin; MFR, methanofuran; CHO-MFR, formylmethanofuran; CH_3-CP, methyl-corrinoid protein; MtaB and MtaC, methanol methyltransferase and corrinoid protein; MttB and MttC, TMA methyltransferase and corrinoid protein; MtbB and MtbC, dimethylamine methyltransferase and corrinoid protein; MtmB and MtmC, monomethylamine methyltransferase and corrinoid protein; Mcr, methyl-CoM reductase; MtaA, methanol:CoM methylase; MtbA, methylamine:CoM methylase; Mcr, methyl-CoM reductase; Mer, methylene-H_4SPT reductase; Mtd, methylene-H_4SPT dehydrogenase; Mch, methenyl-H_4SPT cyclohydrolase; Ftr, formyl-methanofuran:H_4SPT formyltransferase.

In nucleotide metabolism, the heterodimer again was found to bind transcripts corresponding to key enzymes involved in the biosynthesis of nucleotides: Fae/hps bifunctional enzyme and ribose-5-phosphate isomerase (Mbur_1994, region 40 and Mbur_0773, region 10, respectively) that catalyse the two consecutive reactions that produce R5P from the RuMP; adenylosuccinate lyase and phosphoribosylformylglycinamidine synthase II (Mbur_1160, region 21 and Mbur_2051, region 41, respectively) that are involved in the synthesis of IMP, with Mbur_1160 being also involved in the synthesis of AMP from IMP; adenylate kinase (Mbur_0025, region 1) which converts AMP to ADP which is a crucial intermediate step in the formation of ATP; dihydroorotate dehydrogenase subunits K and D (Mbur_1224 and 1225 respectively, region 23) are essential for the formation or orotic acid, and also anaerobic ribonucleoside-triphosphate reductase and corresponding activating (radical SAM) protein (Mbur_2125 and 2126 respectively, region 40) that are responsible for the formation of deoxyribonucleotides from their corresponding ribonucleotides.

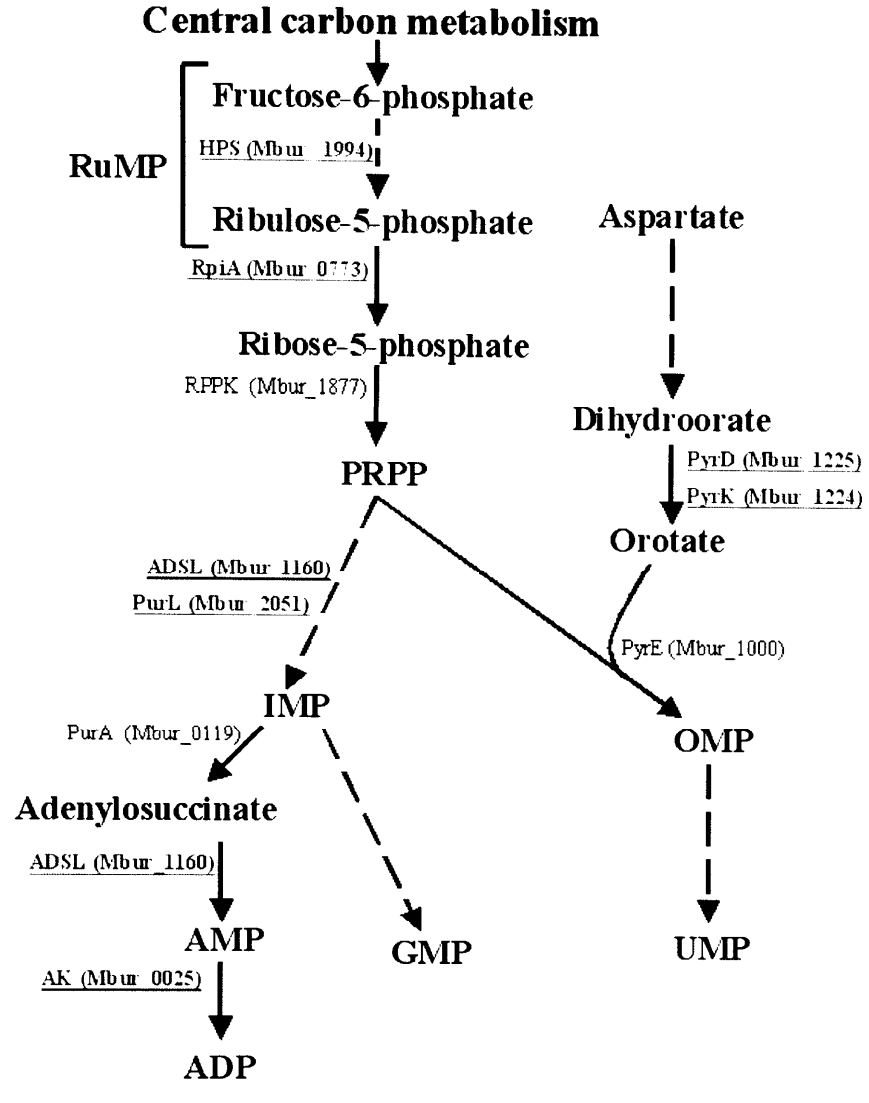


Figure 5. 9: Purine and pyrimidine biosynthesis in *M. burtonii*. Inosine monophosphate (IMP) is the first purine derivative to be synthesised in the purine *de novo* biosynthesis pathway. The starting substrate for biosynthesis of the ribosyl moiety is ribose-5-phosphate (R5P) which in archaea is a product of the ribulose monophosphate pathway (RuMP). This substrate is first activated by phosphorylation to form phosphoribosyl pyrophosphate (PRPP) and eventually converted to IMP after a series of enzymatic reactions. IMP is then converted to adenosine monophosphate (AMP) or either guanosine monophosphate (GMP) (Becerra and Cazlano, 1996; Zhang *et al.*, 2003; Kato *et al.*, 2006). The pyrimidine *de novo* biosynthesis pathway is centered on the formation of orotic acid and its subsequent reaction with PRPP to yield orotidine-5-monophosphate (OMP). OMP is decarboxylated to form uridine monophosphate (UMP), which is the substrate for the formation of UTP that leads also to CTP (Becerra and Cazlano, 1996; Zhang *et al.*, 2003; Kato *et al.*, 2006). In orange are the enzymes whose transcripts were bound by *Mb*RpoE/F: Fae/hps bifunctional enzyme (HPS) and ribose-5-phosphate isomerase (RpiA) that catalyse the two consecutive reactions that produce R5P from the RuMP; adenylosuccinate lyase (ADSL) and phosphoribosylformylglycinamidine synthase II (PurL) that are involved in the synthesis of IMP, with the first being also involved in the synthesis of AMP from IMP; adenylate kinase (AK) which converts AMP to ADP which is a crucial intermediate step in the formation of ATP; dihydroorotate dehydrogenase subunits K and D (PyrK and PyrD, respectively) are essential for the formation or orotic acid. *Mb*RpoE/F also binds to the transcript for cytidylate kinase, which catalyses the conversion of CMP to CDP. Other abbreviations are as follows: RPPK, ribose phosphate pyrophosphokinase; PurA, adenylosuccinate synthase; PyrE, orotate phosphoribosyl tranferase.

*Mb*RpoE/F also bound to transcripts encoding the key enzymes for the synthesis of tetrapyrroles: glutamate-1-semialdehyde 2,1-aminomutase (Mbur_1227, region 23), the protein involved in the final step in the synthesis of ALA; porphobilinogen deaminase (Mbur_1226, region 23), a key enzyme in the production of uroporphyrinogen III, and uroporphyrinogen-III C-methyltransferase (Mbur_1105, region 19) that synthesises precorrin-2 from uroporphyrinogen-III. Precorrin-2 acts as the last common intermediate for the synthesis of heme, cobalamin, siroheme, and coenzyme F_{430}. Each of these cofactors has a different metal ion in the center of the prosthetic group: cobalt for cobalamin, iron for heme and sirohaem and nickel for F_{430}. *Mb*RpoE/F binds the transcripts coding for two components of a cobalt-transport complex involved in the uptake of this metal into the cell for cobalamin biosynthesis (Mbur_1699 and 1700, region 32) and for a ferrous iron transport protein B (Mbur_1872, region 34). Furthermore, with regard to cobalamin, the heterodimer binds to the mRNA coding for N^1-α-phosphoribosyltransferase (Mbur_0191, region 4), the protein responsible for the assembly of the nucleotide loop in the final steps of the pathway. Regarding heme, *Mb*RpoE/F binds the transcripts coding for a protein necessary for transport of heme for cytochrome c assembly (Mbur_0981, region 16).

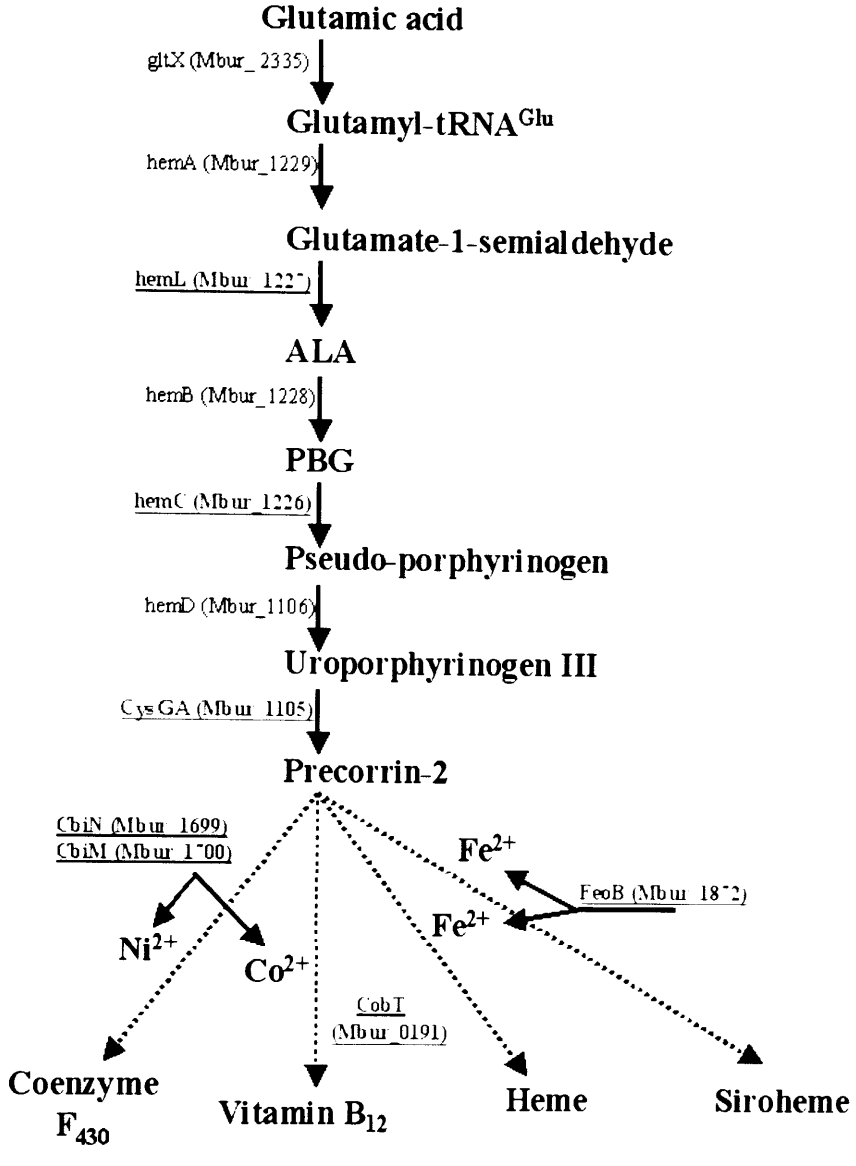


Figure 5. 10 : Synthesis of vitamin B_{12}, coenzyme F_{430}, heme and siroheme from glutamic acid in *M. burtonii*. The tetrapyrroles heme, siroheme, vitamin B_{12} (cobalamin) and F_{430} are all essential cofactors deriving from the precursor uroporphyrinogen III and precorrin-2. Each of these cofactors has a metal ion in the center of the prosthetic group: cobalt for cobalamin, iron for heme and siroheme and nickel for F_{430} (Raux *et al.*, 2000). In orange are the enzymes whose transcripts are bound by the *Mb*RpoE/F: glutamate-1-semialdehyde aminomutase (HemL), the protein involved in the final step in the synthesis of ALA; porphobilinogen deaminase (HemC), a key enzyme in the production of uroporphyrinogen III, and uroporphyrinogen-III C-methyltransferase (CysG) that synthesises precorrin-2 from uroporphyrinogen-III. *Mb*RpoE/F also binds to the transcripts coding for two components of a cobalt-transport complex involved in uptake of Co necessary for cobalamin biosynthesis (CbiM and CbiN), and of a ferrous iron transport protein B (FeoB). For cobalamin, the heterodimer binds to the transcript for N1-α-phosphoribosyltransferase (CobT), the protein responsible for the assembly of the nucleotide loop in the final steps of the pathway. Abbreviations are as follows: ALA, 5-aminolevulinic acid; PBG, porphobilinogen; GltX, glutamyl-tRNA synthetase; HemA, glutamyl-tRNA reductase; HemB, porphobilinogen synthetase; HemD, uroporphyrinogen III synthase.

The heterodimer also binds to the mRNA for Mbur_0288 and 0289 (region 6), both important in the biosynthesis of Fe-S clusters and for Mbur_2182 (region 46), coding for HMP kinase / HMP-P kinase essential for thiamine metabolism. Other targets of *Mb*RpoE/F are the transcripts for Mbur_1813, 1814 and 1815 (region 33), genes coding for serine hydroxymethyltransferase, methylene-THF dehydrogenase/cyclohydrolase and dihydropteroate synthase respectively, all essential in the synthesis and conversion of tetrahydrofolate.

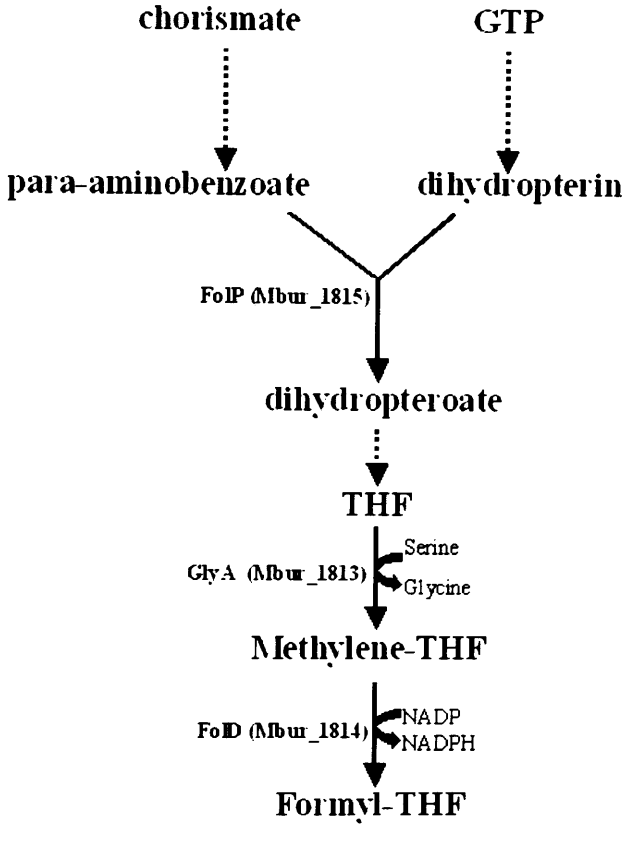


Figure 5. 11: THF synthesis and interconversions. The THF biosynthesis pathway is centered on the condensation of dihydropterin (synthesised from GTP) and para- benzoate (*p*-ABA, synthesised from chorismate) to form dihydropteroate. This reaction is catalysed by the enzyme dihydropteroate synthase (folP). Dihydropteroate is then converted into THF and eventually to formyl-THF via a series of reactions and the enzymes responsible for the last steps of this interconversion are serine hydroxymethyltransferase (GlyA), methylene-THF dehydrogenase/cyclohydrolase (FolD) and formyl-THF synthetase. The transcripts for FolP, FolD and GlyA are bound by *Mb*RpoE/F and are present in the figure.

It is extremely interesting that all these cofactors are involved in methanogenesis and nucleotide metabolism. Cobalamin-dependent methyltransferases catalyse the demethylation of methylated substrates, and transfer of methyl groups to coenzyme M (CoM) (Banerjee and Ragsdale, 2003). Thiamine, via its active form thiamine pyrophosphate, is a cofactor of pyruvate synthase (POR). F_{430} is the prosthetic group of the methylreductase that catalyses the reductive demethylation of methyl-CoM to methane (Ferry and Kastead, 2007). The Rnf complex, inferred to accept electrons from reduced ferredoxin generated in the last step of the oxidation of methyl groups, contains heme (within cytochrome c) (Ferry and Kastead, 2007). Fe-S cluster proteins mediate electron transfers within methanogenesis and downstream metabolic processes (e.g., components of F_{420} dehydrogenase, CoB-CoM heterodisulfide reductase, formyl-methanofuran dehydrogenase, ACDS, Rnf). Also, formyltetrahydrofolate is an essential cofactor for purine biosynthesis, as it required for activity by the enzyme phosphoribosylglycinamide formyltransferase (Mbur_1812) that catalyses the third step in *de novo* purine biosynthesis, the transfer of a formyl group to 5'-phosphoribosylglycinamide (Guillon *et al.*, 1992).

*Mb*RpoE/F binds to the transcripts coding for the prolyl, aspartyl and threonyl-tRNA synthetases (Mbur_0190, region 4; Mbur_0772, region 10; Mbur_1324, region 26 respectively) and also to the one coding for the subunits A and B of a glutamyl-tRNA(Gln) amidotransferase (Mbur_1654 and 1655, region 31) which synthesises Gln-tRNA(Gln) in organisms which lack glutaminyl-tRNA synthetase through the transamidation of mis-acylated Glu-tRNA(Gln). Furthermore it binds to the transcript of Mbur_0028 (region 1) coding for a pseudouridine synthase which catalyses the isomerisation of uridine to pseudouridine (ψ) the most abundant modified nucleoside found in RNA, essential for the synthesis of tRNA, and also to the transcript of Mbur_0961 (region 15), which codes for a tRNA (guanine-N1-)-methyltransferase involved in the maturation of tRNAs.

*Mb*RpoE/F binds to the transcripts corresponding to four genes related to the Sec pathway, which is the most commonly used protein translocation pathway for passage of secretory proteins to their extracellular destinations in all domains of life (Bolhuis, 2004; Pohlschröder *et al.*, 2004). These genes code for the SecY, SecF and SecD subunits

(Mbur_0024, region 1; Mbur_1986 and 1987 region 39, respectively) and a signal recognition docking protein FtsY (Mbur_0110, region 2). The heterodimer also binds the transcripts of a gene coding for a protein involved in the export of lysine (Mbur_0962, region 15).

*Mb*RpoE/F clearly is selective for a defined subset of transcripts *in vitro*. The 118 genes bound by the heterodimer do not represent a random assortment of transcripts without a common function: they belong to specific classes of mRNA and very often they correspond to the proteins at key regulatory sites for each pathway. This suggests that *Mb*RpoE/F is regulating these genes via a specific binding capacity for certain mRNA molecules. This is strongly confirmed by the fact that the mRNA coding for the F subunit is the one that is most strongly bound.

5.4 Conclusion

The 118 genes bound by *Mb*RpoE/F were easily divided in the following categories according to their function (Table 5.11): methanogenesis (44 genes); cofactor biosynthesis (14 genes); nucleotide metabolism (10 genes); transcription (8 genes); translation (7 genes); import/export (7 genes); hypothetical (12 genes).

The fact that these categories are statistically over-represented, that for each group of genes are present the key regulatory genes within the same category and that among the categories it is present a significant number of subunits of the RNA polymerase machinery, E and F included, reinforced the idea that *Mb*RpoE/F may play a regulatory role for these genes and that this function is connected to the preferential binding of the heterodimer for these transcripts.

Table 5. 11: List of the 118 genes whose transcripts were bound by *Mb*RpoE/F divided in categories according to their function.

Name	Product name	
Methanogenesis, 44 genes		
Mbur_0280	FdhD	
Mbur_0281 to 0285	Formylmethanofuran (Mo) dehydrogenase subunits B, D, C, A and F	
Mbur_0605	Pyruvate synthase γ / β -like protein	
Mbur_0858 to 0863	ACDS subunits α, ε, β, Ni, δ and γ 2	
Mbur_0994	Pyrophosphatase (H(+)-PPase) (HppA)	
Mbur_1159	Formylmethanofuran (W) dehydrogenase subunit E like protein	
Mbur_1286 to 1293	F420H2 dehydrogenase subunits O, N, M, L, K, I and H	
Mbur_1378	UPF0132 family protein, part of Rnf complex	
Mbur_1379 to 1384	Rnf electron transport complex proteins B, A, E, G, D, and C	
Mbur_1518 to 1525	H4SPT-methyltransferase subunits H, G, F A, B, C, D and E	
Mbur_2155 to 2158	Pyruvate synthase subunits γ, δ, β and α	
Mbur_2436 and 2437	CoB-CoM heterodisulfide reductase, subunits E and D	
Cofactor biosynthesis, 14 genes		
Mbur_0191	N^1-α-phosphoribosyltransferase	Tetrapyrroles siroheme cobalamin, heme, and cofactor F_{430} biosynthesis
Mbur_0981	ABC transporter, iron permease component	
Mbur_1105	Uroporphyrinogen-III C-methyltransferase (CysGA)	
Mbur_1226	Porphobilinogen deaminase (HemC)	
Mbur_1227	Glutamate-1-semialdehyde 2,1-aminomutase (HemL)	
Mbur_1699	ABC transporter cobalt-binding protein (CbiN)	
Mbur_1700	Cobalamin biosynthesis protein (CbiM)	
Mbur_1872	Ferrous iron transport protein B	
Mbur_0288	Fe-S cluster assembly protein	Fe-S clusters biosynthesis
Mbur_0289	Cysteine desulfurase	
Mbur_1813	Serine hydroxymethyltransferase	Tetrahydrofolate biosynthesis
Mbur_1814	Methylene-THF dehydrogenase/cyclohydrolase (FolD)	
Mbur_1815	Dihydropteroate synthase (FolP)	
Mbur_2182	HMP kinase / HMP-P kinase (ThiD)	Thiamine biosynthesis
Nucleotide metabolism, 10 genes		
Mbur_0025	Adenylate kinase	
Mbur_0027	Cytidylate kinase	
Mbur_0773	D-Ribose-5-phosphate isomerase	
Mbur_1160	Adenylosuccinate lyase (PurB)	
Mbur_1224 and 1225	Dihydroorotate dehydrogenase electron subunits K and D	

Mbur_1994	FAE/HPS bifunctional enzyme	
Mbur_2051	Phosphoribosylformylglycinamidine synthase II (PurL)	
Mbur_2125	Anaerobic ribonucleoside-triphosphate reductase DNA	
Mbur_2126	Radical SAM protein	
Transcription, 8 genes		
Mbur_0904	DNA-directed RNA polymerase subunit F	
Mbur_0963	Transcription initiation factor IIB (TFIIB)	
Mbur_1176 to 1179	DNA-directed RNA polymerase subunits A'', A', B' and B''	
Mbur_1989	Elongator complex protein 3 (Elp3)	
Mbur_2280	DNA-directed RNA polymerase subunit E'	
translation, 7 genes		
Mbur_0028	tRNA pseudouridine synthase B (TruB)	
Mbur_0190	Prolyl-tRNA synthetase	
Mbur_0772	Aspartyl-tRNA synthetase	
Mbur_0961	tRNA (guanine-N(1)-)-methyltransferase	
Mbur_1324	Threonyl-tRNA synthetase	
Mbur_1654 and 1655	Asn /Gln -tRNA amidotransferase subunit B and A	
Import/export, 7 genes		
Mbur_0024	Preprotein translocase SecY subunit	
Mbur_0110	Signal recognition docking protein FtsY	
Mbur_0425	Sodium/solute symporter	
Mbur_0962	Lysine export domain protein	
Mbur_1952	Sodium/solute symporter	
Mbur_1986	Preprotein translocase SecF subunit	
Mbur_1987	Protein-export membrane protein (SecD)	
Central metabolism, 8 genes		
Mbur_0316	Aconitate hydratase 2 (aconitase B) (AcnB)	Tricarboxylic acid and pathway, glycolysis gluconeogenesis
Mbur_1345	Phosphoglycerate kinase (Pgk)	
Mbur_1995	Triosephosphate isomerase (TpiA)	
Mbur_1354	Fructose 1-6-bisphosphatase (GlpX)	
Mbur_1347	Acetyl-CoA synthetase (ADP-forming), α and β subunits	
Mbur_1087	Homoserine dehydrogenase (Hom)	Amino acid metabolism
Mbur_1883	Homoaconitase large subunit (HacA)	
Mbur_2087	L-threonine O-3-phosphate decarboxylase (CobD)	
Other, 8 genes		
Mbur_1906	2-Hydroxyglutaryl-CoA dehydratase component A	
Mbur_0939	FAD dependent oxidoreductase	
Mbur_1088	DNA ligase 1, ATP-dependent (Dnl1)	
Mbur_1353	Radical SAM family protein	
Mbur_1932	3-Hydroxy-3-methylglutaryl-CoA synthase	
Mbur_1933	Acetoacetyl-CoA thiolase	
Mbur_1935	Adenylyl cyclase-family protein (ATP to 3',5'-cyclic AMP)	
Mbur_1988	Replication factor C small subunit	

Hypothetical, 12 genes	
Mbur_0026	Integral membrane protein DUF106
Mbur_0129	Hypothetical protein
Mbur_0850	Hypothetical protein
Mbur_1108	Hypothetical protein
Mbur_1252	Hypothetical protein
Mbur_1344	Hypothetical protein
Mbur_1346	Protein with DRTGG domain
Mbur_1907	Hypothetical protein
Mbur_1908	Protein of unknown function UCP019164
Mbur_1934	DUF35-domain protein
Mbur_2069	Protein of unknown function UPF0142
Mbur_2070	Hypothetical protein

Final Conclusions

The archaeal and the eucaryal transcription apparatus share a number of features that clearly distinguishes it from the less complex system in bacteria (Langer and Zillig, 1995; Thomm, 2007).

Although the structure and function of the RNA polymerase complexes have been elucidated in eucarya and archaea, the role of the dissociable RpoE/F (archaeal homologue of the eucaryotic Rpb4/7) heterodimer is still unclear. Overwhelming data show that in both systems it is involved in the response of the cell to non-optimal growth conditions (Choder, 2004).

According to proteomic and transcriptomic data on *M. burtonii*, *Mb*RpoE/F is involved in the cellular response to the cold, with RpoE being more abundant in cells grown at low (4 °C) compared to high (23 °C) temperature (Goodchild *et al.*, 2004; Campanaro *et al.*, to be submitted). This is in agreement with what has been reported about this heterodimer, where at 4 °C *M. burtonii* grows considerably more slowly than at 23 °C (Allen *et al.*, 2009; Williams *et al.*, 2010). RpoE was found to be involved in the cellular response to the "cold" also in the hyperthermophile *Pyrococcus furiosus*, where it was found to be essential for transcription initiation at relatively low, rather than optimal temperature *in vitro* (Naji *et al.*, 2007).

To understand the role of *Mb*RpoE/F in cold adaptation the genes coding for the two subunits were cloned in the pDuet vector, co-expressed and co-purified. The product consisted exclusively of the heterodimer as assessed by size exclusion chromatography. To facilitate comparative stability and structural analyses for RpoE/F, the genes from *M. jannaschii* were overexpressed in *E. coli* and recombinant proteins purified as described in Todone *et al.* (2001). Biophysical characterisation using circular dichroism (CD), differential scanning calorimetry (DSC) and dynamic fluorescence quenching (DFQ) demonstrated that both the heterodimers are in a native and folded form and that *Mb*RpoE/F

is far less stable and more flexible than *Mj*RpoE/F, which is a common feature for thermal adapted homologues (Siddiqui and Cavicchioli, 2006).

Crystallising conditions were also found for *Mb*RpoE/F and a complete in-house 3.2 Å data set was collected, although it was impossible to solve the structure. A homology model was therefore generated to be compared with the *Mj*RpoE/F crystal structure. This structural comparison resulted in the identification of a number of features that, according to literature, could be the determinants for the increased flexibility and lower thermostability of the psychrophilic homologue. The most interesting of these features is the significantly higher number of non-covalent interactions concentrated in the inter-subunit regions of *Mj*RpoE/F compared to its psychrophilic homologue, a characteristic that has been reported in several thermophilic proteins (Karlström *et al.*, 2006; Vetriani *et al.*, 1998, Rice *et al.*, 2006).

The functional comparison between the two heterodimers, performed using EMSA on a range of ssDNA templates, led ultimately to the most important findings of this study. This data indicated that *Mb*RpoE/F bound specifically to oligo-dC and oligo-dT while *Mj*RpoE/F bound preferentially to oligo-dG and oligo-dT. Although K_b values calculated for the heterodimer:DNA interactions were very high, they clearly showed sequence preference. This is the first report, in either archaea or eucarya, that the E/F heterodimer can bind nucleic acids in specific ways via a preferential affinity for certain sequences: while the ability of RpoE/F (or Rpb4/7) to bind RNA has been experimentally established in many works (Todone *et al.*, 2001; Orlicky *et al.*, 2001; Meka *et al.* 2003, 2005), none of these papers addresses whether the heterodimer has the ability to recognise specific RNA targets.

This point is of capital importance, because in yeast, Rpb4 and Rpb7 have been shown to be involved in the regulation and, more specifically, in the decay of specific classes of mRNA molecules (Lotan *et al.*, 2005 and Pillai *et al.*, 2003; Goler-Baron *et al.*, 2008), but this feature has never been directly associated with sequence-selective RNA binding by the heterodimer. Instead, most studies speculate on the possible role that protein-partners afford

to specificity (*e.g.* Goler-Baron *et al.*, 2008). These assumptions originated from the similarity of specific regions of RpoE and Rbp7 with an OB fold motif that is usually involved in non-specific nucleic acid binding (Todone *et al.*, 2001).

Based on the preliminary data from EMSA, it was hypothesised that the ability to regulate the mRNA level of specific targets could reside in the heterodimer itself, via the differential binding affinity mentioned above. In order to verify this hypothesis, it was decided not to investigate the binding properties of small ssDNA or RNA oligos further, but to test full-length RNA transcripts extracted from growing cultures by affinity purification of the target RNA molecules followed by hybridisation on microarrays designed specifically for *M. burtonii*.The rationale behind this decision was that such an approach would be more biologically relevant compared to EMSA, as full-length transcripts (rather than RNA oligos) would maximise any RNA-protein interaction due to RNA secondary structures. Furthermore, such a strategy allowed the screening of thousands of different sequences at the same time, instead of the very few that can be analysed in each EMSA experiment.

His-tagged *Mb*RpoE/F was incubated with whole cell RNA and complexes were purified using a Ni-NTA column. The bound RNA was eluted, labeled and hybridised to a high-density *M. burtonii* microarray. By mixing only the affinity-purified heterodimer with purified total *M. burtonii* RNA *in vitro*, the binding experiment excluded partner protein(s) mediated interactions, thereby providing a clear assessment of the binding capacity of just *Mb*RpoE/F to cellular mRNA.

The results showed that *Mb*RpoE/F bound to the transcripts corresponding to only 118 of the 2494 *M. burtonii* genes.

The first important consideration about this data is that the transcripts that appeared to be bound by *Mb*RpoE/F are not the most abundant transcripts in the cell, but range from low abundance to very high abundance transcripts. To further confirm this, the transcripts coding for the ribosomal proteins, which are among the most abundant in *M. burtonii* at any condition tested so far, were not bound by the heterodimer at all, but the strongest binding was for the *Mb*RpoF transcript itself, which is expressed at moderate levels.

The vast majority of the 118 genes could easily be divided into distinct categories: methanogenesis, nucleotide metabolism, cofactors biosynthesis, transcription, translation, import/export and others; most importantly, for each category the genes identified appear to code for key regulatory enzymes in the same category.

In methanogenesis, the 44 genes identified code exclusively for the central protein complexes of the pathway: Fpo, Rnf, Hdr, Mtr, Fmd, Fwd, ACDS and POR.

In nucleotide metabolism, the heterodimer again bound the transcripts corresponding to key enzymes involved in the biosynthesis of purines (Fae/hps bifunctional enzyme, ribose-5-phosphate isomerase, adenylosuccinate lyase, adenylate kinase), pyrimidines (dihydroorotate dehydrogenase) and the formation of deoxyribonucleotides from their corresponding ribonucleotides (ribonucleoside-triphosphate reductase). The same is also true for the genes involved in cofactor biosynthesis: the heterodimer bound the transcripts for the key enzymes in tetrapyrrol siroheme, cobalamin and F_{430} production (glutamate-1-semialdehyde 2,1-aminomutase, porphobilinogen deaminase, uroporphyrinogen-III C-methyltransferase and N^1-α-phosphoribosyltransferase) and also for transporters involved in the intake of the metal ions at the center of their prosthetic groups.

*Mb*RpoE/F binds to the transcripts corresponding to key protein in the process of transcription (RNA polymerase subunits A", A', B', B", E, F and the general transcription factor TFB), translation (four aminoacyl-tRNA synthetases among others, and protein export (four central subunits of the Sec pathway: SecY, SecF, SecD and FtsY).

*Mb*RpoE/F clearly is selective for a defined subset of transcripts *in vitro*. These transcripts are not the most abundant in the total extracted RNA, nor do they represent a random assortment of transcripts without a common function: they belong to specific classes of mRNA and very often they correspond to the proteins at key regulatory sites for each pathway. This suggests that *Mb*RpoE/F is regulating these genes via a specific binding capacity for certain mRNA molecules. This is strongly confirmed by the fact that the mRNA coding for the F subunit is the most strongly bound.

As mentioned previously, in yeast, by shuttling between the nucleus and the cytoplasm, *S. cerevisiae* Rpb7/4 regulates mRNA synthesis and decay and thereby controls the cellular abundance of specific mRNA species (Goler-Baron *et al.*, 2008). mRNA species that have been linked to Rpb7- and Rpb4-mediated decay include genes encoding ribosomal proteins, translation initiation factors, aminoacyl tRNA synthetases, and genes from the ribi regulon (required for ribosome biogenesis in yeast) (Pillai *et al.*, 2003; Lotan *et al.*, 2005). Despite the obvious metabolic differences between *S. cerevisiae* and *M. burtonii* (i.e. methanogenesis), it is apparent that some of the specific classes of yeast mRNA are the same as those bound by *Mb*RpoE/F *in vitro*.

This similarity strongly reinforced the microarray data and therefore the idea that the heterodimer is capable of binding RNA in a specific manner. If in archaea the role of the heterodimer is analogous to eucarya, then it would be plausible to assume that its specific role is to directly select a distinct subclass of transcripts, probably immediately after they are synthesised, and direct them to degradation; this in conditions where the cell decreases cell growth, such as at 4 °C for *M. burtonii*. This could explain the reason why the transcript and the protein product of the RpoE gene are more abundant at 4 °C *vs* 23 °C while a significant portion of the genes whose transcripts are bound by it (methanogenesis genes in particular) are generally more abundant at 23 °C *versus* 4 °C (Campanaro *et al.*, to be submitted): if the heterodimer binds those mRNAs and eventually degrades them, their abundance must be inversely proportional. If confirmed, this hypothesis could have a broader significance and explain the role of this heterodimer in respect to different growth conditions not only in archaea, but also in eucarya.

Several eucaryotic and bacterial examples are documented where interaction between the transcribing RNAP and the emerging transcript regulate elongation decisions, and in many cases this binding imposes significant regulation on downstream expression. In bacteria, for example, the domain that covers the RNA exit channel of RNAP is able to contact nascent RNA stem-loop structures (hairpins) that inhibit transcription, and this inhibition can be further enhanced by the multifunctional transcription factor, NusA (Toulokhonov et al., 2001; Shibata *et al.*, 2007). In eucarya, transcription termination can be mediated by the

eukaryotic Xrn2/Rat1 exonucleases, but it has also been proposed that in some cases the transcription of the poly(A) site produces a conformational change in the Pol II elongation complex that results in the termination of the transcription and consequent release of the RNA molecule (Rosonina *et al.*, 2006; Kaneko *et al.*, 2007) .

Regarding mRNA degradation, mechanisms in which specific RNA molecules are shuttled to the P-bodies and thus degraded have been reported for eucarya (Parker and Sheth, 2007; Kulkarni *et al.*, 2010). For example, mRNAs that contain AREs (AU-rich elements) in their 3′-UTRs (untranslated regions) are delivered to the general decay machinery by the TTP family of tandem zinc finger proteins. These proteins bind to the 3′-UTR of the target mRNAs and promote de-adenylation, followed by degradation in the P-bodies (Parker and Sheth, 2007; Kulkarni *et al.*, 2010). Also, aberrant mRNAs that contain a premature termination codon seem to be delivered to the P-bodies by the nonsense-mediated mRNA decay (NMD) factors Upf 1, Upf2 and Upf3 (Kulkarni *et al.*, 2010). Similar shuttling mechanisms have not been reported so far in bacteria and archaea.

The mechanism proposed for the archaeal E/F, if proven to be effective *in vivo*, would be a novel way of regulation compared to the mechanisms described above. The interaction between the polymerase and the transcript would represent the core of a unique cross-talk between the RNA polymerase and the machineries responsible for mRNA decay. It would not just represent a way to regulate the activity of the RNA polymerase or a way to guide specific transcripts to the degradation machinery. This cross-talk would be a central and basic system to maintain appropriate mRNAs levels in the living cell.

It is impossible to speculate whether such a mechanism of regulation could also be present in bacteria. The multi-protein complex that is involved in the degradation of mRNA present in most bacteria is called the degradosome and it shows substantial differences with P-bodies or exosomes (Belasco, 2010); also the RNA polymerase machinery is completely different from the eucaryal and the archaeal one, and presents no homologue to the Rbp4/7 - RpoE/F heterodimers.

The *Mb*RpoE/F mechanism would therefore represent another confirmation that the archaeal transcription system is more closely related to eucarya than to bacteria (Thomm, 2007).

Future work

One idea for a final and definitive proof of the preferential binding of *Mb*RpoE/F for certain mRNA targets, and at the same time validate the entire dataset coming from the microarray work, would be to clone a pair of regions of the *M. burtonii* genome into a T7 vector where RNA can be produced. One region would correspond to a transcript that is strongly bound by the heterodimer and the other to a transcript which does not appear to be bound. In this way it could be possible to produce reasonable amounts of the two transcripts and confirm the preferential binding for one and not the other, possibly by isothermal titration calorimetry (ITC) (Gilbert and Batey, 2009). If such work could be successfully completed, it could be also possible to identify the RNA binding motif via classical methods such as RNA footprinting.

In order to better understand the function of the heterodimer, it would be interesting to use Q-PCR under a range of different conditions (logarithmic *vs* stationary phase, oxidative and non-oxidative stress etc.) and determine the level of abundance of the transcript for the heterodimer in these conditions. This would help to understand how *Mb*RpoE/F is related to stressful conditions. An *in vitro* transcription system, similar to the one described by Darcy *et al.* (1999) for *Methanobacterium thermoautotrophicum*, could also be designed and developed for *M. burtonii* in order to test the transcription of different mRNAs, in the presence and absence of the heterodimer and using different conditions. This could provide a more complete understanding of the preferential binding of the heterodimer for different mRNAs in relation to the other subunits of the RNA polymerase machinery. However, the importance of additional *in vitro* data is questionable, as the limitation of the work described in this thesis is that it is not supported by any *in vivo* experiments. The first priority to validate these experiments would be to determine whether the preferential

binding of *Mb*RpoE/F for a subset of RNA targets also exists *in vivo*, and this could be achieved by RIP-Chip (Mardis, 2007; Townley-Tilson *et al.*, 2006).

Unfortunately, a genetic system is not yet available for *M. burtonii*, and therefore it is not possible to knock-out the RpoE/F genes and determine the effect on the mRNA profile at different temperatures. But at that point it would be much more interesting to apply the same binding strategy described in Chapter 4 to *S. cerevisiae* and determine if the findings on *M. burtonii* also apply to the eucaryal transcription system.

Bibliography

Aiyar, A. 2000. The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol Biol,* 132, 221-41.

Allen, M. A., Lauro, F. M., Williams, T. J., Burg, D., Siddiqui, K. S., De Francisci, D., Chong, K. W., Pilak, O., Chew, H. H., De Maere, M. Z., Ting, L., Katrib, M., Ng, C., Sowers, K. R., Galperin, M. Y., Anderson, I. J., Ivanova, N., Dalin, E., Martinez, M., Lapidus, A., Hauser, L., Land, M., Thomas, T. & Cavicchioli, R. 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J,* 3, 1012-35.

Armache, K. J., Mitterweger, S., Meinhart, A. & Cramer, P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem,* 280, 7131-4.

Ashutosh, T., Prasanna, K. & Bhata R. 2000. An efficient and cost-effective procedure for preparing sample for differential scanning calorimetry experiments *Analytical Chemistry* 284, 406-408.

Bakermans, C. & Nealson, K. H. 2004. Relationship of critical temperature to macromolecular synthesis and growth yield in *Psychrobacter* **cryopegella**. *J Bacteriol,* 186, 2340-5.

Baltscheffsky, M., Schultz, A. & Baltscheffsky, H. 1999. H+ -PPases: a tightly membrane-bound family. *FEBS Lett,* 457, 527-33.

Banerjee, R. & Ragsdale, S. W. 2003. The many faces of vitamin B_{12}: catalysis by cobalamin-dependent enzymes. *Annu Rev Biochem,* 72, 209-47.

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A,* 93, 9188-93.

Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A,* 91, 1609-13.

Baumer, S., Ide, T., Jacobi, C., Johann, A., Gottschalk, G. & Deppenmeier, U. 2000. The $F_{420}H_2$ dehydrogenase from *Methanosarcina mazei* is a Redox-driven proton pump closely related to NADH dehydrogenases. *J Biol Chem,* 275, 17968-73.

Baumer, S., Lentes, S., Gottschalk, G. & Deppenmeier, U. 2002. Identification and analysis of proton-translocating pyrophosphatases in the methanogenic archaeon *Methansarcina mazei. Archaea,* 1, 1-7.

Becerra, A. & Lazcano, A. 1998. The role of gene duplication in the evolution of purine nucleotide salvage pathways. *Origins of Life and Evolution of the Biosphere* 28: 539–553, 1998.

Beinoraviciute-Kellner, R., Lipps, G. & Krauss, G. 2005. In vitro selection of DNA binding sites for ABF1 protein from *Saccharomyces cerevisiae. FEBS Lett,* 579, 4535-40.

Belasco, J.G. 2010. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat Rev Mol Cell Biol* 11, 467-78.

Bell, S. D. & Jackson, S. P. 1998a. Transcription and translation in Archaea: a mosaic of eucaryal and bacterial features. *Trends Microbiol,* 6, 222-8.

Bell, S. D. & Jackson, S. P. 1998b. Transcription in Archaea. *Cold Spring Harb Symp Quant Biol,* 63, 41-51.

Bell, S. D. & Jackson, S. P. 2001. Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol,* 4, 208-13.

Bell, S. D., Magill, C. P. & Jackson, S. P. 2001. Basal and regulated transcription in Archaea. *Biochem Soc Trans,* 29, 392-5.

Bolhuis, A. 2004. The archaeal Sec-dependent protein translocation pathway. *Philos Trans R Soc Lond B Biol Sci,* 359, 919-27.

Briozzo, P., Golinelli-Pimpaneau, B., Gilles, A. M., Gaucher, J. F., Burlacu-Miron, S., Sakamoto, H., Janin, J. & Barzu, O. 1998. Structures of *Escherichia coli* CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity. *Structure,* 6, 1517-27.

Brown, J. R. & Doolittle, W. F. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev,* 61, 456-502.

Buchenau, B., Kahnt, J., Heinemann, I. U., Jahn, D. & Thauer, R. K. 2006. Heme biosynthesis in *Methanosarcina barkeri* via a pathway involving two methylation reactions. *J Bacteriol,* 188, 8666-8.

Burley, S. K. & Petsko, G. A. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science,* 229, 23-8.

Byrne-Steele, M. L., Hughes, R. C. & Ng, J. D. 2009. Recombinant production, cristallisation and preliminary X-ray analysis of PCNA from the psychrophilic archaeon *Methanococcoides burtonii* DSM 6242. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 65, 1131-5.

Cartron, M. L., Maddocks, S., Gillingham, P., Craven, C. J. & Andrews, S. C. 2006. Feo--transport of ferrous iron into bacteria. *Biometals,* 19, 143-57.

Cavicchioli, R. 2006. Cold-adapted archaea. *Nat Rev Microbiol,* 4, 331-43.

Cavicchioli, R., Siddiqui, K. S., Andrews, D. & Sowers, K. R. 2002. Low-temperature extremophiles and their applications. *Current Opinion in Biotechnology,* 13, 253-261.

Chaban, B., Ng, S. Y. & Jarrell, K. F. 2006. Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol,* 52, 73-116.

Chain, P. S., Hu, P., Malfatti, S. A., Radnedge, L., Larimer, F., Vergez, L. M., Worsham, P., Chu, M. C. & Andersen, G. L. 2006. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol,* 188, 4453-63.

Choder, M. 2004. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends in Biochemical Sciences,* 29, 674-81.

Christensen, K. E. & Mackenzie, R. E. 2006. Mitochondrial one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *Bioessays,* 28, 595-605.

Collins, T., D'Amico, S., Georlette, D., Marx, J., Huston, A. L. & Feller, G. 2006. A nondetergent sulfobetaine prevents protein aggregation in microcalorimetric studies. *Analytical Biochemistry,* 352, 299-301.

D'amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M. A., Feller, G. & Gerday, C. 2002. Molecular basis of cold adaptation. *Philos Trans R Soc Lond B Biol Sci,* 357, 917-25.

Dalluge, J. J., Hamamoto, T., Horikoshi, K., Morita, R. Y., Stetter, K. O. & Mccloskey, J. A. 1997. Posttranscriptional modification of tRNA in psychrophilic bacteria. *J Bacteriol,* 179, 1918-23.

Darcy, T.J., Hausner, W., Awery, D.E., Edwards, A.M., Thomm, M. & Reeve, J.N. 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro. Journal of Bacteriology* 181, 4424-9.

De Poorter, L. M., Geerts, W. G., Theuvenet, A. P. & Keltjens, J. T. 2003. Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in *Methanothermobacter thermautotrophicus. Eur J Biochem,* 270, 66-75.

Delong, E. F. 2005. Microbial community genomics in the ocean. *Nat Rev Microbiol,* 3, 459-69.

Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem,* 266, 71-5.

Eggeling, L. & Sahm, H. 2003. New ubiquitous translocators: amino acid export by *Corynebacterium glutamicum* and *Escherichia coli. Arch Microbiol,* 180, 155-60.

Eichler, J. & Moll, R. 2001. The signal recognition particle of Archaea. *Trends Microbiol,* 9, 130-6.

Eng, E. T., Jalilian, A. R., Spasov, K. A. & Unger, V. M. 2008. Characterisation of a novel prokaryotic GDP dissociation inhibitor domain from the G protein coupled membrane protein FeoB. *J Mol Biol,* 375, 1086-97.

Escalante-Semerena, J. C., Woodson, J. D. & Buan, N. R. 2007. Conversion of cobinamide into coenzyme B_{12}. *Tetrapyrroles birth, life and death* M. J. Warren and A. J. Smith. New York, New York, Landes Bioscience: 300-316.

Farago, M., Nahari, T., Hammel, C., Cole, C. N. & Choder, M. 2003. Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Molecular Biology of the Cell,* 14, 2744-55.

Feller, G. & Gerday, C. 2003. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol,* 1, 200-8.

Ferry, J. G. & Kastead, K. A. 2007. Methanogenesis. *Archaea: Molecular and cellular biology*. R. Cavicchioli. Washington, D. C., ASM press: 288-314.

Franzmann, P. D. 1992. A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst Appl Microbiol* 15, 573-81.

Gallivan, J. P. & Dougherty, D. A. 1999. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A*, 96, 9459-64.

Gao, B. & Gupta, R. S. 2007. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics*, 8, 86.

Georlette, D., Blaise, V., Collins, T., D'amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G. & Gerday, C. 2004. Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol Rev*, 28, 25-42.

Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C. & Feller, G. 2003. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J Biol Chem*, 278, 37015-23.

Georlette, D., Jonsson, Z. O., Van Petegem, F., Chessa, J., Van Beeumen, J., Hubscher, U. & Gerday, C. 2000. A DNA ligase from the psychrophile *Pseudoalteromonas haloplanktis* gives insights into the adaptation of proteins to low temperatures. *Eur J Biochem*, 267, 3502-12.

Gianese, G., Argos, P. & Pascarella, S. 2001. Structural adaptation of enzymes to low temperatures. *Protein Eng*, 14, 141-8.

Giaquinto, L., Curmi, P. M. G., Siddiqui, K. S., Poljak, A., Delong, E., Dassarma, S. & Cavicchioli, R. 2007. Structure and function of cold shock proteins in archaea. *Journal of Bacteriology*, 189, 5738-5748.

Gibson, J. A., Miller, M. R., Davies, N. W., Neill, G. P., Nichols, D. S. & Volkman, J. K. 2005a. Unsaturated diether lipids in the psychrotrophic archaeon *Halorubrum lacusprofundi*. *Syst Appl Microbiol*, 28, 19-26.

Gibson, J. A. E., Miller, M. R., Davies, N. W., Neill, G. P., Nichols, D. S. & Volkman, J. K. 2005b. Unsaturated diether lipids in the psychrotrophic archaeon *Halorubrum lacusprofundi*. *Systematic and Applied Microbiology*, 28, 19-26.

Gilbert, J. A., Hill, P. J., Dodd, C. E. R. & Laybourn-Parry, J. 2004. Demonstration of antifreeze protein activity in Antarctic lake bacteria. *Microbiology-Sgm,* 150, 171-180.

Gilbert, S. D. & Batey, R. T. 2009. Monitoring RNA-ligand interactions using isothermal titration calorimetry. *Methods Mol Biol,* 540, 97-114.

Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R. & Choder, M. 2008. Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev,* 22, 2022-7.

Goodchild, A., Raftery, M., Saunders, N. F., Guilhaus, M. & Cavicchioli, R. 2004a. Biology of the cold adapted archaeon, *Methanococcoides burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. *J Proteome Res,* 3, 1164-76.

Goodchild, A., Raftery, M., Saunders, N. F., Guilhaus, M. & Cavicchioli, R. 2005. Cold adaptation of the Antarctic archaeon, *Methanococcoides burtonii* assessed by proteomics using ICAT. *J Proteome Res,* 4, 473-80.

Goodchild, A., Saunders, N. F., Ertan, H., Raftery, M., Guilhaus, M., Curmi, P. M. & Cavicchioli, R. 2004b. A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii. Mol Microbiol,* 53, 309-21.

Graham, R. M., Deery, E. & Warren, M. J. 2007. Vitamin B_{12}: biosynthesis of the corrin ring. *Tetrapyrroles birth, life and death* M. J. Warren and A. J. Smith. New York, New York, Landes Bioscience: 286-299.

Graumann, P. & Marahiel, M. A. 1996. Some like it cold: Response of microorganisms to cold shock. *Archives of Microbiology,* 166, 293-300.

Gu, Z. M., Martindale, D. W. & Lee, B. H. 1992. Isolation and complete sequence of the purL gene encoding FGAM synthase II in *Lactobacillus casei. Gene,* 119, 123-6.

Guillon, J. M., Meinnel, T., Mechulam, Y., Lazennec, C., Blanquet, S. & Fayat, G. 1992. Nucleotides of tRNA governing the specificity of *Escherichia coli* methionyl-tRNA(fMet) formyltransferase. *J Mol Biol,* 224, 359-67.

Hagervall, T. G., Tuohy, T. M., Atkins, J. F. & Bjork, G. R. 1993. Deficiency of 1-methylguanosine in tRNA from *Salmonella typhimurium* induces frameshifting by quadruplet translocation. *J Mol Biol,* 232, 756-65.

Hellman, L. M. & Fried, M. G. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc,* 2, 1849-61.

Hickey, A. J., Conway De Macario, E. & Macario, A. J. 2002. Transcription in the archaea: basal factors, regulation, and stress gene expression. *Crit Rev Biochem Mol Biol,* 37, 199-258.

Hippler, B. & Thauer, R. K. 1999. The energy conserving methyltetrahydromethanopterin:coenzyme M methyltransferase complex from methanogenic archaea: function of the subunit MtrH. *FEBS Lett,* 449, 165-8.

Hirata, A., Kanai, T., Santangelo, T. J., Tajiri, M., Manabe, K., Reeve, J. N., Imanaka, T. & Murakami, K. S. 2008a. Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol Microbiol,* 70, 623-33.

Hirata, A., Klein, B. J. & Murakami, K. S. 2008b. The X-ray crystal structure of RNA polymerase from Archaea. *Nature,* 451, 851-4.

Hoang, C. & Ferre-D'amare, A. R. 2001. Cocrystal structure of a tRNA Psi 55 pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell,* 107, 929-939.

Hochheimer, A., Linder, D., Thauer, R. K. & Hedderich, R. 1996. The molybdenum formylmethanofuran dehydrogenase operon and the tungsten formylmethanofuran dehydrogenase operon from *Methanobacterium thermoautotrophicum*. Structures and transcriptional regulation. *Eur J Biochem,* 242, 156-62.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. & Stetter, K. O. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature,* 417, 63-7.

Hwang, D. M., Dempsey, A., Tan, K. T. & Liew, C. C. 1996. A modular domain of NifU, a nitrogen fixation cluster protein, is highly conserved in evolution. *J Mol Evol,* 43, 536-40.

Ishida, T., Yu, L., Akutsu, H., Ozawa, K., Kawanishi, S., Seto, A., Inubushi, T. & Sano, S. 1998. A primitive pathway of porphyrin biosynthesis and enzymology in *Desulfovibrio vulgaris*. *Proc Natl Acad Sci U S A,* 95, 4853-8.

Jasiak, A. J., Armache, K. J., Martens, B., Jansen, R. P. & Cramer, P. 2006. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Molecular Cell,* 23, 71-81.

Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. 2005. Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem,* 74, 247-81.

Jones, P. G. & Inouye, M. 1994. The cold-shock response--a hot topic. *Mol Microbiol,* 11, 811-8.

Jung, H. 2002. The sodium/substrate symporter family: structural and functional features. *FEBS Lett,* 529, 73-7.

Jurgenson, C. T., Begley, T. P. & Ealick, S. E. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem,* 78, 569-603.

Kammler, M., Schon, C. & Hantke, K. 1993. Characterisation of the ferrous iron uptake system of *Escherichia coli. J Bacteriol,* 175, 6212-9.

Kandror, O. & Goldberg, A. L. 1997. Trigger factor is induced upon cold shock and enhances viability of *Escherichia coli* at low temperatures. *Proc Natl Acad Sci U S A,* 94, 4978-81.

Kaneko, S., Chu, C., Shatkin A.J. & Manley J.L. 2007. Human capping enzyme promotes formation of transcriptional R loops *in vitro. Proc Natl Acad Sci U S A,* 104,17620-5.

Karlstrom, M., Steen, I. H., Madern, D., Fedoy, A. E., Birkeland, N. K. & Ladenstein, R. 2006. The crystal structure of a hyperthermostable subfamily II isocitrate dehydrogenase from *Thermotoga maritima. FEBS J,* 273, 2851-68.

Karner, M. B., Delong, E. F. & Karl, D. M. 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature,* 409, 507-10.

Kath, T., Schmid, R. & Schafer, G. 1993. Identification, cloning, and expression of the gene for adenylate kinase from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius. Arch Biochem Biophys,* 307, 405-10.

Kato, N., Yurimoto, H. & Thauer, R. K. 2006. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Biosci Biotechnol Biochem,* 70, 10-21.

Kletzin, A. & Adams, M. W. 1996. Molecular and phylogenetic characterisation of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima. J Bacteriol,* 178, 248-57.

Kletzin, A. 2007. General characteristics and important model organisms. *Archaea: Molecular and cellular biology.* R. Cavicchioli. Washington, D. C., ASM press: 288-314.

Klotz, M. G., Arp, D. J., Chain, P. S., El-Sheikh, A. F., Hauser, L. J., Hommes, N. G., Larimer, F. W., Malfatti, S. A., Norton, J. M., Poret-Peterson, A. T., Vergez, L. M. & Ward, B. B. 2006. Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceani* ATCC 19707. *Appl Environ Microbiol,* 72, 6299-315.

Koga, Y. & Morii, H. 2007. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol Mol Biol Rev,* 71, 97-120.

Koga, Y., Nishihara, M., Morii, H. & Akagawa-Matsushita, M. 1993. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol Rev,* 57, 164-82.

Koster, S., Kuhlbrandt, W. & Yildiz, O. 2009. Purification, cristallisation and preliminary X-ray diffraction analysis of the FeoB G domain from *Methanococcus jannaschii. Acta Crystallogr Sect F Struct Biol Cryst Commun,* 65, 684-7.

Kranz, R. G., Richard-Fogal, C., Taylor, J. S. & Frawley, E. R. 2009. Cytochrome c biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. *Microbiol Mol Biol Rev,* 73, 510-28, Table of Contents.

Kuhn, C. D., Geiger, S. R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R. & Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell,* 131, 1260-72.

Kulkarni, M., Ozgur, S. & Stoecklin, G. 2010. On track with P-bodies. *Biochem Soc Trans* 38, 242-51.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227, 680-5.

Langer, D., Hain, J., Thuriaux, P. & Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A,* 92, 5768-72.

Leigh, J. A. 1999. Transcriptional regulation in Archaea. *Curr Opin Microbiol,* 2, 131-4.

Leiros, H. K., Willassen, N. P. & Smalas, A. O. 2000. Structural comparison of psychrophilic and mesophilic trypsins. Elucidating the molecular basis of cold-adaptation. *Eur J Biochem,* 267, 1039-49.

Li, Y., Jiang, Z., Chen, H. & Ma, W. J. 2004. A modified quantitative EMSA and its application in the study of RNA--protein interactions. *J Biochem Biophys Methods,* 60, 85-96.

Lim, J., Thomas, T. & Cavicchioli, R. 2000. Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, *Methanococcoides burtonii. J Mol Biol,* 297, 553-67.

Londei, P. 2007. Translation. *Archaea: Molecular and cellular biology.* R. Cavicchioli. Washington, D. C., ASM press: 175-208.

Lotan, R., Bar-on, V. G., Harel-Sharvit, L., Duek, L., Melamed, D. & Choder, M. 2005. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev,* 19, 3004-16.

Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G. & Choder, M. 2007. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol,* 178, 1133-43.

Macgregor, B. J., Moser, D. P., Alm, E. W., Nealson, K. H. & Stahl, D. A. 1997. Crenarchaeota in Lake Michigan sediment. *Appl Environ Microbiol,* 63, 1178-81.

Makarova, K. S. & Koonin, E. V. 2003. Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol,* 4, 115.

Mardis E.R. 2007. ChIP-seq: welcome to the new frontier. *Nat Methods* 8, 613-4.

Mathews, Ii, Krishna, S. S., Schwarzenbacher, R., Mcmullan, D., Jaroszewski, L., Miller, M. D., Abdubek, P., Agarwalla, S., Ambing, E., Axelrod, H. L., Canaves, J. M., Carlton, D., Chiu, H. J., Clayton, T., Didonato, M., Duan, L., Elsliger, M. A., Grzechnik, S. K., Hale, J., Hampton, E., Haugen, J., Jin, K. K., Klock, H. E., Koesema, E., Kovarik, J. S., Kreusch, A., Kuhn, P., Levin, I., Morse, A. T., Nigoghossian, E., Okach, L., Oommachen, S., Paulsen, J., Quijano, K., Reyes, R., Rife, C. L., Spraggon, G., Stevens, R. C., Van Den Bedem, H., White, A., Wolf, G., Xu, Q., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A. & Wilson, I. A. 2006. Crystal structure of phosphoribosylformyl-glycinamidine synthase II, PurS subunit (TM1244) from *Thermotoga maritima* at 1.90 A resolution. *Proteins,* 65, 249-54.

Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40, 658-674.

Medigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P. N., Cheung, F., Cruveiller, S., D'amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E. P., Rouy, Z., Sekowska, A., Tutino, M. L., Vallenet, D., Von Heijne, G. & Danchin, A. 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res,* 15, 1325-35.

Meka, H., Daoust, G., Arnvig, K. B., Werner, F., Brick, P. & Onesti, S. 2003. Structural and functional homology between the RNAP(I) subunits A14/A43 and the archaeal RNAP subunits E/F. *Nucleic Acids Res,* 31, 4391-400.

Meka, H., Werner, F., Cordell, S. C., Onesti, S. & Brick, P. 2005. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res,* 33, 6435-44.

Methe, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X. J., Moult, J., Madupu, R., Nelson, W. C., Dodson, R. J., Brinkac, L. M., Daugherty, S. C., Durkin, A. S., Deboy, R. T., Kolonay, J. F., Sullivan, S. A., Zhou, L. W., Davidsen, T. M., Wu, M., Huston, A. L., Lewis, M., Weaver, B., Weidman, J. F., Khouri, H., Utterback, T. R., Feldblyum, T. V. & Fraser, C. M. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 10913-10918.

Mori, H. & Ito, K. 2001. The Sec protein-translocation pathway. *Trends Microbiol,* 9, 494-500.

Morita, R. Y. 1975. Psychrophilic bacteria. *Bacteriol Rev,* 39, 144-67.

Mrabet, N. T., Van Den Broeck, A., Van Den Brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., Matthijssens, G., Jenkins, J., Chiadmi, M., Van Tilbeurgh, H. & Et Al. 1992. Arginine residues as stabilizing elements in proteins. *Biochemistry,* 31, 2239-53.

Munro, A. W., Girvan, H. M., McLean K. J., Cheesman M. R. & Leys, D. 2007. Heme and hemoproteins. *Tetrapyrroles birth, life and death* M. J. Warren and A. J. Smith. New York, New York, Landes Bioscience: 160-183.

Nagy, M., Lacroute, F. & Thomas, D. 1992. Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *Proc Natl Acad Sci U S A,* 89, 8966-70.

Naji, S., Grunberg, S. & Thomm, M. 2007. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem,* 282, 11047-57.

Nichols, D. S., Miller, M. R., Davies, N. W., Goodchild, A., Raftery, M. & Cavicchioli, R. 2004. Cold adaptation in the Antarctic Archaeon *Methanococcoides burtonii* involves membrane lipid unsaturation. *J Bacteriol,* 186, 8508-15.

Nichols, P. D. & Franzmann, P. D. 1992. Unsaturated Diether Phospholipids in the Antarctic Methanogen *Methanococcoides-Burtonii. Fems Microbiology Letters,* 98, 205-208.

Noon, K. R., Guymon, R., Crain, P. F., Mccloskey, J. A., Thomm, M., Lim, J. & Cavicchioli, R. 2003. Influence of temperature on tRNA modification in archaea: *Methanococcoides burtonii* (optimum growth temperature [Topt], 23 degrees C) and Stetteria hydrogenophila (Topt, 95 degrees C). *J Bacteriol,* 185, 5483-90.

Nordlund, P. & Reichard, P. 2006. Ribonucleotide reductases. *Annu Rev Biochem,* 75, 681-706.

Orita, I., Yurimoto, H., Hirai, R., Kawarabayasi, Y., Sakai, Y. & Kato, N. 2005. The archaeon *Pyrococcus horikoshii* possesses a bifunctional enzyme for formaldehyde fixation via the ribulose monophosphate pathway. *J Bacteriol,* 187, 3636-42.

Orlicky, S. M., Tran, P. T., Sayre, M. H. & Edwards, A. M. 2001. Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J Biol Chem,* 276, 10097-102.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci,* 4, 2411-23.

Paiardini, A., Gianese, G., Bossa, F. & Pascarella, S. 2003. Structural plasticity of thermophilic serine hydroxymethyltransferases. *Proteins,* 50, 122-34.

Paraskevopoulou, C., Fairhurst, S. A., Lowe, D. J., Brick, P. & Onesti, S. 2006. The Elongator subunit Elp3 contains a Fe4S4 cluster and binds S-adenosylmethionine. *Mol Microbiol,* 59, 795-806.

Parker, R. & Sheth, U. 2007. P bodies and the control of mRNA translation and degradation. *Mol Cell,* Mar 25, 635-46.

Pavesi, G., Mauri, G., Stefani, M. & Pesole, G. 2004. RNAProfile: an algorithm for finding conserved secondary structure motifs in unaligned RNA sequences. *Nucleic Acids Res,* 32, 3258-69.

Pernthaler, J., Glockner, F. O., Unterholzner, S., Alfreider, A., Psenner, R. & Amann, R. 1998. Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. *Appl Environ Microbiol,* 64, 4299-306.

Phadtare, S., Alsina, J. & Inouye, M. 1999. Cold shock response and cold-shock proteins. *Current Opinion in Microbiology,* 2, 175-180.

Pillai, B., Verma, J., Abraham, A., Francis, P., Kumar, Y., Tatu, U., Brahmachari, S. K. & Sadhale, P. P. 2003. Whole genome expression profiles of yeast RNA polymerase II core subunit, Rpb4, in stress and nonstress conditions. *J Biol Chem,* 278, 3339-46.

Pohlschroder, M., Dilks, K., Hand, N. J. & Wesley Rose, R. 2004. Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiol Rev,* 28, 3-24.

Price, M. N., Huang, K. H., Alm, E. J. & Arkin, A. P. 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res,* 33, 880-92.

Raux, E., Schubert, H. L. & Warren, M. J. 2000. Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. *Cell Mol Life Sci,* 57, 1880-93.

Reeve, J. N. 2003. Archaeal chromatin and transcription. *Mol Microbiol,* 48, 587-98.

Rice, D. W., Yip, K. S., Stillman, T. J., Britton, K. L., Fuentes, A., Connerton, I., Pasquo, A., Scandura, R. & Engel, P. C. 1996. Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol Rev,* 18, 105-17.

Riley, M., Staley, J. T., Danchin, A., Wang, T. Z., Brettin, T. S., Hauser, L. J., Land, M. L. & Thompson, L. S. 2008. Genomics of an extreme psychrophile, *Psychromonas ingrahamii. BMC Genomics,* 9, -.

Ring, G. & Eichler, J. 2004. Extreme secretion: protein translocation across the archael plasma membrane. *J Bioenerg Biomembr,* 36, 35-45.

Rodionov, D. A., Hebbeln, P., Gelfand, M. S. & Eitinger, T. 2006. Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J Bacteriol,* 188, 317-27.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T. & Vriend, G. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics,* 14, 523-8.

Russell, N. J. 1990. Cold adaptation of microorganisms. *Philos Trans R Soc Lond B Biol Sci,* 326, 595-608, discussion 608-11.

Russell, N. J. 2000. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles,* 4, 83-90.

Russell, N. J. 2007. Psychrophiles: Membrane adaptations. *Physiology and biochemistry of extremophiles.* C. Gerday and N. Glansdorff. Washington D. C., ASM press, Washington D. C.

Russell, N. J. & Fukunaga, N. 1990. A Comparison of Thermal Adaptation of Membrane-Lipids in Psychrophilic and Thermophilic Bacteria. *Fems Microbiology Reviews,* 75, 171-182.

Sakai, S., Imachi, H., Hanada, S., Ohashi, A., Harada, H. & Kamagata, Y. 2008. *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellas ord. nov. *Int J Sys Evol Micro* 58, 929-36.

Saunders, N. F., Goodchild, A., Raftery, M., Guilhaus, M., Curmi, P. M. & Cavicchioli, R. 2005. Predicted roles for hypothetical proteins in the low-temperature expressed proteome of the Antarctic archaeon *Methanococcoides burtonii. J Proteome Res,* 4, 464-72.

Saunders, N. F., Ng, C., Raftery, M., Guilhaus, M., Goodchild, A. & Cavicchioli, R. 2006. Proteomic and computational analysis of secreted proteins with type I signal peptides from the Antarctic archaeon *Methanococcoides burtonii. J Proteome Res,* 5, 2457-64.

Saunders, N. F., Thomas, T., Curmi, P. M., Mattick, J. S., Kuczek, E., Slade, R., Davis, J., Franzmann, P. D., Boone, D., Rusterholtz, K., Feldman, R., Gates, C., Bench, S., Sowers, K., Kadner, K., Aerts, A., Dehal, P., Detter, C., Glavina, T., Lucas, S., Richardson, P., Larimer, F., Hauser, L., Land, M. & Cavicchioli, R. 2003. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii. Genome Res,* 13, 1580-8.

Schleper, C., Delong, E. F., Preston, C. M., Feldman, R. A., Wu, K. Y. & Swanson, R. V. 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum. J Bacteriol,* 180, 5003-9.

Schleper, C., Jurgens, G. & Jonuscheit, M. 2005. Genomic studies of uncultivated archaea. *Nat Rev Microbiol,* 3, 479-88.

Schleper, C., Swanson, R. V., Mathur, E. J. & Delong, E. F. 1997. Characterisation of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum. J Bacteriol,* 179, 7803-11.

Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res,* 31, 3381-5.

Selitrennik, M., Duek, L., Lotan, R. & Choder, M. 2006. Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of *Saccharomyces cerevisiae* RNA polymerase II by two pathways. *Eukaryotic Cell,* 5, 2092-103.

Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H. & Wright, F. 1988. Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res,* 16, 8207-11.

Sharp, P. M. & Li, W. H. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res,* 15, 1281-95.

Sharp, P. M. & Matassi, G. 1994. Codon usage and genome evolution. *Curr Opin Genet Dev,* 4, 851-60.

Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., Shirouzu, M. & Yokoyama, S. 2007. Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem Biophys Res Commun,* 30, 122-8.

Shima, S., Warkentin, E., Thauer, R. K. & Ermler, U. 2002. Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen. *J Biosci Bioeng,* 93, 519-30.

Siddiqui, K. S. & Cavicchioli, R. 2006. Cold-adapted enzymes. *Annu Rev Biochem,* 75, 403-33.

Siddiqui, K. S., Cavicchioli, R. & Thomas, T. 2002. Thermodynamic activation properties of elongation factor 2 (EF-2) proteins from psychrotolerant and thermophilic Archaea. *Extremophiles,* 6, 143-50.

Siddiqui, K. S., Poljak, A., Guilhaus, M., De Francisci, D., Curmi, P. M., Feller, G., D'amico, S., Gerday, C., Uversky, V. N. & Cavicchioli, R. 2006. Role of lysine versus arginine in enzyme cold-adaptation: modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from *Pseudoalteramonas haloplanktis. Proteins,* 64, 486-501.

Sinclair, D. A., Kornfeld, G. D. & Dawes, I. W. 1994. Yeast intragenic transcriptional control: activation and repression sites within the coding region of the *Saccharomyces cerevisiae* LPD1 gene. *Mol Cell Biol,* 14, 214-25.

Sliwinski, M. K. & Goodman, R. M. 2004. Spatial heterogeneity of crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. *Appl Environ Microbiol,* 70, 1811-20.

Smalas, A. O., Leiros, H. K., Os, V. & Willassen, N. P. 2000. Cold adapted enzymes. *Biotechnol Annu Rev,* 6, 1-57.

Soppa, J. 1999. Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol,* 31, 1295-305.

Sowers, K., & Noll, M. 1995. Techniques for anaerobic growth. *Archaea: A laboratory manual.* F. T. Robb, A. R. Place, K. Sowers *et al.*, Cold Spring Harbour Laboratory Press.

Stebbins, C. E., Kaelin, W. G., Jr. & Pavletich, N. P. 1999. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science,* 284, 455-61.

Stenmark, P., Moche, M., Gurmu, D. & Nordlund, P. 2007. The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in *Escherichia coli*: implications for the reductive mechanism. *J Mol Biol,* 373, 48-64.

Tarpgaard, I. H., Boetius, A. & Finster, K. 2006. *Desulfobacter psychrotolerans* sp nov., a new psychrotolerant sulfate-reducing bacterium and descriptions of its physiological response to temperature changes. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology,* 89, 109-124.

Terao, Y., Nakamori, S. & Takagi, H. 2003. Gene dosage effect of L-proline biosynthetic enzymes on L-proline accumulation and freeze tolerance in *Saccharomyces cerevisiae. Applied and Environmental Microbiology,* 69, 6527-6532.

Thomas, T. & Cavicchioli, R. 1998. Archaeal cold-adapted proteins: structural and evolutionary analysis of the elongation factor 2 proteins from psychrophilic, mesophilic and thermophilic methanogens. *Febs Letters,* 439, 281-286.

Thomas, T. & Cavicchioli, R. 2000. Effect of temperature on stability and activity of elongation factor 2 proteins from Antarctic and thermophilic methanogens. *Journal of Bacteriology,* 182, 1328-1332.

Thomas, T. & Cavicchioli, R. 2002. Cold adaptation of archaeal elongation factor 2 (EF-2) proteins. *Current Protein & Peptide Science,* 3, 223-230.

Thomas, T., Kumar, N. & Cavicchioli, R. 2001. Effects of ribosomes and intracellular solutes on activities and stabilities of elongation factor 2 proteins from psychrotolerant and thermophilic methanogens. *J Bacteriol,* 183, 1974-82.

Thomm, M. 2007. Transcription: mechanism and regulation. *Archaea: Molecular and cellular biology.* R. Cavicchioli. Washington, D. C., ASM press: 139-157.

Tina, K. G., Bhadra, R. & Srinivasan, N. 2007. PIC: Protein Interactions Calculator. *Nucleic Acids Res,* 35, W473-6.

Tiwari, A., Kumar, D. P. & Bhat, R. 2000. An Efficient and Cost-Effective Procedure for Preparing Samples for Differential Scanning Calorimetry Experiments. *Analytical Biochemistry,* 284, 406-408.

Todone, F., Brick, P., Werner, F., Weinzierl, R. O. & Onesti, S. 2001. Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Molecular Cell,* 8, 1137-43.

Toth, E. A. & Yeates, T. O. 2000. The structure of adenylosuccinate lyase, an enzyme with dual activity in the de novo purine biosynthetic pathway. *Structure with Folding & Design,* 8, 163-174.

Toulokhonov, I., Artsimovitch, I. & Landick, R. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science,* 292, 730-3.

Townley-Tilson W.H., Pendergrass S.A., Marzluff W.F. & Whitfield M.L. 2006. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. *RNA.*10, 1853-67.

Tronelli, D., Maugini, E., Bossa, F. & Pascarella, S. 2007. Structural adaptation to low temperatures--analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS J,* 274, 4595-608.

Trzebiatowski, J. R., O'toole, G. A. & Escalante-Semerena, J. C. 1994. The cobT gene of *Salmonella typhimurium* encodes the NaMN: 5,6-dimethylbenzimidazole phosphoribosyltransferase responsible for the synthesis of N1-(5-phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole, an intermediate in the synthesis of the nucleotide loop of cobalamin. *J Bacteriol,* 176, 3568-75.

Tusher, V. G., Tibshirani, R. & Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* 98, 5116-21.

Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. & Robb, F. T. 1998. Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc Natl Acad Sci U S A,* 95, 12300-5.

Vorholt, J. A., Vaupel, M. & Thauer, R. K. 1996. A polyferredoxin with eight [4Fe-4S] clusters as a subunit of molybdenum formylmethanofuran dehydrogenase from *Methanosarcina barkeri*. *Eur J Biochem,* 236, 309-17.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph,* 8, 52-6, 29.

Wada, M., Fukunaga, N. & Sasaki, S. 1989. Mechanism of biosynthesis of unsaturated fatty acids in *Pseudomonas sp.* strain E-3, a psychrotrophic bacterium. *J Bacteriol,* 171, 4267-71.

Warren, M. J., Deery, E. & Rose, R. 2007. Biosynthesis of siroheme and coenzyme F_{430}. *Tetrapyrroles birth, life and death* M. J. Warren and A. J. Smith. New York, New York, Landes Bioscience: 343-351.

Welander, P. V. & Metcalf, W. W. 2005. Loss of the mtr operon in Methanosarcina blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. *Proc Natl Acad Sci U S A,* 102, 10664-9.

Werner, F. & Weinzierl, R. O. 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Molecular Cell,* 10, 635-46.

Williams, T. J., Burg, D. W., Ertan, H., Raftery, M. J., Poljak, A., Guilhaus, M. & Cavicchioli, R. 2010a. Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii*. Part II: the effect of different methylated growth substrates. *J Proteome Res,* 9, 653-63.

Williams, T. J., Burg, D. W., Raftery, M. J., Poljak, A., Guilhaus, M., Pilak, O. & Cavicchioli, R. 2010b. Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii*. Part I: the effect of growth temperature. *J Proteome Res,* 9, 640-52.

Woese, C. R. & Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A,* 74, 5088-90.

Woese, C. R., Kandler, O. & 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, 4576-9.

Yamada, T., Kuroda, K., Jitsuyama, Y., Takezawa, D., Arakawa, K. & Fujikawa, S. 2002. Roles of the plasma membrane and the cell wall in the responses of plant cells to freezing. *Planta,* 215, 770-778.

Yamanaka, K. 1999. Cold shock response in *Escherichia coli. J Mol Microbiol Biotechnol,* 1, 193-202.

Yan, H. G. & Tsai, M. D. 1999. Nucleoside monophosphate kinases: Structure, mechanism, and substrate specificity. *Advances in Enzymology, Vol 73,* 73, 103-+.

Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., Narasimhan, S., Kane, D. W., Reinhold, W. C., Lababidi, S., Bussey, K. J., Riss, J., Barrett, J. C. & Weinstein, J. N. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol,* 4, R28.

Zhang, R. G., Andersson, C. E., Savchenko, A., Skarina, T., Evdokimova, E., Beasley, S., Arrowsmith, C. H., Edwards, A. M., Joachimiak, A. & Mowbray, S. L. 2003. Structure of *Escherichia coli* ribose-5-phosphate isomerase: A ubiquitous enzyme of the pentose phosphate pathway and the Calvin cycle. *Structure,* 11, 31-42.

Appendix 1

Microarray data and array design download.

Array design and microarray data for the *Mb*RpoE/F binding experiment were submitted to ArrayExpress database following MIAME compliant.

Array design contains a detailed description of the Agilent platform used in experiments and also all the oligo sequences.

Array Design ID: A-MEXP-1794.adf.txt

RpoE/F binding experiments:

Experiment name: Mbur_RpoEF

ArrayExpress accession: E-MEXP-2625

Specified release date: 2011-03-01 (YYYY-MM-DD)

Reviewers can access microarray description file and data using the following account:

Username: Reviewer_E-MEXP-2625

For the password contact Dr. Stefano Campanaro (stefano.campanaro@gmail.com)

Appendix 2

Transcripts bound by RpoE/F were classified using COG classes and the statistical over-representation evaluated using hypergeometric distribution with "R" statistical package. The details are in Table Appendix 2.

Table Appendix 2 - In columns are reported: (A) COG codes, (B) total number of genes belonging to each COG in the *M. burtonii* microarray, (C) number of genes coding transcripts bound by RpoE/F, (D) percentage of genes coding transcripts bound by RpoE/F respect to the total number of genes, (E) the p-value determined using hypergeometric distribution (classes having p-value lower than 0.05 are in bold character), (F) COG description and (G) COG classes are grouped in more general classes as reported in (http://www.ncbi.nlm.nih.gov/COG/grace/fiew.cgi).

COG	Total number of genes	Genes bound by RpoE/F	Percentage of genes bound by RpoE/F	Hypergeometric (p value)	COG description	COG "supergroups"
C	**149**	**37**	**24,83%**	**0,000**	**Energy production and conversion**	**Metabolism**
H	**141**	**16**	**11,35%**	**0,001**	**Coenzyme transport and metabolism**	**Metabolism**
B	**2**	**1**	**50,00%**	**0,003**	**Chromatin structure and dynamics**	**Information storage and processing**
F	**51**	**6**	**11,76%**	**0,015**	**Nucleotide transport and metabolism**	**Metabolism**
I	**22**	**3**	**13,64%**	**0,025**	**Lipid transport and metabolism**	**Metabolism**
U	**38**	**4**	**10,53%**	**0,045**	**Intracellular trafficking, secretion, and vesicular transport**	**Cellular processes and signaling**
A	1	0	0,00%	0,052	RNA processing and modification	Information storage and processing
G	60	5	8,33%	0,089	Carbohydrate transport and metabolism	Metabolism
K	94	7	7,45%	0,113	Transcription	Information storage and processing
D	13	1	7,69%	0,145	Cell cycle control, cell division, chromosome partitioning	Cellular processes and signaling
E	141	8	5,67%	0,312	Amino acid transport and metabolism	Metabolism
Q	12	0	0,00%	0,475	Secondary metabolites biosynthesis, transport and catabolism	Metabolism
P	92	4	4,35%	0,530	Inorganic ion transport and metabolism	Metabolism
J	165	7	4,24%	0,641	Translation, ribosomal structure and biogenesis	Information storage and processing
S	220	9	4,09%	0,727	Function unknown	Poorly characterized

O	77	2	2,60%	0,777	Posttranslational modification, protein turnover, chaperones	Cellular processes and signaling
V	29	0	0,00%	0,790	Defense mechanisms	Cellular processes and signaling
N	33	0	0,00%	0,831	Cell motility	Cellular processes and signaling
R	225	7	3,11%	0,915	General function prediction only	Poorly characterized
L	104	2	1,92%	0,917	Replication, recombination and repair	Information storage and processing
M	84	0	0,00%	0,990	Cell wall/membrane/envelope biogenesis	Cellular processes and signaling
T	95	0	0,00%	0,994	Signal transduction mechanisms	Cellular processes and signaling
-	549	6	1,09%	1,000	No COG	-

Appendix 3

Transcripts bound by RpoE/F were classified according to GO classes using GoMiner software. In the Table are reported all the GO classes obtained from this analysis. Classes having p-value lower than 0.05 are in bold character.

Table Appendix 3 - In the first column are reported the Gene Ontology ID of each class identified (for example GO:0006091) followed by the description of the Gene Ontology class (for example generation_of_precursor_metabolites_and_energy); in the second colum are reported the total number of genes belonging to each category and present in the *M. burtonii* microarray; in the third column are reported the number of the genes coding transcripts bound by RpoE/F per each GO class; in the fourth column are reported the p-values obtained from statistical analysis performed with GoMiner software.

GO category	Total number of genes	Number of transcripts bound by RpoE/F	p-value
GO:0006091_generation_of_precursor_metabolites_and_energy	**40**	**12**	**3,57E-05**
GO:0006730_one-carbon_compound_metabolic_process	**39**	**11**	**0,000148**
GO:0051186_cofactor_metabolic_process	**65**	**14**	**0,000414**
GO:0006084_acetyl-CoA_metabolic_process	**3**	**3**	**0,000537**
GO:0006814_sodium_ion_transport	**6**	**4**	**0,000562**
GO:0045333_cellular_respiration	**20**	**7**	**0,000632**
GO:0015980_energy_derivation_by_oxidation_of_organic_compounds	**22**	**7**	**0,001214**
GO:0009061_anaerobic_respiration	**19**	**6**	**0,002919**
GO:0015947_methane_metabolic_process	**19**	**6**	**0,002919**
GO:0015948_methanogenesis	**19**	**6**	**0,002919**

GO:0015975_energy_derivation_by_oxidation_of_reduced_inorganic_compounds	**19**	**6**	**0,002919**
GO:0019753_one-carbon_compound_biosynthetic_process	**19**	**6**	**0,002919**
GO:0043446_alkane_metabolic_process	**19**	**6**	**0,002919**
GO:0043447_alkane_biosynthetic_process	**19**	**6**	**0,002919**
GO:0006732_coenzyme_metabolic_process	**27**	**7**	**0,004499**
GO:0046483_heterocycle_metabolic_process	**57**	**11**	**0,004706**
GO:0006753_nucleoside_phosphate_metabolic_process	**50**	**10**	**0,005413**
GO:0009117_nucleotide_metabolic_process	**50**	**10**	**0,005413**
GO:0022900_electron_transport_chain	**10**	**4**	**0,006078**
GO:0022904_respiratory_electron_transport_chain	**10**	**4**	**0,006078**
GO:0042773_ATP_synthesis_coupled_electron_transport	**10**	**4**	**0,006078**
GO:0055114_oxidation_reduction	**10**	**4**	**0,006078**
GO:0055086_nucleobase__nucleoside_and_nucleotide_metabolic_process	**59**	**11**	**0,006218**
GO:0015628_protein_secretion_by_the_type_II_secretion_system	**2**	**2**	**0,006682**
GO:0032787_monocarboxylic_acid_metabolic_process	**6**	**3**	**0,008936**
GO:0006098_pentose-phosphate_shunt	**3**	**2**	**0,018974**
GO:0006740_NADPH_regeneration	**3**	**2**	**0,018974**
GO:0006760_folic_acid_and_derivative_metabolic_process	**3**	**2**	**0,018974**
GO:0009396_folic_acid_and_derivative_biosynthetic_process	**3**	**2**	**0,018974**
GO:0033013_tetrapyrrole_metabolic_process	**37**	**7**	**0,026526**
GO:0033014_tetrapyrrole_biosynthetic_process	**37**	**7**	**0,026526**
GO:0032940_secretion_by_cell	**15**	**4**	**0,028738**
GO:0046903_secretion	**15**	**4**	**0,028738**
GO:0051641_cellular_localization	**15**	**4**	**0,028738**
GO:0051649_establishment_of_cellular_localization	**15**	**4**	**0,028738**
GO:0030001_metal_ion_transport	**31**	**6**	**0,03554**
GO:0006739_NADP_metabolic_process	**4**	**2**	**0,035928**
GO:0006752_group_transfer_coenzyme_metabolic_process	**4**	**2**	**0,035928**
GO:0006119_oxidative_phosphorylation	**17**	**4**	**0,044167**
GO:0008104_protein_localization	**17**	**4**	**0,044167**
GO:0033036_macromolecule_localization	**17**	**4**	**0,044167**
GO:0045184_establishment_of_protein_localization	**17**	**4**	**0,044167**
GO:0009306_protein_secretion	11	3	0,054547
GO:0015672_monovalent_inorganic_cation_transport	26	5	0,05518
GO:0006207_'de_novo'_pyrimidine_base_biosynthetic_process	5	2	0,056715
GO:0019856_pyrimidine_base_biosynthetic_process	5	2	0,056715
GO:0009987_cellular_process	713	65	0,057164
GO:0006778_porphyrin_metabolic_process	35	6	0,060141
GO:0006779_porphyrin_biosynthetic_process	35	6	0,060141
GO:0009165_nucleotide_biosynthetic_process	45	7	0,068361
GO:0051188_cofactor_biosynthetic_process	55	8	0,073591
GO:0006066_alcohol_metabolic_process	20	4	0,074503
GO:0051234_establishment_of_localization	136	16	0,077309
GO:0044249_cellular_biosynthetic_process	302	31	0,07803
GO:0006206_pyrimidine_base_metabolic_process	6	2	0,080604
GO:0006886_intracellular_protein_transport	6	2	0,080604
GO:0046907_intracellular_transport	6	2	0,080604
GO:0006071_glycerol_metabolic_process	1	1	0,08222
GO:0006099_tricarboxylic_acid_cycle	1	1	0,08222

GO:0006171_cAMP_biosynthetic_process	1	1	0.08222
GO:0006222_UMP_biosynthetic_process	1	1	0.08222
GO:0006422_aspartyl-tRNA_aminoacylation	1	1	0.08222
GO:0006433_prolyl-tRNA_aminoacylation	1	1	0.08222
GO:0006435_threonyl-tRNA_aminoacylation	1	1	0.08222
GO:0006544_glycine_metabolic_process	1	1	0.08222
GO:0009052_pentose-phosphate_shunt__non-oxidative_branch	1	1	0.08222
GO:0009060_aerobic_respiration	1	1	0.08222
GO:0009173_pyrimidine_ribonucleoside_monophosphate_metabolic_process	1	1	0.08222
GO:0009174_pyrimidine_ribonucleoside_monophosphate_biosynthetic_process	1	1	0.08222
GO:0009187_cyclic_nucleotide_metabolic_process	1	1	0.08222
GO:0009190_cyclic_nucleotide_biosynthetic_process	1	1	0.08222
GO:0019751_polyol_metabolic_process	1	1	0.08222
GO:0046049_UMP_metabolic_process	1	1	0.08222
GO:0046058_cAMP_metabolic_process	1	1	0.08222
GO:0046356_acetyl-CoA_catabolic_process	1	1	0.08222
GO:0006007_glucose_catabolic_process	13	3	0.083958
GO:0016052_carbohydrate_catabolic_process	13	3	0.083958
GO:0019320_hexose_catabolic_process	13	3	0.083958
GO:0044275_cellular_carbohydrate_catabolic_process	13	3	0.083958
GO:0046164_alcohol_catabolic_process	13	3	0.083958
GO:0046365_monosaccharide_catabolic_process	13	3	0.083958
GO:0044237_cellular_metabolic_process	673	61	0.093958
GO:0051179_localization	142	16	0.105802
GO:0046112_nucleobase_biosynthetic_process	7	2	0.106959
GO:0006006_glucose_metabolic_process	15	3	0.118732
GO:0006220_pyrimidine_nucleotide_metabolic_process	15	3	0.118732
GO:0044265_cellular_macromolecule_catabolic_process	15	3	0.118732
GO:0016043_cellular_component_organization_and_biogenesis	42	6	0.122832
GO:0006766_vitamin_metabolic_process	52	7	0.126526
GO:0006767_water-soluble_vitamin_metabolic_process	52	7	0.126526
GO:0009235_cobalamin_metabolic_process	24	4	0.127505
GO:0009236_cobalamin_biosynthetic_process	24	4	0.127505
GO:0015031_protein_transport	8	2	0.135223
GO:0006812_cation_transport	53	7	0.136449
GO:0006810_transport	128	14	0.152058
GO:0008535_respiratory_chain_complex_IV_assembly	2	1	0.157757
GO:0009109_coenzyme_catabolic_process	2	1	0.157757
GO:0009129_pyrimidine_nucleoside_monophosphate_metabolic_process	2	1	0.157757
GO:0009130_pyrimidine_nucleoside_monophosphate_biosynthetic_process	2	1	0.157757
GO:0009218_pyrimidine_ribonucleotide_metabolic_process	2	1	0.157757
GO:0009220_pyrimidine_ribonucleotide_biosynthetic_process	2	1	0.157757
GO:0015684_ferrous_iron_transport	2	1	0.157757
GO:0019321_pentose_metabolic_process	2	1	0.157757
GO:0051187_cofactor_catabolic_process	2	1	0.157757
GO:0065002_intracellular_protein_transport_across_a_membrane	26	4	0.158654
GO:0044248_cellular_catabolic_process	9	2	0.164914
GO:0009112_nucleobase_metabolic_process	18	3	0.178815
GO:0019318_hexose_metabolic_process			

GO:0006725_aromatic_compound_metabolic_process	37	5	0,181076
GO:0006769_nicotinamide_metabolic_process	10	2	0.195614
GO:0009156_ribonucleoside_monophosphate_biosynthetic_process	10	2	0.195614
GO:0009161_ribonucleoside_monophosphate_metabolic_process	10	2	0.195614
GO:0005996_monosaccharide_metabolic_process	19	3	0,200447
GO:0009259_ribonucleotide_metabolic_process	19	3	0,200447
GO:0009260_ribonucleotide_biosynthetic_process	19	3	0,200447
GO:0006096_glycolysis	11	2	0,226962
GO:0009123_nucleoside_monophosphate_metabolic_process	11	2	0,226962
GO:0009124_nucleoside_monophosphate_biosynthetic_process	11	2	0,226962
GO:0019362_pyridine_nucleotide_metabolic_process	11	2	0,226962
GO:0001522_pseudouridine_synthesis	3	1	0,227149
GO:0006090_pyruvate_metabolic_process	3	1	0,227149
GO:0006094_gluconeogenesis	3	1	0,227149
GO:0007049_cell_cycle	3	1	0,227149
GO:0015886_heme_transport	3	1	0,227149
GO:0016226_iron-sulfur_cluster_assembly	3	1	0,227149
GO:0031119_tRNA_pseudouridine_synthesis	3	1	0,227149
GO:0031163_metallo-sulfur_cluster_assembly	3	1	0,227149
GO:0051181_cofactor_transport	3	1	0,227149
GO:0006418_tRNA_aminoacylation_for_protein_translation	21	3	0,245373
GO:0043038_amino_acid_activation	21	3	0,245373
GO:0043039_tRNA_aminoacylation	21	3	0,245373
GO:0006811_ion_transport	63	7	0,254211
GO:0006733_oxidoreduction_coenzyme_metabolic_process	12	2	0,25865
GO:0009056_catabolic_process	32	4	0,26534
GO:0006163_purine_nucleotide_metabolic_process	22	3	0,268432
GO:0006164_purine_nucleotide_biosynthetic_process	22	3	0,268432
GO:0009058_biosynthetic_process	366	33	0,279204
GO:0006221_pyrimidine_nucleotide_biosynthetic_process	13	2	0,290412
GO:0006352_transcription_initiation	4	1	0,290889
GO:0006563_L-serine_metabolic_process	4	1	0,290889
GO:0006605_protein_targeting	4	1	0,290889
GO:0006612_protein_targeting_to_membrane	4	1	0,290889
GO:0006613_cotranslational_protein_targeting_to_membrane	4	1	0,290889
GO:0006614_SRP-dependent_cotranslational_protein_targeting_to_membrane	4	1	0,290889
GO:0019319_hexose_biosynthetic_process	4	1	0,290889
GO:0045045_secretory_pathway	4	1	0,290889
GO:0045047_protein_targeting_to_ER	4	1	0,290889
GO:0046165_alcohol_biosynthetic_process	4	1	0,290889
GO:0046364_monosaccharide_biosynthetic_process	4	1	0,290889
GO:0065004_protein-DNA_complex_assembly	25	3	0,338726
GO:0009057_macromolecule_catabolic_process	5	1	0,349433
GO:0006120_mitochondrial_electron_transport__NADH_to_ubiquinone	5	1	0,349433
GO:0009069_serine_family_amino_acid_metabolic_process	5	1	0,349433
GO:0042775_organelle_ATP_synthesis_coupled_electron_transport	5	1	0,349433
GO:0051301_cell_division	15	2	0,353301
GO:0015674_di-__tri-valent_inorganic_cation_transport	48	5	0,3594
GO:0009110_vitamin_biosynthetic_process			

GO:0042364_water-soluble_vitamin_biosynthetic_process	48	5	0,3594
GO:0065003_macromolecular_complex_assembly	16	2	0,384082
GO:0006189_'de_novo'_IMP_biosynthetic_process	6	1	0,403199
GO:0006461_protein_complex_assembly	6	1	0,403199
GO:0006551_leucine_metabolic_process	6	1	0,403199
GO:0009098_leucine_biosynthetic_process	6	1	0,403199
GO:0017004_cytochrome_complex_assembly	6	1	0,403199
GO:0043623_cellular_protein_complex_assembly	6	1	0,403199
GO:0000041_transition_metal_ion_transport	17	2	0,41424
GO:0022607_cellular_component_assembly	17	2	0,41424
GO:0006399_tRNA_metabolic_process	40	4	0,420846
GO:0006139_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	278	24	0,42819
GO:0009150_purine_ribonucleotide_metabolic_process	18	2	0,443668
GO:0009152_purine_ribonucleotide_biosynthetic_process	18	2	0,443668
GO:0006188_IMP_biosynthetic_process	7	1	0,452572
GO:0006400_tRNA_modification	7	1	0,452572
GO:0006413_translational_initiation	7	1	0,452572
GO:0006824_cobalt_ion_transport	7	1	0,452572
GO:0006826_iron_ion_transport	7	1	0,452572
GO:0009086_methionine_biosynthetic_process	7	1	0,452572
GO:0022618_protein-RNA_complex_assembly	7	1	0,452572
GO:0046040_IMP_metabolic_process	7	1	0,452572
GO:0044262_cellular_carbohydrate_metabolic_process	30	3	0,454352
GO:0006555_methionine_metabolic_process	8	1	0,497908
GO:0006720_isoprenoid_metabolic_process	8	1	0,497908
GO:0008299_isoprenoid_biosynthetic_process	8	1	0,497908
GO:0009451_RNA_modification	8	1	0,497908
GO:0009108_coenzyme_biosynthetic_process	20	2	0,500016
GO:0000097_sulfur_amino_acid_biosynthetic_process	9	1	0,539532
GO:0006772_thiamin_metabolic_process	9	1	0,539532
GO:0006865_amino_acid_transport	9	1	0,539532
GO:0009126_purine_nucleoside_monophosphate_metabolic_process	9	1	0,539532
GO:0009127_purine_nucleoside_monophosphate_biosynthetic_process	9	1	0,539532
GO:0009167_purine_ribonucleoside_monophosphate_metabolic_process	9	1	0,539532
GO:0009168_purine_ribonucleoside_monophosphate_biosynthetic_process	9	1	0,539532
GO:0009228_thiamin_biosynthetic_process	9	1	0,539532
GO:0015837_amine_transport	9	1	0,539532
GO:0042723_thiamin_and_derivative_metabolic_process	9	1	0,539532
GO:0042724_thiamin_and_derivative_biosynthetic_process	9	1	0,539532
GO:0044272_sulfur_compound_biosynthetic_process	9	1	0,539532
GO:0000096_sulfur_amino_acid_metabolic_process	10	1	0,577745
GO:0009082_branched_chain_family_amino_acid_biosynthetic_process	10	1	0,577745
GO:0006082_organic_acid_metabolic_process	112	9	0,587539
GO:0019752_carboxylic_acid_metabolic_process	112	9	0,587539
GO:0016051_carbohydrate_biosynthetic_process	24	2	0,601355
GO:0000105_histidine_biosynthetic_process	11	1	0,612823
GO:0006547_histidine_metabolic_process	11	1	0,612823
GO:0009075_histidine_family_amino_acid_metabolic_process	11	1	0,612823
GO:0009076_histidine_family_amino_acid_biosynthetic_process	11	1	0,612823

GO:0009081_branched_chain_family_amino_acid_metabolic_process	11	1	0.612823
GO:0015849_organic_acid_transport	11	1	0.612823
GO:0046942_carboxylic_acid_transport	11	1	0.612823
GO:0006350_transcription	91	7	0.638036
GO:0006310_DNA_recombination	12	1	0.645021
GO:0006818_hydrogen_transport	12	1	0.645021
GO:0015992_proton_transport	12	1	0.645021
GO:0006260_DNA_replication	26	2	0.646023
GO:0006790_sulfur_metabolic_process	13	1	0.674572
GO:0005975_carbohydrate_metabolic_process	55	4	0.679649
GO:0009067_aspartate_family_amino_acid_biosynthetic_process	14	1	0.701691
GO:0022613_ribonucleoprotein_complex_biogenesis_and_assembly	14	1	0.701691
GO:0009066_aspartate_family_amino_acid_metabolic_process	15	1	0.726576
GO:0008152_metabolic_process	842	68	0.730278
GO:0016310_phosphorylation	60	4	0.746103
GO:0006793_phosphorus_metabolic_process	61	4	0.758082
GO:0006796_phosphate_metabolic_process	61	4	0.758082
GO:0006520_amino_acid_metabolic_process	104	7	0.775633
GO:0006519_amino_acid_and_derivative_metabolic_process	106	7	0.793034
GO:0009308_amine_metabolic_process	106	7	0.793034
GO:0008033_tRNA_processing	19	1	0.807202
GO:0006807_nitrogen_compound_metabolic_process	110	7	0.824846
GO:0008610_lipid_biosynthetic_process	22	1	0.851793
GO:0044255_cellular_lipid_metabolic_process	23	1	0.864261
GO:0006629_lipid_metabolic_process	24	1	0.875691
GO:0006396_RNA_processing	25	1	0.88617
GO:0006412_translation	106	6	0.89056
GO:0006281_DNA_repair	30	1	0.926821
GO:0006974_response_to_DNA_damage_stimulus	30	1	0.926821
GO:0009719_response_to_endogenous_stimulus	30	1	0.926821
GO:0008652_amino_acid_biosynthetic_process	67	3	0.928214
GO:0009309_amine_biosynthetic_process	68	3	0.932737
GO:0044271_nitrogen_compound_biosynthetic_process	68	3	0.932737
GO:0010467_gene_expression	224	13	0.953898
GO:0009059_macromolecule_biosynthetic_process	123	6	0.955417
GO:0006950_response_to_stress	42	1	0.974909
GO:0016070_RNA_metabolic_process	133	5	0.991231
GO:0006259_DNA_metabolic_process	79	2	0.99262
GO:0050896_response_to_stimulus	56	1	0.992933
GO:0006355_regulation_of_transcription__DNA-dependent	74	1	0.998656
GO:0043412_biopolymer_modification	74	1	0.998656
GO:0051252_regulation_of_RNA_metabolic_process	74	1	0.998656
GO:0045449_regulation_of_transcription	78	1	0.999075
GO:0019219_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	79	1	0.999158
GO:0044267_cellular_protein_metabolic_process	201	7	0.999178
GO:0006351_transcription__DNA-dependent	80	1	0.999234
GO:0019538_protein_metabolic_process	205	7	0.999388
GO:0019222_regulation_of_metabolic_process	83	1	0.999422
GO:0031323_regulation_of_cellular_metabolic_process	83	1	0.999422

GO:0032774_RNA_biosynthetic_process	83	1	0,999422
GO:0010468_regulation_of_gene_expression	86	1	0,999563
GO:0050789_regulation_of_biological_process	87	1	0,999604
GO:0050794_regulation_of_cellular_process	87	1	0,999604
GO:0044260_cellular_macromolecule_metabolic_process	211	7	0,999609
GO:0065007_biological_regulation	99	1	0,999873
GO:0044238_primary_metabolic_process	575	32	0,999896
GO:0043283_biopolymer_metabolic_process	266	7	0,999995
GO:0043170_macromolecule_metabolic_process	400	13	1
GO:0000162_tryptophan_biosynthetic_process	5	0	1
GO:0006284_base-excision_repair	5	0	1
GO:0006526_arginine_biosynthetic_process	5	0	1
GO:0006771_riboflavin_metabolic_process	5	0	1
GO:0006777_Mo-molybdopterin_cofactor_biosynthetic_process	5	0	1
GO:0006935_chemotaxis	5	0	1
GO:0007610_behavior	5	0	1
GO:0007626_locomotory_behavior	5	0	1
GO:0009231_riboflavin_biosynthetic_process	5	0	1
GO:0019720_Mo-molybdopterin_cofactor_metabolic_process	5	0	1
GO:0030163_protein_catabolic_process	5	0	1
GO:0032324_molybdopterin_cofactor_biosynthetic_process	5	0	1
GO:0040029_regulation_of_gene_expression__epigenetic	5	0	1
GO:0042330_taxis	5	0	1
GO:0042435_indole_derivative_biosynthetic_process	5	0	1
GO:0042558_pteridine_and_derivative_metabolic_process	5	0	1
GO:0042559_pteridine_and_derivative_biosynthetic_process	5	0	1
GO:0043414_biopolymer_methylation	5	0	1
GO:0043545_molybdopterin_cofactor_metabolic_process	5	0	1
GO:0046219_indolalkylamine_biosynthetic_process	5	0	1
GO:0000394_RNA_splicing__via_endonucleolytic_cleavage_and_ligation	2	0	1
GO:0000910_cytokinesis	2	0	1
GO:0000917_barrier_septum_formation	2	0	1
GO:0001539_ciliary_or_flagellar_motility	2	0	1
GO:0002376_immune_system_process	2	0	1
GO:0005977_glycogen_metabolic_process	2	0	1
GO:0005978_glycogen_biosynthetic_process	2	0	1
GO:0006012_galactose_metabolic_process	2	0	1
GO:0006040_amino_sugar_metabolic_process	2	0	1
GO:0006041_glucosamine_metabolic_process	2	0	1
GO:0006044_N-acetylglucosamine_metabolic_process	2	0	1
GO:0006047_UDP-N-acetylglucosamine_metabolic_process	2	0	1
GO:0006073_glucan_metabolic_process	2	0	1
GO:0006112_energy_reserve_metabolic_process	2	0	1
GO:0006177_GMP_biosynthetic_process	2	0	1
GO:0006213_pyrimidine_nucleoside_metabolic_process	2	0	1
GO:0006313_transposition__DNA-mediated	2	0	1
GO:0006323_DNA_packaging	2	0	1
GO:0006325_establishment_and_or_maintenance_of_chromatin_architecture	2	0	1
GO:0006333_chromatin_assembly_or_disassembly	2	0	1

GO:0006366_transcription_from_RNA_polymerase_II_promoter	2	0	1
GO:0006367_transcription_initiation_from_RNA_polymerase_II_promoter	2	0	1
GO:0006388_tRNA_splicing	2	0	1
GO:0006415_translational_termination	2	0	1
GO:0006419_alanyl-tRNA_aminoacylation	2	0	1
GO:0006450_regulation_of_translational_fidelity	2	0	1
GO:0006486_protein_amino_acid_glycosylation	2	0	1
GO:0006511_ubiquitin-dependent_protein_catabolic_process	2	0	1
GO:0006549_isoleucine_metabolic_process	2	0	1
GO:0006564_L-serine_biosynthetic_process	2	0	1
GO:0006566_threonine_metabolic_process	2	0	1
GO:0006573_valine_metabolic_process	2	0	1
GO:0006595_polyamine_metabolic_process	2	0	1
GO:0006662_glycerol_ether_metabolic_process	2	0	1
GO:0006800_oxygen_and_reactive_oxygen_species_metabolic_process	2	0	1
GO:0006821_chloride_transport	2	0	1
GO:0006952_defense_response	2	0	1
GO:0006955_immune_response	2	0	1
GO:0007155_cell_adhesion	2	0	1
GO:0008380_RNA_splicing	2	0	1
GO:0009063_amino_acid_catabolic_process	2	0	1
GO:0009065_glutamine_family_amino_acid_catabolic_process	2	0	1
GO:0009088_threonine_biosynthetic_process	2	0	1
GO:0009095_aromatic_amino_acid_family_biosynthetic_process__prephenate_pathway	2	0	1
GO:0009097_isoleucine_biosynthetic_process	2	0	1
GO:0009099_valine_biosynthetic_process	2	0	1
GO:0009100_glycoprotein_metabolic_process	2	0	1
GO:0009101_glycoprotein_biosynthetic_process	2	0	1
GO:0009113_purine_base_biosynthetic_process	2	0	1
GO:0009147_pyrimidine_nucleoside_triphosphate_metabolic_process	2	0	1
GO:0009219_pyrimidine_deoxyribonucleotide_metabolic_process	2	0	1
GO:0009221_pyrimidine_deoxyribonucleotide_biosynthetic_process	2	0	1
GO:0009250_glucan_biosynthetic_process	2	0	1
GO:0009262_deoxyribonucleotide_metabolic_process	2	0	1
GO:0009263_deoxyribonucleotide_biosynthetic_process	2	0	1
GO:0009307_DNA_restriction-modification_system	2	0	1
GO:0009399_nitrogen_fixation	2	0	1
GO:0010035_response_to_inorganic_substance	2	0	1
GO:0010038_response_to_metal_ion	2	0	1
GO:0015009_corrin_metabolic_process	2	0	1
GO:0016053_organic_acid_biosynthetic_process	2	0	1
GO:0016485_protein_processing	2	0	1
GO:0019877_diaminopimelate_biosynthetic_process	2	0	1
GO:0019941_modification-dependent_protein_catabolic_process	2	0	1
GO:0022411_cellular_component_disassembly	2	0	1
GO:0022610_biological_adhesion	2	0	1
GO:0031123_RNA_3'-end_processing	2	0	1
GO:0031497_chromatin_assembly	2	0	1
GO:0032196_transposition	2	0	1

GO:0032506_cytokinetic_process	2	0	1
GO:0032984_macromolecular_complex_disassembly	2	0	1
GO:0042780_tRNA_3'-end_processing	2	0	1
GO:0042816_vitamin_B6_metabolic_process	2	0	1
GO:0042819_vitamin_B6_biosynthetic_process	2	0	1
GO:0043241_protein_complex_disassembly	2	0	1
GO:0043624_cellular_protein_complex_disassembly	2	0	1
GO:0043632_modification-dependent_macromolecule_catabolic_process	2	0	1
GO:0044257_cellular_protein_catabolic_process	2	0	1
GO:0045087_innate_immune_response	2	0	1
GO:0046037_GMP_metabolic_process	2	0	1
GO:0046140_corrin_biosynthetic_process	2	0	1
GO:0046394_carboxylic_acid_biosynthetic_process	2	0	1
GO:0046677_response_to_antibiotic	2	0	1
GO:0051603_proteolysis_involved_in_cellular_protein_catabolic_process	2	0	1
GO:0006265_DNA_topological_change	6	0	1
GO:0006541_glutamine_metabolic_process	6	0	1
GO:0006928_cell_motility	6	0	1
GO:0032259_methylation	6	0	1
GO:0042398_amino_acid_derivative_biosynthetic_process	6	0	1
GO:0042401_biogenic_amine_biosynthetic_process	6	0	1
GO:0042726_riboflavin_and_derivative_metabolic_process	6	0	1
GO:0042727_riboflavin_and_derivative_biosynthetic_process	6	0	1
GO:0042777_plasma_membrane_ATP_synthesis_coupled_proton_transport	6	0	1
GO:0051674_localization_of_cell	6	0	1
GO:0006996_organelle_organization_and_biogenesis	11	0	1
GO:0000051_urea_cycle_intermediate_metabolic_process	7	0	1
GO:0006525_arginine_metabolic_process	7	0	1
GO:0006643_membrane_lipid_metabolic_process	7	0	1
GO:0006644_phospholipid_metabolic_process	7	0	1
GO:0006754_ATP_biosynthetic_process	7	0	1
GO:0008653_lipopolysaccharide_metabolic_process	7	0	1
GO:0008654_phospholipid_biosynthetic_process	7	0	1
GO:0009103_lipopolysaccharide_biosynthetic_process	7	0	1
GO:0009435_NAD_biosynthetic_process	7	0	1
GO:0015985_energy_coupled_proton_transport__down_electrochemical_gradient	7	0	1
GO:0015986_ATP_synthesis_coupled_proton_transport	7	0	1
GO:0019674_NAD_metabolic_process	7	0	1
GO:0046034_ATP_metabolic_process	7	0	1
GO:0046467_membrane_lipid_biosynthetic_process	7	0	1
GO:0006820_anion_transport	9	0	1
GO:0009141_nucleoside_triphosphate_metabolic_process	9	0	1
GO:0009142_nucleoside_triphosphate_biosynthetic_process	9	0	1
GO:0009605_response_to_external_stimulus	9	0	1
GO:0015698_inorganic_anion_transport	9	0	1
GO:0019725_cellular_homeostasis	9	0	1
GO:0001682_tRNA_5'-leader_removal	4	0	1
GO:0006268_DNA_unwinding_during_replication	4	0	1
GO:0006289_nucleotide-excision_repair	4	0	1

GO:0006298_mismatch_repair	4	0	1
GO:0006305_DNA_alkylation	4	0	1
GO:0006306_DNA_methylation	4	0	1
GO:0006308_DNA_catabolic_process	4	0	1
GO:0006364_rRNA_processing	4	0	1
GO:0006414_translational_elongation	4	0	1
GO:0006417_regulation_of_translation	4	0	1
GO:0006560_proline_metabolic_process	4	0	1
GO:0006561_proline_biosynthetic_process	4	0	1
GO:0006817_phosphate_transport	4	0	1
GO:0006827_high-affinity_iron_ion_transport	4	0	1
GO:0009070_serine_family_amino_acid_biosynthetic_process	4	0	1
GO:0009225_nucleotide-sugar_metabolic_process	4	0	1
GO:0009889_regulation_of_biosynthetic_process	4	0	1
GO:0009991_response_to_extracellular_stimulus	4	0	1
GO:0010556_regulation_of_macromolecule_biosynthetic_process	4	0	1
GO:0015893_drug_transport	4	0	1
GO:0016072_rRNA_metabolic_process	4	0	1
GO:0031326_regulation_of_cellular_biosynthetic_process	4	0	1
GO:0031668_cellular_response_to_extracellular_stimulus	4	0	1
GO:0032268_regulation_of_cellular_protein_metabolic_process	4	0	1
GO:0032392_DNA_geometric_change	4	0	1
GO:0032508_DNA_duplex_unwinding	4	0	1
GO:0033554_cellular_response_to_stress	4	0	1
GO:0034220_transmembrane_ion_transport	4	0	1
GO:0042493_response_to_drug	4	0	1
GO:0045005_maintenance_of_fidelity_during_DNA-dependent_DNA_replication	4	0	1
GO:0048878_chemical_homeostasis	4	0	1
GO:0050801_ion_homeostasis	4	0	1
GO:0051246_regulation_of_protein_metabolic_process	4	0	1
GO:0051716_cellular_response_to_stimulus	4	0	1
GO:0055080_cation_homeostasis	4	0	1
GO:0055085_transmembrane_transport	4	0	1
GO:0006575_amino_acid_derivative_metabolic_process	10	0	1
GO:0006576_biogenic_amine_metabolic_process	10	0	1
GO:0009084_glutamine_family_amino_acid_biosynthetic_process	10	0	1
GO:0009116_nucleoside_metabolic_process	10	0	1
GO:0033692_cellular_polysaccharide_biosynthetic_process	10	0	1
GO:0044264_cellular_polysaccharide_metabolic_process	10	0	1
GO:0000103_sulfate_assimilation	1	0	1
GO:0000154_rRNA_modification	1	0	1
GO:0000302_response_to_reactive_oxygen_species	1	0	1
GO:0000738_DNA_catabolic_process__exonucleolytic	1	0	1
GO:0001680_tRNA_3'-terminal_CCA_addition	1	0	1
GO:0003008_system_process	1	0	1
GO:0006000_fructose_metabolic_process	1	0	1
GO:0006011_UDP-glucose_metabolic_process	1	0	1
GO:0006014_D-ribose_metabolic_process	1	0	1
GO:0006020_inositol_metabolic_process	1	0	1

GO:0006021_inositol_biosynthetic_process	1	0	1
GO:0006145_purine_base_catabolic_process	1	0	1
GO:0006146_adenine_catabolic_process	1	0	1
GO:0006183_GTP_biosynthetic_process	1	0	1
GO:0006228_UTP_biosynthetic_process	1	0	1
GO:0006231_dTMP_biosynthetic_process	1	0	1
GO:0006233_dTDP_biosynthetic_process	1	0	1
GO:0006235_dTTP_biosynthetic_process	1	0	1
GO:0006241_CTP_biosynthetic_process	1	0	1
GO:0006270_DNA_replication_initiation	1	0	1
GO:0006275_regulation_of_DNA_replication	1	0	1
GO:0006334_nucleosome_assembly	1	0	1
GO:0006338_chromatin_remodeling	1	0	1
GO:0006342_chromatin_silencing	1	0	1
GO:0006379_mRNA_cleavage	1	0	1
GO:0006397_mRNA_processing	1	0	1
GO:0006420_arginyl-tRNA_aminoacylation	1	0	1
GO:0006424_glutamyl-tRNA_aminoacylation	1	0	1
GO:0006426_glycyl-tRNA_aminoacylation	1	0	1
GO:0006427_histidyl-tRNA_aminoacylation	1	0	1
GO:0006428_isoleucyl-tRNA_aminoacylation	1	0	1
GO:0006429_leucyl-tRNA_aminoacylation	1	0	1
GO:0006430_lysyl-tRNA_aminoacylation	1	0	1
GO:0006431_methionyl-tRNA_aminoacylation	1	0	1
GO:0006434_seryl-tRNA_aminoacylation	1	0	1
GO:0006436_tryptophanyl-tRNA_aminoacylation	1	0	1
GO:0006437_tyrosyl-tRNA_aminoacylation	1	0	1
GO:0006438_valyl-tRNA_aminoacylation	1	0	1
GO:0006465_signal_peptide_processing	1	0	1
GO:0006473_protein_amino_acid_acetylation	1	0	1
GO:0006474_N-terminal_protein_amino_acid_acetylation	1	0	1
GO:0006476_protein_amino_acid_deacetylation	1	0	1
GO:0006479_protein_amino_acid_methylation	1	0	1
GO:0006518_peptide_metabolic_process	1	0	1
GO:0006527_arginine_catabolic_process	1	0	1
GO:0006534_cysteine_metabolic_process	1	0	1
GO:0006535_cysteine_biosynthetic_process_from_serine	1	0	1
GO:0006536_glutamate_metabolic_process	1	0	1
GO:0006537_glutamate_biosynthetic_process	1	0	1
GO:0006542_glutamine_biosynthetic_process	1	0	1
GO:0006543_glutamine_catabolic_process	1	0	1
GO:0006558_L-phenylalanine_metabolic_process	1	0	1
GO:0006570_tyrosine_metabolic_process	1	0	1
GO:0006571_tyrosine_biosynthetic_process	1	0	1
GO:0006591_ornithine_metabolic_process	1	0	1
GO:0006596_polyamine_biosynthetic_process	1	0	1
GO:0006598_polyamine_catabolic_process	1	0	1
GO:0006638_neutral_lipid_metabolic_process	1	0	1
GO:0006639_acylglycerol_metabolic_process	1	0	1

GO:0006742_NADP_catabolic_process	1	0	1
GO:0006746_FADH2_metabolic_process	1	0	1
GO:0006791_sulfur_utilization	1	0	1
GO:0006801_superoxide_metabolic_process	1	0	1
GO:0006808_regulation_of_nitrogen_utilization	1	0	1
GO:0006835_dicarboxylic_acid_transport	1	0	1
GO:0006873_cellular_ion_homeostasis	1	0	1
GO:0006879_cellular_iron_ion_homeostasis	1	0	1
GO:0006979_response_to_oxidative_stress	1	0	1
GO:0007156_homophilic_cell_adhesion	1	0	1
GO:0007242_intracellular_signaling_cascade	1	0	1
GO:0007264_small_GTPase_mediated_signal_transduction	1	0	1
GO:0007600_sensory_perception	1	0	1
GO:0007606_sensory_perception_of_chemical_stimulus	1	0	1
GO:0008213_protein_amino_acid_alkylation	1	0	1
GO:0008216_spermidine_metabolic_process	1	0	1
GO:0008612_peptidyl-lysine_modification_to_hypusine	1	0	1
GO:0008614_pyridoxine_metabolic_process	1	0	1
GO:0008615_pyridoxine_biosynthetic_process	1	0	1
GO:0009094_L-phenylalanine_biosynthetic_process	1	0	1
GO:0009111_vitamin_catabolic_process	1	0	1
GO:0009120_deoxyribonucleoside_metabolic_process	1	0	1
GO:0009132_nucleoside_diphosphate_metabolic_process	1	0	1
GO:0009133_nucleoside_diphosphate_biosynthetic_process	1	0	1
GO:0009138_pyrimidine_nucleoside_diphosphate_metabolic_process	1	0	1
GO:0009139_pyrimidine_nucleoside_diphosphate_biosynthetic_process	1	0	1
GO:0009148_pyrimidine_nucleoside_triphosphate_biosynthetic_process	1	0	1
GO:0009157_deoxyribonucleoside_monophosphate_biosynthetic_process	1	0	1
GO:0009162_deoxyribonucleoside_monophosphate_metabolic_process	1	0	1
GO:0009166_nucleotide_catabolic_process	1	0	1
GO:0009176_pyrimidine_deoxyribonucleoside_monophosphate_metabolic_process	1	0	1
GO:0009177_pyrimidine_deoxyribonucleoside_monophosphate_biosynthetic_process	1	0	1
GO:0009186_deoxyribonucleoside_diphosphate_metabolic_process	1	0	1
GO:0009189_deoxyribonucleoside_diphosphate_biosynthetic_process	1	0	1
GO:0009196_pyrimidine_deoxyribonucleoside_diphosphate_metabolic_process	1	0	1
GO:0009197_pyrimidine_deoxyribonucleoside_diphosphate_biosynthetic_process	1	0	1
GO:0009200_deoxyribonucleoside_triphosphate_metabolic_process	1	0	1
GO:0009202_deoxyribonucleoside_triphosphate_biosynthetic_process	1	0	1
GO:0009208_pyrimidine_ribonucleoside_triphosphate_metabolic_process	1	0	1
GO:0009209_pyrimidine_ribonucleoside_triphosphate_biosynthetic_process	1	0	1
GO:0009211_pyrimidine_deoxyribonucleoside_triphosphate_metabolic_process	1	0	1
GO:0009212_pyrimidine_deoxyribonucleoside_triphosphate_biosynthetic_process	1	0	1
GO:0009266_response_to_temperature_stimulus	1	0	1
GO:0009398_FMN_biosynthetic_process	1	0	1
GO:0009408_response_to_heat	1	0	1
GO:0009628_response_to_abiotic_stimulus	1	0	1
GO:0009636_response_to_toxin	1	0	1
GO:0009892_negative_regulation_of_metabolic_process	1	0	1
GO:0015693_magnesium_ion_transport	1	0	1

GO:0015694_mercury_ion_transport	1	0	1
GO:0015718_monocarboxylic_acid_transport	1	0	1
GO:0015725_gluconate_transport	1	0	1
GO:0015800_acidic_amino_acid_transport	1	0	1
GO:0015813_L-glutamate_transport	1	0	1
GO:0015904_tetracycline_transport	1	0	1
GO:0015936_coenzyme_A_metabolic_process	1	0	1
GO:0015976_carbon_utilization	1	0	1
GO:0015977_carbon_utilization_by_fixation_of_carbon_dioxide	1	0	1
GO:0015979_photosynthesis	1	0	1
GO:0015994_chlorophyll_metabolic_process	1	0	1
GO:0015995_chlorophyll_biosynthetic_process	1	0	1
GO:0016259_selenocysteine_metabolic_process	1	0	1
GO:0016260_selenocysteine_biosynthetic_process	1	0	1
GO:0016337_cell-cell_adhesion	1	0	1
GO:0016458_gene_silencing	1	0	1
GO:0016481_negative_regulation_of_transcription	1	0	1
GO:0016539_intein-mediated_protein_splicing	1	0	1
GO:0016568_chromatin_modification	1	0	1
GO:0017182_peptidyl-diphthamide_metabolic_process	1	0	1
GO:0017183_peptidyl-diphthamide_biosynthetic_process_from_peptidyl-histidine	1	0	1
GO:0018205_peptidyl-lysine_modification	1	0	1
GO:0018298_protein-chromophore_linkage	1	0	1
GO:0018409_peptide_or_protein_amino-terminal_blocking	1	0	1
GO:0019344_cysteine_biosynthetic_process	1	0	1
GO:0019357_nicotinate_nucleotide_biosynthetic_process	1	0	1
GO:0019364_pyridine_nucleotide_catabolic_process	1	0	1
GO:0019430_removal_of_superoxide_radicals	1	0	1
GO:0019740_nitrogen_utilization	1	0	1
GO:0019794_nonprotein_amino_acid_metabolic_process	1	0	1
GO:0030003_cellular_cation_homeostasis	1	0	1
GO:0030005_cellular_di-__tri-valent_inorganic_cation_homeostasis	1	0	1
GO:0030258_lipid_modification	1	0	1
GO:0030259_lipid_glycosylation	1	0	1
GO:0030908_protein_splicing	1	0	1
GO:0031324_negative_regulation_of_cellular_metabolic_process	1	0	1
GO:0031365_N-terminal_protein_amino_acid_modification	1	0	1
GO:0031507_heterochromatin_formation	1	0	1
GO:0032501_multicellular_organismal_process	1	0	1
GO:0042219_amino_acid_derivative_catabolic_process	1	0	1
GO:0042255_ribosome_assembly	1	0	1
GO:0042256_mature_ribosome_assembly	1	0	1
GO:0042262_DNA_protection	1	0	1
GO:0042365_water-soluble_vitamin_catabolic_process	1	0	1
GO:0042402_biogenic_amine_catabolic_process	1	0	1
GO:0042450_arginine_biosynthetic_process_via_ornithine	1	0	1
GO:0042542_response_to_hydrogen_peroxide	1	0	1
GO:0042743_hydrogen_peroxide_metabolic_process	1	0	1
GO:0042744_hydrogen_peroxide_catabolic_process	1	0	1

GO:0042779_removal_of_tRNA_3'-trailer_sequence	1	0	1
GO:0042822_pyridoxal_phosphate_metabolic_process	1	0	1
GO:0042823_pyridoxal_phosphate_biosynthetic_process	1	0	1
GO:0043543_protein_amino_acid_acylation	1	0	1
GO:0045226_extracellular_polysaccharide_biosynthetic_process	1	0	1
GO:0045814_negative_regulation_of_gene_expression__epigenetic	1	0	1
GO:0045892_negative_regulation_of_transcription__DNA-dependent	1	0	1
GO:0045934_negative_regulation_of_nucleobase__nucleoside__nucleotide_and_nucleic_acid_metabolic_process	1	0	1
GO:0046036_CTP_metabolic_process	1	0	1
GO:0046039_GTP_metabolic_process	1	0	1
GO:0046051_UTP_metabolic_process	1	0	1
GO:0046072_dTDP_metabolic_process	1	0	1
GO:0046073_dTMP_metabolic_process	1	0	1
GO:0046075_dTTP_metabolic_process	1	0	1
GO:0046083_adenine_metabolic_process	1	0	1
GO:0046113_nucleobase_catabolic_process	1	0	1
GO:0046125_pyrimidine_deoxyribonucleoside_metabolic_process	1	0	1
GO:0046203_spermidine_catabolic_process	1	0	1
GO:0046339_diacylglycerol_metabolic_process	1	0	1
GO:0046341_CDP-diacylglycerol_metabolic_process	1	0	1
GO:0046379_extracellular_polysaccharide_metabolic_process	1	0	1
GO:0046444_FMN_metabolic_process	1	0	1
GO:0046486_glycerolipid_metabolic_process	1	0	1
GO:0046497_nicotinate_nucleotide_metabolic_process	1	0	1
GO:0046516_hypusine_metabolic_process	1	0	1
GO:0046689_response_to_mercury_ion	1	0	1
GO:0046835_carbohydrate_phosphorylation	1	0	1
GO:0048519_negative_regulation_of_biological_process	1	0	1
GO:0048523_negative_regulation_of_cellular_process	1	0	1
GO:0050787_detoxification_of_mercury_ion	1	0	1
GO:0050877_neurological_system_process	1	0	1
GO:0050983_spermidine_catabolic_process_to_deoxyhypusine__using_deoxyhypusine_synthase	1	0	1
GO:0051052_regulation_of_DNA_metabolic_process	1	0	1
GO:0051171_regulation_of_nitrogen_compound_metabolic_process	1	0	1
GO:0051253_negative_regulation_of_RNA_metabolic_process	1	0	1
GO:0051478_mannosylglycerate_metabolic_process	1	0	1
GO:0051479_mannosylglycerate_biosynthetic_process	1	0	1
GO:0055066_di-__tri-valent_inorganic_cation_homeostasis	1	0	1
GO:0055072_iron_ion_homeostasis	1	0	1
GO:0055082_cellular_chemical_homeostasis	1	0	1
GO:0000160_two-component_signal_transduction_system_(phosphorelay)	41	0	1
GO:0000271_polysaccharide_biosynthetic_process	15	0	1
GO:0005976_polysaccharide_metabolic_process	15	0	1
GO:0006261_DNA-dependent_DNA_replication	13	0	1
GO:0006304_DNA_modification	15	0	1
GO:0006457_protein_folding	12	0	1
GO:0006464_protein_modification_process	50	0	1
GO:0006468_protein_amino_acid_phosphorylation	36	0	1
GO:0006508_proteolysis	25	0	1

GO:0007154_cell_communication	64	0	1
GO:0007165_signal_transduction	60	0	1
GO:0008150_biological_process	973	80	1
GO:0009064_glutamine_family_amino_acid_metabolic_process	17	0	1
GO:0009072_aromatic_amino_acid_family_metabolic_process	17	0	1
GO:0009073_aromatic_amino_acid_family_biosynthetic_process	14	0	1
GO:0018106_peptidyl-histidine_phosphorylation	31	0	1
GO:0018193_peptidyl-amino_acid_modification	34	0	1
GO:0018202_peptidyl-histidine_modification	32	0	1
GO:0019438_aromatic_compound_biosynthetic_process	19	0	1
GO:0042221_response_to_chemical_stimulus	13	0	1
GO:0042592_homeostatic_process	12	0	1
GO:0043284_biopolymer_biosynthetic_process	19	0	1
GO:0043285_biopolymer_catabolic_process	12	0	1
GO:0043648_dicarboxylic_acid_metabolic_process	17	0	1
GO:0043687_post-translational_protein_modification	41	0	1
GO:0046417_chorismate_metabolic_process	14	0	1
GO:0065008_regulation_of_biological_quality	14	0	1
GO:0000272_polysaccharide_catabolic_process	3	0	1
GO:0006144_purine_base_metabolic_process	3	0	1
GO:0006269_DNA_replication__synthesis_of_RNA_primer	3	0	1
GO:0006432_phenylalanyl-tRNA_aminoacylation	3	0	1
GO:0006528_asparagine_metabolic_process	3	0	1
GO:0006529_asparagine_biosynthetic_process	3	0	1
GO:0006553_lysine_metabolic_process	3	0	1
GO:0006825_copper_ion_transport	3	0	1
GO:0006855_multidrug_transport	3	0	1
GO:0006885_regulation_of_pH	3	0	1
GO:0008616_queuosine_biosynthetic_process	3	0	1
GO:0008617_guanosine_metabolic_process	3	0	1
GO:0008618_7-methylguanosine_metabolic_process	3	0	1
GO:0009085_lysine_biosynthetic_process	3	0	1
GO:0009089_lysine_biosynthetic_process_via_diaminopimelate	3	0	1
GO:0009119_ribonucleoside_metabolic_process	3	0	1
GO:0009163_nucleoside_biosynthetic_process	3	0	1
GO:0009310_amine_catabolic_process	3	0	1
GO:0009432_SOS_response	3	0	1
GO:0015074_DNA_integration	3	0	1
GO:0015689_molybdate_ion_transport	3	0	1
GO:0016071_mRNA_metabolic_process	3	0	1
GO:0019748_secondary_metabolic_process	3	0	1
GO:0042278_purine_nucleoside_metabolic_process	3	0	1
GO:0042440_pigment_metabolic_process	3	0	1
GO:0042451_purine_nucleoside_biosynthetic_process	3	0	1
GO:0042455_ribonucleoside_biosynthetic_process	3	0	1
GO:0043413_biopolymer_glycosylation	3	0	1
GO:0044270_nitrogen_compound_catabolic_process	3	0	1
GO:0046114_guanosine_biosynthetic_process	3	0	1
GO:0046116_queuosine_metabolic_process	3	0	1

GO:0046118_7-methylguanosine_biosynthetic_process	3	0	1
GO:0046128_purine_ribonucleoside_metabolic_process	3	0	1
GO:0046129_purine_ribonucleoside_biosynthetic_process	3	0	1
GO:0046148_pigment_biosynthetic_process	3	0	1
GO:0046451_diaminopimelate_metabolic_process	3	0	1
GO:0051258_protein_polymerization	3	0	1
GO:0051276_chromosome_organization_and_biogenesis	3	0	1
GO:0055067_monovalent_inorganic_cation_homeostasis	3	0	1
GO:0006568_tryptophan_metabolic_process	8	0	1
GO:0006586_indolalkylamine_metabolic_process	8	0	1
GO:0006813_potassium_ion_transport	8	0	1
GO:0009144_purine_nucleoside_triphosphate_metabolic_process	8	0	1
GO:0009145_purine_nucleoside_triphosphate_biosynthetic_process	8	0	1
GO:0009199_ribonucleoside_triphosphate_metabolic_process	8	0	1
GO:0009201_ribonucleoside_triphosphate_biosynthetic_process	8	0	1
GO:0009205_purine_ribonucleoside_triphosphate_metabolic_process	8	0	1
GO:0009206_purine_ribonucleoside_triphosphate_biosynthetic_process	8	0	1
GO:0019363_pyridine_nucleotide_biosynthetic_process	8	0	1
GO:0042254_ribosome_biogenesis_and_assembly	8	0	1
GO:0042430_indole_and_derivative_metabolic_process	8	0	1
GO:0042434_indole_derivative_metabolic_process	8	0	1
GO:0045454_cell_redox_homeostasis	8	0	1

FINAL LIBRARY COPY