Functional analysis of *tilS* homologues in Bacteria and Archaea

by

Ruth Stavrum

Thesis submitted for the MSc in Biology

August 2005

Department of Biology

University of Bergen

Acknowledgements

The work for this Master's thesis was performed at the Department of Biology, University of Bergen between February 2003 and August 2005.

First I would like to thank my supervisor, professor dr philos. Nile-Kåre Birkeland for giving me an interesting and challenging project. Your excellent guidance in both the practical and theoretical aspects of this project has given me the knowledge and confidence I need to continue working with science. I would also like to thank dr. scient Ida Helene Steen and PhD student Gyri T. Haugland for showing interest in my project. Your help has been invaluable in particularly the latter part of my project. Also I would like to thank department engineer Marit Madsen and the rest of the laboratory group for all your help. The friendly environment has made the days in the lab a positive experience.

Finally I would like to thank friends and family for all your support during this project.

Ruth Stavrum

August, 2005

Summary

Out of all of the microorganisms sequenced so far about one third of the genes have unknown function. Several studies have shown that information on evolutionary relationships between unknown genes can aid in the prediction of the function of these genes. As a result of all this new information, new methods of identification have been established which sort genes based on sequence similarity between both paralogues and orthologues genes. The Clusters of Orthologous Groups (COG) is one such method, which group genes on the basis of sequence similarity where all groups containing at least three proteins from distant genomes are assumed to belong to the same orthologous group. At the start of this project all, but one of the COGs containing universal genes had been assigned a function. In the last group, COG0037, the genes had been suggested to be ATPases based on conserved motifs. In September, 2003 an article was published where the function of one of the *Escherichia coli* members in COG0037, TilS (previously known as YaeN), had been determined and TilS was shown to be an RNA modification enzyme. Based on this knowledge it was decided to analyse TilS orthologues for similar function. In this work four TilS homologues; AF1595 and AF1321 from *Archaeoglobus fulgidus*, Sso0586 from *Sulfolobus solfataricus* and YdaO from *E. coli* were expressed in *Escherichia coli*, and tested for solubility. Out of the four proteins only one, YdaO proved to be soluble. This protein was

purified by affinity chromatography and analysed further and shown to exhibit ATPase activity and ability to autophosphorylate. An attempt to verify whether this protein is expressed under normal conditions was unsuccessful. An attempt to determine whether the gene coding for YdaO could be used as a signature gene for *E. coli* was also unsuccessful.

Contents

Chapter 1

Introduction

1.1 Genomics

Since the mid-1940s scientists have used a variety of genetic techniques to study the genes in microorganisms. These techniques involved the mapping of the genes as well as studying their function and regulation. The first genome to be sequenced was the 3569-nucleotide RNA genome of the virus MS2 in 1976 [1] and the first DNA genome to be sequenced was the 5386-nucleotide sequence of the virus ϕX174 in 1977 by a group led by Frederick Sanger [2]. In 1992 The Institute for Genomic Research (TIGR) was established whose primary research interests are in structural, functional and comparative analysis of genomes products from a number of organisms including viruses, eubacteria, archaea and eukaryotes[1]. Since then other organisations, such as The Sanger Institute [2] have become involved in genome sequencing. In 1995 TIGR published for the first time the genome sequence for an entire organism *Haemophilus influenzae* [3] and as of this date there are 1496 ongo-

[1] http://www.tigr.org
[2] http://www.sanger.ac.uk/

ing sequencing projects[3]. Included in the first completed microbial projects in addition to *H. influenzae* were *Escherichia coli* [4], *Bacillus subtilis* [5], thermophilic bacteria, 5 representatives of the *Archaea* domain and the first eukaryotic organism, *Saccharomyces cerevisiae* [6]. Out of all of the microorganisms sequenced so far about one third of the genes have unknown function. There is also some degree of uncertainty about several of the genes where the function have been inferred by the use of homology [7]. Several studies have shown that information on evolutionary relationships between unknown genes can aid in the prediction of the function of these genes. As a result of all this new information, new methods of identification have been established which sort genes based on sequence similarity between both paralogues and orthologues genes [8]. One method developed by Tatusov *et. al.* (1997) called Clusters of Orthologous Groups (COG) group genes on the basis of sequence similarity where all groups containing at least three proteins from distant genomes are assumed to belong to the same orthologous group [9]. Orthology (described in figure 1.1) is defined as two genes from two different species that derive from a single gene in the last common ancestor of the species [10, 11]

1.2 Clusters of Orthologous Groups

As of today there are 14059 groups in the COG database including 104101 proteins from 43 genomes of bacteria, archaea and fungi[4]. Most of the orthologous groups have representatives from all the genomes and contain one or more experimentally characterised functions. However, a study done by Harris *et al.*, 2003 shows that out of 3100 COGs analysed only 80 were found

[3]http://www.genomesonline.org/
[4]http://www.ncbi.nlm.nih.gov/COG/

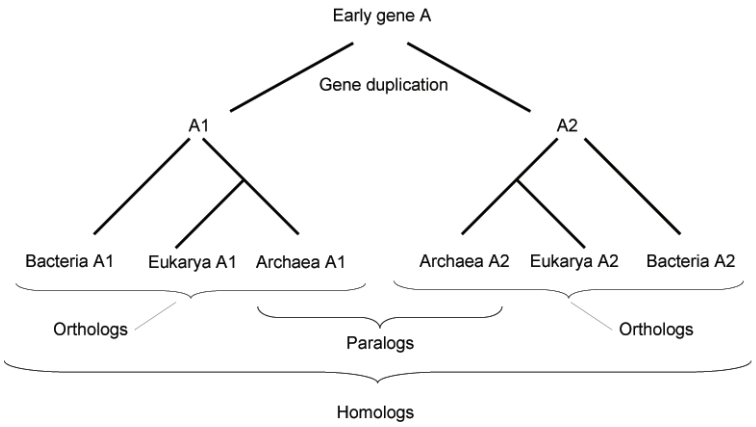

Figure 1.1: Orthologs and Paralogs are two types of homologous sequences. Orthology describes genes in different species that derive from a common ancestor. Orthologous genes may or may not have the same function. Paralogy describes homologous genes within a single species that diverged by gene duplication.

to occur in all organisms and 50 of these shared the same phylogenetic relationships as rRNA [12]. In some of the COGs the genes are grouped based on conserved motifs. One such group is COG0037, which includes 77 genes from all eubacterial, eukaryotic and archaebacterial completed genomes. Most of them appear to have one or two universal genes belonging to this group, implicating an important function (table 1.1). A phylogenetic analysis by Stranden (2002) showed five highly conserved regions between the members of COG0037. Although the members of the COGs are grouped based on sequence similarity 10 sequences showed poor alignment with three or more of the conserved regions and was therefore excluded from further analysis. An NJ-bootstrap tree (figure 1.2) was constructed based on a multiple alignment of the remaining 67 sequences and shows three main phylogenetic groups; Bacteria (subdivided into Bacteria I and Bacteria II), Eukarya and Archaea.

One method of determining the importance of a gene is to disrupt the open reading frame (ORF) and see whether the organism is still viable. This was done for *Saccharomyces cerevisiae* [14], *Mycoplasma genitalium*/*Mycoplasma pneumoniae* [15] and *E. coli* [13] represented in COG0037 with the ORF's

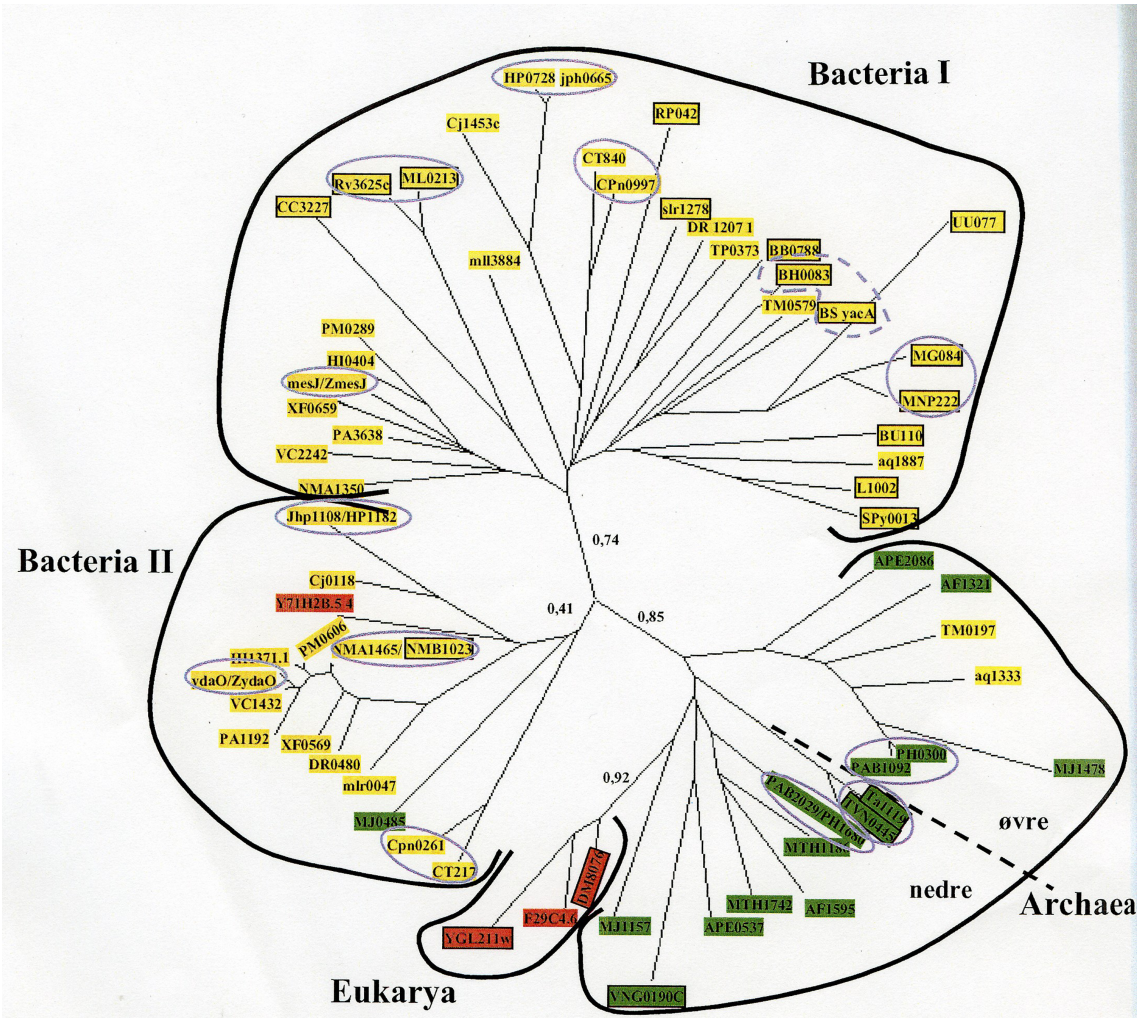

Figure 1.2: NJ bootstrap tree for the alignment COG0037^{-10} constructed in ClustalX. Positions with gap in the alignment are excluded and the possibility of multiple substitutions have been accounted for. The numbers in bold are the bootstrap values for the main branches. The open reading frames (ORFs) have been designated different colours depending on the domain the organisms belong to. Yellow represent Bacteria, green Archaea and red Eukarya. The ORFs surrounded by a black box descend from organisms represented with only one open reading frames within COG0037. *ydaO* and MesJ are circled in blue. [13]

YGL211w, *MG084/MNP222* and *yaeN/ydaO*, respectively. In the three studies the disruption of the genes *YGL211w*, *MG084/MNP222* and *yaeN* was not possible, indicating that the genes are essential. A deletion in the gene *ydaO* did not seem to affect the growth of the organisms suggesting that this gene is not essential in *E. coli*.

The proteins in COG0037 were assumed to be ATPases belonging to

the PP-loop superfamily taking part in the control of the cell cycle. This assumption was based on an unpublished comment in GeneBank file GB-BCT:ECMESJORF, accession number: Z50870, "Identification of a new *Escherichia coli* cell cycle gene" by Pichoff and Bouche, 1995 [16]. However, through personal contact with Sebastein Pichoff we have learned that this citation is wrong and that although several attempts have been made to correct this mistake, the annotation still lingers in the databases (GeneBank, EMBL, DDBJ, COG and Swiss-Prot). The function of the proteins in this COG remained unknown until 2003, when Soma *et al.* proved that the *E. coli* *yaeN* gene codes for an tRNA modification enzyme with ATPase activity.

1.3 tRNA

There are four types of RNA; tRNA, mRNA, rRNA and snRNA. Translation of the genetic information requires decoding of the codons (a triplet made up of three nucleic acids) by direct interaction between the tRNA and the nucleic acids on the mRNA. There are more than 20 different tRNA molecules, which all have between 74 and 95 nucleotides [17]. As shown in figure 1.3 there are 4 arms in tRNA: the acceptor, D, T pseudouridine C (TΨC) and anticodon. Occasionally tRNA molecules have an extra or "variable loop". The acceptor stem and TΨC-arm stack to each other to form a continuous α-helix, while the D-arm and anticodon arm stack to form another continuous helix. The characteristically L-shape tertiary structure (figure 1.4) is formed by a 90°cross between the two RNA double helices [18]. There is no physical force that specifically pair an amino acid with the anticodon. Instead the anticodon forms a hydrogen bond with the codon in a base-specific manner which allows the aminoacyl-tRNA synthetases to pair the cognate

tRNA and amino acids in a correct combination. The tRNA is first transcribed as a precursor with 5' and 3' extension. Then many modifications are incorporated into various portions (mainly the core formed by the D-loop and TΨC-loop, and the anticodon loop) of this unmodified transcript by a number of modification enzymes which results in the mature tRNA. Finally, aminoacyl- tRNA synthetases specifically recognize the mature tRNA, and ligate the corresponding amino acid onto the 5'-CCA terminus [19].

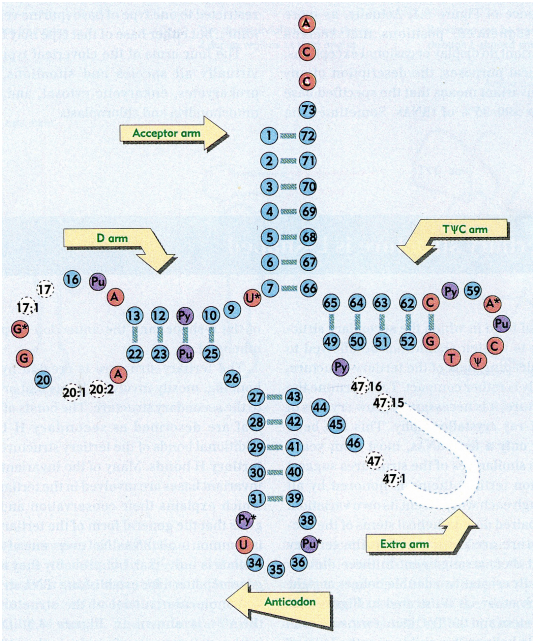

Figure 1.3: Overview of tRNA showing the codon/anticodon binding site,the D- and TΨC loops, the discriminator and the 3'CCA acceptor site [17].

1.3.1 tRNA^{ile}

In *E. coli* there are two forms of isoleucine tRNA; $tRNA^{ile}_{major}$ ($tRNA^{ile}_{1}$), which recognises the codons AUU and AUC, and $tRNA^{ile}_{minor}$ ($tRNA^{ile}_{2}$) which recognises AUA only [20]. The *E. coli* $tRNA^{ile}_{2}$, which is essential for the specific recognition of the codon AUA is encoded by the gene *ileX* [21]. This

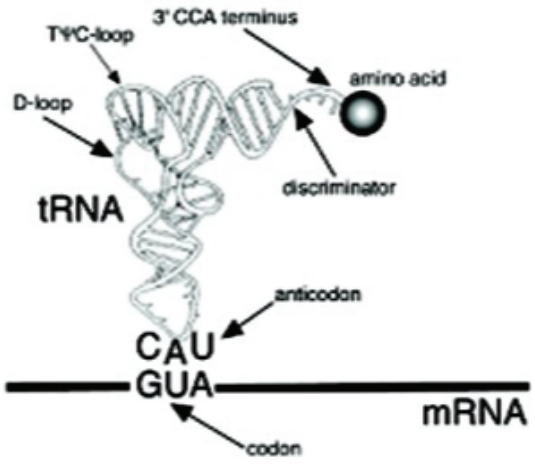

Figure 1.4: Overview of the tertiary structure of tRNA formed by the 90°cross between two RNA double helices. [19].

tRNA is unusual in that it contains a CAU anticodon which normally is the anticodon for methionine (AUG). Muramatsu *et. al.* (1988) proved that the $tRNA^{ile}_{minor}$ contained a modified nucleoside (N+) in the first position of the anticodon (N+34) [20]. In the same study this modified nucleoside was shown to be a cytidine derivative with lysine moiety. The mechanism involved in this modification remained unknown until Soma *et. al.* (2003) identified the enzyme responsible for both the amino acid and codon specificity modifications [22]. This enzyme, TilS (also called $tRNA^{ile}$-lysidine synthetase, coded for by *tilS*, previously known as *yaeN*) catalyses the formation of k^2C34 exclusively in the CAU-containing precursor $tRNA^{ile}$ and was identified in both *E. coli* and *B. subtilis* [22]. Figure 1.5 describes the tRNA-modification performed by this enzyme.

In *E. coli* the AUA codon is used at a frequency of less than 0.4% ([24] and references therein) and is the least frequently used codon after AGG and CTA [25]. Interestingly, a codon usage table created from the international DNA sequence databases [26] shows that in the pathogenic Entero Aggregative *E. coli* (EAEC) strain the AUA codon usage is 1.85%, which is nearly 5

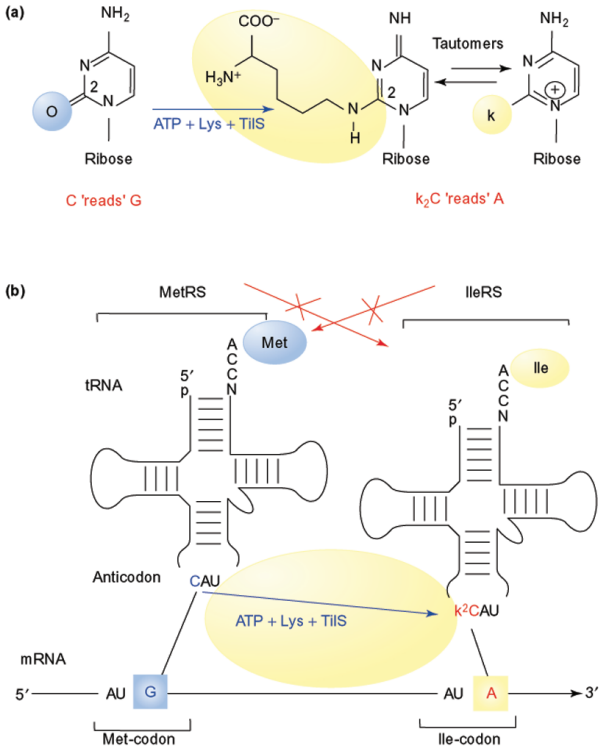

Figure 1.5: Structure of lysidine in tRNA and its implication on tRNA aminoacylation and codon recognition. (a) Enzymatic conversion of cytidine to lysidine[(4-amino)-2-(N^6 -lysino)-1-(D-ribofuranosyl)pyrimidine]. Lysidine can exist in tautomeric forms (K refers to lysine moiety). k^2C in the first wobble position of anticodon can only pair with A in the third position of the codon. (b) The enzymatic C34-to-k^2C34 conversion in the minor *E. coli* $tRNA^{ile}$ (k^2CAU) has two important consequences: (i) it inhibits the MetRS (methionyl-tRNA synthetase) to bind Met, allowing the IleRS (Isoleucyl-tRNA synthetase) to bind Ile instead. (ii) it prohibits the recognition of G and, instead, allows the recognition of A as the third base of the codon during translation on the ribosome. Figure is from [23]

times as much as the non-pathogenic *E. coli* K-12 reference strain. Also the Shiga-toxin VT2-encoding bacteriophage phage EH297 which integrates into the Entero Hemorrhagic *E. coli* (EHEC) genome causing hemolytic colitis (HC) or hemolytic uremic syndrome (HUS) [27] has an AUA codon usage of 1.36%, which is more than three times as much as the *E. coli* K-12 reference strain. Another interesting factor is that although the *E. coli* O157:H7 strain carries the structural genes for both the Shiga toxin 1 (*stx1*) and Shiga toxin 2 (*stx2*) only the *stx2* gene is known to also carry the nucleotide sequence for the *ileX* gene upstream of *stx2*. The *stx1* gene not only has a much lower

AUA codon usage than *stx2* (0.8%), but also lack the *ileX* segment [28]. The reason for this is unclear, although the presence of the tRNA gene upstream of the *stx2* gene may provide a suitable sequence for integration by homologous recombination ensuring proper expression of the virulence genes. A similar observation has been made for the $tRNA^{arg}_{AGA}$ sequence in *stx2A* gene where all the genes from the different *E. coli* strains contained the sequence for this rare tRNA [29]. Also in the uropathogenic *E. coli* strain 536 a similar thing has been reported where the pathogenicity island II (PAI II) is integrated into the tRNA gene *leuX* [30, 31, 32], which controls the expression of type 1 fimbria production and flagellar formation [33, 34, 35]. In an analysis of the codon usage in the *stx2A* gene in further detail it was found that codons which are rarely used in *E. coli* K12, such as AGA for arginine, ACA for threonine, AAU for asparagine and UCA for serine, are also used frequently [29]. A schematic presentation of the codon usage for *E. coli* and a few other organisms based on PCA (Principal Component Analysis) can be found in figure 1.6.

1.4 Sequence and structure analysis of *E. coli* TilS

TilS is a 48.5 kDa protein belonging to the mesJ protein family, which consists entirely of eubacterial orthologs [22]. It forms a homodimer with each subunit consisting of the N-terminal dinucleotide-binding fold domain (NTD) and the C-terminal globular domain (CTD) connected by a long α-helical linker (see figure 1.7) [18]. A sequence alignment performed by Soma *et al.* (2003) revealed two highly conserved motifs. The first motif, SGGXDS (X being any hydrophobic amino acid) contains a conserved P loop motif in the N-

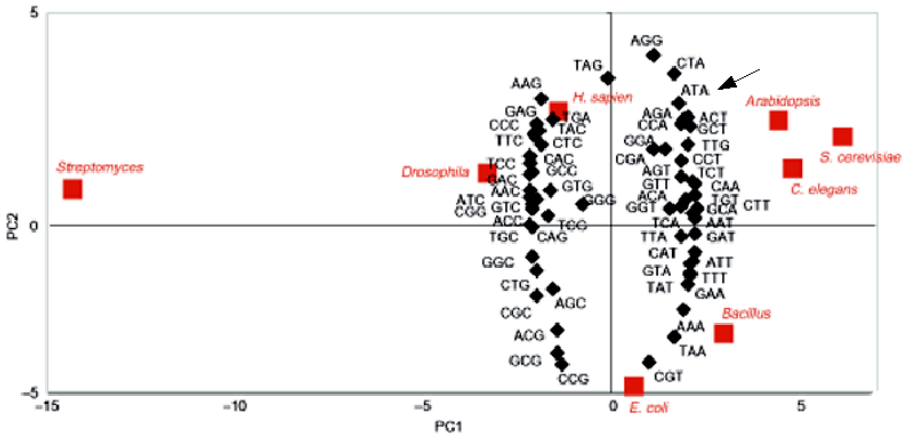

Figure 1.6: Graphical representation of codon usage space. Principal component analysis (PCA) involves a mathematical procedure that transforms several correlated variables (here codon frequencies) into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The frequencies with which each codon is used in all proteins of eight commonly studied organisms were tabulated (8 rows/organisms 62 columns/codons) and subjected to PCA to produce a map of codon usage space. The two codons ATG and TGG that uniquely encode Met and Trp, respectively, have been omitted. Two dimensions were identified that accounted for 70% (PC1) and 12% (PC2) of the total codon variability information, respectively. The black diamonds represent the loads (i.e. the contribution of each codon to the two principal component dimensions; e.g. codons GAT and CAG contribute nothing to PC2 but have approximately equal negative and positive contributions to PC1). The values of the codon loads have been normalized to that of the organism distribution. The red squares show the preferences of each organism plotted within this space. The further away from an organism a codon is plotted, the less frequent that codon appears in the organism. Figure from [25] modified in OpenOffice.org.

terminal region known to participate in the binding and hydrolysis of the $\alpha-\beta$ phosphate bond by ATP pyrophosphatase (PPi synthetase) family [36]. The second conserved motif contains residue R160 which is predicted to interact with the γ phosphate of ATP. In the non-conserved C-terminal region there are two CTD domains (CTD1 and CTD2) [22].

1.5 Aims

The genes in COG0037 are grouped based on sequence similarity. *E. coli* is represented with two open reading frames in this group, *yaeN* and *ydaO* and studies by Stranden (2002) indicates that the gene *yaeN* is essential to

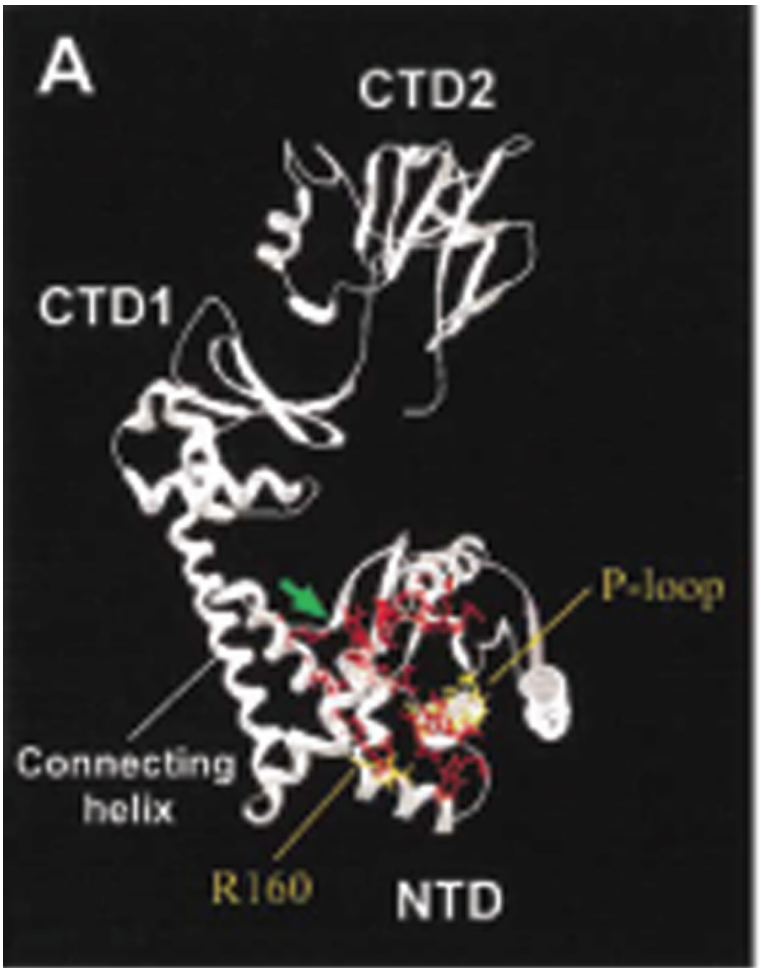

Figure 1.7: A ribbon diagram of *E. coli* TilS (mesJ) protein showing the conserved NTD domain (red), the P loop and R160 residues (yellow), the connecting helix and CTD1 and CTD2 domains. Figure from [22].

E. coli. As the genes in this COG group are grouped based on predicted functions it is of interest to learn more about the function of these proteins.

The aims of this project was divided into two parts:

1. Construction of a conditional mutant in order to assess the implications of up- and down regulation of the transcription of *yaeN* in *E. coli*.

2. Analysis of MesJ paralogs in *E. coli* and Archaea.

Several methods for controlling transcription have been developed such as insertion inactivation by transposons [37] or different substitution plasmids [38, 39, 40]. The method of choice in order to construct a conditional mutant was to use the plasmid pKO3 as a substitution vector as this vector allows site-specific deletions and insertions into the *E. coli* chromosome without causing polar effects [38]. After insertion of a deleted version of the gene into the *E. coli* chromosome by homologous recombination, the organism would be transformed with the expression vector pBAD containing a functional version of *yaeN*. The promotor for transcription of the inserted gene in this in this expression vector is tightly controlled by the amount of arabinose present in the medium, and hence the regulation of transcription would be controlled.

The second object of this project was to look at *yaeN* homologs in both *E. coli* and Archaea in order to detect conserved regions which would be functionally important for the proteins. The *yaeN* homologues chosen for this work were *AF1595* and *AF1321* from *Archaeoglobus fulgidus*, *Sso0586* from *Solfulobus solfataricus* and *ydaO* from *E. coli*. Both *E. coli* and *A. fulgidus* are represented with two open reading frames in COG0037, whereas *S. solfataricus* is represented with only one. Since *S. solfataricus* only has one homologue it was of interest to see whether this gene is transcribed and if so, whether it is functional. As for *A. fulgidus* and *E. coli*, which have two homologues each it would be interesting to see if both the genes were transcribed and if their proteins were functional.

Table 1.1: List of organisms in COG0037 and the open reading frames they are represented with.

Organisms	Open reading frames	Organisms	Open reading frames
Aeropyrum pernix	APE0537, APE1799 and APE2086	*Bacillus cereus*	BC4247
Agrobacterium tumefaciens	AGRC3410, AGRL2252 and Atu3711	*Bacillus halodurans*	BH0083
Aquifex aeolicus	aq1333 and aq1887	*Bacillus subtilis*	Bsu0067
Archaeoglobus fulgidus	**AF1321 and AF1595**	*Bordetella bronchiseptica*	BB1738
Bacillus anthracis	BA0652, BA4922 and BA4475	*Bordetella parapertussis*	BPP4146
Borrelia burgdorferi	BB0788	*Pseudomonas putida KT2440*	PP1608
Bradyrhizobium japonicum	bll7147	*Pseudomonas syringae*	PSPTO1551
Brucella melitensis	BMEI0342 and BMEI1132	*Pyrobaculum aerophilum*	PAE2295 and PAE3646
Brucella suis 1330	BR0831	*Pyrococcus abyssi*	PAB2029, PAB1992, PAB1092 and PAB1152
Buchnera aphidicola Sg	BUsg103	*Pyrococcus furiosus*	PF1758
Buchnera sp	BU110	*Pyrococcus horikoshii*	PH0300, PH1608, PH1680 and PH1968
Campylobacter jejuni	Cj0118, Cj1316c, Cj1324 and Cj1453c	*Ralstonia solanacearum*	RS01046 and RSc1170
Caulobacter crescentus	CC3227	*Rhodopseudomonas palustris CGA009*	RPA1125
Chlamydia muridarum	TC0228	*Rickettsia conorii*	RC0067
Chlamydia trachomatis	CT217 and CT840	*Rickettsia prowazekii*	RP042
Chlamydophila caviae	CCA00764	*Salmonella typhi*	STY1412 and t1558
Chlamydophila pneumoniae	CP0858, CPn0261, CPn0997, CPj0997 and CpB1035	*Salmonella typhimurium*	STM0236 and STM1654
Chromobacterium violaceum	CV4131	*Shewanella oneidensis*	SO2354
Clostridium acetobutylicum	CAC3204	*Shigella flexneri*	SF1825 and S1448
Clostridium perfringens	CPE2472	*Sinorhizobium meliloti*	SMc00534 and SMc02940
Corynebacterium efficiens YS-314	CE2544	*Staphylococcus aureus*	SAV0509, MW0464 and SA0467
Corynebacterium glutamicum	NCgl2605	*Staphylococcus epidermidis*	SE2274
Deinococcus radiodurans	DR0480	*Streptococcus pneumoniae*	spr0010 and SP0011
Escherichia coli K12	**mesJ and ydaO**	*Streptococcus pyogenes*	SPy0013
Escherichia coli O157H7	ZmesJ, ZydaO, ECs0190 and ECs1928	***Sulfolobus solfataricus***	**SSO0586**
Fusobacterium nucleatum	FN1679, FN1977 and FN0868	*Synechocystis PCC6803*	slr1278
Haemophilus ducreyi 35000HP	HD1716	*Thermoplasma acidophilum*	Ta1119
Haemophilus influenzae	HI0404 and HI1371.1	*Thermoplasma volcanium*	TVN0445
Halobacterium sp	VNG0190C	*Thermotoga maritima*	TM0197 and TM0579
Helicobacter hepaticus	HH1612	*Treponema pallidum*	TP0373
Helicobacter pylori	HP0728, HP1182, jhp0665 and jhp1108	*Tropheryma whipplei*	TW171 and TWT590
Lactococcus lactis	L1002	*Ureaplasma urealyticum*	UU077
Mesorhizobium loti	mlr0047 and mll3884	*Vibrio cholerae*	VC1432 and VC2242
Methanobacterium thermoautotrophicum	MTH1186 and MTH1742	*Xylella fastidiosa*	XF0569, XF0659 and PD1514
Methanococcus jannaschii	MJ0485, MJ1016, MJ1157, MJ1478 and MJ1599	*Yersinia pestis*	YPO1062, YPO2335 and y1998
Mycobacterium tuberculosis	MT3727 and Rv3625c	*Neisseria meningitidis*	NMB1023, NMA1350 and NMA1465
Mycoplasma genitalium	MG084	*Nostoc sp*	all0932
Mycoplasma penetrans	MYPE1460 and MYPE5070	*Pasteurella multocida*	PM0289 and PM0606
Mycoplasma pneumoniae	MPN222	*Photorhabdus luminescens*	plu2575
Mycoplasma pulmonis	MYPU7800	*Pseudomonas aeruginosa*	PA1192, PA3150 and PA3638
Methanopyrus kandleri	MK0051, MK0144, MK0756, MK1054, MK1103, MK1691 and MK1692	*Mycobacterium bovis*	Mb3649c
Methanosarcina acetivorans	MA1029, MA1974, MA3614 and MA3716	*Mycobacterium leprae*	ML0213
Methanosarcina mazei	MM0613		

Chapter 2

Materials and methods

2.1 Organisms used in this project

2.1.1 *Escherichia coli*

E. coli was first isolated in 1885 by Theodor Escherich, a German bacteriologist, as a microorganism normally found in the digestive tract. It is known as the best known prokaryote and the structure and function of *E. coli* has been used as the archetype of all living organisms. The reason why it is so well-known is that with its short generation time and simple nutrition requirements, it is very easy to grow in laboratories. *E. coli* is 1.1 - 1.5 µm wide x 2 - 6 µm long and has a generation time of 21 min at 40°C. Another advantage with *E. coli* is that since it is haploid it is easy to identify different mutations. Even though the *E. coli* genome was sequenced in 1997 and it has been studied quite intensely for 50 years, there is still a lot we do not know about this organism; about 30% of the 4406 open reading frames in *E. coli* still have unknown function [41].

2.1.2 *Archaeoglobus fulgidus*

Archaeoglobus fulgidus is a strict anaerobic and hyperthermophilic organism found in hydrothermal vents and oil reservoars. With its 2,178,400 base pairs it was the first sulphate-reducing organism to have its genome sequence determined([42] and references therein). Out of 2,436 open reading frames (ORFs) 651 ORFs encodes functionally uncharacterized yet conserved proteins. *A. fulgidus* VC-16 is the type strain of the *Archaeoglobales*. The cells are irregular spheres with a glycoprotein envelope and monopolar flagella. Although optimum growth temperature is at 83°C growth also occurs between 60 and 95°C. Minimum division time at 83°C is 4 hours. The organism grows organoheterotrophically using a variety of carbon and energy sources, but can grow lithoautotrophically on hydrogen, thiosulphate and carbon dioxide.

2.1.3 *Sulfolobus solfataricus*

The genome of *Sulfolobus solfataricus* was first sequenced in 2001 and contains 2992245 base pairs and 2994 protein coding genes [43]. *S. solfataricus* is a sulphur-oxidising chrenarchaeon which grows optimally at 80°C and pH 2-4. It is the most widely studied organisms of the chrenarchaeal branch of the Archaea serving as a model organism for research on mechanisms of DNA replication, cell cycle, chromosomal integration, transcription, RNA processing and translation ([43] and references therein).

2.1.4 Strains and their respective open reading frames used in this work

Organisms	Open reading frames
Escherichia coli K-12 MG1655	*mesJ* and *ydaO*
Archaeoglobus fulgidus VC16	*AF1595* and *AF1321*
Solfulobus solfataricus	*Sso0586*

E. coli TOP10 was used for cloning and sequencing.
E. coli Strain BL21-CodonPluss[DE3]-RIL was used for expression of recombinant proteins.

2.2 Culturing of the strains

Aseptic technique was applied at all stages when working with the different microorganisms.

Liquid cultures were incubated on a shaking platform.
Media used for the culturing of the different *Escherichia coli* strains were:

LA-medium (Luria-Bertani agar medium): 10% tryptone, 0.5% yeast extract, 10% NaCl, 15% agar
LB-medium (Luria-Bertani medium): 10% tryptone, 0.5% yeast extract, 10% NaCl
Low salt LA/LB medium with 5% sucrose: prepared as LA/LB, but with 0.5% NaCl
SOC medium: 2% tryptone, 0.5% yeast extract, 0.5% NaCl, 2.5M KCl,

10mM $MgCl_2$, 20mM glucose

A. fulgidus was cultured in AG medium (0.01% glucose, 0.015% glycerol, 0.5% peptone, 0.5% yeast extract, 0.2& malt extract, 0.7% $CaCO_3$).

DNA/RNA from *S. solfataricus* was kindly donated by Melanie Jonuscheit

2.3 Isolation of chromosomal DNA

The isolation of chromosomal DNA requires several steps such as culturing, harvesting, cell lysis, separation of DNA from other cell components and concentration of DNA. The following protocol uses phenol/chloroform for the extraction of the DNA.

- 3 ml of an overnight culture was centrifuged at 13000xg for 1 min.
- The supernatant was discarded and the pellet resuspended in 0.5 ml TE buffer pH 8.0 and 0.5 ml equilibrated phenol.
- The sample was mixed gently for 5 min and centrifuged for 1 min at 13000xg.
- The liquid phase was transferred to a new tube and added an equal volume of chloroform.
- The sample was mixed gently and centrifuged for 1 min at 13000xg before transferring the liquid phase to a new tube containing 0.5 μl 3M pH 4.6 NaAc, 5 μl 5NaCl and 0.7 volume isopropanol.
- The mixture was gently agitated and centrifuged for 15 min at 13000xg.
- The pellet was then washed twice with 70% ethanol and air-dried before being resuspended in 50 μl TE pH 8.0 or MilliQ water.

2.4 RNA isolation

The RNA was isolated from a culture when the cells were in the exponential phase. For the isolation the SV Total RNA Isolation System (Promega) kit was used according to the manufacturers instructions.

2.5 Polymerase chain reaction

Polymerase chain reaction (PCR) is the process in which multiple copies of a target region of DNA are made. There are three major steps involved: denaturation; the melting of the double stranded DNA to make single stranded DNA and to disassociate DNA/primers, annealing; primer binds to single stranded DNA, extension; DNA polymerase extends from the primer binding site. The annealing temperature is based on the primers (4°C per GC's and 2°C per AT's). The time of the extension cycle depends on the length of the DNA fragment to be copied; 1 minute extension time for $\leq$1 Kb and 15 seconds for every additional 1 Kb.

2.5.1 Primers

Primers are short oligonucleotides to which the DNA polymerase can attach the first deoxyribonucleotide and initiate DNA synthesis. The primers are designed so that they are partly or completely complimentary to the flanking regions of the DNA to be amplified. When the primers have hybridized to the template DNA elongation will proceed in the 5' $\rightarrow$ 3' direction. The primers used in this work are listed in table 2.1:

Table 2.1: List of primers used in this work

Primers	Oligonucleotide sequence
yaeNA	5'-GGATCCCGGAAGCTTGCGGTCTGC-3'
yaeND	5'-GTCGACGACTACGCTGCGGCAAACTTC-3'
yaeN2A	5'-CGGGATCCCAAAGAGAGCCGCACAGTACTC-3'
yaeN2D	5'-GCGTCGACGTTTATCTTCGTAGGCTTAGACT-3'
ydaOAI	5'-GGATCCACTGACCGGGCATCATCGC-3'
ydaODII	5'-GTCGACCGGTCGTACAGCTCGTGCAG-3'
yaeN-for	5'-AGTGAAACGATGACACTCAGCCTCAATAGAC-3'
yaeN-rev.	5'-GGCTGGCGTTTAACTAAGCGTTTTCTGCCTC-3'
araC-for.	5'-CCATTCGCGAGCCTCCG-3'
araC-rev.	5'-GAGTGTCATCGTTTCACTCCATCCAAAAAAA-3'
araB-for.	5'-CTTAGTTAAACGCCAGCCAGAAACAAC-3'
araB-rev.	5'-CTTATAGAGTCGCAACGGCCT-3'
ydaO-for.	5'-GGGAATTCCATATGCAAGAAAATCAACAAAT-3'
ydaO-rev.	5'-CCTTAAGCTCGAGTTTCACTTCAACCACATT-3'
AF1595-for.	5'-GGGAATTCCATATGATTACCTGCTCCAAGTGCAGCA-3'
AF1595-rev.	5'-CCTTAAGCTCGAGATCTACGCCCCTGAGCTCCTCAA-3'
AF1321-for.	5'-GGGAATTCCATATGAAGTGCAAAAAGTGCGGAAGAA-3'
AF1321-rev.	5'-CCTTAAGCTCGAGGCTCAGCCTCTCCCTCAGCCT-3'
Sso0586-for.	5'-GGGAATTCCATATGATTTGTGATAATTGTAAGACGCGAG-3'
Sso0586-rev.	5'-CCTTAAGCTCGAGCGATATAGGTAAGTATTTTTGATATTCTTGAGG-3'

2.5.2 DNA-polymerases

A DNA-polymerase is an enzyme that synthesizes a new strand of the DNA in the 5' → 3' direction using an antiparallel DNA strand as template [44]. Depending on the purpose of the PCR products different DNA-polymerases were used.

Taq DNA-polymerase(Promega)

Taq DNA-polymerase is a thermostable enzyme of approximately 94 kDa isolated from *Thermus aquaticus*. This enzyme replicates DNA at 74°C and exhibits a half-life of 40 minutes at 95°C. The enzyme catalyzes the polymerization of nucleotides into duplex DNA in the 5 → 3 direction in the presence of magnesium and also possesses a 5 → 3 exonuclease proofreading activity.

PLATINUM *Taq* DNA polymerase High Fidelity (Promega)

Platinum *Taq* DNA Polymerase is ideal for automatic hot-start amplification of DNA fragments with improved specificity. It is derived from recombinant *Taq* DNA polymerase by binding of a thermolabile inhibitor containing monoclonal antibodies to *Taq* DNA polymerase. During the initial denaturation step of PCR the inhibitor is denatured and active *Taq* DNA polymerase is released into the reaction.

Pfu Turbo DNA polymerase (Promega)

Pfu DNA Polymerase is a thermostable enzyme of approximately 92 kDa isolated from *Pyrococcus furiosus*. The enzyme replicates DNA at 75°C, catalyzing the polymerization of nucleotides into duplex DNA in the 5- 3 direction in the presence of magnesium. *Pfu* DNA Polymerase also possesses 3- 5 exonuclease proofreading activity. Base mis-insertions that may

occur during polymerization are rapidly excised by the proofreading activity. Consequently, *Pfu* DNA Polymerase is recommended for use in PCR and primer extension reactions that require high-fidelity synthesis. *Pfu* DNA Polymerase-generated PCR fragments are blunt-ended.

Phusion **High-Fidelity DNA Polymerase**

Phusion is a high fidelity DNA polymerase that offers extreme performance for all PCR applications with an error rate of $4.4\text{x}10^{-7}$. It possesses 5'→3' DNA polymerase activity and 3'→5' exonuclease activity and generates blunt ends in the amplification products.

2.6 Reverse-Transcription PCR (RT-PCR)

RT-PCR is a technique in which an RNA strand is first reverse transcribed into its DNA complement (cDNA) through an RNA-dependent DNAse, reverse transcriptase. A complementary DNA strand is then synthesised through the use of a deoxyoligonucleotide primer and a DNA-dependent DNA polymerase. This new double-stranded DNA is then amplified via PCR as described earlier. A two-step RT-PCR protocol was followed. First the reverse transcription mixture was prepared as listed in table 2.2 and the mixture was incubated for 1 hour at 55°C. In the second step 2μl of the RT reaction was used as a template in a PCR reaction as described in table 2.3.

2.7 Purification of PCR product

When the PCR reaction gave a correct product which was to be used for later work the product was purified by using the StrataPrep^{TM} PCR Purifi-

Table 2.2: The first step in a two-step RT-PCR procedure.

Component	Amount (μl)
5X RT Buffer	5
0.1 mM DTT	1
10 mM dNTP	1
3' Primer	2
RNA	5
Superscript^{TM} $III RT (200U/\mu l)$ enzyme	1
dH_2O	6

Table 2.3: The second step in a two-step RT-PCR procedure.

Component	Amount (μl)
dH_2O	37.5
10X Buffer	5
10mM dNTP	1
5' Primer	2
3' Primer	2
cDNA	2
Phusion DNA polymerase	0.5
Total	50

cation Kit according to the manufacturers instructions. For the elution of the product TE buffer pH 8.0 or MilliQ water was used.

2.8 Electrophoresis

Electrophoresis is a method used to separate molecules (DNA or proteins) electrophoretically according to their size and charge.

2.8.1 Agarose gel electrophoreses

The phosphate groups that together with the sugar groups make up the backbone of the DNA are negatively charged. When the DNA is placed on a field with an electric current it will begin to migrate towards the positive electrode. The agarose gel is a cross-linked, three-dimensional matrix. When the DNA molecules are pulled to the positive end by the current, the smaller molecules are able to navigate faster through the agarose gel than the larger ones and hence make it further down the agarose gel. The gel is stained with ethidium bromide in order to visualize how these DNA molecules resolved into bands along the gel.

2.8.2 SDS-PAGE

This is a method that separates proteins according to size. SDS (sodium dodecyl sulfate) is a charged detergent that can dissolve hydrophobic molecules. When exposed to SDS, the proteins become denatured and negatively charged. They will therefore migrate towards the positive pole when exposed to an electric field. The environment of choice is polyacrylamide, which is a polymer of acrylamide monomers. When this polymer is formed, it turns into a gel made of a mesh of tunnels of fibers.

The gels were prepared as listed in table 2.4 and run at 190 V for 45 minutes in TGS buffer in a Mini-PROTEAN3 Cell (BIO-RAD). The bands were visualised by UV after staining and destaining in Coomassie Brilliant Blue R-250 and 10% acetic acid/ 20% ethanol, respectively. Pictures were taken using GeneSnap 6.00.26 (Synoptics Ltd) and ChemiGenius Bio Imaging system (Syngene).

Table 2.4: Preparation of SDS-PAGE gels (2 gels)

Components	Separating gel (12%)	Stacking gel (4%)
dH_2O	3.35 ml	6.1 ml
1.5 M Tris-HCl pH 8.8	2.5 ml	-
0.5 M Tris-HCl pH 6.8	-	2.5 ml
10% (w/w) SDS	100 μl	100 μl
Acrylamide/Bis (37.5:1)	4 ml	1.3 ml
10% Ammonium presulphate	50 μl	50 μl
TEMED	5 μl	10 μl

2.8.3 Molecular weight standards

1Kb DNA Ladder (Invitrogen)

The 1Kb molecular weight standard (figure 2.1) was used to determine the size of linear DNA fragments from 500 bp to 16 kb, and to estimate the amount of DNA present in a sample.

Supercoiled Plasmid DNA Standard (Invitrogen)

The Supercoiled Standard (figure 2.2) was used to determine the size of the supercoiled plasmids DNA from 2-16 kb and estimate the amount of plasmid DNA present in the sample.

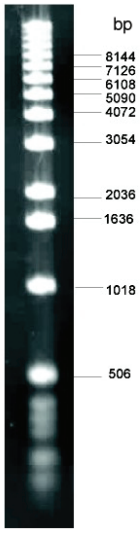

Figure 2.1: 1 Kb DNA Ladder

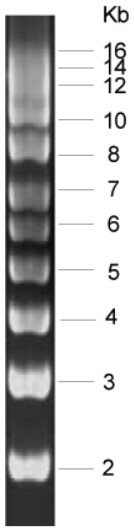

Figure 2.2: Supercoiled standard (www.invitrogen.com)

Low-range SDS-PAGE standard (BIO-RAD)

Low-range SDS-PAGE standard was used to determine the approximate molecular weight of the proteins.

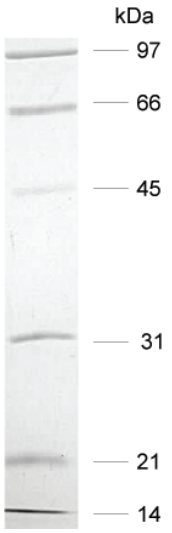

Figure 2.3: Low-range SDS-PAGE standard

2.9 Vectors

2.9.1 Substitution vector pKO3

The pKO3 vector is an integration vector with a temperature sensitive replication system. At 42°C (non-permissive temperature) the vector containing an insert integrates into the chromosome by homologous recombination, creating a tandem duplication. When shifted to the permissive temperature (30°C) the pSC101 replication origin in the integrated plasmid is activated and the vector is excised from the chromosome. To select for the loss of vector sequence from the cell, the *Bacillus subtilis* gene *sacB* is incorporated into the vector, since expression of *sacB* in the presence of sucrose is lethal to *E. coli*. This vector also carries the gene for chloramphenicol resistance, *cat* as a positive marker. A cloning site with the restriction sites for *Bam*H1, *Not*1, *Sal*1 *and Sma*1 is also included in order to clone DNA fragments flanked by one or two of these restriction sites (figure 2.4).

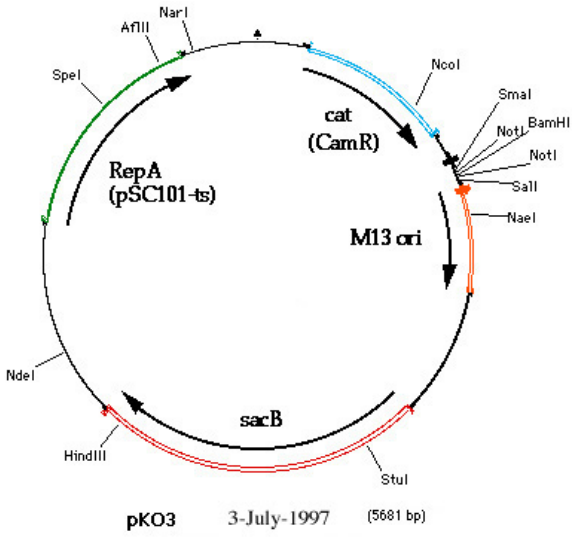

Figure 2.4: Substitution vector pKO3. Figure from [45]

2.9.2 Sequencing vector pCR4-TOPO (Invitrogen)

The pCR4-TOPO sequencing vector (figure 2.5) is a 3957 bp vector with a single 3'-deoxythymidine (T) overhang for cloning of *Taq* polymerase-amplified PCR products. It is commonly used for cloning and sequencing of PCR products. The *Taq* polymerase has a terminal transferase activity, which adds a single deoxyadenosine (A) to the 3'-end of the PCR products, which allows the PCR product to efficiently ligate with the vector with the aid of Topoisomerase I from the *Vaccinia* virus. As positive selection markers the vector carries the genes for ampicillin and kanamycin resistance and the lethal *E. coli* gene *ccdB* which is fused to the C-terminus end of LacZα. Ligation of the PCR product disrupts the expression of the *lacZα-ccdB* gene fusion permitting growth of only positive recombinant clones.

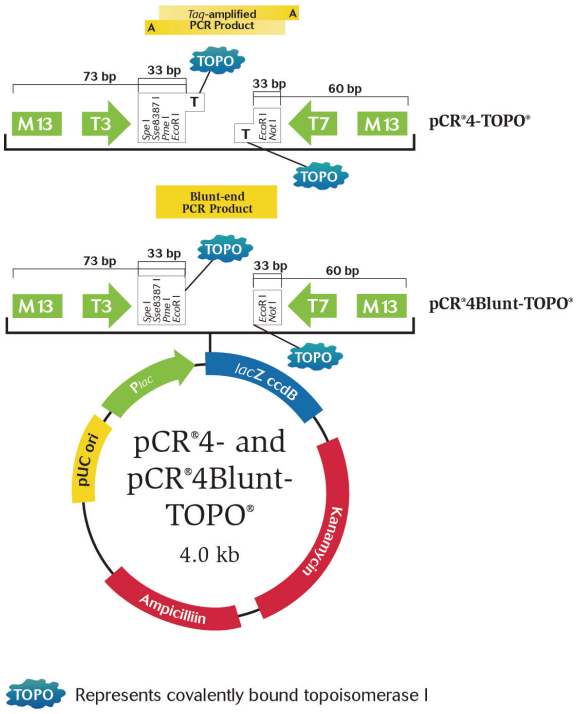

Figure 2.5: Sequencing vector pCR4-TOPO. (www.invitrogen.com)

2.9.3 Cloning vector pACYC184

The pACYC184 plasmid (figure 2.6) is a low-copy number, circular, 4245 bp cloning vector which contains the replication origin, *rep*, responsible for the replication of the plasmid, the *tet* gene encoding tetracycline resistance protein and the *cat* gene coding for chloramphenicol acetyl transferase that confers resistance to chloramphenicol.

2.9.4 Expression vector pBAD/HisA (Invitrogen)

The pBAD/HisA (figure 2.7) is a 4102 bp circular expression vector. Proteins made by expression of genes in this vector acquire an N-terminal tag of six histidines (his-tag). The vector carries the ampicillin resistance gene as a positive selective marker, in addition to the *araC* gene and *araBAD* promoter

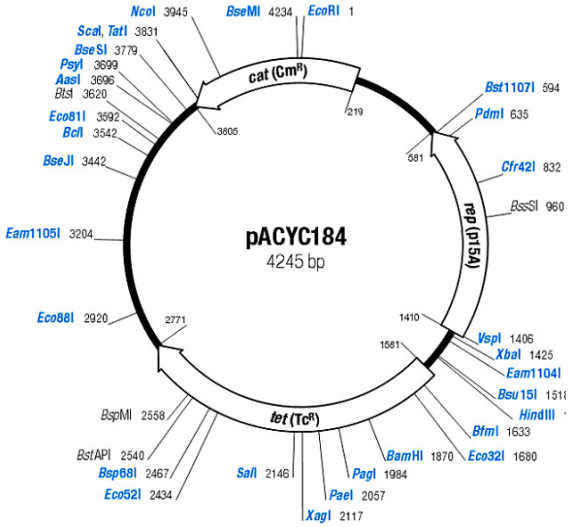

Figure 2.6: Expression vector pACYC184 (http://www.fermentas.com)

(P_{BAD}) in order to regulate expression of the cloned gene. The promoter is up- and down regulated by the product of the gene *araC*. *AraC* is a transcriptional regulator that forms a complex with arabinose. In the absence of arabinose, the *araC* dimer binds $araO_2$ (repressor) and $araI_1$ (inducer) forming a 210 bp loop which leads to a complete inhibition of transcription. In the presence of arabinose, the dimer is released from $araO_2$ and binds $araI_1$ and $araI_2$ leading to transcription. By adding glucose or glucose-6-phosphate expression is switched off through the catabolite repression system.

2.9.5 Expression vector pET-30a (Novagen)

The pET-30a-c(+) is an expression vector carrying an N-terminal HisTag/thrombin/STag/enterokinas configuration plus an optional C-terminal HisTag sequence. It has a T7 promotor and terminator and carries the gene for kanamycin resistance (illustrated in figure 2.8).

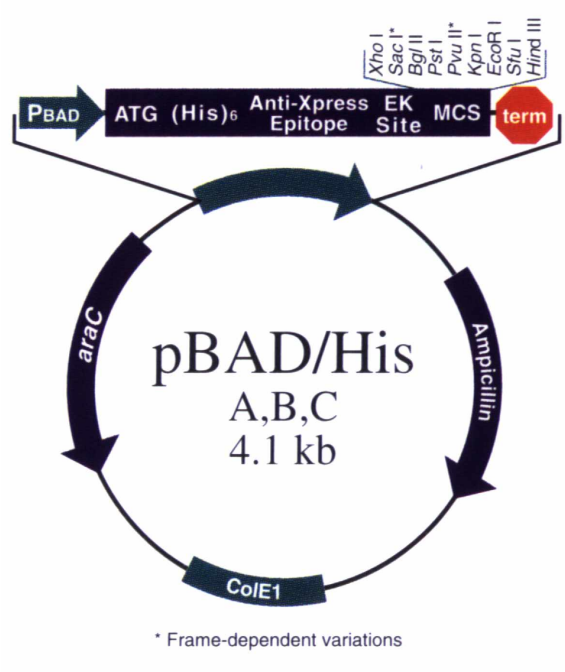

Figure 2.7: Expression vector pBAD/hisA. (http://www.invitrogen.com)

2.10 Cutting of DNA with restriction enzymes (RE).

Restriction enzymes were first discovered by Smith, Wilcox and Kelly in 1968 who isolated and characterised the *Hind*II enzyme from *Haemophilus influenzae*. A restriction enzyme is a bacterial DNA-cutting enzyme that recognises foreign DNA and cuts at short, specific pallindrome sequences. Methylation of the same sequences by the host protects host DNA from being cut by it's own enzymes. The different RE's recognise different sequences, but they all produce a product which can serve as a substrate for a DNA-ligase. An overview of the restriction enzymes used in this work and their cutting sites are listed in table 2.5

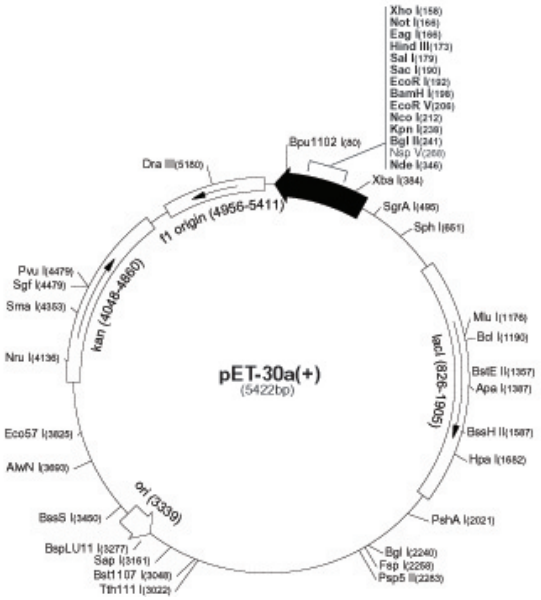

Figure 2.8: Expression vector pET-30a-c(+). (http://www.emdbiosciences.com)

Table 2.5: Specific sequences recognised by the restriction enzymes used in this project.

Restriction enzyme	**Recognition sequence**
*Bam*HI	5'...G$^{\downarrow}$GATCC...3'
*Sal*I	5'...G$^{\downarrow}$TCGAC...3'
*Nde*I	5'...CA$^{\downarrow}$TATG...3'
*Xho*I	5'...C$^{\downarrow}$TCGAG...3'

2.11 T4 DNA Ligation and transformation

2.11.1 T4 DNA Ligation

DNA ligation involves creating a phosphodiester bond between the 3' hydroxyl of one nucleotide and the 5' phosphate of another. T4 DNA ligase originates from the T4 bacteriophage and ligates DNA fragments with overhanging cohesive ends. T4 DNA ligase can also ligate blunt ends, however, this requires a larger amount of ligase present in the mixture.

2.11.2 Transformation

Transformation is the process where free DNA is incorporated into a recipient cell and causes genetic change. A number of bacteria, both Gram positive and Gram negative, have been found to be naturally transformable, however some bacteria have to be made competent by artificial means in order to take up external DNA. There are generally two ways of transforming cells; chemical transformation and electroporation.

Preparation of electrocompetent cells

- 10 ml of LB medium was inoculated and incubated at 37°C over night on a shaking platform.
- 1 ml from the overnight culture was transferred and incubated under the same conditions until a cell density of 0,5-0,8 at OD_{600} was obtained.
- 10 ml of LB-medium was inoculated with a small amount of cells and incubated at 37°C overnight on a shaking platform.
- Transferred 1 ml of overnight culture to 100 ml of LB-medium and incubated overnight under same conditions.
- The cells were then transferred to a chilled, sterile centrifuge tube and kept on ice for 20 mins before being collected by centrifugation at 4000xg for 15 mins at 4°C.
- The supernatant was removed and the cells washed with 100 ml and 50 ml ice-cold, autoclaved dH_2O.
- Cells were then washed with 2 ml ice-cold 10% ice-cold Glycerol before being resuspended in 10% Glycerol.

- The cells were then transferred to eppendorf tubes in aliqvotes of 50μl and frozen at -70°C. technique (REF).

Electroporation

Electroporation is a technique where cells are exposed to pulsed electrical fields to open small pores in their membranes. When small DNA molecules are present outside the membrane during the electric pulse they can enter the cell through these pores. This method has been proven to be more effective than the chemical transformation

- 50 μl electrocompetent cells were thawed on ice and added 1 μl plasmid. Cells were incubated on ice for 1 minute.

- Cells were then transferred to an electroporation cuvette and electroporation was performed at 2.5 kV, 200Ω, 25 µf. Time constant varied between 4.0 and 4.9 seconds.

- After electroporation cells were immediately added 1ml of room-temperature SOC medium and incubated for 1 hour at 37°C with shaking.

- Cells were then plated out on LA low-salt medium containing the antibiotic the plasmid selected for.

2.12 Isolation of plasmid

Depending on whether a low-copy number or high-copy number plasmid was used different extraction methods were used.

2.12.1 Rapid plasmid isolation

When the only purpose of the plasmid extraction was to check for the presence of plasmid and estimate its size the rapid isolation method was used. This method is a modified version of a rapid alkaline extraction procedure first described by Vogelstein and Gillespie, 1979 [46].

2.12.2 StrataPrep^{TM} Plasmid Miniprep Kit

When high-copy number plasmids were used and the purified plasmid were to be used for further work like PCR, sequencing and cloning, this kit was used according to the manufacturers description. TE-buffer (pH 8.0) or MilliQ water were used as elution buffers.

2.12.3 Large-scale plasmid purification

When a low-copy number plasmid was used a larger batch of culture was needed in order to get a higher yield. To isolate plasmids from a large culture the QIAGEN Plasmid Midi Kit (100) was used. Plasmids were isolated according to the manufacturers instructions.

2.12.4 Gel Extraction

When gel-fractionated DNA or plasmids of a certain size were needed, the DNA was first run on an 0.7% agarose gel. The DNA was then extracted using StrataPrep DNA Gel Extraction Kit from Stratagene according to the manufacturers instructions.

2.13 Sequencing

The dideoxynucleotide method was developed by Sanger in 1977. In this procedure the sequence is determined by making a copy of the single-stranded DNA using a modified DNA polymerase, which then uses deoxyribonucleotide triphosphates (dNTP) as substrates and adds them to a primer. The incubation mixture contains small amounts of the fluorescent-labelled dideoxy analogs of the dNTPs (ddNTP*) which lacks the 3'-hydroxyl group and thus inhibiting lengthening of the chain. Through repetitive cycles of denaturing, hybridisation and synthesis a pool of DNA fragments with primer defined 5'-ends and variable 3' ends, determined by the position of the incorporated ddNTP*, is produced. These DNA fragments are then separated by polyacrylamidegel electrophoresis. As the different ddNTP*s fluoresce at different wavelengths the detector at the end of the system will detect which of the four ddNTP*s have caused the termination of the fragment and thus determine the sequence of the fragments. Big-Dye 3.1 sequencing kit was used for th sequencing and table 2.6 shows the reagents and amounts used for the sequencing of PCR products.

Table 2.6: Sequencing mixture

Reagent	Quantity
Big-Dye 3.1	1 μl
Sequencing buffer	1 μl
Template	200 ng
Primer	3.2 pmol
Deionised water	x.x μl[1].
Total volume	10 μl

[1]Add water for a final concentration of 10μl

2.14 Construction of an Arabinose-dependent mutant.

2.14.1 Cross-over PCR

In order to make a construct for an integrative transformation fusion PCR was used [47, 48]. Briefly, 3 separate PCR products were produced using 3 different sets of primers. The gene to be assessed was tagged at both ends by primers with a 3 codon overhang with homology to the genes on either side. The up- and downstream genes were tagged with a 3 codon overhang with homology to the gene to be assessed. The PCR products were gel-purified and the fusion of the three PCR products was performed using reaction mixtures and programs as described in tables 2.7- 2.11

A schematic overview of the construction of the fused PCR products are shown in figure 2.9.

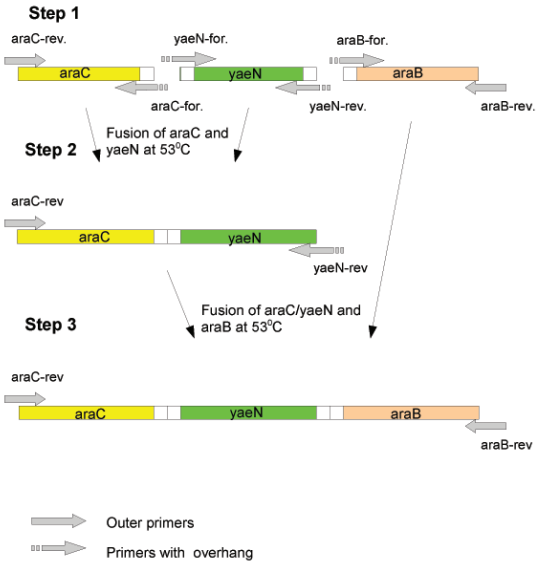

Figure 2.9: Crossover PCR. Three separate PCR products were constructed. 3'-end and 5'-end of the primers upstream and downstream, respectively, were tagged with an overhang with homology to the gene to be assessed. Also the primers to the gene to be assessed were tagged with overhangs with homology to the genes on either side. The three PCR products were then fused into one in two separate operations.

Table 2.7: PCR Mix for cross-over PCR using *Taq* HiFi Polymerase

Components	Amount (μl)
suH_2O	xx.x
25mM $MgCl_2$	4.0
Taq-buffer	5.0
8mM dNTP	1.0
10µM forward primers	1. 0
10µM reverse primers	1.0
20-150ng DNA	x.x
Taq (5U/μl)	0.2
Total volume	50

Table 2.8: Program for PCR with *Taq* HiFi polymerase

Temperature	Time	Cycles
96°C	2 min	1
96°C 53°C 68°C	0.3 min 0.3min 1.5 min	30
68°C	5 min	1
6°C	∞	1

Table 2.9: PCR mix for fusion of two PCR products

Components	Amount (μl)
suH_2O	xx.x
50mM $MgSO_4$	2.0
Taq HiFi buffer	5.0
8mM dNTP	1.0
c. 200ng PCR1 product	x.x
c. 200ng PCR2product	x.x
Taq HiFi (5U/ μl)	0.2
Total volume	50

2 μl of 10mM outer primers were added to the fused PCR products before running the amplification of fused PCR products (figure 2.11).

Table 2.10: Program for Fusion PCR using *Taq* HiFi Polymerase.

Temperature	Time	Cycles
96°C	2 min	1
96°C 53°C 68°C	0.3 min 0.3 min 3 min	5

Table 2.11: Amplification of fused PCR products

Temperature	Time	Cycles
96°C	2 min	1
96°C 53°C 68°C	0.3 min 0.3min 1.5min	20
68°C	10 min	1
6°C	∞	1

2.15 Cloning and characterisation of mesJ homologs from *E. coli* and Archaea.

2.15.1 Cloning

The definition of molecular cloning is to isolate a fragment of DNA and obtaining multiple copies of it usually by transforming it into a bacterium by using a vector.

PCR

The open reading frames *ydaO* from *E. coli* (ydaO-for./rev.), *AF1595* and *AF1321* from *A.fulgidus* (AF1595-for./rev. and AF1321-for./rev. respectively) and *Sso0586* (Sso0586-for./rev.) from *S. solfataricus* were amplified using *Phusion* High-Fidelity DNA Polymerase. The PCR mixture and program are listed in tables 2.12 and 2.13, respectively. PCR products were

purified using the Stratagene PCR Purification Kit according to the manufacturers instructions. PCR product was eluted with 40 μl MilliQ water.

Table 2.12: PCR Mix for PCR using *Phusion* Polymerase

Components	Amount (μl)
suH_2O	32.5
Phusion-buffer (5X)	10.0
10mM dNTP	1.0
2x10µM primers	1.0
30ng/μl DNA	5.0
Phusion (2U/μl)	0.5
Total volume	50

Table 2.13: PCR Program using *Phusion* Polymerase

Temperature	Time	Cycles
98°C	30 sec	1
98°C 54°C 72°C	10 sec 30 sec 30 sec	30
72°C	10 min	1
6°C	∞	1

Restriction digestion

The vector and the purified PCR products were cut with the restriction enzymes *Nde*I and *Xho*I (table 2.14). Two sets of restriction digest mixtures were set up, one with the vector to be cut and one with the PCR products.

The RE/PCR cutting mixture was incubated for 2 hours at 37°C. After 1 hour incubation another 1μl restriction enzymes were added. At the end of the incubation restriction enzymes were heat inactivated at 65°C for 15 minutes. The digested PCR products were run on a 0.8% agarose gel and

Table 2.14: Mixture for cutting with restriction enzymes (RE's)

Components	Amount (μl)
PCR-product in SuH_2O	40
Buffer D (Promega)	5
BSA (1mg/ml)	5
*Nde*I (U, Promega)	1
*Xho*I (U, Promega)	1
Total	52

were gel-purified using the Stratagene Gel extraction Kit according to the manufacturers instructions.

T4 DNA Ligation and transformation

The digested vector and insert were ligated using T4 DNA Ligase (Promega). The DNA ligation mixture (table 2.15) was incubated at 4°C overnight and 1/10 of the ligated DNA was mixed with 50 μl electrocompetent Top 10 cells and electroporated as described earlier.

The electroporated cells were then plated on LA media containing the appropriate antibiotics and incubated overnight at 37°C.

Table 2.15: T4DNA ligation mixture

Component	Amount (μl)
MilliQ water	3
Vector	3
Insert	2
T4DNA Ligase Buffer (10x)	1
T4DNA Ligase	1
Total volume	10

The following day colonies from the overnight plates were streaked for single colonies and incubated under the same conditions. 6 colonies were picked and the plasmids isolated and sequenced in order to confirm that the

inserts were correct.

2.15.2 Pilot expression and solubility testing of mesJ homologs

Pilot expression

- Plasmids containing the correct insert were transformed into *E. coli* Codon+ cells by electroporation, plated on LA medium containing the appropriate antibiotics and incubated overnight at 37°C. The following day colonies were picked and streaked for the isolation of single colonies.
- A starter culture in LB medium containing the appropriate antibiotics and inoculated with the transformed Codon+ cells was set up and incubated overnight at 37°C.
- 300 ml LB + antibiotics was inoculated with 100 ml of the starter culture and incubated at 37°C until OD_{600} was between 0.4 and 0.5.
- After taking a subsample for later analysis 1 ml of the culture was spun down at 13000 x g and the pellet frozen at -20°C.
- 300 μl 1M IPTG was added to the 300 ml culture (final concentration 1mM IPTG) in order to start expressing the protein. OD_{600} was read at 1, 2 and 4 hours after initiation. After each reading subsamples of the cultures were spun down and the pellets frozen at -20°C.
- The pellets from the subsamples were resuspended in 50 μl 10mM Tris (pH=8.0) and an equal amount of testbuffer was added.
- The samples were then heat-treated for 5 min at 100°C and placed on ice.

- The samples were run on an SDS-PAGE gel.

Testing of solubility

- After ended incubation the culture was transferred to 6 x 50 ml Falcon tubes and centrifuged at 7500 rpm for 15 min and the pellets were kept at -20°C.
- In order to test for solubility the pellets were resuspended in 3 ml of the test buffer and sonicated for 3 x 2 mins. The parameters were set to: Output control = 3, Duty cycle = 25% and timer = 2 min.
- After sonication the suspension was inspected using a light microscope to verify that the cells were lysed.
- In order to check for the presence of expressed protein a sub-sample was collected before centrifugation.
- 1.5 ml of the sonicated (Branson Sonifier 250) cells was transferred to an Eppendorf tube and centrifuged at 10 000 rpm for 15 mins and the supernatant containing the protein was transferred to a new tube.
- The supernatant was then analysed by SDS-PAGE (20 μl sample + 20 μl testbuffer) with the pre-centrifugation sub-sample as a positive control.

Test buffers:

1. Lysisbuffer: 50mM HCl, pH=8.0, 0.5M NaCl, 20% glycerol
2. KPO^4: 50mMKPO^4, pH=7.5, 0.1M NaCl
3. Tris/HCl: 50mM Tris, pH=7.7, 50mM NaCl

2.15.3 Protein purification

His-tagged *E. coli* YdaO, expressed in *E. coli* Codon+, were resuspended in Lysisbuffer with 20 % glycerol and disrupted by a French pressure cell at 1000 psi in High position. The sample was then centrifuged at 10000 x g for 15 min to remove cell debris. The cleared supernatant was then sterilized by filtration through first a 0.45µm then 0.2µm filter (FP30/0.2 CA-S, Schleicher & Schuell) before being applied to a BD TALON^{TM} Metal Affinity Resin column equilibrated with a TALON wash buffer (50mM Tris, pH=7.7, 0.3M NaCl, 10mM immidazole) using a pump system from BIO-RAD. The column was washed with TALON elution buffer (20mM NaP, pH=7.5, 1mM β-mercaptoetanol, 10% glycerol, 0.3M NaCl, 150mM immidazole) until all the unbound proteins were flushed out and A_{280} was zero. His-tagged YdaO protein was eluted with a buffer B. Fractions of 0.5 ml were collected and analysed by SDS-PAGE. Fractions containing pure YdaO protein were pooled and filter sterilized (0.2µm).

2.15.4 Amplification of *ydaO* from environmental *E. coli* strains from Bangladesh

DNA from 32 strains isolated from different sites in Bangladesh and one type strain (K12-MG1655) was amplified by PCR using *Pfu* Turbo polymerase. The PCR was set up according to the protocol described earlier for PCR with *Pfu* Turbo polymerase. The PCR product were run on an 1%agarose gel. PCR products of the correct size were purified using the Stratagene PCR purification kit and sequenced using the Bid-Dye 3.1 kit

2.16 ATPase activity assay

The ability of *E. coli* YdaO protein to hydrolyse ATP was measured using thin layer chromatography as described by Kasiviswanathan *et al.* (2004) [49]. This method is based on the chromatographic separation of PO_4^{2-} and ADP/ATP by the capillary action of a solvent moving upwards (figure 2.10.)

15μl mixtures containing 25 mM Hepes-NaOH, pH 7.5, 1 mM DTT, 100 μg/ml BSA, 1.5 nmol ATP, 1μCi[γ-^{32}P]ATP (3000 Ci/mmol) and different concentrations of purified protein were incubated with or without the presence of tRNA and incubated for 1 hour at 37°C. 100 ng and 500 ng control protein, TaMCM, with or without the presence of ssDNA was incubated at 1 hour at 60°C. 1 μl aliquots of the mixtures were spotted onto a TLC membrane (PEI cellulose F, Merck VWR) and the bottom part of the membrane was placed in a solvent containing 1 M formic acid and 0.5 M LiCl until the solvent had almost reached the top. The position and amount of dissociated PO_4^{2-} and ADP/ATP was determined using PhosphoImager FLA-2000 (Fuji Photo Film co., Ltd.).

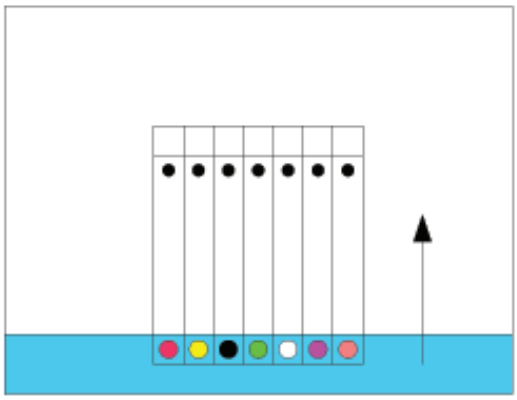

Figure 2.10: Reaction mixtures spotted onto TCL membrane. As the solvent move upwards by capillary action the dissociated PO_4^{2-} is separated from the ADP/ATP

Chapter 3

Results

3.1 Construction of conditional mutant

Several methods were used in order to control the transcription of *yaeN*.

3.1.1 pKO3 as substitution plasmid

A previous attempt to construct a conditional mutant using pKO3 as a substitution plasmid was made by Stranden (2002). This experiment was unsuccessful due to unknown reasons. At the start of this project the experiment was repeated as a control experiment. A brief review of the procedure and results are given below: *Escherichia coli* MG1655 was transformed with the substitution plasmid pKO3 containing an insert, yaeN2-, which has a deleted copy of *yaeN*. Colonies were observed after incubation at 30°C, but not at 42°C. The colonies cultured at 30°C were analysed to see if they still carried the plasmid. An 1% agarose gel showed signs of the plasmid being present, but no clear bands. The presence of both wild-type and the deleted copy of the gene was tested for by PCR using primers yaeNA+yaeND and yaeN2A+yaeN2D, but only the wild-type gave product. In a last attempt to

check whether the bacteria were able to carry the plasmid, *E. coli* MG1655 was transformed with a pKO3yaeN^{control} plasmid and pKO3yaeN2- separately. The pKO3yaeN^{control} plasmid contains an insert with a silent point mutation in *yaeN* as proof that mutagenesis is possible in *yaeN*. The colonies from the transformation were small in size. An inoculum was set up and plasmids were extracted, but no bands were observed.

3.1.2 Fusion-PCR

As two separate experiments failed to produce a conditional mutant using pKO3 as a substitution plasmid, it was decided to try a different approach. The idea here was to insert the *yaeN* gene into the araBAD operon by homologous recombination using a pBAD plasmid, containing the a fusion-PCR insert, as vector. The method for creating the fusion PCR product is explained in figure 2.9. If the vector integrated into the chromosome the wild-type *yaeN* would then be deleted allowing total control over transcription of *yaeN* throught the AraC regulator. All the primers had *Bam*HI/*Sal*I as flanking restriction sites. PCR with the yaeNfor./yaeNrev. resulted in a band about 1.3 Kb with pKO3yaeN^{control} as template. The primers araCfor./araCrev. and araBfor./araBrev. both gave a PCR product of 0.6 Kb with chromosomal DNA as template and the primers yaeNfor./yaeNrev. gave a PCR product of 1.3 Kb (figure 3.1). The fusion-PCR with outer-primers araCrev./yaeNrev. gave a band of 2.5 Kb instead of 1.9 which was expected (figure 3.2). A series of dilutions was set up with different concentrations of the two PCR products to see if this would produce a fusion-product of the correct size, but the product was still 2.5 Kb.

The fusion product from primers araCfor./yaeNrev. and PCR product from primers araBfor./araBrev. were sequenced. The sequencing results

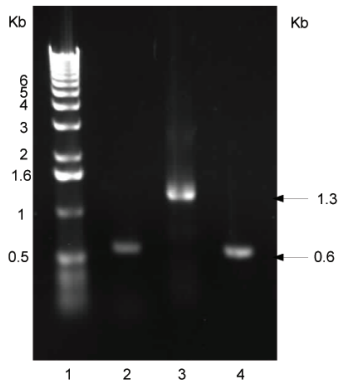

Figure 3.1: PCR product used for fusion PCR: Lane 1; 1Kb DNA Ladder, lane 2; product from primers araCfor. + araCrev., lane 3; *yaeN* and lane 4; PCR product from primers araB.for. + araBrev.

came out satisfactory for the araB product, but not for araC/yaeN fusion product. The araC/yaeN fusion product was inserted into a TOPO-vector using a TOPO-cloning kit in an attempt to get better sequencing results, but as shown in figure 3.2 the plasmid was not of the correct size after the transformation, suggesting the cloning experiment had not worked.

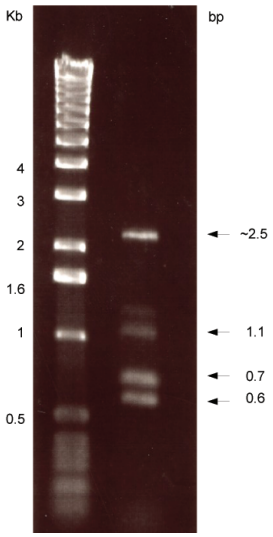

Figure 3.2: Analysis of fusion product between araC/yaeN. Lane 1; 1 Kb DNA ladder, lane 2; fused araC/yaeN product.

In another attempt to fuse *yaeN* with *araC/araB* a new set of primers with flanking restriction sites for *Bam*HI/*Sal*I and *Kpn*I/*Nde*I were used to

amplify *araB* and *yaeN* respectively. The PCR products (yaeN^{res}; 1.3 Kb and araB^{res}; 0.6 Kb) were inserted into a TOPO vector and transformed into *E. coli* TOP10 cells and incubated overnight. 4 colonies were picked from each set of clones and the plasmids were extracted. Bands of 5.3 Kb (TOPO/yaeN^{res}) and 4.6 (TOPO/araB^{res}) were expected, but no plasmids of the correct size were observed (figure 3.3). The most dominating bands were 2.3 Kb, which was smaller than the vector alone and several weak bands were seen which were too large. Another vector was chosen, pACYC184, which contained the same restriction sites as yaeN^{res} and araB^{res}. Both the PCR products and the vector were cut with the appropriate RE's, but once again the plasmids were of the incorrect size after ligation (results not shown) and at this point the experiment was terminated.

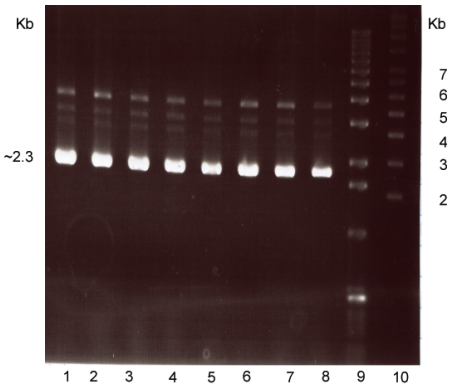

Figure 3.3: Gel picture of extracted plasmids from *E. coli* TOP10 cells. Lanes 1-4; TOPO-vector+yaeN, lanes 5-8; TOPO-vector+araB, lane 9, 1Kb DNA ladder and lane 10; Supercoiled DNA ladder.

3.1.3 Co-transformation of *E. coli* MG1655 with pKO3yaeN2- and pBADMesj

E. coli MG1655 was transformed with pKO3yaeN2-, which has a deleted copy of *yaeN*, and pBADMesj simultaneously and incubated overnight at

42°C. The pKO3yaeN2- would insert into the host chromosome by homologous recombination, whilst *yaeN* would be expressed from the pBADMesJ plasmid. Single colonies were streaked on plates containing 5% sucrose to select for clones that had excised the pKO3 vector from the chromosome as this vector contains the lethal *sucB* gene. Clones that grew in the presence of sucrose were streaked on plates containing arabinose in order to induce transcription of MesJ from the araBAD promoter through the AraC regulator. Clones that grew in the presence of ampicillin, sucrose and arabinose were streaked on plates containing arabinose, ampicillin and chloramphenicol to select for clones that had lost the pKO3 vector. However, no colonies were observed that grew only in the presence of ampicillin suggesting they still carried the pKO3 vector.

3.2 Cloning and characterisation of mesJ homologs in *E. coli* and Archaea

After several unsuccessful attempts to construct a conditional mutant, Soma *et al.* (2003) reported in an article published in September that year that the gene *yaeN*, renamed *tilS* coded for an RNA modifying enzyme. Due to this new information it was decided to focus on *yaeN* paralogs and determine whether they had the same function. *AF1595* and *AF1321* from *A. fulgidus*, *Sso0586* from *S. solfataricus* and *ydaO* from *E. coli* were amplified using *Phusion* High-Fidelity DNA Polymerase. All the primers had flanking restriction sites for *Nde*I (5' end) and *Xho*I (3' end). The PCR products were purified and both the PCR products and the pET-30a expression vector were digested with the restriction enzymes (RE) *Nde*I and *Xho*I (figure 3.4).

Digested and purified PCR products were ligated with the vector. Plas-

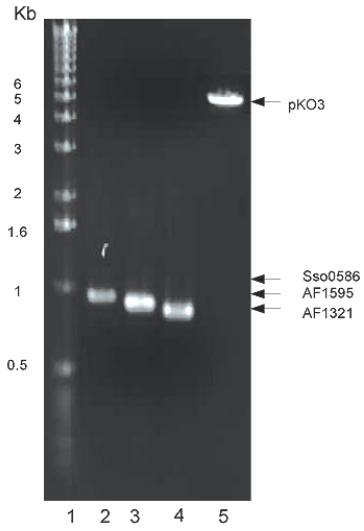

Figure 3.4: Gel purified PCR products and expression vector pET-30a digested with restriction enzymes *Nde*I and *Xho*I: Lane 1; 1Kb DNA Ladder, lane 2; RE digested *Sso0586*, lane 3; RE digested *AF1595*, lane 4; RE digested *AF1321*, lane 5; RE digested pKO3

mids of the correct size (figure 3.5) were cut with the same RE's to test for correct inserts before being sequenced in order to make sure the amplified genes were still functional (figure 3.6).

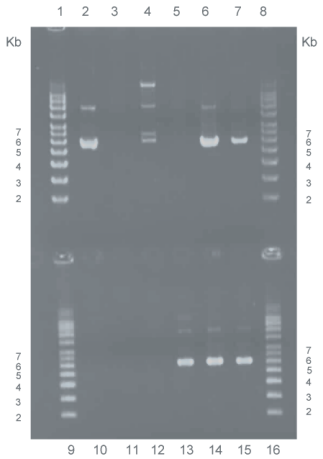

Figure 3.5: Purified plasmids from transformed TOP10 cells: Lanes 1, 8, 9 and 16; Supercoiled standard, lanes 2-4; pET-30a+*AF1595*, lanes 5-7; pET-30a+*AF1321*, lanes 10-12; pET-30a+*AF1321* and lanes 13-15; pET-30a+*Sso0586*.

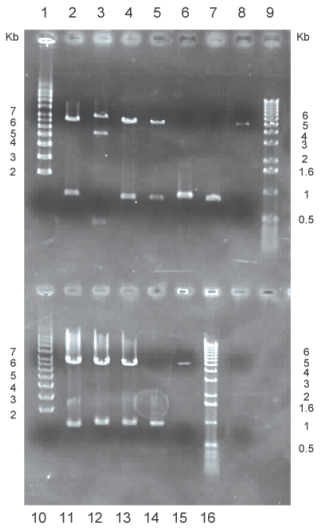

Figure 3.6: *Nde*I/*Xho*I digested purified plasmids: Lanes 1 and 10; Supercoiled standard, lanes 9 and 16; 1Kb DNA Ladder, lanes 2-3; digested pET-30a+*AF1595*, lanes 4-5; digested pET-30a+*AF1321*, lane 6, digested PCR prod. of *AF1595*, lane 7; digested PCR prod. of *AF1321*, lane 8; digested pET-30a, lanes 11-13; digested pET-30a+*Sso0586*, lane 14; digested PCR prod. of *Sso0586* and lane 15; digested pET-30a.

3.2.1 Pilot expression and solubility testing of proteins

Pilot expression

Vector pET-30a containing the correct inserts were transformed into *E. coli* Codon+ cells. The inserted genes were expressed and the level of expression were analysed by SDS-PAGE. Figure 3.7 shows the expression of AF1321 and Sso0586 proteins at the time of initiation of expression, and after 1, 2 and 4 hours.

As the pilot expressions of AF1321 and Sso06586 were satisfactory it was decided to carry on with transformation and expression of YdaO and AF1595 and screen all four proteins for solubility.

Solubility testing

The expressed AF1321 and Sso0586 proteins were first tested for solubility in lysis-, KPO^4- and Tris/HCl buffers. However, as can be seen in figure 3.8,

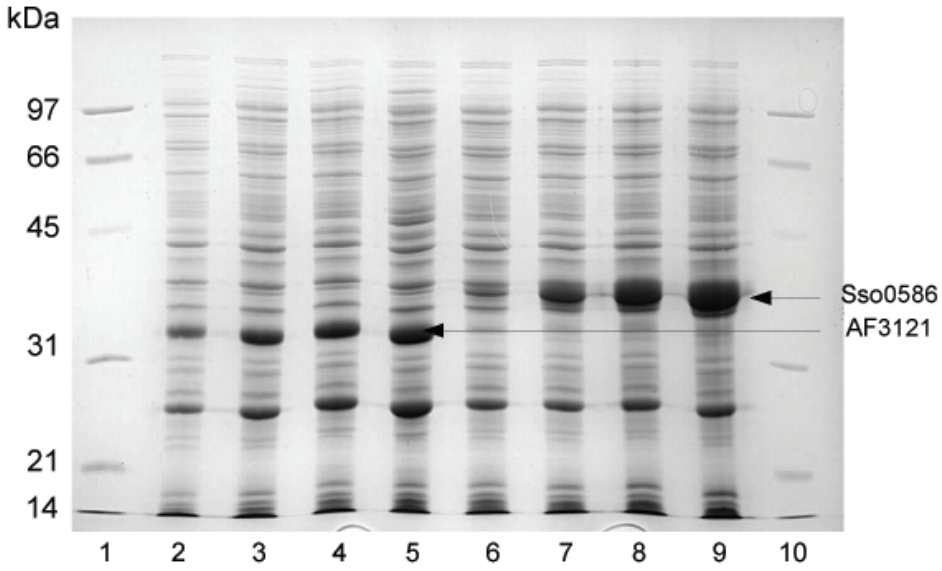

Figure 3.7: Pilot expression of AF1321 and Sso0586. Lanes 1 and 10; Low-range SDS-PAGE standard, lanes 2-5; AF1321 1, 2 and 4 hours, respectively, lanes 6-9; Sso0586 1, 2 and 4 hours, respectively.

the drastic reduction in protein concentration of the correct molecular size indicates that the proteins precipitated in all three buffers.

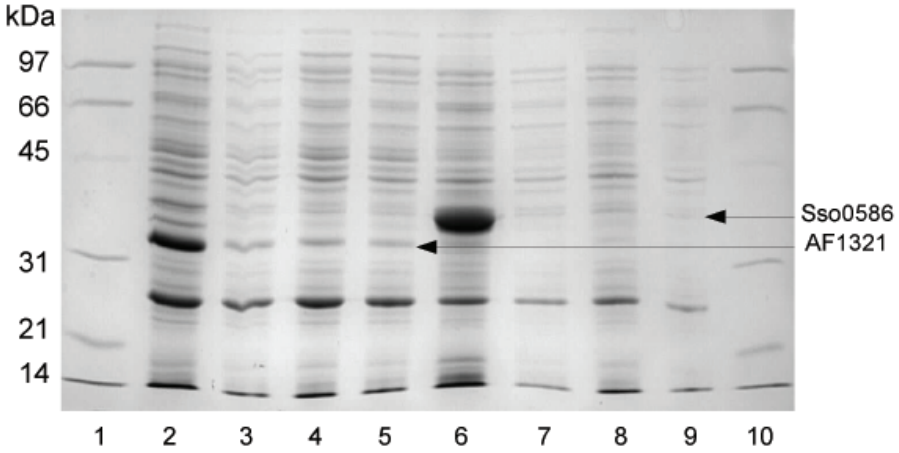

Figure 3.8: Solubility testing of proteins AF1321 and Sso0586. Lanes 1 and 10; Low-range SDS-PAGE standard, lanes 2 and 6; pre-centrifugation sub-sample of AF1321 and Sso0586 respectively, lanes 3-5; AF1321 in buffers 1, 2 and 3 respectively, lanes 7-9; Sso0586 in buffers 1, 2 and 3 respectively.

In order to verify that the proteins precipitated all the cells were lysed in lysis-buffer and an SDS-PAGE was prepared for each of the four proteins,

AF1595, AF1321, Sso0586 and YdaO. On these gels both the supernatant and the pellet was applied. As shown in figures 3.9 - 3.12 the presence of expressed proteins in all the pellets except for YdaO indicates that proteins AF1595, AF1321 and Sso0586 are not soluble in all the buffers tested for, whereas YdaO is soluble in lysis-buffer.

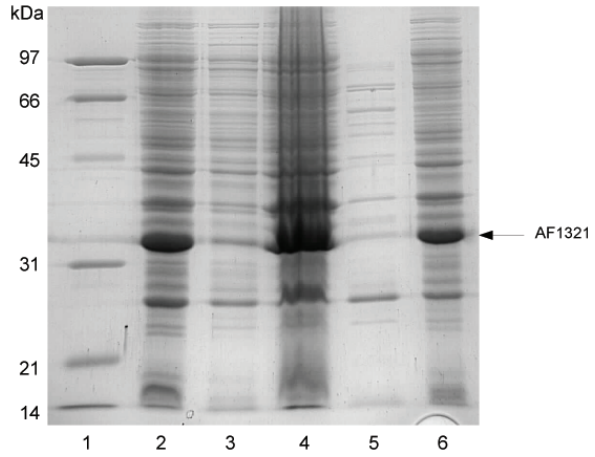

Figure 3.9: Solubility testing of protein AF1321 in Lysis-buffer. Lane 1; Low-range SDS-PAGE standard, lane 2; pre-centrifugation sub-sample of AF1321, lanes 3 and 4; supernatant and pellet, respectively, after sonication and centrifugation, lanes 5 and 6; heat-treated supernatant and pellet, respectively after sonication and centrifugation, lane 7; blank.

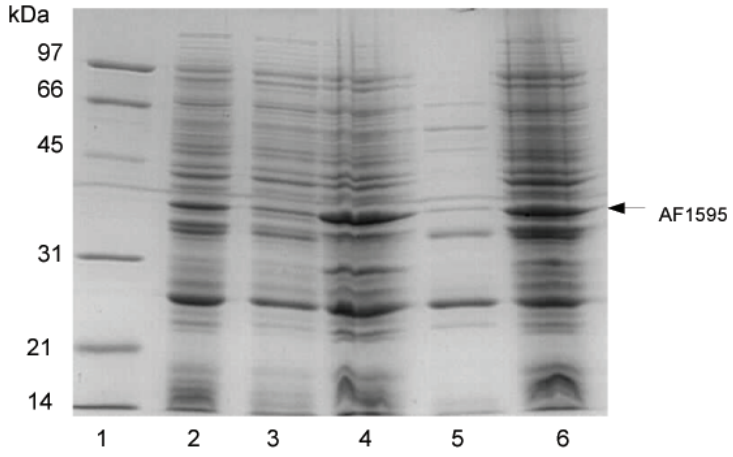

Figure 3.10: Solubility testing of protein AF1595 in Lysis-buffer. Lanes 1 and 8; Low-range SDS-PAGE standard, lane 2; pre-centrifugation sub-sample of AF1595, lanes 3 and 4; supernatant and pellet, respectively, after sonication and centrifugation, lanes 5 and 6; heat-treated supernatant and pellet, respectively after sonication and centrifugation, lane 7; blank.

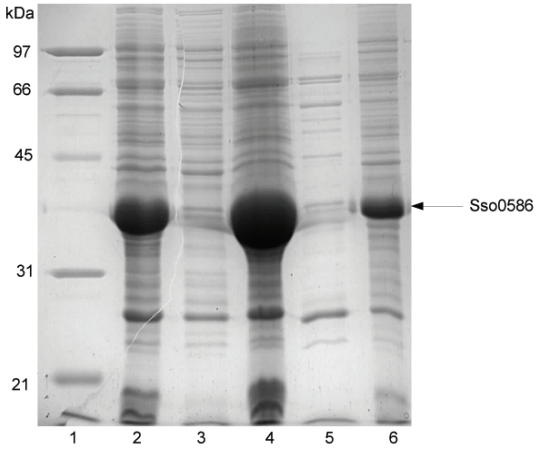

Figure 3.11: Solubility testing of protein Sso0586 in Lysis-buffer. Lanes 1 and 7; Low-range SDS-PAGE standard, lane 2; pre-centrifugation sub-sample of Sso0586, lanes 3 and 4; supernatant and pellet, respectively, after sonication and centrifugation, lanes 5 and 6; heat-treated supernatant and pellet, respectively after sonication and centrifugation.

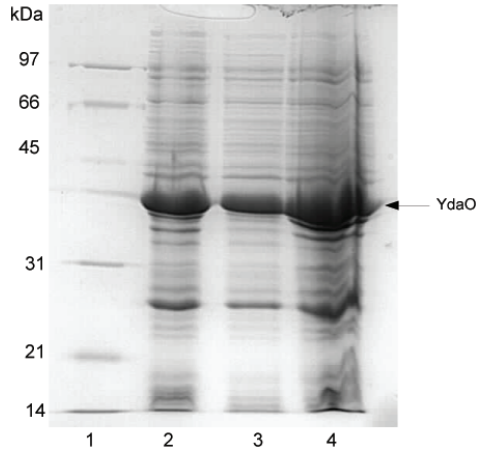

Figure 3.12: Solubility testing of mesJ (ydaO) in Lysis-buffer. Lane 1; Low-range SDS-PAGE standard, lane 2; pre-centrifugation sub-sample of ydaO, lane 3 and 4; supernatant and pellet, respectively, after sonication and centrifugation.

3.3 YdaO ATPase assay

E. coli YdaO was purified, as described previously, and analysed by SDS-PAGE (figure 3.13). The purified protein was also run a native PAGE to verify that the sample was not contaminated with other proteins. The native PAGE showed (not shown) that the sample contained only YdaO and

indicated that it might be a monomer.

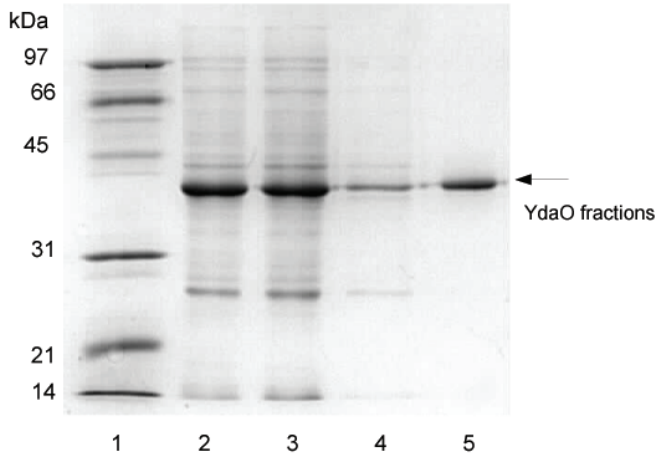

Figure 3.13: Purification of YdaO. Lane 1; Low-range SDS-PAGE standard, lane 2; After centrifugation, lane 3 unbound proteins, lane 4; first wash, lane 5; purified YdaO

As TilS was reported to have ATPase activity, it was decided to analyse YdaO to see if it possessed the ability to hydrolyse ATP. A pilot ATPase assay was set up using TaMCM (*Thermoplasma acidophilum* Mini Chromosome Maintenance protein) as a positive control. Figure 3.14 shows that YdaO very efficiently hydrolyses ATP at both 100 ng and 500 ng.

In order learn more about the efficiency of YdaO, the protein was expressed in a 1 litre batch and purified (figure 3.15) as described earlier . The ATPase assay was repeated using protein concentrations ranging from 0-500 ng at 50 ng intervals (figure 3.16). A graph was drawn based on data extrapolated from this assay and the the efficiency of the enzyme was calculated (figure 3.17). YdaO was shown to hydrolyse ATP at a rate of 55 pmol ATP/min/pmol protein.

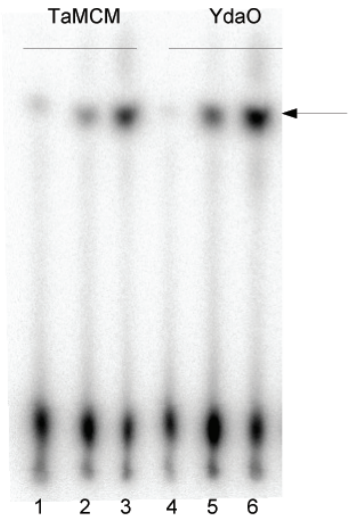

Figure 3.14: YdaO ATPase assay. Lanes 1-3; 0, 100 and 500 ng of TaMCM, respectively, lanes 4-6; 0, 100 and 500 ng of YdaO protein, respectively.

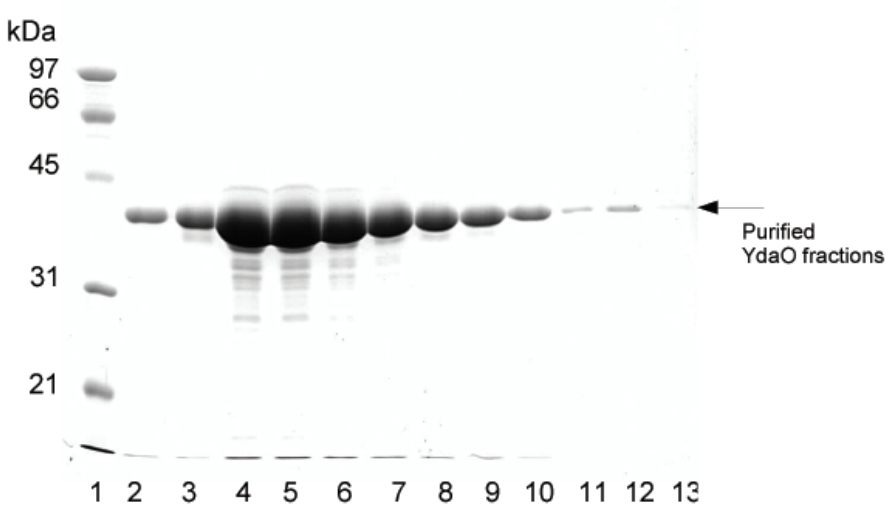

Figure 3.15: Lane 1; Low-range SDS-PAGE standard, lanes 2-13; fractions from the purification of YdaO.

3.4 Amplification of *ydaO* from environmental isolates of *E. coli* from Bangladesh

The gene *ydaO* was amplified from 32 *E. coli* strains isolated from different sites in Bangladesh and the K12 MG1655 reference strain using primers that

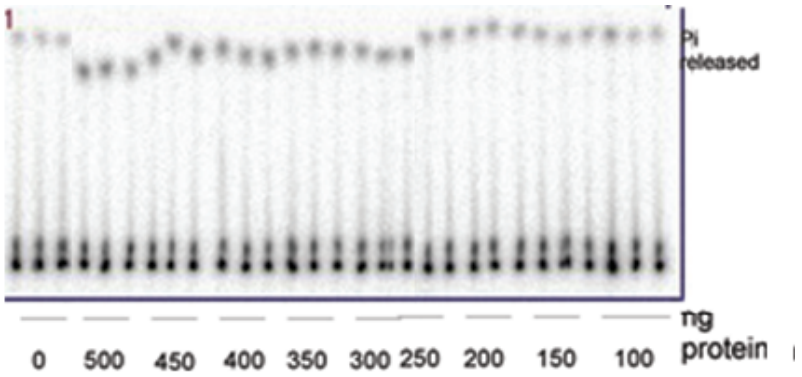

Figure 3.16: ATPase assay mixture containing increasing amounts of YdaO. Lanes 1-3; negative control without protein, lanes 4-30; Different amounts of YdaO starting at 500 ng and decreasing at 50 ng intervals. Each sample occurs as triplicate.

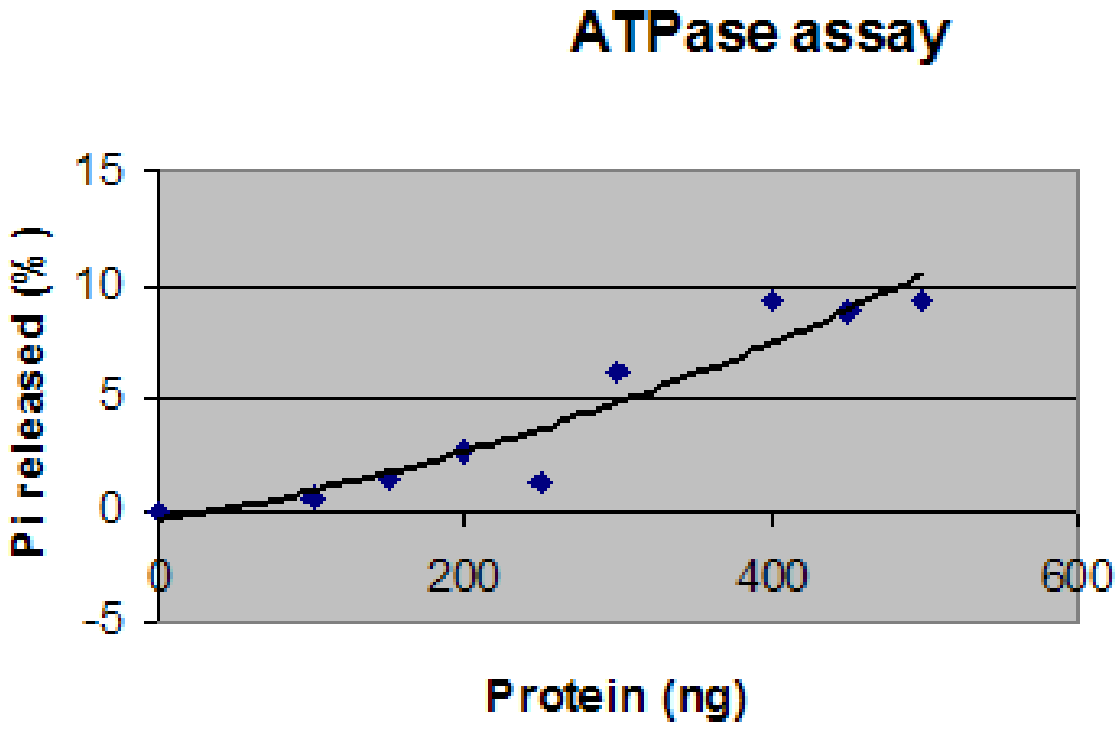

Figure 3.17: A graph was drawn based on the data extrapolated from the quantification results from the YdaO ATPase assay. The Y-axis show the percentage Pi released per hour and the X-axis show amount of protein (ng) ranging from 0-600 ng at 25 ng intervals.

annealed to flanking regions of *ydaO* (ydaOAI/ydaODII). Out of 32 strains only 16 gave product and several of these strains also gave more that one

product (figure 3.18). New amplifications of 23 of the strains were performed using primers ydaO-for./ydaO-rev. from which 10 strains gave a correct product (figure 3.19). Six of the PCR products from ydaO-for./ydaO-rev. were sequenced, but only a few of the sequencing results were satisfactory. Some of the strains which previously seemed to lack the gene *ydaO* when the flanking primers were used, gave product when ydaO-for./ydaO-rev. were used. Due to time restraints it was not possible to optimise the PCR protocol nor re-sequence the PCR products in order to get more sequences and the acquired sequencing results were therefore excluded from the thesis.

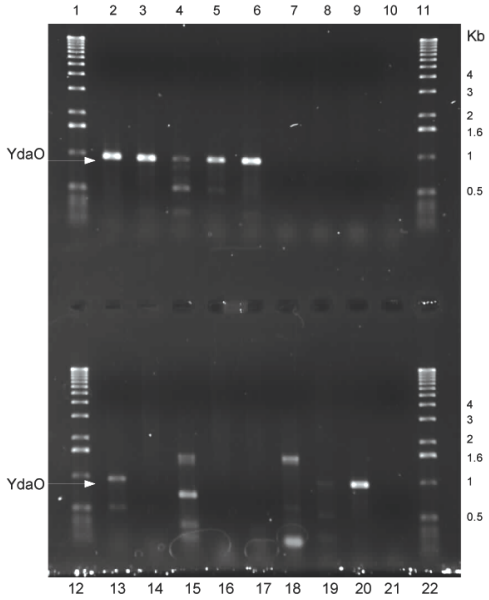

Figure 3.18: Amplified *ydaO* from environmental isolates from Bangladesh and K12 MG1655 using primers ydaOAI/ydaOdII. Lanes 1, 11, 12 and 22; 1Kb DNA ladder, lanes 2-10, and 13-20; amplified *ydaO* from various isolates, lane 21; amplified *ydaO* from *E. coli* K12 MG1655.

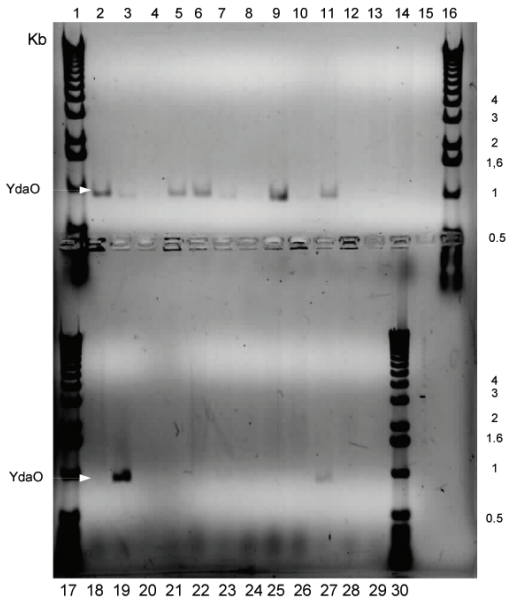

Figure 3.19: Amplified *ydaO* from environmental isolates from Bangladesh using primers ydaO-for. and ydaO-rev. Lanes 1, 16, 17 and 30; 1 Kb DNA ladder, lanes 2-15 and 18-29; amplified *ydaO* from various isolates.

3.5 Sequence and structure analysis of *E. coli* YdaO and Archaeal YdaO homologs

The protein sequences of TilS, YdaO, AF1595, AF1321 and Sso0586 were aligned using ClustalX and manually adjusted in GeneDoc. The conserved regions are marked in red and the less conserved regions are marked in blue. Figure 3.20 shows a more conserved N-terminal with the conserved PP-loop domain and residue R160 (marked in red). In the C-terminal there are only a few semi-conserved regions (marked in blue).

The conserved PP-loop and residue R150 from were visualised on a 3D-structure model of TilS downloaded from the Protein Data Bank (PDB). The conserved domains are marked in yellow (figure 3.21)

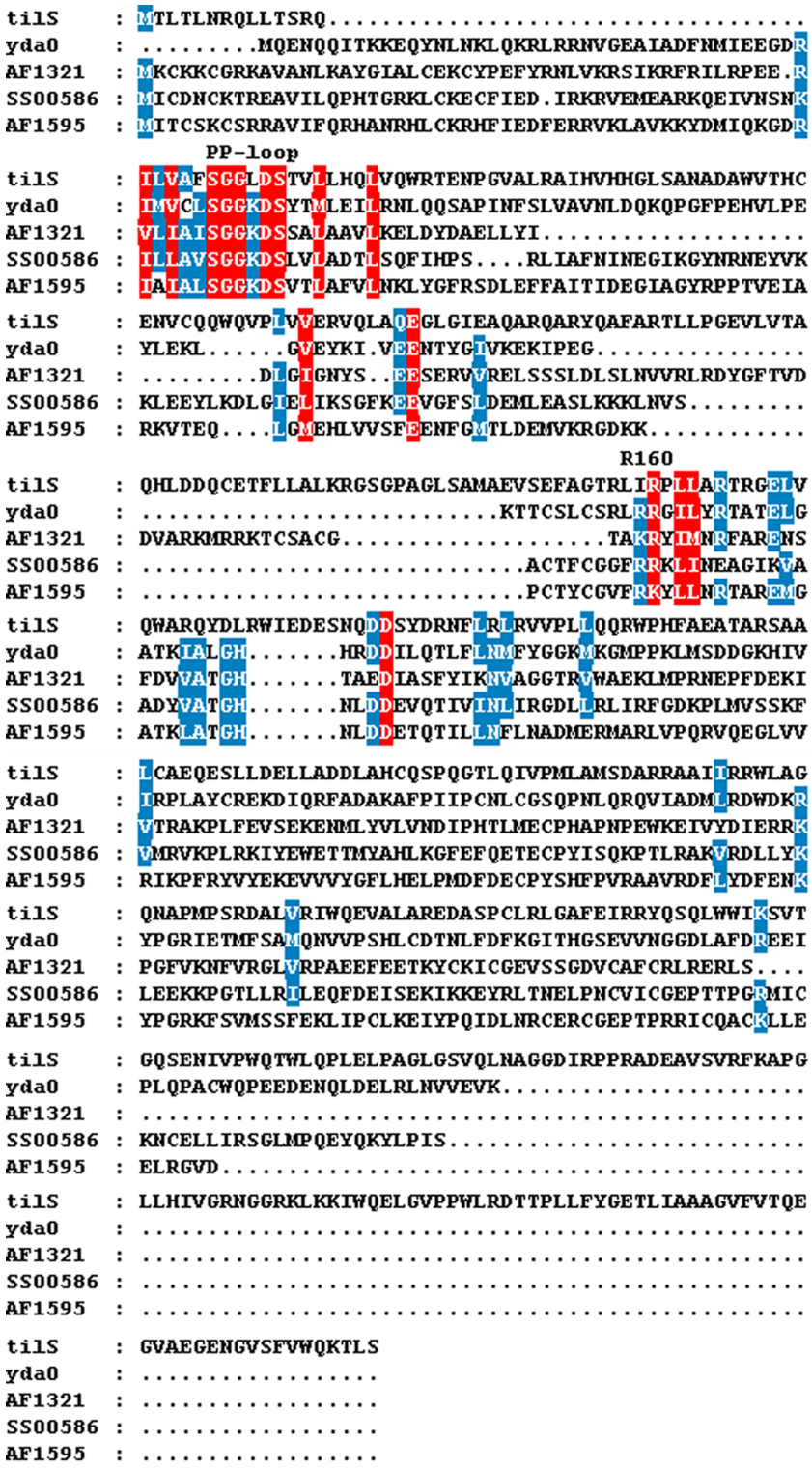

Figure 3.20: A multiple alignment of the MesJ homologs TilS, YdaO, Af1595, AF1321 and Sso0586, showing conserved domains in red and semi-conserved regions in blue. The sequences were aligned using ClustalX and manually adjusted in GeneDoc.

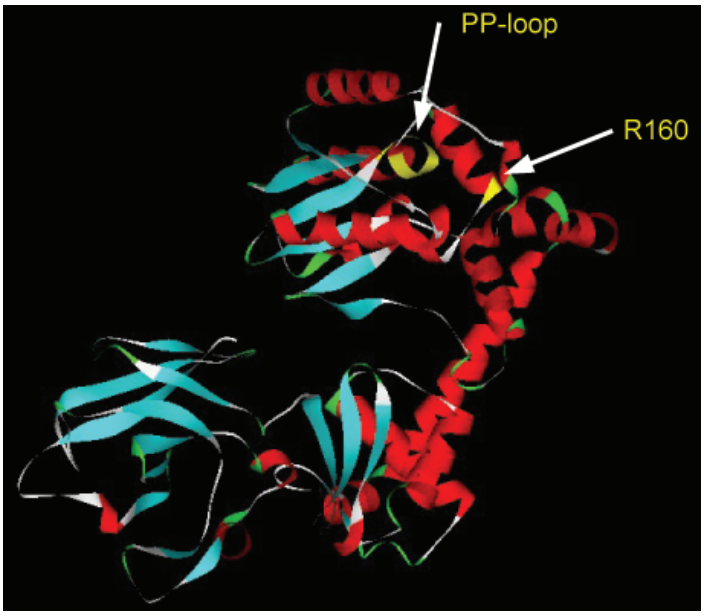

Figure 3.21: The crystal structure of *E. coli* TilS. A ribbon diagram showing the conserved PP-loop and residue R160 marked in yellow. the coordinates of *E. coli* TilS (1NI5) were obtained from the Protein Databank and visualised using DS ViewerPro 6.0.

Chapter 4

Discussion

With the genome sequencing project came vast amount of information about many organisms, but a decade later the function of one third of all the genes sequenced so far still remains unknown. As a result of all this new information new methods of identification were developed which sort genes based on sequence similarity between paralogue and orthologue genes [8]. Paralogy refers to the relationship between genes that have originated by gene duplication, whereas orthodoxy describes genes that originated by speciation [50]. One method of identification called Clusters of Orthologous Groups (COG), group genes on the basis of sequence similarity where all groups containing at least three proteins from distant genomes are assumed to belong to the same orthologuos group [9]. At the start of this project all, but one of the COGs containing universally expressed genes had been assigned a function. In the last group, COG0037, the genes were, based on conserved motifs, suggested to be ATPases. A phylogenetic analysis of COG0037 by Stranden (2002), showed that the genes in COG0037 form three main groups; Bacteria (subdivided into Bacteria I and Bacteria II), Eukarya and Archaea (figure 1.2). The aim of this project was to gather more information on the

genes belonging to COG0037 using *E. coli* as a model organism.

4.1 Construction of conditional mutant

E. coli is represented in COG0037 with two open reading frames, *yaeN* and *ydaO*. Experiments performed by Stranden (2002) indicated that the gene *yaeN* is essential, whereas *ydaO* is non-essential. In order to assess the implications of up- and down regulation of the transcription of *yaeN* an attempt to construct a conditional mutant was made. By substituting the wild-type gene with a deleted copy on the chromosome the gene could be expressed from a vector with an inducible promotor. The pBAD/HisA plasmid which is based on the pBAD promotor and AraC-regulator allows a tightly controlled regulation of the inserted gene by the L-arabinose inducer [51, 52]. This system allows a rapid up- and down regulation (induction after 1 min and repression after 5 min), and a dimming-effect of the expression can be achieved by adding different amounts of L-arabinose [52]. The AraC/pBAD system has been shown in various studies to be suitable for analysis of essential/non-essential genes [53, 52]. The substitution of the wild-type *yaeN* with a deleted copy was not successful as the cells were not viable with a temporary integrated substitution plasmid on the chromosome. Stranden (2002) also reported on this problem when attempting to construct a conditional mutant. The reason for this is still unclear as previous constructions of a deletion mutant and silent mutations were possible [13]. As the transformed clones failed to grow at the non-permissive temperature it is reasonable to assume that the substitution of the wild-type with the deleted copy by homologous recombination had failed and the vector had been lost and that the colonies observed at 30°C might be the result of spontaneous mutations resulting in

an antibiotic resistance. This assumption was backed-up by the absence of the plasmid on an agarose gel. The reasons for the failure to substitute the two genes remains unclear. In a second attempt to create a conditional mutant *E. coli* was transformed with both pKO3yaeN2- and pBADMesJ vectors simultaneously. The idea here was to delete the chromosomal *yaeN*, whilst expressing the protein from the pBAD expression vector. Since pKO3yaeN2- contains only the flanking regions of *yaeN*, and pBADMesJ contains only the *yaeN* sequence there should be no problems with homologous recombination between the two plasmids. Although this experiments seemed to work better, the failure of the transformed clones to grow in the presence of ampicillin, but not chloramphenicol suggest that the pKO3 substitution vector had been lost. As discussed by Stranden (2002) the expression vector and the substitution vector contains different ori (origin of replication): pBAD/HisA; pMB1-replicon (pUC-based ori) and pKO3; pSC101-replicon they should be able to co-exist in one cell. One possible explanation why this problem arises could be that the two vectors start competing for the cellular components responsible for the recombination. This type of competition has previously been seen when the RecBCD function in *E. coli* cells was replaced with the Red-system from bacteriophage λ before transformation with linear DNA [54]. The RecBCD-function is a multifunctional enzyme essential for homologous recombination in *E. coli* by non-specifically degrading the linear doublestranded DNA during the process of unwinding the DNA [55, 56]. Due to the difficulties with the pKO3 vector an attempt to insert the *yaeN* gene into the araBAD operon of the pBAD vector was made. By creating a fusion product containing *yaeN* with *araC* and *araB* up- and downstream, respectively, the idea was to insert this product into the araBAD operon of the *E. coli* chromosome and control the transcription of the gene by the addition

of arabinose in the medium. However, due to technical difficulties it was not possible to fuse the products together in a correct manner and the attempts to create a conditional mutant was terminated. With the publication of the function of the gene *yaeN* (renamed *tilS*) by Soma *et al.* (September, 2003) a decision was made to end the work on *yaeN* altogether and focus on *ydaO*.

4.2 Sequence and structure analysis of *E. coli* YdaO and Archaeal YdaO homologues

As stated above, homologues to the MesJ protein are present in all organisms in all three domains of life. While genes diversify throughout evolution, regions that are important for the function of the proteins are conserved. A sequence analysis of YdaO, AF1595, AF1321 and Sso0586 revealed 4 conserved motifs. As YdaO has been proven to be non-essential in *E. coli* [13] it was of interest to investigate whether the sequence varied between different *E. coli* strains, and also to determine whether this gene could be used as a signature sequence for *E. coli*. *ydaO* from 32 environmental strains isolated from various sites in Bangladesh were amplified. Out of the 32 strains only 16 gave a product when using primers that annealed to flanking regions. The absence of a PCR product could be a result of non-optimal conditions for the PCR or that *ydaO* and/or it's flanking regions were deleted from the genome. A hybridisation experiment has shown that there are deletions of genes flanking *ydaO*, such as *ydaO, ydaL, ydaM and ydaN*, in *E. coli* K-12 MG1655 [57], and that the gene *dbpA*, which is an immediate neighbour of *ydaO* have shown variable detection results. As these genes are present in the complete genome sequence for *E. coli*K-12 MG1655 any deletions must have taken place after the sequence was published in 1997. As the lack of PCR

products from some of the strains could be a result of deletion of the flanking regions of *ydaO* a second set of primers were used. These primers amplify only the *ydaO* open reading frame from the start to the stop codon. If a region flanking *ydaO* was deleted, but *ydaO* was still present, these primers should produce a product. Out of 23 strains 10 gave a product with the new primers. One of these products had not produced a product with the old primer set. This could be due to either experimental variations or the fact that the site of primer annealing flanking *ydaO* were deleted. Some of the PCR products from the new primer set were sequenced, but only two of the sequencing reactions gave satisfactory results. Due to time restraint it was not possible to optimise the PCR protocol or re-sequence the PCR products, so the sequencing results obtained were excluded from the thesis.

A multiple alignment was performed using GeneDoc and conserved regions were located. Conserved regions have previously been reported with the PP-loop domain and residue R160 being particularly important for the function of the enzyme [22]. The PP-loop is known to participate in the binding and hydrolysis of the $\alpha-\beta$ phosphate bond of ATP pyrophosphatase (PPi synthetase) family such as GMP- and NAD synthetase and residue R160 is predicted to interact with the γ phosphate of ATP. Together these two motifs are thought to be an ATP binding site.

4.3 Characterisation of mesJ orthologs from *E. coli* and Archaea

As *ydaO* shows a much higher sense of homology to the archaeal genes in the COG0037 group, it was decided to include two archaeal organisms into the project, one with two MesJ orthologues and one with one orthologue. The

organisms chosen were *A. fulgidus*, represented in COG0037 with *AF1595* and *AF1321*, and *S. solfataricus*, represented with *Sso0586*. All four genes were cloned and expressed in *E. coli* BL21(DE3) before being tested for solubility. Out of all four proteins only YdaO proved to be soluble. There are several factors that could contribute to the archaeal proteins being insoluble. Expression of most non-bacterial, recombinant proteins in bacteria often result in insoluble non-functional inclusion bodies, especially at high translation rates. Inclusion bodies occur by deposition of misfolded or partially folded polypeptides through the construction of hydrophobic patches and the consequent intermolecular interactions [58]. Bacterial chaperons, which are involved in conformational processing during heat-shock or other conformational stresses, can assist recombinant proteins ([59] and references therein). Some of these chaperons aid in the folding process whereas others prevent protein aggregation. However, as chaperons are limiting in bacterial cells, the co-expression of selected chaperone-encoding genes together with the recombinant protein could result in improved folding and enhanced solubility [60]. This approach has, however, produced irregular results, suggesting that the chaperons might be limiting for different polypeptides. The aggregation of proteins can also be controlled to a certain extent by altering the incubation conditions, such as temperature [61] or through reduced recombinant expression rate [62]. Incubation temperatures around 30°C have also shown to increase the activity and expression of chaperons [63]. Both a lower incubation temperature and a reduced concentration of IPTG were applied in an attempt to decrease aggregation, but the proteins remained insoluble. Several alterations to the method could be done in order to improve protein solubility. As a reduction in cellular protein concentration favours folding, an even lower incubation temperature could be applied. The co-

expression of the target protein together with chaperons Cpn60 and Cpn10 from the psychrophilic bacterium *Oleispira antartica* allows for protein expression and folding at 4°C [63]. Using this procedure the specific activity of the purified protein increased by 180 fold compared to enzymes prepared at 37°C. Also numerous specialised host strains have been developed to overcome the difficulties related to high level protein expression [64]. Amongst these are two mutant *E. coli* strains C43(DE3) and C41(DE3), which derives from the BL21(DE3) strain. These mutant strains allow over-expression of some globular and membrane proteins unable to be expresses in the parent strain. As the archaeal proteins chosen for this work normally are exposed to much higher temperatures, heat-stimulation of the proteins after lysation of the cells was therefore applied. Again the proteins precipitated proving the proteins to be insoluble. Although much work has been done in order to improve the solubility of proteins, the 'trial and error' method is still the only way of reaching the goal.

Since YdaO proved to be soluble, further analysis on this protein was performed. As the homologues *E. coli* TilS protein had been shown to be an RNA-modification enzyme with ATPase activity [22], YdaO was also tested for enzyme activity. An ATPase activity assay was performed using radioactively labeled ATP. The hydrolysis of ATP proved YdaO to be very active, suggesting that it is a functional protein. The graph shows that the hydrolysis of ATP is linear up to 500 ng protein added. The dissociation of PO_4^{2-} from ADP/ATP also supports the identification of the conserved PP-loop and R160 by the multiple alignment 3.20.

(Stranden, 2002) showed that YdaO is non-essential. It was therefore of interest to see whether this protein was expressed under normal conditions. However, it was not possible to obtain reliable results to verify whether this

protein is expressed, hence this question remains unanswered.

4.4 Conclusion

In conclusion, the results obtained in this work show that the *E. coli* protein YdaO, which is a member of COG0037, is a functional enzyme with ATPase activity. It was, however, not possible to verify whether this enzyme is expressed under normal conditions. Due to the insolubility of the archaeal orthologues, it was not possible to determine whether these proteins are functional, or to say anything about their function. However, based on the conserved PP-loop motif identified in the multiple alignment, it is reasonable to suggest that if functional, these proteins also possess ATPase activity.

4.5 Further work

Further work should include testing the TilS homologues for similar function as TilS. This includes determination of which bases are modified by these enzymes and which amino acid they use as substrate. It would also be important to verify which corresponding bases the modified bases pair with. The determinination of the structure of the TilS homologues by crystallisation and a verification of the conditions under which these proteins are expressed would also be of great interest.

Bibliography

[1] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. Complete nucleotide sequence of bacteriophage ms2 rna: primary and secondary structure of the replicase gene. *Nature*, 260(5551):500–7.

[2] F. Sanger, A.R. Coulsen, T. Friedman, G.M. Air, B.G. Barrel, N.L. Brown, J.C. Fiddes, C.A. Hutchison, P.M. Slocombe, and M. Smith. The nucleotide sequence of ϕx174. *J. Mol. Biol*, 162:729–773, 1978.

[3] R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* rd. *Science*, 269(5223):496–512, 1995.

[4] F. R. Blattner, 3rd Plunkett, G., C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* k-12. *Science*, 277(5331):1453–74, 1997.

[5] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, A. Danchin, and et al. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657):249–56, 1997.

[6] C.M. Fraser, J.A. Eisen, and S.L. Salzberg. Microbial genome sequencing. *Nature*, 406:799–803, 2000.

[7] D.A. Relman and S. Falkow. The meaning and impact of the human genome sequence for microbiology. *Trends in Microbiology*, 9:206–208, 2001.

[8] J.W. Thornton and R. DeSalle. Gene family evolution and homology: Genomics meet phylogenetics. *Annu. Rev. Genomics Hum Genet.*, 1:41–73, 2000.

[9] R.L. Tatusov, E.V. Koonin, and D.J. Lipman. A genomic perspective on protein families. *Science*, 278:631–637, 1997.

[10] W. M. Fitch. Distinguishing homologous from analogous proteins. *Syst Zool*, 19(2):99–113, 1970.

[11] E. L. Sonnhammer and E. V. Koonin. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet*, 18(12):619–20, 2002.

[12] J. K. Harris, S. T. Kelley, G. B. Spiegelman, and N. R. Pace. The genetic core of the universal ancestor. *Genome Res*, 13(3):407–12.

[13] I. Stranden. *Hovedfagsoppgave: Nye universelle genfunksjoner? Funksjonell analyse av en gen-gruppe med ukjent funksjon ved hjelp av Escherichia coli som modellsystem.* Institutt for Mikrobiologi, Universitetet i Bergen, 2002.

[14] E. A. Winzeler, D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Veronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, and R. W. Davis. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285(5429):901–6, 1999.

[15] C. A. Hutchison, S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter. Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, 286(5447):2165–9, 1999.

[16] M. Kail, E. Juttner, and D. Vaux. Lambda clone b22 contains a 7676 bp genomic fragment of saccharomyces cerevisiae chromosome vii spanning the vam7-spm2 intergenic region and containing three novel transcribed open reading frames. *Yeast*, 12(8):799–807, 1996.

[17] B. Lewin. *Genes.* Oxford University Press, fifth edition, 1994.

[18] K. Nakanishi, S. Fukai, Y. Ikeuchi, A. Soma, Y. Sekine, T. Suzuki, and O. Nureki. Structural basis for lysidine formation by atp pyrophosphatase accompanied by a lysine-specific loop and a trna-recognition domain. *Proc Natl Acad Sci U S A*, 102(21):7487–92, 2005.

[19] K. Nakanishi and O. Nureki. Recent progress of structural biology of trna processing and modification. *Mol Cells*, 19(2):157–66, 2005.

[20] T. Muramatsu, S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, and T. Miyazawa. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine trna from *Escherichia coli*. *J Biol Chem*, 263(19):9261–7, 1988.

[21] T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, and S. Yokoyama. Codon and amino-acid specificities of a transfer rna are both converted by a single post-transcriptional modification. *Nature*, 336(6195):179–81.

[22] A. Soma, Y. Ikeuchi, S. Kanemasa, K. Kobayashi, N. Ogasawara, T. Ote, J. Kato, K. Watanabe, Y. Sekine, and T. Suzuki. An rna-modifying enzyme that governs both the codon and amino acid specificities of isoleucine trna. *Molecular Cell*, 12:689–698, September 2003.

[23] H Grosjean and G.R Bjoerk. Enzymatic conversion of cytidine to lysidine in anticodon of bacterial trna^{Ile} an alternative way of rna editing. *Trends in Biochemical Sciences*, 29(4):165–168, April 2004.

[24] K. Nakanishi, S. Fukai, Y. Ikeuchi, A. Soma, Y. Sekine, T. Suzuki, and O. Nureki. Structural basis for lysidine formation by atp pyrophos-

phatase accompanied by a lysine-specific loop and a trna-recognition domain. *Proc Natl Acad Sci U S A*, 102(21):7487–92, 2005.

[25] C. Gustafsson, S. Govindarajan, and J. Minshull. Codon bias and heterologous protein expression. *Trends Biotechnol*, 22(7):346–53, 2004.

[26] Y. Nakamura, T. Gojobori, and T. Ikemura. Codon usage tabulated from international dna sequence databases: status for the year 2000. *Nucleic Acids Res*, 28(1):292, 2000.

[27] H. De Greve, C. Qizhi, F. Deboeck, and J. P. Hernalsteens. The shiga-toxin vt2-encoding bacteriophage varphi297 integrates at a distinct position in the *Escherichia coli* genome. *Biochim Biophys Acta*, 1579(2-3):196–202, 2002.

[28] H. Schmidt, J. Scheef, C. Janetzki-Mittmann, M. Datz, and H. Karch. An ilex trna gene is located close to the shiga toxin ii operon in enterohemorrhagic *Escherichia coli* o157 and non-o157 strains. *FEMS Microbiol Lett*, 149(1):39–44, 1997.

[29] N. Kanjo and H. Inokuchi. Genes for trna(arg) located in the upstream region of the shiga toxin ii operon in enterohemorrhagic *Escherichia coli* o157:h7. *DNA Res*, 6(1):71–3, 1999.

[30] J. Hacker, L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb Pathog*, 8(3):213–25, 1990.

[31] G. Blum, M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker. Excision of large dna regions termed pathogenicity islands

from trna-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun*, 62(2):606–14, 1994.

[32] S. Knapp, J. Hacker, T. Jarchau, and W. Goebel. Large, unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* o6 strain 536. *J Bacteriol*, 168(1):22–30, 1986.

[33] M. Susa, B. Kreft, G. Wasenauer, A. Ritter, J. Hacker, and R. Marre. Influence of cloned trna genes from a uropathogenic *Escherichia coli* strain on adherence to primary human renal tubular epithelial cells and nephropathogenicity in rats. *Infect Immun*, 64(12):5390–4, 1996.

[34] A. Ritter, D. L. Gally, P. B. Olsen, U. Dobrindt, A. Friedrich, P. Klemm, and J. Hacker. The pai-associated leux specific trna5(leu) affects type 1 fimbriation in pathogenic *Escherichia coli* by control of fimb recombinase expression. *Mol Microbiol*, 25(5):871–82, 1997.

[35] A. Ritter, G. Blum, L. Emody, M. Kerenyi, A. Bock, B. Neuhierl, W. Rabsch, F. Scheutz, and J. Hacker. trna genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. *Mol Microbiol*, 17(1):109–21, 1995.

[36] P. Bork and E. V. Koonin. A p-loop-like motif in a widespread atp pyrophosphatase domain: implications for the evolution of sequence motifs and enzyme activity. *Proteins*, 20(4):347–55, 1994.

[37] H. Mori, K. Isono, T. Horiuchi, and T. Miki. Functional genomics of *Escherichia coli* in japan. *Res Microbiol*, 151(2):121–8, 2000.

[38] A. J. Link, D. Phillips, and G. M. Church. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia*

coli: application to open reading frame characterization. *J Bacteriol*, 179(20):6228–37, 1997.

[39] C. Kato, R. Ohmiya, and T. Mizuno. A rapid method for disrupting genes in the *Escherichia coli* genome. *Biosci Biotechnol Biochem*, 62(9):1826–9, 1998.

[40] C. Freiberg, B. Wieland, F. Spaltmann, K. Ehlert, H. Brotz, and H. Labischinski. Identification of novel essential *Escherichia coli* genes conserved among pathogenic bacteria. *J Mol Microbiol Biotechnol*, 3(3):483–9, 2001.

[41] M. Riley and M.H. Serres. Interim report in genomics of *Escherichia coli*. *Ann. Rev. Microbiol*, 54:341–441, 2000.

[42] H. P. Klenk, R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. D'Andrea, C. Bowman, C. Fujii, S. A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, 390(6658):364–70, 1997.

[43] Q. She, R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost. The complete genome of the crenarchaeon *Sulfolobus solfataricus* p2. *Proc Natl Acad Sci U S A*, 98(14):7835–40, 2001.

[44] M.T. Madigan, J.M. Martinko, and J. Parker. *Brock Biology of Microorganisms*. Prentice Hall, tenth edition, 2003.

[45] A.J. Link, d. Phillips, and G.M. Church. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: Application to open reading frame characterization. *J. Bacteriology*, 179:6228–6237, 1997.

[46] B. Vogelstein and D. Gillespie. Citation classic - a rapid alkaline extraction procedure for screening recombinant plasmid dna. *Proc. Nat. Acad. Sci.*, 76:615–619, 1979.

[47] S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1):51–9, 1989.

[48] A. N. Warrens, M. D. Jones, and R. I. Lechler. Splicing by overlap extension by pcr using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene*, 186(1):29–35, 1997.

[49] R. Kasiviswanathan, J. H. Shin, E. Melamud, and Z. Kelman. Biochemical characterization of the *Methanothermobacter thermautotrophicus* minichromosome maintenance (mcm) helicase n-terminal domains. *J Biol Chem*, 279(27):28358–66, 2004.

[50] G. Theissen. Secret life of genes. *Nature*, 415(6873):741, 2002.

[51] N. Lee, C. Francklyn, and E. P. Hamilton. Arabinose-induced binding of arac protein to arai2 activates the arabad operon promoter. *Proc Natl Acad Sci U S A*, 84(24):8814–8, 1987.

[52] L. M. Guzman, D. Belin, M. J. Carson, and J. Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *J Bacteriol*, 177(14):4121–30, 1995.

[53] M. J. Carson, J. Barondess, and J. Beckwith. The ftsq protein of *Escherichia coli*: membrane topology, abundance, and cell division phenotypes due to overproduction and insertion mutations. *J Bacteriol*, 173(7):2187–95, 1991.

[54] K. C. Murphy. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol*, 180(8):2063–71, 1998.

[55] D. A. Dixon and S. C. Kowalczykowski. The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* recbcd enzyme. *Cell*, 73(1):87–96, 1993.

[56] D. A. Dixon and S. C. Kowalczykowski. Role of the *Escherichia coli* recombination hotspot, chi, in recabcd-dependent homologous pairing. *J Biol Chem*, 270(27):16360–70, 1995.

[57] C. Rosenow, R. M. Saxena, M. Durst, and T. R. Gingeras. Prokaryotic rna preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic Acids Res*, 29(22):E112, 2001.

[58] P. Valax and G. Georgiou. Molecular characterization of beta-lactamase inclusion bodies produced in *Escherichia coli*. 1. composition. *Biotechnol Prog*, 9(5):539–47, 1993.

[59] A. Villaverde and M. M. Carrio. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett*, 25(17):1385–95, 2003.

[60] J. G. Wall and A. Pluckthun. Effects of overexpressing folding modulators on the in vivo folding of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol*, 6(5):507–16, 1995.

[61] J. J. Chalmers, E. Kim, J. N. Telford, E. Y. Wong, W. C. Tacon, M. L. Shuler, and D. B. Wilson. Effects of temperature on *Escherichia coli* overproducing beta-lactamase or human epidermal growth factor. *Appl Environ Microbiol*, 56(1):104–11, 1990.

[62] C. A. Galloway, M. P. Sowden, and H. C. Smith. Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. *Biotechniques*, 34(3):524–6, 528, 530, 2003.

[63] M. Ferrer, T. N. Chernikova, M. M. Yakimov, P. N. Golyshin, and K. N. Timmis. Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat Biotechnol*, 21(11):1266–7, 2003.

[64] H. P. Sorensen and K. K. Mortensen. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact*, 4(1):1, 2005.