TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

LOHITH PATURU

Growth and gas-phase dependent transcription analysis of fermentation pathway in *Caloramator celer*

MASTER OF SCIENCE THESIS

Subject approved by Department Council
On 4th April, 2012

Examiners: Professor Matti Karp
Ph.D. Alessandro Ciranna

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Program in Science and Bioengineering
PATURU, LOHITH: Growth and gas-phase dependent transcription analysis of fermentation pathway in *Caloramator celer.*
Master of Science Thesis, 58 pages, 5 appendix pages
March 2015
Major: Biotechnology
Examiners: Professor Matti Karp, Ph.D. Alessandro Ciranna
Keywords: Hydrogen, gene expression, Hydrogenase, RT-PCR, transcription

Caloramator celer strain JW/YL-NZ35, formerly known as *Thermobrachium celere* is a unique alkalithermophilic organism that has the potential to solve energy problems in the future because of its hydrogen production mechanism. However in order to use this organism in industrial scale, more intrinsic understanding of the genetic and metabolic functioning of the organism is needed. The aim of this work is to undertake a transcriptional analysis of the growth and gas phase dependant fermentation pathway in this organism. This is done in order to better understand the genetic factors involved in hydrogen production and to monitor the pattern of gene expression in different physiological states of the cell. This transcription analysis was done using a Real time quantitative PCR that uses a fluorescence molecule for real time detection. This is a very sensitive process that gives reliable results. Two different housekeeping genes *recA* and *polC* were used as normalizing factors in order to calculate the relative gene expression and $2^{-\Delta\Delta C}{}_{T}$ method was used for this purpose. Through this analysis, the interdependence of formate, acetate and ethanol pathways with each other, and their combined effects on hydrogen production was studied. It was seen that formate and ethanol pathway is in direct competition with both the hydrogenases. Downregulation of hydrogenase genes because of ethanol and formate pathways was seen. It was also found that the cultures grown in artificially stressed gas phase media had a widespread downregultaion of its genes. The soluble metabolite was measured and it was correlated with the transcription data. Decrease of pH and increase of hydrogen partial pressure were the leading cause for the downregulation of *mbhL* and *hydA* hydrogenase genes. Increase in hydrogen partial pressure also led to increased ethanol production and *adhE* gene upregulation.

PREFACE

This thesis work has been conducted by Department of Chemistry and Bioengineering of Tampere University of Technology, Tampere, Finland.

Past four years has been one of the best periods in my life. The calmness of Finnish atmosphere helped me greatly in my study. So, I cannot start without thanking Tampere University of Technology and Finnish education system for providing a place of study without any Tuition fees. I will always remember the high quality educational environment and flexibility offered in course structure. Secondly, I would like to thank my Professor Matti Karp for providing me with the opportunity for thesis work in his lab and for his constant encouragement and support in my academic endeavors. And finally I am very grateful to my supervisor Alessandro Ciranna for his continued guidance in experiments, patient teaching, result analysis and thesis writing.

I dedicate this work to my parents, who have always been a strong pillar of moral and emotional support in completing my thesis.

Tampere,

Lohith Paturu

TABLE OF CONTENTS

LIST OF TERMS AND ABBREVIATIONS

dNTP	deoxyribonucleotide
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ssDNA	Single stranded DNA
dsDNA	Double stranded DNA
kDa	Kilodalton
FTIR	Fourier transform infrared spectroscopy
mRNA	Messenger RNA
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RT PCR	Reverse transcription polymerase chain reaction
cDNA	Complementary DNA
UV	Ultra violet
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
PFL	Pyruvate formate lyase
PFLae	Pyruvate formate lyase activating enzyme
FNOR	NADH:ferredoxin oxidoreductase
ADH	Alcohol dehydrogenase
ACK	Acetate kinase
CoA	Coenzyme A
PTA	Phosphotransacetylase
PFOR	Pyruvate oxidoreductase
Fd H_2ase	Ferredoxin-dependent NiFe hydrogenase
NADH H_2ase	NADH-dependent FeFe hydrogenase
O.D	Optical density
GC	Gas chromatography
HPLC	High performance liquid chromatography

1. INTRODUCTION

Regulation of a gene is of prime importance to all organisms. Our understanding of genetics has indeed come a long way since the Darwinian era. At that time when the idea of gene was still not conceived, scientists held to the notion that traits were passed on from parents to off springs in an entirely random manner by an unexplained mechanism. Then it was Gregor Johann Mendel who laid the foundation for modern genetics. His laws of inheritance gave a mathematical background for the hereditary phenomenon. Though he was not able to apply his laws to all systems, they were only a start. In the year 1902 Theodor Boveri and Walter Sutton in accordance with Mendelian laws proposed that chromosomes were responsible for carrying hereditary information in life forms [1, 2].

It was through Avery–MacLeod–McCarty experiment that was conducted in 1944 that proved that DNA was that genetic material [3]. Subsequently the double helical structure of the DNA was later revealed by Watson and Crick that ushered in a new era for genetics. With the structure of DNA deciphered, the template was used to explain a number of phenomenons. However this model was not sufficient in explaining the protein production and various other regulatory principles. To answer this, the now followed central dogma model was explained whose core idea was that the information stored in the DNA was transferred to another nucleic acid called RNA through a process called transcription. This RNA was then used as the template to produce a protein through the process called translation. As a result of this, a central system of Replication - Transcription – Translation was formed.

Within a few years, different types of RNA like ribosomal, messenger, transfer etc. were discovered. Their structure was deciphered and its subunits plotted. The role and relation of these RNA with each other was studied over the past 40 years. A series of well researched data gave a clear understanding of how these RNA functioned in order to produce a protein. In the year 1968 Nirenberg and Gobind Khorana gave the triplet codons of all twenty amino acid in terms of nucleotides for which they received the Nobel Prize [4]. During the course of time it was apparent that the transcription and the translation models between prokaryotes and eukaryotes had obvious deviations from each other. Both of these had a series of gene regulatory mechanisms with the

eukaryotic system having developed a more complex mechanism of gene regulation and post translational modification system. With this knowledge genetic engineering and metabolic engineering gained importance which has facilitated various breakthroughs in the field of medicine, agriculture and in manufacturing of organic components.

Biological applications aside, there is a growing sense of urgency for tackling the global energy problem. Our extensive dependency on fossil fuels has already cost the planet with accelerated global warming. Even though alternative energy sources like solar and wind have been known for decades, they could not be implemented on a popular scale because of their limiting cost factor. This energy problem can only be solved by a combination of multiple sustainable energy systems. Hydrogen has the potential to provide an energy solution because of its clean nature and high calorific value. Despite hydrogen being the most abundant element in Earth, its large scale use for energy generation is hampered because of difficulty in finding sufficient hydrogen in gaseous form. About 40% of global hydrogen production is from natural gas, 30% from oil sources, 18% from coal and only 1% from biomass [5].

As far as biological processes are concerned, there are photosynthetic and fermentative methods of hydrogen production. Of this, hydrogen production through dark fermentation is much simpler and also has higher hydrogen production rate than photosynthetic method [6]. Our organism of interest *Caloramator celer* is an alkalithermophilic organism which has many distinct advantages over its fellow biohydrogen producers. It's simple metabolic pathway coupled with its thermophilic nature makes it one of the most efficient producers of hydrogen. However there is still a long way to go before a large scale industrial output is feasible. Moreover, hydrogen partial pressure in the gas phase is one of the core limiting factors of hydrogen production by this bacteria. Therefore in order to improve upon the organism, metabolic and genetic engineering has to be employed. Through metabolic engineering, pathways that are problematic to hydrogen production can be blocked. Genetic engineering can also be employed to block pathways or introduce new ones.

However in order to do this, sound knowledge regarding the functioning of the pathways and the genetic elements involved is needed. Transcriptional analysis of the growth stage of the organism takes us one step closer to this understanding. Transcriptional analysis is basically the monitoring of levels of gene expression by the organism. The expression of different genes under different physiological conditions and stresses helps us in engineering a suitable complex. This research work aims to

understand the workings of this bacterium through transcription analysis. Moreover in order to better understand the role of hydrogen partial pressure in inhibiting hydrogen production, transcriptional analysis is done to monitor changes from a genetic standpoint. With better understanding of the pathways, the bacteria can be guided with characteristics such as resistance to pH drop, more tolerance to hydrogen partial pressure, resistance to toxic end products and capability to use a desired energy source instead of glucose.

2. THEORETICAL BACKGROUND

2.1 *Caloramator celer*

2.1.1 A brief description of *Caloramator celer*

Caloramator celer strain JW/YL-NZ35, formerly known as *Thermobrachium celere* (equivalent to ATCC 700318 and DSM 8682) [7], is a Gram-positive, strictly anaerobic and alkalithermophilic bacterium isolated from New Zealand in the Ohinemutu hot spring sediments region. *C. celer* has one of the lowest doubling times reported to be near 10 minutes [8, 9] when grown at a suitable optimal growth temperature of 67°C and an optimal pH of 8.2 at 67°C. *C. celer's* metabolic pathway goes through dark fermentation during which C6 sugars are converted to acetate, ethanol, formate, butyrate, hydrogen, and CO_2 as principal metabolites in which acetate occupies a major chunk of 78% of total aqueous metabolite composition during optimal hydrogen production [10]. It is also capable of producing significant quantities of ethanol and formate in modified growth conditions. A number of studies have shown the capability of *C. celer* to have excellent hydrogen production yield because of its thermophilic nature. The hydrogen producing characteristic was verified both in a mixed microbial community [11] and in pure culture [9, 10]. An increased yield of ethanol and formate can hamper the hydrogen production efficiency of *C. celer* as it taps into the reservoir of reduced cofactor pools (reduced ferredoxin and NADH) which will otherwise be used for hydrogen production by the hydrogenases. Therefore, when hydrogen is the desired end product, the accumulation of ethanol and formate has to be controlled [14]. To achieve this end metabolic and genetic engineering can be employed to manipulate the metabolic pathway. Transcriptional analysis of the genes involved in the pathway can shed more light regarding the level of expression of various genes during different stages of growth of the cell.

In terms of hydrogen production *C. celer* has a number of advantages when compared to other mesophilic counterparts because of its high temperature and high pH functioning. As a result of this contamination by hydrogen consuming methanogens and other growth competing organisms can be minimised [16] and the cost of sterilization of bio reactors can be reduced [11]. It was also found that thermophilic hydrogen producing bacteria in general had more H_2 yield when compared to its mesophilic counterparts since conversion of carbon sources into hydrogen is thermodynamically favourable which leads to higher efficiency at higher temperatures [17]. This can also indicate a selective hydrogen evolution by these species [11].

2.1.2 Genetic analysis of *C. celer*

The genome of *C. celer* was sequenced with Illumina HiSeq 2000 in order to obtain Paired end scripts from both long and as well as short fragment libraries. Also in order to get longer single end reads 454 sequencing was done. MIRA [18, 21] was used for the assembly of genome followed by scaffolding and contig extension using SSPACE [19, 21]. Annotation was done using RAST server [20, 21] followed by intermittent BLAST analysis. The total size of the genome assembly is 2,644,756 bp, organized in 56 scaffolds and 162 contigs. On the basis of annotation the G-C content of the genome is 31.3%. A total of 2,381 protein-coding sequences including 151 RNAs are present in the genome. Enzymes involved in the regeneration of NAD^+ and oxidized ferredoxin through proton reduction were found to be coded by three operons [21].

It was also found that the subunits of Fe-Fe hydrogenases had 47% to 67% identity with *Thermoanaerobacter tengcongensis* [21, 22] and the subunits of Ni-Fe hydrogenase show 30 to 54% identity to *Pyrococcus furiosus* [21, 23]. The primers needed for the transcriptional analysis in this organism were designed with reference to this draft genome.

2.1.3 Metabolic pathway of *C. celer*

2.1.3.1 Dark Fermentation

Dark fermentation is the process of anaerobic digestion of organic matter to bio hydrogen. Unlike photoremediation, dark fermentation does need the presence of light and it has been shown to have high hydrogen production rates when compared to photosynthetic methods [6]. It is an acidogenic process in which organic acids like acetate and formate are produced along with alcohols and gases like H_2 and CO_2 as end products [14]. There are various bacteria with a characteristic potential for hydrogen production ranging mesophilic to thermophilic bacteria. A wide arsenal of organic compounds like glucose, simple carbohydrates or complex polymers like starch and cellulose can be used to process dark fermentation. In recent times studies have been done to study the viability of using organic wastes and sludge for bio hydrogen production [5]. Hydrogen production by using cheap and largely abundant agro-industrial residues as substrate can offer distinct benefits in environmental and economic facets as shown in Figure 2.1 [24 - 27].

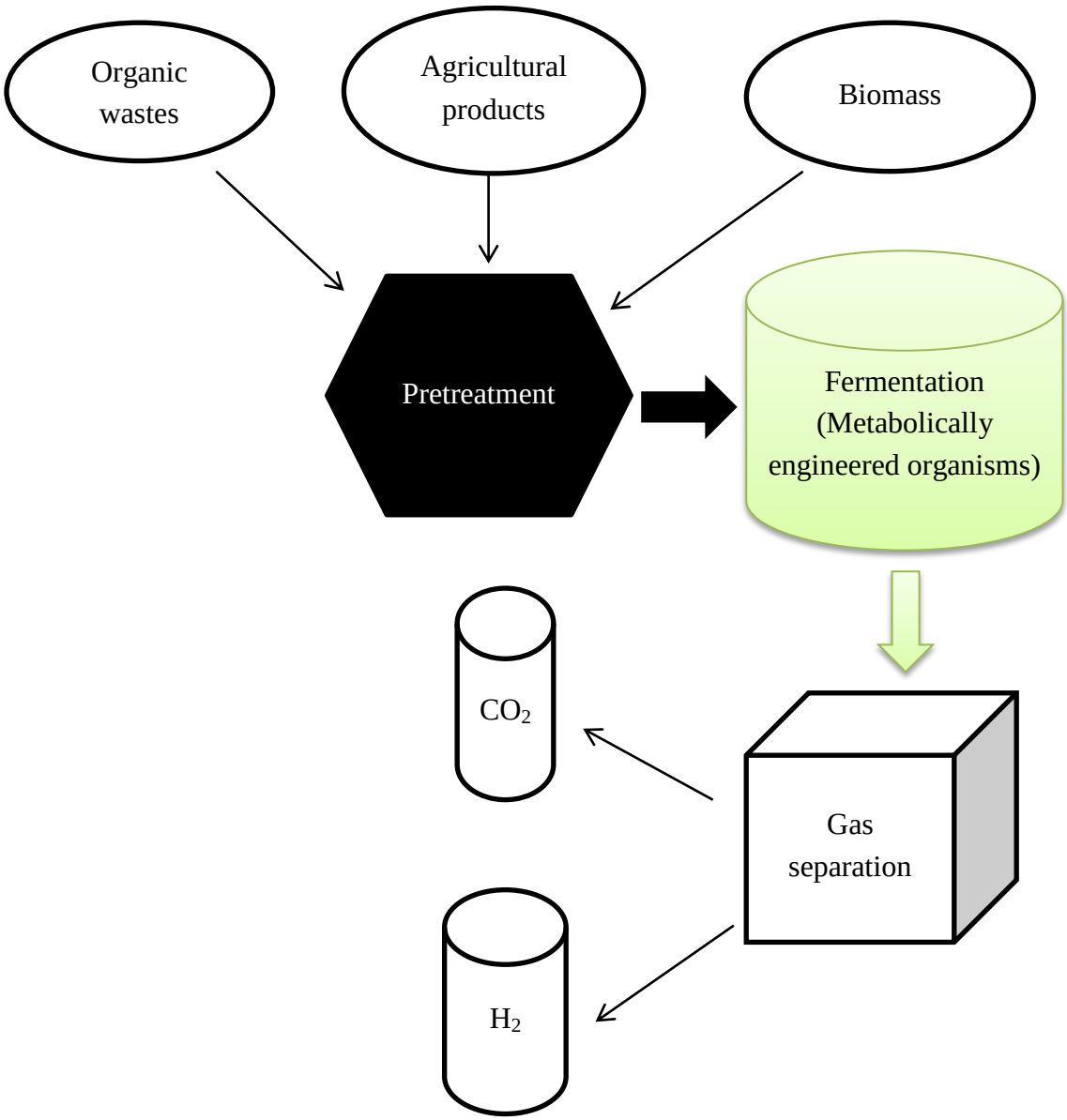

Figure 2.1. *An economically viable model of hydrogen production by dark fermentation as adapted from [28].*

Hydrogen produced through dark fermentation requires either facultative or strict anaerobes and in our case *C. celer* is strictly anaerobic. The metabolic pathway of this bacterium governs the production of acetate, formate, ethanol, hydrogen and CO_2 can be promoted or inhibited depending on the desired end product. An important enzyme of the metabolic pathway without which the biohydrogen process is not possible, is Hydrogenase. Presently three types Hydrogenases are involved in the hydrogen production process; nitrogenase, Fe-Fe hydrogenase, and NiFe hydrogenase. [28]. *C. celer* has only two types of hydrogen involved in its pathway; Ferredoxin-dependent NiFe hydrogenase and NADH-dependent FeFe hydrogenase. [21] These two hydrogenases are in active competition with ethanol and formate pathway because of their shared need for reduced cofactors as shown in Figure 2.2. Pyruvate formate lyase, acetate kinase and alcohol dehydrogenase are the enzymes responsible for the

production of formate, acetate and ethanol repectively. High substrate volumes need to be converted to hydrogen if an economically and environmentally sustainable hydrogen production process has to be developed.

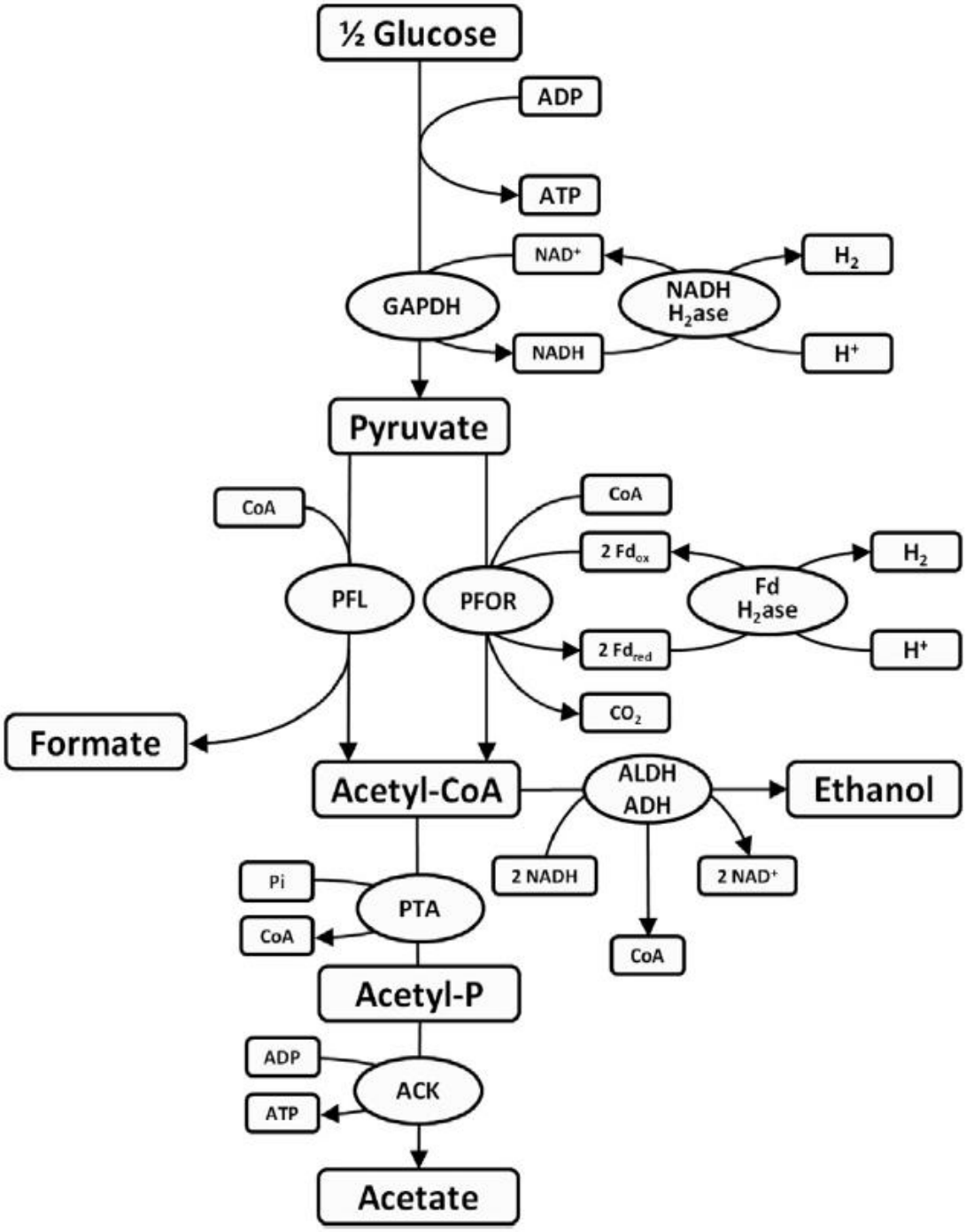

Figure 2.2. *The metabolic pathway of C. celer described in this diagram. The genes involved are shown here and the active competition from formate and ethanol pathways with hydrogenase enzymes for the reservoir of reduced cofactor pools is illustrated [14].*

2.1.3.2 FeFe – Hydrogenases

Hydrogenases are metallo-enzymes that facilitate hydrogen production in organisms [29, 30]. Both FeFe and NiFe hydrogenases have catalytic bi-nuclear metal centers in which a tholiage bridge facilitates the connection of one nickel with one iron or two iron with each other [30]. FTIR spectroscopy has shown that FeFe Hydrogenase from *Desulfovibrio desulfuricans* has $Fe(CO)_x$ unit in the active site where iron atoms also bind CN^- ligands as seen in Figure 2.3. [12, 13, 33].

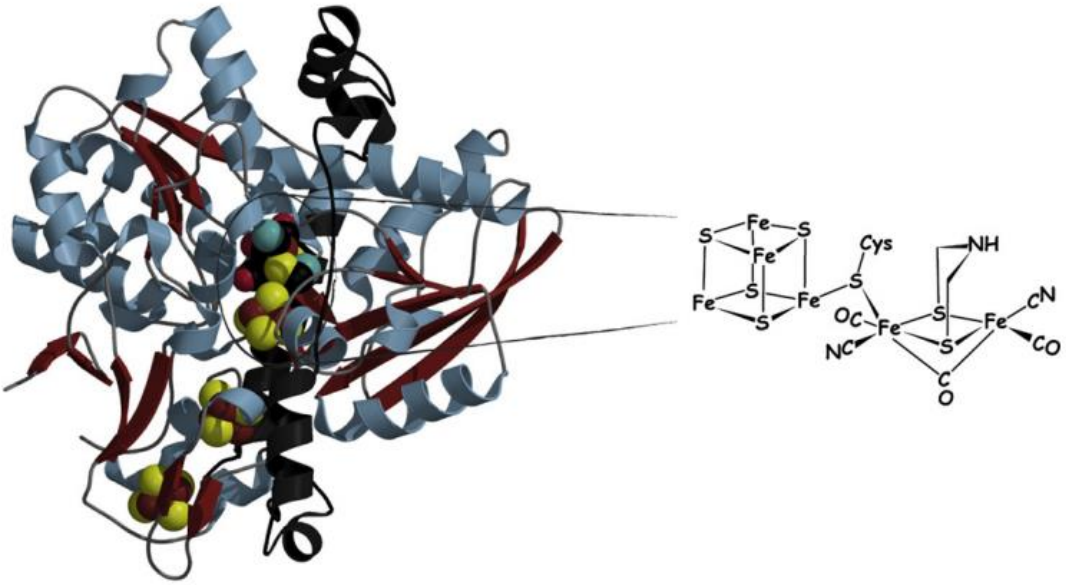

Figure 2.3. *Structure of the FeFe-hydrogenase and a depiction of the H-cluster in its active oxidized state [30].*

In all [FeFe]-hydrogenases, a classical [4Fe–4S] cluster is linked through a cysteine ligand with the dinuclear Fe center, leading to the formation of H-cluster, as shown in Fig. 2.3.[30]. In addition the cysteine ligand that manages the [4Fe–4S] sub cluster is also the only protein ligand to the binuclear sub cluster. It has also been established through infrared spectroscopic studies that one CO and one CN ligand is connected to each Fe atom which then connects to another Fe atom through extra CO ligand bridging [31 - 33]. Accessory ferredoxin like clusters ([4Fe–4S] and [2Fe–2S]) that act as electron transport centers by connecting the active site to the protein surface are also contained in the hydrogenase [30].

A complex post translational process is responsible for the formation of a functional hydrogenase protein. The process involves the formation of dithiolate, CO and CN bridging ligands, fabrication of di-Fe active site and its addition into the protein containing [4Fe-4S] element of the H-cluster and finally the assembly of the supplementary [Fe-S] domains [30]. The [FeFe]- hydrogenase polypeptide has to include the H-cluster and the accessory [Fe–S] clusters in order to be active after its

synthesis. During this assemble it is imperative that [4Fe-4S] element of the H cluster be assembled before the inclusion of di-Iron sub cluster into the hydrogenase [34]. When compared to NiFe hydrogenases, FeFe hydrogenases have superior efficiency in the production of molecular hydrogen [35].

2.1.3.3 NiFe – Hydrogenases

NiFe hydrogenases are probably the most widely found and they contain a NiFe active site. They are more oxygen tolerant when compared to that of the FeFe counterparts [15]. They are mainly found in prokaryotes and lesser eukaryotes. The smaller subunit is 28.8 kDA while the larger subunit is 62.5kDA when the [NiFe]-hydrogenase from *Desulfovibrio vulgaris* was studied [36]. The Fe-S cluster is also present in NiFe hydrogenases at a distance of 12 Å from the active site [43] and it coordinates the connection of nickel active site to the exterior of the protein complex as shown in Fig. 2.4 [15].

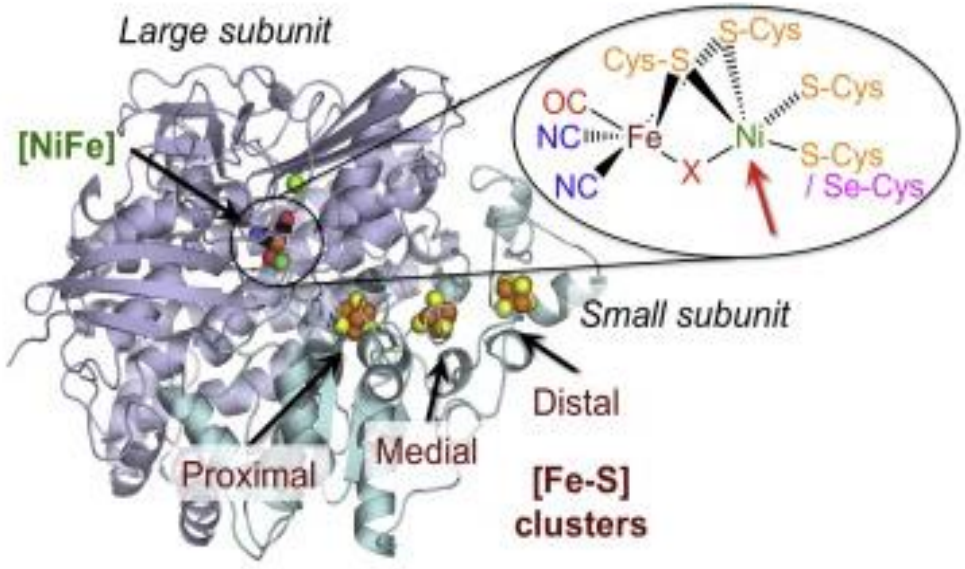

Figure 2.4. *The structure of Ni Fe hydrogenase consists of two subunits; large and small subunit. The metallo cluster active site is present in the large subunit, surrounded by three 4S domains from small subunit [15].*

2.1.3.4 Limiting factors of hydrogen production in *C. celer*

When aiming for an industrial scale application of biohydrogen production importance has to be given to intrinsic factors that affect the bioprocess. Hydrogen partial pressure in the gas phase is one of the core limiting factors of hydrogen production. In both mesophiles and thermophiles, it has been noticed that with increased H_2 production rates, the concentration of H_2 in the headspace and the liquid space also increases rapidly and as a result the reaction becomes thermodynamically unfavourable [37- 39, 10]. Following the metabolic pathway shifts towards the production of unfavourable end products [40, 9]. In addition to this a low hydrogen partial pressure is very crucial for the oxidation of NADH to NAD^+ which can ensure a higher H_2 production [41, 9].

pH is another limiting factor that can affect the efficiency of hydrogen production [42]. It has been observed in *C. celer* that during its prolonged dark fermentation there is an accumulation of volatile fatty acids which results in the drop of pH much below the optimal range [10].

Dark fermentation by *C. celer is an* acidogenic process which produces organic acids that are toxic to the bacterium. The cellular metabolism is impaired due to the inhibitory effects of the organic acids [44, 14]. The nonpolar undissociated form can release protons in the cytoplasm and increases the cellular upkeep energy [45, 14], whereas, the polar dissociated form increases the ionic strength of the medium and causes cell lysis [46] polar dissociated form. In addition to this the presence of increased concentrations alcohols like ethanol and butanol can denture biological molecules and exert pressure on the cell membrane [44, 14]. These compounds at subinhibitory concentrations can alter the distribution of carbon and electron flow [14].

Another limiting factor to hydrogen production in *C. celer* is related to the pathway concerned with ethanol and formate production. An increased yield of ethanol and formate can hamper the hydrogen production efficiency of *C. celer* as it taps into the reservoir of reduced cofactor pools (reduced ferredoxin and NADH) which will otherwise be used for hydrogen production by the hydrogenases. Therefore, when hydrogen is the desired end product, the accumulation of ethanol and formate has to be controlled [14]. To achieve his end metabolic and genetic engineering can be employed to manipulate the metabolic pathway. Transcriptional analysis of the genes involved in the pathway can shed more light regarding the level of expression of various genes during different stages of growth of the cell.

2.2 Overview of transcription

Transcription is the process in which a segment of DNA is copied into RNA to initiate the gene expression process. RNA polymerase is the principal enzyme that runs this process. During the transcription process the DNA sequence is read by the RNAP which produces a single stranded primary transcript or mRNA [47]. In case of prokaryotes there are three important steps to this process; initiation, elongation and termination.

Bacterial transcription as shown in Figure 2.5, starts with the initiation process when the RNA polymerase starts analyzing the promoter of a gene and binds to it [48]. RNA polymerase is a core enzyme consisting of five subunits. For a successful initiation the core enzyme has to be associated with the sigma factor to form a holoenzyme that helps in finding the correct -35 and -10 base pairs downstream of promoter sequences [49]. A number of base pairs leading from the -10 region are broken when the RNAP/promoter complex goes through a conformational change in order to form a bubble in which the two DNA strands have been split apart. This bubble is called the open complex and is usually 17 base pairs long. By using one DNA strand as a template the RNA synthesis is started by adding complementary base pairs to the template [47].

In the elongation phase more nucleotides are added to the expanding RNA chain as the RNA polymerase continues unzipping the DNA double helix as it moves along the template. As this process continues the open complex also moves. The average speed of this transcription is 40bp per second. 5' tail end of the growing RNA chain is separated from the base paired DNA template strand. After the initial 8bp are synthesized, the sigma unit detaches while the core enzyme synthesizes the RNA china alone [48].

This process of elongation continues until the RNAP encounters a code for termination. In prokaryotes there are two types of transcription termination; Rho independent and Rho dependent termination. This Rho independent termination also called the intrinsic termination takes place when the synthesized RNA transcript forms a hair pin loop that is G-C rich. However, as for Rho dependent termination as the name implies, is dependent on a particular protein called Rho factor. This Rho factor binds to the end of the newly synthesized RNA transcript and slides along its entire length until it reaches the open complex, where it cause the termination of transcription process [47]. This newly synthesized mRNA then undergoes translation in order to produce the proteins. Therefore when the gene expression is more, there is more mRNA concentration and when the gene expression is less there is less mRNA concentration. Using real time qPCR and RT PCR, the concentration of the mRNA can be determined during a particular growth which in turn lets us know the level of gene expression at different time points of bacterial growth stage.

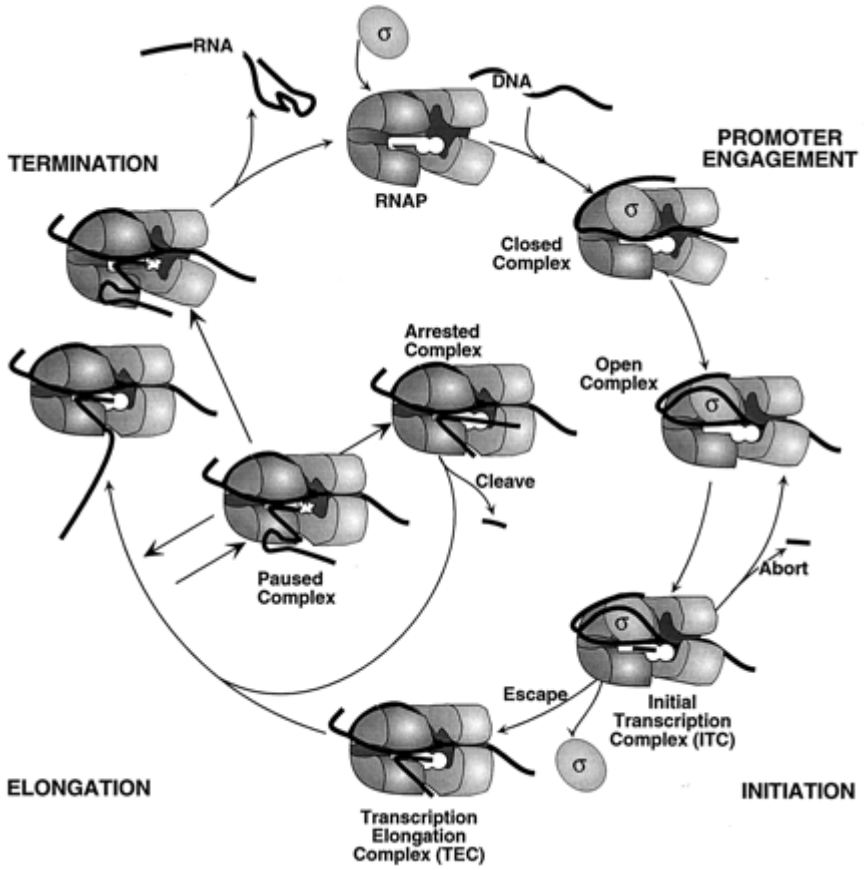

Figure 2.5. *The process of transcription beginning from the initiation phase to its end at the termination phase is shown. The mRNA produced here is translated to synthesize the corresponding protein [47].*

2.3 Transcriptional analysis

The mRNA from the transcription process has to be quantified in order to know the level of gene expression. In order to aid us in this process we make use of a variety of PCR techniques; qPCR and RT PCR.

2.3.1 Real time PCR

The real-time polymerase chain reaction (real-time PCR) which is a modified version of normal PCR was first introduced by Higuchi in 1992 and has since been extensively used in real time analysis [50, 51]. Real-time PCR uses a sensitive Fluorescent technology which allows it to precisely quantify target nucleic acids even if the initial concentration is very low. The amount of initial concentration determines speed at which the amplified target reaches a threshold detection level. Since its introduction real-time PCR has been extensively used various multi-disciplinary fields. Its applications cover quantifying viral load in patients [53], genotyping [52, 54, 55], and determining gene copy number in cancerous tissues [56, 57]. However, the most common use for this technology ¨ is to study the gene expression levels at specific time points by combining it with a procedure called reverse transcription [58].

Real-time PCR uses fluorescence-based technology to measure the quantity of amplicon produced during each cycle of amplification. Instead of measuring the quantity of amplicon at the end of the cycle Real-time PCR can find the amount of amplicon produced at the exponential phase of the PCR cycle itself [59]. The accumulating product in the PCR is labeled and monitored in real time through a fluorescently tagged substrate. Therefore in contrast to conventional PCR, real time PCR has many advantages like increased speed, no need for Post PCR agarose gel electrophoresis for detection of any amplification and higher sensitivity [60, 61].

SYBR® Green is a nonspecific fluorescent asymmetrical cyanine dye [62]. The emission maxima for SYBR Green 1 is around that of fluorescein (around 520 nm). The SYBR Green family of dyes follows the principle according to which they go through a 1000-fold increase in their fluorescence output [63] when they bind upon a dsDNA as shown in Fig. 2.6. As a result of this as the amount of dsDNA magnifies in the reaction mix, there will also be an increase in the fluorescent output that is detected by the machine as seen in Fig. 2.7 [59]. This can also prove to be problematic in unprepared scenarios as they do differentiate between different dsDNA strands. Therefore precautions have to be undertaken to prevent any contamination that can challenge the validity of the test [60].

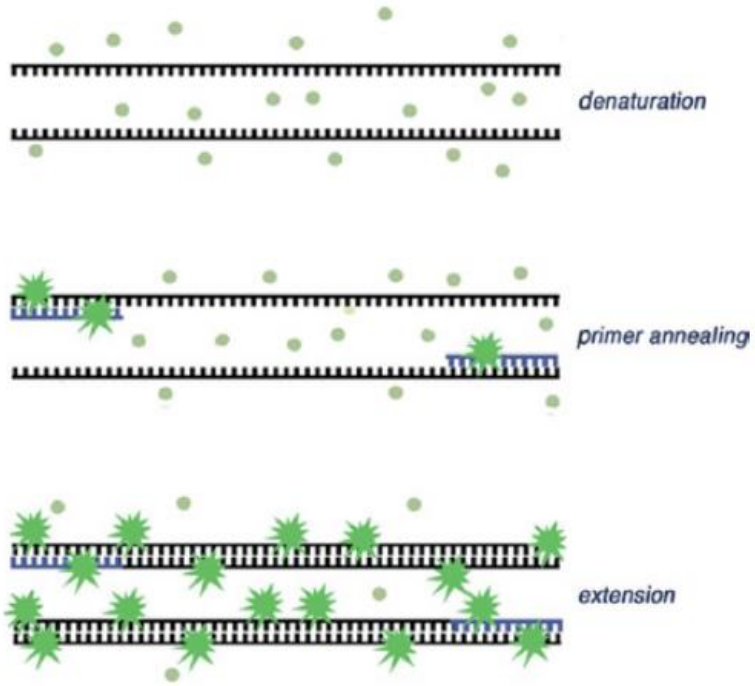

Figure 2.6. *SYBR green during the process of qPCR amplification in which the dots denote the SYBR green that is unattached to the minor groove of the dsDNA and therefore is unexcited to emit fluorescence. This dye attaches only to the minor groove of a dsDNA and then gives a 1000 fold increase in fluorescence which is shown as stars in this image[59].*

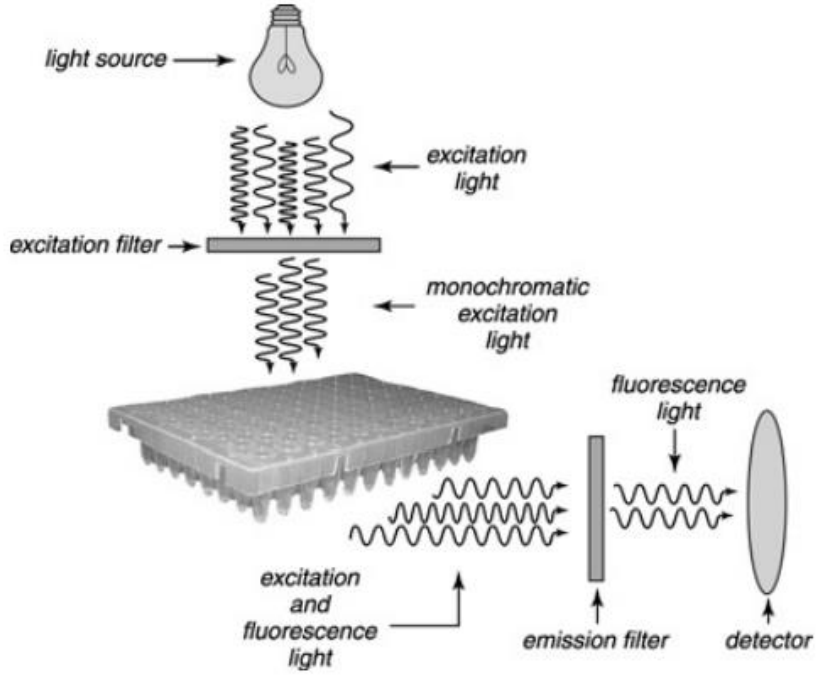

Figure 2.7. *The real time PCR along with the optical design for the detection of fluorescence, which contains an emission filter based system in which the measurement takes place from a micro-titration plate.[59].*

2.3.2 Reverse transcription PCR

The process of RNA to DNA transcription is called reverse transcription. Here a RNA-dependent DNA polymerase enzyme called reverse transcriptase is used to convert the RNA template into a cDNA which is then used as a template and amplified in a PCR. Reverse trancriptase enzyme was discovered from retro viruses which uses it for its replication [67]. Since the discovery of this process a number of uses have been attributed to this. Transcripts of practically any gene can be detected now [64], tolerance to RNA degradation [65], cancer detection in which the mRNA transcripts can be used as specific bio-markers [66]. RNA is less stable than DNA because of its susceptibility to RNases and its ubiquitous nature. Purified RNA is also vulnerable to impromptu degradation in solution [59]. Therefore the mRNA obtained from the RNA extraction process can be reverse transcribed to its complementary dsDNA which can be stored for a longer time without degradation.

Practically, there are many variations of reverse-transcription but in all; the prime step is the conversion of mRNA into a cDNA template through a reaction catalyzed by reverse-transcriptase as shown in (Fig .2.8). A DNA structure known as an oligodeoxynucleotide primer is essentially hybridized to a complementary mRNA that allows the RT enzyme to extend the primer and synthesize a complementary DNA strand. This oligodeoxynucleotide primer sequence can be engineered to bind to a specific target gene through a synthetic antisense oligonucleotide or to all of the mRNA in a sample through a universal primer. Either oligo (dT) primers or random hexamers can be used. However oligo (dT) primers are not suitable for bacterial mRNA because of absence of a stable poly A+ tail. Therefore in our case random hexamers are used to convert bacterial mRNA to cDNA [59] which requires that prior to the RT reaction, rRNA is completely removed from total RNA.

There are several reverse-transcriptase enzymes that are obtained from retroviruses [67] of which the most common are encoded by the avian myeloblastosis virus (AMV) and murine leukemia virus (MMLV). These RT enzymes have a special property called RNase H which monitors the RNA-DNA complex that arises during the reverse transcription reaction and then degrades the RNA. In order to circumvent this reverse transcriptases are genetically designed to lack RNase H activity. A one step RT-PCR process where, all the reagents and enzymes are added in a single vial is preferred as it reduces experimental variation.

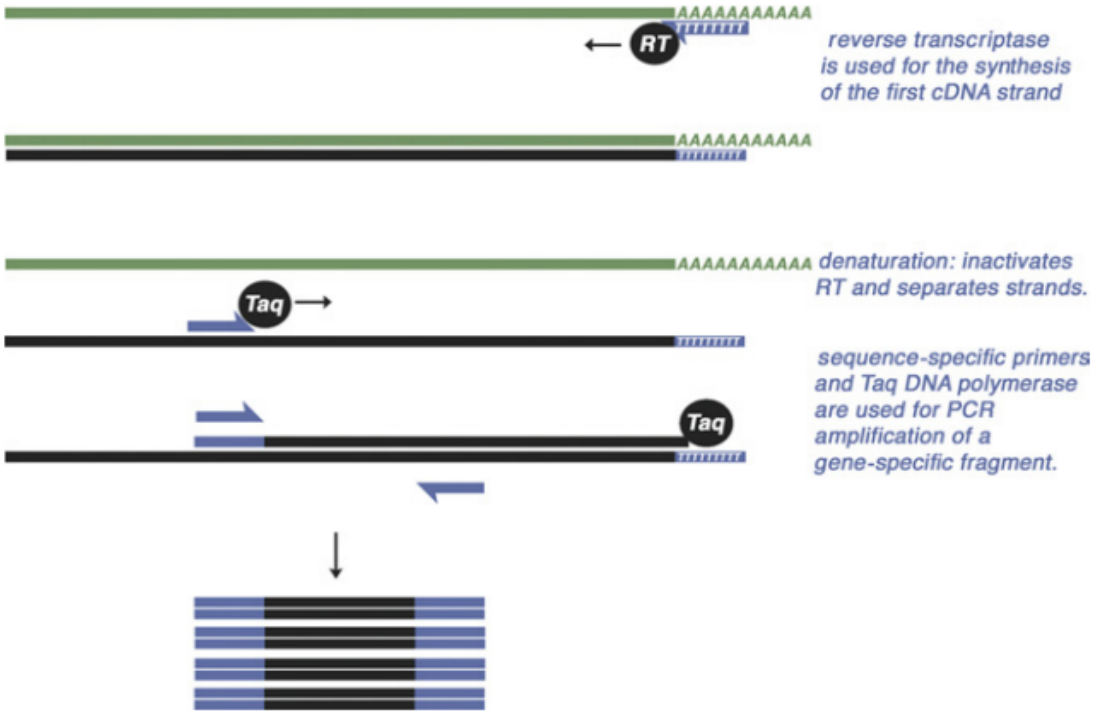

Figure 2.8. *Steps involved in the RT PCR process is shown here. mRNA is the gray line and DNA is black. Taq is a type of thermostable DNA polymerase and RT is the reverse transcriptase enzyme [59].*

2.3.3 Limiting factors to transcriptional analysis

Optimal primer design is very important in real time PCR in order to get reliable results. A Successful run for RT-PCR needs to have a primer with the ability to amplify a short product (<300 bp), that is explicit to the mRNA. A successful design requires the sequence of the interested genomic DNA and also the direct alignment with the cDNA coding sequence. During the design phase of primers care should be taken to eliminate the possibility of primer dimer structures because of the complementary sequence between the two primers or with the same [59]. The annealing temperature of the two primers should be matched. They should either flank a large intron or mount a small intron. An equal GC content between the two primers should be maintained. Once designed and run in a real time PCR the efficiency of these primers should be high and in an acceptable range to each other so the gene expression data can be compared with each other. The result can also be hampered if the primer concentrations are not optimally used. This can lead to the production of nonspecific products like primer

dimers. It is always desirable to use those concentrations that give the lowest C_T values [59].

Another limiting factor to the experimental validity of the real time PCR results is through DNA contamination. As we know about the extreme sensitivity of real time PCR setup because of its use of fluorescent mechanism for detection, even a very small amount of DNA presence can give a significant amplified signal at later stages [68]. As the gene is intrinsic to the genomic DNA, with its specific primer the gene in the DNA will get amplified whether it is expressed or not, which could give false expression in samples [69]. In order to avoid this, an effective Dnase treatment procedure has to be undertaken after the RNA extraction process. An RNase free DNase should be used for the process. RT^- control samples should be used after the RT PCR procedure in order to verify DNA contamination. Any amplification from the RT^- control samples before the 35^{th} cycle can be deemed as contamination by genomic DNA. The DNase has to be denatured by heat treatment after its use, so that there is no interference after the RT PCR process.

Similar to DNA contamination RNase contamination is also a crucial factor. As analysis of gene expression is the desired goal accurate initial concentrations of mRNA is very crucial to be maintained [68]. Contamination by RNase can lead to degradation of RNA which will represent a lower gene expression level than normal. RNases or ribonuleases are type nuclease that degrades RNA. Most organisms produce this as means of protection against viruses. They are very widespread and not as easily denatured as a DNase [70]. Special protocol is followed for RNase removal and all reagents and materials used has to be certified RNase free. And the RNA samples have to be stored at -80°C to avoid RNA degradation. Therefore it is imperative that the RT PCR process be initiated at the earliest.

2.4 Housekeeping genes as normalization factor

Housekeeping genes are constitutive in their function and are ubiquitously expressed in all the cells of an organism under normal physiological conditions [71]. They are responsible for the maintenance of basic cellular functions in a cell. The nature of gene expression studies especially those that rely on qPCR to provide supplementary analysis may run the risk of garnering insufficient data quality [72] because of variance in between samples, efficiency of reverse transcription, extraction of mRNA and finally in the PCR itself [73]. The technique of real-time RT-PCR has to be monitored at various stages because of problems or errors associated with RNA isolation procedure, sample storage, poorly selected primers, and inappropriate statistical backing [74]. It had

already been observed in multiple instances that even though uniform weight and volume in sample collection was achieved, there was no guarantee for presence of identical amount of matrix for the reaction setup [60]. This dilemma is further amplified when possibility of errors in pipetting and transferring of samples. To circumvent these issues a normalizing factor is needed in order to monitor the true level of expression of genes [60] and proper design of the experiment and use of replicates to give credibility to the data.

A reference gene can be used as a normalizing factor. However, a reliable reference gene has to meet several criteria [75, 60] like maintaining an uniform expression level by being unaffected by experimental factors and different physiological states of the organism, to have identical threshold value with the gene of interest, while also having the ability to show variability during differences in technology and protocols so that the object of research can also take in to account any variation in the quantity of genetic material (60). Housekeeping genes seem to qualify for these strict criteria because of their inherent function. It is also generally accepted that use two different housekeeping genes is important in order to avoid large errors associated with a single gene and also to increase the accuracy and resolution of the results (76).

During the selection of housekeeping genes for the target organism, the genome of the organism needs to be analyzed so that the sequence and structure of the interested genes be recognized for primer design to amplify the actual amount of mRNA both individually and in samples (77). Primers should also be designed for each possible variation of mRNA present in the bacterium while also neglecting fragment that contain gene polymorphism. The amplified product should also be tested to see if it is of specific size.

2.4.1 polC (DNA polymerase III gene) as housekeeping gene

DNA polymerase is an integral part of DNA replication process that helps to build DNA strand by assembling nucleotides. During the process of replication the DNA polymerase enzyme reads the old strand and creates new sister strands to match the existing ones. To add a copy of original DNA into the daughter cells during cell division, DNA polymerase is needed. When synthesizing new DNA, the new strand elongates only in the 5'-3' direction because DNA polymerase can add nucleotides only to the 3' end of the strand. No known DNA polymerase is capable of de novo synthesis of new chain.

There are a number of families of DNA polymerase (I to V) with varying domain structures [Figure 2.10], of which our gene is of DNA polymerase C family

(polC). Family C polymerases are the most common in prokaryotes and they form a holoenzyme complex as shown in Figure 2.9. The holoenzyme consists of three assemblies; the clamp-loading complex, pol III core unit and the beta sliding clamp processivity factor. The pol III core consists of three subunits α, θ, and δ, whose functions are polymerase activity, stabilization factor for δ and exonuleolytic proof reader. In order to permit for the processing of both lagging and leading strand, the holoenzyme contains two cores [78] along with duplicate beta sliding clamp [79].

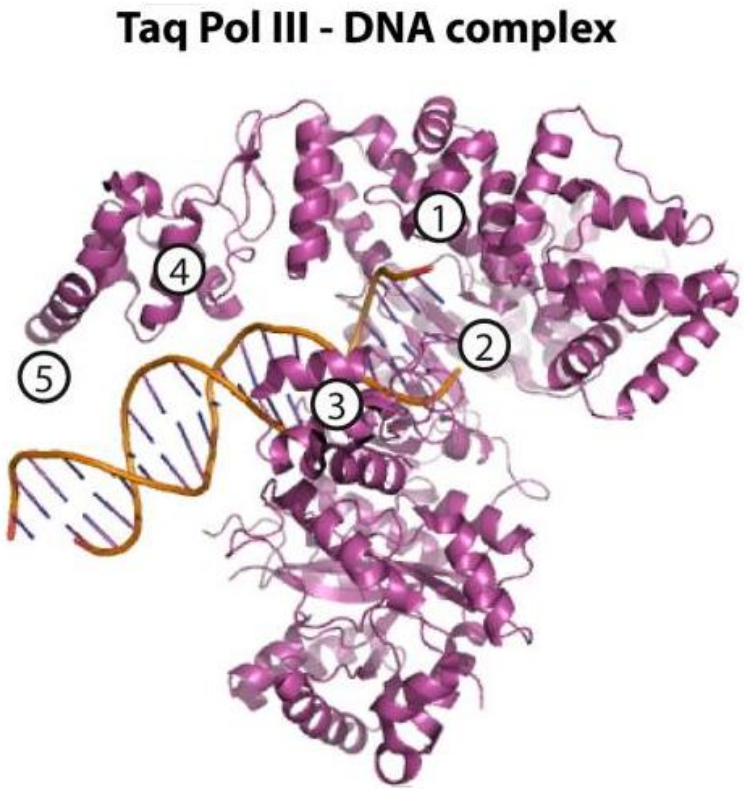

Figure 2.9. *A schematic diagram showing the clamp loading complex of Taq polymerase III along with all the subunits involved. The holoenzyme contains two cores along with duplicate beta sliding clamp. In the Figure 1 denotes single-stranded template DNA entering the Taq polymerase III which is guided by the fingers domain; polymerase active site is shown as 2; while the thumb domain comes into contact with the DNA minor groove as it exits the polymerase active site; 4 represents the elongated fingers domain which contacts the ds region of the DNA; and finally the binding site for the DNA sliding clamp is represented by 5. [87, 88]*

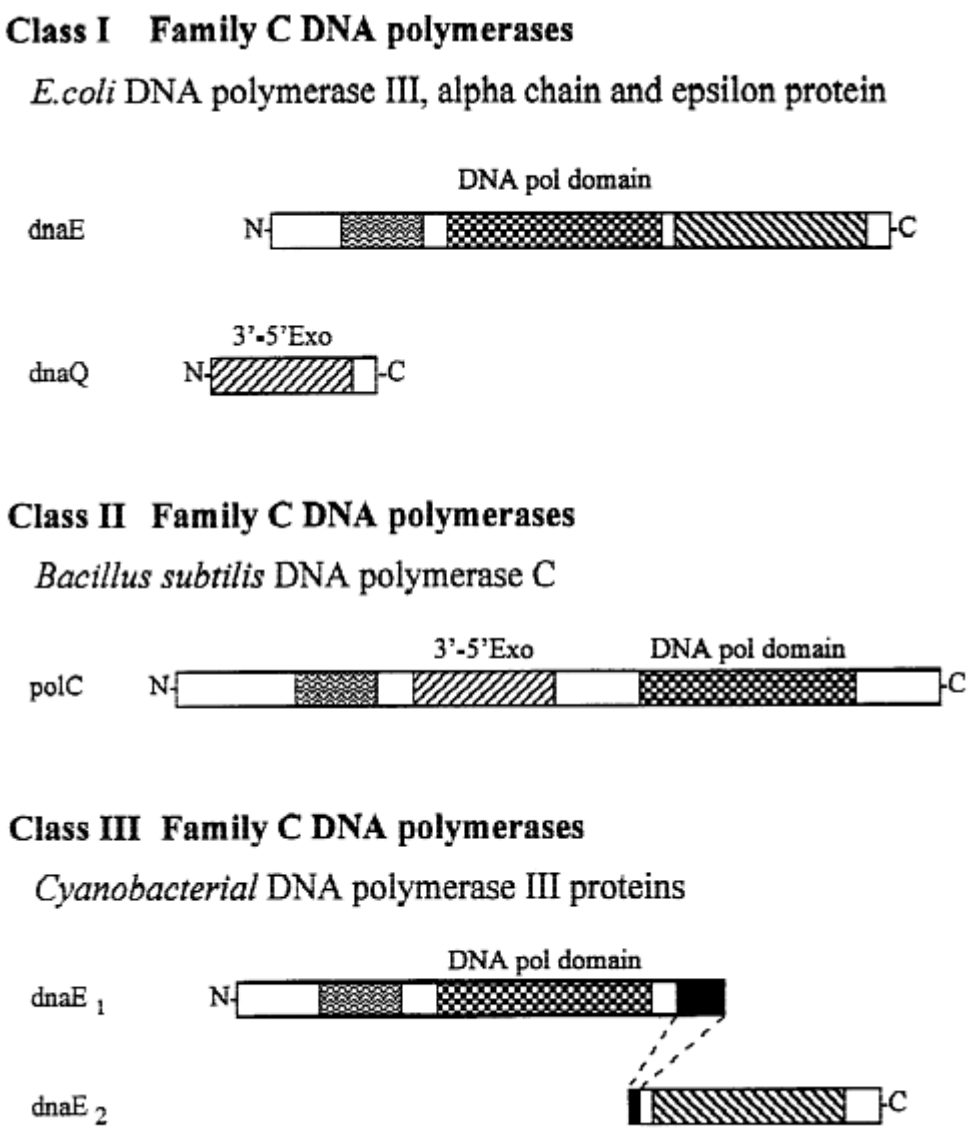

Figure 2.10. *Figure showing the domain structure of the three different classes of family C DNA polymerases. In class I depiction, the dnaE protein is α subunit and dnaQ protein is ε subunit of DNA polymerase III holoenzyme. The class II family consists of a 3' to 5' exonuclease domain. The two proteins coded by $dnaE_1$ and $dnaE_2$ in the class III family can pursue trans splicing to form a functional DNA polC. [80]*

The primary function of DNA polymerases is to manufacture an exact copy of the genome during the cell division process. Apart from this polymerization process, they are also found to be involved in the control of cell cycle [89- 91]. DNA repair is perhaps the next vital function of these enzymes. As the genome of the cell is repeatedly exposed to environmental and mutagenic factors, damage to DNA can occur through ways such as modification of DNA bases, destruction of strands, and backbone alterations. As a result of these unfavourable mutations the transcription process can be

disrupted thereby affecting the cellular functions. In order to avoid this DNA polymerases involve in multiple repair mechanisms [91, 92]. These functions stress the importance of DNA polymerase in a cellular system. As a result of their vital functions and their need for constant maintenance of genetic material, they act as housekeeping genes. A DNA polymerase gene is usually ubiquitously expressed in all physiological conditions of the cell.

2.4.2 recA as housekeeping gene

Bacterial DNA recombination protein (recA) is a maintenance protein whose primary function is DNA repair. This protein which belongs to a family of ATPases, manages homologous recombination in order to ensure genomic integrity and diversity [81]. The recA protein along with ATP attaches to a single stranded DNA (ssDNA) and conforms into a helical filament. This filament later attached to a double stranded DNA (dsDNA) and scans for presence of any homology. Once found it helps in the formation of a new heteroduplex by exchanging complementary strand [81].

The RecA protein has multiple DNA binding sites and can therefore independently hold either single stranded or double stranded DNA through which it can actuate a synapsis reaction between the complementary region of single stranded DNA and the DNA double helix [85, 86]. After this an exchange of strands among the two DNA double helices takes place. After this a process called branching migration starts in which an unpaired area of one single strand supersedes a paired region of the other single strand. In this movement only the branching point is moved while the total number of base pairs is kept constant. Unidirectional branch migration is catalysed by the recA through which complete recombination takes place leading to the formation of a 1000 bp long heteroduplex DNA. RecA which is a DNA-dependent ATPase contains a binding site for ATP [81] as shown in Figure 2.11. When ATP is bound to recA, the protein is more tightly associated with the DNA then when ADP is bound. In addition one ADP-aluminium fluoride-Mg (ADP-Alf_4-Mg) for each recA promoter is present in all DNA bound structures.

As a result of these functions, the cell's diversity and stability is assured. Due to the pervasive nature of this function recA is ever present all over the cell during multiple physiological stages. Therefore, recA is a strong example of housekeeping gene.

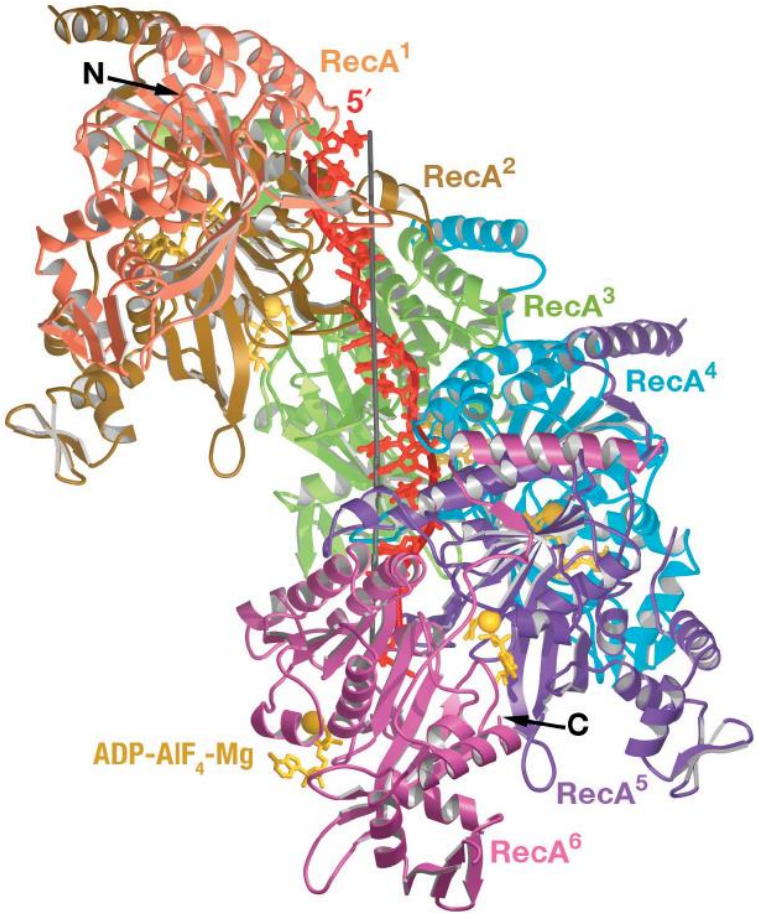

Figure 2.11. *Structure of the recA nucleoprotein filament ($RecA_6$–$(ADP\text{-}AlF_4\text{-}Mg)_6$–$(dT)_{18}$). Here dT_{18} is the 18- nucleotide oligo-deoxythymidine ssDNA. The DNA backbone is represented by the red filament in the center. The 6 different recA promoters are coloured and numbered. The ssDNA is numbered and it starts with 5' end. [81]*

2.5 $2^{-\Delta\Delta C}{}_T$ method of normalization

In order to reliably analyze the data from qPCR, two different methods exist; absolute quantification which determines input copy number of the transcript and relative quantification which shows the gene expression level relative to a reference group. In situations where an absolute copy number of the transcript is required, absolute quantification can be used [82]. However in our case relative quantification technique is sufficient because we are only interested in gene expression levels. In order to quantify the gene expression level through real time PCR certain assumptions and validation tests are required, which is provided by Applied biosystems user bulletin [82]. Using $2^{-\Delta\Delta C}{}_T$ method for analyzing gene expression data is already well established in previous studies [82- 84]. Nevertheless, it is imperative that we understand the derivation of 2^{-}

$^{\Delta\Delta C}{}_{T}$ method along with the assumptions involved, so that we understand all the factors involved and can use this method competently in our analysis. This derivation is obtained through literature by Kenneth *et al*., 2001

The exponential amplification of PCR is shown through the given below,

$$X_n = X_o \times (1 + E_X)^n$$

Where, X_n is the number of target molecules at nth cycle of reaction, X_o is the initial quantity of target molecules, E_X is the efficiency of amplification and n denotes the number of cycles.

$$X_T = X_o \times (1 + E_X)^{C_{T,X}} = K_X$$

Here C_T is the threshold number, X_T is the threshold number of target, K_X a constant and $C_{T,X}$ denotes threshold cycle of target amplification. Now a similar equation regarding the reference gene reaction is given below where, R is associated with the reference gene.

$$R_T = R_o \times (1 + E_R)^{C_{T,R}} = K_R$$

Now dividing X_T by R_T we get the following equation,

$$\frac{X_T}{R_T} = \frac{X_O \times (1+E_X)^{C_{T,X}}}{R_O \times (1+E_R)^{C_{T,R}}} = \frac{K_X}{K_R} = K$$

During real time amplification, X_T and R_T are influenced by the reporter dye used in the probe, along with characteristics like fluorescence properties of the probe, its purity, and efficiency of the probe cleavage. As result of this the constant K need not be equal to 1. Therefore now assuming that the efficiencies of both reference and target are same,

$$E_X = E_R = E,$$

$$\frac{X_O}{R_O} \times (1 + E)^{C_{T,X} - C_{T,R}} = K,$$

$$X_N \times (1 + E)^{\Delta C_T} = K,$$

Where, X_N denotes the normalized quantity of $\frac{X_O}{R_O}$ and ΔC_T is the difference between $(C_{T,X} - C_{T,R})$. Therefore after rearranging we get,

$$X_N = K \times (1 + E)^{-\Delta C_T}$$

In the final step, X_N for any sample q is divided by X_N for the calibrator cb.

$$\frac{X_{N.q}}{X_{N,cb}} = \frac{K\times(1+E)^{-\Delta C_{T,q}}}{K\times(1+E)^{-\Delta C_{T,cb}}} = (1 + E)^{-\Delta\Delta C_T}$$

Also here,

$$-\Delta\Delta C_T = -(\Delta C_T, q - \Delta C_T, cb).$$

Those templates designed to be less that 150bp and for which the Mg^{2+} concentrations and the primers are efficiently optimized, the efficiency is near 1. Therefore the quantity of target, normalized to a reference and relative to a calibrator is shown as,

Amount of target = $2^{-\Delta\Delta C_T}$. [82]

3. MATERIALS AND METHODS

3.1 Materials Used

Caloramator celer strain JW/YL-NZ35, formerly known as *Thermobrachium celere* (equivalent to DSMZ 8682), was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany) was used throughout the study for DNA and RNA extraction to map gene expression levels. $Na_2S.9H_2O$; cystein-HCl; KH_2PO_4; Na_2HPO_4; $FeSO_4.7H_2O$; KCl; $(NH_4)_2SO_4$; NH_4Cl; $MgCl_2.6H_2O$; $CaCl_2.6H_2O$; resazurin; yeast extract; tryptone; and glucose was obtained from Sigma-Aldrich, Germany.

Primers were ordered from Thermo scientific. Genomic DNA extraction kit from GeneJet, Thermo Scientific, Wilmington, USA and RNA extraction from PureLink® RNA Mini Kit (Invitrogen, Carlsbad, USA) were used. For RT PCR the Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Wilmington, USA) was used. Amplification grade DNase I (Sigma, St. Louis, USA) was used for DNA removal. 2 ml natural flat cap microcentriguge tubes (STARLAB, Hamgurg, Germany), optical adhesive covers (Applied Biosystems, Carlsbad, USA), pipette tips topOne (STARLAB, Hamgurg, Germany), Innova 44 incubator shaker series (Brunswick scientific, Eppendorf, Boulevard, USA) were also used.

3.2 Medium and culture

C. celer was cultivated in a modified ATCC 2072 medium containing (g/l): $Na_2S.9H_2O$ 0.13; cystein-HCl 0.13; KH_2PO_4 1.24; Na_2HPO_4 5.79; $FeSO_4.7H_2O$ 0.2; KCl 1; $(NH_4)_2SO_4$ 0.5; NH_4Cl 0.5; $MgCl_2.6H_2O$ 0.1; $CaCl_2.6H_2O$ 0.11; resazurin 0.001; yeast extract 2; tryptone 2; glucose 10. After this an additional 10 ml/l of both multi vitamin solution and trace element solution (ATCC medium No. 2072, American Type Culture Collection) were added to the medium. Pure N_2 gas (Oy AGA Ab, Espoo, Finland) was flushed into the medium for 20 min to make it anaerobic. It was then discharged to serum bottles, while being flushed by pure N_2 gas and then autoclaved for 15 min at 121°C. In addition to this, separate stock solutions were prepared anaerobically under N_2 atmosphere, containing glucose, $MgCl_2.6H_2O$ and $CaCl_2.6H_2O$, $FeSO_4.7H_2O$, cystein-HCl and $Na_2S.9H_2O$. vitamins were sterilized separately using a 2.2 µm sterile syringe filter (VWR international,USA) and added anaerobically to the medium before

the inoculation at the required concentration. After the autoclaving of the medium, its pH was adjusted to 8.2 at 67°C with anaerobic and sterile 3M NaOH solution. The pH was measured with a pH330i pH meter (WTW, Weilheim, Germany) which was rigged with a Slimtrode pH electrode (Hamilton, Bonaduz, Switzerland). Syringes (BD plastipack, Finland) with needles (BD microlance, Finland) were used for all anaerobic operations. For the gas phase experiment an extra step was incorporated where the serum bottles containing the medium was purged with pure hydrogen (Oy AGA Ab, Espoo, Finland) and CO_2 (Oy AGA Ab, Espoo, Finland) respectively for approximately 15 minutes before they were inoculated.

The serum bottles were inoculated with *C. celer* at 5% v/v and incubated at 67°C and 150 rpm. As for the growth phase and gas phase experiment the 0^{th} samples were collected and then samples for RNA extraction procedure was also collected at specific time points. In addition to this air samples from head space and liquid samples were collected at specific time points for gas chromatography (GC) and high performance liquid chromatography (HPLC). For gas chromatography a GC-2014 gas chromatograph (Shimadzu, Kyoto, Japan) having thermal conductivity detector and a Porapak N column (80/100 mesh) was used. In this setup N_2 was used as carrier gas and the temperatures of detector, column and injector were 110°C, 80°C and 110°C, respectively. As for the HPLC system a LC-20AC prominence liquid chromatograph prepped with a DGU-20A5 prominence degasser, RID-10A refractive index detector, and a CBM-20A prominence communications bus module (Shimadzu, Kyoto, Japan). In addition to this, the column was a 30-cm Rezex RHM Monosaccharide H^+ (8%) column (Phenomenex, Alleroed, Denmark) incubated at 40°C.

3.3 Primer preparation

All primers used in this experiment were designed the Primer express software (Applied Biosystems, Carlsbad, USA). These primers were synthesized and obtained from Thermo scientific. An initial 100µM mother stock of each primer was prepared by adding the required volume of TE buffer. Then finally a 5µM stock was prepared by dilution, which will be used for further PCR reactions.

Table 3.1. *The table showing the information of all the primers tested.*

Primer name	Sequence (5'-3')	Gene target	Putative enzyme
pfl_f	GATATGATGTTTCAAGACCAGCAAAG	*pfl*	Pyruvate-formate lyase
pfl_r	ACTTACTCTACCTAGTGACATAGCAGCAC		
pflae_f	GGAGGCGGAGTTACACTTTCAG	*pfl-ae*	Pyruvate-formate lyase activating enzyme

pflae_r	TGTGAATTCCTTCTTCCTTACACCT		
por1_f	ACCACTTGATCTTCCAAACCACTAC	*porA*	Pyruvate oxidoreductase
por1_r	TTCAGCAGCAACTTCAAGAATTACTC		
por2_f	GTTTACTCAAATACAGGTGGTCAATCA	*por*	Pyruvate oxidoreductase
por2_r	GCTCATTGCTATCATTCCAAGGT		
pta_f	GGTTGCTGAATTAAAGGCTCCA	*pta*	Phosphotransacetylase
pta_r	GCTCCTGCTAATCTTTGAACTAACTTG		
acka_f	CAGGAGTTCTTGGTATTTCAGGTGTA	*ack*	Acetate kinase
acka_r	AATGGAATACATCAAGTGCTAGTTGTG		
hyd1_f	TGCAGCAGACCTTACAATAATGGA	*hydA*	NADH-dependent FeFe hydrogenase
hyd1_r	ACAGCAGCTTGTCATAAGTGGAAG		
hyd2_f	GACGCTGTGAGACAATGTGTAATG	*hydA*	NADH-dependent FeFe hydrogenase
hyd2_r	AGGCAGGTCCTACAACTGTTTCA		
mbhl_f	GACAGGCGTTAGAAAACCTTCCA	*mbhL*	Ferredoxin-dependent NiFe hydrogenase
mbhl_r	TGGTGCTTCGTGTCTTGCTGTA		
mbhj_f	GTATCACTAGGATCATGCCCAAGA	*mbhJ*	Ferredoxin-dependent NiFe hydrogenase
mbhj_r	GCCATTATAGCTTCTGGTTTTGGA		
adhe_f	CAGTTAAAGCTGGAGCACCAAAG	*adhE*	Alcohol dehydrogenase
adhe_r	GAAGAGTAGGCAGCCTTTACCATTC		
bdh_f	GAACAGAGGTTACAAGGGCATCAG	*bdh*	Butanol dehydrogenase
bdh_r	CATTCCTGTTTCAGCAACAACTTC		
polc_f	TGGATGGAGTAACATCTGCAACA	*polC*	DNA polymerase
polc_r	CATAGCTTCAGGGAATGCTTGA		
reca_f	GAAATGGGAGATGCTTTTGTAGGA	*recA*	Bacterial DNA recombination protein
reca_r	TTCAGGACTACCAAACATAACACCAA		

Fnora_f	TTGGTAAAAGGGTTGCTGTGATAG	*fnor*	NADH:ferredoxin oxidoreductase
Fnora_r	CTTGCTGGCATCTCATCTTCAG		

3.3.1 Optimal annealing temperature and primer efficiency

An optimal annealing temperature for each primer was determined by using a temperature gradient method where the primers were run in the StepOne™ Plus Real-Time PCR System (Applied Biosystems, Carlsbad, USA) in temperature ranging from 57 to 65°C. The temperature that gave the lowest C_t was selected as the optimal temperature. Using this optimal annealing temperature, the primer efficiency for each primer was determined using serial dilutions of quantified total genomic DNA. For all these PCR runs Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific, Wilmington, USA) was used. The PCR parameters are as follows: initial holding stage at 95°C for 10 minutes, 40 cycles of 15 seconds denaturation at 95°C, 1minute annealing and elongation at optimal temperature of each primer. In the end the primer specificity was tested with melting curves after real time amplification and then run in electrophoresis by 1% agarose gel (Bio Rad laboratories Inc., Finland)

3.4 RNA extraction

The cells were inoculated and grown and their concentrations were periodically determined by using absorbance spectrophotometry at 600nm with an ultraspec 500 pro spectrophometer (Amersham bisciences, munich, Germany). All the pipettes and surfaces were cleaned using the RNase AWAY® solution (Molecular bioproducts, Mexico). Half volume (500µl) of 0.1 mm glass beads (Sigma Aldrich, Germany) in a specially allocated tube. 70% ethanol was prepared according to volume needed. At the required time point the cells were extracted (about 1ml) and harvested and pelleted by centrifuging at 8000xg for 1min at 4˚C. 1 ml of the TRIzol® reagent (Invitrogen, Carlsbad, USA) was added per 1 x 10^7 of bacterial cells and allowed to lyse by pipetting up and down repetitively. The cells were not washed in any other solution before TRIzol® treatment to avoid mRNA degradation. The TRIzol® - cell mixture was transferred in the tube containing the glass beads. The cells were then mechanically disrupted in a Mini-BeadBeater-16 (BioSpec, Bartlesville, USA) with 6 cycles of 30 seconds with intervals of 60 seconds in ice between cycles. Then it was centrifuged at 14,000g for 3 min at 4˚C. After that the RNA isolation was done using a PureLink® RNA Mini Kit (Invitrogen, Carlsbad, USA). The instuctions specified by the manufacturer was followed for the extraction process. The concentration of the RNA was checked using a Nanodrop 2000 spectrophotometer (Thermo scientific, Wilmington, USA).

3.5 Preparation for RT PCR and qPCR

Before undertaking the PCR process, the sample has to undergo DNase treatment in order to remove any contaminating DNA. For this about 2µg of total RNA is taken in 16µl of water in an RNase free PCR tube to which 2µl of 10x reaction buffer and 2µl of amplification grade DNase I (AMP-D1 DNase kit, SIGMA, St. Louis, USA) (1 unit/µl) was added. It was mixed gently and incubated for 15 minutes at room temperature, after which 2µl of stop solution was added to bind calcium and magnesium ions and to inactivate the DNase I enzyme. Finally it was heated at 70 ˚C for 10 minutes to denature both the DNAse I and the RNA and then chilled on ice. After the DNase treatment, about 11µl of the sample was taken from the treated tube and transferred into a new RNase free PCR tube (RT^-). The original tube was taken as RT and another new RNase free PCR tube negative control.

Table 3.2. *The reagents were added according to the order shown below*

µl	RT	NC	RT^-
5x Reaction mix	4	4	4
Enzyme mix	2	2	-
RNA (total 1µg)	11	-	11
RNase free Water	3	14	5

With a Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Wilmington, USA) the cDNA was synthesized from 1 µg of total RNA. The PCR parameters are as follows; initial incubation at 10 min at 25˚C followed by 15 minute at 50˚C. The reaction is terminated by heating at 85˚C for 5minutes. The product of this reaction was then used in qPCR. For all qPCR runs, Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific, Wilmington, USA) was used along with specific primers, product from RT PCT, and nuclease free water for a total volume of 20µl.

4. RESULTS

Results are presented as different parts in this section which include the determination of primer efficiency and optimum annealing temperature for primers, HPLC and gas chromatography results, and relative gene expression in growth phase and gas phase of *Caloramator celer*. The relative gene expression is represented through two endogenous reference genes recA and polC and the geometric mean of these two genes is also used for reference.

4.1 Primer efficiency and optimum annealing temperature

From Table 4.1 we can check the primer efficiency and optimal annealing temperatures of all the primers used. The annealing temperatures of all the primers were in between 61 and 62°C. As for the primer efficiency a majority of primers had an efficiency of more than 90%. Apart from one primer, the rest were in a $\pm$10% with each other. Therefore, this ensures a minimum variation in the experimental results.

Table 4.1. *Optimum annealing temperature and primer efficiency of primers used in this study.*

Primer (r+f)	Optimum annealing temperature (°C)	Primer efficiency (%)
PFL	61	94.07
PORb	61	94.5
PTA	61	89.76
ACKA	62	94.71
Hyd1	62	95
Hyd2	62	94.3
mbhL	61	92.4
mbhJ	61	91.83
PFLae	62	93.73
PORa	61	96.47
ADH1	61	88.34
ADH2	61.5	94.38
polC	61	100.44
RPOC	62	86.60
DNAA	62	97.87
Reca	62	93.82

Fnora	61	92
BADH	62	95.16

4.2 RT PCR from total RNA

The gel below (Fig. 4.1) shows the successful amplification of all the genes involved in transcription process. The DNase treated purified RNA after being run in the RT PCR was again run individually with separate primers. This gives a visual indication that the primers function properly and that the mRNA of the genes involved in the transcription process was successfully isolated. The controls also signify the absence of any DNA contamination. The RNA here was isolated at the mid-log phase of the growth cycle.

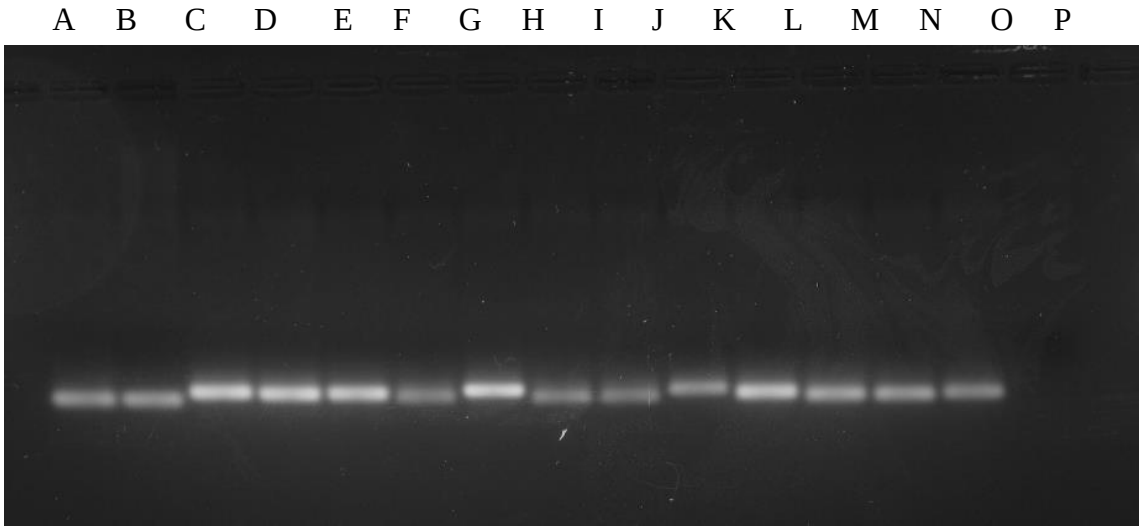

Figure 4.1. *The gel below shows the successful amplification of all the genes involved in transcription process. Where lane A to P denote 16S gDNA, 16S cDNA, pfl, pforb, pta, adh, mbhH, pfora, adhE, fnora, acka, hyd1, hyd2, no RT control and No cDNA control.*

4.3 Growth plot of *C. celer* for both growth and gas phase experiment

The growth plot of *C. celer* is shown in the figures (Fig. 4.2 and Fig 4.3) below. The samples were taken periodically and its O.D at 600nm was checked. At suitable cell densities the time point was noted and the extraction took place. All cultures were grown in replicates for both the experiments and their standard deviation is also included below.

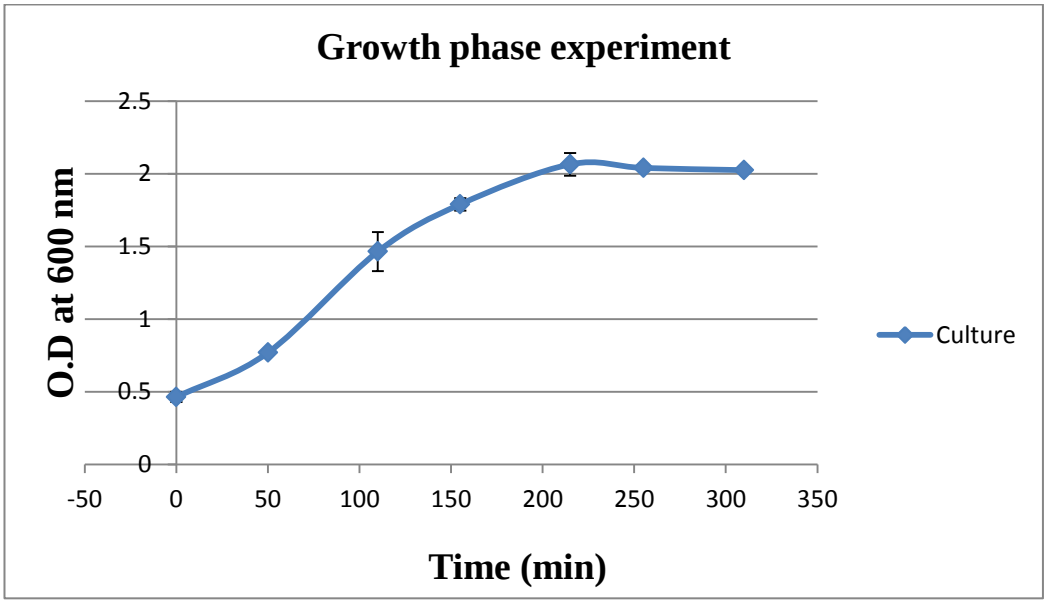

Figure 4.2. *Graph showing the growth plot of C. celer with time for the growth phase experiment. The extraction for log, mid log and stationary points was done at 50th, 150th and 250th minute.*

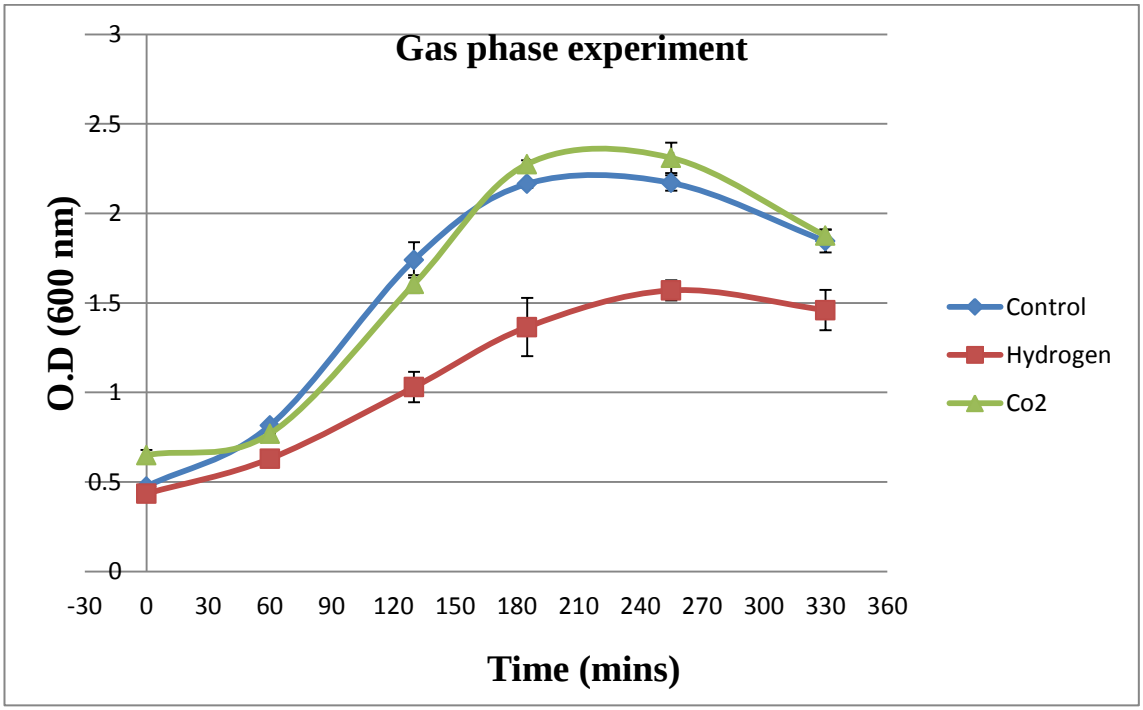

Figure 4.3. *Graph showing the growth plot of C. celer with time for the gas phase experiment. The extraction point for control, hydrogen gas phase culture and CO_2 gas phase culture was at 130th, 130th and 185th minute.*

4.4 Relative gene expression in growth phase experiment of *C. celer*

Using $2^{-\Delta\Delta C}{}_T$ method the relative gene expressions of various genes were plotted. Samples collected from three different time points were compared to monitor the varying gene expression levels. Three different endogenous reference systems were used; recA (Fig. 4.4), polC (Fig. 4.5) and geometric mean and average of polC (Fig. 4.6) and recA. From the results (Figures 4.4, 4.5 and 4.6) we can notice mid log phase has the most gene expression while stationary phase has down regulation of gene expression. Gene expression results in Figures 4.4 and 4.5 follows similar pattern and their average in Fig. 4.6 shows minimal variance. From Fig. 4.6 we can see that *pfl*, *pflae*, *pta* and *ack*, are upregulated in log and mid log phase when compared to 0^{th} sample, with their mid log phase having greater upregulation. Their stationary phase is downregulated. Genes like *adhE, adh2* and *fnora* are upregulated in all the three growth phases when compared to the 0^{th} sample, with their mid log phase having the highest gene expression. In contrary, the genes such as *porA, porB, hydA, mbhL, mbhJ* and *badh* have widespread downregulation in all the growth phases.

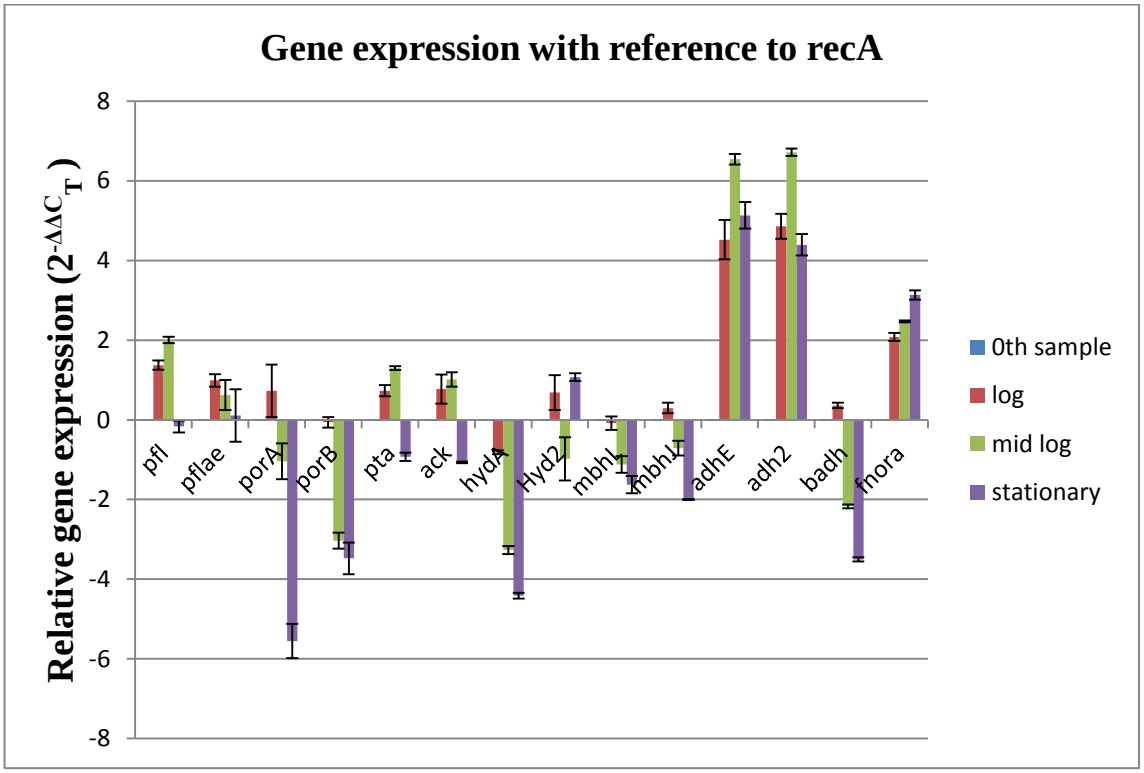

Figure 4.4. *Graph showing relative gene expression in three samples that were obtained during three significant growth phases. $2^{-\Delta\Delta C}{}_T$ method is used for the normalization of the data. In this graph recA housekeeping gene is used as a reference gene. Negative value denotes downregulation of gene expression when compared to other genes at that specific time. Two different replicates were used and their standard deviation plotted.*

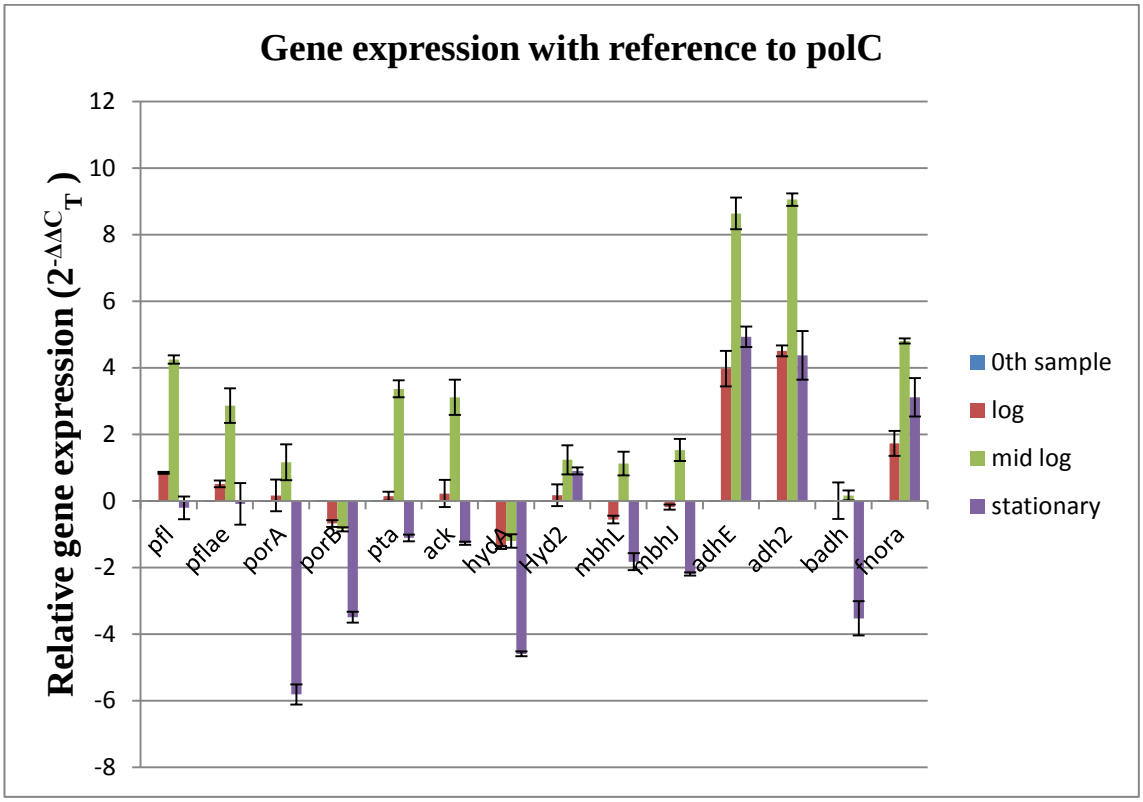

Figure 4.5. *Graph showing relative gene expression in three samples that were obtained during three significant growth phases.* $2^{-\Delta\Delta C}{}_T$ *method is used for the normalization of the data. In this graph polC housekeeping gene is used as a reference gene. Negative value denotes downregulation of gene expression when compared to other genes at that specific time while the positive value denotes upregulation. Two different replicates were used and their standard deviation plotted.*

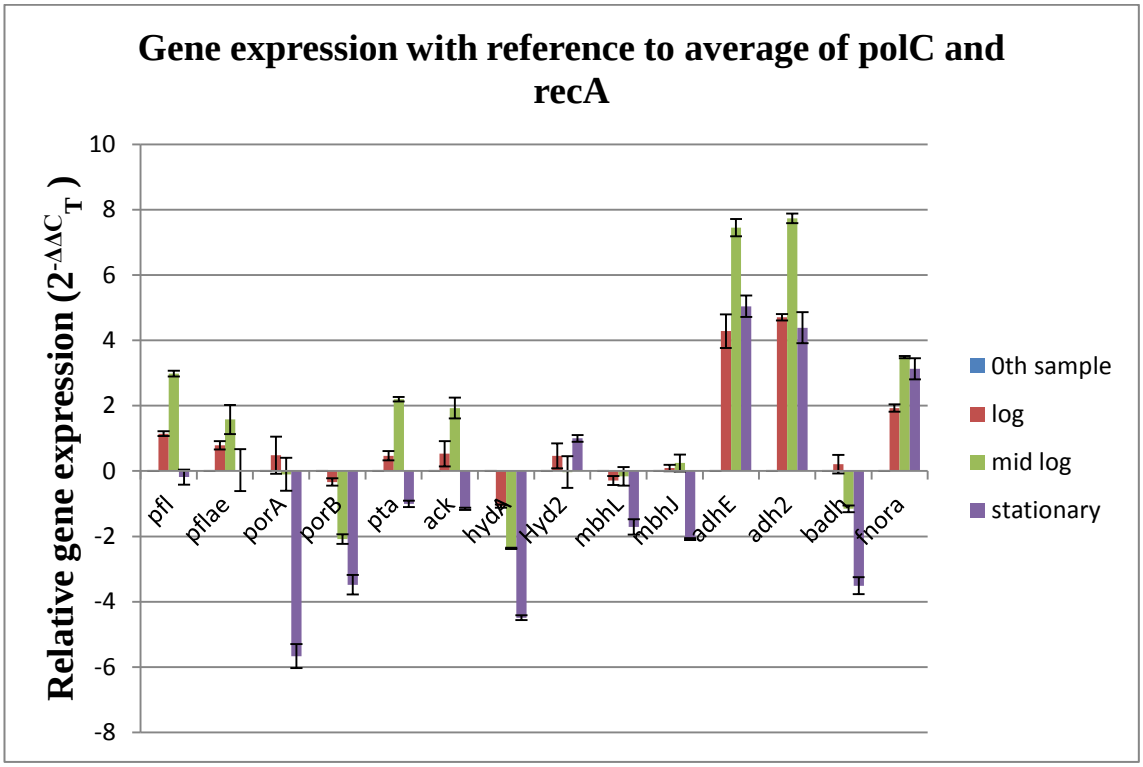

Figure 4.6. *Graph showing relative gene expression in three samples that were obtained during three significant growth phases.* $2^{-\Delta\Delta C_T}$ *method is used for the normalization of the data. In this graph geometric mean between polC and recA housekeeping genes were used and their average was used for plotting. Negative value denotes down-regulation of gene expression when compared to other genes at that specific time. Two different replicates were used and their standard deviation plotted.*

4.5 Product distribution in growth phase experiment of *C. celer*

The metabolic products accumulated with time are given in the graphs below (Figures 4.7, 4.8), along with the yield of the individual products (Table 4.2). All the values contain standard deviation from both the replicates and their variance seems to be negligible. Fig. 4.7 shows the glucose consumption and metabolite accumulation in the growth culture. Glucose consumption is rapid between 50th and 150th minute of growth, and then it slows down after that. This can also be correlated with the formate, acetate and ethanol accumulation where, there is rapid accumulation between 50th and 150th minute, after which the values are maintained. Butyrate is absent. It is also true for hydrogen as shown in Fig. 4.8 where, after 150th minute there is a dip in the rate of accumulation.

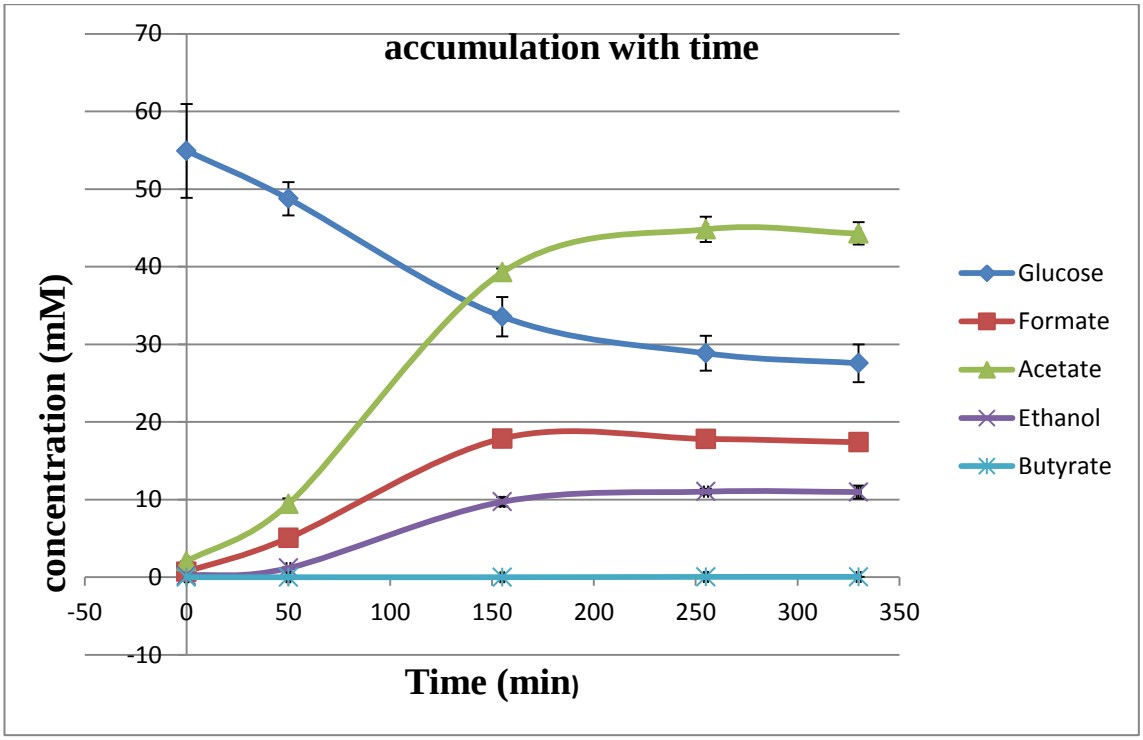

Figure 4.7. *This graph shows the accumulation of products with time, and the consumption of glucose with time.*

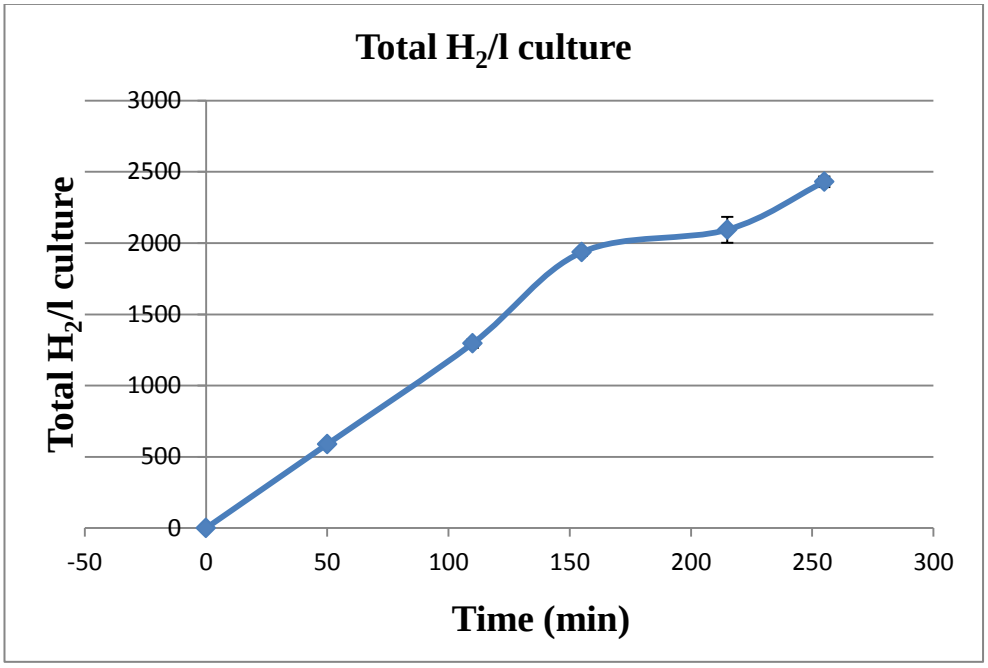

Figure 4.8. *This graph shows the total hydrogen per liter culture that was produced with time.*

Table 4.2. *Yields of all the products formed is given below along with standard deviation.*

Product	Yield	Standard deviation
Formate	0.61	0.0888
Acetate	1.55	0.2543
Ethanol	0.39	0.0835
Hydrogen	3.21	0.3719

4.6 Relative gene expression in gas phase experiment of *C. celer*

Using $2^{-\Delta\Delta C}{}_T$ method the relative gene expressions of various genes were plotted. Samples collected from both the hydrogen and CO_2 gas phase replicate at specific time points were compared to monitor the varying gene expression levels. Three different endogenous reference systems were used; recA, polC and geometric mean and average of polC and recA. From the results (Fig. 4.9, 4.10, 4.11) we can notice when compared to the control the level of gene expression in both hydrogen and CO_2 gas phase replicate is very much reduced. Samples with CO_2 headspace seem to have fared much worser than samples with hydrogen headspace. The variance between polC and recA seems to be significant; however the overall context is maintained as explained in discussion. From Fig. 4.9 we can see that almost all genes are downregulated in both the CO_2 and hydrogen gas phase cultures with respect to the control. An exception to this is the *porA, mbhL* and *hyd2* genes from hydrogen gas phase culture which are upregulated. However this is not the case for Fig. 4.10 where, all the genes from hydrogen gas phase culture seem to be upregulated, while all the genes from CO_2 gas phase culture are downregulated. Fig. 4.11 can give better clarity to this analysis as it uses the average of both Fig. 4.9 and 4.10. This fine-tuned data from Fig. 4.11 shows that all the genes from both the gas phases are significantly downregulated when compared to the control.

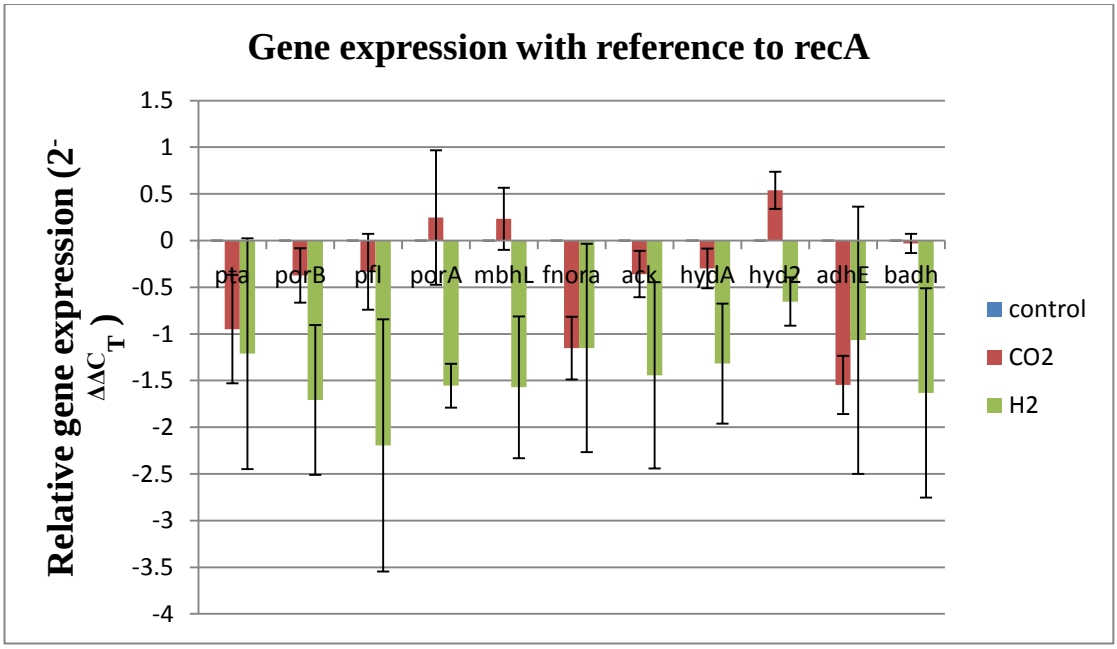

Figure 4.9. *Graph showing relative gene expression in hydrogen and CO_2 head space samples that were obtained during specific timepoints. $2^{-\Delta\Delta C}{}_T$ method is used for the normalization of the data. In this graph recA housekeeping gene is used as a reference gene. Negative value denotes downregulation of gene expression when compared to other genes at that specific time. Two different replicates were used and their standard deviation plotted.*

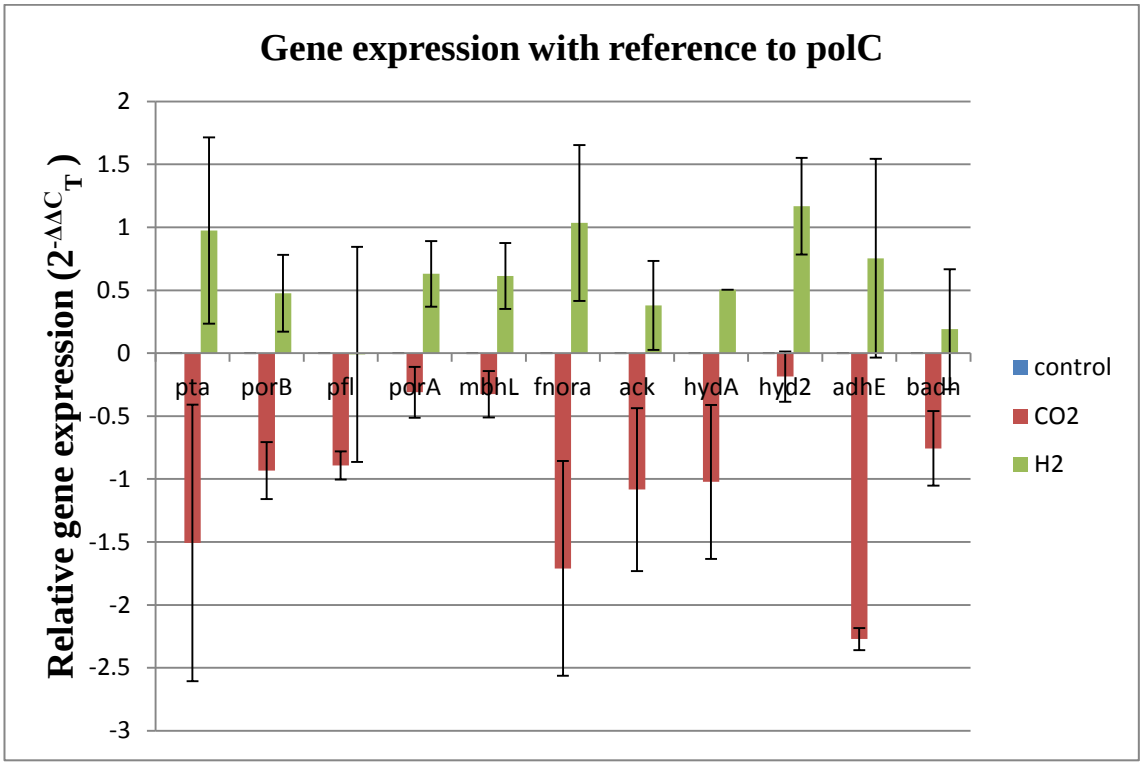

Figure 4.10. *Graph showing relative gene expression in hydrogen and CO_2 head space samples that were obtained during specific timepoints. $2^{-\Delta\Delta C_T}$ method is used for the normalization of the data. In this graph polC housekeeping gene is used as a reference gene. Negative value denotes downregulation of gene expression when compared to other genes at that specific time while the positive values denotes upregulation of gene expression. Two different replicates were used and their standard deviation plotted.*

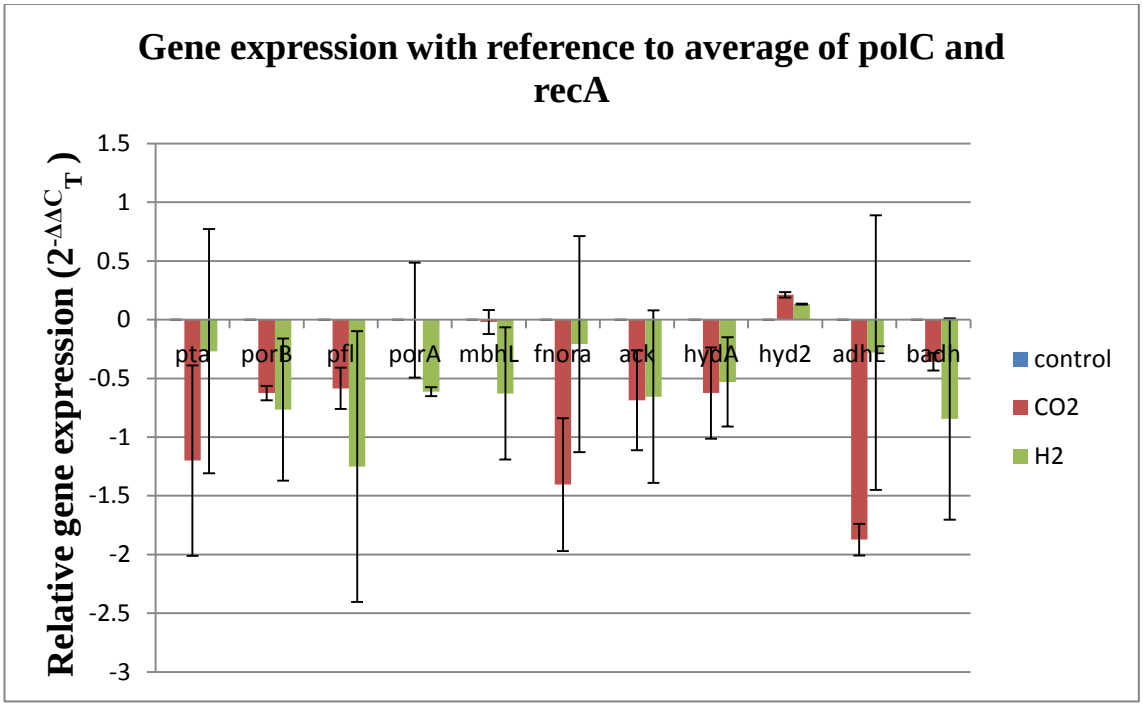

Figure 4.11. *Graph showing relative gene expression in hydrogen and CO_2 head space samples that were obtained during specific time points. $2^{-\Delta\Delta C}{}_T$ method is used for the normalization of the data. In this graph geometric mean between polC and recA house-keeping genes were used and their average was used for plotting. Negative value denotes downregulation of gene expression when compared to other genes at that specific time. Two different replicates were used and their standard deviation plotted.*

4.7 Product distribution in gas phase experiment of *C. celer*

The metabolic products accumulated with time are given in the graphs below (Figures 4.12, 4.13, 4.14 and 4.15). From Fig. 4.12 it is seen that the formate accumulation is highest for the control at 17.7mM. The formate accumulation peaks at 150th minute for control and CO_2 gas phase culture whereas, for hydrogen gasphase culture it peaks at 210th minute. The acetate accumulation (Fig. 4.13) and glucose consumption (Fig. 4.15) results correlate with each other and show that the control samples have higher consumption and accumulation when compared to that of the hydrogen gas phase cultures. However Fig. 4.14 shows that hydrogen gas phase cultures have highest ethanol accumulation at 11.8mM, while control only has 8.4mM of ethanol accumulation. The metabolic yields are given in Fig. 4.16 and it shows that hydrogen gas phase samples have highest ethanol yield and the control has highest yields of acetate and formate. .

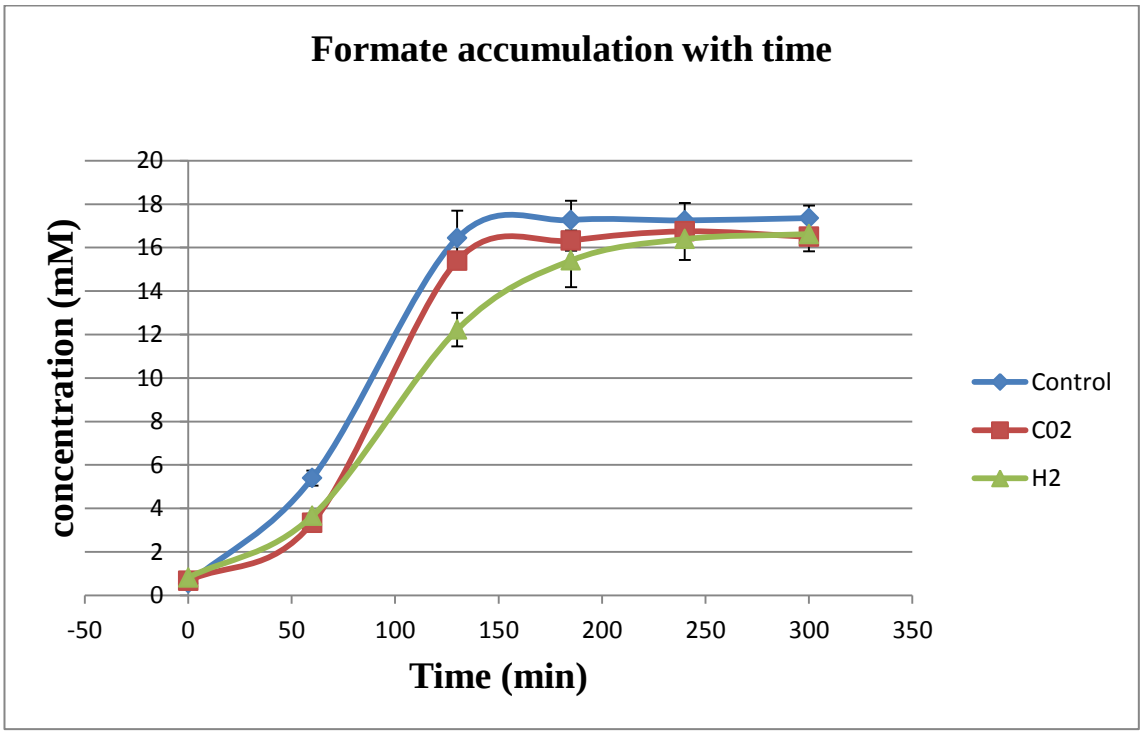

Figure 4.12. *Graph shows the accumulation of formate with time in control, hydrogen and CO_2 head space samples.*

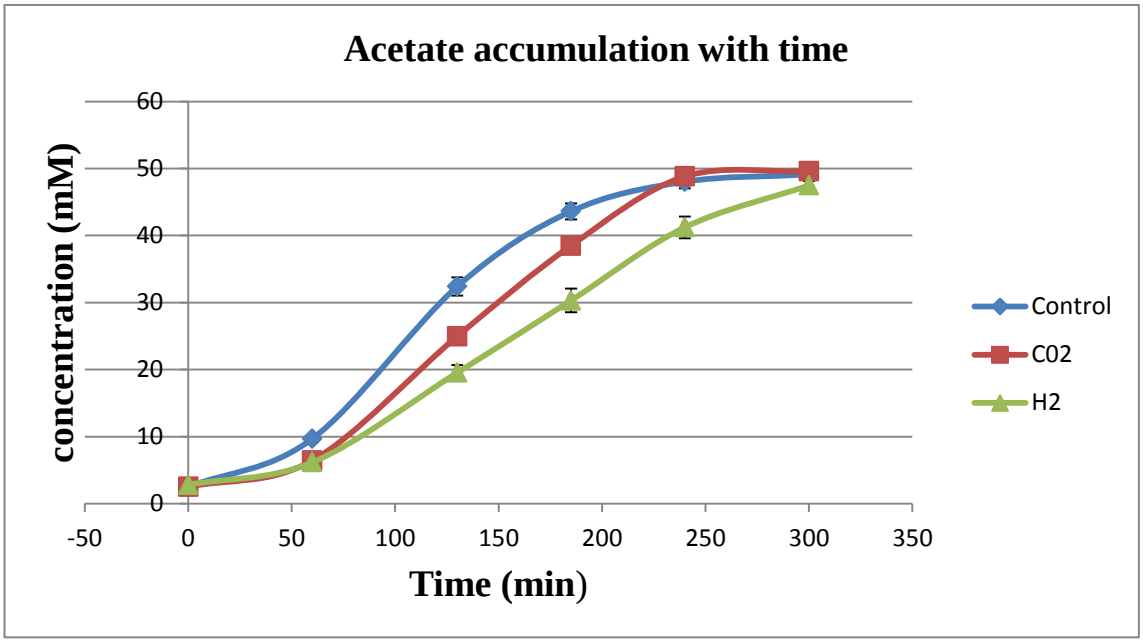

Figure 4.13. *Graph shows the accumulation of acetate with time in control, hydrogen and CO_2 head space samples.*

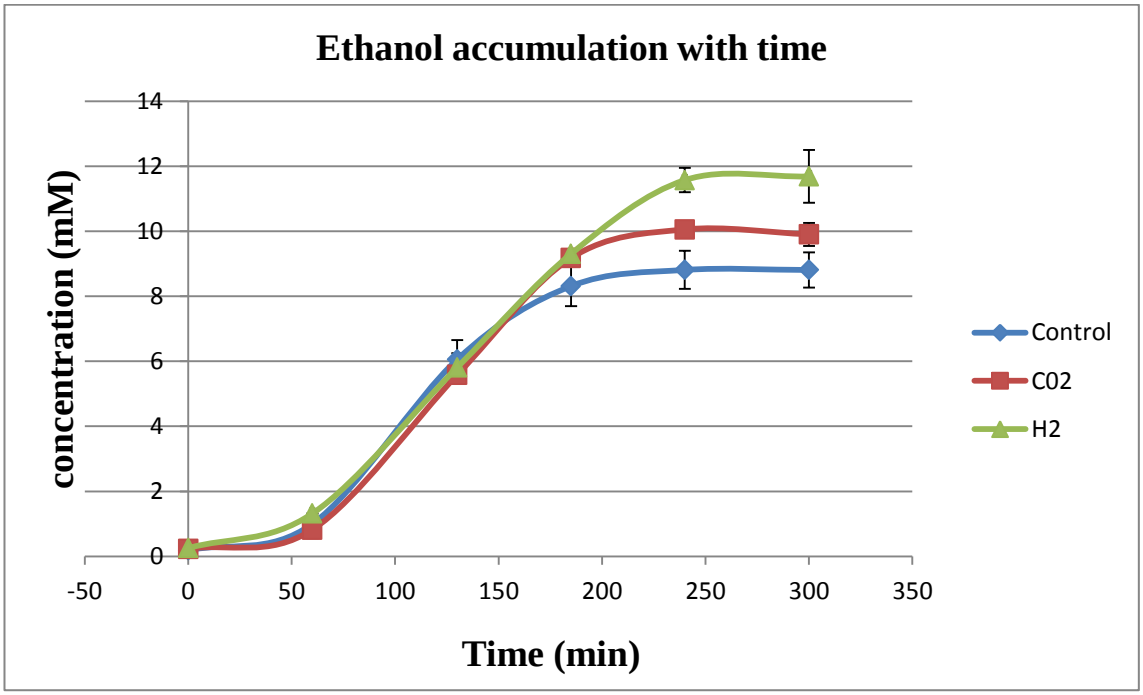

Figure 4.14. *Graph shows the accumulation of ethanol with time in control, hydrogen and CO_2 head space samples. It seems that hydrogen heapspace sample has higher ethanol accumulation*

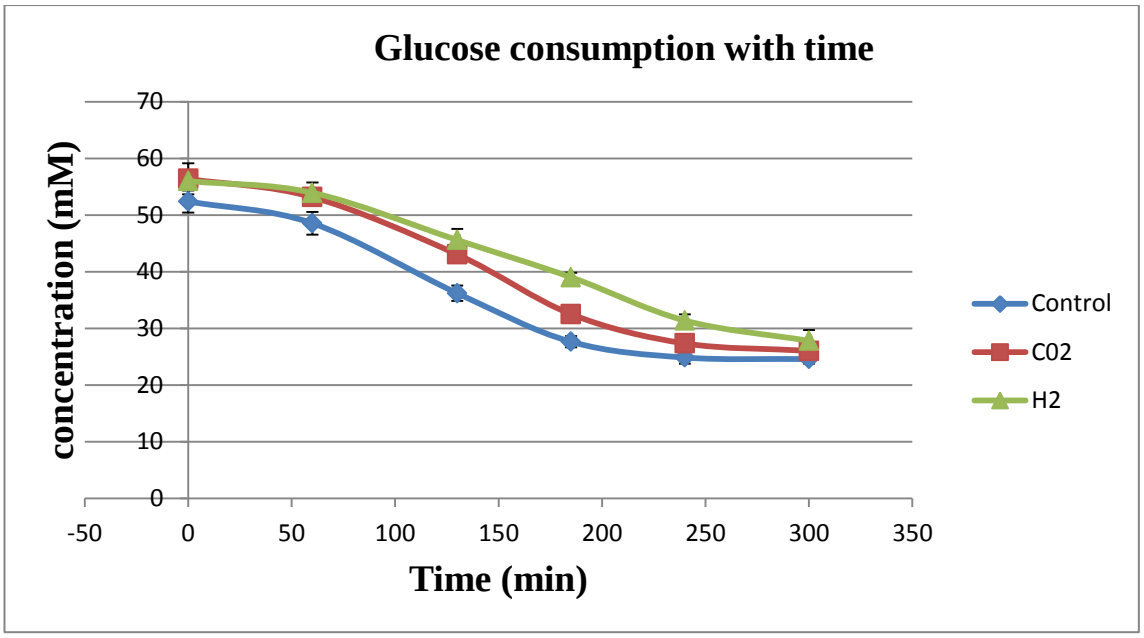

Figure 4.15. *Graph shows the consumption of glucose with time in control, hydrogen and CO_2 head space samples.*

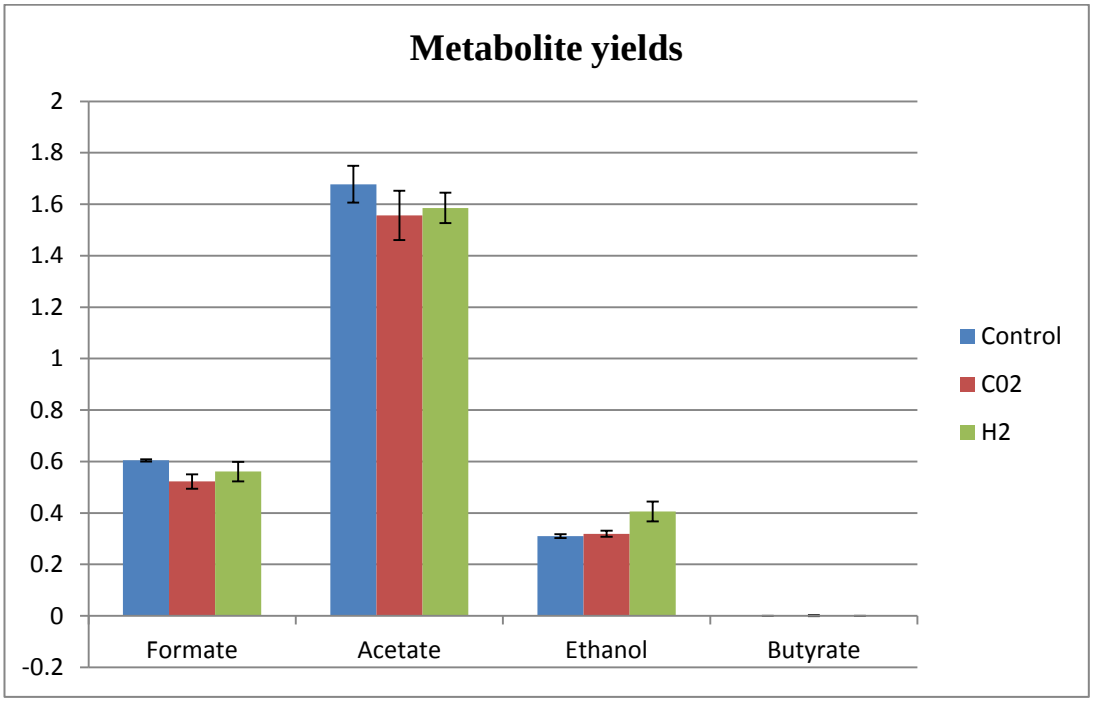

Figure 4.16. *Graph shows Yields of all the products formed along with the standard deviation. As observed the yield of ethanol is more in hydrogen flushed head space sample.*

5. DISCUSSION

5.1 Statistical significance of the results

As seen from Table 4.1 the annealing temperature and the primer efficiencies of the primers were found to be in acceptable levels. Also Figure 4.1 shows the successful amplification of all the genes involved in transcription process along with no contamination from intermediary processes. The use of two replicates for each experiment and inclusion of two different housekeeping genes as a normalizing factor has helped in validating the significance of the results [60, 82]. Figures 4.4, 4.5 and 4.6, which are related to the growth phase analysis experiment, shows a level of intrinsic correlation between the two housekeeping genes. It is important to take into account that the $2^{-\Delta\Delta C}{}_{T}$ method deduces relative gene expression through an arbitrary value by comparing with control and reference gene [82]. Therefore the result is more coherent if the objective is to find out if a target gene is upregulated or downregulated than the control, rather than finding the absolute percentage of expression. For example, in the growth phase experiment the normalized results of both recA and polC correlate with each other in general terms of expression. When the average of these two normalizing genes was used, a better resolution in terms of gene expression was obtained while the standard deviation was still relevant. Figure 4.6 was used to understand the levels of gene expression in the growth phase experiment. However, in special circumstances (gas phase experiment) such as modified physiological conditions, the normalized results of gene expression had to be separately studied for each house keeping gene before combining their average for statistical relevance.

As seen from Figures 4.6 and 4.7, the recA and polC normalized relative gene expression results from hydrogen gas phase experiment seems to vary between each other. For this analysis the nature of the housekeeping gene has to be noted which defines that they are ubiquitously expressed in all the cells of an organism under normal physiological conditions [71]. However, here the cells were inoculated in an artificially stressed medium that had high pre inoculation concentration of CO_2 and H_2 in the headspace. Therefore in this case the use of two different housekeeping genes for normalizing the relative gene expression is important in order to resolve the deviation in analysis and give significant results.

5.2 Growth dependent transcriptional analysis of fermentative pathway

Figure 4.6 depicting the relative gene expression of samples collected from log, mid log and stationary growth phase of *C. celer* is used for the analysis. Pyruvate formate lyase (PFL) which produces formate from pyruvate by utilising coenzyme A (CoA) as a substrate is encoded by *pfl* gene and the activating enzyme *pflae.* An analogy between *pfl* and *pflae* can be seen here where they seem to be directly in proportional with each other in terms of gene regulation [14]. Both the genes are upregulated in log and mid log growth phase with mid log having the peak expression. From there the gene expression for both the genes is then downregulated in the stationary phase.

Pyruvate and CoA are the limiting factors governing formate synthesis. As a result, formate formation is aspected by ethanol, acetate and ferrredoxin dependant hydrogen pathway [14]. Therefore, Phosphotransacetylase (PTA) and alcohol dehydrogenase (adh) which are coded by *pta* and *adhE* respectively, are upregulated (mid log) in tandem with *pfl* and *pflae.* Acetate which is the primary metabolite that accounts for 78% of the aggregate soluble metabolite in the growth medium [14, 10] is produced through PTA and acetate kinase (ACK) enzymes where, ACK is coded by *ack* gene. Both *pta* and *ack* genes are upregulated equally in both log and mid log phase and it is comparable to that of *pfl* and *pflae*. The upregulation of acetate production genes is also comparable to that of the glucose consumption and acetate accumulation as seen in figure 4.7 where the consumption and accumulation is dramatic in the mid log phase. Both of these genes are then downregulated in the stationary phase. Acetyl CoA is the end product of formate and ferredoxin dependant hydrogen pathway and used as a substrate by PTA to carry forward the acetate production. Therefore, an increased gene expression of *pfl* in the log and mid log phase ensures and abundant supply of Acetyl CoA which in turn leads to upregulation of *pta* and *ack.*

Acetyl CoA is also the substrate needed for ethanol synthesis. Alcohol dehydrogenase (ADH) is the enzyme involved in this process and it is coded by *adhE* gene. This synthesis competes for NADH, which is also used by the NADH dependant hydrogenase (hydA) [14]; and releases NAD^+ and CoA. The *adhE* gene seems to be significantly upregulated in the log and mid log phase. A good supply of acetyl CoA ensures a continuous manufacture of ethanol and thereby ensuring steady CoA release needed for formate and ferredoxin dependent hydrogen production. This may also be partly responsible for upregulation of genes related to formate synthesis. However unlike acetate and formate synthesis, *adhE* gene is still upregulated in the stationary phase.

Ferredoxin-dependent NiFe hydrogenase (Fd H_2ase) is one of the two enzymes in *C. celer* that is responsible for hydrogen production. It is coded by *mbhL* and *mbhJ*.

Both of these genes seem to be partially downregulated in log and mid log phase. This can be attributed to direct competition in acquiring CoA by the formate pathway [14]. This is also evident in the case of *porA* gene that codes for Pyruvate ferredoxin oxidoreductase (PFOR) that is needed in the ferredoxin dependent hydrogen production process. As the formate pathway competes for CoA with PFOR, its gene expression is maintained at a minimum when compared to that of *pfl*. Therefore, an observation can be made that increased upregulation of *pfl* leads to a downregulation of *porA* and *mbhJ*.

NADH-dependent FeFe hydrogenase (NADH H_2ase) is the other hydrogenase in *C.celer*, and it is coded by *hydA*. This gene is increasingly downregulated with time between log, mid log and stationary phase. Competition for NADH between ADH and NADH H_2ase might be a contributing factor especially between mid-log and stationary phase. In addition to that, as the hydrogen partial pressure in the system increases with time, the inhibition to hydrogen production also increases, thereby leading to increased downregulation of all the genes involved in hydrogen production process. In addition to that a low hydrogen partial pressure is very crucial for the oxidation of NADH to NAD^+ which can ensure a higher H_2 production [41, 9]. As the concentration of H_2 in the headspace and the liquid space increases rapidly, the reaction becomes thermodynamically unfavourable [37- 39, 10]. Following which the metabolic pathway shifts towards the production of reduced end products like ethanol [40, 9]. This could explain the reason behind the continued downregulation of *hydA* and *mbhL* hydrogenase genes with time while the *adhE* gene is upregulated with time. The effect of pH on hydrogen and other metabolites is discussed below.

5.3 Gas dependent transcriptional analysis of fermentative pathway

The cells grown in artificially stressed conditions have given interesting results in terms of transcriptional analysis. From Figure 4.3 we can see that the culture inoculated in hydrogen injected medium grows much slower than that of control and its mid log extraction was at 185^{th} minute compared to that of control and CO_2 phase which was at 130^{th} minute. Relative gene expression results from figure 4.8 show a severe downregulation of majority of genes in both CO_2 and hydrogen gas phase cultures when compared to that of the control. With the help of HPLC results, an attempt was made to correlate these with that of the relative gene expression results.

The *pfl* gene was downregulated in both the cultures grown at CO_2 and hydrogen gas phase when relatively compared to that of the control. The gene from hydrogen gas phase culture was found to be more downregulated than that of the CO_2 gas phase. It was also evident in the HPLC results (Figure 4.11) which showed that the formate

accumulation for hydrogen gas phase culture at its point of extraction was much lower than CO_2 gas phase culture while control had the highest formate accumulation. This trend also continues for acetate synthesis where *pta* and *ack* genes for both the gas phase cultures were downregulated. This correlated with the HPLC results (Figures 4.13 and 4.15) which showed that acetate accumulation and glucose consumption was much lower than control with hydrogen gas phase culture having the lowest values. Since acetyl CoA is vital for acetate synthesis, the downregulation of genes in formate pathway might have contributed to the reduced gene expression of *pta* and *ack* genes which in turn affects the supply of CoA. A better explanation can also be indicated through ethanol pathway because ADH requires acetyl CoA as substrate for ethanol synthesis.

Ethanol synthesis in the two gas phase cultures has taken a turn for the interesting because the HPLC results (Figure 4.14) shows that both the cultures have higher ethanol accumulation than that of the control, with hydrogen phase culture having the highest accumulation. This is especially true at the end of the growth phase. Even though *adhE* gene is downregulated for both the gas phases at its point of extraction, the cells in the gas phase under severe physiological stress might have switched to ethanol pathway. Therefore as time passed from mid log phase, the acetate pathway for the cells in gas phase culture would have been downregulated even more because of increased gene expression of *adhE* gene in order to produce more ethanol. As discussed above the increased H_2 concentration in the headspace has favored the ethanol production due to which the ethanol concentration is highest in the hydrogen gas phase experiment [40, 9].

These gas phase experiments provided valuable insights into genes involved in hydrogen production. It was seen that *hydA* was downregulated for both the gas phase cultures in an equal manner because of the thermodynamically unfavorable process due to increased H_2 concentration [37- 39, 10], which in turn led to increased ethanol production. However in the case of *mbhL* gene, it was seen that the gene was more downregulated in the presence of hydrogen in the culture. This is because in the CO_2 gas phase culture the *mbhL* gene was expressed equal to that of the control, while in the H_2 gas phase culture that same gene was significantly downregulated. This shows that this gene is also sensitive to the presence of hydrogen in the system. The *fnor* gene which codes for NADH:ferredoxin oxidoreductase depends on the hydrogen partial pressure [41, 9]. The down regulation of this gene can be attributed to this aspect.

The importance of pH in the system cannot be overstated. When the pH decreases the hydrogen production is hampered [42, 10]. Usually in prolonged fermentation, volatile fatty acids are produced which drops the pH well below the optimal range [10]. This may be the cause for downregulation of hydrogenase genes over prolonged periods for both the growth phase and gas phase experiment. However,

as the pH decreases more, the cellular function is disrupted and other soluble metabolite pathways are also inhibited. This may explain why the genes are downregulated in the stationary phase. This is especially true for the CO_2 gas phase experiment where, its addition into the serum bottle might have contributed to pH drop. This drop in pH has contributed to all its genes being downregulated.

6. CONCLUSION

Through this work the metabolic interdependency within the fermentative pathway of *C. celer* was studied. The intricate relationship between acetate, formate and ethanol pathways has shown the limiting factors involved in their individual synthesis and their direct competition on hydrogen production was analyzed through transcription. Formate and ethanol pathway competed with the hydrogen production pathway and this caused down regulation of *mbhL* and *hydA* genes. The effect of different physiological states on the cell was studied through the route of relative gene expression. The importance of pH and hydrogen partial pressure was studied from a transcription point of view. It was also found that the cultures grown in artificially stressed gas phase media had a widespread downregulation of its genes, this valuable data can be used in future for metabolic engineering.

In order to produce an efficient hydrogen producing system, limiting factors have to be removed. However this study has shown that removal or suppression of interdependent pathways can negatively affect the desired product production. Therefore future attempts at metabolic engineering have to make sure that balance is reached. This can be circumvented if genetic engineering is used to provide alternate efficient pathways. More research can be undertaken to develop metabolite resistant strains that can resist the inhibition due to the buildup of toxic metabolites. *C. celer* has to undergo more work in order to make it industrially viable for hydrogen production. Nonetheless, *C. celer* provides interesting new possibilities with respect to clean energy production.

REFERENCES

[1] Baltzer, F. 'Theodor Boveri'. *Science,* 1964. 144: 809–815.

[2] McKusick, V.A. Walter S. Sutton and the physical basis of Mendelism. *Bull Hist Med,* 1960. 34: 487–497.

[3] Avery, O.T., MacLeod, C.M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of *Pneumococcal* types. *J. Exp. Med,* 1944. 79: 137–158.

[4] Daneholt, B., & Annika, R. Advanced Information: RNA interference. The Nobel Prize in Physiology or Medicine. *Nobelprize.org. Nobel media AB 2014.* 2006. [Accessed on 06.04.2013]. Available at: http://www.nobelprize.org/nobel_prizes/medicine/laureates/2006/advanced.html.

[5] Sinha, P., Pandey, A. An evaluative report and challenges for fermentative biohydrogen production. *Journal of Hydrogen Energy,* 2011. 36: 7460–78.

[6] Das, D., Veziroğlu, T.N. Hydrogen production by biological processes: a survey of literature. *International Journal of Hydrogen Energy,* 2001. 26: 13–28.

[7] Baena, S., Patel, B.K. 2009. Genus V. Caloramator Collins, Lawson, Willems, Cordoba, Fernández-Garayzábal, Garcia, Cai, Hippe and Farrow 1994, 812^{VP} emend. Chrisostomos, Patel, Dwivedi and Denman 1996, 497: 834–838. *In* De Vos, P., Garrity, G., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., Whitman, W.B (ed). *Bergey's manual of systematic bacteriology*, 2nd ed, vol 3. *The Firmicutes*. Springer-Verlag, New York.

[8] Engle, M., Li, Y., Rainey, F., DeBlois, S., Mai, V., Reichert, A., Mayer, F., Messner, O., Wiegel, J. *Thermobrachium celere* gen. nov., sp. nov., a rapidly growing thermophilic, alkalitolerant, and proteolytic obligate anaerobe. *Int. J. Syst. Bacteriol*, 1996. 46: 1025–1033.

[9] Ciranna, A., Santala, V., Karp, M. Biohydrogen production in alkalithermophilic conditions: Thermobrachium celere as a case study. *Bioresource Technology,* 2011. 102: 8714–8722.

[10] Ciranna, A., Santala, V., Karp, M. Enhancing biohydrogen production of the alkalithermophile *Thermobrachium celere*. *Int. Journal of Hydrogen Energy,* 2012. 37: 5550–5558.

[11] Koskinen, P.E.P., Lay, C., Puhakka, J.A., Lin, P., Wu, S.Y., Orlygsson, J. High-efficiency hydrogen production by an anaerobic, thermophilic enrichment culture from an Icelandic hot spring. *Biotechnol. Bioeng*, 2008. 4: 665–678.

[12] Volbeda, A., Charon, M.H., Piras, C., Hatchikian, E.C., Frey, M., Fontecilla-camps J.C. Crystal-structure of the nickel–iron hydrogenase from *Desulfovibrio gigas*. *Nature*, 1995. 373: 580–587.

[13] Hiromoto, T., Ataka, K., Pilak, O., Vogt, S., Stagni, M.S., Meyer-Klaucke, W., et al. The crystal structure of C176A mutated [Fe]-hydrogenase suggests an acyl-iron ligation in the active site iron complex. *FEBS Lett*, 2009. 583(3): 585–590.

[14] Ciranna, A., Ferrari, R., Santala, V., Karp, M. Inhibitory effects of substrate and soluble end products on biohydrogen production of the alkalithermophile *Caloramator celer*: Kinetic, metabolic and transcription analyses. *Int. Journal of Hydrogen Energy*, 2014. 39: 6391–6401.

[15] Shafaat, H.S., Rüdiger, O., Ogata, H., Lubitz, W. [NiFe] hydrogenases: A common active site for hydrogen metabolism under diverse conditions. *Biochimica et Biophysica Acta*, 2013. 1827: 986-1002.

[16] Chou, C.J., Jenney Jr., F.E., Michael, W.W., Adams, M.W.W., Kelly, R.M. Hydrogenesis in hyperthermophilic microorganisms: implications for biofuels. *Metab. Eng*, 2008. 10: 394–404.

[17] Kengen, S.W.M., Goorissen, H.P., Verhaart, M.R.A., Stams, A.J.M., Van Niel, E.W.J., Claassen, P.A.M. Biological Hydrogen Production by Anaerobic Microorganisms. In: Soetaert, W., Verdamme, E.J. (Eds.), Biofuels. *John Wiley & Sons*, 2008, Chichester. 197–221.

[18] Chevreux, B., Wetter, T., Suhai, S. Genomic sequence assembly using trace signals and additional sequence information. Computer science and biology: proceedings of the *German conference on bioinformatics (GCB)*, 1999, Hannover, Germany. 45–56.

[19] Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, 2011. 27: 578–579.

[20] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. The RAST server: rapid annotations using subsystems technology. *BMC Genomics*, 2008. 9: 75.

[21] Ciranna, A., Larjo, A., Kivisto, A., Santala, V., Roos, C., Karp, M. Draft genome sequence of the hydrogen and ethanol producing anaerobic alkalithermophilic bacterium *Caloramator celer*. *Genome Announc.* 2013; 1(4):e00471e3.

[22] Soboh, B., Linder, D., Hedderich, R. A multisubunit membrane bound [NiFe]-hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. *Microbiology*, 2004. 150: 2451–2463.

[23] Sapra, R., Verhagen, M.F., Adams, M.W. Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol*, 2000. 182: 3423–3428.

[24] Wukovits, W., Drljo, A., Hilby, E., Friedl, A. Integration of Biohydrogen Production with Heat and Power Generation from Biomass Residues. *Chemical Engineering Transactions*, 2013. 35: 1003–1008.

[25] Guo, X.M., Trably, E., Latrille, E., Carrère, H., Steyer, J. Hydrogen production from agricultural waste by dark fermentation: A review. *International Journal of Hydrogen Energy*, 2010. 35: 10660–73.

[26] Kotsopoulos, T.A., Fotidis, I.A., Tsolakis, N., Martzopoulos, G.G. Biohydrogen production from pig slurry in a CSTR reactor system with mixed cultures under hyper-thermophilic temperature (70ºC). *Biomass Bioenergy*, 2009. 33: 1168-74.

[27] Vijayaraghavan, K., & Ahmad, D. Biohydrogen generation from palm oil mill effluent using anaerobic contact filter. *Journal of Hydrogen Energy*, 2006. 31: 1284–91.

[28] Hallenbeck, P.C., & Benemann, J.R. Biological hydrogen production; fundamentals and limiting processes. *International Journal of Hydrogen Energy*, 2002. 27: 1185–1193.

[29] Adams M.W. The structure and mechanism of iron-hydrogenases. *Biochim Biophys Acta*, 1990. 1020: 115–45.

[30] Nicolet, Y., Fontecilla-Camps, J.C., Fontecave, M. Maturation of [FeFe]-hydrogenases: Structures and mechanisms. *International journal of Hydrogen energy*, 2010. 35: 10750–10760.

[31] Pierik, A.J., Hulstein, M., Hagen, W.R., Albracht, S.P.J. A low-spin iron with CN and CO as intrinsic ligands forms the core of the active site in [Fe]-hydrogenases. *Eur J Biochem*, 1998. 258: 572–8.

[32] Bagley, K.A., Vangarderen, C.J., Chen, M., Duin, E.C., Albracht, S.P.J., Woodruff, W.H. Infrared studies on the interaction of carbon-monoxide with divalent nickel in hydrogenase from Chromatium vinosum. *Biochemistry*, 1994. 33: 9229–9236.

[33] Nicolet, Y., De Lacey, A.L., Vernede, X., Fernandez, V.M., Hatchikian, E.C., Fontecilla-Camps, J.C. Crystallographic and FTIR spectroscopic evidence of changes in Fe coordination upon reduction of the active site of the Fe-only hydrogenase from *Desulfovibrio desulfuricans*. *J Am Chem Soc*, 2001. 123: 1596–601.

[34] Mulder, D.W., Ortillo, D.O., Gardenghi, D.J., Naumov, A.V., Ruebush, S.S., Szilagyi, R.K., et al. Activation of $HydA^{\Delta EFG}$ requires a preformed [4Fe–4S] cluster. *Biochemistry*, 2009. 48: 6240–6248.

[35] Madden, C., Vaughn, M.D., Díez-Pérez, I., Brown, K.A., King, P.W., Gust, D., Moore, A.L., Moore, T.A. Catalytic Turnover of [FeFe]-Hydrogenase Based on Single-Molecule Imaging. *J. Am. Chem. Soc*, 2012. 134: 1577–1582.

[36] Ogata, H., Mizoguchi, Y., Mizuno, N., Miki, K., Adachi, S., Yasuoka, N., Yagi, T., Yamauchi, O., Hirota, S., Higuchi, Y. Structural Studies of the Carbon Monoxide Complex of [NiFe]-hydrogenase from *Desulfovibrio vulgaris* Miyazaki F: Suggestion for the Initial Activation Site for dihydrogen. *Journal of the American Chemical Society*, 2002. 124: 11628–11635.

[37] Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M., Yukawa, H. Enhanced hydrogen production from glucose using ldh- and frd-inactivated *Escherichia coli* strains. *Appl Environ Microbiol.*, 2006. 73: 67–72.

[38] De Vrije, T., Mars, A.E., Budde, M.A., Lai, M.H., Dijkema, C., De Waard, P., et al. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. *Appl Microbiol Biotechnol.*, 2007. 74: 1358–1367.

[39] Willquist, K., Zeidan, A.A., van Niel, E.W.J. Physiological characteristics of the extreme thermophile *Caldicellulosiruptor saccharolyticus*: an efficient hydrogen cell factory. *Microb Cell Fact.*, 2010. 9: 89.

[40] Levin, D.B., Pitt, L., Love, M. Biohydrogen production: prospects and limitations to practical application. *Int. J. Hydrogen Energy*, 2004. 29: 173–185.

[41] Mandal, B., Nath, K., Das, D. Improvement of biohydrogen production under decreased partial pressure of H_2 by *Enterobacter cloacae*. *Biotechnol. Lett.*, 2006. 28: 831–835.

[42] Wang, J., Wan, W. Factors influencing fermentative hydrogen production: a review. *International Journal of Hydrogen Energy*, 2009. 34: 799–811.

[43] Ogata, H., Lubitz, W., Higuchi, Y. [NiFe] hydrogenases: structural and spectroscopic studies of the reaction mechanism. *Dalton Trans.*, 2009. 37: 7577–7587.

[44] Nicolaou, S.A., Gaida, S.M., Papoutsakis, E.T. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation. *Metab Eng*, 2010. 12: 307–331.

[45] Jones, D.T., Woods, D.R. Acetone-butanol fermentation revisited. *Microbiol Rev.*, 1986. 50: 484–524.

[46] Van Niel, E.W.J., Claassen, P.A.M., Stams, A.J.M. Substrate and product inhibition of hydrogen production by the extreme thermophile, *Caldicellulosiruptor saccharolyticus*. *Biotechnol Bioeng.*, 2003. 81: 255–262.

[47] Mooney, R.A., Artsimovitch, I., Landick, R. Minireview: Information Processing by RNA Polymerase: Recognition of Regulatory Signals during RNA Chain elongation. *Journal of Bacteriology*, 1998. 180: 3265–3275

[48] Berg, J.M., Tymoczko, J.L., Stryer, L. Biochemistry (6^{th} ed.). *W. H. Freeman*, 2006, San Francisco. 850p.

[49] Raven, P., Johnson, G., Mason, K., Losos, J., Singer, S. Biology (9th ed.). *McGraw-Hill*, 2010, New York. 278–301.

[50] Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, 1992. 10: 413–417.

[51] Higuchi, R., Fockler, C., Dollinger, G., Watson, R. Kinetic PCR: Real time monitoring of DNA amplification reactions. *Biotechnology*, 1993. 11: 1026–1030.

[52] Alker, A.P., Mwapasa, V., Meshnick, S.R. Rapid real-time PCR genotyping of mutations associated with sulfadoxine-pyrimethamine resistance in *Plasmodium falciparum*. *Antimicrob. Agents Chemother*., 2004. 48: 2924–2929.

[53] Ward, C.L., Dempsey, M.H., Ring, C.J., Kempson, R.E., Zhang, L., Gor, D., Snowden, B.W., Tisdale, M. Design and performance testing of quantitative real time PCR assays for Influenza A and B viral load measurement. *J. Clin. Virol.*, 2004. 29: 179–188.

[54] Cheng, J., Zhang, Y., Li, Q. Real-time PCR genotyping using displacing probes. *Nucl. Acids Res*., 2004. 32: 54–61.

[55] Gibson, N.J. The use of real-time PCR methods in DNA sequence variation analysis. *Clin. Chim. Acta*., 2006. 363: 32–47.

[56] Bi´eche, I., Olivi, M., Champ´eme, M.H., Vidaud, D., Lidereau, R., Vidaud, M. Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer. *Int. J. Cancer*, 1998. 78: 661–666.

[57] Koenigshoff, M., Wilhelm, J., Bohle, R.M., Pingoud, A., Hahn, M. HER-2/neu gene copy number quantified by real-time PCR: Comparison of gene amplification heterozygosity, and immune-histochemical status in breast cancer tissue. *Clin. Chem*., 2003. 49: 219–229.

[58] Gibson, U.E., Heid, C.A., Williams, P.M. A novel method for real time quantitative RT-PCR. *Genome Res*., 1996. 6: 995–1001.

[59] Gallaghar, S.R., Wiley, E.A., Fraga, D., Meulia, T., Fenster, S. Real time PCR in Current Protocols Essential Laboratory Techniques. *John Wiley and* Sons, 2008, New Jersey. 800p.

[60] Kozera, B., & Rapacz, M. Reference genes in real-time PCR. *J Appl Genetics*, 2013. 54: 391–406.

[61] Gachon, C., Mingam, A., Charrier, B. Real-time PCR: what relevance to plant studies? *J Exp Bot*., 2004. 55: 1445–1454

[62] Bustin, S.A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol*., 2002. 29: 23–39.

[63] Huggett, J., Bustin, S. Standardisation and reporting for nucleic acid quantification. *Accred Qual Assur*., 2011. 16: 399–405

[64] Deepak, S., Kottapalli, K., Rakwal, R., et al. Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. *Curr. Genomics*, 2007. 8(4): 234–51.

[65] Burchill, S.A., Lewis, I.J., Selby, P. Improved methods using the reverse transcriptase polymerase chain reaction to detect tumor cells. *Br. J. Cancer*, 1999. 79: 971–977.

[66] Xi, L., Nicastri, D.G., El-Hefnawy, T., Hughes, S.J., Luketich, J.D., Godfrey, T.E. Optimal markers for real-time quantitative reverse transcription PCR detection of circulating tumor cells from melanoma, breast, colon, esophageal, head and neck, and lung cancers. *Clin. Chem.*, 2007. 53 (7): 1206–15.

[67] Tu, X., Das, K., Han, Q., Bauman, J.D., Clark, A.D., Hou, X., Frenkel, Y.V., Gaffney, B.L., Jones, R.A., Boyer, P.L., Hughes, S.H., Sarafianos, S.G., Arnold, E. Structural basis of HIV-1 resistance to AZT by excision. *Nat. Struct. Mol. Biol*., 2010. 17(10): 1202–1209.

[68] Pfaffl, M.W. Quantification strategies in real-time PCR. A-Z of quantitative PCR. *International University Line*, 2004, California. 87–112

[69] Sanders, R., Mason, D.J., Foy, C.A., Huggett, J.F. Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Anal Bioanal Chem.*, 2014. 406(26): 6471–6483.

[70] Luhtala, N., Parker, R. T2 Family ribonucleases: Ancient enzymes with diverse roles. *Trends in Biochemical Sciences*, 2010. 35(5): 253.

[71] Zhu, J., He, F., Hu, S., Yu, J. On the nature of human housekeeping genes. *Trends in genetics*, 2008. 24(10): 481–484.

[72] Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*., 2009. 55: 611–622.

[73] Huggett, J., Dheda, K., Bustin, S., Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun*., 2005. 6: 279–284

[74] Mallona, I., Lischewski, S., Weiss, J., Hause, B., Egea-Cortines, M. Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMCPlant Biol*., 2010. 10:4.

[75] Chervoneva, I., Li, Y., Schulz, S., Croker, S., Wilson, C., Waldman, S.A., Hyslop, T. Selection of optimal reference genes for normalization in quantitative RT-PCR. *BMC Bioinforma*., 2010. 11:253.

[76] Nicot, N., Hausman, J.F., Hoffmann, L., Evers, D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot*., 2005. 56(421): 2907–2914.

[77] Andersen, C.L., Jensen, J.L., Ørntoft, T.F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*., 2004. 64: 5245–5250.

[78] Banach-Orlowska, M., Fijalkowska, I.J., Schaaper, R.M., Jonczyk, P. DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in *Escherichia coli*. *Mol. Microbiol.*, 2005 58(1): 61–70.

[79] Olson, M.W., Dallmann, H.G., McHenry, C.S. DnaX complex of *Escherichia coli* DNA polymerase III holoenzyme. The chi psi complex functions by increasing the affinity of tau and gamma for delta.delta' to a physiologically relevant range. *J. Biol. Chem.*,1995. 270(49): 29570–29577.

[80] Huang, Y.P., Ito, J. DNA Polymerase C of the Thermophilic Bacterium *Thermus aquaticus*: Classification and Phylogenetic Analysis of the Family C DNA Polymerases. *J Mol Evol*., 1999. 48: 756–769.

[81] Chen, Z., Yang, H., Pavletich, N.P. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature*, 2008. 453: 489–494.

[82] Livak, K.J., Schmittgen, T.D., Analysis of Relative Gene Expression Data Using Real- Time Quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ Method. *Methods*, 2001. 25: 402–408.

[83] Winer, J., Jung, C.K., Shackel, I., Williams, P. M. Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Anal. Biochem*., 1999. 270, 41–49.

[84] Schmittgen, T.D., Zakrajsek, B.A., Mills, A.G., Gorn, V., Singer, M.J., Reed, M.W. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.*, 2000. 285: 194–204.

[85] Savir, Y. & Tlusty, T. RecA-mediated homology search as a nearly optimal signal detection system. *Molecular Cell*, 2010. 40(3): 388–96.

[86] De Vlaminck, I.,van Loenhout, M.T.J., Zweifel, L., den Blanken, J., Hooning, K., Hage, S., Kerssemakers, J., Dekker, C. Mechanism of Homology Recognition in DNA Recombination from Dual-Molecule Experiments. *Molecular Cell*, 2012. 46(5): 616–624.

[87] Wing, R.A., Bailey, S., Steitz, T.A. Insights into the replisome from the structure of a ternary complex of the DNA polymerase III -subunit. *J Mol Biol*, 2008. 382: 859–869.

[88] Lamers, M.H., & Mike O'Donnell. A consensus view of DNA binding by the C family of replicative DNA polymerases. Proc Natl Acad Sci., 2008. 105(52): 20565–20566.

[89] D'Urso, G., Grallert, B., Nurse, P. DNA polymerase alpha, a component of the replication initiation complex, is essential for the checkpoint coupling S phase to mitosis in fission yeast. *J. Cell Sci*., 1995. 108: 3109–3118.

[90] Navas, T.A., Zhou, Z., Elledge, S.J. DNA polymerase epsilon links the DNA replication machinery to the S phase checkpoint. *Cell*, 1995. 80: 29–39.

[91] Miguel, G.D., & Bebenek, K. Multiple functions of DNA polymerases. *CRC Crit Rev Plant Sci*., 2007. 26(2): 105–122.

[92] Bray, C.M., West, C.E. DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. *New Phytol*., 2005. 168: 511–528.